Transects, quadrats, or points? What is the best combination to get a precise estimation of a coral community?

Luis M Montilla Corresp., 1 , **Emy Miyazawa** 1 , **Alfredo Ascanio** 1 , **María López-Hernández** 1 , **Gloria Mariño-Briceño** 1 , **Zlatka Rebolledo** 1 , **Andreína Rivera** 1 , **Daniela S. Mancilla** 1 , **Alejandra Verde** 1 , **Aldo Croquer** Corresp. 1

[1] Experimental Ecology Laboratory, Universidad Simón Bolívar, Caracas, Miranda, Venezuela

Corresponding Authors: Luis M Montilla, Aldo Croquer
Email address: luismmontilla@usb.ve, acroquer@usb.ve

The characteristics of coral reef sampling and monitoring are highly variable, with numbers of units and sampling effort varying from one study to another. Numerous works have been carried out to determine an appropriate effect size through statistical power, however, always from a univariate perspective. In this work, we aimed to assess the multivariate standard error of a series of reefs in Venezuela, sampled between 2017 and 2018, and also, to evaluate the consequences of using different combinations of points, quadrats, and transects over this error. For this, the multivariate standard error of 36 sites previously sampled was estimated, using four 30m-transects with 15 photo-quadrats each and 25 random points per quadrat. We obtained that the multivariate standard error was highly variable between sites and is not correlated with the univariate standard error nor with the richness of species. Then, a subset of sites were re-annotated using 100 uniformly distributed points, which allowed the simulation of different numbers of transects per site, quadrats per transect and points per quadrat using resampling techniques. The magnitude of the multivariate standard error stabilized by adding more transects, however, adding more quadrats or points does not improve the estimate. For this case study, when comparing between sampling with 10 transects, 10 quadrats per transect and 25 points per quadrat; and the original data for Venezuela, we find that the error is reduced by half. We recommend the use of multivariate standard error in reef monitoring programs, in particular when conducting pilot surveys to optimize the estimation of the community structure.

Transects, quadrats, or points? What is the best combination to get a precise estimation of a coral community?

Luis M. Montilla[1], **Emy Miyazawa**[1], **Alfredo Ascanio**[1], **María López-Hernández**[1], **Gloria Mariño-Briceño**[1], **Zlatka Rebolledo**[1], **Andreína Rivera**[1], **Daniela S. Mancilla**[1], **Alejandra Verde**[1], **and Aldo Cróquer**[1]

[1]**Experimental Ecology Laboratory, Simón Bolívar University, Venezuela**

Corresponding author:
Aldo Cróquer[1]

Email address: acroquer@usb.ve

ABSTRACT

The characteristics of coral reef sampling and monitoring are highly variable, with numbers of units and sampling effort varying from one study to another. Numerous works have been carried out to determine an appropriate effect size through statistical power, however, always from a univariate perspective. In this work, we aimed to assess the pseudo multivariate dissimilarity-based standard error (MultSE) of a series of reefs in Venezuela, sampled between 2017 and 2018, and also, to evaluate the consequences of using different combinations of points, quadrats, and transects over this error. For this, the MultSE of 36 sites previously sampled was estimated, using four 30m-transects with 15 photo-quadrats each and 25 random points per quadrat. We obtained that the MultSE was highly variable between sites and is not correlated with the univariate standard error nor with the richness of species. Then, a subset of sites was re-annotated using 100 uniformly distributed points, which allowed the simulation of different numbers of transects per site, quadrats per transect and points per quadrat using resampling techniques. The magnitude of the MultSE stabilized by adding more transects, however, adding more quadrats or points does not improve the estimate. For this case study, when comparing between sampling with 10 transects, 10 quadrats per transect and 25 points per quadrat; and the original data for Venezuela, we find that the error is reduced by half. We recommend the use of MultSE in reef monitoring programs, in particular when conducting pilot surveys to optimize the estimation of the community structure.

INTRODUCTION

The intrinsic value of coral reefs and their relevance in terms of services provided to human societies makes them an object of constant observation, however because some ecological processes operating in the reefs occur at large spatial and temporal scales, coral scientist face the challenge of obtaining data while keeping a compromise between high precision, reproducibility, and statistical power, with low cost and time (Aronson et al., 1994). Several technological advances have allowed a reduction in data variability derived from multiple human observers, e.g. the use of photo- and video-transects instead of *in situ* benthic characterization (Leujak and Ormond, 2007); the use of ROV's instead of divers (Lam et al., 2006); or the use of artificial intelligence to assist the annotation of photo-quadrats (Beijbom et al., 2015; Williams et al., 2019; González-Rivero et al., 2016).

Coral reef typical sampling and monitoring have usually relied on plot methods, like belt transects; plotless methods, like line-intercept method; or a combination like the point-intercept method. The choice and the sampling effort have varied from study to study or from program to program, e.g. Aronson et al. (1994) proposed ten 25m transects with about 50 quadrats containing 10 random points, for sampling spur-and-groove habitats. The CARICOMP protocol relied on the chain method to identify the substrate underneath each chain link on ten 10m transects (CARICOMP, 2001); while AGRRA protocol gets estimates from six 10m transects for shallow reefs (Lang et al., 2010). All these methods have had different amounts of popularity, cost/benefit ratios, and levels of precision associated to the estimation of coral cover (Nadon and Stirling, 2006; Leujak and Ormond, 2007).

Statistical power has received particular attention in this matter as a tool to decide when a set of conditions allowed the researchers to detect appropriate effect sizes (Aronson et al., 1994; Brown et al., 2004; Lam et al., 2006; Leujak and Ormond, 2007; Molloy et al., 2013; Houk and Van Woesik, 2006); which typically have been studied as univariate analysis of total or mean coral cover (or other particular substrates). This kind of criteria are used even if the research question is related to multivariate cases, which are particularly relevant because it is also important to understand changes occurring not only in the cover of the substrate, but also in the assemblage and functional structure of corals and other reef organisms (Wulff, 2001; Alvarez-Filip et al., 2013).

Anderson and Santana-Garcon (2015) proposed a multivariate approach for estimating the precision of sampling when it is of interest to perform a dissimilarity-based multivariate analysis: the **pseudo multivariate dissimilarity-based standard error (MultSE)**.

In this study we wanted to answer what values of multSE were achieved in a series of scleractinian coral assemblage surveys conducted in 36 sites of Venezuela?; and after knowing the reference values, what would be the effect of using different combinations of sampling strategies, including number of points, quadrats, and transects over the multSE? This with the aim of comparing the best sampling protocol and potentially improve precision for future samplings of these communities.

METHODS

We estimated the multivariate standard error (Anderson and Santana-Garcon, 2015)

We used a dataset from 36 previously surveyed sites on seven localities of Venezuela between 2017 and 2018 (detailed in Supp. 1) following a variation of the GCRMN monitoring protocol: instead of using five transects per site, we used four 30m transects (to increase the number of sampling sites), 15 photo-quadrats per transect placed one another meter, and 25 random points per quadrat (GCRMN, 2016). The photo-quadrats were annotated using Photoquad (Trygonis and Sini, 2012); this generated a matrix of 143 observations x 43 species.

We calculated the MultSE using the scripts available as supplementary material in Anderson and Santana-Garcon (2015), based on a Bray-Curtis dissimilarity matrix. Next, we chose the 12 sites with the highest interquantile range (5%-95%) of the MultSE and re-annotated the transects using 100 uniformly distributed points. This allowed us to create a new data set using resampling techniques accounting for different number of transects per site, quadrats per transect and points per quadrat.

Resampled datasets

Resampling techniques allowed us to create new data sets containing up to 20 simulated transects per site, with 5 to 25 quadrats per transect, and evaluating 25, 50, and 75 points per quadrat. We constructed a script to generate resampled data sets containing transects with all possible combination of number of transects per site, quadrats per transects, and points per quadrats desired. Broadly, from an input specifying the desired parameters:

1. Given the number of desired transects per site (*nbT*), the algorithm lists the available transects and adds repetitions of the same set of transects if *nbT* is greater than the actual available transects, until *nbT* is reached; for each simulated transect, the real transects factors and raw CSV file path are extracted.

2. To simulate the desired number of quadrats per simulated transect (*nbQ*), the algorithm reads each transect raw CSV file and, depending on the number of real quadrats (*NQ*) makes a sampling with (if $nbQ > NQ$) or without replacement (if $nbQ \leq NQ$) of the observed quadrats.

3. Finally, for points resampling within each simulated quadrat, the algorithm makes a random sampling without replacement, considering the desired number of points (*nbP*), which could not be higher than the observed number of points per quadrat (100 points per quadrat in this study). The algorithm flowchart can be observed in figure 1.

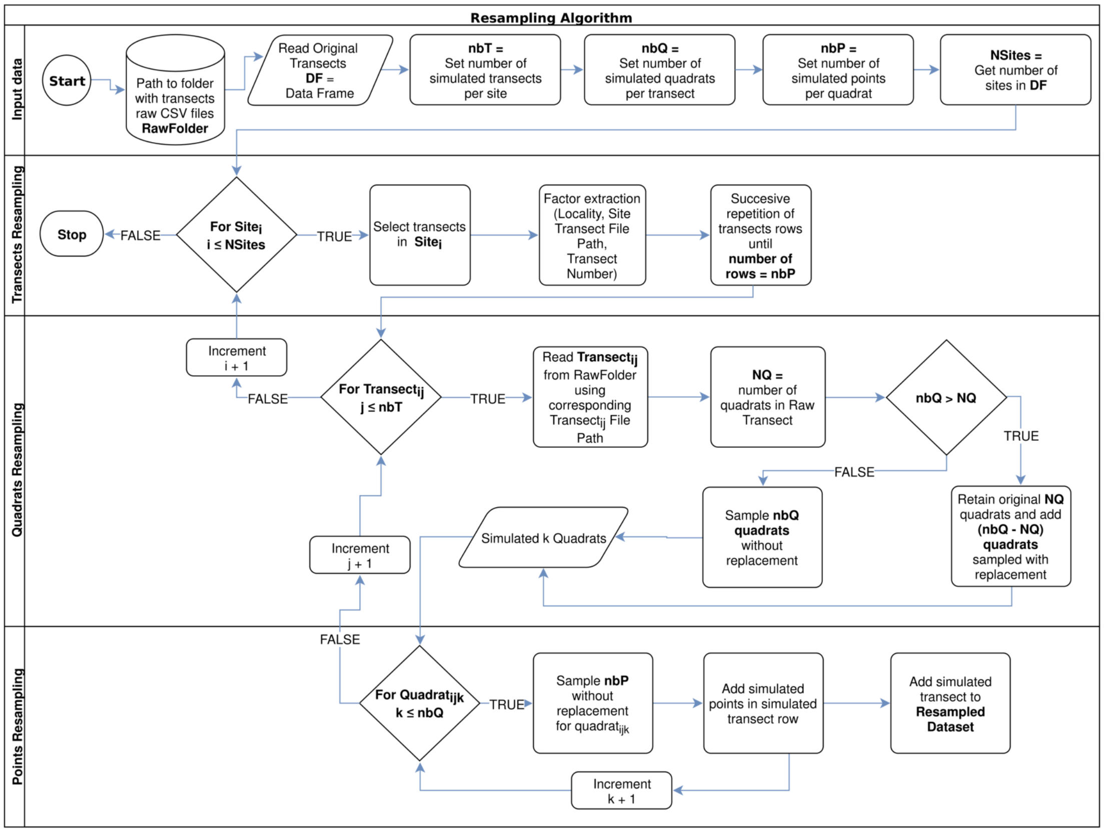
Resampling Algorithm
Input data
Start
Path to folder with transects raw CSV files RawFolder
Read Original Transects DF = Data Frame
nbT = Set number of simulated transects per site
nbQ = Set number of simulated quadrats per transect
nbP = Set number of simulated points per quadrat
NSites = Get number of sites in DF
Transects Resampling
Stop
FALSE
For Site_i i ≤ NSites
TRUE
Select transects in Site_i
Factor extraction (Locality, Site Transect File Path, Transect Number)
Succesive repetition of transects rows until number of rows = nbP
Quadrats Resampling
Increment i + 1
FALSE
For Transect_ij j ≤ nbT
TRUE
Read Transect_ij from RawFolder using corresponding Transect_ij File Path
NQ = number of quadrats in Raw Transect
nbQ > NQ
TRUE
FALSE
Sample nbQ quadrats without replacement
Retain original NQ quadrats and add (nbQ - NQ) quadrats sampled with replacement
Simulated k Quadrats
Increment j + 1
Points Resampling
FALSE
For Quadrat_ijk k ≤ nbQ
TRUE
Sample nbP without replacement for quadrat_ijk
Add simulated points in simulated transect row
Add simulated transect to Resampled Dataset
Increment k + 1

Figure 1. Resampling algorithm flowchart used to simulate new data per sampled site, according to a desired number of transects per site (nbT), number of quadrats per transect (nbQ), and number of points per quadrats (nbP).

Statistical analysis

We used a linear regression on Box-Cox-transformed data to test effect of each variable on the MultSE. We performed all the data manipulation and statistical analyses with R (R Core team, 2019). All our data and scripts are available as a research compendium at https://github.com/luismmontilla/coral_muse.

RESULTS

The MultSE of the field data was variable among sites and locations. Some locations had relatively uniform errors among all their sites, while other cases, like Los Roques and Mochima, included sites with largely different values; additionally, some of the highest multivariate error values had also associated the largest interquartile ranges (Fig. 2). These patterns in general were positively, but lowly correlated to the standard error of the mean coral cover and species richness (r = 0.14 and r = 0.18 respectively, Fig. 3). The mean value of our field data was 0.16 ± 0.03, however this value is purely referential for future studies using this metric under the same conditions.

As expected, the magnitude of the MultSE stabilized with the addition of more transects to the data set, however, analyzing more points per quadrat or using more or fewer quadrats didn't improve considerably the estimation (Fig. 4). The linear regression confirmed this, showing that an increase in the number of transects is the best approach to achieve lower error values ($Estimate = -0.26$, $t = -23.72$, $p = 2x10^{-16}$). Despite all the sources of variation having low p-values, the largest reduction in the estimator was observed for transects alone. Unexpectedly, increasing the number of points and quadrats seem to slightly increase the error ($Estimate = 0.029$, $t = 3.47$, $p = 5.26x10^{-4}$, Fig. 5). Selecting a value of ten transects, the value of the MultSE stabilizes near 0.1, representing a reduction of 0.05 units from the average value.

Considering the result of the regression, we used a comparison of ten transects, ten quadrats per transects, and 25 points per quadrat to compare against the original data using the selected resampled sites.

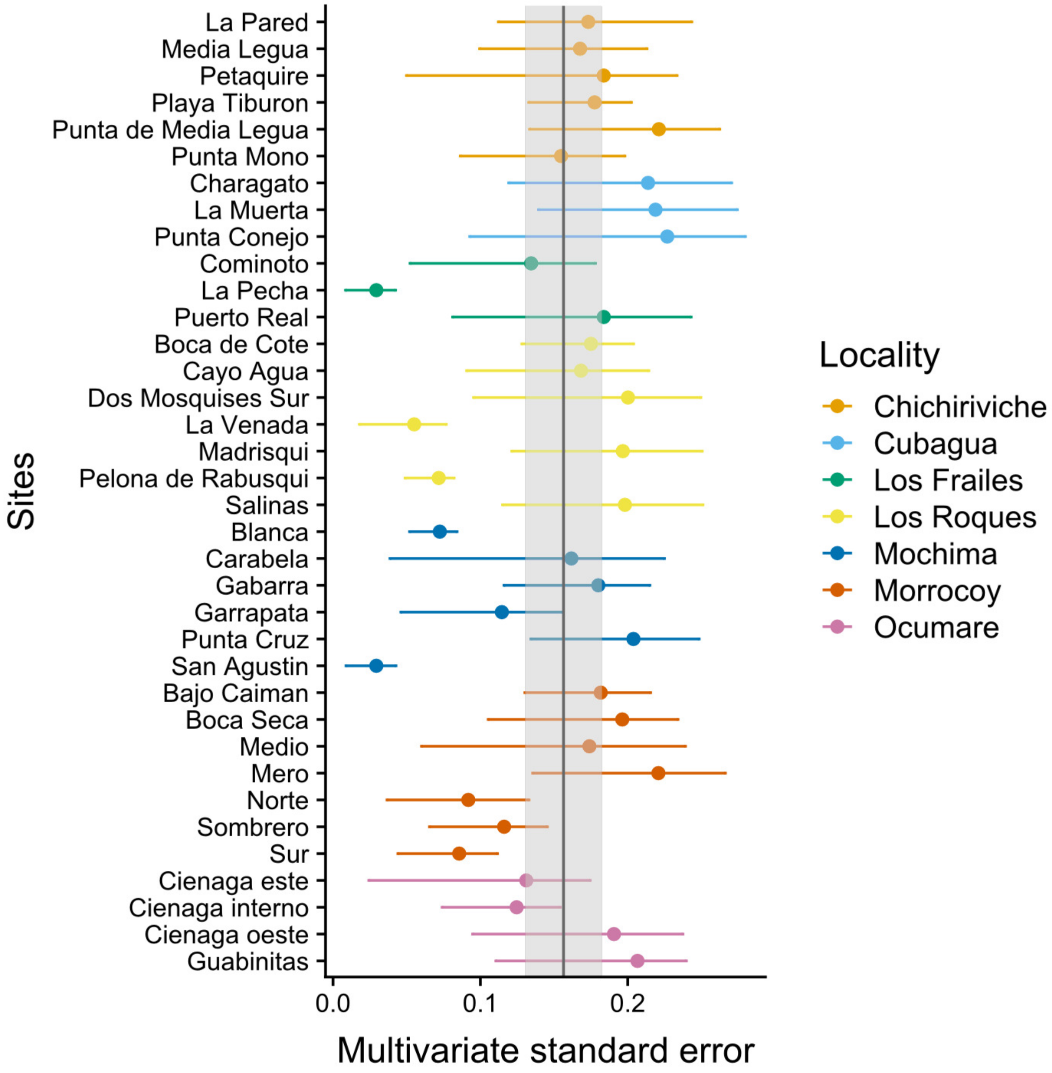

Figure 2. Achieved MultSE for each site. Error bars represent 2.5 and 97.5 percentiles. The vertical band represents the mean±se.

In this case, the new error was about half the original estimation in the field (for the subset of sites) using four transects with 15 quadrats per transect (Fig. 6). Furthermore, the inclusion of more transects had an effect on the multivariate ordination of the sites; a principal coordinates analysis with both sampling schemes showed that using ten transects reduced the centroids standard deviation overlap, potentially making easier to discriminate between actual different coral assemblages in these sites (Supp. 2).

DISCUSSION

Here we evaluated the potential of the pseudo multivariate dissimilarity-based standard error as a tool to determine the appropriate number of transects to sample coral assemblages. Though this method is still under development and has its constrains (Anderson and Santana-Garcon, 2015), it also provides additional information often overlooked when the research question is related to studying assemblage patterns. We suggest that using at least ten transects instead of four (given that the same dissimilarity measure is used) improves the estimation of the coral assemblages from Venezuela; this number can be partially compensated by reducing the number of quadrats per transect from 15 to 10 quadrats (i.e. 20m transects), and keeping 25 random points per quadrat. To implement this scenario, there are some additional costs that would have to be considered, for example, this would imply increasing the time

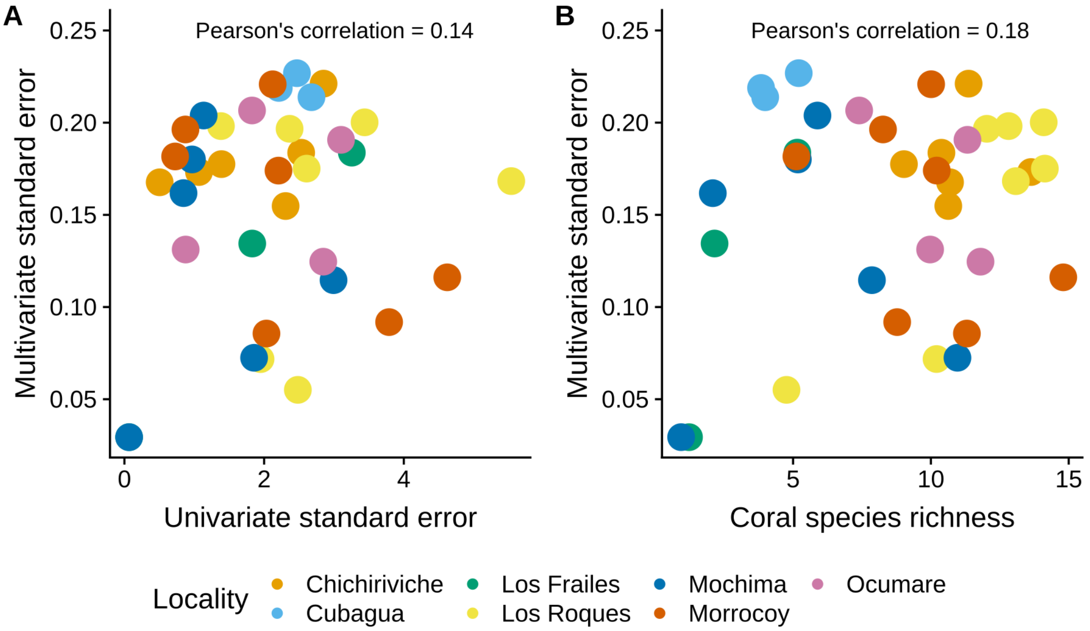

Figure 3. Comparison of the MultSE with A) the standard error of mean cover by site, and B) coral species richness.

spent in the field for transect deployment, but this also could be compensated in successive surveys if the researchers use fixed-transects (Molloy et al., 2013).

Additionally, for our proposed scheme, the reduction in the number of photoquadrats per transect still results in a net increase of photoquadrats to be analyzed per site, but the impact of this specific step would depend on the size and training of the identification team, the required resolution of the data according to the question, and the tools used to generate the dataset. In any case, it is important to highlight that in different locations in would be necessary to implement a similar approach to our proposal here and verify the required number of transects necessary to improve the estimation of the coral assemblage.

Our results also coincide with the findings of Molloy et al. (2013), about the importance of increasing the number of transects instead of quadrats and/or points per quadrat. However, our own comparison indicated that univariate precision of coral cover is uncorrelated from MultSE, contributing to the idea that this estimation reveals valuable information, specially if the research question is related to compare coral assemblages. We also found that spatial variability had an effect on the MultSE; at the scale of sites that we used –separated by hundred of meters– there were notable differences at the achieved precision, making more necessary the evaluation of the appropriate number of transects to be used. The behaviour of this metric at larger scales still remains to be explored, just as the result on more biodiverse localities e.g. the Indo-Pacific in contrast with the Caribbean.

One example of a practical use of this tool is its incorporation into monitoring programs right at the beginning, assessing the precision of a set of the same number of transects for all the sites; with the results, the researchers can consider redistribute the sampling effort to increase the precision where needed. This can also be performed under an adaptive perspective *sensu* Lindenmayer and Likens (2009); introducing new questions about the coral community would require to take data about coral species instead of only coral cover, keeping the same logistics in the field. In this case, the MultSE can be used to assess the precision of the estimation of the community over time and increase the number of transects, or redistribute the sampling efforts if necessary, and in general, to assess other monitoring tools that rely on multivariate data like multivariate control charts (Anderson and Thompson, 2004) can benefit of this approximation.

To conclude, monitoring program normally put emphasis on fulfilling three pillars: (1) standardization of their sampling protocols, (2) intense training to reduce observer bias and (3) unification of sampling effort. While the practical value of unification is indisputable, we recommend the evaluation of MultSE

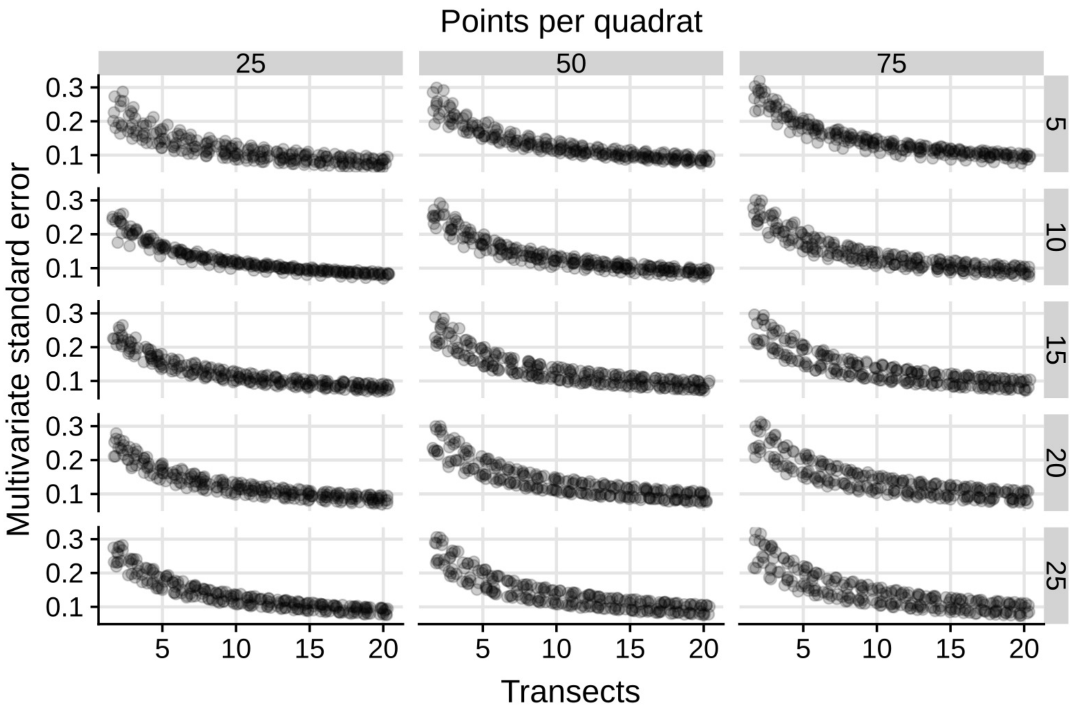

Figure 4. Multivariate standard error for a combination of different number of quadrats (rows), points per quadrat (columns), and transects (x-axis).

in coral reef monitoring programs, especially in pilot surveys to optimize the estimation of the coral assemblage across Caribbean geographies.

ACKNOWLEDGMENTS

We wish to thank Dr. Rita Peachey, the 39th AMLC scientific meeting organizing committee, and the Waitt Foundation.

REFERENCES

Alvarez-Filip, L., Carricart-Ganivet, J. P., Horta-Puga, G., and Iglesias-Prieto, R. (2013). Shifts in coral-assemblage composition do not ensure persistence of reef functionality. *Scientific reports*, 3:3486.

Anderson, M. J. and Santana-Garcon, J. (2015). Measures of precision for dissimilarity-based multivariate analysis of ecological communities. *Ecology letters*, 18(1):66–73.

Anderson, M. J. and Thompson, A. A. (2004). Multivariate control charts for ecological and environmental monitoring. *Ecological Applications*, 14(6):1921–1935.

Aronson, R. B., Edmunds, P. J., Precht, W. F., Swanson, D. W., and Levitan, D. R. (1994). Large-scale, long-term monitoring of Caribbean coral reefs: Simple, quick, inexpensive techniques. *Atoll Research Bulletin*, 421(421):1–19.

Beijbom, O., Edmunds, P. J., Roelfsema, C., Smith, J., Kline, D. I., Neal, B. P., Dunlap, M. J., Moriarty, V., Fan, T. Y., Tan, C. J., Chan, S., Treibitz, T., Gamst, A., Mitchell, B. G., and Kriegman, D. (2015). Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation. *PLoS ONE*, 10(7):1–22.

Brown, E. K., Cox, E., Jokiel, P. L. P. L., Rodgers, S. K., Smith, W. R., Tissot, B. N., Coles, S. L. S. L., and Hultquist, J. (2004). Development of Benthic Sampling Methods for the Coral Reef Assessment and Monitoring Program (CRAMP) in Hawai'i. *Pacific Science*, 58(2):145–158.

CARICOMP (2001). Caribbean Coastal Marine Productivity (CARICOMP) Methods Manual: Levels 1 and 2. Manual of Methods for Mapping and Monitoring of Physical and Biological Parameters in the Coastal Zone of the Caribbean. Technical report, Kingston.

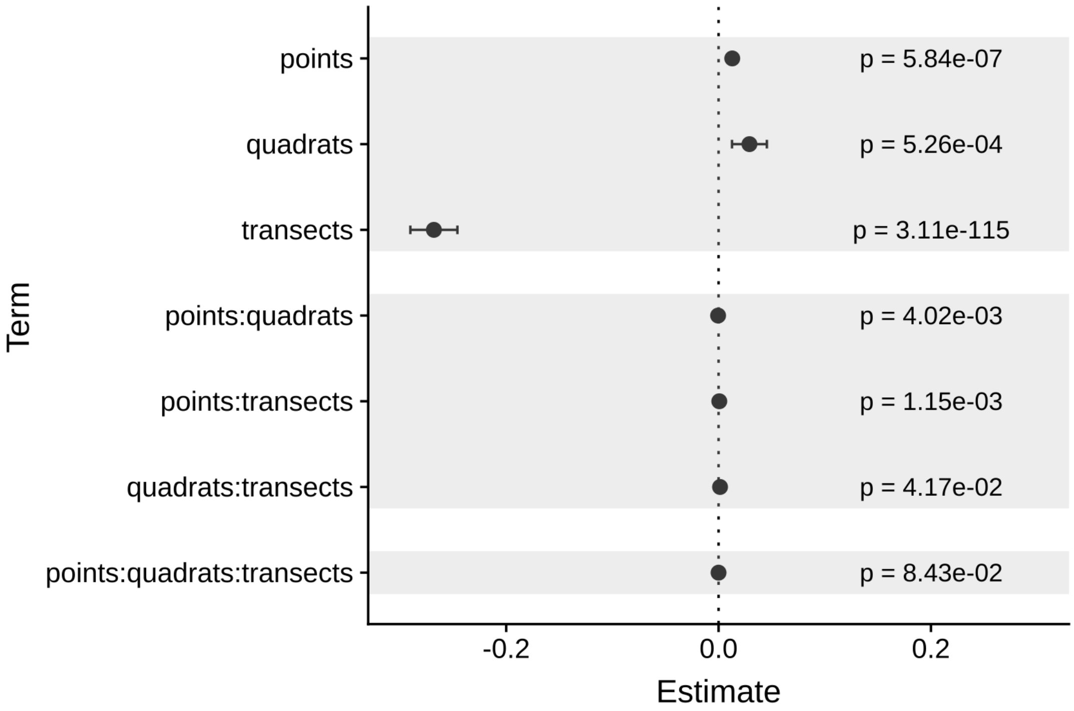

Figure 5. Coefficients of each source of variation for the linear regression of MultSE. Negative values imply that an increase of a unit in the respective source of variation, reduces the value of the multivariate standard error.

GCRMN (2016). GCRMN-Caribbean guidelines for coral reef biophysical monitoring. Technical Report November, United Nations Environment Programme, Miami, USA.

González-Rivero, M., Beijbom, O., Rodriguez-Ramirez, A., Holtrop, T., González-Marrero, Y., Ganase, A., Roelfsema, C., Phinn, S., and Hoegh-Guldberg, O. (2016). Scaling up Ecological Measurements of Coral Reefs Using Semi-Automated Field Image Collection and Analysis. *Remote Sensing*, 8(1):30.

Houk, P. and Van Woesik, R. (2006). Coral Reef Benthic Video Surveys Facilitate Long-Term Monitoring in the Commonwealth of the Northern Mariana Islands: Toward an Optimal Sampling Strategy. *Pacific Science*, 60(2):177–189.

Lam, K., Shin, P. K., Bradbeer, R., Randall, D., Ku, K. K., Hodgson, P., and Cheung, S. G. (2006). A comparison of video and point intercept transect methods for monitoring subtropical coral communities. *Journal of Experimental Marine Biology and Ecology*, 333(1):115–128.

Lang, J., Marks, K., Kramer, P. A., Kramer, P. R., and Ginsburg, R. (2010). AGRRA protocols v 5.4. Technical report.

Leujak, W. and Ormond, R. F. G. (2007). Comparative accuracy and efficiency of six coral community survey methods. *Journal of Experimental Marine Biology and Ecology*, 351(1-2):168–187.

Lindenmayer, D. B. and Likens, G. E. (2009). Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in Ecology and Evolution*, 24(9):482–486.

Molloy, P. P., Evanson, M., Nellas, A. C., Rist, J. L., Marcus, J. E., Koldewey, H. J., and Vincent, A. C. (2013). How much sampling does it take to detect trends in coral-reef habitat using photoquadrat surveys? *Aquatic Conservation: Marine and Freshwater Ecosystems*, 23(6):820–837.

Nadon, M. O. and Stirling, G. (2006). Field and simulation analyses of visual methods for sampling coral cover. *Coral Reefs*, 25(2):177–185.

R Core team (2019). *R: A language environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Trygonis, V. and Sini, M. (2012). PhotoQuad: A dedicated seabed image processing software, and a comparative error analysis of four photoquadrat methods. *Journal of Experimental Marine Biology and Ecology*, 424-425:99–108.

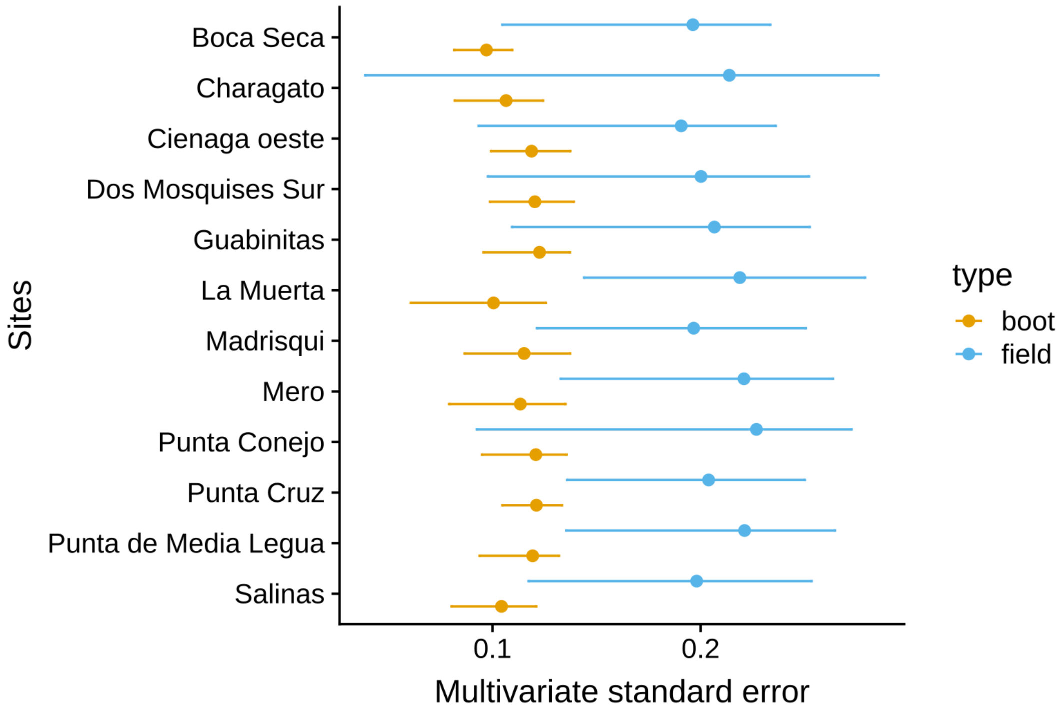

Figure 6. Comparison of the MultSE for a combination of ten transects, ten quadrats, and 25 random points per quadrat and original sampling scheme.

Williams, I. D., Couch, C. S., Beijbom, O., Oliver, T. A., Vargas-Angel, B., Schumacher, B. D., and Brainard, R. E. (2019). Leveraging Automated Image Analysis Tools to Transform Our Capacity to Assess Status and Trends of Coral Reefs. *Frontiers in Marine Science*, 6(April):1–14.

Wulff, J. (2001). Assessing and monitoring coral reef sponges: Why and how? *Bulletin of Marine Science*, 69(2):831–846.