Chapter 2: Sampling Design for a Fish Community Survey

2.1 OBJECTIVES

Surveys of the fish community provide characterization and monitoring information relevant to sanctuary management, such as species diversity, relative abundance, size-distribution and spatial patterns. Surveys can take many forms ranging from complete censuses to samples of opportunity. A comprehensive census of the entire FGBNMS fish community cannot be accomplished because the fish community is too large. The next best option is to sample the fish community. If completed properly, samples provide necessary estimates of population parameters and uncertainty.

Sampling requires a significant investment of planning to ensure that limited fiscal and personnel resources are used effectively. Planning should resolve the location, timing, number of samples and measurements to be taken. Collectively, this information is known as the sampling design. The first and most important step of generating a sampling design is an explicit statement of objectives. Establishing sampling design objectives is an iterative process and should be revisited in concordance with updates to the sanctuary management plan where management needs are outlined. As this work was undertaken prior to the development of this plan, the following objectives were established in consultation with sanctuary management staff during a meeting in May 2006:

1) Determine long-term changes in fish community structure using metrics of diversity, density and trophic ratios;
2) Determine long-term changes in density and mean-size of selected economically important taxa;
3) Determine the relationship between physical measures such as habitat type, depth, slope and geographic location with the associated fish community using metrics of fish diversity, density and biomass; and
4) Find better ways to collect information such that the probability of detecting change over time or space is increased.

2.2 BASIC STRATEGY

There are many useful sampling designs, each with distinct advantages and disadvantages. Only probabilistic sampling designs are considered herein, because they are the only type to allow reliable inferences to be made from sample units to the sampled population and quantify uncertainty. Simple random sampling (SRS) is the simplest and most fundamental probability-based survey design. The SRS design considers all sample units equal (i.e., all sample units have the same probability of being selected) and thus is appropriate for situations where there is no spatial structure in the variance of investigated metrics or in situations where no prior knowledge exists regarding this structure.

The assumption of homogenous spatial structure in fish communities is rarely met. More often fish species show a strong association with benthic habitats, depths, salinity and other environmental covariates (Ault et al., 1999; Kendall et al., 2003) and thus are heterogeneous. Communities with a heterogeneous spatial distribution can be sampled more effectively if the population can be divided into internally homogeneous groups. This is the goal of stratified random sampling (StRS). A StRS design may divide the survey domain (study area to be sampled) into regions of relatively homogeneous variance called strata and by sampling more intensively in highly-variable strata, a StRS design can achieve more precise results than a SRS design using the same sample size. The ultimate effect is that the likelihood of detecting spatial and temporal changes in observed metrics is increased.

One method to divide the fish community into strata is to parse the population by environmental covariates of the fish community (see Section 2.6.2). This process requires maps of environmental variables at appropriate spatial scales. Benthic habitat maps are ideal because they can integrate multiple environmental variables and have a proven track record in fish sampling designs in both Florida (Ault et al., 2001) and the U.S. Virgin Islands (USVI; Menza et al., 2006).

Incorporating an understanding of costs associated with sampling, such as transportation, equipment rental and time can also improve the efficiency of sampling. For instance, sampling more intensively in areas with lower sampling costs can increase sample size while keeping costs the same. In the marine environment, sampling costs are strongly linked to depth, because sampling in deeper environments often requires special equipment and/or training, and more time (see Section 2.4). Consequently, one of the principal methods to decrease costs is to stratify by depth.

In this report, we recommend stratifying by both sampling cost and environmental covariates to produce high-precision estimates of population and community metrics at a minimum of cost. At the broadest spatial scale, dividing the sanctuary into areas representing the coral caps and areas representing the remaining deep habitats will drastically reduce costs. In addition, this division would allow the entire coral cap community to be sampled with scuba using the same underwater census methods (i.e., diver belt-transect), while a more expensive, alternative method can be used to survey the deeper habitats.

A further division of the coral caps into areas which can be sampled using conventional scuba and those that must use technical scuba is beneficial as well. This would ensure greater than 90% of the coral cap community can be sampled with relatively cost-effective conventional scuba equipment. In addition to reducing sampling cost, these divisions allow each area (strata) to be monitored independently, in case they cannot be sampled together. For instance, we recommend monitoring the shallow portions of the coral caps even if the deeper regions cannot be monitored due to monetary or time constraints. This would provide at least some data for long-term monitoring.

The division of the aforementioned three strata was completed using a fine-scale, half-meter resolution bathymetric model (source: sanctuary staff) for the East Flower Garden Bank (EB) and West Flower Garden Bank (WB). The EB and WB coral caps were divided using the 33.5 m (110 ft) isobath into the Shallow Coral Cap (SCC) and Deep Coral Cap (DCC) strata and the remaining area within sanctuary boundaries was designated as the Deep Non-Coral Cap (DNCC) stratum (Figure 2.1). These three mutually-exclusive areas exhaustively cover the entire fish community in the FGBNMS.

In this report we discuss the iterative process of developing a sampling design for the SCC (see Section 2.6). A separate report should discuss sampling the DCC and DNCC. If technical diving is an option, the process of selecting a sampling design recommended in this report may be useful for the DCC as well. The DNCC will require different technologies and methods and will likely be much more costly.

2.3 TARGET POPULATION

The target population is limited to the SCC, which includes the areas of East and West Banks readily surveyed using conventional scuba diving techniques (<33.5 m). It is also limited to those species of visible, diurnally-active fish typically associated with the reef and less adequately describes the more pelagic, small, cryptic and/or nocturnal species.

The SCC encompasses a total area of 1.09 km^2 and is approximately 0.5% of the entire FGBNMS area. Approximately two thirds of the SCC is part of the EB and the remaining third is part of the WB. The SCC component of the EB is further divided into two areas, with the majority of area in a contiguous southern section (Figure 2.1, insets 2 and 3). Although the SCC represents a small fraction of the FGBNMS, it is of great importance to sanctuary managers because of its distinct fish and coral communities and use by sanctuary visitors.

2.4 DATA TO BE COLLECTED (portions excerpted from Menza et al., 2006)

In order to meet the objectives stated in Section 2.1, collection of the following information relating to the fish community is essential: identification to the lowest possible taxonomic classification of each individual, abundance and size-frequency.

Concurrently collected physical data can be assimilated in a survey design to improve survey performance. It is important to collect information on parameters used in maps such as habitat type, depth, slope, and bank location in order that these factors can be analyzed to determine their respective roles in structuring the fish communities. As demonstrated in the Florida Keys and USVI, an accurate benthic habitat map can be effectively utilized to meet this need (Ault et al., 2001; Menza et al., 2006).

In addition, information describing the associated benthic habitat should be collected to further assist in the interpretation of the fish data. This complementary data should include: abiotic information (substrate type, habitat type), biotic information (corals, sponges, algae) and anthropogenic information (marine debris).

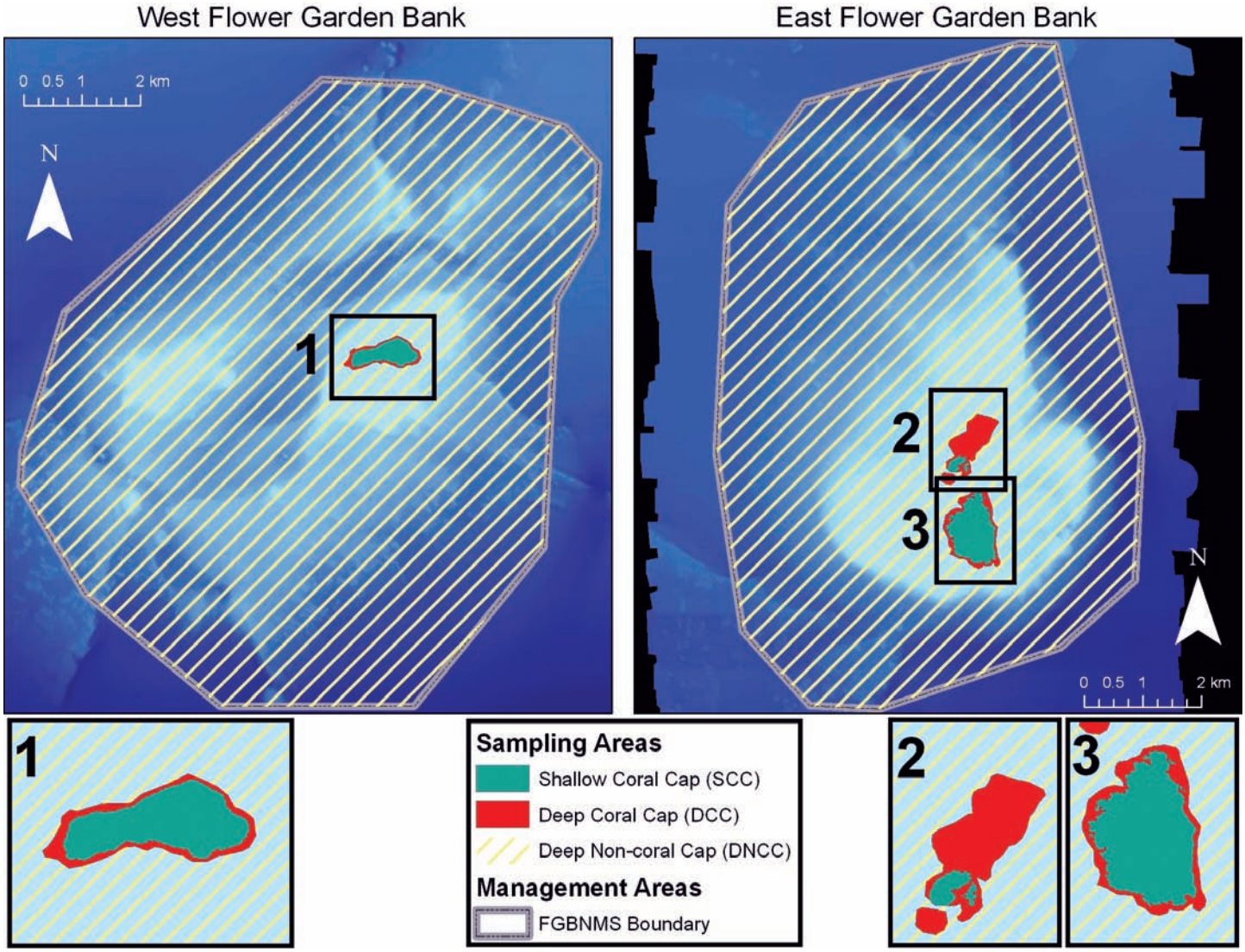

Figure 2.1. The East Flower Garden Bank (EB) and West Flower Garden Bank (WB) divided into three sampling areas. The insets, numbered 1 through 3, represent three geographic regions, not strata.

Detailed methods for collection of both fish and benthic composition data are available in Appendices A and B, respectively

2.5 SURVEY TECHNOLOGIES

As interest in ocean exploration, characterization, and monitoring advances, so must the methods and technologies employed for data collection. These technologies provide a range of various methods and approaches to data collection and monitoring. This section discusses applicable technologies used to sample and monitor coral reef communities along a transect using visual underwater census methods. Four factors should be considered when approaching each technology: 1) the cost of equipment and operation; 2) the quality and quantity of data collected; 3) the practicality of the method with regards to the overall goals of the mission or project; and 4) safety. The two technologies discussed below have been chosen because they satisfy the objectives listed in Section 2.1, are simple and cost effective, and have an extensive history of use. Additionally, they meet management concerns requiring the use of non-destructive assessment methods in the sanctuary.

2.5.1 Basic Scuba Diving

Scuba diving is one of the most common methods used for data collection, sampling, monitoring, and studying coral reef communities. Two common gas mixtures used for basic diving are a standard air mixture and an oxygen enriched air mixture. Standard air (normal atmospheric air) is comprised of approximately 21% oxygen and 78% nitrogen and 1% other; while Enriched Air (or “nitrox”) generally contains between 32-36% oxygen.

Advantages

Using scuba to conduct survey work has many operational advantages. *In situ* observations cannot be replaced by video or still photographic images, because a diver has the ability to visualize the survey area multi-dimensionally, identify distance and lengths, and investigate holes or crevices. Also, the preponderance of studies which have used basic scuba allows standardized comparisons of results. Another advantage is the relatively simple nature of basic scuba operations which requires minimal technical or operational support, easy gear assembly and breakdown, and widely available training.

The technical advantages of diving with nitrox over standard air are numerous, particularly when conducting multiple dives over more than one day (Mastro and Dinsmore, 1989). Compared to diving with air, using nitrox increases bottom-time and shortens surface intervals (PADI, 2003). Nitrox also has an advantage over air by reducing decompression requirements and the occurrence of decompression sickness (DCS, or the "bends"). By replacing nitrogen with oxygen in the gas mixture, nitrogen build-up in the body is reduced due to the increased oxygen in the breathing mixture (Wells, 1989).

Disadvantages

There are both technical and operational disadvantages to using scuba as well. Diving to depths of 35 m and greater or for extended times using air or Nitrox is achievable, but not practical or recommended. Most organizations and educational programs have a diving limit of 33-37 m without advanced or technical certifications. Many of these restrictions are based on health risks that can be potentially serious, such as DCS, arterial gas embolism and oxygen toxicity.

Operationally, the disadvantages of conducting visual surveys include inter-diver variability in data collection and the effect of human activity on the organisms surveyed. For fish estimates, data variability can include: under or over estimation of size, quantity and distance (Edgar et al., 2004). Much of this variability can be explained by diver experience (Harvey et al., 2004). Also, the presence of human activity can create audible and physical disturbances that can serve to attract or deter potential species of interest (Lobel, 2001; Harvey et al., 2004).

2.5.2 Technical Scuba Diving

Technical diving exceeds the scope and limits of recreational diving requiring additional training, equipment and extensive experience. Many organizations and programs provide training for technical diving. Technical diving can include the use of common gas mixtures such as "trimix" and "heliox" gases, as well as closed-circuit systems ("rebreathers"). The operational advantages and disadvantages of survey data collection using these technical diving methods are the same as those mentioned in Section 2.5.1; however, there are additional considerations which are outline below.

Advantages

The advantages of diving on mixed gases are extended bottom-times and greater maximum operating depths. This is made possible by the multiple gases utilized and the gas ratio combinations both of which can be adjusted based on depth and dive time to meet the requirements of the sampling design. The advantages of rebreathers include a longer bottom-time, shorter topside operations, cheaper costs and fewer imposed disturbances on habitats and organisms (the latter results in a more accurate representation of species composition and abundance). A comparison study of open versus closed-circuit diving conducted by Parrish and Pyle (2002) demonstrated that divers using open-circuit equipment required more gas, more preparation time and spent more on consumable gases (i.e., oxygen, air, heliox premix).

Disadvantages

One of the disadvantages for all three of these types of technical diving is the additional equipment and training required for the use of the equipment. Rebreathers and trimix/heliox require extra equipment maintenance and calibration, and can be costly in the short-term. The same medical risks involved with standard and enriched air diving pertain to these three technical diving operations too. However, working with technical equipment and diving deeper than recreational depths also increases decompression times and the potential for accidents to occur. Additionally, since divers must attain additional levels of certification and diving platforms must have the equipment onboard to support closed-circuit system diving, these requirements may impose further restrictions on the proposed sampling design.

2.6 SAMPLING DESIGN

A common first step for characterizing and monitoring any natural resource, including a fish community, is to initially select an uncomplicated sampling design (e.g., simple random) given the current understanding of the community. Then as data are gathered, more efficient sampling designs can be compared to the original sampling design, and adopted if proven effective. This iterative process ensures collected data are used to produce efficient sampling designs as data properties are better understood.

To ensure fish community data are collected efficiently, six distinct sampling designs were identified and then compared based on estimated precision of population and community metrics. Analyses were undertaken using fish data collected in 2006. These data were collected by CCMA and ONMS personnel, using a stratified random design (measurement methods are described in Appendices A and B). Four strata were used to parse sample sites (see Appendix C). Strata were composed of areas with differing bathymetric slope (i.e., steep versus flat) and geographic location (i.e., EB versus WB; Figure 2.2). A total of 73 samples were collected using randomly positioned belt-transects within each stratum. Each transect was allocated among strata proportional to area, with a minimum of five samples allocated to each to ensure an adequate sample size for computing precise strata metrics (i.e., reducing standard error).

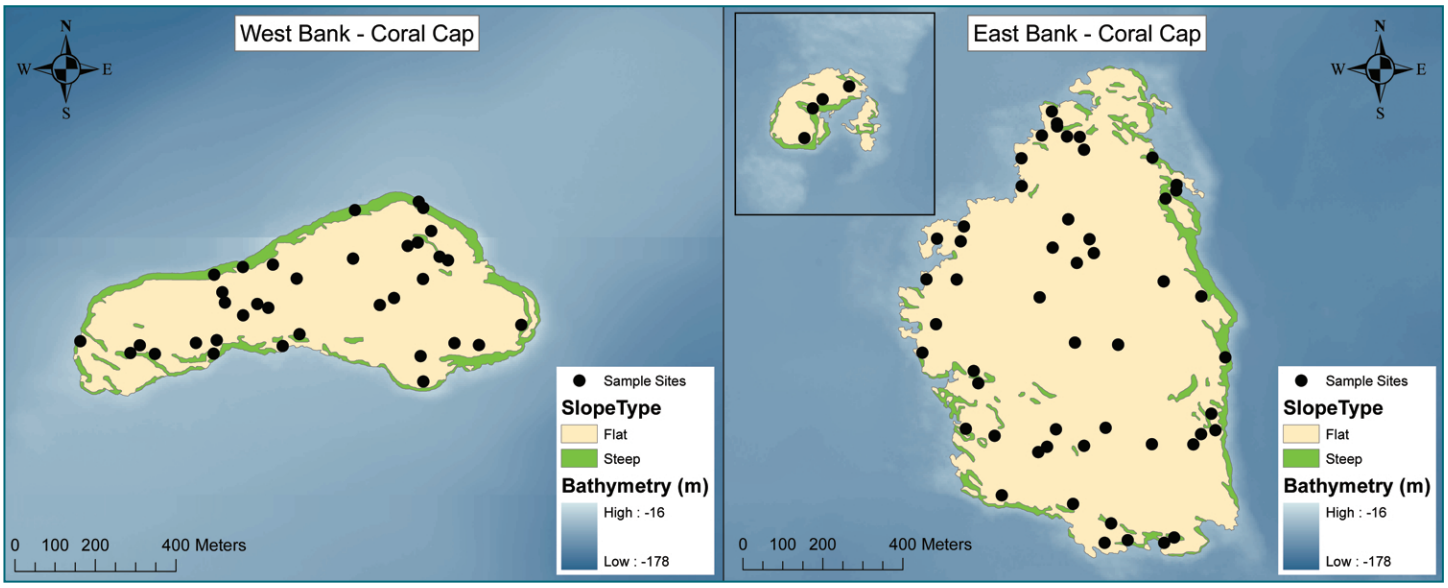

Figure 2.2. Location of samples and strata from the 2006 field mission. At each sample location fish and benthic habitat data were collected.

Results from candidate sampling design analysis (Section 2.6.2) were used to choose a sampling design for the 2007 field mission. It was anticipated that the data gathered in 2007 would be used in a second round of candidate design analysis, but field operations were cancelled prematurely and provided insufficient data to complete this task.

The following sections describe the process and outcome of design analysis using 2006 data.

2.6.1 Candidate Design Identification

Six candidate sampling designs were identified to compare design performance. Four distinct stratified sampling designs were produced using different combinations of geographic location, depth and benthic habitat to parse the survey domain. The fifth design was the sampling design used in 2006 (StRS-2006) and the sixth was a simple random sampling design (SRS) used to assess design efficiency without stratification.

To ensure the designs would satisfy all sanctuary objectives (Section 2.1), an assortment of population and community metrics (e.g., species density, species composition) and fish assemblages (e.g., all fish species, herbivores and specific species) were used to evaluate design performance (Table 2.1.). The combinations of metrics and fish assemblages used in the evaluation process are hereafter referred to as indices.

Number of species, herbivore:piscivore (H:P) ratios, density and biomass of all species combined are community-level metrics and were chosen to reflect community regime shifts associated with major environmental perturbations or fishing pressure. Number of species is simply the number of distinct species observed at a site. The H:P ratios are determined from the densities and biomass of each trophic group. Assignment to a particular trophic group was determined using FishBase (Froese and Pauley, 2007). Total density is simply the sum of all individuals of each species observed within a sample unit. Biomass is the sum of biomass from all fish species. Details of the computations are provided by Menza et al. (2006). Density, biomass and average size of groupers are assemblage-level metrics and were chosen to reflect changes in fishing pressure. The grouper assemblage was a subset of species in the Serranidae family, consisting of all species in the *Mycteroperca, Cephalopholis, Epinephelus* and *Dermatolepis* genera. Average size was computed from the midpoint of size-class data collected in the field. Density was computed as defined previously, but constrained to the subset of grouper species defined above. The remaining indices (density of grey snapper [*Lutjanus griseus*], yellowmouth grouper [*Mycteroperca interstitialis*], tiger grouper [*Mycteroperca tigris*] and marbled grouper [*Dermatolepis inermis*]) are species-level metrics and were chosen to provide data on key taxa of interest to sanctuary managers. Again, density was computed as defined previously, but only for individuals of each corresponding species.

Table 2.1. Indices used in the sampling design evaluation process.

Index
Number of species
Density (No. / 100m^2)
All species
Groupers
Herbivore:Piscivore Ratio
Grey snapper (*Lutjanus griseus*)
Yellowmouth grouper (*Mycteroperca interstitialis*)
Tiger grouper (*Mycteroperca tigris*)
Marbled grouper (*Dermatolepis inermis*)
Average Size (cm)
Groupers
Biomass (g / 100m^2)
All species
Groupers
Herbivore:Piscivore Ratio

Two of the potential environmental covariates investigated for stratified designs, slope and depth, were continuous variables. To include these variables in analyses they were first grouped into discrete categories. Since stratified sampling designs are most effective when the measurements of interest (i.e., indices) are divided into internally homogenous groups, an analysis of variance was undertaken to identify suitable breakpoints by which to categorize these two continuous variables. The analysis of variance was accomplished using recursive partitioning (function Partition: JMP© by SAS Institute Inc., 2000). This process recursively divided each index into groups such that the ratio of variance within groups to among groups was minimized. A non-parametric analysis of variance test (Kruskal-Wallis ANOVA) was used to indicate which groupings were significantly different (Kruskal-Wallis; $p<0.10$; Table 2.2). A Type I error probability of 0.10 was used instead of 0.05, because this was an exploratory analysis and Type I error was not a critical concern. The results suggest a division by slope at 37° and by depth at 30 or 32 m provide suitable breakpoints for indices. To ensure a parsimonious stratification scheme the depth breakpoints were averaged to 31 m, thus one stratum was composed of all areas deeper than 31 m and another for areas shallower than 31 m.

Table 2.2. Breakpoints in slope and depth determined from analysis of variance of nine fish indices. Breakpoints with an asterisk denote a significant difference using ANOVA.

Index	Slope (°)	Depth (m)
Number of species	42	32
Density (No. / 100m^2)		
All species	39	32
Groupers	17	32
Herbivore:Piscivore Ratio	18	32*
Grey snapper (*L. griseus*)	37*	32
Yellowmouth grouper (*M. interstitialis*)	42	27
Tiger grouper (*M. tigris*)	16	32
Marbled grouper (*D. inermis*)	4	32
Average Size (cm)		
Groupers	42	30*
Biomass (g / 100m^2)		
All species	14	22
Groupers	14	22
Herbivore:Piscivore Ratio	31	32*

Benthic habitat strata were taken from a benthic habitat map of the coral caps (Appendix C, Map 2). Two distinct benthic habitats were identified: low-relief coral (mixture of *Madracis* and rubble) and high-relief coral (mixture of boulder and plate corals). Geographic location was simply used to divide the shallow coral cap community into an EB and a WB category.

Kruskal-Wallis ANOVA was also used to indicate if covariate groupings were significantly different (Kruskal-Wallis; $p<0.10$; Table 2.3). The decision of which variables to use in generating strata for the candidate sampling designs balanced potential increases in efficiency (as shown by ANOVA results), logistical value in the field, likely consistency over time and parsimony. For these reasons, geographic location, depth and benthic habitat were selected for stratification schemes and slope was not.

Candidate sampling designs and variables used to delineate strata are listed in Table 2.4. All stratification schemes for candidate sampling designs, except for the SRS used geographic location to divide banks, because geographic location simplifies field logistics. The composite design (STRS-Composite) divided high-relief coral habitats into two depth categories (shallow and deep) and a single low-relief coral habitat category. Low-relief coral habitat was not divided by depth, because of the small sampling area involved in shallow areas. A simple t-test was used to show deep and shallow low-relief coral habitat strata were not different (t-test; $p>0.10$).

Table 2.3. Results from non-parametric analysis of variance of nine fish indices among groups (ANOVA levels) derived from geographic location, depth, slope and benthic habitat type. Values represent the probability that data from each group come from the same population (p-value). Asterisks denote probabilities lower than 0.10.

Index	Geographic Location	Benthic Habitat	Slope	Depth
Number of species	0.12	0.04*	0.60	0.25
Density (No. / 100m^2)				
All species	0.47	0.54	0.36	0.38
Groupers	0.13	0.07*	0.32	0.60
Herbivore:Piscivore Ratio	0.49	0.31	0.24	0.04*
Grey snapper	0.03*	0.24	0.07*	0.47
Yellowmouth grouper	0.48	0.10*	0.77	0.49
Tiger grouper	0.86	0.57	0.16	0.86
Marbled grouper	0.81	0.80	0.84	0.39
Average Size (cm)				
Groupers	0.25	0.03*	0.12	0.01*
Biomass (g / 100m^2)				
All species	0.51	0.66	0.11	0.81
Groupers	0.08*	0.06*	0.52	0.64
Herbivore:Piscivore Ratio	0.84	0.18	0.26	0.09*

Table 2.4. List and description of identified sampling designs.

Sampling Design	Variables used to delineate strata
SRS	None
StRS-Bank	Geographic location
StRS-2006	Geographic location, slope
StRs-Depth	Geographic location, depth
StRS-Benthic Habitat	Geographic location, benthic habitat
StRS-Composite (used in 2007)	Geographic location, benthic habitat, depth

2.6.2 Candidate Sampling Design Comparison

Design performance was evaluated using the sample size required to obtain a coefficient of variation (CV) of 10% for each index. For these comparisons the precise value of the CV was not important; rather it was important to keep the CV constant to apply a standardized statistic in comparisons. Cochran (1977) describes the process of post-stratification and corresponding computations for both SRS and StRS designs. Post-stratification analysis is required because the domains used to parse the sampling frame are not the same as the strata used for obtaining data (except for StRS-2006). A limitation of post-stratification analysis is that estimates of variance for any given fish metric are not technically valid for data collected under a different stratification scheme (except for SRS). Thus the results of this section are for comparative analysis only and are not a re-estimation of variance.

Sample size requirements were computed for a specified CV, because unlike other performance measures such as a confidence interval, the computation of a CV does not make any assumptions concerning data structure (e.g., Normal distribution). If needed, CVs of a given size can be explained in terms of confidence intervals of a given size because they are different by a common factor. Sample size requirements were determined using the methods described in Cochran (1977) for a StRS design. Under a presumed optimal allocation scheme sample size is given by

$$n^* = \frac{\left(\sum_{j=1}^{J} W_j s_j\right)^2}{\left(CV\left[\overline{X}_{st}\right]\overline{X}_{st}\right)^2 + \frac{1}{\sum_{j=1}^{J} N_j}\sum_{j=1}^{J} W_j s_j^2}$$

where s_j is the stratum standard deviation, $CV[\bar{X}_{st}]$ is the desired coefficient of variation (i.e., 0.10) and $\bar{X}_{st}$ is the survey-wide mean. A fundamental requirement for the computation is accurate stratum weighting factors. These are computed using

$$W_j = \frac{N_j M_j}{\sum_{j=1}^{J} N_j M_j} \approx \frac{A_j}{\sum_{j=1}^{J} A_j}$$

where N_j is the number of sample units in stratum j, M_j is the number of transects which can fit in each sample unit, A_j is the area of a transect in stratum j, and J is the number of strata. Estimates for the mean and variance within a particular stratum were computed from transects within a given stratum. Stratum designations were determined using the intersection of the sample design map and transect coordinates in a GIS (ESRI, 2006).

Sample size requirements of each tested design are provided in Table 2.5. Results suggest StRS-Composite is the optimal choice for sampling the SCC. Not only was StRS-Composite the best design for the most number of indices, but StRS-Composite possessed the lowest sum of sample size requirements from all indices and had the lowest maximum.

Table 2.5. The sample size requirements needed to obtain a CV of 10% for nine reef fish indices using six distinct sampling designs. Numbers in bold represent the minimum sample size requirement for a given index.

Community Index	SRS	StRS 2006	StRS Bank	StRS Depth	StRS Benthic Habitat	StRS Composite
Number of species	2	2	2	2	2	**2**
Density (No. / 100m^2)						
All species	37	32	**30**	37	36	34
Groupers	74	76	74	72	76	**70**
Herbivore:Piscivore Ratio	121	123	119	**90**	127	109
Grey snapper	730	710	764	727	825	**574**
Yellowmouth grouper	158	167	157	167	155	**143**
Tiger grouper	234	239	242	252	239	**216**
Marbled grouper	676	741	704	861	**703**	731
Average Size (cm)						
Groupers	36	36	**35**	54	48	45
Biomass (g / 100m^2)						
All species	186	456	146	130	446	**129**
Groupers	842	**145**	455	455	1465	425
Herbivore:Piscivore Ratio	817	550	588	**237**	913	530
Totals	3913	3277	3316	3084	5035	3008
Maximums	842	741	764	861	1465	731
Averages	356	298	301	280	458	273

StRS-Composite proved superior to alternative candidate designs, because it ultimately could satisfy the multiple and diverse objectives selected by sanctuary managers (Section 2.1) at a minimum of cost. Although StRS-Composite was optimal, the large sample sizes required for some metrics (e.g., marbled grouper, grey snapper) are greater than what typically can be afforded during most sampling missions. Consequently, managers and researchers have three options. They can relax precision requirements (i.e., CV>10%), or use a different, less variable proxy (e.g., presence-absence) or continue to look for more cost effective sampling designs. These options are discussed further in Section 2.6.4.

A principal reason why StRS-Composite was optimal is that most groupers and snappers were sighted along the margins of East and West Banks (see Section 3). These areas are dissimilar from the remaining SCC, because they are characterized by benthic habitat transition zones.

Data gathered in 2007 was to be used in a similar analysis of the spatial relationships among fish indices and environmental covariates, but due to low sample size and lack of data for most strata, only a cursory analysis was possible. A total of 70 surveys were scheduled, but the field mission was cut short by severe weather associated with Hurricane Humberto and only 32 surveys were taken (Figure 2.3). No samples were taken from the WB and samples collected on the EB were biased towards the southern areas of the coral cap.

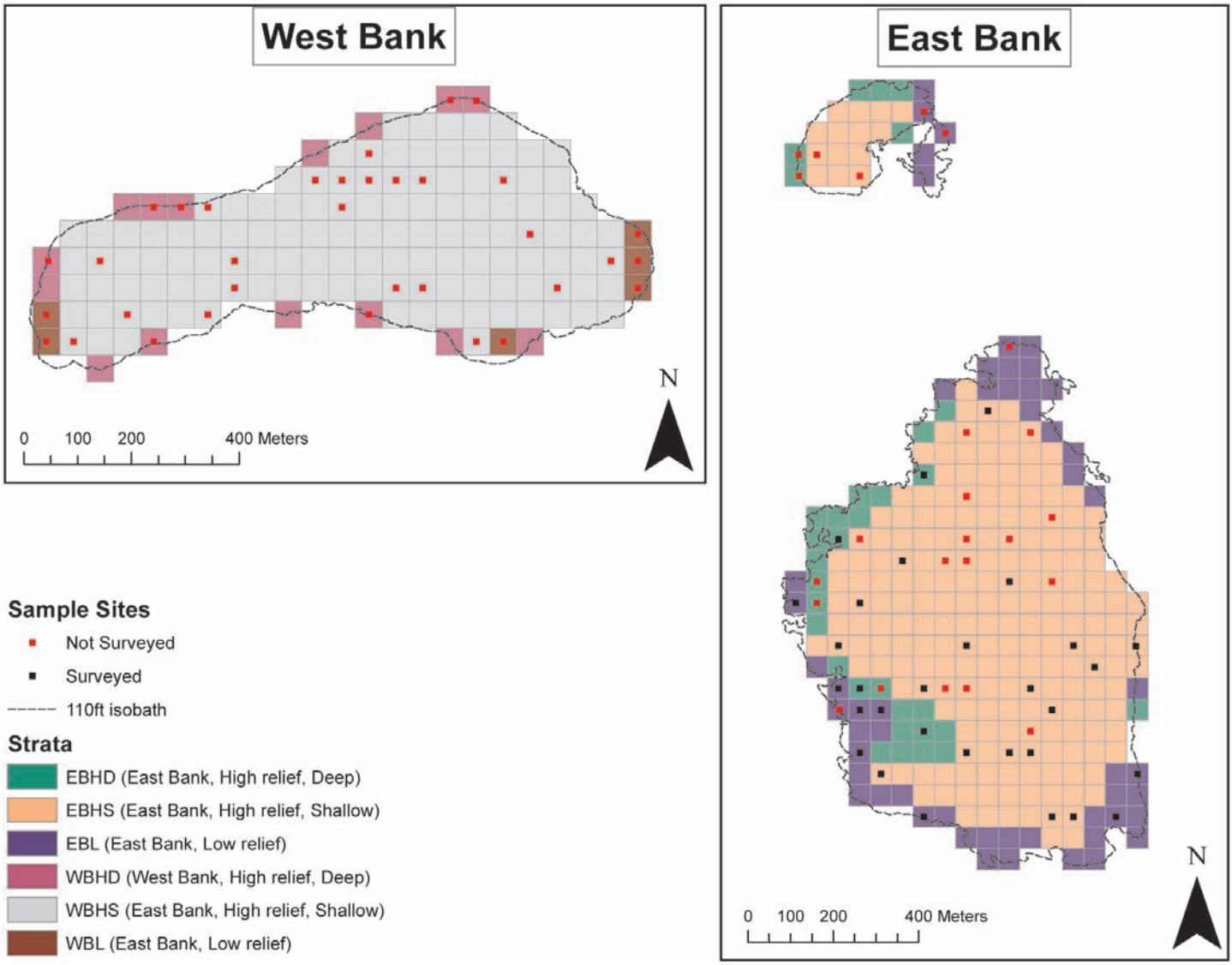

Figure 2.3. Location of samples and strata from the 2007 field mission.

The sampling design used in 2007 (see Appendix C) was different than the one used in 2006 in two respects. First, a different set of strata were used. The strata used in 2007 were based on the optimal candidate design from analysis of 2006 data. Second, the method used to position sample sites was different. In 2006, sites were selected by randomly placing geographic coordinates within a polygon representing a given strata. In 2007 a uniform distribution of points separated by 50 m was overlaid on the coral caps and a random selection of these points was taken. A separation of 50 m was used to ensure 25 m transects in neighboring sample units would not overlap. A detailed comparison of these methods and the rationale behind using the latter are described in Menza et al. (2007). The approach used in 2007 has the advantage of incorporating a habitat's sphere of influence, ensuring sample units are exhaustive and mutually-exclusive, and providing a means to update the design more easily. This sample frame is easily adapted to new stratification schemes. For example, should a new marine protected area (MPA) be established on one of the banks, a new objective can be devised to assess its effectiveness. All sample units (i.e., points) within the spatial limits of the MPA could be incorporated into new strata (or a new stratum), while all the remaining sampling units retain their original stratum designations. Then community and population metrics computed from strata with and without the MPA can be compared in order to assess the new management regime's effectiveness.

Multiple discriminant analysis (MDA) was used to assess index differences among strata using data collected in 2007. Only strata on the east bank were investigated due to the absence of data from the west bank. The results show a moderate separation between low-relief habitat and high-relief habitats, but only a marginal division between shallow and deep high-relief habitats (Figure 2.4). Based on these results the current strata should be maintained, but as new data are gathered the strata should be reassessed, especially with respect to differences among shallow and deep high-relief habitats. If future data do not show clear differences among shallow and deep habitats, aggregation is warranted.

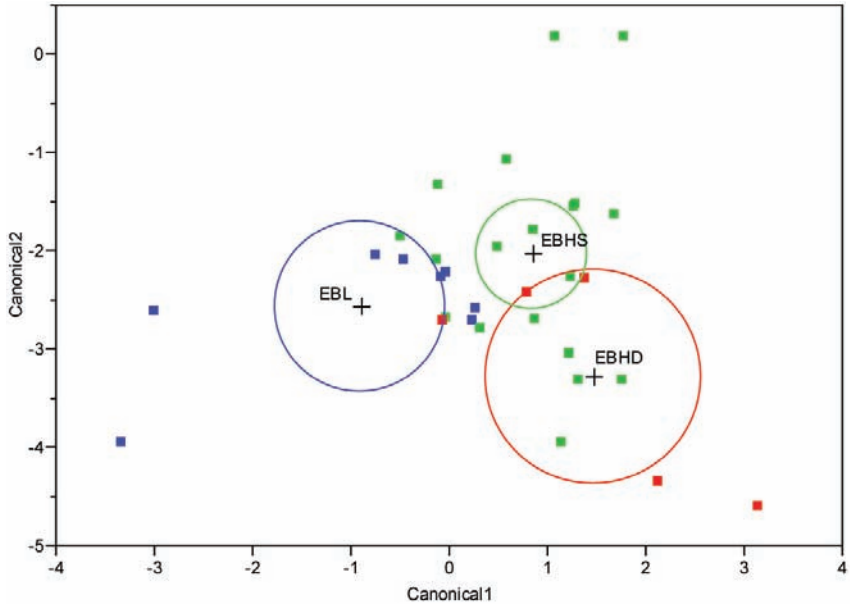

Figure 2.4. A canonical plot of multivariate discriminant analysis. Strata designations are EBL (Blue)– East Bank low-relief, EBHS (Green)– East Bank high-relief shallow, and EBHD– East Bank high-relief deep. Circles correspond to 95% confidence limits of multivariate means.

2.6.3 Example of Implementation

In this section we provide an example of implementing StRS-Composite to obtain a sample of survey sites. The generated sample coordinates can be used for a future mission. It is important to rerun the randomization process once the sample coordinates presented here are used to eliminate sampling bias.

In this example the sample frame constructed for the 2007 field mission was used for sample site selection. Ideally desired levels of sampling precision and statistical power are used to set the most efficient sample size for a sampling design. Unfortunately, this process is not feasible in many circumstances, because many important metrics require samples size which cannot be afforded. Consequently, sample size is set by fiscal and logistical constraints. Total sample size is generally between 50 and 100 and is strongly related to the number of divers and time available for diving. In 2006 CCMA and sanctuary staff were able to obtain 73 samples in six days using five buddy pairs. On average each diver conducted three dives per day. These numbers can be used to infer an estimate of sample size given personnel, equipment and logistical constraints for future missions. In this example, we used a sample size of 70.

Once a total sample size is determined it must be allocated among strata. The simplest allocation scheme and the one recommended here is sample size proportional to area. In this method, samples are allocated among strata proportional to their size (Table 2.6). Thus, large strata (e.g., East Bank high relief-Shallow [EBHS]) receive more samples than small strata (e.g., East Bank high relief- Deep [EBHD]). The Neyman allocation scheme is an alternative which can increase the precision of population and community estimates if accurate strata variance estimates are known. The latter is not recommended here, because the lack of data increases uncertainty when identifying highly variable strata for multiple indices.

Table 2.6. Sample allocation of 70 sample sites.

Strata	Area (m^2)	Weights	Proportional Allocation	Adjusted Allocation
EBHD	15,303	0.01	1	6
EBHS	579,115	0.12	37	26
EBL	128,826	0.53	8	10
WBHD	38,155	0.03	2	6
WBHS	322,074	0.30	21	17
WBL	7,503	0.01	0	5
Totals	1,090,976	1.00	70	70

One problem with the allocation scheme is that not all strata garner adequate samples to obtain an estimate for variance (e.g., WBL in Table 2.6). An adjusted form of the allocation scheme, which reduces the total n by $H*X$, where H is the number of strata, and then adds X samples to each strata after allocation, is used to ensure an adequate sample size within each stratum (Table 2.6). At a minimum X must be 2 to obtain an estimate of variance, but an X of 5 will reduce the standard error in small strata. The adjusted stratum sample size in Table 2.6 uses an X of 5.

The final step before field operations can take place is the random selection of samples among strata based on the allocation identified in the previous step. A tool was developed by CCMA to help in this capacity (see http://ccma.nos.noaa.gov/ecosystems/sanctuaries/fgb_nms.html). An alternative method is to use a program with a random selection procedure such as SAS or an equivalent statistical package. It is important to include all sample units within a given stratum when a random selection is being made. This ensures a selection bias is not incorporated into the sampling design. Figure 2.5 shows the location of survey sites of a stratified random sample which can be used in a future survey (see Appendix D for sample coordinates).

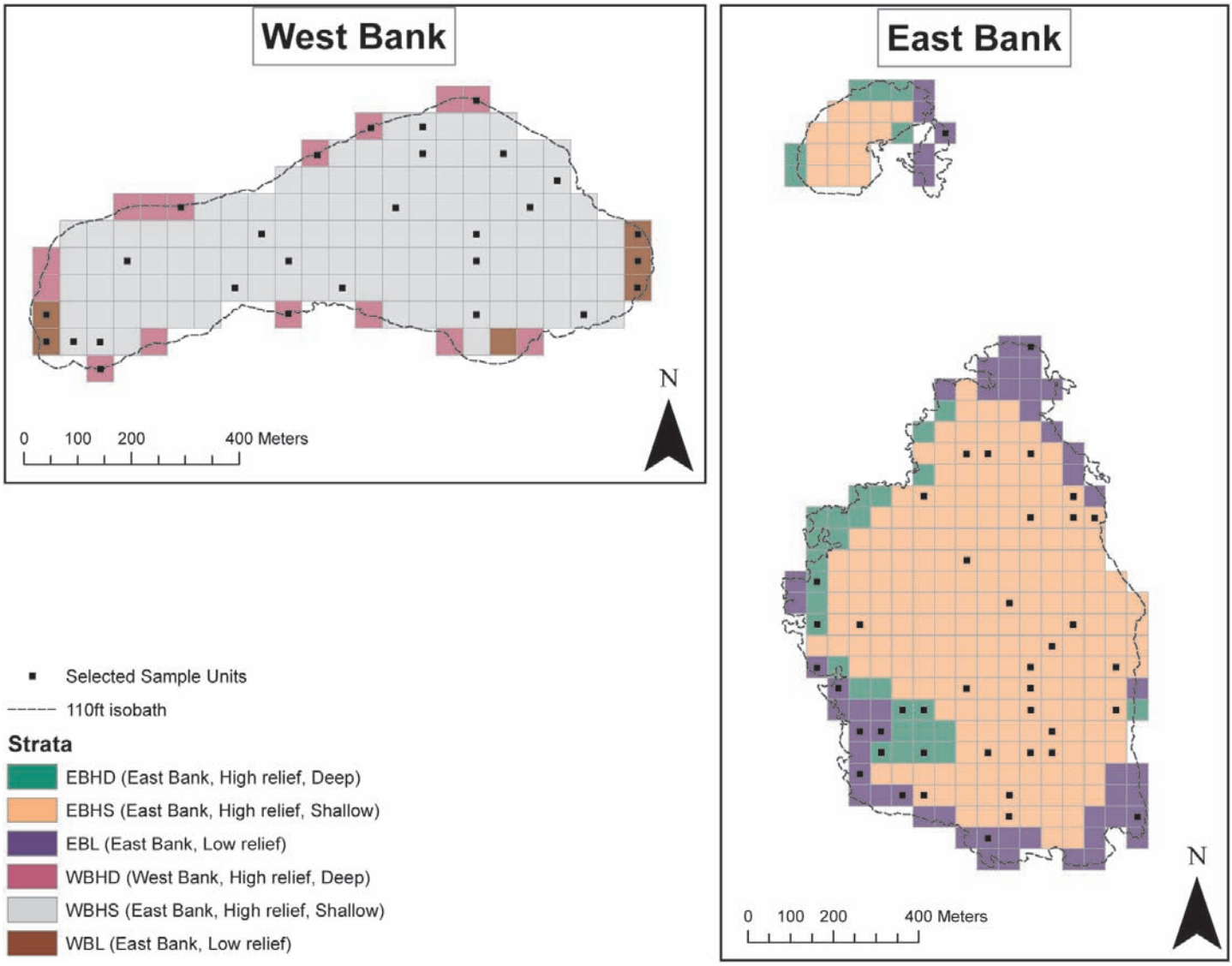

Figure 2.5. Sample example selected from strata of the shallow coral cap.

Once survey data is collected it must be stored and analyzed. Information on data storage and distribution is detailed in Menza et al (2006). Data analysis typically focuses on a set of simple descriptive statistics computed to assess populations and communities and identify temporal changes. It is important to keep in mind that when statistics are computed using multiple strata they must be weighted appropriately. Several basic sampling references (e.g., Cochran, 1977; Lohr, 1999) provide information on how to calculate sampling weights and use them to define descriptive statistics. Menza et al. (2006) define a set of descriptive statistics commonly used to assess reef fish communities and populations, and identify appropriate computations.

2.6.4 Future Direction

Although StRS-Composite was chosen as the optimal design, sample size requirements varied over two orders of magnitude depending on the reef fish index in question. Design selection must be tempered with realistic projections of maximum sample size. Given likely logistical constraints a maximum sample size will lie between 50 and 100. All community-level and assemblage-level indices will likely be sampled with a CV<10% when n≤100, but all species-level indices require n>150 and some require n>600 to ensure CV<10%. These large

sample sizes are effectively impossible given likely logistical and monetary constraints of sampling on the SCC. To provide suitable data, managers may relax precision requirements (i.e., CV>10%), use a different less variable metric (e.g., presence-absence), or continue to look for more cost effective sampling designs.

The reason for high CVs and large sample size requirements is the heavily skewed distribution of data with many zeros. Zeros occur where a species was not observed within a sample and are common for rare species. More abundant species typically have fewer zeros, lower CVs and consequently lower sample size requirements (e.g., Table 2.7). The probability of detecting an item can be increased by using either a different sampling design or survey method.

Table 2.7. Coefficient of variation (CV) and sample size requirements (n) to obtain a CV=10% for density of six species of reef fish. Data from 2006 sampling mission.

Fish Index	% Occurrence	CV	n
Density – Bluehead wrasse (*Thalassoma bifasciatum*)	100	0.08	44
Density – Spanish hogfish (*Bodianus rufus*)	97	0.08	47
Density – Redband parrotfish (*Sparisoma aurofrenatum*)	89	0.09	64
Density – Yellowhead wrasse (*Halichoeres garnoti*)	85	0.12	94
Density – Queen parrotfish (*Scarus vetula*)	64	0.16	169
Density – Marbled grouper (*Dermatolepis inermis*)	14	0.31	607

Most researchers agree that a stratified design is necessary for sampling rare species (Thompson, 2002), but an alternative to the candidate designs investigated in this first stage of analysis may help ensure rare species are sampled better. One alternative is to use adaptive sampling, but the approach has acknowledged difficulties in large areas like the SCC and modifying field logistics on the fly can be difficult (Thompson and Seber, 1996). Alternatively, as more data becomes available better relationships among fish indices and environmental variables may emerge. Incorporating these relationships into a stratified design will increase design performance.

Alternate survey methods may also be employed to generate more precise survey estimates with less zeros. The most practical modification is a survey targeting rare species. This may increase sample size, allow larger areas to be surveyed, and increase the probability of detection of groupers and snappers. A two-part survey, one targeting the entire community (such as described in this study) and another survey targeting large species, is an option. These two phases could be completed simultaneously or one immediately after the other during a single dive. In addition, a larger sample plot may increase the probability of detection of rare species and decrease CVs; however an increase in plot size must be balanced with the time needed to survey the plot. If the increase in bottom time and required surface intervals decrease total sample size, the positive effect may be negated. Finally, a survey method which does not use divers (e.g., ROV, submersible, bait camera) could be used, but the sampling costs associated with these methods will likely be higher.

2.7 SUMMARY AND RECOMMENDATIONS

- Use a stratified-random sampling design. Results indicate depth and benthic habitat are covariates of several fish indices and can be used to effectively sample the coral caps.
- The development of a sampling design is an iterative process. As new data are gathered the strata should be reassessed, especially with respect to differences among shallow and deep high-relief habitats. If future data do not show clear differences among shallow and deep habitats, aggregation is warranted.
- A sample size between 60 sites and 100 sites is adequate to survey community-level and assemblage-level indices, but not species-level indices (e.g., marbled grouper [*Dermatolepis inermis*] density).
- To adequately survey species level metrics consider a different approach: both sampling design and survey method. Potential changes are suggested in Section 2.6.4
- Continue to acquire samples throughout the SCC and among all habitats. These data will allow a more refined analysis of fish-habitat relationships; improve the stratification scheme; and quantify changes associated with natural population variability, changes in management strategy, or environmental and anthropogenic impacts.
- Survey the remainder of the sanctuary not included in the SCC to provide comprehensive population estimates for the sanctuary and to identify linkages.

2.8 REFERENCES

Ault, J.S., G.A. Diaz, S.G. Smith, J. Luo, and J.E. Serafy. 1999. An efficient sampling survey design to estimate pink shrimp population abundance in Biscayne Bay, Florida. North Am. J. Fish. Manage. 19(3): 696-712.

Ault, J.S., S.G. Smith, G.A. Meester, J. Luo, and J.A. Bohnsack. 2001. Site characterization for Biscayne National Park: assessment of fisheries resources and habitats. NOAA Technical Memorandum NMFS SEFSC 468. Miami, FL. 185 pp.

Cochran, W.G. 1977. Sampling Techniques, 3rd edition. John Wiley & Sons, Inc., New York. 428 pp.

Edgar, G.J., N.S. Barrett, and A.L. Morton. 2004. Biases associated with the use of underwater visual census techniques to quantify the density and size-structure of fish populations. J. Exp. Mar. Biol. Ecol. 308: 269-290.

Environmental Systems Research Institute, Inc (ESRI). 2006. ArcGIS 9.0. Redlands, CA.

Froese, R. and D. Pauly (eds.). 2007. FishBase. World Wide Web electronic publication. www.fishbase.org, version (10/2007).

Harvey, E., D. Fletcher, M.R. Shortis, and G.A. Kendrick. 2004. A comparison of underwater visual distance estimates made by scuba divers and a stereo-video system: implications for underwater visual census of reef fish abundance. Mar. Freshw. Res. 55: 573-580.

Kendall, M.S., J.D. Christensen, and Z. Hillis-Starr. 2003. Multi-scale data used to analyze the spatial distribution of French grunts, *Haemulon flavolineatum*, relative to hard and soft bottom in a benthic landscape. Environ. Biol. Fish. 66(1): 19-26.

Lobel, P.S. 2001. Fish bioacoustics and behavior: passive acoustic detection and the application of a closed-circuit rebreather for field study. Mar. Technol. Soc. J. 35: 19-28.

Mastro, S.J. and D.A. Dinsmore. 1989. The operational advantages of enriched air nitrox versus air for research diving. Mar. Technol. Soc. J. 23: 42-46.

Menza, C., J. Ault, J. Beets, J. Bohnsack, C. Caldow, J. Christensen, A. Friedlander, C. Jeffrey, M. Kendall, J. Luo, M. Monaco, S. Smith, and K. Woody. 2006. A Guide to Monitoring Reef Fish in the National Park Service's South Florida / Caribbean Network. NOAA Technical Memorandum NOS NCCOS 39. Silver Spring, MD. 166 pp.

Menza, C., C. Caldow, and M. Monaco. 2007. Analysis of Using a Sample Frame and Subsampling For Reef Fish Surveys. Unpublished Report.

PADI. 2003. PADI Enriched air diving manual, Imperial Edition. PADI, Rancho Santa Margarita, CA.

Parrish, F.A. and R.L. Pyle. 2002. Field comparison of open-circuit scuba to closed-circuit rebreathers for deep mixed-gas diving operations. Mar. Technol. Soc. J. 36: 13-22.

Lohr, S.L. 1999. Sampling: Design and Analysis. Duxbury Press. Pacific Grove, CA. 512 pp.

SAS Institute. 2000. JMP© Statistics and Graphics Guide, Version 4.00. SAS Institute, Inc. Cary, NC, USA.

Thompson, S.K. 2002. Sampling, 2nd edition. John Wiley & Sons, Inc., New York, 343pp.

Thompson, S.K. and G.A.F. Seber. 1996. Adaptive Sampling. John Wiley & Sons, Inc., New York, 265pp.

Wells, J. 1989. Nitrox diving within NOAA: history, applications, and future. pp. 31-35. In: R. Hamilton, D. Crosson, and A. Hulbert (eds.). Workshop on Enriched Air Nitrox Diving. National Undersea Research Program Research Report 89(1). 135 pp.