

Standard Operating Procedure (SOP) #19

Statistical Data Analysis and Reporting

Version 2.0 (July 28, 2021)

Change History

New Version #	Revision Date	Author	Changes Made	Reasons for Change	Previous Version #
2.0	7/28/2021	Kim Weisenborn, Sheila McKenna	Updates to content. Changes to statical software and reporting.	To update outdated external links. To update statistical programs and reporting requirements.	1.0

Only changes in this specific SOP will be logged here. Version numbers increase incrementally by hundredths (e.g., version 1.01, version 1.02) for minor changes. Major revisions should be designated with the next whole number (e.g., version 2.0, 3.0, 4.0). Record the previous version number, date of revision, author of the revision, changes made, and reason for the change along with the new version number.

Purpose

This SOP addresses data analysis and reporting for the Marine Fish Vital Sign in the Pacific Island Network (PACN). Data analyses for (1) species richness, (2) abundance, (3) biomass, and (4) diversity are outlined. Integrated reporting of these four components is also outlined. This protocol outlines two initial steps of data analysis at each of the unit, landscape, and trend-assessment levels. These steps summarize and determine the range of variation for some parameters (based on present knowledge). Additional, multi-vital sign integration, synthesis, analysis, and reporting is not addressed in this SOP at present. The use of products identified in this protocol does not imply endorsement, effectiveness, or warranty by NPS.

Analysis Overview

Data analysis is defined here as the steps by which observations of the environment are processed for interpretation and synthesis into meaningful information that is accessible to managers. Data analysis also includes quality assurance and control efforts that occur in the field, during exploratory analysis, in summarization, during interpretation, and when drawing conclusions.

Analytical Strategies

Two initial steps are identified in data analysis for all Marine Fish Monitoring data: summarization and establishing the range of variation. These steps are encompassed in the larger construct of data management and data stewardship which is discussed in the data management SOPs #10 – 18. Data

analysis for (1) species richness, (2) abundance, (3) biomass, and (4) diversity incorporate these two initial steps at each of the unit, landscape, trend-assessment, and synthesis levels.

1. **Summarization:** This step includes identification and documentation of outliers, missing data, quality control and quality assurance issues (see SOP #13 Data Entry and Verification and SOP #14 Post Season Data Quality Review and Certification), mean or median values, variance estimates, and other descriptive statistics. Data visualization (see Cleveland 1993, 1994, Tufte 1983, 1990, 1997) and exploratory data analyses are an integral component of this process.
2. **Range of Variation:** This step includes establishing expected boundaries for response variables, typically as quantiles, and identifying data that falls outside these expected bounds. Establishing a range of normal variation may take years or decades for many parameters, but providing managers and decision-makers with the best available information will strengthen adaptive management feedback.

Analytical Approach

We identify three basic levels of analytical methods for our monitoring data: unit level, landscape level, and trend assessment level (Table SOP 19.1). The unit level, the most basic element, addresses sample design questions regarding use of finite or infinite populations, and the establishment of relative, absolute, or index response variables. The landscape level identifies the spatial nature of sampling and the temporal allocation of samples, such as the revisit scheme in use. The trend assessment level integrates unit or landscape level data over time, where typically some form of regression or repeated measures ANOVA is used to identify the slope or trend. Finally, synthesis examines patterns within and across Vital Signs, and with other ecological factors, to gain broad insight on ecological processes and integrity. Unit level, landscape level, and trend assessment level are addressed here. Synthesis will be addressed within and across multiple Vital Signs and is therefore left for network level or broader scientific consideration in the future.

Table SOP 19.1. Three approaches at the park level for analyzing Marine Fish Vital Sign monitoring data.

Level of Analysis	Description	Responsible Party
Unit Level	Quality assurance, quality control, and calculation of individual, site specific statistics from monitoring data	Marine Biological Technician with oversight by PACN Marine Ecologist and with assistance from Park Ecologists.
	Step 1 (Summarization): Measures of mean, median, variation, and other basic statistics. Include graphical presentation of data.	
	Step 2 (Range of Variation): Establish historical or expected range of values, relation to relevant regulatory levels, confidence estimates.	
	Indices or other site-specific metrics may also be developed.	
Landscape Level	Integration of unit level data across appropriate landscapes to address ecological status.	Marine Biological Technician with oversight by PACN Marine Ecologist and with assistance from Park Ecologists.
	Step 1: Integration of unit level summarization results across multiple locations and sites.	
	Step 2: Integration of unit level variation results across multiple locations and sites. Additional refinement of spatial pattern analyses.	
	Sample design, assumptions about the target population, actual data relationships, and evolving status determination methods will guide selection of appropriate methods.	
Trend Assessment Level	Evaluation of Vital Sign trends over time to detect change. Typically, some form of regression or repeated measures ANOVA will be among the methods used.	PACN Marine Ecologist and Marine Biological Technician with assistance from Park Marine Ecologists.
	Step 1: Integration of landscape level summarization over time.	
	Step 2: Integration of landscape level variation results over time. Includes establishing a direction and rate of change or variation that may be used to provide early warnings of trends in resources condition. Confidence levels of documenting trend will be established.	
	Parametric, nonparametric approaches and models will be used. Trend assessment will also include accounting for influence from drivers and stressors as much as possible.	

The analyses outlined below will be applied to each of the four primary components of this Vital Sign: (1) species richness, (2) abundance, (3) biomass, and (4) diversity.

Analysis Procedures

Table SOP 19.2 summarizes these analytical procedures for the parameters measured at the park level. Analytical procedures should only be conducted by an individual with advanced statistical training and experience with R¹, PRIMER-e®², and ArcGIS Pro®³

¹ R Core Team 2021; <https://www.R-project.org/> (accessed 28 July 2021)

² Clarke and Warwick 2001; <https://primer-e.com> (accessed 28 July 2021)

³ <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview> (accessed 28 July 2021)

Table SOP 19.2. Summary of analytical procedures including parameters measured at the park level.

Level of Analysis	Data Analysis Approach	Frequency of Analysis	Responsible Party
Unit	Summarization and range-of-variation for annual fish species richness, abundance, biomass, and diversity by transect for each park.	Annually, within three months of field data collection for each of the three levels	PACN Marine Ecologist or designee
Landscape	Annual status of fish species richness, abundance, biomass, and diversity for area of interest (e.g., entire park like KAHO, individual islands in NPSA or individual units in WAPA)	Annually, within three months of field data collection for each of the three levels	PACN Marine Ecologist or designee
Trend Assessment	Multi-year change in fish species richness, abundance, biomass, and diversity for both unit level data (transects/sites) and landscape level areas of interest	Annually, within three months of field data collection for each of the three levels	PACN Marine Ecologist or designee

Fish Species Richness: This section will reflect the following organization. Parameters or response variables of interest for the three levels are fish species list and the number of fish species by transect.

Unit Level Analysis: This level is applied for all parameters in the current year only. Text and Graphic Plots by individual transect of mean (or median), standard deviation (or skew and kurtosis), outliers, normal probability plot, and frequency histogram of the distribution. If multiple observers have been collecting data, the summaries should be grouped by observer to evaluate potential observer bias.

Landscape Level Analysis: This level is applied to each parameter of interest in the current year only. Text and graphic plots grouped by area(s) of interest (e.g., multiple transects on individual islands in NPSA or individual park units in WAPA) of mean (or median), standard deviation (or skew and kurtosis), outliers, normal probability plot, and frequency histogram of the distribution. Graphical map of park that displays values for species richness at each of the transects.

Trend Assessment Level Analyses: This level is applied to the current year, previous year, five year, ten year, and other to-be-determined intervals for each parameter of interest. Text and graphic plots grouped by specific transects of mean (or median), standard deviation (or skew and kurtosis). Change detection will include repeated measures ANOVA, regression, time-series evaluation (typically requires 50 or more data points [i.e., years]), or other appropriate technique. Text and graphic plots grouped by area(s) of interest (e.g., entire park such as KAHO, individual islands in NPSA or individual park units in WAPA) of: mean (or median), standard deviation (or skew and kurtosis) of fish variables. It includes a graphical map by park that displays changes in fish species richness by transect.

Abundance: The abundance section will follow the same procedures as for fish species richness except that the response variables of interest for the three levels are taxon list and abundance level ($\text{No. hectare}^{-1} \text{ year}^{-1}$) for each transect.

Biomass: The biomass section will follow the same procedures as for fish species richness except that the response variable of interest for the three levels is biomass level ($\text{metric tons year}^{-1}$) for each transect.

Diversity: The diversity section will follow the same procedures as for fish species richness except that the response variable of interest for the three levels is diversity (H' from the Shannon-Weiner diversity index) for each transect.

Data Analysis and Reporting Overview

Reports identified in Table SOP 19.3 are to be based on collected data and other science-based documentation as per Data Analysis and Reporting Requirements Memo (Gallo 2018⁴) and NPS-75, Natural Resources Inventory and Monitoring guideline⁵. These reports will provide context and explanations for the results presented, as well as include any additional analyses that may result from initial investigations. Pertinent detailed excerpts from Gallo 2018 modified for the Marine Fish Monitoring Protocol are provided below.

Table SOP 19.3. Summary of data analysis and reporting requirements.

Type of Report	Timing	Example Analyses	Purpose	Required?
Published dataset	Annual (or every year data are collected).	Data review and certification.	Ensure data integrity, security, availability, and transparency.	Yes
Resource brief, status report, or update to interactive visualizer	Annual (or interval defined in protocol).	Limited to analyses that are cited in protocol.	Inform stakeholders, maintain relevance	No
Baseline status (and possibly trend if applicable)	Completed once, after one cycle of data collection or not more than 5 years.	Summary or simple statistics of protocol measures and variability to establish definitions/ benchmarks of comparison for future data analyses.	Provide the baseline (if appropriate) for comparison for later analysis. Evaluate metrics relative to benchmarks.	Yes
Protocol evaluation	After every data collection cycle; starting no later than after the second cycle of data collection or 10 years of protocol implementation	Power analyses or simulations to determine if data collected are adequate to meet the objectives of the protocol.	Determine whether protocol objectives, standards, or methods require modification.	Yes
Synthesis or trend report	At least once every three cycles per protocol, starting no later than after three cycles of data collection or 15 years of protocol implementation.	Exploratory analysis to identify or hypothesize relationships between monitoring, inventory, or other datasets and consequent implications for NPS resource management. Analyses to detect linear trends or non-linear changes over time.	Detect or document trends or lack of trends. Address questions, assumptions, or hypotheses defined in protocol objectives or vital signs monitoring plans. Establish or justify new assessment thresholds, management triggers, or recommended management actions.	Yes

⁴ <https://irma.nps.gov/DataStore/Reference/Profile/2252934> (accessed 28 July 2021)

⁵ <https://irma.nps.gov/DataStore/Reference/Profile/622933> (accessed 28 July 2021)

Reporting Requirements

- All PACN Marine Fish Monitoring data must be published on an annual basis and made easily accessible to parks and the public (when appropriate).
- Formal Marine Fish Monitoring reports will only be written to convey something worth reporting, and not because of a perceived requirement that they must produce formal reports every year.
- PACN has opted to publish resource briefs for the Marine Fish Vital Sign in lieu of formal annual data reports to describe results or unusual results or findings, without publishing a data report, given the following requirements are met:
 - Resource briefs are used for routine updates. They cannot be used to communicate significant findings that should be presented in a peer-reviewed report with full and transparent documentation of methods for data collection, processing, and analysis. They may not be used to communicate results of methods or analyses that have not been described in an approved monitoring protocol or peer-reviewed paper or report.
 - All data and associated metadata used as the basis for the brief must be published in the NPS Data Store and cited.
 - Those data must be in an “Accepted” or “Certified” status.
 - Once cited as the basis for a brief or report, datasets cannot be deleted or altered, but they may be versioned if errors are found.
 - Published datasets must have a digital object identifier. Datasets published before NPS has the capability to assign digital object identifiers will be given one as soon as possible.
- Peer Review: All analyses, briefs, or reports must receive an appropriate level of peer review before publication or use by park managers. See the NPS 2017 Peer review, publication, and administrative recordkeeping requirements for scholarly and science information disseminated by the NPS Inventory and Monitoring Division⁶.

⁶ <https://irma.nps.gov/DataStore/Reference/Profile/2242656> (accessed 28 July 2021)

Literature Cited

- Clarke, K., and R. Warwick. 2001. Change in marine communities: an approach to statistical analysis and interpretation. 2nd edition. Primer-E Ltd, Plymouth, UK.
- Cleveland, W.S. 1993. Visualizing data. Hobart Press. Summit, New Jersey, USA.
- Cleveland, W.S. 1994. The elements of graphing data. Hobart Press. Summit, New Jersey, USA.
- Gallo, K. 2018. Data Analysis and Reporting Requirements. Memorandum to Inventory and Monitoring Division Regional and Network Program Managers. 4 pgs.
<https://irma.nps.gov/DataStore/Reference/Profile/2252934> (accessed 28 July 2021).
- National Park Service. 1992. Natural Resources Inventory and Monitoring guideline. NPS-75. National Park Service, Natural Resource Stewardship and Science. Washington, DC.
- National Park Service. 2017. Peer review, publication, and administrative recordkeeping requirements for scholarly and science information disseminated by the NPS Inventory and Monitoring Division. Inventory and Monitoring Division. Fort Collins, CO.
<https://irma.nps.gov/DataStore/Reference/Profile/2242656> (accessed 28 July 2021).
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> (accessed 28 July 2021).
- Tufte, E.R. 1983. The visual display of quantitative information. Graphics Press. Cheshire, Connecticut, USA.
- Tufte, E.R. 1990. Envisioning information. Graphics Press. Cheshire, Connecticut, USA.
- Tufte, E.R. 1997. Visual explanations. Images and quantities, evidence and narrative. Graphics Press. Cheshire, Connecticut, USA.