

South Florida/Caribbean Network

National Park Service
Inventory and Monitoring Division



Standard Operating Procedure CR20

QA/QC Procedures—Version 2.1

Please cite this as:

South Florida/Caribbean Network. 2025. QA/QC procedures—Version 2.1. South Florida/Caribbean Network Standard Operating Procedure NPS/SFCN/SOP—CR20. South Florida/Caribbean Network, Palmetto Bay, Florida.

Summary

This SOP details the steps taken to perform Quality Assurance (QA) and Quality Control (QC) procedures on the data for this protocol, which relate to the Data Processing Levels (DPL) that are assigned to each trip.

QA controls are discussed, including those that are built into the iPad database/application as well as those in the IRMA SFCN Coral Reef Monitoring application. A broader set of QC steps are also described including a verification process before import into the SQL database, the selection and analysis (by a second analyzer) of 10% of video transects, a QC report to compare video transects, and additional QC procedures in the IRMA application, R dashboard, and SQL queries.

Revision Log

Revision Date	Author	Changes Made	Reason for Change	New Version #
1999	Jeff Miller	All checks of data entered occurred in original data entry into Excel, generally consisting of number of points entered matching number of video frames, all codes were valid, and checking that summary category of results (stony coral, macroalgae, gorgonians, etc.) were generally consistent with previous years.	Basic checks of data	1.00
2005	Andrea Atkinson	QA/QC of Site vs LocationID, Year vs EventID, Number of transects analyzed, number of points analyzed on a transect done by exporting to Excel spreadsheet and conducting checks through basic Excel functions.	With the growing number of sites, occasional data entry problems caused mismatches that were not being caught during data entry.	1.01
2008	Brian Witcher	Pilot Analysis of 10% of videography transects by a second person for DRTO sites.	Tested as I&M data managers group was recommending a 10% check on data.	1.02
2010	Brian Witcher	Analysis of 10% of videography transects by a second person added as part of routine processing. At least half of the transects should have high coral cover to allow checking of coral species ID.	Maintains consistency of identification among staff in overall benthic community ID and coral species ID and provides a 10% quality check on data.	1.03
2014	Brian Witcher	Database checks were added for Site vs LocationID, Year vs EventID, Number of transects analyzed, number of points analyzed at the point of data entry and a summary report was created eliminating the need for further checks in Excel.	Pro-actively catching such errors at data entry is more efficient and saves analysis time later.	1.04

Revision Date	Author	Changes Made	Reason for Change	New Version #
2015	Andrea Atkinson	Procedures documented in a written SOP.	Steps were performed previously but not written as an SOP. Tied to I & M 2.00 version of protocol.	2.0
2021	Jeff Miller and Judd Patterson	SOP name changed from 20 to CR20. Reorganized the SOP to draw distinctions between QA and QC. Expanded QC steps to those used on the website, coral dashboard, and SSMS. Steps added for generating the video transects to undergo QC. Steps added for use of IRMA data entry portal. Removed instructions related to Access database.	Renamed according to regional guidance. QA and QC have long been combined, but it is helpful to differentiate. We are actively developing and refining QC procedures. Random transects are now selected with an R script. Updated data entry and reporting method. Use of an iPad removes the need for data entry and moves many QA steps to the field.	2.1
2023	Mike Feeley	<i>Diadema</i> pilot study, clarify disease QC.	To assess the QA of the <i>Diadema</i> counts.	2.1

Contents

	Page
Introduction.....	5
Data Processing Levels	5
Quality Assurance	5
QA of Transect Covariates	5
QA of Coral Video Data.....	6
Verification of Transect Covariates.....	8
Disease Verification	8
Species List Verification	9
CDR Verification.....	9
Quality Control	10
QC of Coral Video Data	10
Phase 1 – Database Administrator.....	10
Phase 2 - Analyst.....	11
QC Report for Coral Video Data.....	13
Additional QC Procedures.....	16
SFCN Coral Reef Monitoring Application Review	17
R Dashboard Data Warnings	17
SQL Server Management Studio Queries	18
R Markdown QC	18
Literature Cited	26

Introduction

Quality assurance (QA) efforts are preventative in nature and are intended to stop incorrect data from making it into databases and analyses. Quality control (QC) efforts are intended to locate incorrect data so they can be corrected or flagged. Within this protocol these two distinct procedures have sometimes been lumped together and referred to as QA/QC.

Data Processing Levels

IMD Certification Guidelines for data products utilize the Data Processing Level (DPL) as a standardized way to define the degree of data validation or transformation that has been performed on a data set (NPS 2016). There are three DPL levels that are currently available for use for this protocol:

- Raw – Data in their state of original collection have been added to the database and server. These data generally have been collected following some QA procedures, but have not undergone any QC procedures.
- Provisional – Data that have undergone all planned QA procedures, have been entered into the database and passed verification procedures to ensure they match source data (field forms, etc.). These data have not yet undergone all QC procedures.
- Accepted – Data have undergone all QA and QC procedures as defined in the published protocol and associated SOPs.

The SQL Coral Database tracks DPL's for four separate components of each monitoring trip. These components are the video analysis, disease data, coral species data, and colony counts/*Diadema*/rugosity (also referred to as CDR). Data defaults to a 'Raw' DPL when imported/entered, and must be manually changed to 'Provisional' and 'Accepted' at the appropriate steps. Once data has reached the Accepted level, all subsequent changes to any value must be tracked and properly annotated in an edit log.

Note: When a Quality Assurance Plan (QAP) has been published for this protocol, the Accepted level will be retired in favor of Pre-certified and Certified.

Quality Assurance

QA of Transect Covariates

The best time to prevent incorrect data entry is during collection, while the diver is still in the field and the data collection event can be most easily recalled. In the case of the sampled covariates along the transects (rugosity, urchin counts, coral disease lesions, coral colony counts, and species lists), these data are collected in the FileMaker Go application on an underwater iPad. The database used by this application has drop-down lists, buttons, acceptable data ranges, and warning messages that limit incorrect entries. There is also a 100% review of entered data to check for both completeness and any

immediate outliers. Data sheets are used only if the iPad is unavailable or fails, and when possible are immediately entered to the iPad after the dive to ensure the data are passed through the same QA controls. If datasheets are used and an iPad is not available for entry even back in the office, the data are entered directly into the IRMA coral application.

After entry in the field database, the data are exported from FileMaker Pro to intermediate Excel spreadsheets (one for disease, one for species lists, and a third for colony counts/*Diadema*/rugosity) which will eventually be imported into the SQL database. Before the import occurs, these spreadsheets are sent to the person who did the initial data collection for a 100% review of entered data to check for both completeness and any immediate outliers. These exports are stored on the server in the following folder:

X:\SFCN\Vital_Signs\Marine_benthic_communities\CoralMonitoring\data\iPad\YYYY

Each trip is a subfolder, with individual exports named according to the appropriate component of the monitoring:

- CDR_export_YYYYMMDD_XXXX.xlsx
- Disease_export_YYYYMMDD_XXXX.xlsx
- Species_export_YYYYMMDD_XXXX.xlsx

These exports are reviewed for mistakes. Care is paid to detecting duplicate transects, mistakes in incrementing colony numbers or lesion letters, or field notes indicating that a value needs to be adjusted. The purpose of this review is to catch any errors that may have slipped past the existing FileMaker Go QA procedures, and in many cases may lead to refinements in digital data collection. Missing values are explained in the trip report and logged in a spreadsheet kept on the server. That spreadsheet, named ‘Coral_Database_Data_Gaps.xlsx’ contains all known data gaps and associated notes, and is on the server at:

X:\SFCN\Vital_Signs\Marine_benthic_communities\CoralMonitoring\data

For example, if bottom time didn’t allow colony counts on transect 7, those fields will be blank. Note the reason for the lack of count in the trip report and the ‘Coral_Database_Data_Gaps.xlsx.’ If *Diadema*, colony count, or rugosity are **not** collected, those cells will be **blank**, not zero.

When this review is complete, the spreadsheets are returned to Data Management with any corrections so they can be imported into the SQL database. A small set of QA checks, such as coral and disease code verification, are performed automatically during the import process.

QA of Coral Video Data

While entering each point (A–J) on each benthic image (SOP CR12 *Conducting Analysis of HD Images—Version 2.1* [SFCN 2025a]), the drop-down list of acceptable taxon codes is shown and is automatically limited to potential matches based on the entered characters. Only known taxon codes are acceptable, and a completely invalid four-letter code will not be saved. The autofill feature can be

useful, but analysts also need to understand its behavior. If a code matches the letters typed, then pressing enter or tab will select that entry. An example is shown in Figure 1, below, where typing 'mca' and pressing tab will select the MCAV code without needing to type the final "v." All points must be entered before the analyst can advance to the next frame. This assures there are no omitted points in a frame.

Note : Video Frame data will be saved automatically when the cell is changed.

Frame: 1 of 1 ◀ ▶ Add Frame Delete Last Frame

Point	Species Code	Bleaching	Disease	Notes
A	mca		No	
B	MCAV - Montastrea cavernosa - coral		No	
C			No	
D			No	
E			No	
F			No	
G			No	
H			No	
I			No	
J			No	

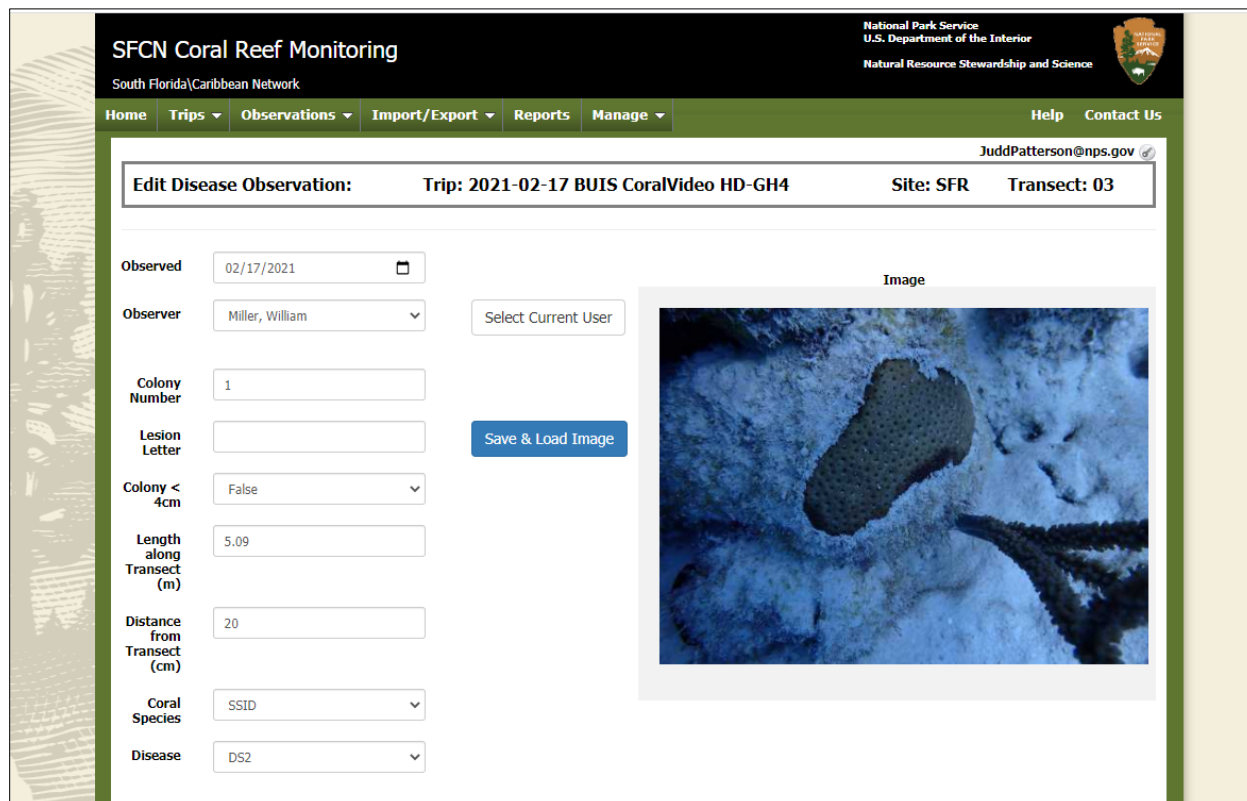
Figure 1. Screen capture showing selection to QA List.

Verification of Transect Covariates

The coral disease, species lists, and CDR components of the database support the tracking of both a Verifier (a name) and a Verified Date. After the import or entry of each component, the following steps are taken to complete verification.

Disease Verification

Go to each transect and click the “Edit Disease” link to review the coral disease lesions. A diver other than the one who did the original data collection reviews the data fields and the associated photo (see Figure 2).



The screenshot shows the 'SFCN Coral Reef Monitoring' web application. The header includes the National Park Service logo and the text 'South Florida/Caribbean Network'. The navigation bar has links for Home, Trips, Observations, Import/Export, Reports, Manage, Help, and Contact Us. The user is logged in as JuddPatterson@nps.gov. The main form is titled 'Edit Disease Observation' and shows data for Trip: 2021-02-17 BUIS CoralVideo HD-GH4, Site: SFR, and Transect: 03. The form fields on the left include: Observed (02/17/2021), Observer (Miller, William), Colony Number (1), Lesion Letter (empty), Colony < 4cm (False), Length along Transect (m) (5.09), Distance from Transect (cm) (20), Coral Species (SSID), and Disease (DS2). A 'Save & Load Image' button is located next to the Lesion Letter field. On the right, a thumbnail photo labeled 'Image' shows a coral colony with a dark, irregular lesion.

Figure 2. Screen capture showing a disease observation with data along the left and thumbnail photo on the right.

If an entry has lesion information but the thumbnail photograph does not appear to the right of the data, the issue must be investigated. The photographs are stored on the SFCN shared drive and made available to the IRMA online application. The actual storage location is:

X:\SFCN\Vital_Signs\Marine_benthic_communities\CoralMonitoring\images\disease

The disease photo folders and files are named according to a standard (see SOP CR24 *iPad Preparation, Data Entry, and Download—Version 1.0* [SFCN 2025b]). The most common reasons for a missing disease photograph are improperly named photographs on the server, the SFCN server is not reachable due to a network/internet issue, or the subfolder of ‘web’ images has not been

created. This ‘web’ subfolder provides smaller thumbnails for quicker load times. Every effort needs to be made, in coordination with the Data Manager, to ensure that the photographs load within the application so that the verification can continue.

Once the photograph is visible, review it to confirm the entered coral species, disease type, and associated metrics (e.g., length and width) are correct. All recorded disease metrics are also verified by a second independent observer. Disease lesion length and width measurements will be proofed by scaling the recorded lesion dimensions to published average species polyp size. This is an approximate verification of lesion size but will identify gross errors. Any necessary changes are made in concurrence with the Marine Ecologist. Once the original entries are confirmed, the Verifier field is filled in by clicking the ‘Select Current User’ button, and the Verified Date field is filled in by clicking the ‘Insert Current Date’ button (see Figure 3).

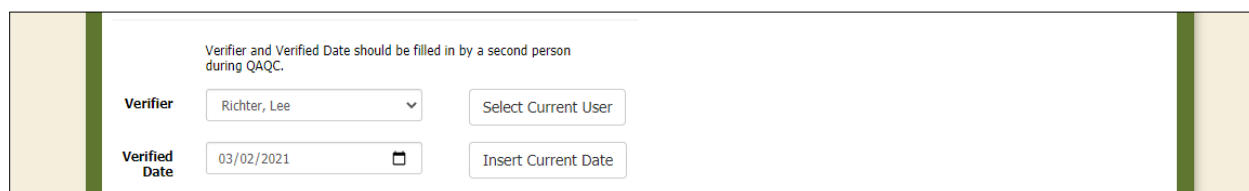
A screenshot of a web application interface. At the top, a message reads: "Verifier and Verified Date should be filled in by a second person during QAQC." Below this, there are two rows of input fields. The first row is labeled "Verifier" and contains a dropdown menu with "Richter, Lee" selected, and a button labeled "Select Current User". The second row is labeled "Verified Date" and contains a text input field with "03/02/2021" and a calendar icon, and a button labeled "Insert Current Date". The interface has a light beige background with green vertical bars on the left and right sides.

Figure 3. Screen capture showing the Verifier and Verified Date fields and associated buttons.

Species List Verification

Go to each transect and click the “Edit Species” link to review the coral species lists. Once changes have been made, or the original entries confirmed, the Verifier field is filled in by clicking the ‘Select Current User’ button, and the Verified Date field is filled in by clicking the ‘Insert Current Date’ button.

CDR Verification

Go to each transect and click the “Edit CDR” link to review the stony coral counts, *Diadema* count, and rugosity values that were collected. Pay particular attention to large discrepancies between the two stony coral count numbers and confer with the individual divers and the Marine Ecologist as necessary. Once changes have been made, or the original entries confirmed, the Verifier field is filled in by clicking the ‘Select Current User’ button, and the Verified Date field is filled in by clicking the ‘Insert Current Date’ button. A pilot study is planned for one year of sampling after publication of the protocol and SOPs to assess the accuracy of *Diadema* counts. The pilot study will consist of recording total *Diadema* counts by two independent observers on all transects on all sites for one year. In addition, double observer *Diadema* counts will be performed on 1–2 transects per site or a maximum of 10% of transects per site that have one or more *Diadema* urchins present. The purpose of the pilot study is to assess the accuracy of our single observer *Diadema* counts across of a variety of coral reef habitat in Florida and the U.S. Virgin Islands.

Quality Control

QC of Coral Video Data

One of the primary QC techniques used for coral video is the selection of approximately 10% of transects to be reanalyzed by a second person. The results of this exercise are compared through an automated report. These steps are broken into a preliminary phase for the Data Manager, and a second phase for the analyzer.

Note: Some of the labels and names in the online web application currently confuse QA/QC terms, and many QC operations are labeled as QA. These will be cleaned up in a future release.

Phase 1 – Database Administrator

1. For the site(s) of interest, use the Coral Dashboard (SOP CR21 *Data Summaries, Analysis, and Reporting—Version 3.0* [SFCN 2025c]) to determine the transects from the most recent visit to that site that had 10% or more of live coral cover. These will be used below to ensure that at least one transect has high coral cover. If all transects at a site fall below 10%, then drop the ‘high coral cover’ threshold to 5% and continue.
2. Open RStudio and browse to open the ‘Coral_QC_Rnd_Transect_Selection.R’ file. It is located here:
 - a. X:\SFCN\Vital_Signs\Marine_benthic_communities\CoralMonitoring\analysis\R\QC_select_transects
3. The R script is meant to be run for one site at a time (or for the entire park in the case of Dry Tortugas National Park [DRTO]). Near the top of the script there is a large block of code that lists all the sites and their corresponding transects. Choose a single site of interest and uncomment the vectors ‘transects’ and transects_high’ for that site. Ensure that only one such pair is uncommented in the script.
 - a. For example, if running this for HA, uncomment ‘transects’ and ‘transects_high’ for HA and ensure the other site vectors (for NF, MB, TK, etc.) are commented out.

Note: For Dry Tortugas National Park, where the sites visited can vary a lot by year, be sure to carefully customize the vectors to represent the sites visited during the year of interest.

4. Review the ‘transects_high’ vector that was uncommented in step 3 to ensure it only contains transects with high coral cover that were identified in step 1. Remove or add transects to this list as necessary.
5. Run the entire script.
 - a. The script will first print the fully random transect(s).
 - b. The script will then print the high coral cover random transect(s).

- Open the IRMA SFCN Coral Monitoring application (see SOP CR13 *Coral Reef Video Transect Database User's Guide—Version 3.0* [SFCN 2025d]) and navigate to the site visit of interest. On the 'Trip Transects' page, put a check mark in the 'QA' column for the transects identified in step 5. An 'Edit QA' action will appear immediately next to this checkbox and is described in Phase 2 below.

Note: This step must be performed by an 'Administrator' in the coral database. Typically, this is the Data Manager, Assistant Data Manager, and Marine Ecologist.

Phase 2 - Analyst

- Open the on-line IRMA SFCN Coral Monitoring application (see SOP CR13 [SFCN 2025d]).
- Select Trips, from the drop-down menu, then QAQC list (Figure 4).

SFCN Coral Reef Monitoring
South Florida\Caribbean Network

National Park Service
U.S. Department of the Interior
Natural Resource Stewardship and Science

Home Trips Observations Reports Help Contact Us

Existing Trips
Create New Trip
QAQC List

WMiller@nps.gov

Welcome to SFCN Coral Reef Monitoring

Year Summary Year **2019** Park **-- Select Park --**

Park (Site)	Trips to Site	Transects	Video	Disease	Species	CDR
Biscayne National Park (AM)	1	20	22	20	20	20
Biscayne National Park (BB)	1	20	22	20	20	20
Buck Island Reef National Monument (BAR01)	1	4	4	4	4	4
Buck Island Reef National Monument (BAR02)	1	4	5	4	4	4
Buck Island Reef National Monument (BAR03)	1	4	5	4	4	4
Buck Island Reef National Monument (BAR04)	1	4	4	4	4	4

Figure 4. Screen capture showing selection of QAQC List.

- Use the filters for Year, Trip, and Site to limit down to the trip where QC needs to be entered or reviewed (Figure 5).

SFCN Coral Reef Monitoring

South Florida\Caribbean Network

National Park Service
U.S. Department of the Interior
Natural Resource Stewardship and Science



HomeTripsObservationsReports

HelpContact Us

WMiller@nps.gov

QAQC List

Filters:

Year2019

Trip-- Select Trip --

Site-- Select Site --

Clear Filters

Trip Name	Year	Site	Transect	QA Action
2019-02-12 BUIS CoralVideo HD-GH4	2019	SFR	03	Edit QA
2019-02-12 BUIS CoralVideo HD-GH4	2019	SFR	07	Edit QA
2019-01-31 VIIS CoralVideo HD-GH4	2019	HA	09	Enter QA
2019-01-31 VIIS CoralVideo HD-GH4	2019	HA	15	Enter QA
2019-05-03 BUIS CoralVideo HD-GH4	2019	BAR02	04	Edit QA
2019-05-03 BUIS CoralVideo HD-GH4	2019	BAR03	01	Edit QA
2019-03-26 VIIS CoralVideo HD-GH4	2019	NF	01	Edit QA
2019-03-26 VIIS CoralVideo HD-GH4	2019	NF	16	Edit QA

Figure 5. Screen capture showing selection to QAQC Site and Transect.

4. If partial or complete QC data exist, the “QA Action” will indicate “Edit QA.” If no data yet exist, then “Enter QA” will be shown. In either case, clicking on this action will show the video data entry form (see Figure 6).
 - a. For QA Date – enter the date the QA analysis was begun for this transect (or use the “Insert Current Date” select box), **not** the date the transect was sampled/filmed.
 - b. For QA Analyzer, enter your name (or use the “Select Current User” select box).
 - c. Enter the number of frames that are being analyzed in this transect.
 - d. Click the box “Create QA Observation”. For more details, see instructions in SOP CR13 (SFCN 2025d).

SFCN Coral Reef Monitoring
South Florida\Caribbean Network

National Park Service
U.S. Department of the Interior
Natural Resource Stewardship and Science

Home | Trips | Observations | Reports | Help | Contact Us

WMiller@nps.gov

Create QA Observation: Trip: 2019-01-31 VIIS CoralVideo HD-GH4 Site: HA Transect: 09

QA Date:

QA Analyzer:

Number of Frames:

Video Frames Note : Video Frame data will be saved automatically when the cell is changed.

Figure 6. Screen capture showing “Enter QA” data entry form.

5. The images for analysis are opened and displayed following SOP CR12 (SFCN 2025a). Once a transect is complete or the appropriate edits have been made, return to QAQC list, and complete the next transect. When all QC entries are complete you can close the IRMA data portal or go on to review the QAQC Report (below).

QC Report for Coral Video Data

This report is run by the video analyst after the QC transects have been completed. The results are reviewed by the Protocol Lead to determine if any further work or training is necessary.

1. From within the IRMA SFCN Coral Reef Monitoring portal, select the Reports drop-down menu, then select ‘CoralVideo QAQC Comparison’ (Figure 7).

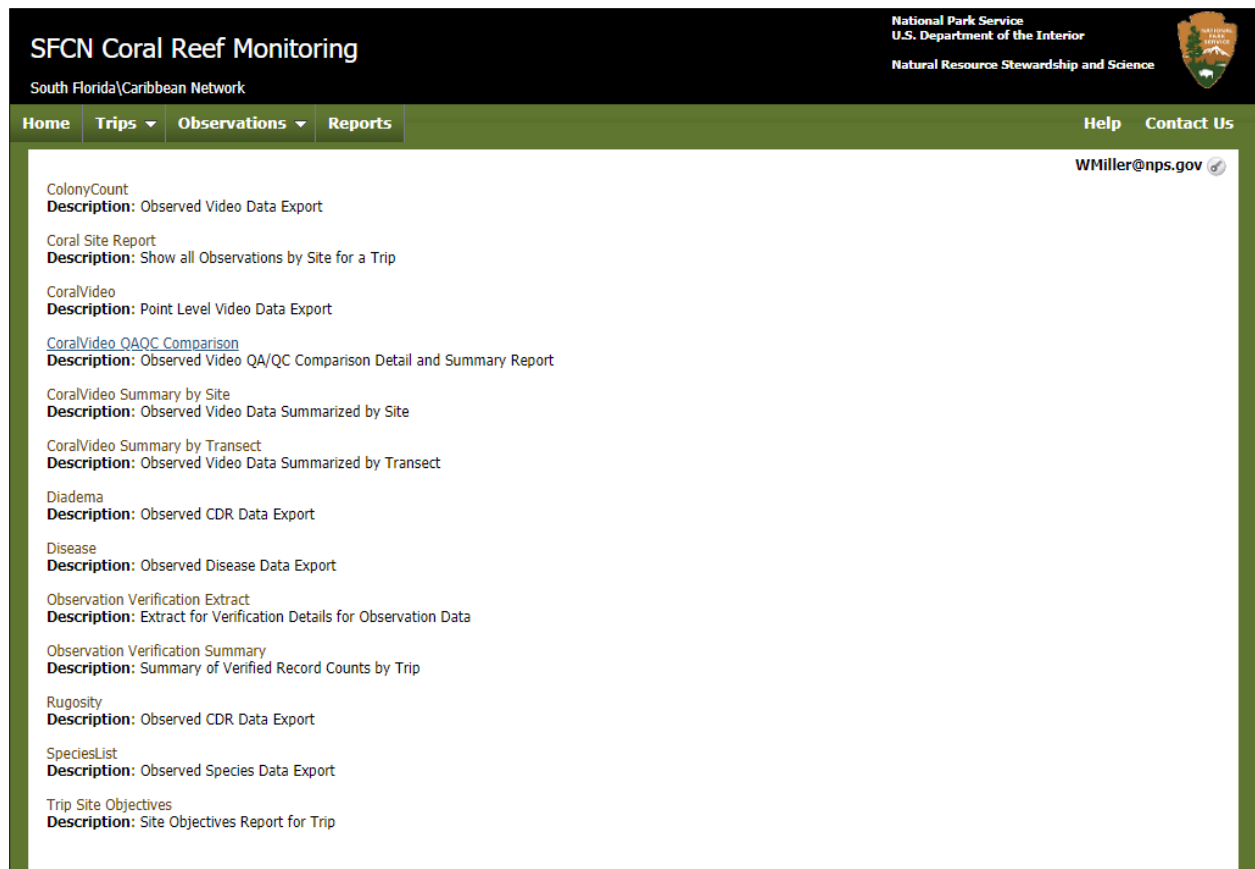


Figure 7. Screen capture showing “CoralVideo QAQC Comparison” selection.

2. A report form opens (Figure 8). Fill in the appropriate choices for Park, Year, and QAQC Trip Name. There are two options for reporting. There is a summary, which indicates the numeric and percentage of dots by category, and the detail shows the dot by dot comparison with mismatches shown in red. Select the desired report type, then click ‘View Report.’

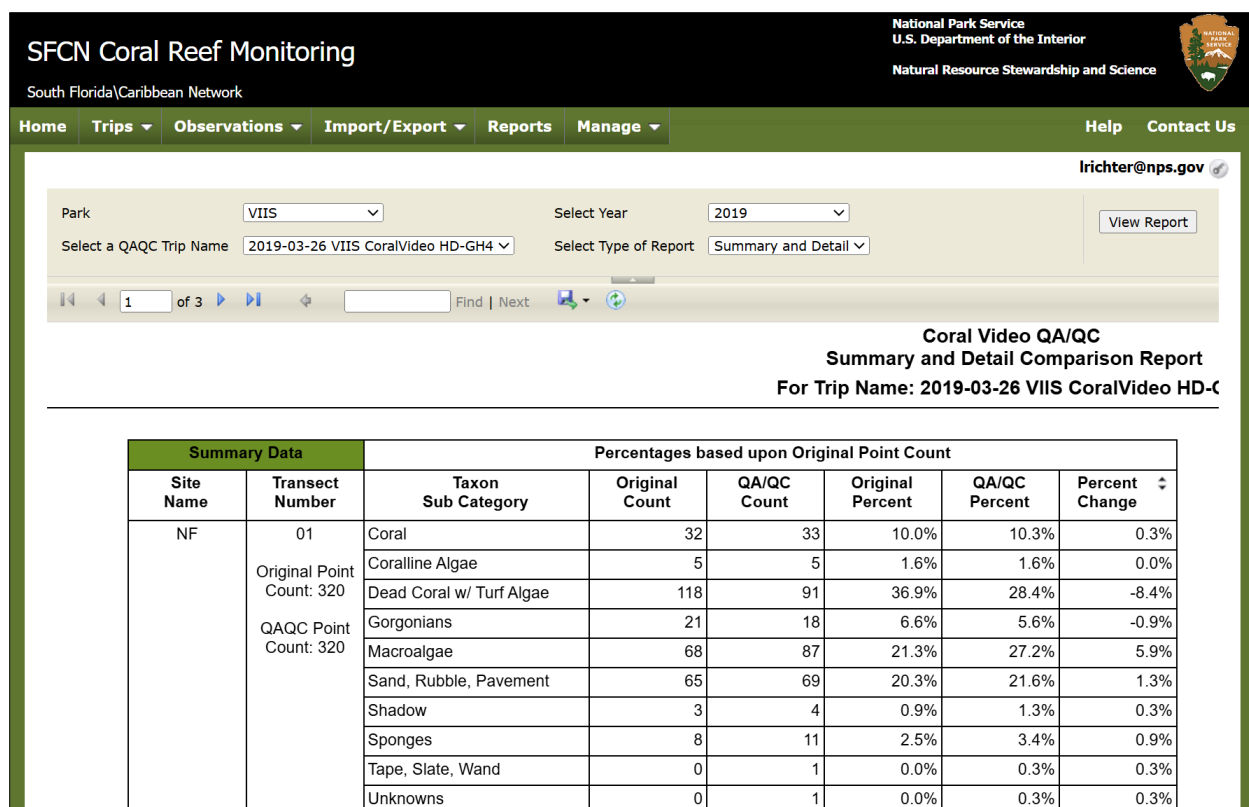


Figure 8. Screen capture showing QA/QC report summary.

- Once the reports display, they can be exported by clicking on the disk icon in the upper middle of the screen (Figure 9). There are several format options, and the CSV option is very useful because it copies the information such that it can be brought into an Excel sheet without formatting. From there, the comparison of the details can be made.

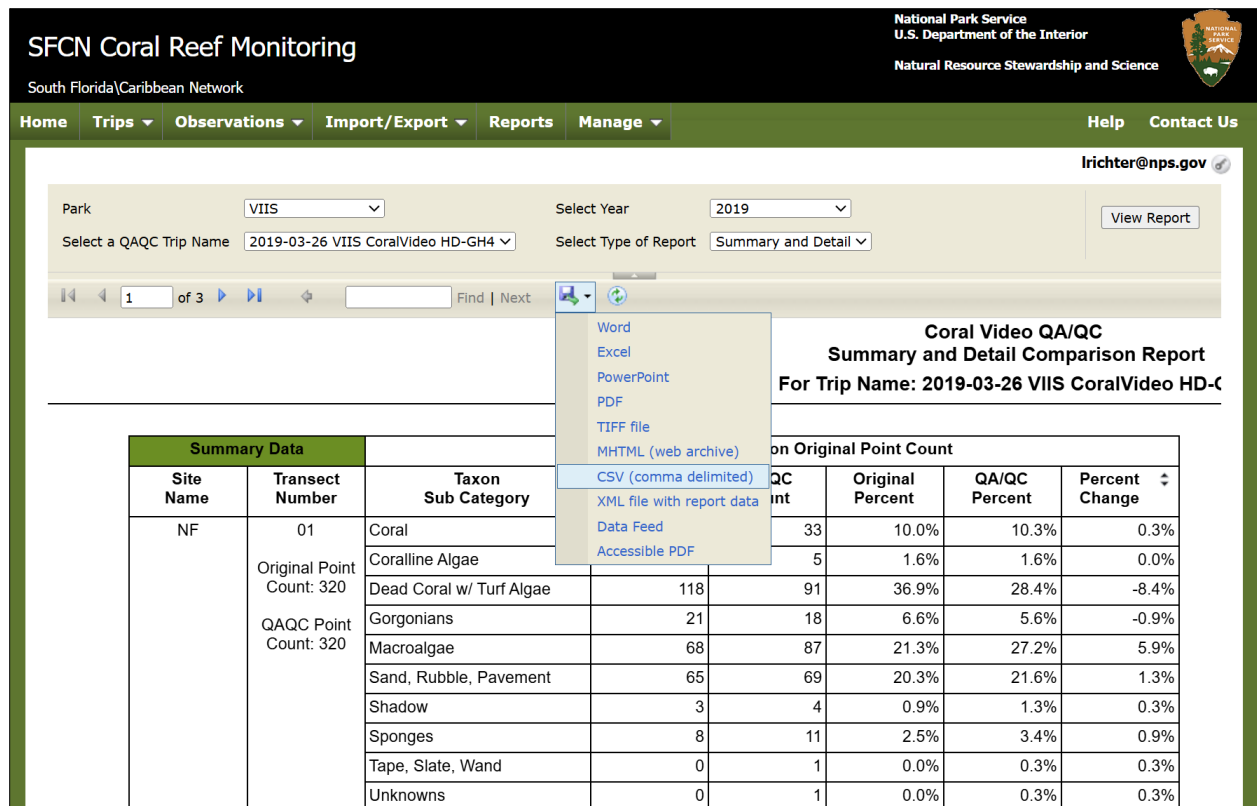


Figure 9. Screen capture showing QA/QC report summary and detail. Middle of screen shows the save as or export options for this report (CSV selected).

4. Evaluating the QA/QC report incorporates a variety of comparisons. A 10% change threshold is applied for all Taxon Sub Categories **with the exception** of Coral, Gorgonians, and Sponges, which are set at 5%. Evaluations are conducted by the original analyst. Once complete, a short written summary of the results is provided to the Marine Ecologist.
 - a. Do the coral points closely match? Open the images containing coral to see if the discrepancy can be explained (point on the corner or edge) or is an error.
 - b. Evaluate the magnitude of the algal components; are the relative magnitude of each similar? For example, is most of this category macroalgae or turf algae?
 - c. Remaining components can be directly compared for similarity. Sponges, Gorgonians, Coralline algae, and Cyanobacteria can all be visually evaluated for explanation of substantial differences. The number of dots assigned to SHADOW or UNKNOWN often vary among observers.

Additional QC Procedures

After all data have been entered and imported into the database, there are additional QC steps that the Protocol Lead and Data Manager take through the Coral Dashboard (SOP CR21 [SFCN 2025c]), R scripts, and SQL Server Management Studio (SSMS) to summarize and explore the dataset. This

process attempts to highlight issues such as unexpected data gaps, outliers, and unusual trends. In some cases, it might be possible to correct these errors, however when that is not possible, the data are flagged appropriately to inform future users. These dataset reviews are completed annually before each component of a trip can have the DPL changed to ‘Accepted’ or ‘Certified.’

SFCN Coral Reef Monitoring Application Review

The simplest QC check is to ensure annual data completeness. This is handled initially by filtering the main page on the IRMA SFCN Coral Reef Monitoring Application to a park unit and reviewing the number of transects for each category under the ‘Video,’ ‘Disease,’ ‘Species,’ and ‘CDR’ labels. Each category should match the number of transects at that site, except for the Video, which will also include any QA/QC transects (e.g., BISC AM will show 22 out of 20 transects for Video because there are 20 transects plus two additional QA/QC transects for this site). Click on any of the ‘Park (Site)’ labels to see individual trips. Click on ‘Go to Transects’ to drill down an additional level and review each transect.

R Dashboard Data Warnings

Run the R dashboard (see SOP 21 [SFCN 2025c]) and after creating/importing a summary RDS file, switch to the ‘Data Warnings’ tab. This will display several data checks that are completed. These include:

- Missing taxon codes or recorded as “No Taxon” – All “dots” should be filled in with a valid TaxonCode. Any that were left blank will be shown here.
- Bleaching codes assigned to non-corals – Bleaching codes only apply to corals. Any that were accidentally assigned to non-corals will be shown here.
- Site visits with transect counts other than 4 or 20 – Sites have 4 or 20 transects. Any other number typically indicates that a transect was skipped.
- Transect surveys with a total taxon count not divisible by 10 – Each image has 10 “dots” that need to be analyzed, so the total number of taxa is always divisible by 10. A count that is not divisible by 10 indicates skipped entries that need to be reviewed.
- Transect surveys with taxon counts less than 200 or greater than 480 – Over time this is the typical range for transect level taxon counts. Any number above or below this needs review.
- Transect surveys with shadow exceeding 5% of the total taxon counts – This check is experimental and is an attempt to detect transects with video or analysis problems. A value above 5% is not necessarily bad, but particularly high numbers are reviewed to ensure that the video quality is adequate and that the analyzer is experienced.
- Transect surveys with unknown (e.g., “UNK”) exceeding 5% of total taxon counts – This check is also experimental and is an attempt to detect transects with video or analysis problems. A value above 5% is not necessarily bad, but particularly high numbers are reviewed to ensure that the video quality is adequate and that the analyzer is experienced.

SQL Server Management Studio Queries

Additionally, there are custom SQL queries that are used to quickly identify issues in the database. These queries are stored on the server in the following folder:

X:\SFCN\Vital_Signs\Marine_benthic_communities\CoralMonitoring\data\database_queries

Using these queries has several advantages over the R Dashboard mentioned above. By running directly on the SQL database, they run very quickly and there is no need for intermediate CSV files. Changes made in the IRMA coral application can be reviewed immediately by refreshing the query.

To run these queries, open SSMS and logon to the IRMA server (INP2300VIRMASQL\IRMA) utilizing an account that has been given read-only permission to the database. Open the following queries one by one and execute them, review the results, and work with the analyzers to fill in any blanks or make appropriate adjustments:

- QC_Trips_with_Blank_Analyzer.sql – This query pulls up trips where the video analyzer has not been recorded. This is relatively common for analysis done prior to 2013, but should no longer occur. Review any recent entries from this query and work with the analyzers to fix.
- QC_Blank_Taxon.sql – This query searches for any TaxonCodes that may have been left blank. Review any recent entries from this query and work with the analyzers to fix.
- QC_Bleached_NonCorals.sql – This query looks for non-corals that have been given a bleaching code during the video analysis. Review any entries from this query and work with the analyzer to fix.
- QC_CoralDots_vs_SpeciesList.sql – This query creates a list of coral codes identified both during video analysis and the in-situ species lists. These results provide a way to compare the coral species encountered using these two methods. Any species missing from the species lists, but encountered in the video analysis are reviewed to determine if they should be added to the species list.

R Markdown QC

The exploration of principles and techniques of Exploratory Data Analysis (EDA) led to a new tool that has taken shape in R Markdown. This tool is a hybrid QC/EDA tool that is intended to grow through time to include most (if not all) required QC procedures for the entire coral dataset. It is built upon the experience gained from the long process of cleaning this 20+ year dataset, and if run frequently it will allow items to be corrected or flagged closer to the time of creation. It also includes several dynamic graphs and heatmaps that update based upon drop-down options and require exploration in order to reveal relevant patterns.

This tool depends upon several exports that are pulled directly from the SFCN_Coral database and are saved as R dataset (RDS) files. These preparation steps are meant to be run by a Data Manager who has read-only access to the SQL database and the ability to establish an ODBC connection. Database permission can be modified by contacting the current NPS Data Center database

administrator. Once these RDS files have been created, anyone with R and RStudio can run the ‘Viewing Steps’ and interact with the results.

Preparation Steps

1. To download a new database snapshot, first obtain read-only access to the production version of the SFCN_Coral database that runs in the data center. As of 2021 this server is: INP2300VIRMASQL\IRMA.
2. Open ‘ODBC Data Sources (64-bit)’ in Windows to create a new data source that can be accessed by R (see Figure 10).

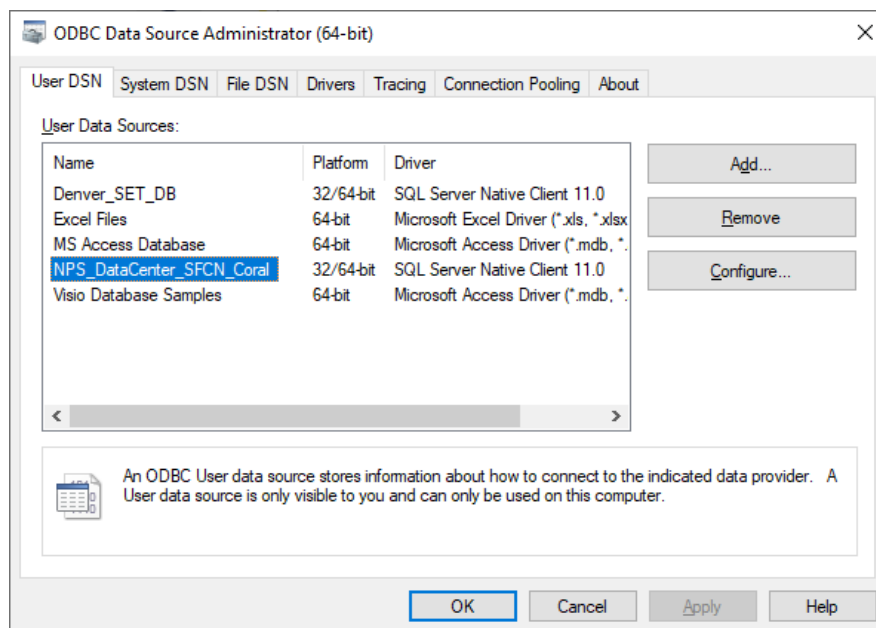


Figure 10. Screen capture showing the ‘ODBC Data Source Administrator (64-bit).’

- a. Click ‘Add...’ to add a new data source.
- b. Choose ‘SQL Server Native Client 11.0’ or the current version of SQL and choose ‘Finish’.
- c. Fill in the Name as ‘NPS_DataCenter_SFCN_Coral’, Description as ‘NPS Data Center - SFCN_Coral Database’, and Server as ‘INP2300VIRMASQL\IRMA’. Click Next.
- d. Choose to login ‘With Integrated Windows authentication’ so that it uses your current PIV card information for authentication. Click Next.
- e. Choose ‘Changed the default database to:’ and select ‘SFCN_Coral’ from the options. The other default options are acceptable, but ensure that ‘Application intent’ is set to READONLY. Click Next and then Finish.

- f. The connection test screen will appear (see Figure 9). Click the ‘Test Data Source’ button and if everything works you’ll see a “TESTS COMPLETED SUCCESSFULLY!” message (see Figure 11). Click ‘Ok’ three times to complete the process.

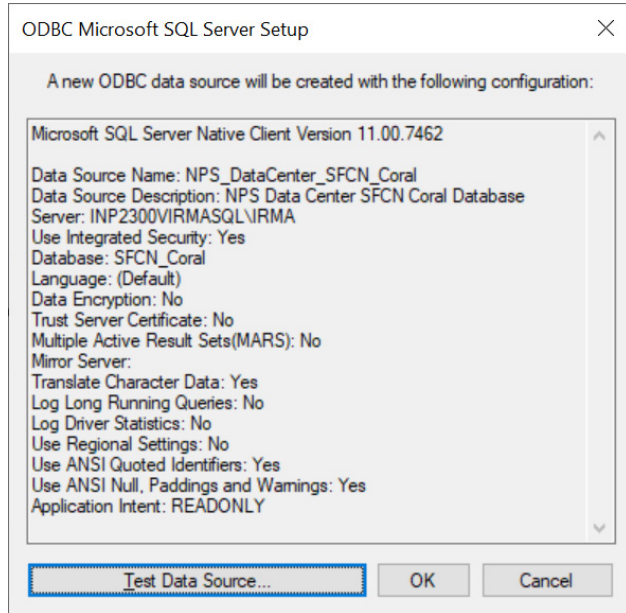


Figure 11. Screen capture showing ‘Test Data Source’ option.

3. Open the R script named ‘Coral_SQL_data_import.R’ from this folder:

X:\SFCN\Vital_Signs\Marine_benthic_communities\CoralMonitoring\analysis\R\QC_R_Markdown

4. This script contains 5 SQL statements. Each is run in succession and the results are saved to RDS files in a ‘data’ subfolder with a ‘YYYYMMDD’ date stamp. As of 2021 it takes approximately five minutes for the script to run. The script can be closed as soon it is complete.

Viewing Steps

1. To run this tool, open the latest version of the R script ‘QC_dashboard_YYYYMMDD.Rmd’ from:

X:\SFCN\Vital_Signs\Marine_benthic_communities\CoralMonitoring\analysis\R\QC_R_Markdown

2. Look in the first section of code and there is a Parameters section that can be adjusted to alter the scope of the QC checks.
 - a. StartYear and EndYear: The first and last year to include. Leave blank to include all years in the input files.

- b. ParkList: A vector of one or many park codes to include.
 - c. DiseasePhotoCheck: TRUE means that the name of each disease photo is calculated based on the lesion data and then a check is run to find that photo on the server. This takes a few extra minutes to run (especially over the VPN), so setting this value to FALSE will skip this check to save time.
 - d. SnapshotDate: This script relies on predownloaded RDS file “snapshots” of the database that are stored on the server. Changing this date tells the script to look for RDS files with a particular date stamp (see ‘Preparation Steps’ above).
3. Within RStudio, click the ‘Run Document’ option in the upper-left side of the screen (see Figure 12) to start the script.

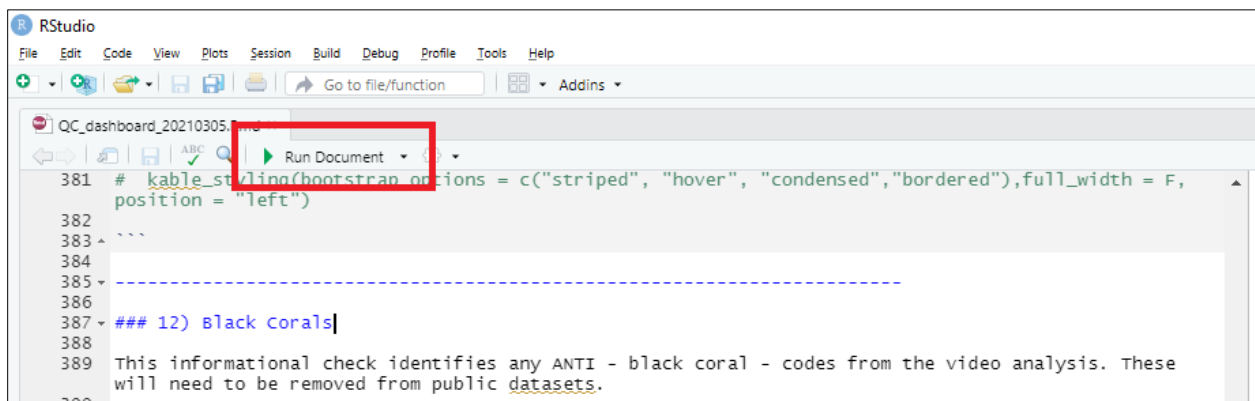


Figure 12. Screen capture showing the R Shiny QC check results that appears when running the QC R Markdown script. This can be moved to a web browser by clicking the ‘Open in Browser’ option in the upper left.

4. A new window should pop up, and after a few minutes of performing the calculations, the ‘SFCN Coral QC Checks’ results should appear (see Figure 13).

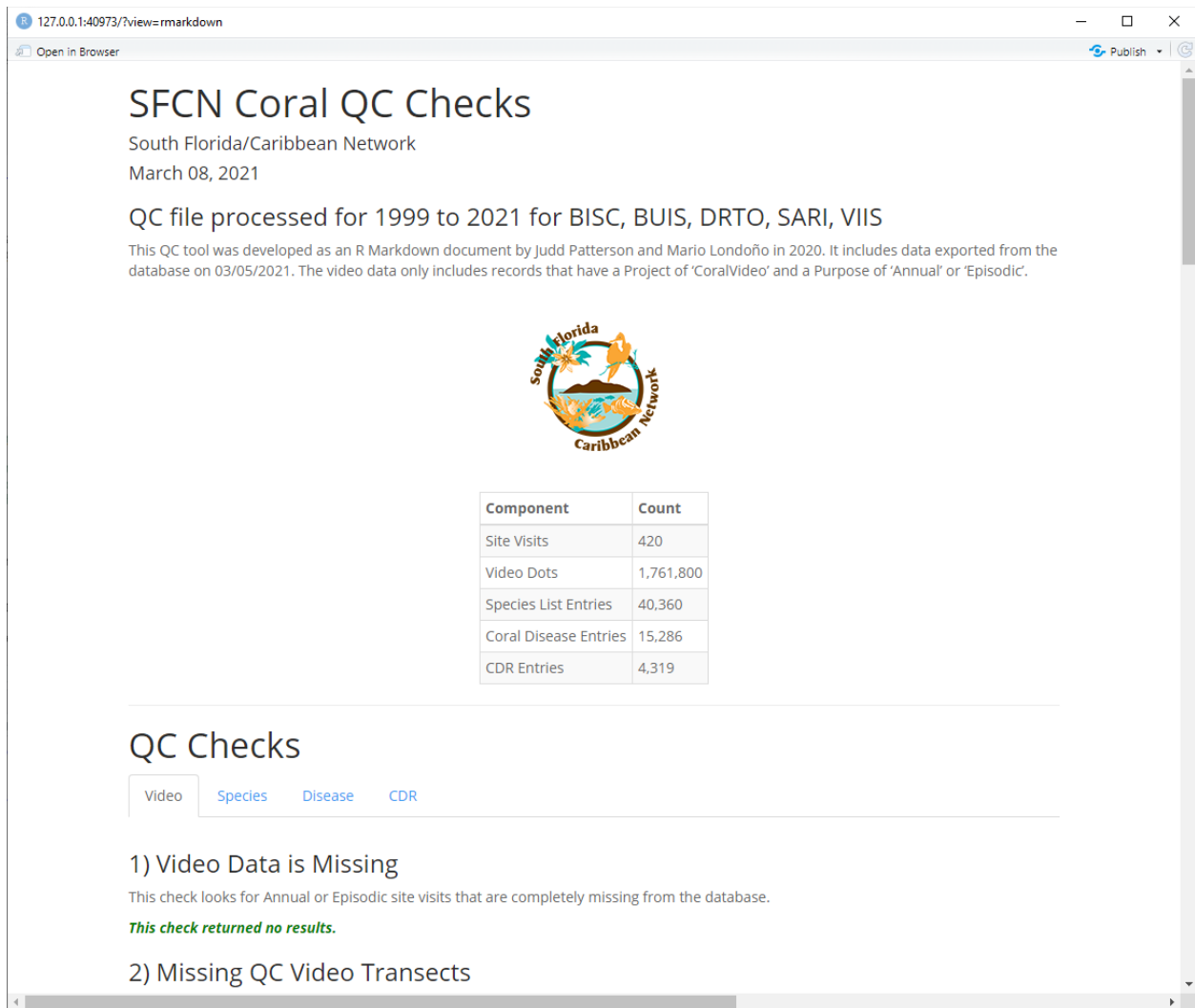


Figure 13. Screen capture showing the R Shiny QC check results that appears when running the QC R Markdown script. This can be moved to a web browser by clicking the 'Open in Browser' option in the upper left.

5. Under the 'QC Checks' heading there are tabs for Video, Species, Disease, and CDR that organize the results. Some checks offer tabular results, which can be easily copied to the clipboard or exported. Other checks offer dynamic graphs that can be updated to view the dataset in a new way. Work through each check, and determine the necessary corrections, flagging, or other documentation that are appropriate.

The QC/EDA checks included in this R Markdown tool include the following:

- Video
 1. Missing Video transects – Checks for Annual or Episodic site visits with no video transects entered into the database.

2. Missing QC Video transects – Calculates the total number of video transects and QC video transects entered for each site visit, and shows where the 10% requirement has not been met.
 3. Dot Count Per Image – Checks for video frames that do not have 10 points/dots.
 4. Blank TaxonCode – Checks for entries with an NA or ‘No Taxon’ for TaxonCode.
 5. High/Low Frame Counts – Checks for an unexpected number for frames/images per transect. This check calculates the interquartile range (IQR). Outliers are values below $Q1 - (1.5 \times IQR)$ and $Q3 + (1.5 \times IQR)$.
 6. Bleached Non-Corals – Checks for non-corals with a bleaching code.
 7. Unexpected Transect Count – Checks for site visits with a transect count other than 4 or 20.
 8. Transects with Blank Analyzer.sql – This query pulls up trips where the video analyzer has not been recorded. This is relatively common for analysis done prior to 2013, but should no longer occur. Review any recent entries from this query and work with the analyzers to fix.
 9. Mismatch between Frame Count and ‘Number of Frames’ – Calculates the actual number of entered frames, compares that to the ‘Number of Frames’ field, and displays any mismatches. Errors could either be in what was entered or the entry of the ‘Number of Frames.’
 10. Bleaching Codes Used – Displays a summary table of bleaching codes used by year.
 11. Taxon Codes Used – Displays a summary table of TaxonCode entries used by year.
 12. Black Corals – Checks for video point with a TaxonCode = ‘ANTI,’ which are protected records that must be handled according to protected data memos.
- Species Lists
 1. Species Presence through Time (Dynamic) – A dynamic graph (choose a site) shows the percentage of transects that held certain coral species over time.
 2. Black Corals – Checks for species lists with a TaxonCode = “ANTI,” which are protected records that must be handled according to protected data memos.
 3. Potential Missing Species on Transects – This check compares the in situ coral species lists to the corals identified through the video analysis and flags any differences at the transect level.

4. Potential Missing Species at Sites – This check compares the in situ coral species lists to the corals identified through the video analysis and flags any differences at the site level.
 5. Transects with AGRA – Checks for transects that have AGRA (*Agaricia grahamae*), a coral species that has sometimes been identified incorrectly.
 6. Unverified Species Lists – Checks for species lists that have not been marked as “Verified” in the database.
- Disease
 1. Double Disease Free Transect – Checks for transects with more than one “Transect Disease Free” entry.
 2. Not Disease Free – Checks for transects with both the “Transect Disease Free” entry as well as disease lesions. Both can’t be true, so one must be in error.
 3. Empty Disease Entry – Checks for transects without disease lesions that are missing a “Transect Disease Free” entry.
 4. Missing Discoloration Disease Values – Checks for disease entries where either ‘Number of Blemishes’ or ‘% Discolored Tissue’ is zero.
 5. Potential Tissue Loss Lesion Issues – Checks for disease lesions with a length or width greater than 1 meter (3.3 feet [ft]).
 6. Potential Discoloration Lesion Issues – Checks for disease lesions where ‘% Discolored Tissue’ is greater than or equal to 90%.
 7. Both Tissue Loss and Discoloration Metrics – Checks for disease lesions that have both tissue loss (length & width) and discoloration metrics (% discolored & number of blemishes).
 8. Unlinked Disease Photos – Checks for missing disease photos on Z drive.
 9. Missing Tissue Loss Values – Checks for disease lesions that have a length or width of zero.
 10. Unknown Disease Type – Checks for disease lesions that use the UNK disease code.
 11. DS Disease Type – Checks for disease lesions that use the DS code. This older code has been replaced by more detailed combinations (e.g., DS2 or DS124).
 12. WBLO Disease Type – Checks for disease lesions that use the WBLO disease code, which is currently restricted to Dry Tortugas National Park.

13. YB with Wrong Metrics – Checks for YB disease lesions that use the ‘% Discolored’ and ‘# of Blemishes’ instead of length x width.
 14. Unverified Disease Lesions – Checks for disease entries that have not been marked as verified in the database.
 15. Distance from Transect (Dynamic) – This dynamic density plot (choose a year) shows patterns in the distance from transect and may provide insight into search effort.
 16. Disease Types by Park (Dynamic) – This dynamic bar chart (choose a park) shows patterns in the disease types recorded over time.
- CDR
 1. Colony Count Comparison (Dynamic) – This dynamic line graph (choose a site and transect) shows a comparison of colony count 1, colony count 2, and their average.
 2. Potential *Diadema* Outliers – Checks for *Diadema* counts that are greater than or equal to 10.
 3. Potential Colony Count Outliers – Checks for stony coral counts that are greater than or equal to 200.
 4. Stony Coral Count Equal to Zero – Checks for stony coral counts equal to zero. These may represent early, imported data that should be NA instead.
 5. Potential Rugosity Outliers – Checks for rugosity greater than 26 meters.
 6. Highly Variable Colony Counts – On transects where two observers counted coral colonies, this checks for situations where the ratio of high/low stony coral counts is greater than 1.9.
 7. Potential Fire Coral Outliers – Checks for fire coral counts that are outliers.
 8. Fire Coral Counts Equal to Zero – Checks for fire coral counts equal to zero. These may represent early, imported data that should be NA instead.
 9. Unverified CDR Entries – Checks for CDR entries that have not been marked as verified.
 10. Schedule Rugosity Soon – Calculates the length of time since the last rugosity measurement and shows those greater than 3.5 years to assist in planning for upcoming trips. Some retired sites are excluded.

Literature Cited

- National Park Service. 2016. Certification guidelines for Inventory and Monitoring data products. National Park Service, Inventory and Monitoring Division. Fort Collins, Colo.
<https://irma.nps.gov/DataStore/Reference/Profile/2227397> (Last accessed April 2025).
- South Florida/Caribbean Network (SFCN). 2025a. Conducting analysis of HD images—Version 2.1. South Florida/Caribbean Network Standard Operating Procedure NPS/SFCN/SOP—CR12. South Florida/Caribbean Network, Palmetto Bay, Florida. Available at:
<https://irma.nps.gov/DataStore/Reference/Profile/2309551> (last accessed April 2025).
- South Florida/Caribbean Network (SFCN). 2025b. iPad Preparation, Data Entry, and Download—Version 1.0. South Florida/Caribbean Network Standard Operating Procedure NPS/SFCN/SOP—CR24. South Florida/Caribbean Network, Palmetto Bay, Florida. Available at:
<https://irma.nps.gov/DataStore/Reference/Profile/2309559> (last accessed April 2025).
- South Florida/Caribbean Network (SFCN). 2025c. Data summaries, analysis, and reporting—Version 3.0. South Florida/Caribbean Network Standard Operating Procedure NPS/SFCN/SOP—CR21. South Florida/Caribbean Network, Palmetto Bay, Florida. Available at:
<https://irma.nps.gov/DataStore/Reference/Profile/2309556> (last accessed April 2025).
- South Florida/Caribbean Network (SFCN). 2025d. Coral reef video transect database user's guide—Version 3.0. South Florida/Caribbean Network Standard Operating Procedure NPS/SFCN/SOP—CR13. South Florida/Caribbean Network, Palmetto Bay, Florida. Available at:
<https://irma.nps.gov/DataStore/Reference/Profile/2309552> (last accessed April 2025).