South Florida/Caribbean Network

National Park Service
Inventory and Monitoring Division

Standard Operating Procedure RF05

Data Analyses and Reporting—Version 1.1

Please cite this as:

South Florida/Caribbean Network. 2025. Data Analyses and Reporting—Version 1.1. South Florida/Caribbean Network Standard Operating Procedure NPS/SFCN/SOP—RF05. South Florida/Caribbean Network, Palmetto Bay, Florida.

Purpose

This standard operating procedure (SOP) describes standard data analyses for fisheries indices such as relative density and size in the exploited phase of the stock and important reef fish ecological indices such as species richness, biomass and proportion of occurrence.

Revision History Log

Revision Date	Author	Changes Made	Reason for Change	New Version #
3/24/2020	Mike Feeley	Added single-stage formulas and updates related to this survey design change	Fish surveys are a single-stage design in Florida and the USVI	1.1

Contents

Page

Introduction 3

Fishery Assessment and Calculation of Metrics 4

Relative Density 5

Abundance 6

Average Size in the Exploited Phase ($\bar{L}$) 7

Descriptive Ecological Analysis 9

Relative Proportion of Occurrence 9

Species Richness 10

Assessing Changes to Reef Fish Metrics 11

Multi-Species Fishery Stock Assessments 12

NCRMP Analysis R-package 12

Installation 13

Example Workflows 13

Basic Workflow 13

Length Classes and Length Bins 16

Biomass Estimates 17

List of Functions 19

Importing Data Sets 20

Analysis Ready Data Variables 21

Raw Data Variables 23

Literature Cited 27

Introduction

The types of data analyses and reporting derived from the sample data will be specific to the types of management questions and resource goals that are being addressed. The objectives are to monitor reef fish species richness and the relative densities, proportion occurrence, average size in the exploited phase, and size structure of target species and determine changes in these parameters over time across "analysis" strata within the Florida Keys and USVI regions, specific sub-regions, and inside versus outside different marine management zones (NPS, FKNMS). Analyses below the sub-region level may require enhanced sampling to be effective. These estimates are relative rather than absolute since we do not know the true capture efficiency of a visual survey for a particular fish species. We are not likely counting 100% of all fish present within the cylinder in a given time. Counts may be higher on average for some species and lower on average for other species. Therefore, proportion occurrence, species richness and density data will be reported per secondary sampling unit (SSU). Data summaries of density abundance are considered relative population totals (or index) and not an absolute value. Estimates of relative density and abundance are most appropriate for target species, and may not be accurate or reliable for rare, highly mobile, or cryptic species.

This SOP describes suggested analyses that should be applied to the data annually and before sampling commences in each subsequent season. The intent of the RVC survey design is to track population metrics of exploited and target reef fish species over time (e.g., annual, biennial) and space (e.g., inside/outside marine reserves). The types of changes in population metrics that we have observed over the past 20 years of RVC surveys have been in response to management interventions (e.g., pre- and post-implementation of no-take marine reserves) or environmental disturbances (e.g., pre- and post-hurricanes). These analyses are better suited using the standard survey design methods described below, rather than a model-based GLM-style regression analysis. Note that model-based methods are not suitable for trend analyses of stratified random surveys because the underlying data assumptions of simple random sampling over the survey frame are not satisfied.

Data sets, which are managed by NOAA-SEFSC, include RVC sample data, stratum data, taxonomic data and benthic data available by region (Florida Keys, Tortugas, St. Thomas/St. John and St. Croix). Analysis ready datasets and data attributes for South Florida and the USVI (as well as Puerto Rico and Southeast Florida) are shared on-line at the South Florida National Coral Reef Monitoring Program data portal (https://grunt.sefsc.noaa.gov/rvc_analysis20/), as well as archived through NOAA's National Centers for Environmental Information (NCEI). Data collected the previous year will generally be made available for download by August of the following year. All data collected in South Florida and the Tortugas since 1999, and since 2017 in the U.S Virgin Islands, are available as comma-separated value (CSV) spreadsheet files in uncompressed or compressed zip files. Prior to 2017, the belt-transect method was used to conduct fish surveys in the U.S Virgin Islands. The benthic data include physical habitat covariates that are made available for general use.

Fishery Assessment and Calculation of Metrics

One of the objectives of the Florida Keys reef fish monitoring program is to evaluate changes in the population structures of commercially important fish species. The six focal species are: Hogfish (*Lachnolaimus maximus*), Mutton Snapper (*Lutjanus analis*), Gray Snapper (*L. griseus*), Yellowtail Snapper (*Ocyurus chrysurus*), Red Grouper (*Epinephelus morio*), and Black Grouper (*Mycteroperca bonaci*). In the U.S. Virgin Islands, the focal species are a mix of commercial species and several ecosystem health indicator species. Several large-bodied snapper and grouper species are significantly overfished and are consequently not encountered routinely on visual surveys are no longer sustainable or important commercial fisheries. Unlike the Florida Keys, parrotfishes and surgeon fishes are commercially important fish species, and are also recognized as important grazers of macroalgae on the reef. The eight focal species in the U.S. Virgin Islands are: Coney (*Cephalopholis fulva*), Red Hind (*Epinephelus guttatus*) Yellowtail Snapper (*Ocyurus chrysurus*), French Grunt (*Haemulon flavolineatum),* Foureye Butterflyfish (*Chaetodon capistratus*), Stoplight Parrotfish (*Sparisoma viride*), Blue Tang (*Acanthurus coeruleus*) and Queen Triggerfish (*Balistes vetula*). These fish species are all either ecologically and/or economically important, but their populations may fluctuate in response to changes in the environment and/or fishing pressure. Therefore, a key component of data analysis consists of calculating population metrics for each of these species and comparing them among management zones relative to sampling in years past. Note that a true performance evaluation of a management zone needs to consider the habitat composition of the management unit. Differences in habitat composition between areas (e.g., inside and outside of protected areas) can drive results and need to be accounted for.

Robust estimators of population "health" for the focal species above include the relative density of species and the average length of the exploited phase individuals, $\bar{L}$ (Ault et al. 2001; Ault et al. 2005). These two biological indicators are calculated and compared to previous years and between management zones with consideration to habitat. Mean density and $\bar{L}$ may also be considered - spatially across the domain. Density and abundance measures are relative since we do not know the true capture efficiency of a visual survey for a particular fish species.

We recognize that the nature of fish behavior may introduce an element of bias that may affect the accuracy and precision of visual census surveys (Pais and Cabral 2017). It is difficult to select one standardized visual method or survey area configuration that is a best fit for all species, conditions and habitats encountered in field surveys. Part of this is attributed to the non-instantaneous nature of underwater surveys since fish can enter a survey area that were not there at the start (Pais and Cabral 2018). Thus, the RVC point count utilizes a short period and large area to minimize potential bias of fish behavior avoidance and/or attraction. Fish recorded in period 1 (< 5 minutes) and period 2 (5–10 minutes) are used in all analyses. Additionally, cryptic species (Gobidae, Labrisomidae, Chaenopsidae, Blennidae, Tripterygiidae, Callionymidae, Opistognathidae, Apogonidae, Ogcocephalidae, Scorpaenidae, Antennariidae, Syngathidae, Synodonitae, Uranoscopidae) recorded in period 3 (> 10 minutes) are included in species richness analyses under the assumption that they

were always present within the cylinder during the survey and did not swim into it while the survey was underway.

Relative Density

Relative mean fish density can be estimated for the entire domain (i.e., Florida Keys) or for a specified analysis stratum of interest (i.e., management zone, geographic sub-region). Fish density, *D*, is the number of individuals observed per diver station (i.e., individuals observed/SSU). However, because of dive safety considerations, the actual fish density for each diver station within a primary unit is the average density for the buddy team of divers (usually two, but sometimes three). There are typically two diver stations (second-stage units) sampled per primary unit. The mean density ($\overline{D}_h$) in stratum *h,* where D_{hj} is the density in second-stage unit *j* in stratum *h* and n_h is the number of primary units sampled in stratum *h* is estimated as:

$$\overline{D}_h = \frac{1}{n_h}\sum_j D_{hj}$$ SOP RF01.2

To calculate the overall stratified mean density for the domain ($\overline{D}_{st}$), stratum densities and a weighting factor are used:

$$\overline{D}_{st} = \sum_h \left(w_h \overline{D_h}\right)$$ SOP RF01.6

where w_h is the stratum weighting factor.

$$w_h = \frac{N_h}{\sum_h N_h}$$ SOP RF01.5

where N_h is the total possible number of primary units in a stratum. The area of a stationary point count cylinder is approximately 177 square meters (1,905 square feet [ft^2]) and the total area of a primary unit is 2,500 square meters (50 × 50 meters [164 × 164 feet {ft}]). Therefore, there is the equivalent of the area of fourteen secondary sampling units within each primary sampling unit. The total possible number of primary units in a stratum, N_h, is dependent on the area that the stratum extends over, and so values for N_h are derived directly from the GIS benthic maps.

Sample variance of density in stratum *h* is calculated among primary units in stratum *h*:

$$s_h^2 = \frac{1}{n_h - 1}\sum_{j=1}^{n_h}\left(D_{hj} - \overline{D}_h\right)^2$$ SOP RF01.3

Then the variance of the mean density for stratum *h* is:

$$var[\bar{D}_h] = \left(1 - \frac{n_h}{N_h}\right)\frac{s_h^2}{n_h}$$ SOP RF05.1

And finally, the variance for the domain wide mean density variance $var\left[\overline{D}_{st}\right]$ is calculated using the weighting factors for each stratum.

$$var[\bar{D}_{st}] = \sum_h w_h^2 \, var[\bar{D}_h] \qquad \text{SOP RF05.2}$$

Standard error $SE\left[\overline{\overline{D}}_{st}\right]$ and the coefficient of variation of domain wide mean density $CV[\overline{D}_{st}]$ are:

$$SE[\overline{D}_{st}] = \sqrt{var[\overline{D}_{st}]} \qquad \text{SOP RF05.3}$$

$$CV[\overline{D}_{st}] = \frac{SE[\overline{D}_{st}]}{\overline{D}_{st}} \qquad \text{SOP RF05.4}$$

Overall domain-wide mean density and stratum-level mean density should be calculated separately for juvenile (size of first recruitment to the size of first maturity) and adult (size of first sexual maturity to maximum size) or exploited phases of each focal species, as these species are known to exhibit ontogenetic patterns among reef habitats through time. Calculations of mean densities should then be compared with consideration to changes through time and potential spatial patterns.

Abundance

Domain wide total relative abundance values ($\hat{Y}_{st}$) for each species (i.e., estimated total number of fish in an area) can be estimated by sub-region and management zone after stratum densities are calculated. Domain-wide relative abundance and associated variance are obtained by summing the respective stratum estimates over all strata. First relative abundance values for each stratum are calculated:

$$\hat{Y}_{h=(\overline{D}_h)(N_h)} \qquad \text{SOP RF05.5}$$

N_h values can be for management zones or sub-regions. Then relative abundances are summed for all strata to get a total relative abundance.

$$\hat{Y}_{st} = \sum_h \hat{Y}_h \qquad \text{SOP RF05.6}$$

Survey wide variance of mean relative abundance ($var\left[\hat{Y}_{st}\right]$) can be calculated as follows:

First stratum variance of mean relative abundance is calculated:

$$var\left[\hat{Y}_h\right] = var[\overline{D}_h] \, (N_h)^2 \qquad \text{SOP RF05.7}$$

The survey wide variance of mean relative abundance is then calculated by summing up the variances of each stratum.

$$\text{var}\left[\hat{Y}_{st}\right] = \sum_h \text{var}\left[\hat{Y}_h\right] \qquad \text{SOP RF05.8}$$

The standard error and CV are calculated as follows:

$$SE\left[\hat{Y}_{st}\right] = \sqrt{\text{var}\left[\hat{Y}_{st}\right]}$$ SOP RF05.9

$$CV[\overline{\hat{Y}}_{st}] = \frac{SE[\overline{\hat{Y}}_{st}]}{\overline{\hat{Y}}_{st}}$$ SOP RF05.10

Average Size in the Exploited Phase ($\bar{L}$)

The average size (length or weight) in the exploited phase of the stock (Beverton and Holt 1957) is a physiologically-based indicator that is sensitive to direct and indirect stress on marine ecosystems (Ault et al. 2005). It is highly correlated with average population size, and so reflects the rate of fishing mortality operating in the fishery. As fishing mortality rate increases, $\bar{L}$ decreases at a rate proportional to the population-dynamic tolerance of a stock (Ault et al. 2005). $\bar{L}$ will decrease until it is nearly equal to the minimum size of first capture regulated by fishery management. By its definition, there also exists a value of $\bar{L}$ corresponding to a population size that produces maximum sustainable yields on a continuing basis.

$\bar{L}$ at time *t* is defined as:

$$\overline{L}(t) = \frac{F(t)\int_{a_c}^{a_\alpha} N(a,t)L(a,t)da}{F(t)\int_{a_c}^{a_\alpha} N(a,t)da}$$

$$\overline{\mathrm{L}}(\mathrm{t}) = \frac{\mathrm{F}(\mathrm{t})\int_{\mathrm{a_c}}^{\mathrm{a_\lambda}} \mathrm{N}(\mathrm{a},\mathrm{t})\mathrm{L}(\mathrm{a},\mathrm{t})\mathrm{da}}{\mathrm{F}(\mathrm{t})\int_{\mathrm{a_c}}^{\mathrm{a_\lambda}} \mathrm{N}(\mathrm{a},\mathrm{t})\mathrm{da}}$$

where a_c is the minimum age at first capture, a_λ is the oldest age in the stock, *N(a,t)* is the abundance for age class *a*, *L(a,t)* is the length-at-age, and *F(t)* is the instantaneous fishing mortality rate at time *t*.

The average size in the exploited phase ($\bar{L}$) of focal species is calculated at the end of each survey season. This metric is calculated by averaging the size of all fish that are greater than or equal to the exploited phase size limit. In Florida, this is available in the taxonomic variables and defined as the minimum length of capture (L_C) in centimeters. In the Virgin Islands, the only species with a L_C is Yellowtail Snapper (*Ocyurus chrysurus*); therefore, the exploited phase of all other species is recognized as L_M, the average fork length of maturity in centimeters.

$\bar{L}$ is defined as the average fork length of a fish species in the exploited phase following standard statistical procedures:

$$\bar{L} = \frac{1}{n_{exp}} \sum L \qquad \text{SOP RF05.11}$$

where n_{exp} is defined as the number of fish in the exploited phase and L is the individual fork length of an exploited phase fish. Non-normality may be corrected by a log transformation if necessary. The average size of all fish, juveniles or adults are also calculated by substituting the respective value of n for that phase of the stock.

For the survey however, the design ratio-of-means estimator becomes,

$$\bar{L}_{st} = \frac{\bar{Y}_{st}}{\bar{X}_{st}} \qquad \text{SOP RF05.12}$$

where X is the density of exploited phase fish and Y is defined as the sum of lengths of the exploited phase fish. The stratum estimates of X and Y are,

$$\bar{X}_h = \frac{1}{n_h} \sum_j X_{hj} \qquad \text{SOP RF05.13}$$

and

$$\bar{Y}_h = \frac{1}{n_h} \sum_j Y_{hj} \qquad \text{SOP RF05.14}$$

Where x_{hj} is the number of exploited phase fish in sample unit j in stratum h, Y_{hj} is the sum of lengths of exploited phase fish in the sample unit j in stratum h and n_h is the number of units sampled in stratum h. With the domain estimates of X and Y further defined as,

$$\bar{X}_{st} = \sum_h w_h \bar{X}_h \qquad \text{SOP RF05.15}$$

$$\bar{Y}_{st} = \sum_h w_h \bar{Y}_h \qquad \text{SOP RF05.16}$$

Finally, the domain wide estimate of the average size in the exploited phase ($\bar{L}st$) becomes

$$\bar{L}_{st} = \frac{\bar{Y}_{st}}{\bar{X}_{st}} \qquad \text{SOP RF05.17}$$

Sample variance in stratum h is,

$$s_h^2 = \frac{\sum_j (Y_{hj} - \bar{L}_{st} X_{hj})^2}{\bar{n}_h - 1} \qquad \text{SOP RF05.18}$$

Variance of the average size in the exploited phase ($\bar{L}$) in stratum h is,

$$var[\bar{L}_h] = \left(1 - \frac{n_h}{N_h}\right)\frac{s_h^2}{n_h \bar{x}_h}$$ SOP RF05.19

Variance of the survey domain is,

$$var[\bar{L}_{st}] = \sum_h w_h^2 var[\bar{L}_h]$$ SOP RF05.20

and the standard error (*SE*) for the survey domain is,

$$SE[\bar{L}_{st}] = \sqrt{var[\bar{L}_{st}]}$$ SOP RF05.21

Descriptive Ecological Analysis

A basic descriptive ecological analysis is also completed annually that includes a species list and proportion of occurrence of all fish species observed as well as indices of fish biodiversity. Species composition is assessed using the indices proportion of occurrence, species richness, and diversity. Proportion of occurrence is a straightforward measure of the percent of samples where a species was recorded. Richness is calculated as the number of species observed.

Relative Proportion of Occurrence

The relative proportion of occurrence of each species is generated by sub-region and park. First, the total number of possible secondary sample units in a stratum (N_h) is determined. The total number of primary units (grid cells) in a stratum is dependent on stratum area (A_h), so values for are derived directly from GIS grid layers.

$$N_h = \frac{A_h}{177\ \mathrm{m}^2}$$

SOP RF05.22

Where the stratum area (A_h) is the total primary units in stratum h * 2,500 square meters (26,910 ft^2).

Proportion occurrence for each species is calculated first by stratum ($\bar{P}_h$), then multiplied by a weighting factor (W_h) and summed to provide a domain estimate ($\bar{P}_{st}$).

$$\bar{P}_h = \frac{1}{n_h}\sum_j P_{hj}$$ SOP RF05.23

w_h is the stratum weighting factor. This weighting factor is calculated for each stratum as:

$$w_h = \frac{N_h}{\sum_h N_h}$$ SOP RF01.5

and

$$\bar{P}_{st} = \sum_h w_h \bar{P}_h$$ SOP RF05.24

Variance and standard error are also calculated for the mean proportion occurrence each year. Variance of the mean is first calculated for each stratum, $var[\bar{P}_h]$).

$$var[\overline{P}_h] = \left(1 - \frac{n_h}{N_h}\right) \times \frac{\left(\frac{n_h}{(n_h-1)}\right) \times \overline{P}_h \times (1-\overline{P}_h)}{n_h}$$ SOP RF05.25

Total management zone variance $var[\overline{P}_{st}]$) is calculated by summing all stratum variance values that have been multiplied by the squared stratum weighting factor.

$$\text{var}[\bar{P}_{st}] = \sum_h var[\bar{P}_h]\, w_h^2$$ SOP RF05.26

Standard error for proportion of occurrence can also be calculated:

$$SE[\overline{P}_{st}] = \sqrt{var[\bar{P}_{st}]}$$ SOP RF05.27

Species Richness

In addition to proportion occurrence for all species, a total count of species observed throughout all surveys as well as mean species richness (defined here as average number of species observed per SSU) can be calculated. The mean species richness ($\overline{Sp}_h$) for each stratum (h) within the entire domain or specified region of interest is first calculated. Where n_h is the number of primary units sampled in stratum h of each sample unit j:

$$\overline{Sp}_h = \frac{1}{n_h} \sum_h sp_{hj}$$ SOP RF05.28

To calculate the overall stratified mean species richness for the domain ($\overline{Sp}_{st}$) stratum densities and a weighting factor are used:

$$\overline{Sp}_{st} = \sum_h w_h \overline{Sp}_h$$ SOP RF05.29

The sample variance of species richness in stratum h ($\overline{Sp}_h^2$), variance of the mean richness by strata ($var[\overline{Sp}_h]$) and variance of domain wide mean species richness ($var[\overline{Sp}_{st}]$) can be calculated as:

$$\overline{Sp}_h^2 = \frac{1}{n_h - 1} \sum_i \left(\overline{Sp}_{hi} - \overline{Sp}_h\right)^2$$ SOP RF05.30

$$\text{var}\left[\overline{Sp}_h\right] = \left(1 - \frac{n_h}{N_h}\right) \frac{\overline{Sp}_h^2}{n_h}$$ SOP RF05.31

$$\text{var}\left[\overline{Sp}_{st}\right] = \sum w_h^2 \,\text{var}\left[\overline{Sp}_h\right]$$ SOP RF05.32

Assessing Changes to Reef Fish Metrics

An advantageous property of statistical sampling theory is that survey design estimates, such as stratum density and the variance of stratum density, do not require knowledge of the underlying probability distribution (e.g., normal, gamma, etc.) of the respective population metric, (e.g., observations of stratum density, y_{hi}) (Cochran 1977). A second property, based on large sample theory, is that survey design estimates (e.g., density) are normally distributed due to the central limit theorem if sample size is large. These properties facilitate the analysis of survey design estimates among times or areas (e.g., monitoring density over time, assessing MPA effectiveness). With large enough samples the distribution of the estimate is approximately normal and thus confidence intervals (CI) can be determined (Neyman 1934; Hansen 1987). The method of stratified random sampling allows a consistent estimate of the average of a variable, whatever the properties of the population, and allows an estimate of precision with CI without depending on the validity of the assumed model (Neyman 1934; Hansen 1987).

A simple, straightforward approach to performing statistical tests for differences among survey estimates for a time or area are confidence intervals (CI). Cochran (1977) and Lohr (2010), or in the technical documentation for the mainline statistical packages, e.g., SAS PROC SURVEYMEANS demonstrate the practice of using standard errors from survey estimates to construct confidence intervals for statistical evaluation. This is the basis for the t-test procedure described in the documentation entitled "SAS PROC SURVEYMEANS, Details, Statistical Computations". If the sample is relatively large (n > 100) and has a Normal distribution, the survey mean, $\overline{y}_{st}$, will lie within a CI bounded by:

$$\overline{y}_{st} \pm t_{\alpha/k,df} SE[\overline{y}_{st}]$$

With a probability of α, the Type I error rate, and where *t* is the critical value of Student's *t*-distribution, and degrees of freedom are *d.f.* = $\sum_h (n_h - 1)$. Most reef metrics do not possess a Normal distribution which means the Type I error rate will not equal α. Cochran (1977) states α will be very close to what is expected if:

$$n > 25\ G_1^2$$

where G_1^2 is Fisher's measure of skewness. If a sample is too small and the population is heavily skewed, transforming the data (e.g., log [y_{st} + 1], [y_{st}]^2) may help. With the cooperative effort of the interagency program, adequate sample sizes are typically not an issue. Degrees of freedom are computed following the "PROC SURVEYMEANS" documentation, which states they "*calculate the degrees of freedom for the t-test as the number of clusters minus the number of strata. If there are no clusters, then the degrees of freedom equal the number of observations, minus the number of strata."* Using n-H to compute the degrees of freedom for stratified random designs is also stated and shown in worked examples in numerous places in Lohr (2010), e.g., Section 3.2 — Theory of Stratified Sampling, and Section 4.5 — Ratio Estimation with Stratified Samples. This is the same procedure for constructing confidence intervals and conducting t-tests found in Ault et al. (2006, 2013). CIs can be used to test the hypothesis that samples were drawn from the same population, as is done to assess temporal change or determine MPA effectiveness. Cochran (1977), Sokal and Rohlf (2001),

and Zar (1996) describe methods using CIs to test for differences among means. The RVC surveys are designed to produce two basic types of estimates: (1) frame-wide population metrics (e.g., relative abundance); and, (2) population dynamic metrics. These are considered for two separate management sub-domains of the survey frame: (i) open to fishing; and, (ii) closed to fishing. As such, the basic statistical testing boils down to comparing metrics at various time points in the survey history, or comparing open versus closed spatial management zones as was illustrated in Ault et al. (2006, 2013) and Smith et al. (2011). Thus, multiple comparisons are not necessary for the standard types of analyses for which the RVC surveys were designed. A useful relationship stemming from Equation SOP RF05.16 is that the 95% CI for a population metric is approximately twice the CV, because the value of t for $\alpha = 0.05$ and $d.f. > 20$ is approximately 2. Thus, for example, a stratified random sampling survey that provides a domain-wide estimate of abundance with a CV of 15% would be able to statistically detect a minimum change of 30% in population abundance between survey time periods (with a Type I error rate = 0.05).

Multi-Species Fishery Stock Assessments

It is important to quantitatively assess stocks of exploited reef species across the sampling domain and evaluate their status relative to established Federal and International benchmarks for sustainable management of fishery resources. Calculation of mean fish density by life stage and $\bar{L}$ for each of the focal species is described above. Population dynamic parameters such as natural mortality rate, fishing mortality rate, spawning stock biomass, and spawning potential ratio may also be derived from the data set annually or when needed. Methodologies for computing these parameters from the data can be found in the scientific literature; for example, in Ault and Ehrhardt (1991), Ehrhardt and Ault (1992), Ault and Olson (1996), Ault et al. (2008, 2014, 2019).

Additional life history information for species observed in the Reef Visual Census is available at the NOAA-SEFSC data portal. This includes scientific and common names for species observed, the species code (first three letters of the generic name, and first four letters of the specific name) and life history data for select species: L_M - median length at maturity; L_C - minimum length at capture; W_{LEN_A} - linear coefficient of the allometric growth equation in g/mm and W_{LEN_B} - exponential coefficient of the allometric growth equation.

$W = a\, L^b$

NCRMP Analysis R-package

An updated National Coral Reef Monitoring Program (NCRMP) reef fish analysis R-package is available at the NOAA-SEFSC data portal (https://grunt.sefsc.noaa.gov/rvc_analysis20/). The NCRMP R-package is shared on the code hosting platform GitHub (https://github.com/jeremiaheb/rvc) (https://github.com/jeremiaheb/rvc) and is maintained by the NOAA-SEFSC data manager. The new version (v1.0.0) allows analyses for samples taken under a one-stage or a two-stage design. The Florida Keys and the Dry Tortugas were a two-stage design through 2018. Since 2020, Florida has been sampled under a one-stage design. The USVI regions, St. Thomas/St. John and St. Croix, have always been a one-stage survey design. Since variance

calculations are different for one and two-stage survey designs, the new code tests for the stage level and adjusts the formulas accordingly.

The fish analysis R-package is designed to conveniently access and analyze the RVC data. It can be used to import RVC data into R and produce commonly used summary statistics such as relative fish density, relative proportion of occurrence, length frequencies and biomass which can then be output to CSV files or turned into graphs.

Installation

The updated NCRMP fish analysis R-package can be installed through the RStudio console by typing the following into R.

```
install.packages('devtools')

devtools::install_github('jeremiaheb/rvc')
```

A restart of RStudio is required after you have completed installing the package. This will install the newest version of the RVC fish analysis R-package (v 1.0.0). If you have previously analyzed data in an existing R project using the previous package, you will run into an error stating that 'STAGE_LEVEL' is not present in your stratum file. To fix this either:

1. Download the data again (after updating the RVC fish package) using the getRVCData() function call (see Basic Workflow section below). This will bring in the new NTOT file that states if it is from a one-stage or two-stage design.

2. Manually add the STAGE_LEVEL field to your stratum file with the correct value (1 or 2).

Example Workflows

The following examples are available in the NCRMP Reef Visual Census statistical package in R project compiled by Harry Ganz, Andrew Jo and Jeremiah Blondeau and hosted on GitHub.

Basic Workflow

```
library(rvc)

## Download desired Years/Regions data from server

keys11_14 = getRvcData(years = 2011:2014, regions = c("FLA KEYS", "DRY TORT"))


## Take a look at object structure

str(keys11_14, 1) # 3 data.frames


## Make a list of species
```

```
## You can use full scientific names, common names, or
## species codes (first 3 letters of genus, and first 4 of species)
## Only scientific names are case-sensitive
spcs = c("Epinephelus morio", "Black Grouper", "STE VARI")

## Calculate statistics for entire sampling domain
ddens = getDomainDensity(keys11_14, species = spcs)
dabun = getDomainAbundance(keys11_14, species = spcs)
docc = getDomainOccurrence(keys11_14, species = spcs)

## Stratum Level Estimates for Red Grouper in 2011 in
## The Florida Keys
sdens = getStratumDensity(keys11_14, species = "EPI MORI",
                          years = 2011, regions = "FLA KEYS")
```

Using Filters and Grouping Species

Filtering data can be used to select a subset of the data.

```
require(rvc)

## Get Data from the Florida Keys in 2011-2012
fk11_12 = getRvcData(years = 2011:2012, regions = "FLA KEYS")

## Display occurrence for hogfish both inside and outside
## protected areas separately
lm_prot = getDomainOccurrence(fk11_12, species = "LAC MAXI",
                              merge_protected = FALSE)
## Yellowtail occurrence only in protected areas
yt_prot = getDomainOccurrence(fk11_12, species = "Ocyurus chrysurus",
```

```
 is_protected = TRUE)

## Red grouper density when present

rg_present = getDomainDensity(fk11_12, species = "Red Grouper",

 when_present = TRUE)

## Only select data from 2011

mut_2011 = getDomainAbundance(fk11_12, species = "Lut anal", years = 2011)
```

The group parameter can be used to calculate a statistic per group instead of per species. This can be useful in calculating biomass or abundance per genus or family, or for a trophic grouping.

```
require(rvc)

## Get data for the Dry Tortugas from 2000 - 2006

dt2000_8 = getRvcData(years = 2000:2006, region = "DRY TORT")


## Create a lookup table for snapper/grouper families

lookup = subset(dt2000_8$taxonomic_data, FAMILY %in% c("Lutjanidae", "Serranidae"),

  select = c("SPECIES_CD", "FAMILY"))


## Calculate snapper/grouper abundance

sngpAbun = getDomainAbundance(dt2000_8, species = lookup$SPECIES_CD, group = lookup)


## You can also create your own lookup tables, with species

## names in the first column and group names in the second

lookup2 = data.frame(species = c("Lachnolaimus maximus", "Epinephelus morio"),

group = rep("Custom Grouping",2))


sngpBiom = getDomainBiomass(dt2000_8, species = lookup2$species, group = lookup2)
```

Available Filters:

1. merge_protected: A boolean indicating whether protected areas should be combined in calculating the statistic (TRUE), the default, or should be calculated separately (FALSE).

2. strata: character vector of strata codes to keep.
3. status: character vector of protected statuses to keep.
4. is_protected: A boolean indicating whether only samples from protected areas should be kept (TRUE) or only unprotected areas (FALSE).
5. years: A numeric vector of years to keep.
6. regions: A character vector of region codes to keep. DRY TORT - Dry Tortugas, FLA KEYS: Florida Keys, STTSTJ: St. Thomas / St. John, USVI and STX: St. Croix, USVI
7. when_present: A boolean indicating whether only samples in which individuals are seen should be kept (TRUE) or not (FALSE).
8. group: A lookup table (data.frame) - the first column of which is a list of species codes, scientific names, or common names, and the second column of which is a list of names by which to group the species (e.g. Family names, trophic groups). If this option is used instead of the statistic being calculated per species, it will be calculated per group.

Length Classes and Length Bins

Any statistic in the RVC R-package can be calculated for length classes or in length bins by providing the length_bins option. Be aware that using length bins can make calculating statistics significantly slower.

There are three options in calculating length specific statistics:

1. You can use keywords "lc" and "lm" to calculate statistics above and below the minimum length-at-capture and median length-at-maturity, respectively, assuming the values are available in the database.
2. You can pass in your own number or sequence of numbers which will serve as breakpoints, in cm, above, below, and between which the statistic will be calculated for all species in the species argument combined. This is the most useful when binning data.
3. You can pass in a custom lookup table with species names and breakpoints in centimeters for each species. The lookup table will be used to calculate the statistic above and below the breakpoint for each species separately.

Note: Only option 2 can be used in the getDomainLengthFrequency and getStratumLengthFreqency functions.

```
require(rvc)

## Get 2013 SEFCRI data from server

sf2013 = getRvcData(years = 2013, region = "SEFCRI")
```

```
## Calculate length frequencies for Mangrove Snapper
## for lengths from 0-100cm by 1 cm bins
lglf = getDomainLengthFrequency(sf2013, species = "Lut gris",
                                length_bins = seq(0,100,1))
## Abundance of yellowtail above/below length at maturity (from database)
ytabun = getDomainAbundance(sf2013, species = "OCY CHRY", length_bins = "lm")

## Abundance of yellowtail with 26cm cutpoint
ytabun2 = getDomainAbundance(sf2013, species = "OCY CHRY",
                             length_bins = 26)
## Create a custom lookup table and use it to set length classes
## for Mutton and Mangrove Snapper when calculating densities
lb = data.frame(SPECIES_CD = c("LUT ANAL", "LUT GRIS"), CUT = c(40, 30))
snappers2013 = getDomainDensity(sf2013, c("LUT ANAL", "LUT GRIS"),
  length_bins = lb)

## White Grunt abundance from 0-50cm in 1cm bins
## This calculation may take a few minutes
hp2013 = getDomainAbundance(sf2013, species = "HAE PLUM",
                             seq(0,50,1))
```

Biomass Estimates

Biomass per secondary sampling unit and total biomass for the sampling domain can be calculated using allometric growth parameters. The 'a' and 'b' coefficients must be for conversion from millimeters to grams, even though the final output is in kilograms (see function help files for more information).

```
require(rvc)
dt2012 = getRvcData(years = 2012, regions = "DRY TORT")
```

```
## Calculate biomass per SSU (177m^2) and Total Biomass (per domain)
## for Black and Red Grouper Using data from server
groupers = c("EPI MORI", "MYC BONA")
dbiom = getDomainBiomass(dt2012, species = groupers)
dtbiom = getDomainTotalBiomass(dt2012, species = groupers)
## To take a look at the growth_parameters used
gp1 = subset(dt2012$taxonomic_data, SPECIES_CD %in% toupper(groupers),
  select = c("SPECIES_CD", "WLEN_A", "WLEN_B"))

## Using the rfishbase package, enter custom lookup table for
## Black and Red grouper and recalculate growth parameters
## NOTE: The weighting methods used here are not sufficient where records are
## doubtful or growth parameters vary by location
require(rfishbase)
require(plyr)
## Get length-weight parameters from fishbase
lw = length_weight(c("Epinephelus morio", "Mycteroperca bonaci"))
## Create a table of parameters
## Note the unit conversion of 'a' from cm->g to mm->g by
## Dividing 'a' by 10^b
gp2 = ddply(lw, c("sciname"), summarize,
  WLEN_A = exp(weighted.mean(log(a/10^b), Number)), # Weighted geometric mean
  WLEN_B = weighted.mean(b, Number)) # Weighted arithmetic mean
## Generate species codes from scientific names
gp2$SPECIES_CD = toupper(paste(substr(gp2$sciname, 1, 3),
sapply(strsplit(gp2$sciname, " "), function(x){substr(x[[2]],1,4)}), sep = " "))
## Use custom table in biomass functions
dbiom2 = getDomainBiomass(dt2012, species = groupers, growth_parameters = gp2)
```

```
dtbiom2 = getDomainTotalBiomass(dt2012, species = groupers,
   growth_parameters = gp2)

## If you would like to apply the same growth_parameters to multiple species
## or are only using one species, you can use a list instead of a lookup table
## In this case we can get the biomass of Cocoa Damselfish
## by inputing the growth parameters as a list
gp3 = list(WLEN_A = 2.346e-5, WLEN_B = 2.98)
dbiom3 = getDomainBiomass(dt2012, species = "STE VARI",
 growth_parameters = gp3)
dtbiom3 = getDomainTotalBiomass(dt2012, species = "STE VARI",
 growth_parameters = gp3)
```

List of Functions

- getSampleData - Retrieves sample data from server as data.frame.
- getStratumData - Retrieves stratum information from server as data.frame.
- getBenthicData - Retrieves benthic data from server as data.frame.
- getTaxonomicData - Retrieves taxonomic and life history data from server as data.frame.
- getRvcData - Retrieves sample, stratum, and taxonomic data from server as list.
- getStratumDensity - Calculates average density per secondary sampling unit for each stratum.
- getDomainDensity - Calculates average density per secondary sampling unit for each sampling domain.
- getStratumOccurrence - Calculates average occurrence per secondary sampling unit for each stratum.
- getDomainOccurrence - Calculates average occurrence per secondary sampling unit for each sampling domain.
- getStratumAbundance - Calculates total number of individuals per stratum.
- getDomainAbundance - Calculates total number of individuals per sampling domain.

- getStratumBiomass - Calculates average biomass per secondary sampling for each stratum.
- getDomainBiomass - Calculates average biomass per secondary sampling unit for each sampling domain.
- getStratumTotalBiomass - Calculates total biomass per stratum.
- getDomainTotalBiomass - Calculates total biomass per sampling domain.
- getStratumLengthFreqency - Calculates length frequency per stratum.
- getDomainLengthFrequency - Calculates length frequency per sampling domain.

Importing Data Sets

The sample data or benthic data can be imported directly from the server and can also be saved as a data.frame. The data are divided into four parts:

1. Sample Data: Diver collected counts and geographic information relating to the fish survey.
2. Stratum data: Meta information on the number of primary sample units per stratum and the size of each primary sample unit.
3. Benthic Data: Data collected by diver on the vertical relief, and composition of the benthos at each secondary sampling unit.
4. Taxonomic Data: Taxonomic and life history information for species observed in the Reef Visual Census.

The functions to access this data are shown in the example below:

```
## Get taxonomic data for all species as a data.frame

taxonomic_data = getTaxonomicData()


## Get sample data for a range of years and regions

keys00_14 = getSampleData(years = 2000:2014, regions = c("DRY TORT", "FLA KEYS"))


## Get stratum info for a range of years and regions

ntot00_14 = getStratumData(years = 2000:2014, regions = c("DRY TORT", "FLA KEYS"))


## Get benthic data for the Florida Keys in 2011

benthic11 = getBenthicData(years = 2011, regions = "FLA KEYS")
```

Analysis Ready Data Variables

A list of analysis ready Sample, Stratum, Taxonomic and Benthic data variables are also available at the NOAA-SEFSC data portal and presented in Tables 1 to 3.

Table 1. Analysis Ready Sample Data Variables.

Variable	Description
PRIMARY_SAMPLE_UNIT	A code indicating the primary sample unit in which a sample was collected.
YEAR	A number indicating the calendar year.
MONTH	A number indicating the month of the year.
DAY	A number indicating the day of the month (EST).
STATION_NR	A number indicating the secondary sampling unit within a given primary sample unit.
LAT_DEGREES	Latitude of secondary sampling unit in decimal degrees.
LON_DEGREES	Longitude of secondary sampling unit in decimal degrees.
DEPTH	Average depth, in meters, of secondary sampling unit.
UNDERWATER_VISIBILITY	Visibility, in meters, at secondary sampling unit.
MAPGRID_NR	A unique numerical identifier indicating the primary sample grid number.
HABITAT_CD	A code indicating the habitat type.
ZONE_NR	A code indicating the distance offshore: 1 - Inshore, 2 - Midchannel, 3 - Offshore, 4 - Fore-reef.
SUBREGION_NR	A number indicating the subregion.
MPA_NR	A number identifying the marine protected area in which the sample was collected. Zero indicates unprotected status.
SPECIES_NR	A number indicating the species for a sample.
SPECIES_CD	A code indicating the species for a sample. Consists of the first three letters of the generic name and the first four of the specific name.
LEN	The length, in cm, of a sample.
NUM	The number of individuals of a given species and length observed in a secondary sampling unit.
TIME_SEEN	A number indicating when, during sampling, an individual was observed. 1: In the first five minutes, 2: From 5-10 minutes, 3: After 10 minutes.
PROT	A boolean value indicating whether a sample was in a protected area (1), or not (0).
STRAT	A code indicating the stratum in which a sample was taken. Differs by region.

Variable	Description
REGION	A code indicating the region in which a sample was taken. DRY TORT: Dry Tortugas, FLA KEYS: Florida Keys, STTSTJ: St. Thomas / St. John, USVI and STX: St. Croix, USVI

Table 2. Analysis Ready Stratum Data Variables (NTOT file)

Variable	Description
REGION	A code indicating the region. DRY TORT: Dry Tortugas, FLA KEYS: Florida Keys, STTSTJ: St. Thomas / St. John, USVI and STX: St. Croix, USVI
YEAR	A number indicating the calendar year.
PROT	A boolean indicating the protected status
ADMIN	Code indicating a named managed area. Caribbean and GOM regions only.
STRAT	A code indicating the stratum. Differs by region.
NTOT	The number of possible primary sample units for a given year, region, stratum, and protected status.
GRID_SIZE	The length (in meters) to a side of a primary sample unit for a given year, region, stratum, and protected status.
STAGE_LEVEL	Indicates 1 or 2 stage design.

Table 3. Analysis Ready Taxonomic Data Variables

Variable	Description
SPECIES_CD	The first three letters of the generic name and first four of the specific name.
FAMILY	The family name.
SCINAME	The scientific name.
COMNAME	The common name.
LC	Minimum length at capture, in centimeters.
LM	Median length at maturity, in centimeters.
WLEN_A	The linear coefficient of the allometric growth equation in grams per millimeter (g/mm)
WLEN_B	The exponential coefficient of the allometric growth equation.

Raw Data Variables

A list of raw diver-collected Sample, Substrate and Species variables are also available at the NOAA-NCEI data archive and presented in Tables 4 to 6. The sample and substrate data are made available to predict fish density and size as a function of the available covariates. The modeling of function correlations on fish metrics are not a formal part of this protocol.

Tables 4. Raw Sample Data Variables. Diver level sample/site information collected using RVC.

Variable	Description
MASTER_SAMPLE_CD	unique sample identifier consisting of year, month, day, sample time, diver number
DIVER_NR	primary diver number
BUDDY_NR	diver number of primary diver's buddy
SAMPLE_TYPE_NR	all sample types are 1, indicating Bohnsack-Bannerot stationary point count
divhab	integer representing habitat name
habname	habitat name
YEAR	year
MONTH	month
DAY	day
divbegh	dive begin hour
divbegm	dive begin minute
divendh	dive end hour
divendm	dive end minute
SAMPLE_START_TIME_HOURS	sample start hour
SAMPLE_START_TIME_MINUTES	sample start minute
SAMPLE_END_TIME_HOURS	sample end hour
SAMPLE_END_TIME_MINUTES	sample end minute
divdepth	max dive depth
SAMPLE_DEPTH	Sample depth, in feet
FISHING_GEAR	field indicating fishing gear or debris present
fieldid	ID of sample consisting of primary sample unit, station number and replicate
UNDERWATER_VISIBILITY	Visibility underwater in meters
WATER_TEMP	underwater temperature in f
CURRENT	underwater current low, medium, high
SAMPLE_DESCR	diver notes

Tables 5. Raw Substrate Data Variables. Diver level substrate information collected using the RVC methodology.

Variable	Description
MASTER_SAMPLE_CD	unique sample identifier consisting of year, month, day, sample time, diver number
SUBSTRATE_MAX_DEPTH	max depth (f) in diver cylinder
SUBSTRATE_MIN_DEPTH	min depth (f) in diver cylinder
HARD_VERTICAL_RELIEF	max height of hard substrate (m)
SOFT_VERTICAL_RELIEF	max height of soft substrate (m)
h_relcat_0	proportion of hard substrate with height from 0-20 cm
h_relcat_1	proportion of hard substrate with height from >20 - 50 cm
h_relcat_2	proportion of hard substrate with height from >50 - 100 cm
h_relcat_3	proportion of hard substrate with height from >100 - 150 cm
h_relcat_4	proportion of hard substrate with height from >150 cm
s_relcat_0	proportion of soft substrate with height from 0-20 cm
s_relcat_1	proportion of soft substrate with height from >20 - 50 cm
s_relcat_2	proportion of soft substrate with height from >50 - 100 cm
s_relcat_3	proportion of soft substrate with height from >100 - 150 cm
s_relcat_4	proportion of soft substrate with height from >150 cm
SAND_PERCENTAGE	abiotic cover sand percentage
HARDB_PERCENTAGE	abiotic cover hardbottom percentage
RUBBLE_PERCENTAGE	abiotic cover rubble percentage
SAND_BARE	biotic cover percentage in sand that is bare
SAND_MACRO_ALGAE	biotic cover percentage in sand that has macro algae
SAND_SEAGRASS	biotic cover percentage in sand that is seagrass
SAND_SPONGE	biotic cover percentage in sand that is sponges
s_pcov_oth1_lab	diver labeled other category
s_pcov_oth2_lab	diver labeled other category
s_pcov_oth1	biotic cover percentage in sand that is labeled other_1
s_pcov_oth2	biotic cover percentage in sand that is labeled other_2
HARDBOTTOM_ALGAL_TURF	biotic cover percentage on hardbottom that has turf algae

Variable	Description
HARDBOTTOM_MACRO_ALGAE	biotic cover percentage on hardbottom that has macro algae
HARDBOTTOM_LIVE_CORAL	biotic cover percentage on hardbottom that has live coral
HARDBOTTOM_OCTOCORAL	biotic cover percentage on hardbottom that has octocoral
HARDBOTTOM_SPONGE	biotic cover percentage on hardbottom that is sponges
h_pcov_oth1_lab	diver labeled other category
h_pcov_oth2_lab	diver labeled other category
h_pcov_oth1	biotic cover percentage on hardbottom that is labeled other_1
h_pcov_oth2	biotic cover percentage on hardbottom that is labeled other_2

Tables 6. Raw Species Data Variables. Diver level species information collected using the RVC methodology.

Variable	Description
MASTER_SAMPLE_CD	unique sample identifier consisting of year, month, day, sample time, diver number
SPECIES_CD	A code indicating the species for a sample. Consists of the first three letters of the genus and the first four of the species name.
NUMBER_OF_INDIVIDUALS	The number of individuals of a given species and length observed in a secondary sampling unit.
AVERAGE_LENGTH	average length of most of the species from number_of_individuals (cm)
MIN_LENGTH	minimum length of species (cm)
MAX_LENGTH	maximum length of species (cm)
TIME_SEEN	A number indicating when, during sampling, an individual was observed.
instance_field	legacy field referencing order of species entered into database
instance_key	legacy field referencing time period that species entered into database

Literature Cited

Ault, J. S., J. A. Bohnsack, S. G. Smith, and J. G. Luo. 2005. Towards sustainable multispecies fisheries in the Florida, USA, coral reef ecosystem. Bulletin of Marine Science **76**:595–622.

Ault, J. S. and N. M. Ehrhardt. 1991. Correction to the Beverton and Hold Z-estimator for truncated catch length-frequency distributions. ICLARM Fishbyte **9**:37–39.

Ault, J. S. and D. B. Olson. 1996. A multicohort stock production model. Transactions of the American Fisheries Society **125**:343–363.

Ault, J. S., S. G. Smith, J. A. Bohnsack, J. Luo, D. E. Harper, and D. B. McClellan. 2006. Building sustainable fisheries in Florida's coral reef ecosystem: positive signs in the Dry Tortugas. Bulletin of Marine Science 78: 633–654.

Ault, J. S., S. G. Smith, J. A. Bohnsack, J. Luo, M. H. Stevens, G. T. DiNardo, M. W. Johnson, D. R. Bryan. 2019. Length-based risk analysis for assessing sustainability of data-limited tropical reef fisheries ICES Journal of Marine Science **76(1)**:165–180.

Ault, J. S., Smith, S. G., J. A. Bohnsack, J. Luo, N. Zurcher, D. B. McClellan, T. A. Ziegler, D. E. Hallac, M. Patterson, M. W. Feeley, B. I. Ruttenberg, J. Hunt, D. Kimball, and B. Causey. 2013. Assessing coral reef fish population and community changes in response to marine reserves in the Dry Tortugas, Florida, USA. Fisheries Research 144: 28–37.

Ault, J. S., S. G. Smith, J. A. Browder, W. Nuttle, E. C. Franklin, J. Luo, G. T. DiNardo, J. A. Bohnsack. 2014. Indicators for assessing the ecological dynamics and sustainability of southern Florida's coral reef and coastal fisheries. Ecological Indicators **44**:164–172.

Ault, J. S., S. G. Smith, J. Luo, M. E. Monaco, and R. S. Appeldoorn. 2008. Length-based assessment of sustainability benchmarks for coral reef fishes in Puerto Rico. Environmental Conservation **35**(3):221–231.

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. 185 pp.

Beverton, R. J. H. and S. J. Holt. 1957. On the dynamics of exploited fish populations. Fishery Investigation Series II, Ministry of Agriculture Fisheries and Food, London **XIX**:1–533.

Cochran, W. G. 1977. Sampling techniques, 3rd Ed. Wiley, New York.

Ehrhardt, N. M. and J. S. Ault. 1992. Analysis of 2 length-based mortality models applied to bounded catch length frequencies. Transactions of the American Fisheries Society **121**:115–122.

Hansen, M. H. 1987. Some history and reminiscences on survey sampling. Statistical Science **2(2)**:180–190.

Lohr, S.L. 2010. Sampling: design and analysis. 2nd ed. Brooks/Cole. Boston.

Neyman, J. (1934) On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection (with discussion 607-625). Journal Royal Statistical Society 97:558–606.

Pais, M. P., H. N. Cabral. 2017. Fish behaviour effects on the accuracy and precision of underwater visual census surveys. A virtual ecologist approach using an individual-based model. Ecological Modelling **346**: 58–69.

Pais M. P., H. N. Cabral. 2018. Effect of underwater visual survey methodology on bias and precision of fish counts: a simulation approach. PeerJ **6**:e5378 https://doi.org/10.7717/peerj.5378

Sokal, R. R. and F. J. Rohlf. 2001. *Biometry; the principles and practice of statistics in biological research*. 3rd edition. W.H. Freeman and Company, New York.

Zar, J. H. 1996. Biostatistical Analysis. 3rd edition. Simon & Schuster, Upper Saddle River, N.J.