Comparing Coral Demographic Surveys From In Situ Observations and Structure-From-Motion Imagery Shows Low Methodological Bias

Courtney Couch[1,2], Rhonda Suka[1,2], Thomas Oliver[2], Mia Lamirand[1,2], Mollie Asbury[2], Corinne Amir[1,2], Bernardo Vargas-Angel[1,2], Morgan Winston[1,2], Brittany Huntington[1,2], Frances Lichowski[1,2], Ariel Halperin[1,2], Andrew Gray[1,2], Joao Garriques[1,2], Raymond Boland[2], Noah Pomeroy[1,2], Jennifer Samson[2]

[1] Joint Institute for Marine and Atmospheric Research
University of Hawaii
1000 Pope Road
Honolulu, Hawaii 96822

[2] Pacific Islands Fisheries Science Center
National Marine Fisheries Service
1845 Wasp Boulevard
Honolulu, HI 96818

January 2021

NOAA Administrative Report H-21-01
https://doi.org/10.25923/307k-fk90

About this report

Pacific Islands Fisheries Science Center administrative reports are issued to promptly disseminate scientific and technical information to marine resource managers, scientists, and the general public. Their contents cover a range of topics, including biological and economic research, stock assessment, trends in fisheries, and other subjects. Administrative reports typically have not been reviewed outside the Center; therefore, they are considered informal publications. The material presented in administrative reports may later be published in the formal scientific literature after more rigorous verification, editing, and peer review.

Other publications are free to cite PIFSC administrative reports as they wish, provided the informal nature of the contents is clearly indicated and proper credit is given to the author(s).

Recommended citation

Couch C, Suka R, Oliver T, Lamirand M, Asbury M, Amir C, Vargas-Angel B, Winston M, Huntington B, Lichowski F, Halperin A, Gray A, Garriques J, Boland R, Pomeroy N, Samson J. 2021. Comparing coral demographic surveys from in situ observations and structure-from-motion imagery shows low methodological bias. PIFSC Administrative Report, H-21-01, 61 p. doi:10.25923/307k-fk90

Copies of this report are available from

Pacific Islands Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
1845 Wasp Boulevard, Building #176
Honolulu, Hawaii 96818

Or online at

https://repository.library.noaa.gov/

Contents

List of Tables

List of Figures

Executive Summary

NOAA's Ecosystem Sciences Division (ESD) at the Pacific Islands Fisheries Science Center (PIFSC) executes the NOAA Coral Reef Conservation Program's National Coral Reef Monitoring Program (NCRMP) across 40 primary islands, atolls, and shallow banks in the U.S. Pacific. Through funding from CRCP, ESD works to understand how reefs are responding, in both time and space, to local and global disturbances, and what factors enhance reef resilience.

Assessments of benthic cover and coral demography are key components of ESD's Pacific-NCRMP and have historically been conducted using in situ visual surveys and photoquadrats by a team of three to four divers. With NOAA's aging vessel infrastructure and increasing complexities of field missions, ESD is exploring innovative technologies to monitor more efficiently benthic communities. One approach gaining traction in reef sciences is photogrammetry, also known as Structure from Motion (SfM), which generates 3D reconstructions of a scene from overlapping images. SfM is a low-cost method, compared to techniques such as light detection and ranging (LIDAR). Here, we quantitatively compare data generated from in situ surveys to SfM-derived metrics for assessing coral demography to evaluate whether ESD can maintain continuity in our long-term NCRMP data sets if the program transitions to SfM survey methods.

The four objectives of this study are the following:

1) compare error between methods to within method observer error,
2) test for methodological bias between SfM and in situ visual surveys, and
3) identify strengths and weaknesses and compare costs of both methods, and
4) identify steps needed to transition to SfM for future benthic monitoring.

To determine whether the level of variability between methods is comparable to the normal variability between observers, multiple divers and SfM annotators surveyed the same 43 benthic plots. In situ surveys and SfM imagery collection were conducted during the 2019 main Hawaiian Islands NCRMP mission. Colony density, average colony length, average partial mortality, and prevalence of disease and compromised health states were recorded for each survey and method type. Our results suggest that the level of variability between methods is comparable to the variability between divers underwater for all metrics and SfM annotator variability was lower than diver variability for both estimates of colony length and colony density. To test the comparability between methods across a broader range of habitats, depths and conditions, one diver and one SfM annotator surveyed the same 104 sites. Overall, our results suggest that a majority of the metrics do not vary significantly between methodological approaches nor between methods across habitats and depths. Metrics derived from in situ and SfM surveys differed in estimates of partial mortality, disease and bleaching prevalence, though simple improvements in image quality for SfM, continued training of annotators to improve consistency of these challenging metrics and summarizing data at scales larger than site will likely yield better alignment between the two methods.

As ESD and NCRMP explore transitioning from in situ surveys to SfM, we will need to weigh the variety of strengths and weaknesses of each method, as well as differences in personnel and equipment costs. Monitoring coral demography with SfM provides a number of advantages over in situ methods such as reduced survey effort, finer-scale and more accurate measurements of structural complexity, and the ability to re-evaluate SfM imagery to control observer error. However, several limitations should be considered before investing in this approach such as the following: the considerable amount of time to extract data relative to in situ methods, the risk of taking poor quality imagery leading to poor annotation, and SfM's inability to capture all of the crevices and overhangs that can be observed underwater.

The primary difference in costs between these two methods is in personnel time. SfM methods for raw data collection in the field are more cost-effective, reducing both field time and field staff costs. However, SfM requires substantial post-processing costs absent from in situ methods. Post-modeling annotation of SfM imagery results in a 34% increase in personnel costs to meet NCRMP's data requirements. SfM currently also requires an estimated hardware investment of ~$120,000 every 5 years.

If ESD transitions to SfM for NCRMP benthic monitoring, a hybrid approach should be considered. Such an approach would involve extracting metrics such as density, colony size and old partial mortality from SfM imagery, while recording recent dead and incidences of disease, bleaching and other compromised health states using in situ observations. This would leverage the strengths of both methods while still reducing field costs. ESD should continue to reduce inter-diver and inter-annotator variability through rigorous training. Improved machine learning and automated processing methods would also further reduce long-term costs associated with manual annotation.

Introduction

As the threats to coral reefs mount, scientists and managers are looking for innovative ways to increase the scope, scale, and efficiency of coral reef monitoring. Monitoring changes in coral communities provides key information about ecosystem function and resilience of reefs. Over the last 5 decades, coral reef ecologists have utilized a broad range of in situ visual and imaging survey methods to quantify various aspects of coral demography and benthic features. Historically, visual methods such as line point intercept (Choat and Bellwood 1985), point intercept transect (English et al. 1994), quadrat (Conand et al. 1999), timed swims (Donnelly et al. 2003), and belt transect (Connell et al. 1997) have been widely used. While these surveys allow divers to leave the water with data in hand, they can be time consuming underwater, require specialized training, and visual observations made by a single diver cannot be verified or re-evaluated. Video transects and photoquadrats (English et al. 1994) have become more common in the last two decades with the increased accessibility of digital cameras. These methods are more efficient underwater and require less specialized skills, but necessitate significant post processing and are typically only used to quantify benthic cover. However, a single underwater photograph only captures a small area of reef in two dimensions and does not allow for accurate measurements of benthic features.

NOAA's Ecosystem Sciences Division (ESD) in the Pacific Islands Fisheries Science Center (PIFSC) is funded by the NOAA Coral Reef Conservation Program (CRCP) to monitor the status and trends of coral reefs across 40 Pacific islands, atolls, and shallow banks as part of the National Coral Reef Monitoring Program (NCRMP). The NCRMP surveys reefs in the Hawaiian Archipelago (including Papahānaumokuākea Marine National Monument), the Mariana Archipelago (Guam and the Commonwealth of the Northern Mariana Islands, including the Marianas Trench Marine National Monument), American Samoa (including Rose Atoll Marine National Monument), and the Pacific Remote Islands Marine National Monument (Figure 1). Since 2000, ESD has used a variety of benthic methods, including line point intercept (NOAA Pacific Islands Fisheries Science Center 2020a), benthic photo-quadrat (NOAA Pacific Islands Fisheries Science Center 2020b), and belt transect surveys (NOAA Pacific Islands Fisheries Science Center 2015) to quantify coral demography and benthic cover. Each year ESD visits 400–600 sites, which is a huge undertaking, requiring a large infrastructure and complex organization to complete. ESD relies heavily on the small fleet of NOAA's seagoing vessels, which are heavily tasked and nearing the end of their operational lifespan. This continued demand on increasingly limited resources requires innovative solutions for monitoring benthic communities with reduced field teams.

Figure 1. Map of geographic areas (in white) where ESD conducts coral reef monitoring surveys in the Pacific.

An emerging photogrammetry technology called Structure-from-Motion (SfM) is gaining popularity in coral reef research (Burns et al. 2015; Casella et al. 2017; Bayley et al. 2019; Fox et al. 2019; Obura et al. 2019) and offers a potential opportunity to continue collecting coral demographic metrics such as density, size structure, health, and partial mortality while reducing field costs. SfM uses overlapping 2D imagery to create a 3D point cloud. Each image pixel value is incorporated into the point cloud surface to generate a 2D mosaic image. This technique allows researchers to study the reef from the polyp to reef-level. To date, the majority of studies on coral reefs utilizing SfM methodology has focused on quantifying structural complexity (Figueira et al. 2015; Burns et al. 2015; Storlazzi et al. 2016; Ferrari et al. 2017; Fukunaga et al. 2020; Torres-Pulliza et al. 2020). Others have used SfM in small-scale studies to quantity disease and bleaching (Palma 2016; Fox et al. 2019; Voss et al. 2019; Burns et al. 2020), quantify spatial clustering of corals (Edwards et al. 2017; Pedersen et al. 2019), measure coral growth (Lange and Perry 2020), and quantify size frequency distributions (Hernández-Landa RC and R. 2020). However, no one has tested whether the standard suite of demographic metrics (e.g., density, colony size, partial mortality, and prevalence of altered health states) extracted from SfM imagery are consistent with in situ observations across large gradients of community structure, depth, and reef complexity.

ESD's SfM journey has been several years in the making. The program first became involved with SfM in 2013 when Dr. Stuart Sandin's research team from the Scripps Institution of Oceanography began conducting SfM surveys at fixed sites during the Pacific NCRMP cruises.[1] With additional guidance from Scripps colleagues and Dr. John Burns from the University of Hawaii at Hilo, ESD started actively exploring the possibility of integrating SfM into the stratified random sampling design, with a long-term goal of eventually replacing the standard in situ benthic surveys. In 2019, after a year of methods development and standardizing procedures (Suka et al. 2019), SfM surveys were conducted at all benthic sites during the main Hawaiian

[1] Part of the 100 Island Challenge project (http://100islandchallenge.org/overview/).

Islands NCRMP mission alongside in situ surveys. To ensure data continuity, SfM was implemented in a manner that replicated the in situ survey approach.

The **four objectives** of this study were to:

1) compare error between methods to within method observer error,
2) test for methodological bias between SfM and in situ visual surveys, and
3) identify strengths and weaknesses as well as compare costs of both methods, and
4) identify steps needed if ESD decides to transition to SfM for future NCRMP benthic monitoring, while maintaining continuity with long-term data sets.

Methods

To incorporate SfM into the existing stratified random sampling design and replicate the in situ visual surveys, we conducted a series of pilot surveys. These surveys allowed us to determine the best equipment for image collection (Appendix A & B), optimal size and shape of the survey area (Appendix C), optimal survey time needed (Appendix D), optimal camera position (Appendix E) and the impact of environmental conditions on image collection (Appendix F). Through this process, we developed a robust protocol for image collection that also met the constraints of mission logistics (Suka et al. 2019).

In situ collection

From April through July 2019, ESD conducted 180 NCRMP benthic surveys across the main Hawaiian Islands. Site selection was based on a one-stage stratified random survey design stratified by depth bin (shallow: 0–6 m; mid-depth: 6–18 m; and deep: 18–30 m) and sub-island sectors. At each site, one 18-m transect line was deployed along the associated depth contour. Visual observations were recorded using standard NCRMP methods (See Winston et al. 2020 for further details on methods) within each of four segments, each 1-m wide and spaced at 2.5 intervals along the transect (at 0–2.5 m, 5–7.5 m, 10–12.5 m, and 15–17.5 m). Staggered segments were surveyed to better capture spatial variability across the transect line (see details on how data were summarized in the Data Analysis section). Only three segments were surveyed at deep sites due to bottom time limitations. For each adult coral colony, maximum diameter, ID (to lowest taxonomic level), morphology, and percent old (denuded skeleton colonized by turf or other organisms) and recent (recently denuded skeleton not yet colonized) mortality were recorded. Divers also recorded any disease or condition, taking note of the type of disease and the colony surface area affected by disease. Bleaching severity was also recorded (1 = just starting to lose pigmentation, 2 = significant pigmentation loss, 3 = full loss of pigmentation). Juvenile coral colonies (< 5 cm max diameter) were recorded within the first 1 × 1 m portion of the first 3 segments due to limited survey time. For juvenile colonies, ID, and max diameter were recorded (Winston et al. 2020). Of the 180 sites, only 104 were used for the methodological comparison due to the lengthy SfM annotation process (Figure 2). This subset of sites represents the broadest range of islands, depths, habitat complexities, coral cover, and diving conditions. Within these 104 sites, 43 haphazardly chosen segments across 28 transects/sites were re-surveyed by a different diver during the same dive to create a replicate in situ observation for the error comparison analysis (Figure 2).

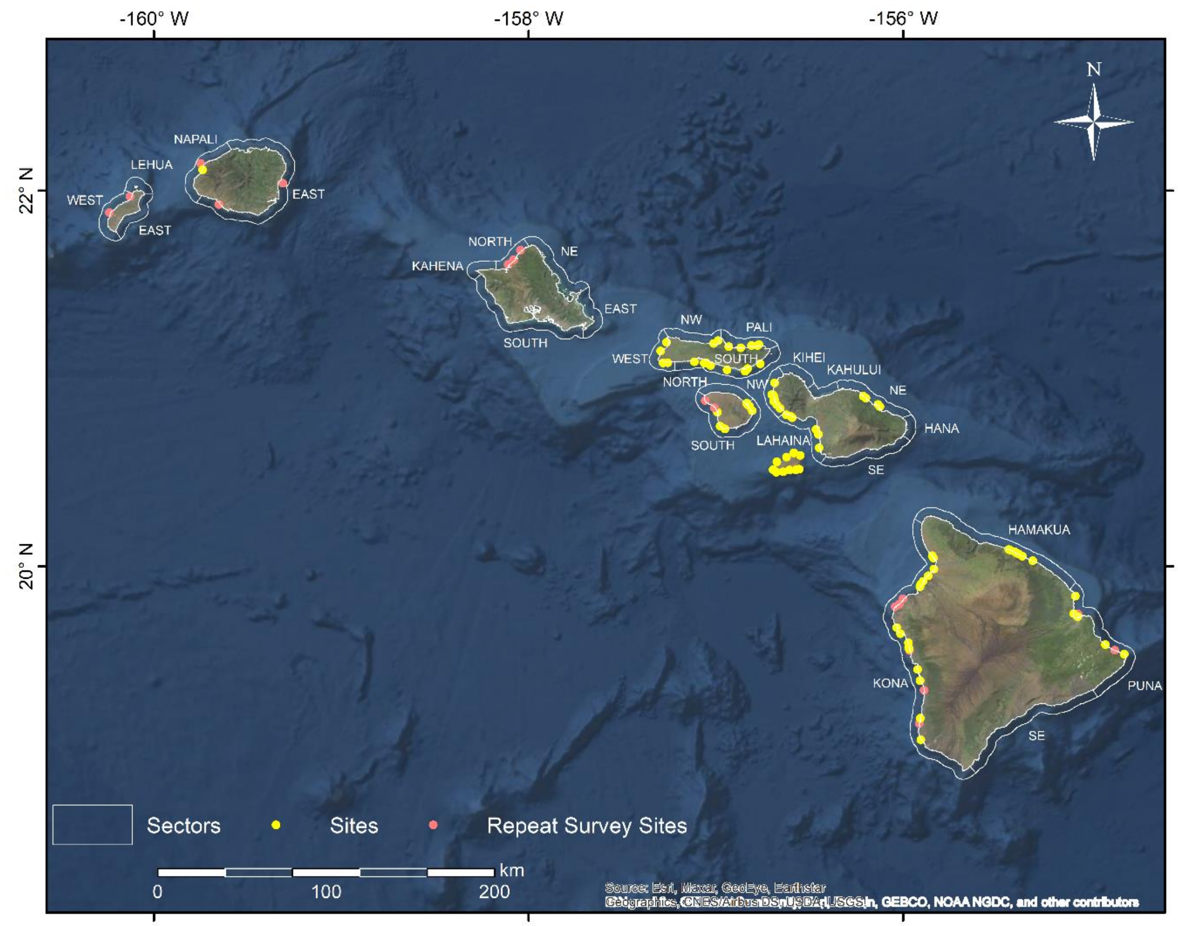

Figure 2. Location of 104 survey sites (yellow), sites with 28 repeated surveys indicated in orange and white outlines indicating sub-island sectors.

SfM image collection

In conjunction with the standard in situ visual surveys, SfM image collection was also conducted at each site. Scale bar markers, also known as Ground Control Points (GCPs), were placed at the beginning of each segment at least 0.5 m away from the transect line. The depth and position along the transect line of each GCP was recorded on a datasheet. The SfM survey was conducted over a 3 × 20 m area with the transect running down the middle at depths of 1–20 m (Figure 3) and a 3 × 13 m area at depths >20 m. This imaged area allowed us to capture the 3–4 segments discussed above as well and an adequate buffer around the segments to ensure that colonies that extended outside the segments were fully captured in the imagery. JPEG images were taken continuously by using an entry-level digital SLR camera (Canon EOS Rebel SL2, Ikelite underwater housing with 6” dome port) with an 18–55 mm lens set at 18 mm after color balancing with an 18% grey card (Appendix A) along the transect to the 20-m mark. Swimming back and forth with 0.5-m spacing between passes, while maintaining a 1-m distance from the seafloor, the photographer swam three passes on each side of the transect (for a total of six passes) to produce the total image area. This swim pattern allowed for ~60% side overlap and

~80% forward overlap of images. Imagery was manually QC'd to ensure only quality imagery (e.g., no overexposed or blue imagery) were included in the models.

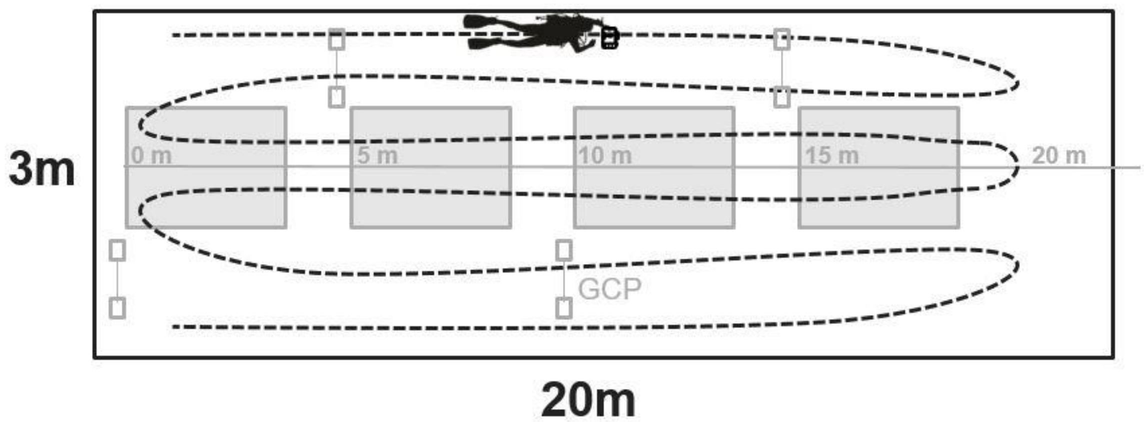

Figure 3. Graphic of benthic survey plot with in situ visual survey segments (in gray). The SfM swim path is indicated by the black dashed line, covering a 3 m x 20 m area.

SfM model generation and data extraction

A 3D model from each site was generated using Agisoft Metashape software (AgiSoft Metashape Professional Version 1.6.1). The workflow sequence included aligning images, building, and exporting the 3D dense point cloud (DPC) following parameters described by Burns et al. (2015). The DPC was then brought into Viscore software (Petrovic et al. 2014; Naughton et al. 2015) and then scaled and oriented using the GCP information. A geometrically accurate 2D projection of the DPC (hereafter referred to as an orthoprojection) and a scale grid were exported from Viscore and were then brought into ArcMap 10.6.1 for manual colony annotation. See (Suka et al. 2019) for detailed information.

In ArcMap, each site was set up for annotation by defining the ratio of the scale of the orthoprojection using the scale grid, manually digitizing the transect and segments as a shapefile, and setting up the attribute table in a geodatabase to mirror the in situ visual survey database. During annotation, the original JPEG imagery was viewed alongside the orthoprojection with the Viscore Image View feature to see fine scale colony details and observe colonies from multiple angles. Annotators were encouraged to speak with each other during the annotation process.

To record and extract data from the orthoprojection image, each coral colony within each segment was annotated following the in situ visual survey methods. Each colony was measured by digitizing a line across the maximum diameter of the colony. Coral ID (to lowest taxonomic level), morphology, disease/condition, and percent estimates of old and recent mortality were recorded. Annotators relied on the underlying original imagery to not only identify colonies and conditions, but also locate colonies not visible in the orthoprojection (e.g., under ledges; Figure 4A, B). Due to the 2D and grainy nature of the orthoprojection, using the underlying imagery was crucial to achieving the highest data quality and comparability to in situ surveys. In addition to this standard annotation, a subset of segments that were replicated in situ by divers were also annotated twice by different SfM annotators to create replicate data sets for both methods.

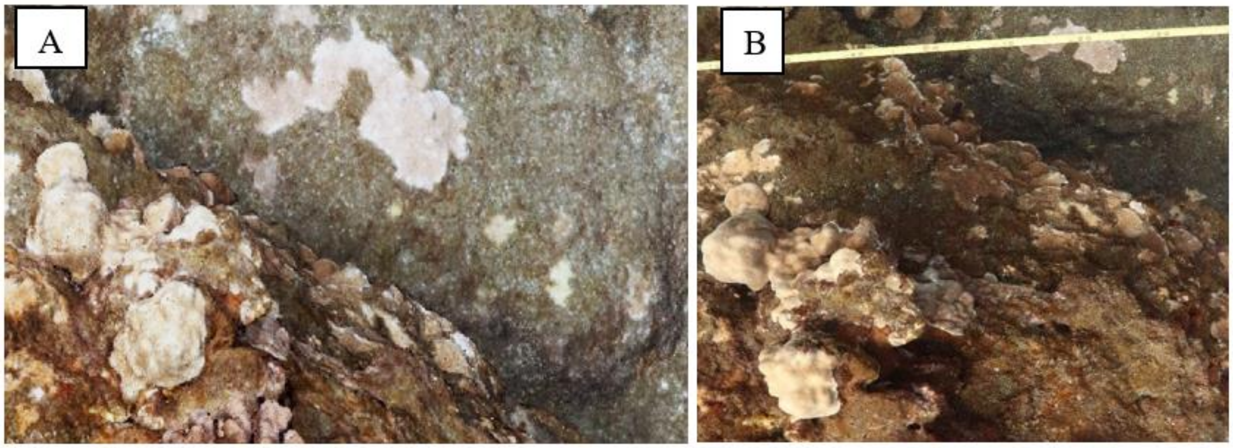

Figure 4. Orthoprojection image (A) and original image from camera (B) . Coral colonies near the center of the image are obscured in the orthoprojection, but visible in the original image.

Quality control of the extracted SfM data was carried out in four stages.

I. QC Script: R scripts with specific queries were used to identify data entry errors (e.g., misspelled species names, data in the incorrect column) and were corrected in the ArcGIS geodatabase.
II. Spot Check: Site-level metrics were examined across annotators to identify potential issues for annotators to spot check and correct. If persistent issues were identified for a given annotator (e.g. not identifying specific conditions), that annotator reviewed each site they annotated to correct these issues.
III. Crosscheck: Annotators reviewed a subset (10% of the annotated segments) of randomly selected segments (stratified by annotator) that they did not originally annotate to establish individual annotator error rates for each metric. Errors were recorded in a separate csv file, but not corrected in the main ArcGIS geodatabase. Errors were not corrected because this step was intended to be fast and efficient error-check process.
IV. Belt corrections: If an annotator had ≥10% error rate for a given metric another annotator reviewed all of that annotator’s sites and corrected the metric in question.

Data analysis

Eight demographic metrics were summarized as follows:

- adult colony density (number of colonies ≥5 cm per m^2),
- juvenile colony density (number of colonies 0.7 to 5 cm per m^2),
- average maximum adult colony diameter,
- average percent old partial mortality,
- average percent recent mortality,
- prevalence of acute diseases (percent of colonies with acute disease),

- chronic disease prevalence (percent of colonies with chronic disease),
- and bleaching prevalence (percent of colonies with a bleaching severity >2).

Acute disease is defined as diseases such as tissue loss syndrome and white syndrome that cause progressive morality of coral tissue. Chronic diseases include skeletal growth anomalies and fungal infection that typically do not cause progressive coral mortality. Colonies with bleaching severity 1 were not included in this analysis due to challenges associated with consistently identifying very low levels of bleaching severity.

To compare error between methods to within method observer error for the eight demographic metrics, data were summarized at the segment-level for all scleractinians combined. Error was calculated as the absolute difference in values (between methods or observers) divided by overall mean and then scaled from 0 to 1 so that we could compare the relative level of error across metrics (termed "midpoint scaled mean absolute error"). We calculated error for three different types of comparisons for the 43 paired plots sampled by both SfM and in situ methods. "Diver observer error" represents the difference between divers for a given demographic metric (Figure5A). "SfM observer error" represents the difference in error for a given metric between SfM annotators (Figure 5B). "Method Error" is the difference between methods for all possible combinations of method x observer divided by the mean of the absolute difference across all method x observer comparisons for a given metric (Figure 5C). We summarize each of these error distributions using the mean and standard error of the mean. We used nonparametric Kruskal-Wallis tests to test for differences between the three errors for metrics that did not meet assumptions of normality and equal variance.

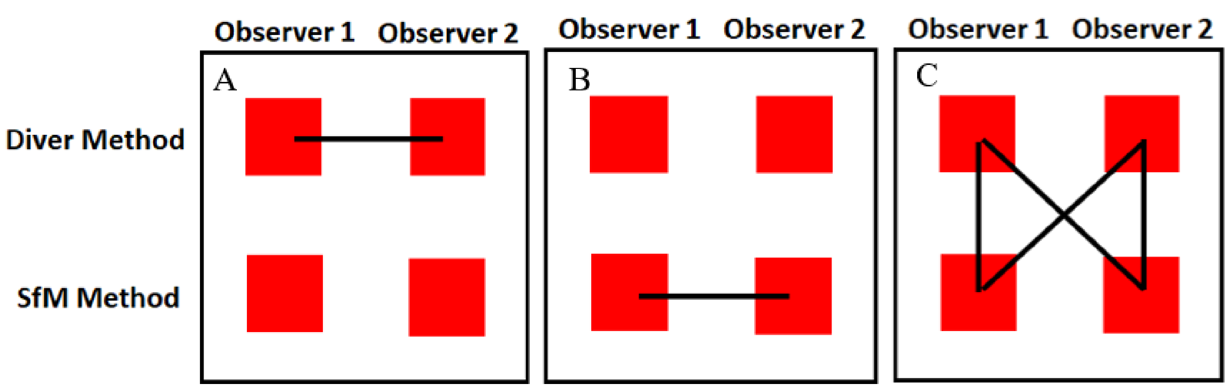

Figure 5. Graphic depicting comparisons (A) between observers for the in water observers (in situ divers), (B) between observers for the SfM imagery (SfM), and (C) between the methods of diver-derived data (in situ method) and SfM image-derived data (SfM method).

To test for differences between methods in the coral demographic metrics, data were summarized at the site-level at 104 sites that were surveyed by 1 diver and 1 SfM annotator. Within a site, only segments that were surveyed in both methods were included and then pooled to the site-level (5–10 m^2 of reef area/site). Data were pooled to the site-level because segments are not considered independent samples, but rather a means to capture spatial variability and this is the lowest spatial resolution that ESD typically summarizes NCRMP data. Results are presented for the eight metrics for total scleractinian corals combined and for adult and juvenile density of the three dominant coral genera (*Porites*, *Montipora*, and *Pocillopora)*. Each metric was tested for normality and equal variance. Measures of adult density, juvenile density, and average old partial mortality were square root transformed. Average colony diameter and *Porites* adult and juvenile density were log transformed. For each of these metrics, we established a series of linear mixed effects models (LMMs) to test three hypotheses: method type, method type × habitat type, and method × maximum depth. These variables were treated as fixed effects and sub-island sector was treated as a random effect. We hypothesize that our metric may vary between methods across different habitat types (aggregate reef, patch reef, pavement, rock and boulder, and rubble) as well as across different depths given issues of imaging colonies in high reef complexity, limited bottom time for divers working on deeper reefs, and lower image quality on deeper reefs. To assess the significance of fixed effects, we refit each model using maximum likelihood estimation (ML) and applied likelihood ratio tests (LRTs) (Zuur et al. 2009). Fixed effects that were not significant were sequentially dropped from models. The resulting best-fit models were refit using REML to estimate the fixed-effects parameters and associated effect sizes. Average recent dead, all prevalence metrics, and *Montipora* and *Pocillopora* density could not be transformed and therefore were only tested for overall difference between methods using nonparametric Wilcoxon Rank Tests for each metric.

To determine whether we had adequate sampling to detect significant methodological bias, we ran power analyses for each of the metrics assuming a 2-tailed t-test, the standard deviation of the untransformed or transformed metric (see transformation type above), power of 0.8, $\alpha = 0.05$, and an N from 3 to 350 samples. On each of the resulting curves, we selected the observed sample size (102–104) and compared it to the calculated mean absolute error (MAE), which is the mean absolute difference between methods.

Results

How does Methodological Error Compare to Observer Error?

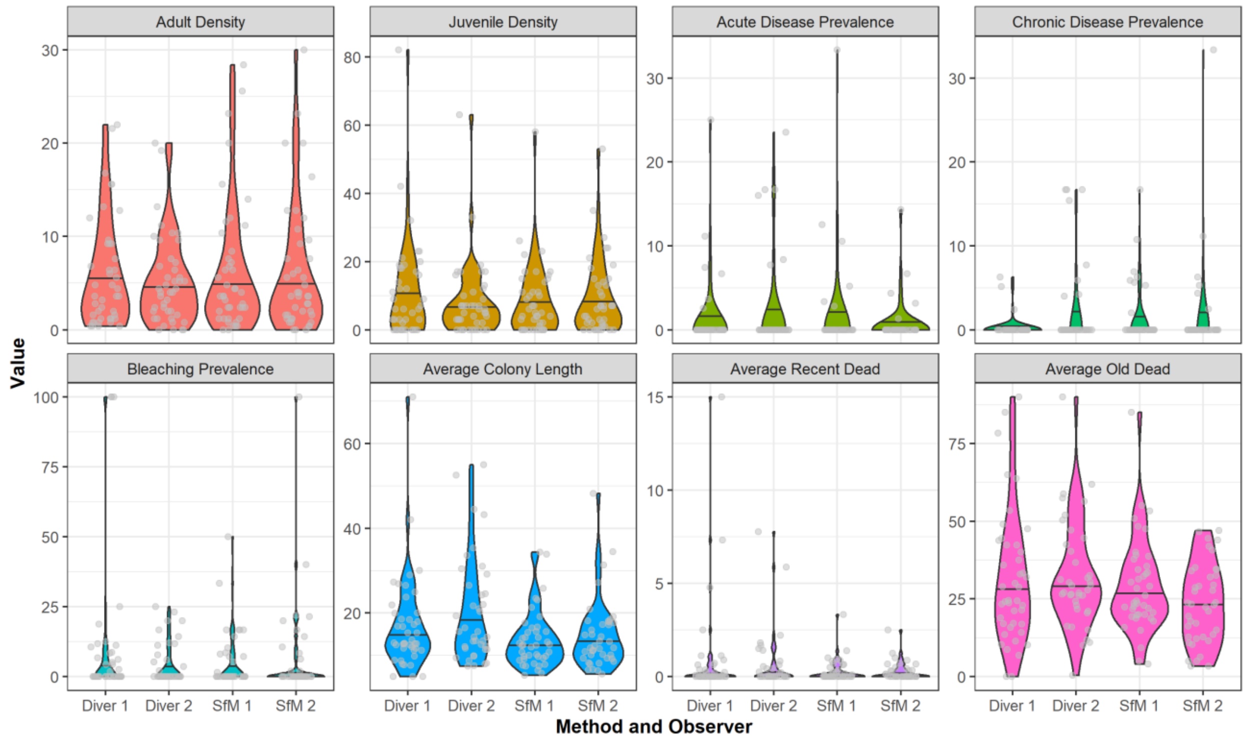

Figure 6. Violin plots that show the distribution and skew of 8 demographic metrics across 43 segments surveyed by 2 divers (Diver 1, Diver 2) and 2 SfM annotators (SfM 1, SfM 2) . Plots show the median for each metric and jitter of raw segment data (grey points). A total of 7 divers and two SfM annotators collected data across the 43 segments, so Diver 1, Diver 2, SfM 1 and SfM 2 were randomized.

To provide a general sense of how different metrics shift across observers and methods, Figure 6 illustrates the probability distributions and raw segment-level data of eight demographic metrics across the subset of 43 segments surveyed by two in situ divers and two SfM annotators. The shape of the distribution illustrates how each metric varied between observers as well as the range of values.

Although there is some minor variation in acute disease, chronic disease, bleaching and average old dead, the general shape of the distribution is retained among observers/methods across all eight metrics. The density metrics (adult, juvenile density) and average colony length each show moderate median values with substantial right skew, i.e., most values are small, but there is a long tail of larger values for each metric. The prevalence metrics (acute disease, chronic disease, bleaching) all show dominance by zero values (i.e., zero inflation) with a scattering of positive values, and recent dead shows a very small median (close but not equal to zero) with a similar scattering of large outlier values. Average old dead shows both the least skewed distributions and most variation among observers/methods of all metrics. It is important to note for all of these comparisons, that the identities of observers within each method were randomly assigned to observer "1" or "2" (i.e., Diver 1 vs. Diver 2, SfM 1 vs. SfM 2). Therefore, comparisons between observers within a method highlight general variation among multiple observers but do not reflect the tendencies of a single, human observer.

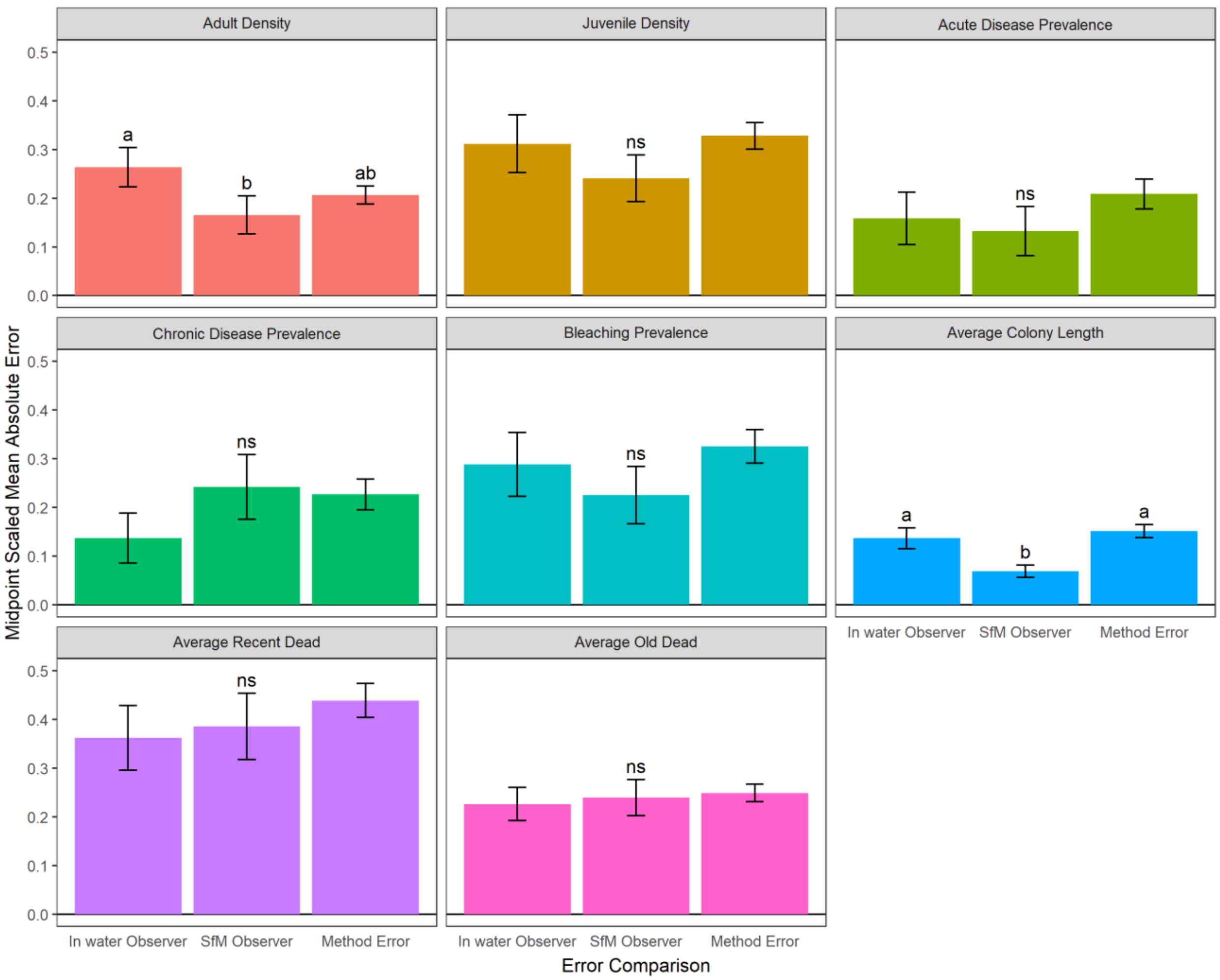

Figure 7. A comparison of the in situ observer error (difference between divers) and SfM observer error (difference between SfM annotators) to method error (difference between methods) for each coral demographic metric . Error is represented as the midpoint scaled mean absolute error (±SE) from 43 segments that were surveyed by multiple divers and SfM annotators. Letters represent results of post-hoc tests for each metric using Dunn's tests with Benjamini and Hochberg multiple test corrections. ns = not significant (adjusted p > 0.05).

To understand whether the absolute difference between methods is greater or smaller than the difference between observers, we compare the level of method error to both kinds of observer error (Figure 7). For adult density, diver observer error was significantly higher than SfM error, but was not different then method error (Figure 7, Appendix G). Similarly, for average colony diameter, there was no difference between diver error and method error, but SfM observer error was significantly lower than the other error comparisons (Figure 7, Appendix G). These results suggest that the difference between methods for adult density and colony length were just as variable as what we normally see between divers underwater, but SfM annotators were more consistent in scoring than divers. For the other metrics there was no significant difference between the three error types. Overall, this suggests that while there may be variability between methods, it is consistent with the level of variability we have between divers. Given the low sample size (43 segments) and low prevalence of diseases and bleaching as well as recent partial mortality (Figure 7) we may not be fully capturing an accurate estimate of error.

Is there a Methodological Bias in Coral Demographic Metrics at the Site-level?

Adult Density of Total Scleractinians

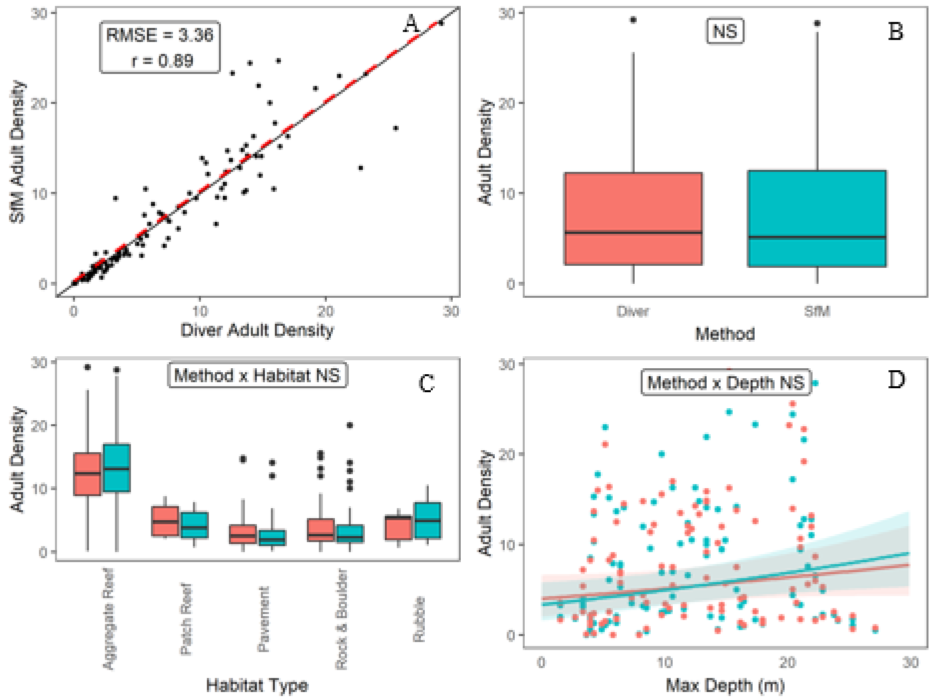

Figure 8. 1:1 Plot of the paired site-level SfM adult colony density vs. Diver adult colony density (points). Black line is 1:1 line, red dashed line is linear regression line; (B) Boxplot of adult density by method type; (C) Boxplot of adult density by method and habitat type; (D) Marginal effects plot of adult density by depth and maximum depth (m) including predicted values (lines) and confidence intervals of predictions (shaded areas) from linear mixed effects models. For plots B, C and D, the results of the LRTs are included for the different fixed effects (Method, Method x Habitat or Method x Depth) from the LMMs. NS: p >0.05

At the site-level, adult colony density showed a strong correlation between methods with a low RMSE and more variability above 15 colonies/m^2 (Figure 8A). We did not detect a significant methodological difference using LMMs (Figure 8B, Appendix H), nor was there a significant interaction between method and habitat (Figure 8C, Appendix H), or method and depth (Figure 8D, Appendix H).

Juvenile Density of Total Scleractinians

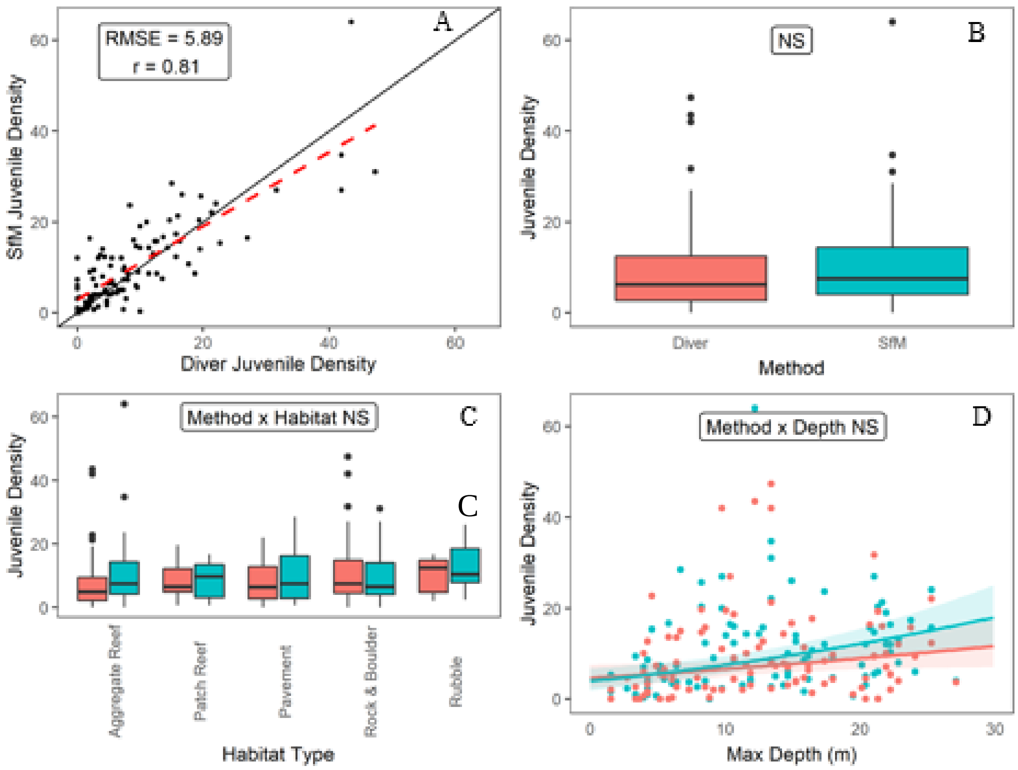

Figure 9. (A) 1:1 Plot of site-level SfM juvenile colony density vs. Diver juvenile colony density (points). Black line is 1:1 line, red dashed line is linear regression line; (B) Boxplot of juvenile density by method type; (C) Boxplot of juvenile density by method and habitat type; (D) Marginal effects plot of juvenile density by depth and maximum depth (m) including predicted values (lines) and confidence intervals of predictions (shaded areas) from linear mixed effects models. For plots B, C and D, the results of the LRTs are included for the different fixed effects (Method, Depth, Habitat, Method x Habitat or Method x Depth) from the LMMs. NS: p > 0.05.

Juvenile colony density strongly correlated between methods with a low RMSE (Figure 9A). It appears that SfM may be slightly underestimating juveniles at higher densities when compared to in situ surveys, although we did not detect a significant methodological difference (Figures 9A, B, Appendix H). More observations are needed at higher densities to resolve this. Juvenile density did not vary significantly as a function of method and habitat (Figure 9C, Appendix H). While there was no significant interaction of method and depth, SfM annotators recorded slightly higher juvenile density with increasing depth, with predicted SfM juvenile density approximately 6 colonies/m^2 higher than in situ surveys at deep sites (Figure 9D, Appendix H).

Adult Colony Diameter of Total Scleractinians

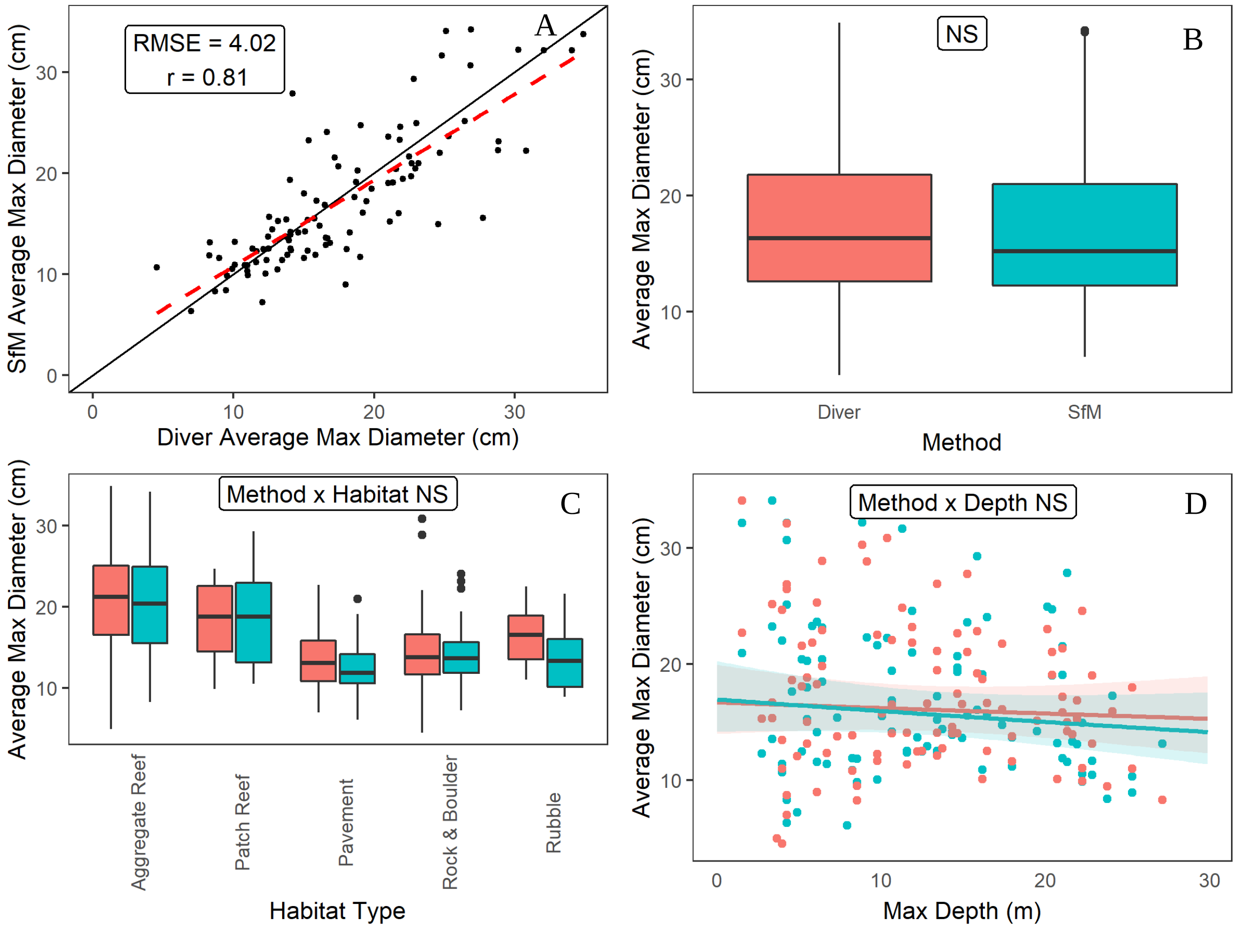

Figure 10. (A) 1:1 Plot of site-level SfM adult average maximum colony diameter vs. diver adult average maximum colony diameter (points). Black line is 1:1 line, red dashed line is linear regression line; (B) Boxplot of average maximum diameter by method type; (D) Marginal effects plot of average maximum diameter by depth and maximum depth (m) including predicted values (lines) and confidence intervals of predictions (shaded areas) from linear mixed effects models. For plots B, C, and D, the results of the LRTs are included for the different fixed effects (Method, Method x Habitat or Method x Depth) from the LMMs. NS: p> 0.05

Adult average maximum diameter strongly correlated with a low RMSE between methods (Figure 10A), and we did not detect a significant methodological difference (Figure 10B, Appendix H). There was no significant interaction of method and habitat (Figure 10C), nor was the interaction of method and depth significant (Figure 10D).

Average Old and Recent Partial Mortality of Total Scleractinians

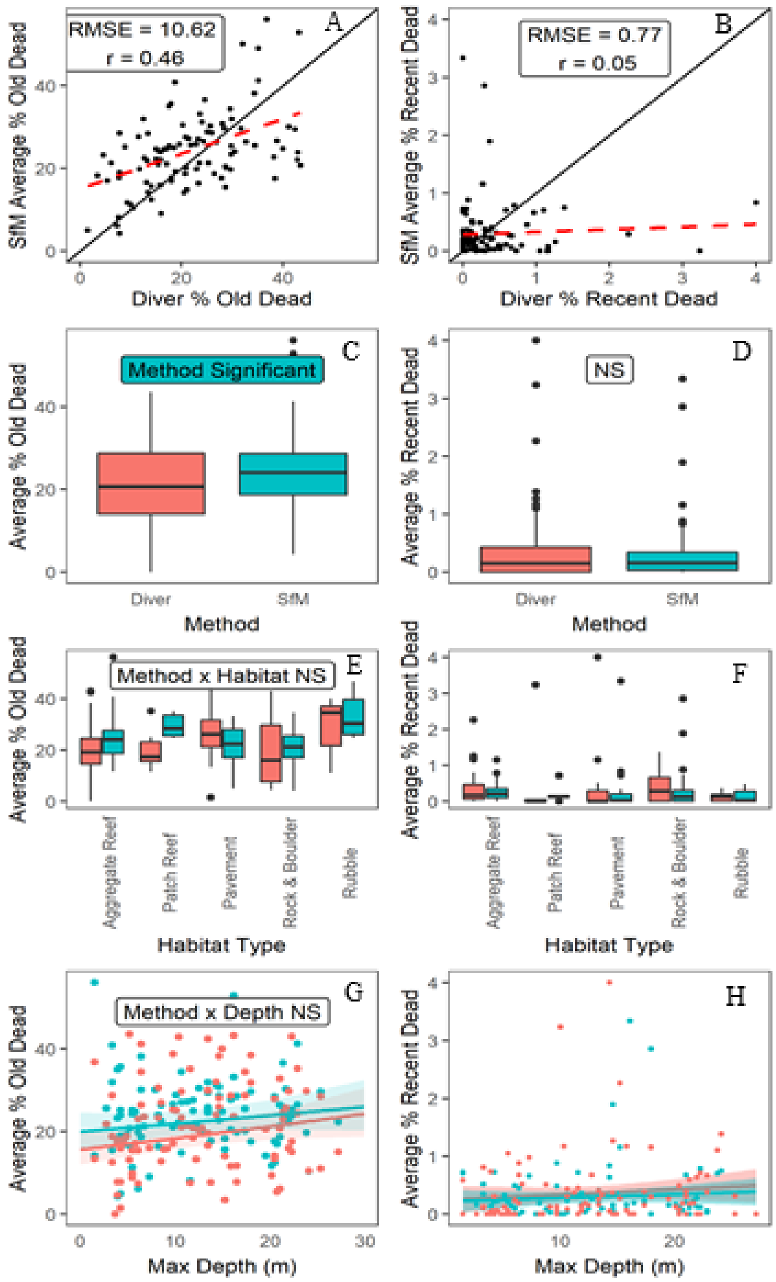

Figure 11. (A,B) 1:1 Plot of site-level SfM adult average percent partial mortality vs. Diver adult average percent old partial mortality (points) for old and recent mortality, respectively. Black line is 1:1 line, red dashed line is linear regression line; (C, D) Boxplot of average mortality by method type for old and recent mortality, respectively; (E, F) Boxplot of average mortality by method and habitat type for old and recent mortality; (G) Marginal effects plot of average old mortality by depth and maximum depth (m) including predicted values (lines) and confidence intervals of predictions (shaded areas) from linear mixed effects models. (H) Linear regressions (lines) of average recent partial mortality by depth and maximum depth (m) standard error (shaded areas). For plots C, E, and G the results of the LRTs are included for the different fixed effects (Method, Method

x Habitat or Method x Depth) from the LMMs. NS: $p > 0.05$. For recent mortality, the difference between methods overall (F) or by each habitat (H) were tested using nonparametric Wilcoxon tests and Benjamini and Hochberg multiple test corrections.

Adult average percent old dead was only moderately correlated with a high RMSE between methods (Figure 11A). SfM percent old partial mortality was significantly higher than in situ, but the median difference between methods was only 3% (Figure 11C, Appendix H). Although SfM percent old dead was higher than diver old dead in patch reef habitats, the interaction of method and habitat was not significant (Figure 11E, Appendix H). There was no significant interaction between method and depth (Figure 11G, Appendix H).

Adult average percent recent dead was not correlated between methods despite the low RMSE (Figure 11B), which is likely due to the high proportion of values less than 1%. There were no significant difference between methods (Figure 11D, Appendix H). We did not detect a significant difference between methods for each habitat when using separate Wilcoxon tests for each habitat type (Figure 11F, Appendix H). While the recent dead was slightly more correlated with depth for SfM annotators (Spearman rho = 0.19) compared to in situ data (Spearman rho = 0.05), the differences do not appear meaningful (Figure 11H). At 8.6% of the sites, divers recorded recent mortality that was not recorded by SfM annotators, and at 18.2% of sites, SfM annotators recorded recent mortality that divers did not record.

Acute Disease, Chronic Disease and Bleaching Prevalence of Total Scleractinians

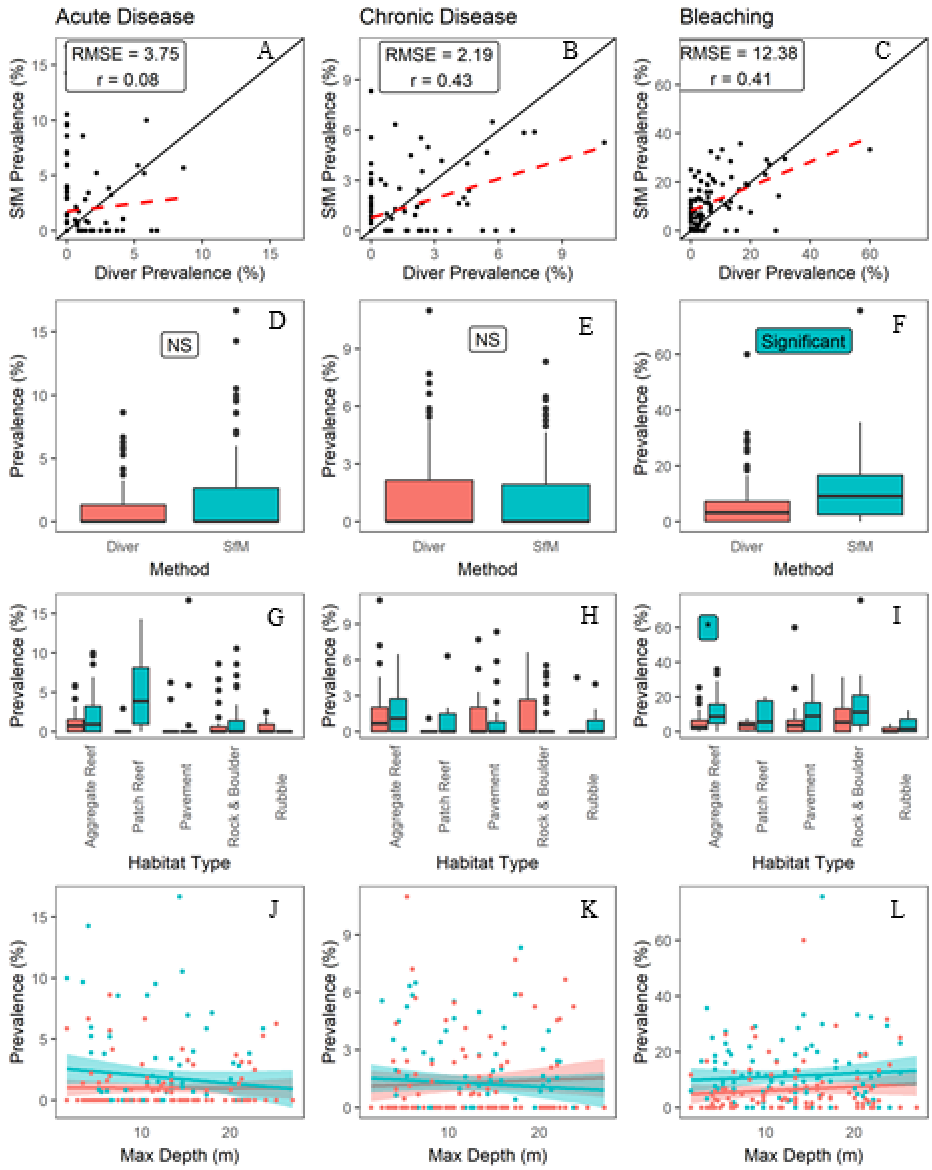

Figure 12. (A, B, C) 1:1 Plots of site-level SfM prevalence vs. Diver prevalence (points) for acute disease, chronic disease and bleaching, respectively . Black line is 1:1 line, red dashed line is linear regression line; (D, E, F) Boxplots of prevalence by method type for acute disease, chronic disease and bleaching, respectively. Significance (α= 0.05) of method was tested using a nonparametric Wilcoxon Test. NS: p > 0.05; (G, H, I) Boxplots of prevalence by method and habitat type for acute disease, chronic disease and bleaching, respectively. * = significant difference between methods for a given habitat type when tested with separate nonparametric Wilcoxon tests; (J, K, L) Linear regressions (lines) of prevalence by depth and maximum depth (m) standard error (shaded areas) for acute disease, chronic disease and bleaching, respectively.

Acute disease prevalence was weakly correlated between methods (Figure 12A). While prevalence was slightly higher for SfM compared to divers, there was no significant difference between methods (Figure 12D, Appendix H). We did not detect a significant difference between methods for each habitat when using separate Wilcoxon tests for each habitat type (Figure 12G, Appendix H). Prevalence was similarly poorly correlated with depth for both methods, suggesting there is not a significant interaction of method and depth (Figure 12J SfM: Spearman rho = - 0.07; Diver: Spearman rho = 0.04). At 15.4% of the sites, divers recorded disease that was not recorded by SfM annotators and 21.1% of sites, SfM annotators recorded disease that divers did not record.

Chronic disease prevalence was moderately correlated between methods with a relatively low RMSE (Figure 12B). Chronic disease prevalence was not significantly different between methods (Figure 12E, Appendix H). We did not detect a significant difference between methods for each habitat when using separate Wilcoxon tests for each habitat type (Figure 12H, Appendix H). Prevalence was similarly correlated with depth for both SfM (Spearman rho = - 0.004) and in situ methods (Spearman rho = 0.06), suggesting that there is no interaction of method and depth (Figure 12K). At 15.4% of the sites, divers recorded disease that was not recorded by SfM annotators and 15.4% of sites, SfM annotators recorded disease that divers did not record.

Bleaching prevalence was moderately correlated between methods with a higher RMSE (Figure 12C). SfM annotators recorded significantly higher bleaching compared to divers (Figure 12F, Appendix H). When considering each habitat separately, SfM bleaching prevalence was significantly higher than diver prevalence on aggregate reefs and prevalence did not vary between methods for the other habitats (Figure 12I, Appendix H). Prevalence was similarly correlated with depth for both SfM (Spearman rho = 0.07) and in situ methods (Spearman rho = 0.08), suggesting that there is no interaction of method and depth (Figure 12L). At 10.6% of the sites, divers recorded bleaching that was not recorded by SfM annotators and 18.2% of sites, SfM annotators recorded bleaching that divers did not record.

Density of Dominant Coral Genera

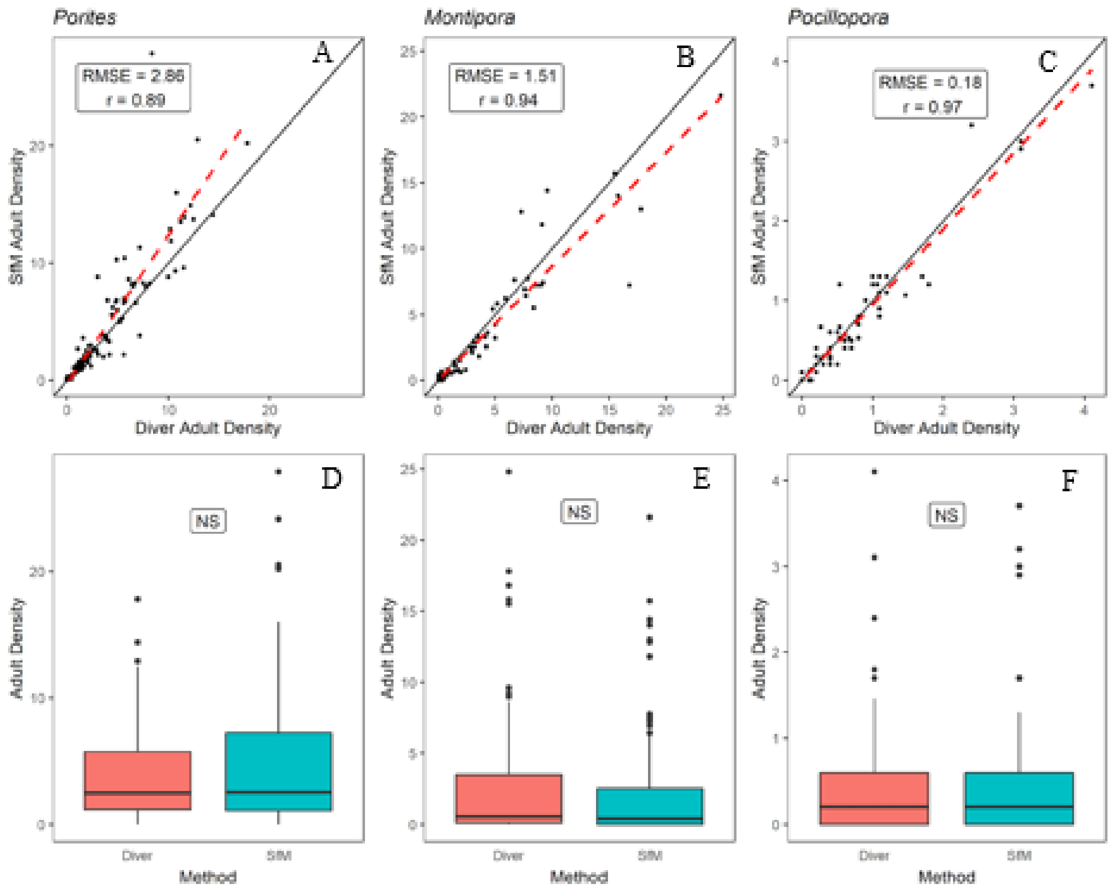

Figure 13. (A, B, C) 1:1 Plots of site-level SfM adult colony density vs. Diver adult colony density (points) for Porites, Montipora, and Pocillopora , respectively. Black line is 1:1 line, red dashed line is linear regression line; (D, E, F) Boxplots of adult colony density by method type for *Porites, Montipora,* and *Pocillopora*, respectively. α = 0.05. For *Porites*, significance of method was tested using a LRT of a LMM with Sector as a random effect. For *Montipora* and *Pocillopora*, significance of method was tested using a nonparametric Wilcoxon Test.

For all three of the dominant coral genera in the main Hawaiian Islands, adult colony density strongly correlated between methods, with the greatest correlation observed in *Pocillopora* (Figures 13 A, B, C). Adult density did not vary significantly between methods for any of the dominant genera (Figures 13 D, E, F. Appendix H). However, it does appear that SfM may be overestimating adult *Porites* relative to divers at densities >10 colonies/m^2 and there is more variability between methods for *Montipora* densities >10 colonies/m^2 (Figures 13A, B). More observations are needed at higher colony density sites.

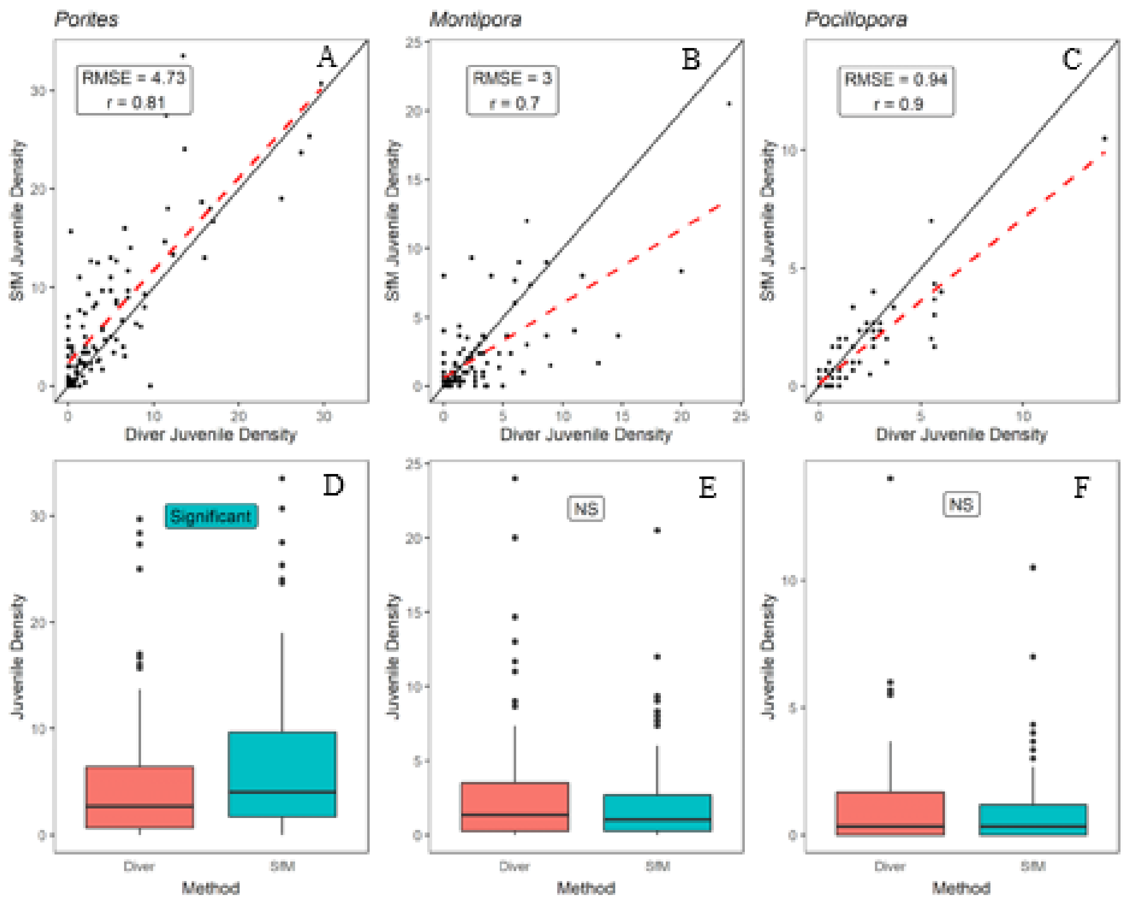

Figure 14. (A, B, C) 1:1 Plots of site-level SfM juvenile colony density vs. Diver adult colony density (points) for Porites, Montipora, and Pocillopora , respectively. Black line is 1:1 line, red dashed line is linear regression line; (D, E, F) Boxplots of juvenile colony density by method type for *Porites, Montipora*, and *Pocillopora*, respectively. α= 0.05. For *Porites*, significance of method was tested using a LRT of a LMM with Sector as a random effect. For *Montipora* and *Pocillopora*, significance of method was tested using a nonparametric Wilcoxon Test.

Juvenile colony density of the three dominant genera were moderately to strongly correlated between methods, with the weakest correlation observed in *Montipora* (Figures 14 A, B, C). SfM annotators observed significantly more juvenile *Porites* than divers, particularly at sites with low juvenile density (Figure 14 D, Appendix H). While *Montipora* and *Pocillopora* juvenile density was slightly lower for SfM compared to divers, we did not detect a significant difference between methods (Figures 14 E, F, Appendix H).

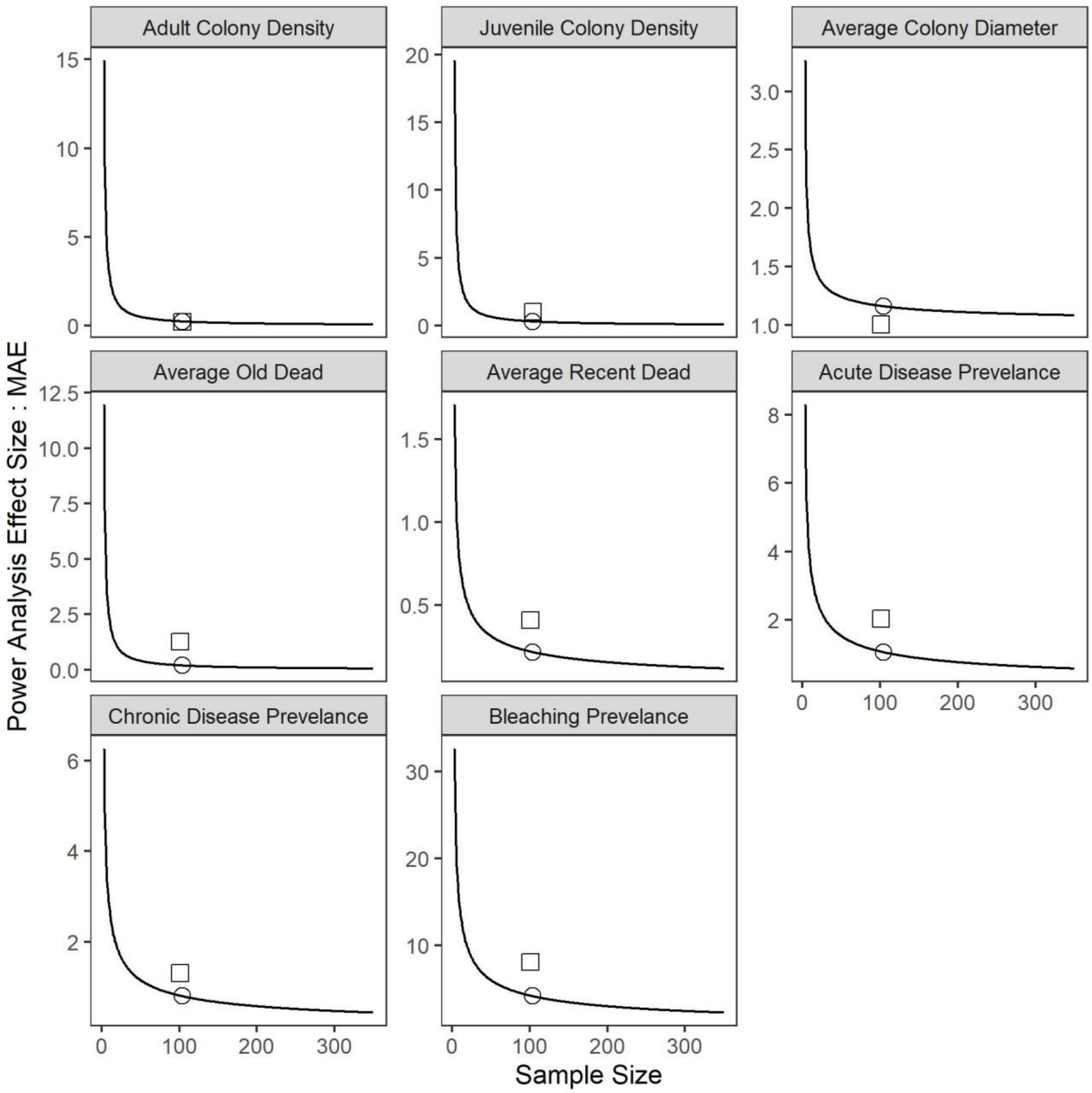

Figure 15. Power analysis to detect significant methodological bias assuming a 2-tailed t-test , the standard deviation of the metric (adult density, juvenile density, average colony diameter, and average old dead were transformed then back transformed, the remaining metrics were untransformed), power of 0.8, α = 0.05, and an N from 3 to 350 samples. Circles represent the effect size at n = 104 and the squares represent the mean difference between methods (MAE — mean absolute error) measured in this study.

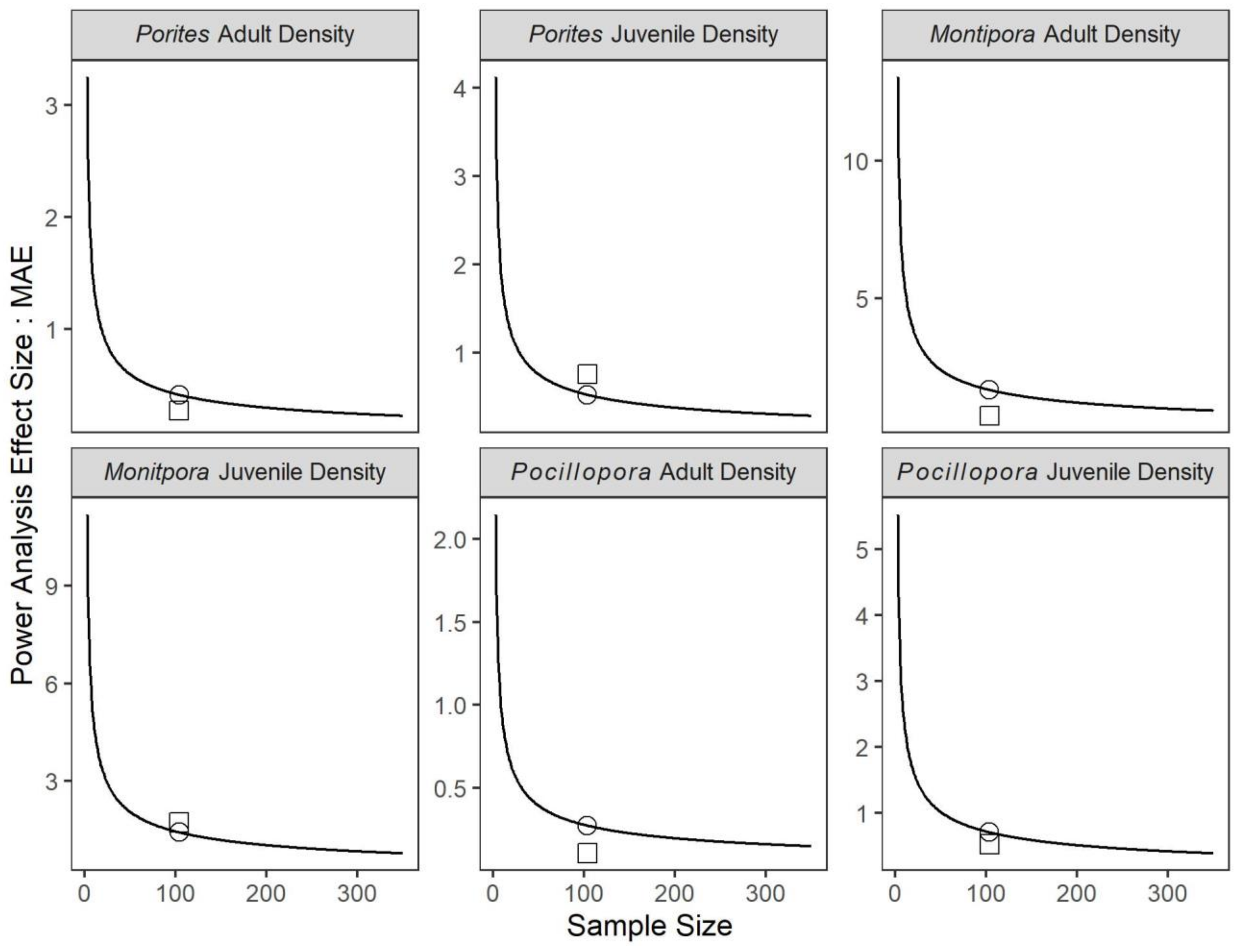

Figure 16. Power analysis to detect significant methodological bias assuming a 2-tailed t-test , the standard deviation of the metric (Porites adult and juvenile density were transformed then back transformed, the remaining density metrics were untransformed), power of 0.8, α = 0.05, and an N from 3 to 350 samples. Circles represent the effect size at n = 104 and the squares represent the mean difference between methods (MAE — mean absolute error) measured in this study.

Differences across methods for 6 of the 8 total scleractinian metrics and 5 of the 6 dominant genera metrics showed no significant difference from zero (Figures 8, 9, 10, 11, 12, 13, 14) and overall, the between-method error (MAE) was very low for all metrics (Figures 15, 16). The power analysis (Figures 15, 16) suggests that we have an adequate sample size to detect a significant difference between methods, especially as 7 of the 8 observed MAEs for total scleractinians and 3 of the 6 dominant taxa metrics show, non-significant estimates at or above our power analysis effect size. This suggests that the consistent pattern of small and non-significant differences between methods is not due to a lack of statistical power but instead due to well-supported similarities in the results across methodologies.

Discussion

The scientific community is increasingly looking to SfM to expand the spatial scale of reef monitoring and improve the efficiency of field data collection by replacing visual surveys with SfM while maintaining legacy data sets. Our study is the first to compare quantitatively standard in situ coral demographic survey methods from an existing long-term monitoring program to data generated from SfM across a range of islands, habitat types, and depths.

Overall, our results suggest that a majority of the metrics we investigated do not vary significantly between methodological approaches. For those that do, including percent old partial mortality, bleaching prevalence and Porites juvenile density, the difference between methods is minimal.

Demographic metrics that are comparable between methods

Overall, our results suggest that density of all adult colonies, adult density of the three dominant coral genera (*Porites*, *Montipora* and *Pocillopora*) and adult colony diameter, showed a strong correlation between methods and did not vary significantly between methods, even across habitats and depths (Figure 14). We saw more variability between methods at sites with higher colony densities, indicating that this kind of analysis would benefit from more observations at high densities. Identifying the boundaries of colonies is a fundamental challenge of these types of coral demographic surveys, regardless of whether surveys are conducted underwater or behind a computer. As colonial organisms, coral colonies can fragment into tissue patches. ESD NCRMP methods dictate that observers identify colonies by lumping together tissue fragments of a similar color and morphology on the original skeletal structure into one colony. Enumerating and sizing colonies can be challenging when partial mortality is not recent and colonies are densely aggregated. The role of some level of observer error in these patterns is also supported by the fact that we reported that between diver variability in adult density and adult maximum diameter is comparable to differences between methods. With lower between-observer error, SfM may provide an opportunity to reduce observer error compared to divers by allowing annotators to converse during annotation, revisit plots and correct errors, which is not possible for in situ observers (Figure 7).

Quantifying juveniles can be challenging given their small size and cryptic nature, sometimes preferring crevices and vertical surfaces to exposed substrates (Babcock and Mundy 1996; Edmunds et al. 2004). For these reasons, we hypothesized that SfM may underestimate juvenile density. However, our results indicate that there is no significant difference in density for all taxa combined between methods, nor between methods across habitats and depths. At the genus level some patterns emerge for the dominant taxa. Juvenile colony density for *Porites* among SfM observers was significantly higher than recorded by divers. *Porites* juveniles tend to be more inconspicuous than other juvenile taxa, often blending in with the substrate due to their muted color and low profile, especially when colonies are < 2 cm. In fact, when we quality checked the SfM annotations, the most common issue across all annotators was missing juveniles. While we were able to review and correct this in the SfM data set, we were not able to correct missing juveniles in the in situ data set. With that being said, the difference between methods in *Porites* juvenile density was only 3 colonies/m^2. For *Montipora* juveniles, there was no significant difference between methods, but there was more variability between methods at higher colony densities. Fragmentation is very common in this taxa, so similar to adult density, this variability may be associated with inter-observer variability in identifying colonies especially in high-density populations.

Demographic metrics that require additional adjustment

It is not surprising that metrics that rely on count data and measurements such as colony density and size showed a stronger correlation between methods than metrics that rely on visual subjective estimates of extent such as percent partial mortality and colony health.

Old partial mortality was significantly higher using SfM compared to in situ methods, and did not vary significantly between methods across habitats and depths. However, the absolute difference in old mortality estimates between methods (MAE) was low, i.e., only 1.27% (Figure 15) relative to the overall mean of 23% averaged across both methods. Given that the level of method error is comparable to observer error (Figure 6), we hypothesize that much of the variability between methods at the site-level is driven by inter-observer variation. Estimating old partial mortality can be challenging due to the coarse nature of this metric (recorded in 5% increments) and the need to set "100%" as the previous live extent of now partially dead colony. This latter estimation of old colony bound is particularly challenging in communities where corals are highly fragmented and/or branching.

Of the demographic metrics that ESD typically reports for NCRMP, recent partial mortality, prevalence of acute (causes progressive mortality) and chronic (typically does not cause mortality) diseases, and bleaching present the largest challenge. Our study did not detect a significant difference between methods nor the interaction of method and habitat or depth for these metrics. Several recent studies have tested comparability of in situ methods to imagery-based methods to extract coral health information. Contrary to our study, Page et al. (2016) found that disease prevalence recorded by in situ divers was six times higher compared to imagery at 12 coral reef sites in Northwestern Australia. However, unlike our study, Page et al. (2016) used photoquadrat images that were taken from one angle perpendicular to the colony surface, which likely resulted in missed lesions that would have been better captured by SfM given its 3D nature. Burns et al. (2020) used SfM to test the comparability of in situ and SfM methods on 895 colonies across two sites on Hawaii Island. At the colony-scale, they found that in situ assessments had a better ability to identify diseases correctly and conditions (sensitivity), but a lower ability to detect colonies correctly without a lesion (specificity) than SfM. They conclude that in situ surveys should not be considered the gold standard method given its lower specificity, and SfM should be considered as a viable replacement for in situ observations because the size of the difference between the methods was modest. As suggested by Burns et al., (2020) inter-observer variability and the subjective nature of coral health assessments likely explain the overall weak correlation between the two methods for recent dead mortality and prevalence metrics. The weak relationship between the methods may also be related to suboptimal image quality at some sites, which can impair disease and lesion identification. Due to the generally uncommon nature of diseases, assessing patterns at the site-level for partial mortality and prevalence is often challenging with the many sites having very low or zero prevalence of lesions. We recommend for future comparative analyses in other regions that we consider stratifying across sectors in addition to capturing the variability across conditions so that we can assess partial mortality at a coarser spatial resolution.

Bleaching prevalence was significantly higher in SfM assessments compared to in situ with a mean absolute error of 8.07%. We also found that bleaching was significantly higher in SfM on aggregate reefs. One possible explanation for this pattern is that aggregate reefs generally have higher colony density, which means that divers are more task loaded and may be prone to overlooking low/moderate levels of bleaching. These results differ slightly from previous methodological comparison studies which found bleaching was no different between in situ and imagery-based methods, given it should be one of the most easily identifiable conditions (Page et al. 2016; Burns et al. 2020). In our study, we captured moderate-to-severe levels of bleaching and while severe bleaching is easily identifiable, corals vary widely in their extent and severity of bleaching both within and across taxa. With our sites distributed across a large range of observers, depths, lighting conditions, and habitats, it is unsurprising that we are seeing more variability between the methods. One explanation for the higher level of bleaching in SfM assessments is that divers are more conservative about how they are identifying lower levels of bleaching severity. Alternatively, SfM images may have appeared overexposed resulting in higher estimates of bleaching. However, we believe this had a minimal effect on our data as all imagery was color corrected at depth and overexposed/poor quality imagery was removed during the QC process.

In summary, we saw little overall evidence of stable methodological bias between in situ collection and SfM imagery in the metrics we employ for benthic demographic surveys. In those cases in which there were distinctions between methods, we also saw large inter-observer variation or high variation at the segment/site scale.

Strengths and weaknesses of SfM and in situ coral demographic surveys

Both in situ and SfM survey methods have a variety of strengths and weaknesses (Table 1) and which method you choose depends on the research question and logistical constraints. In situ visual methods have a variety of strengths over SfM. In-water surveys allow divers to look at corals at the polyp-level from all possible angles, which is important for observing taxa located in crevices, under overhangs, or covered by sand or algae. It also allows scientists to exit the water with data that require minimal post-processing, resulting in summarized data on the order of weeks to months. This is important for providing timely data to managers to inform decision making. However, in situ monitoring also has several weaknesses. Surveys generally take longer (an average of 45 min for a three-person dive team). Given the longer dive time and the number of metrics collected underwater, it can be more strenuous to survey in poor conditions and diver fatigue can impact data quality. With limited bottom time, measurements of size and habitat complexity are generally measured or estimated at coarse levels. Due to limitations in bottom time, the reef area that can be surveyed by divers is often reduced when surveying deeper reefs. Lastly, visual survey data does not allow divers to verify or re-evaluate the benthos, which can lead to unmitigated observer error.

Conducting SfM surveys is generally more efficient and consistent across sites (approx. 25 min for a two-person dive team), which reduces field duration and therefore costs. SfM also allows annotators to take more accurate measurements of metrics such as colony size and structural complexity. Contrary to in situ methods, SfM annotators do not have time limitations, can take breaks when fatigued, and the area sampled is only limited by the extent of the reef that is imaged. Annotators also have the opportunity to speak with each other during data collection and improve consistency, thus reducing observer error. The SfM method allows, for the first time, the ability to quantify in a statistically rigorous way and correct observer error by reviewing the original imagery. By allowing divers and annotators to annotate the same area of reef, and then discuss differences in observations, SfM provides a unique opportunity to improve benthic training.

One of the primary weaknesses of SfM for demographic surveys is the time it takes to extract data from the imagery, which is on the order of months to a year depending on the number of sites, leading to delays in data dissemination. Another weakness is that annotators are limited by the quality and coverage of the imagery. If the imagery is poor quality or has poor overlap then it may be difficult to achieve polyp-level detail or fully capture all of the colonies. This can make species and coral health identification challenging. Furthermore, even with good image coverage, SfM cannot capture all of the crevices and overhangs that divers can assess in in situ surveys. SfM is also likely to miss colonies that are covered by sediment and macroalgae.

Table 1. Comparison of in situ and SfM strengths and weaknesses.

Metric	In situ	SfM
Data extraction	Quick	Lengthy
Permanent visual record	No	Yes
Survey team needed (# divers)	Moderate (3)	Small (2)
Benthic training needed	Same	Same
Survey in poor conditions	Difficult and strenuous	Moderate to easy
Bottom time	30 min–1 hr	10–25 min
Cryptic corals and diseases	Often easy to identify	Sometimes difficult to see
Sizing	Accurate	Highly accurate
Identification	Relatively easy	More difficult for some species
Visual observation	3D in situ	2D image
Data verification & re-evaluation	No	Yes
Area surveyed	Limited by bottom time	Limited by image size
Complexity metrics	Estimated & generalized	Accurate & detailed

Cost-benefit Analysis

In this section we provide a breakdown and comparison of the estimated time costs of generating the standard suite of coral demographic and benthic metrics each year for in situ surveys and SfM surveys (Table 2, 3). These costs have been tailored to meet ESD and NCRMP's programmatic requirements. Please contact authors if you are an external partner interested in identifying costs specific to your program. We also provide a summary of the hardware costs for SfM (Table 4). For the purposes of this comparative analysis, we assume that data will be collected from an average of 200 sites per year to meet NCRMP benthic allocations regardless of method. This number may vary depending on the region and year. We provide a breakdown of the amount of time generally needed to complete different tasks by site and for an average year.

It is important to note that we have conducted significant SfM research and development over the last year. As a result, tasks often took longer than they are likely to take in the future now that we have developed the process for collecting, managing, and extracting data from imagery. Furthermore, we will continue to identify ways to streamline data management and utilize emerging tools in artificial intelligence (AI) in an effort to increase efficiency. The time estimates provided in Table 2 reflect a conservative assessment of time following the research and development stage and involve a mixture of well-trained and new staff. We assume that we will have the same staff need for both methods and that staff may have a range of expertise.

Personnel Cost

Training & Cruise Preparation

Training is an important component of benthic surveys. It takes time to develop a team of benthic analysts that is consistent and accurate. The amount of time to train staff is equivalent between the two methods. Weeks of classroom and in-water training and calibration are required for both methods prior to each cruise. SfM image collection training can be conducted in just a few hours and involve staff not trained in benthic monitoring. However, SfM image annotation requires the same in depth training in taxonomic identification and demographic methods as an in situ diver to annotate imagery accurately regardless of whether it is conducted by ESD staff or outsourced to partners.

A significant amount of preparation is needed prior to each cruise. This includes gathering and purchasing gear, calculating the site allocation, generating maps, printing data sheets and more. The preparation time is similar between methods.

Field Time

At the benthic site-level, it requires on average 45 min with three divers to collect in situ NCRMP benthic data, compared to an average of 25 min with two divers to complete a SfM survey. In situ surveys require a 3-diver team with 1 backup diver that will need a total of 67 days to complete 200 sites and a total 2,133 personnel hours. This total time estimate does not include non-dive days associated with a cruise. SfM requires a 2-diver team with 1 backup diver and would require 40 days to complete and 960 personnel hours. Overall when comparing the total time needed during dive days to collect data from 200 sites, there is a 55% reduction in field time needed by switching SfM.

Processing Time:

Processing time is substantial for SfM and minimal for in situ methods. The SfM processing time to build the 3D point cloud and orthoprojection varies considerably based on the size of the plot, structural complexity of the plot, how many photos are taken, number of sites, and computer configuration. The hands-on processing time to extract data from an orthoprojection includes 1.5 personnel hours per site to set up the Agisoft projects to align the images and generate the dense point cloud (3D model), scale the models and generate the orthroprojection in Viscore. A majority of the processing time per site (8 hours) is required to manually delineate colonies in ArcGIS and QC delineations. With the processing infrastructure we have developed within PIFSC, we are able to generate 8 models simultaneously in approximately 17 hours (hands-off time). It will take approximately 3 weeks of time to generate the 3D and 2D products from 200 sites on our 2 most powerful servers (assuming continuous generation of models) and 1900 personnel hours to annotate these sites using 100% manual annotation.

Data Summary and Archiving Time:

Through the use of R scripts, it takes the same amount of time to run the final data QC and generate strata, sector and island-level data estimates regardless of method (Table 2). Archiving benthic data is the same for both methods, but we anticipate that more time will be needed to archive the SfM imagery and associated products. We are currently unable to provide an accurate time estimate because NCEI is still developing the best strategy for organizing and transferring the volume of data we anticipate for NCRMP SfM surveys.

Summary of Personnel Costs:

In summary, tasks such as training, cruise preparation, generation of data summaries, and data archiving are very similar between the two methods. The primary difference between these two methods is seen in the field and processing time. Field costs are reduced by 55% by switching to SfM under the proposed model. However, this method requires an estimated 1900 personnel hours of processing time that does not exist for in situ surveys. Overall, this would result in a 34% increase in personnel hours by switching from in situ to SfM using the current methods and data pipeline. A majority of this increase is due to the annotation costs of SfM. The efficiency of annotation can be improved in several ways. Most immediately, improvements in data entry processing in SfM and improved training of our annotators will dramatically reduce QC errors and speed up data entry during annotation. Given the rapidly advancing field of SfM and artificial intelligence (AI), and our desire to reduce the human annotation burden, it is important to continue partnering with computer scientists to research, test and implement new technological approaches to improve the efficiency of the SfM processing pipeline (see Section below on *Increasing efficiency through innovation and technology*).

Table 2. A comparison of the average hands-on time, per year, for field data collection and data processing between in situ and SfM surveys. Annual time estimates include time to generate data for 200 benthic sites. Although the number of personnel needed may vary by method, the general level of expertise is similar for both methods.

Field time/site & year	In situ	SfM	Personnel Needed
Bottom time (h)	0:45	0:25	Researcher, MERC, MERT
# divers needed/site*	4	3	
# of sites/day	3	5	
# of dive days/year	67	40	
Total field hours/year*	2133	960	
Processing time/site & year	In situ	SfM	
Hands-on time/site to generate 3D and 2D products (h)	0	1:30	MERC or MERT
Demographic data extraction & QC hands-on/site (h)	0	8:00	MERC and MERT
Total hands-on personnel time/year	0	1900	
Overall personnel hours for field collection and data extraction	**2,133**	**2,860**	

*Includes 1 backup diver
** Total field hours/year = (200 sites/ # sites per day) x # of divers x 8 hours
Researcher = PhD level scientist
MERC = Marine Ecosystem Research Coordinator (masters level)
MERT = Marine Ecosystem Research Technician (bachelors)

Equipment Cost

Equipment costs are associated with both methods, but the equipment costs for SfM is approximately 19 times higher than the in situ method. Table 3 provides a comparison of costs for each method divided by general category and initial, annual, and every 5-year purchases. For camera purchases, in situ surveys require an initial purchase of Canon G9× cameras (or equivalent point and shoot camera) and housings to photograph colonies that divers have questions about and require an annual purchase of one new camera/housing. For SfM, four new Canon Rebel SL2 cameras and housings are initially required. We estimate that half of the camera shutters will need to be rebuilt each year, and all cameras and housings will need replacement every 5th year. Field gear SfM costs are higher for one-time purchases of markers not associated with in situ surveys. The most expensive component of equipment costs for SfM are computers and servers. Computer costs for in situ surveys are minimal with two back-up hard drives purchased every 5 years. Computer costs for SfM are substantially more with an initial purchase of a processing server with GPU acceleration, a Synology server for cruise data storage, an in-house storage server, one high-powered workstation for annotation, backup hard drives, monitors and six Agisoft licenses. Annual cost includes additional hard drive storage and replacement parts for computers. These items will continue to improve in performance and drop in price over time. Because the actual lifespan of these items can be variable, complete replacement for these systems are included in the 5-year cost estimate, but it is possible that the equipment may last longer than 5 years.

Table 3. List of equipment and cost for in situ and SfM data collection and processing.

Item	In situ **Cost ($)**	SfM **Cost ($)**
Cameras and accessories		
Initial	4760.00	10,492.00
Each year	644.00	1,124.00
Every 5 years*	0	7,036.00
Field Gear		
Each year	235.00	240.00
Every 5 years*	445.00	435.00
Computers		
Initial	200.00**	7,828.00
Each year	0	5,000.00
Every 5 years*	200.00	48,169.00
Total	6,284.00	120,324.00

* 5-year costs include annual costs.
**Initial computer costs for in situ surveys do not include laptops that all ESD staff are issues and are needed to summarize data.

Recommendations

The feasibility of replacing standard in situ coral demographic surveys with SfM surveys while maintaining continuity with legacy NCRMP data is dependent on improving training of our benthic analysts, exploring a hybrid approach to benthic monitoring, and balancing field and post-processing costs.

Overall, our results suggest that a majority of the metrics do not vary significantly between methodological approaches. For those that do, including percent old partial mortality, bleaching prevalence and *Porites* juvenile density, the difference between methods is minimal. Furthermore, the level of variability between methods is comparable to the variability we normally see between divers. With that being said, the correlation between methods was weakest for recent partial mortality and disease and bleaching prevalence, which is likely the result of inter-observer variability, sensitivity of the metric to rare events at the segment/site scale (i.e., very low prevalence condition metrics), and poor image quality at some sites.

- **A hybrid approach**: One method that would both prioritize metric quality and balance in-field and annotation costs is a hybrid approach to benthic monitoring. This would involve extracting metrics such as density, colony size and old partial mortality from SfM, but recording recent dead and incidences of disease, bleaching and other compromised health states using in situ observations. This would allow ESD to use SfM to minimize field time while using higher-accuracy in situ surveys to obtain the colony-level detail needed to assess coral health, and thereby reduce post-collection annotation time. We estimate that divers would still be able to survey 5 sites/day so field costs would not change from what is proposed in Table 2. However, we estimate that it would save SfM annotators 1 hr of data extraction time. While the hybrid method has not been tested and personnel costs could vary, we estimate that this would result in a 24% increase by switching from in situ to SfM instead of a 34% increase as proposed above. Some limitations of this approach are that it still requires at least one highly trained benthic specialist in the field, and unless effectively merged with fish surveys (as SfM belt have been) will require separate specialized benthic and fish survey teams.

- **Reduce observer error**: To implement most effectively this hybrid strategy, we recommend that ESD continue efforts to improve consistency between divers and annotators. This should include a field component as well as a classroom component. We recommend expanding the current pre-cruise training and calibration to incorporate a rigorous SfM component where all annotators and divers annotate the same subset of sites, conduct QC of their peer's annotations, quantify and compare error rates, and then have a detailed discussion about specific issues. This would require an additional two weeks of training for each staff member.

- **Strive for high image quality:** It is paramount that divers continue collecting quality imagery using guidelines listed in the SOP (Suka et al. 2019) and Appendices. In addition to properly color balanced and sharpen images, our annotations would also benefit from shooting from more than one angle. This would allow divers to capture colonies on vertical or concave surfaces and reduce gaps in the model. When time permits, we recommend that plots with high rugosity be photographed from a minimum of two angles. Taking extra care photographing these sites will ensure all surfaces are captured.

- **Improve efficiency of data extraction**: One of the largest hurdles to overcome with SfM is the significant annotation and post-processing necessary to extract data. In the short-term, we recommend developing a tool to allow annotators to enter directly SfM data into the established benthic database from ArcGIS rather than entering data into a geodatabase. We estimate that this would take 2–3 months of personnel time. This will allow ESD to leverage the efficient data entry tool and drastically reduce time spent correcting database QC errors. We also see value in developing a script that will allow a server to generate continuously 3D models minus manual human interaction (approximately 1 month of time). Having an automated script will speed the model run time considerably. In the long-term, we recommend working with the computer science community to develop 'human-in-the-loop' annotation tools to reduce manual annotation time (see "Increasing efficiency through innovation and technology section" below).

Increasing efficiency through innovation and technology

The rapidly expanding field of Artificial Intelligence (AI) has the potential to reduce significantly the amount of human interaction time required to extract data from SfM imagery. While AI tools are potentially revolutionary for scaling up coral reef monitoring, they cannot fully replace humans at this stage and tools should continue to leverage human expertise by employing 'human-in-the-loop' approaches. Currently, this challenge is being tackled from a variety of angles by approaches such as CoralNet, a widely used machine-learning image analysis tool for point classification (Beijbom et al. 2015); an encoder-decoder Convolutional Neural Network (CNN) for semantic segmentation leveraging human annotated sparse points (Alonso et al. 2017); TagLab, an interactive semantic segmentation tool that integrates CNN results and previous human labelling (Pavoni et al. 2019); the use of bounding boxes to more efficiently identify benthic features for humans to annotate (Mandel et al. 2019; Modasshir and Rekleitis 2020); NemoNet, a CNN approach with a citizen-scientist videogame to generate training data (Chirayath and Li 2019); and AI challenges that invite computer scientists to develop completely novel automated solutions to delineation (Ionescu et al. 2019). These advances will help move annotation from a mostly hands-on, time-intensive approach to a semi-automated workflow, and shift annotation from 2D to 3D space.

Advances are being made to employ the 3D qualities of SfM to examine large- and small-scale variations in habitat complexity (rugosity, fractal dimension, and height range). These approaches can provide novel and highly detailed measurements of the complexity across spatial and temporal scales. At sites with co-located benthic and fish data, these metrics can be used to explore complexity thresholds that drive fish abundance and biodiversity and trends between coral health and fish trophic levels. These new approaches would expand ESD's data streams through statistically rigorous methods to calculate robust complexity metrics (Fukunaga et al. 2020; Torres-Pulliza et al. 2020) to improve our understanding of which biotic and abiotic factors affect structural complexity and how structural complexity is changing with anthropogenic disturbances.

The use of Unmanned Systems (UxS) is expanding the reach of coral research (Wheeler et al. 2005; Mortensen et al. 2008; Freiwald et al. 2009). UxS technology is becoming smaller and more affordable while capabilities and sensors continue to improve. These developments offer an opportunity to extend survey capabilities to greater depths (expanded bottom time at depth and reaching beyond safe scuba diving depths) and provide the flexibility to survey in the absence of divers. SfM has already been implemented on UxS to conduct coral reef research (Pizarro et al. 2009; Price et al. 2019). While significant R&D is still needed to assess the utility of UxS to assess coral reef systems, implementing a combination of these two technologies could provide new opportunities for benthic data collection.

Aerial or satellite imagery is a powerful approach to expand the survey area of shallow coral reefs (Collin et al. 2018). Hyperspectral aerial imagery is promising an enhanced level of visual data that can record subtle differences in the color spectrum and has been recently applied to coral reefs (Asner et al. 2020). These methods require accurate in situ imagery to properly ground truth and characterize benthic features at a high taxonomic level (Thompson et al. 2017). Some success has been made in incorporating a geolocation device into SfM surveys to collect accurate site positioning that allows expansion of benthic sampling data beyond what a diver can achieve (Hatcher et al. 2020).

The hardware and software investment needed to generate the dense point clouds, 2D products, and derived benthic metrics efficiently may exceed the means of most small monitoring programs. To address this challenge, current SfM practitioners should identify opportunities to develop infrastructure for cloud processing, data sharing and data storage. Cloud networks allow flexibility in processing power that can dynamically expand and contract to meet surges in processing needs (e.g., post-cruise processing) and Cloud networks make data widely accessible to collaborators earlier in the data processing pipeline. Collaborations with groups that have abundant computing power resources (NASA and NOAA HPC centers) may be an alternative to Cloud computing for quicker implementation.

As cameras, computing power, algorithms, large data storage, data sharing, remote systems, and software advances, new and improved methods will develop to allow us to increasingly leverage SfM digital data in the future. These advances will expand the reach of coral research and advance the capabilities across the coral research community that will feed back into NOAA's data collections and expand overall knowledge of coral reef ecosystems.

Acknowledgements

We would like to thank Raymond Boland, Noah Pomeroy, and the crew of the NOAA Ship *Oscar Elton Sette* for providing field support during the main Hawaiian Islands Rapid Assessment and Monitoring Cruise. We would also like to thank Michael Akridge, Brooke Olenski, Kevin Trick and Annette DesRochers for assistance with development and maintenance of the data management pipeline. We thank Stuart Sandin, John Burns, Nicole Pederson, Vid Petrovic and Clinton Edwards for support with field testing, methods development and implementation of SfM. This work is funded by the NOAA Coral Reef Conservation Program, as part of the Pacific-National Coral Reef Monitoring Program.

Literature Cited

Alonso I, Cambra A, Muñoz A, Treibitz T, Murillo AC. 2017. Coral-Segmentation: Training Dense Labeling Models with Sparse Ground Truth. In: Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017. Vol. 2018-January.

Asner GP, Vaughn NR, Balzotti C, Brodrick PG, Heckler J. 2020. High-Resolution Reef Bathymetry and Coral Habitat Complexity from Airborne Imaging Spectroscopy.". Remote Sens. 12(2):310. https://doi.org/10.3390/rs12020310.

Babcock R, Mundy C. 1996. Coral recruitment: Consequences of settlement choice for early growth and survivorship in two scleractinians. J Exp Mar Bio Ecol. 206(1):179–201. doi:10.1016/S0022-0981(96)02622-6.

Bayley DTI, Mogg A, Koldewey H, Purvis A. 2019. Capturing complexity: field-testing the use of ‘structure from motion’ derived virtual models to replicate standard measures of reef physical structure. PeerJ. 7:e6540. doi:10.7717/peerj.6540.

Beijbom O, Edmunds PJ, Roelfsema C, Smith J, Kline DI, Neal BP, Dunlap MJ, Moriarty V, Fan TY, Tan CJ, et al. 2015. Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation. PLoS One. 10(7). doi:10.1371/journal.pone.0130312.

Burns JHR, Delparte D, Gates R, Takabayashi M. 2015. Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs. PeerJ.:3:e1077. doi:10.7717/peerj.1077.

Burns JHR, Weyenberg G, Mandel T, Ferreira SB, Gotshalk D, Kinoshita CK, Marshall MJ, Del Moral NA, Murphy SJ, Pascoe KH, et al. 2020. A Comparison of the Diagnostic Accuracy of in-situ and Digital Image-Based Assessments of Coral Health and Disease. Front Mar Sci. 7:304. doi:10.3389/fmars.2020.00304.

Casella E, Collin A, Harris D, Ferse S, Bejarano S, Parravicini V, Hench JL, Rovere A. 2017. Mapping coral reefs using consumer-grade drones and structure from motion photogrammetry techniques. Coral Reefs. 36(1):269–275. doi:10.1007/s00338-016-1522-0.

Chirayath V, Li A. 2019. Next-Generation Optical Sensing Technologies for Exploring Ocean Worlds—NASA FluidCam, MiDAR, and NeMO-Net. Front Mar Sci. 6. doi:10.3389/fmars.2019.00521.

Choat JH, Bellwood DR. 1985. Interactions amongst herbivorous fishes on a coral reef: influence of spatial variation. Mar Biol. 89(3):221–234. doi:10.1007/BF00393655.

Collin A, Ramambason C, Pastol Y, Casella E, Rovere A, Thiault L, Espiau B, Siu G, Lerouvreur F, Nakamura N, et al. 2018. Very high resolution mapping of coral reef state using airborne bathymetric LiDAR surface-intensity and drone imagery. Int J Remote

Sens. 39(17):5676–5688. doi:10.1080/01431161.2018.1500072.

Conand C, Chabanet P, Quod J-P, Bigot L. 1999. Guidelines for coral reef monitoring in the south-west region of the Indian Ocean. https://www.researchgate.net/publication/259890494_Guidelines_for_coral_reef_monitoring_in_the_south-west_region_of_the_Indian_Ocean.

Connell JH, Hughes TP, Wallace CC. 1997. A 30-year study of coral abundance, recruitment, and disturbance at several scales in space and time. Ecol Monogr. 67(4). doi:10.1890/0012-9615(1997)067[0461:AYSOCA]2.0.CO;2.

Donnelly R, Neville D, Mous P. 2003. Report on a rapid ecological assessment of the Raja Ampat islands, Papua , Eastern Indonesia, held October 30 - November 22, 2002. www.komodonationalpark.org.

Edmunds PJ, Bruno J, Carlon D. 2004. Effects of depth and microhabitat on growth and survivorship of juvenile corals in the Florida Keys. Mar Ecol Ser - MAR ECOL-PROGR SER. 278:115–124. doi:10.3354/meps278115.

Edwards CB, Eynaud Y, Williams GJ, Pedersen NE, Zgliczynski BJ, Gleason ACR, Smith JE, Sandin SA. 2017. Large-area imaging reveals biologically driven non-random spatial patterns of corals at a remote reef. Coral Reefs. 36(4):1291–1305. doi:10.1007/s00338-017-1624-3.

English S, Wilkinson C, Baker V. 1994. Survey Manual for Tropical Marine Resources. https://www.aims.gov.au/sites/default/files/Survey Manual-sm01.pdf.

Ferrari R, Figueira WF, Pratchett MS, Boube T, Adam A, Kobelkowsky-Vidrio T, Doo SS, Atwood TB, Byrne M. 2017. 3D photogrammetry quantifies growth and external erosion of individual coral colonies and skeletons. Sci Rep. 7(1):16737. doi:10.1038/s41598-017-16408-z.

Figueira W., Ferrari R., Weatherby E., Porter A., Hawes S., Byrne M. 2015. Accuracy and Precision of Habitat Structural Complexity Metrics Derived from Underwater Photogrammetry. Remote Sens. 7:16883–16900. doi:10.3390/rs71215859.

Fox MD, Carter AL, Edwards CB, Takeshita Y, Johnson MD, Petrovic V, Amir CG, Sala E, Sandin SA, Smith JE. 2019. Limited coral mortality following acute thermal stress and widespread bleaching on Palmyra Atoll, central Pacific. Coral Reefs. 38(4):701–712. doi:10.1007/s00338-019-01796-7.

Freiwald A, Beuck L, Ruggeberg A, Taviani M, Hebbeln D. 2009. The White Coral community in the Central Mediterranean Sea Revealed by ROV Surveys. Oceanography. 22(1):58–74.

Fukunaga A., Burns JHR., Pascoe KH., Kosaki RK. 2020. Associations between Benthic Cover and Habitat Complexity Metrics Obtained from 3D Reconstruction of Coral Reefs at Different Resolutions. Remote Sens. 12:1011. doi:10.3390/rs12061011.

Hatcher GA, Warrick JA, Ritchie AC, Dailey ET, Zawada DG, Kranenburg C, Yates KK. 2020. Accurate Bathymetric Maps From Underwater Digital Imagery Without Ground Control. Front Mar Sci. 7:525. doi:10.3389/fmars.2020.00525.

Hernández-Landa RC B-FE, R. R-N. 2020. Size-frequency distribution of coral assemblages in insular shallow reefs of the Mexican Caribbean using underwater photogrammetry. PeerJ. 8(e8957). doi:10.7717/peerj.8957.

Ionescu B, Müller H, Péteri R, Cid YD, Liauchuk V, Kovalev V, Klimuk D, Tarasau A, Abacha A Ben, Hasan SA, et al. 2019. ImageCLEF 2019: Multimedia Retrieval in Medicine, Lifelogging, Security and Nature. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 11696 LNCS.

Lange ID, Perry CT. 2020. A quick, easy and non-invasive method to quantify coral growth rates using photogrammetry and 3D model comparisons. Methods Ecol Evol. 11(6):714–726. doi:10.1111/2041-210X.13388.

Mandel T, Gotshalk D, Del Moral N, Wilde D, Marshall M, Spengler A, Murphy S, Wells B, Burns J, Weyenberg G. 2019. Balancing human and machine performance when analyzing image cover. In: Proceedings - 6th Annual Conference on Computational Science and Computational Intelligence, CSCI 2019.

Modasshir M, Rekleitis I. 2020. Enhancing Coral Reef Monitoring Utilizing a Deep Semi-Supervised Learning Approach.

Mortensen PB, Buhl-Mortensen L, Gebruk A V, Krylova EM. 2008. Occurrence of deep-water corals on the Mid-Atlantic Ridge based on MAR-ECO data. Deep Sea Res Part II Top Stud Oceanogr. 55(1):142–152. doi:10.1016/j.dsr2.2007.09.018.

NOAA Pacific Islands Fisheries Science Center. 2015. Pacific Reef Assessment and Monitoring Program: Belt Transect Surveys for Demographics and Condition of Corals at U.S. Pacific Reefs from 2007 to 2012 from 2010-06-15 to 2010-08-15. https://inport.nmfs.noaa.gov/inport/item/25260.

NOAA Pacific Islands Fisheries Science Center ESD. 2020a. Pacific Reef Monitoring and Assessment Program: Line-Point-Intercept (LPI) Surveys for Benthic Cover at Coral Reef Sites across the Pacific Ocean from 2005 to 2012 from 2010-06-15 to 2010-08-15. https://www.fisheries.noaa.gov/inport/item/25261.

NOAA Pacific Islands Fisheries Science Center ESD. 2020b. National Coral Reef Monitoring Program: Benthic Images Collected from Stratified Random Sites (StRS) across the Hawaiian Archipelago from 2013 to 2019 from 2010-06-15 to 2010-08-15. https://www.fisheries.noaa.gov/inport/item/25247.

Obura D, Aeby G, Amornthammarong N, Appeltans W, Bax N, Joe B, Brainard RE, Chan S, Pamela F, Gordon TAC, et al. 2019. Coral Reef Monitoring, Reef Assessment Technologies, and Ecosystem-Based Management. Front Mar Sci. 6:580.

doi:10.3389/fmars.2019.00580.

Page CA, Field SN, Pollock FJ, Lamb JB, Shedrawi G, Wilson SK. 2016. Assessing coral health and disease from digital photographs and in situ surveys. Environ Monit Assess. 189(1):18. doi:10.1007/s10661-016-5743-z.

Palma N. 2016. 3D Modeling of Coral Reefs Using Structure-from-Motion Photogrammetry as an Inventory and Monitoring Tool. https://scholarship.miami.edu/discovery/fulldisplay/alma991031447804602976/01UOML_INST:ResearchRepository.

Pavoni G, Corsini M, Palma M, Scopigno R. 2019. A Validation Tool For Improving Semantic Segmentation of Complex Natural Structures. In: Eurographics. https://diglib.eg.org/handle/10.2312/egs20191014.

Pedersen NE, Edwards CB, Eynaud Y, Gleason ACR, Smith JE, Sandin SA. 2019. The influence of habitat and adults on the spatial distribution of juvenile corals. Ecography (Cop). 42(10):1703–1713. doi:10.1111/ecog.04520.

Pizarro O, Eustice RM, Singh H. 2009. Large Area 3-D Reconstructions From Underwater Optical Surveys. IEEE J Ocean Eng. 34(2):150–169. doi:10.1109/JOE.2009.2016071.

Price DM, Robert K, Callaway A, Lo lacono C, Hall RA, Huvenne VAI. 2019. Using 3D photogrammetry from ROV video to quantify cold-water coral reef structural complexity and investigate its influence on biodiversity and community assemblage. Coral Reefs. 38(5):1007–1021. doi:10.1007/s00338-019-01827-3.

Storlazzi CD, Dartnell P, Hatcher GA, Gibbs AE. 2016. End of the chain? Rugosity and fine-scale bathymetry from existing underwater digital imagery using structure-from-motion (SfM) technology. Coral Reefs. 35(3). doi:10.1007/s00338-016-1462-8.

Suka R, Asbury M, Gray AE, Winston M, Oliver T, Couch CS. 2019. Processing Photomosaic Imagery of Coral Reefs Using Structure-from-Motion Standard Operating Procedures. doi:10.25923/h2q8-jv47.

Thompson DR, Hochberg EJ, Asner GP, Green RO, Knapp DE, Gao B-C, Garcia R, Gierach M, Lee Z, Maritorena S, et al. 2017. Airborne mapping of benthic reflectance spectra with Bayesian linear mixtures. Remote Sens Environ. 200:18–30. doi:10.1016/j.rse.2017.07.030.

Torres-Pulliza D, Dornelas MA, Pizarro O, Bewley M, Blowes SA, Boutros N, Brambilla V, Chase TJ, Frank G, Friedman A, et al. 2020. A geometric basis for surface habitat complexity and biodiversity. bioRxiv. doi:10.1101/2020.02.03.929521.

Voss J, Shilling E, Combs I. 2019. Intervention and fate tracking for corals Florida, affected by stony coral tissue loss disease in the northern Florida Reef Tract. https://floridadep.gov/sites/default/files/Voss SEFL Disease Intervention 2019_FDEP FINAL Report.Fully508compliant.pdf.

Wheeler AJ, Thiede TB, Klages M, Grehan A, Monteys FX. 2005. Deep-water coral mounds on the Porcupine Bank, Irish Margin: preliminary results from the Polarstern ARK-XIX/3a ROV cruise. In: Freiwald A. RJM, editor. Cold-Water Corals and Ecosystems. Erlangen Earth Conference Series. Springer Berlin Heidelberg.

Zuur AF, Ieno EN, Walker NJ, Anatoly A. Saveliev, Smith GM. 2009. Mixed Effects Models and Extensions in Ecology with R. Springer Science & Business Media.

Appendices

Appendix A. Camera Selection

Aside from excellent image quality, the camera selection for SfM image collection was based on four important factors. (1) The camera must be small enough to allow divers to complete other tasks underwater and use a very limited space on a small boat deck. (2) The camera must be affordable and durable as it will be used every day for months and life span would be shorter than normal. (3) The camera must be easy to use underwater, and (4) the camera and housing must allow white balancing and continuous shooting while underwater. Although all the cameras we tested (Figure 17A Canon SL2, Figure 17B Sony a6300 and Figure 17C Canon G9×) had the ability to white balance in situ, the accessibility and quality of white-balancing varied between models. The detail and color of the Canon SL2 proved to be the best of the three cameras tested underwater (Figure 17A).

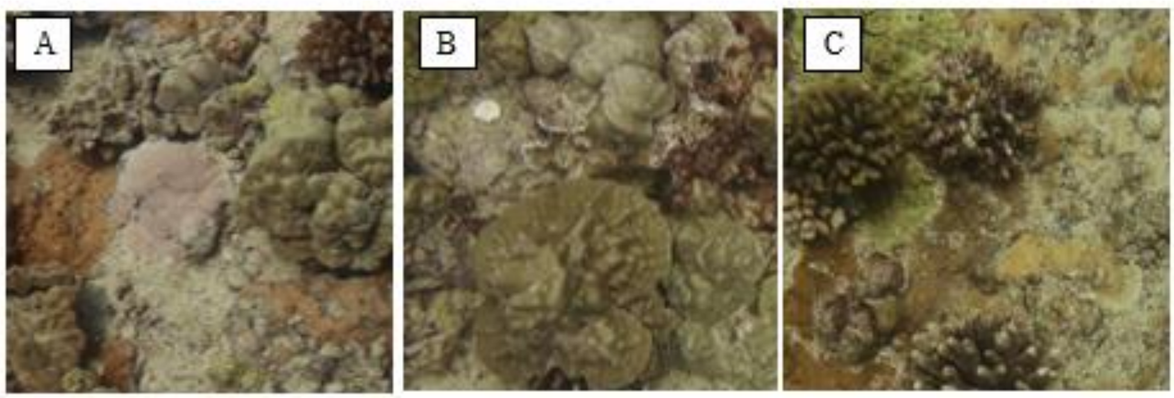

Figure 17. Close-up examples of image quality from Canon SL2 (A), Sony a6300 (B), and Canon G9x (C).

Appendix B. Accurate Image Color

Poor color imagery makes identification and analysis of coral colonies challenging (Figure 18, A). True color imagery is key to producing reliable and consistent annotations. Having accurate and consistent color values is also critical for any automated processing where pixel RGB color values are evaluated. Although image-processing software provides tools to correct for color issues after image collection (Figure 18 B), using the in-camera white-balancing options produce the best color results (Figure 18 C). Using an industry-standard 18% gray card at depth at every site to white balance produces the best color results.

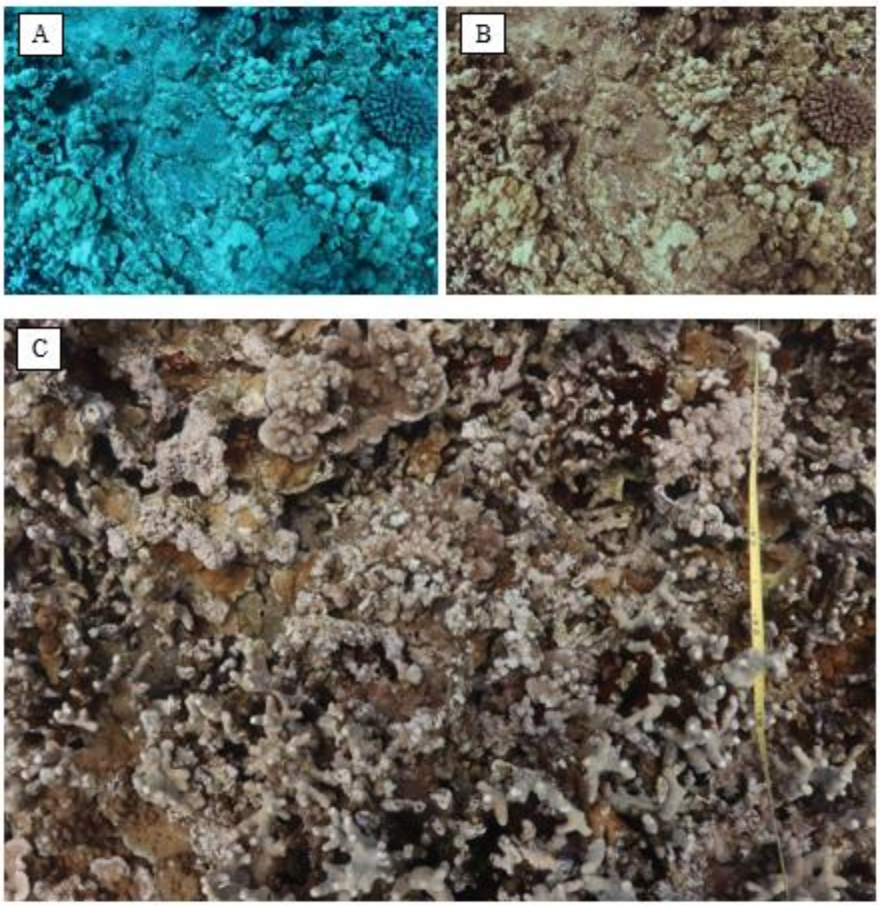

Figure 18. Image without any in-water white balance (A), post-processing color correction (B), image taken using proper in-water white balance (C).

Appendix C. Selecting Size and Shape of Plot

The size and shape of SfM survey areas should be determined based on the research question. To collect coral demographic data for adult and juvenile colonies, plots need to be large enough to capture a full colony size range, detailed enough to capture juvenile colony features, but small enough to minimize time underwater. Estimated swim time for different size plots can be calculated to determine the ideal plot size that fits within logistical constraints (Figure 19). ESD SfM benthic surveys covered a 3 × 20 m area, which replicates the historical in situ survey area (four segments of 1 × 2.5 m along the transect line). The SfM survey area includes a buffer around the area of interest to avoid any distortion that may occur at the edges of the resulting orthoprojection (see below for discussion of this issue).

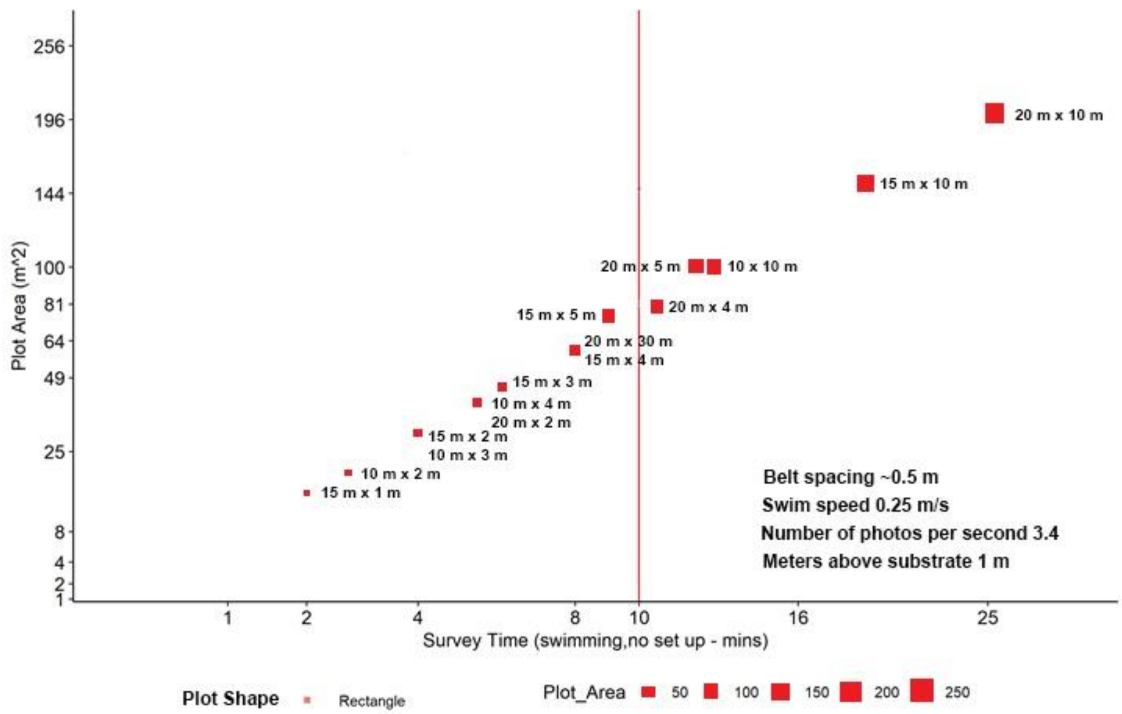

Figure 19. Calculated swim time for different size plots.

There is a concern that long, narrow SfM plots have more geometric distortion than shorter, wider plots. This distortion occurs around the edges of the orthoprojection image because the geometry of the model is not constrained on all sides when the original images are aligned. This lack of constraint may allow geometric error to propagate across narrow models. To evaluate this potential error, we photographed a 10 × 10 m plot at 3 sites then selected photographs from the same image set at each site to create models that were 10 × 10 m, 6 × 10 m, and 2 × 10 m and ran each image set independently. We selected 9 reef features from each site, 3 across each end and 3 across the middle. We measured the same 9 features on each model for each site. Results from ANOVAs of each site indicate that there is no significant difference in the measurement of features between models of different widths for any of the sites (Figure 20, Lapakahi: $F = 0.012$, $df = 1$, $p = 0.915$. Kealakekua: $F = 0.006$, $df = 1$, $p = 0.937$. Wailea: $F = 0.01$, $df = 1$, $p = 0.923$).

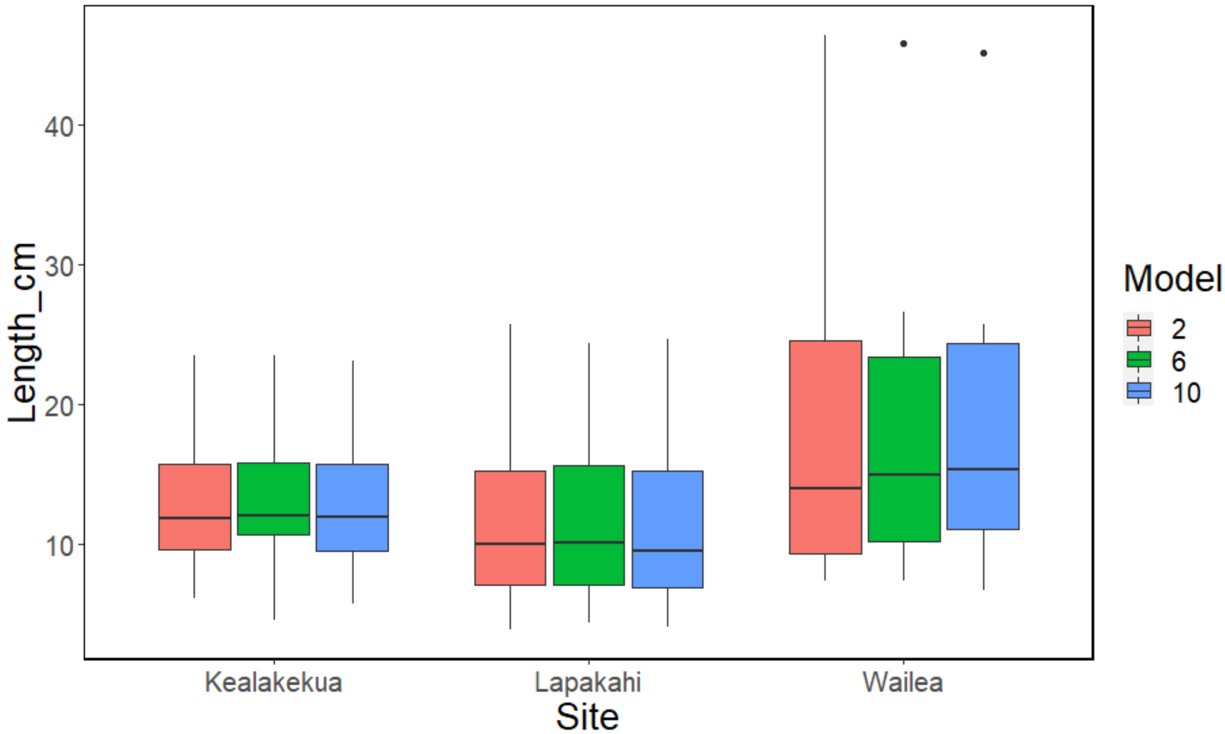

Figure 20. Box plots of three Hawaii Island sites (Kealakekua, Lapakahi, and Wailea) comparing length measurements (cm) of the same bottom features between models with different widths (pink = 2×10 m, green = 6×10 m, and blue = 10×10 m) with no significant difference between features measured ($p > 0.05$).

Data volume is impacted by survey size due to the increase in the number of images collected and should be considered when determining plot size. For example, the data generated from 200 benthic sites using SfM is approximately 6.6 TB (Table 4) with the largest portions of this data set being the original imagery and 3D models (Agisoft data). Data volume varies based on the size of the plot, site complexity, swim speed, and number of sites surveyed.

Table 4. Breakdown of average data size for each step in SfM image processing for 200 benthic sites.

Plot type	# sites	Image data size per site (GB)	Agisoft data size per site (GB)	Arc data size per site (GB)	Total data size per site (GB)	Total data size for cruise
3 × 20 m belts	200	15	16	1	33	13 TB

Appendix D. Optimizing Survey Time

Achieving finer detail in SfM products requires careful image collection. This can be achieved by increasing the number of images collected, taking images from multiple angles, adjusting camera settings, slowing swim speed and collecting images closer to the substrate. To examine the differences in quality and efficiency between image collection methods that fit within operational constraints, we collected imagery at 4 sites, first swimming a single pass back and forth over the plot (Figure 21A), then swimming a second pass perpendicular to the first (double pass) to create a cross hatch pattern (Figure 21B). The double pass increased the number of images taken and increased the number of angles captured throughout the survey area. The images were processed independently to generate models using the single pass images then using the double pass imagery. Double pass models took more than twice as long to run compared to the single pass models (Figure 22) with no clear visual differences in model detail or quality. The RMS reprojection error was low in both models ($\leq$ 1.5 pix) indicating that a single pass is sufficient to produce a quality model and significantly reduces field and processing time.

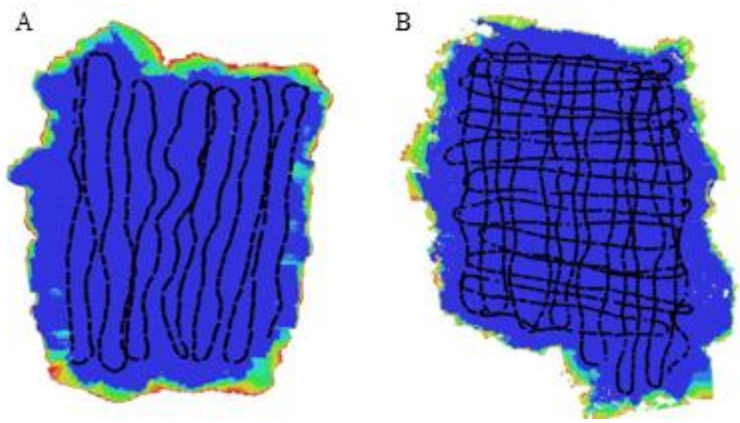

Figure 21. Graphic of the survey plot with single pass survey (A) and double pass survey (B) . Black dots indicate the position of the camera during image capture, or swim path. Blue area indicates high image overlap, red areas indicate low image overlap.

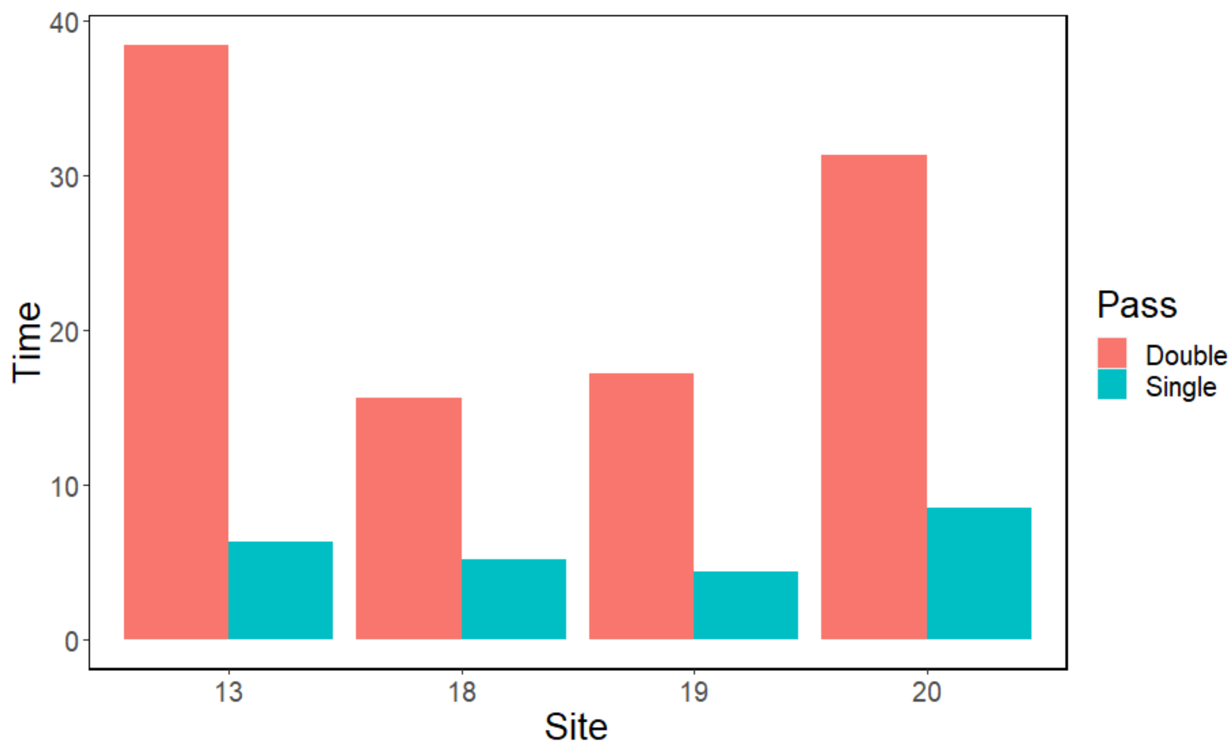

Figure 22. Model run time (minutes) for a single swim pass (green) and a double swim pass (pink) at 4 sites.

SfM survey time can be shortened by swimming at a faster speed, but it introduces blur in images (Figure 23A), even with camera settings optimized. Image blur is a critical problem because it cannot be corrected in post processing. A 3×20 m plot can be completed within 10 minutes even if the photographer swims at a very slow speed. A slow swim speed will greatly improve image quality (Figure 23B). Swimming slow is especially important in overcast or surge conditions and at deeper depths where light is limited.

Figure 23. A close-up view of coral colony taken during a SfM survey . Image A is taken at a high swim speed, image B is taken at a slow swim speed.

Appendix E. Camera Position

Image detail and the quality of the resulting SfM model decreases as the distance between the camera and seafloor increases. However, taking images closer to the seafloor reduces the area captured in each image, requiring collection of more images to achieve appropriate overlap (Figure 24A). Testing with the Canon SL2 revealed that feature details deteriorated rapidly beyond a 1-m distance from the substrate (Figure 24B, C).

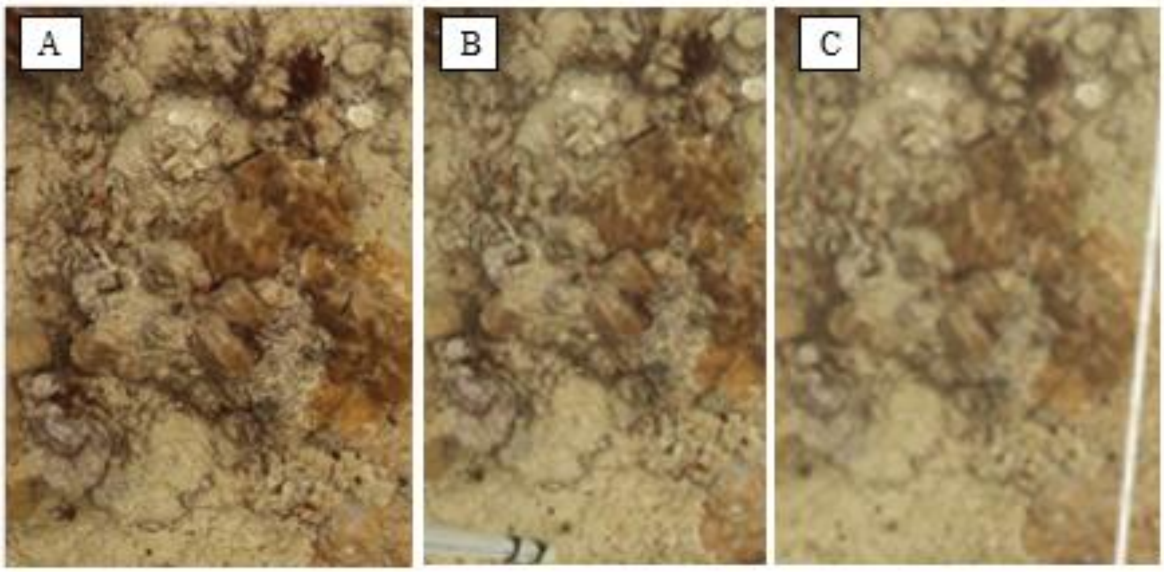

Figure 24. Close-up of images showing the decline in image resolution with distance from the substrate . Images were taken at a 1 m (A), 1.5 m (B), and 2 m (C) distance from the substrate using a Canon SL2 camera.

Appendix F. Environmental Conditions

When images are collected in very challenging environmental conditions such as surge and high turbidity (Figure 25A) SfM models render surprisingly well (Figure 25B). Objects that move during image collection cannot be matched when overlapping photographs are compared and aligned during model generation, so those objects are no longer visible in the resulting orthoprojection image. Because of this, it is common for the transect tape and suspended sediment to be missing from the orthoprojection image, producing better overall results than the original images. Due to the challenges of collecting in situ data under these conditions, the data extracted from the SfM models may be better than in situ observations.

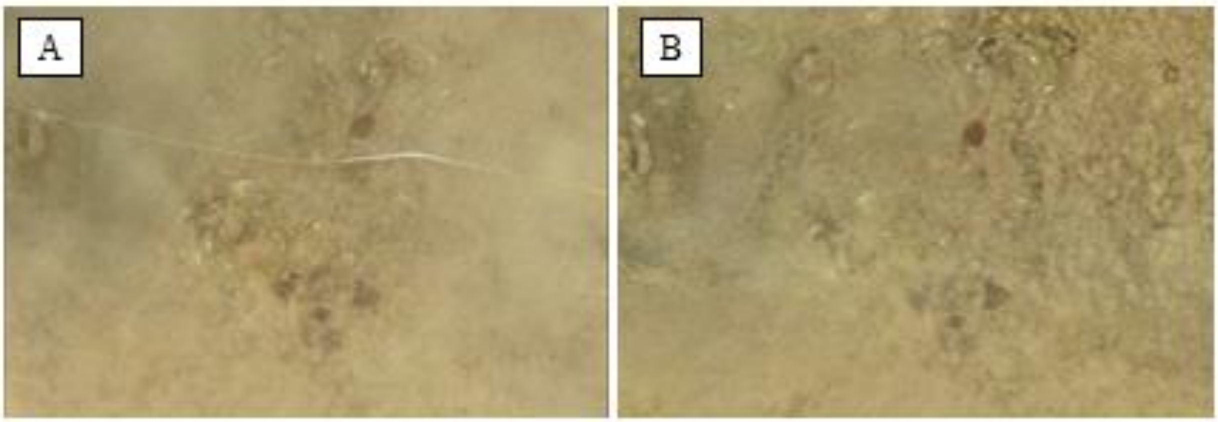

Figure 25. Original image from camera (A) with suspended particulates obscuring view of substrate, and resulting orthoprojection image (B) with improved view of substrate.

Appendix G. Table of Statistical Results for Error Comparison.

Response Variable	Test	Fixed effect	Likelihood Ratio	*p*
Adult Colony Density	LMM	Method × Habitat	0.837	0.933
		Method × Depth	0.512	0.474
		Method	0.001	0.978
Juvenile Colony Density	LMM	Method × Habitat	3.795	0.434
		Method × Depth	1.419	0.234
		Method	1.823	0.177
Average Colony Diameter	LMM	Method × Habitat	1.643	0.801
		Method × Depth	0.387	0.534
		Method	0.152	0.697
Average Old Dead	LMM	Method × Habitat	4.479	0.345
		Method × Depth	0.338	0.561
		Method	**6.507**	**0.011**
Average Recent Dead	Wilcoxon	Method		0.992
		Method (Aggregate Reef)		0.709
		Method (Patch Reef)		0.158
		Method (Pavement)		0.741
		Method (Rock and Boulder)		0.382
		Method (Rubble)		0.842
Acute Disease Prevalence	Wilcoxon	Method		0.194
		Method (Aggregate Reef)		0.288
		Method (Patch Reef)		0.288
		Method (Pavement)		0.877
		Method (Rock and Boulder)		0.882
		Method (Rubble)		0.288
Chronic Disease Prevalence	Wilcoxon	Method		0.960
		Method (Aggregate Reef)		0.722
		Method (Patch Reef)		0.722
		Method (Pavement)		0.722
		Method (Rock and Boulder)		0.722
		Method (Rubble)		0.722
Bleaching Disease Prevalence	Wilcoxon	**Method**		**0.000**
		Method (Aggregate Reef)		**0.014**
		Method (Patch Reef)		0.678
		Method (Pavement)		0.301
		Method (Rock and Boulder)		0.154
		Method (Rubble)		0.516

Bold indicates fixed effect is significant ($p < 0.05$).

Appendix H. Table of Statistical Results for Method Bias.

Response Variable	χ^2	df	*p*
Adult Colony Density	**9.855**	**2**	**0.007**
Juvenile Colony Density	2.344	2	0.310
Average Colony Diameter	**12.487**	**2**	**0.002**
Average Old Dead	0.305	2	0.859
Average Recent Dead	1.102	2	0.577
Acute Disease Prevalence	0.631	2	0.729
Chronic Disease Prevalence	1.542	2	0.463
Bleaching Disease Prevalence	1.196	2	0.550