

SETr Workflow Guide

Is marsh surface tracking sea level change?

Developing tools and visualizations for NERRS Sentinel Site Data

In this guide:

1. Background on the SETr project and goals [3](#)
2. Where to find important files in your Reserve's SETr folder [5](#)
3. Workflow instructions [7](#)
4. Details on data & metadata files and QA/QC codes [15](#)
5. Contents of the entire SETr folder [23](#)
6. Appendix I: Troubleshooting Tips [25](#)
7. Appendix II: Suggested best practices for data collection [27](#)
8. Appendix III: Statistical considerations [31](#)

This work was sponsored by the National Estuarine Research Reserve System Science Collaborative, which supports collaborative research that addresses coastal management problems important to the reserves. The Science Collaborative is funded by the National Oceanic and Atmospheric Administration and managed by the University of Michigan Water Center (NAI4NOS4190145).

Project Partners:



This page intentionally left empty for clean double-sided printing

1. Background on SETr project and goals

Can tidal wetland surfaces keep up with sea level rise? This is a critical coastal management question, around which many monitoring programs have been developed. Surface Elevation Tables (SETs) are widely used to collect data on the dynamics of tidal wetland communities, through precise measurements of marsh surface over time. Most sites within the network of 29 National Estuarine Research Reserves (NERRs) have been measuring SETs for several years, resulting in a need for tools to process, analyze, and communicate about surface elevation change. This project focused on developing such tools, within a larger workflow, using R, RStudio, and associated packages: a QA/QC app to interact with graphs and tables of raw data and easily identify points that need inspection; click-of-a-button report generation to produce basic analyses of rates of change and comparisons to sea level rise; and clearly annotated visualizations that can be used to communicate the results to both technical and non-technical audiences. We chose these open-source tools so the developed workflow and products would be shareable, reproducible, and useful well into the future, as additional years of data are generated. Input from technical and outreach-oriented working groups has ensured that these tools are appropriate and useful to a variety of end-users. This project has helped standardize and increase the usability of NERR SET data by providing Reserves the tools they need to analyze, understand, and communicate these important, underutilized datasets.

Data and metadata files from 15 Reserves have been reformatted into a heavily-discussed, consistent spreadsheet format. This is a major step forward for system-wide SET data management, but does not replace the need for a standardized, system-wide database. With this consistency, however, it will be substantially easier to move forward when a Sentinel Site database becomes a funded system-wide priority.

Much thought has gone into data formatting, code, and user interfaces for these product outputs; our intent is for this workflow to be understandable and easily repeated, even for people who are uncomfortable with R software. The project was meant to not only perform analyses on past data, but to create a workflow that can be used in the future for data entry, QA/QC, and repeating these analyses on the growing datasets.

This guide is meant to be a thorough introduction to the file directory and intended workflow.

Instructions and explanations have also been provided in pieces of these tools where they make sense to help with flow. For example, each R script that is meant to be run by end-users has instructions at the top of the script. Output documents contain explanations of graphics and analyses. The technical output document also contains file paths to all saved graphics; it serves as a graphical table-of-contents so figures can be re-used as appropriate.

Major outputs are produced in Microsoft Word, .docx format. This allowed the SETr team to provide basic framework and wording through the R code, but the documents are easily customized if end-users want to add additional site-specific information or remove redundant graphics or tables.

- **SET_Analyses** is a technical report on the rate of elevation change at each SET in a reserve. The document includes graphics showing change over time, graphical comparisons to sea level rise, a spatial view of the changes within a reserve, and tables of all relevant numbers and levels of uncertainty. This document is aimed at technical Reserve staff that already understand SETs and their importance.

- **SET_Outreach** is a report aimed at Reserve outreach staff. This document contains more detail SETs, and steps through graphs a bit more slowly to build up the concepts so broad audiences can understand them. The SETr team has provided background and context in this document, but it still needs to be supplemented by other information based on the goals of the specific outreach person using it. We encourage reserve outreach staff to talk to reserve technical staff about the specific information contained in this document, how to interpret it, and what stories you can tell with the data.

Project Team

Project Lead:

Kim Cressman, Grand Bay NERR
kimberly.cressman@dmr.ms.gov

Collaborative Leads:

Suzanne Shull, Padilla Bay NERR
Margo Posten, Grand Bay NERR

Partners:

Kristin Evans, Mission Aransas NERR
Brook Russell, Clemson University
Jenni Schmitt, South Slough NERR
Kari St. Laurent, Delaware NERR
Megan Tyrrell, Waquoit Bay NERR

2. Where to find important files in your Reserve's SETr folder

Enter your latest measurements

data/raw/Gndset.xlsx (substituting your Reserve code for 'GND')

Update your metadata

metadata/Gndset_metadata.xlsx

Update Sea Level Rise information

metadata/slr_rates.csv

Find the list of QA/QC codes

metadata/user_defined_inputs.xlsx, qaqc_codes tab

Specify which QA/QC codes you want to exclude from analyses

metadata/user_defined_inputs.xlsx, qaqc_codes tab

Specify other options for your graphs and analyses

metadata/user_defined_inputs.xlsx, general tab

Find the technical analysis

R_output/analysis/SET_Analyses_YYYY-MM-DD.docx

Find the outreach document

R_output/outreach_doc/SET_Outreach_YYYY-MM-DD.docx

This page intentionally left empty for clean double-sided printing

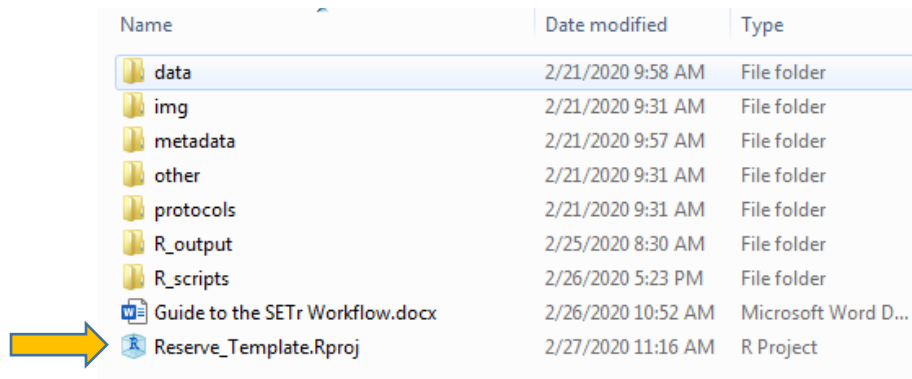
3. Workflow Instructions

Example file names use the reserve code **GND**; in reality the files will use your 3-letter reserve code. Instructions on running scripts are included at the top of each.

I. Set up the project

Necessary once per computer, or if your R/RStudio installations are not current

1. Make sure a current version of R software (3.5.3 or higher) is installed.
 - If you do not have R, you can download it for your operating system at <https://cran.r-project.org/>
2. Make sure a current version of RStudio (1.2 or higher) is installed.
 - Use the free, Open Source License version of RStudio Desktop at this link: <https://rstudio.com/products/rstudio/download/>
3. Unzip the folder you were provided with; details on everything in the directory are in [section 5](#) of this document.
4. In the top level of your Reserve's folder is a file called **Reserve_Template.Rproj** (Figure 1). Open this; it will open a session of RStudio.



Name	Date modified	Type
data	2/21/2020 9:58 AM	File folder
img	2/21/2020 9:31 AM	File folder
metadata	2/21/2020 9:57 AM	File folder
other	2/21/2020 9:31 AM	File folder
protocols	2/21/2020 9:31 AM	File folder
R_output	2/25/2020 8:30 AM	File folder
R_scripts	2/26/2020 5:23 PM	File folder
Guide to the SETr Workflow.docx	2/26/2020 10:52 AM	Microsoft Word D...
Reserve_Template.Rproj	2/27/2020 11:16 AM	R Project

Figure 1: RStudio Project file in SETr Reserve folder.

5. In the lower right pane of RStudio, click on the "Files" tab. Navigate to the "R_scripts" folder (Figure 2b).
6. Click on the file **00a_install_packages.R** (Figure 2c). This will open in the top left pane of RStudio (Figure 2a). Follow the instructions at the top of the script to run it; this will install all necessary R packages for this workflow. Note that this could take several minutes to half an hour or more, depending on your computer's current configuration.
7. Next, click on the file **00b_check_packages.R** (Figure 2c). Follow the instructions at the top of the script to run it. If something is wrong, the script will give you suggestions on how to fix it. Then re-run the script to see if it worked. When the output in the console (bottom left pane) says "All required packages are installed and loading properly!", you're ready-to-go.

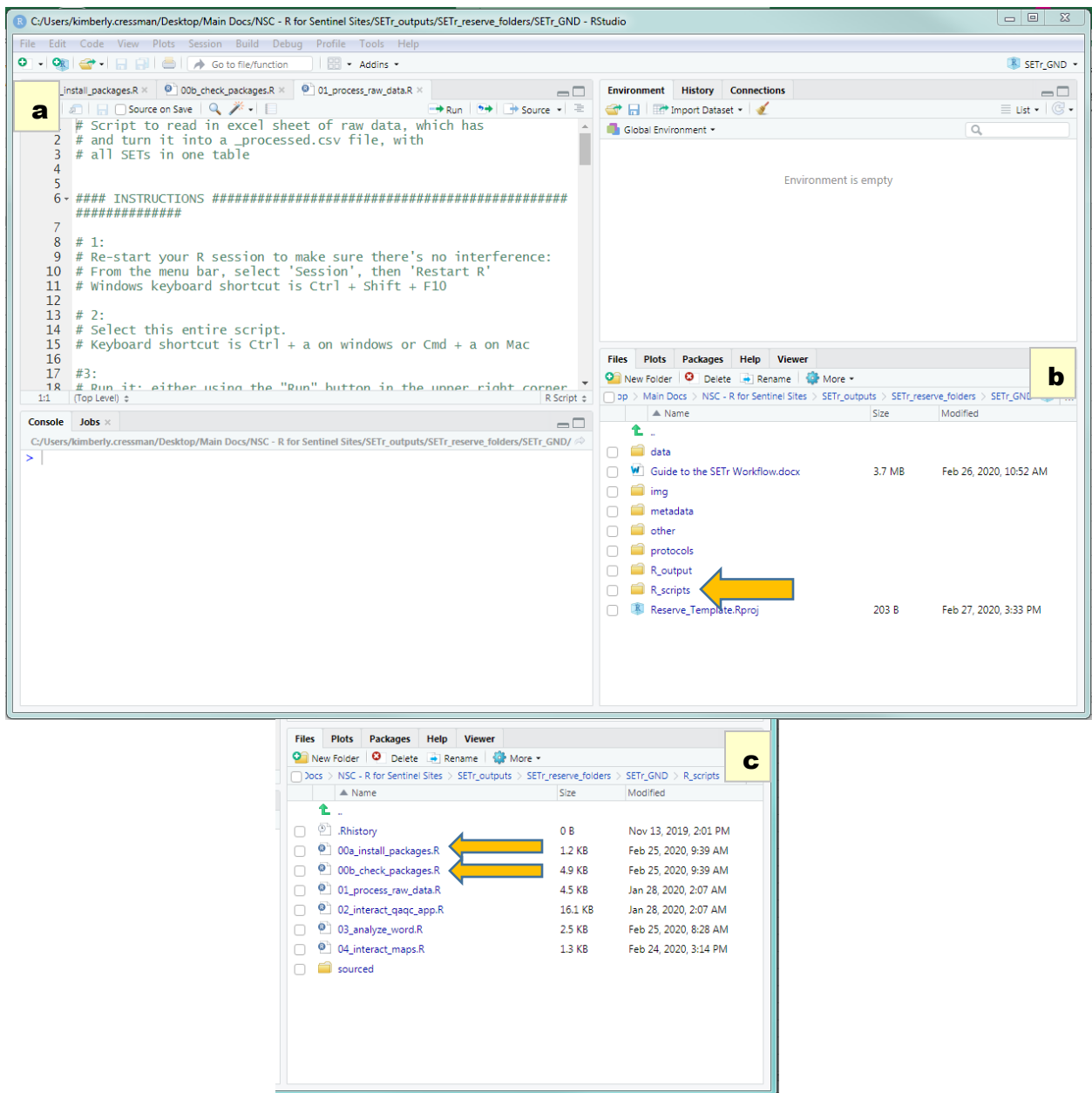


Figure 2: Active RStudio session. Upper left pane (a) is the source pane, where open scripts appear; lower right pane (b) is file explorer window. Clicking on “R_scripts” (arrow in b), will pull up a file list as in (c). The package installation and checking scripts are marked with arrows in c.

II. Enter new SET data

Yearly (or the same frequency at which you take measurements)

- Enter data into the raw data spreadsheet, **gndset.xlsx**, found in the directory **data/raw**.
 - From this point forward, file names will be listed after the directory, e.g. **data/raw/gndset.xlsx**
 - Note that the file you have is named with your reserve code, and you won't have to change it. 'gnd' is used throughout this document as an example.

- The file name **must** end with “set.xls” or “set.xlsx”.
 - ONLY KEEP ONE raw file per complete SETr directory. The scripts automatically choose a data file to work on based on what is contained in the folders; if there is more than one matching “...set.xls.”, they will choose the first alphabetically. If you have multiple reserve components and want the data in separate spreadsheets, you must copy the entire SETr folder and keep one for each component.
 - Don’t forget to close this file when you’re finished updating it
2. A detailed breakdown of this spreadsheet is in [section 4](#) of this document.

III. Run a processing script

EVERY TIME you update the raw data – essential step for scripts to run properly

1. Open the RStudio Project file: **Reserve_Template.Rproj**
2. Open and run: **R_scripts/01_process_raw_data.R**
 - a. script purpose:
 - i. pull in your raw data and reformat it into a computer-friendly version
 - ii. the file produced by this script will be used in all further analyses
 - b. files used by script:
 - i. data/raw/gndset.xlsx
 - ii. R_scripts/sourced/000_functions.R
 - c. files produced by script:
 - i. data/processed/gndset_processed.csv

IV. Run interactive QA/QC app

Identify potential errors in data recording/entry and recognize influential events that could impact data, so corrections can be made in the raw data file and documentation added as necessary.

1. Open the RStudio Project file: **Reserve_Template.Rproj**
2. Open and run: **R_scripts/02_interact_qaqc_app.R**
 - a. script purpose:
 - i. allow you to interact with data in various ways (Figure 3):
 1. Raw Data tab (Figure 3a)
 - a. This tab contains a graphic for pin readings averaged to the arm level for each date, followed by a graphic with one panel per arm with each panel containing all 9 pin readings for each date.
 - b. Use it to look for obvious outliers or pins that might have been read in the wrong arm position.
 2. Incremental Calcs tab (Figure 3b)
 - a. The changes calculated here are the [reading for a date] – [previous reading], for each pin. In the Lynch et al. 2015 NPS protocols, they recommend examining any incremental changes that fall outside +/- 25mm of the previous reading.
 - b. The graphic contains one panel for each arm position, with 9 pins in each panel.
 - c. The red lines start at +/- 25mm, and this can be changed with the slider bar at the top.
 - d. A check-box allows you to make a table of points exceeding the threshold chosen above.

- e. Again, look for outliers. They can lead to finding either errors in data entry or interesting events that influenced the data!
3. Cumulative Calcs tab (Figure 3c)
 - a. Every pin reading had the reading on the first measurement date subtracted from it, to get change-since-baseline for each pin on each date. Then pins on an arm were averaged together on each date, and the arm positions were averaged together to yield overall average change-since-baseline for the SET.
 - b. A check box allows you to “overlay a linear regression” – use this to check for general linearity of your change. If you see curvature or obvious changepoints, be very cautious about rate calculations generated through linear models.
- ii. in any graphic, hover over a point to get more information (date, arm position, and pin number of the reading in question)
- iii. use the drop-down menu and date boxes above the tab positions to choose a SET to examine, and limit the date range if desired. The default date range uses the first and last dates in the dataset.
- iv. change the graphics as you like with options on the left sidebar – point size, axis limits, etc.
- b. files used by script:
 - i. data/processed/gndset_processed.csv
 - ii. R_scripts/sourced/000_functions.R
- c. files produced by script:
 - i. none

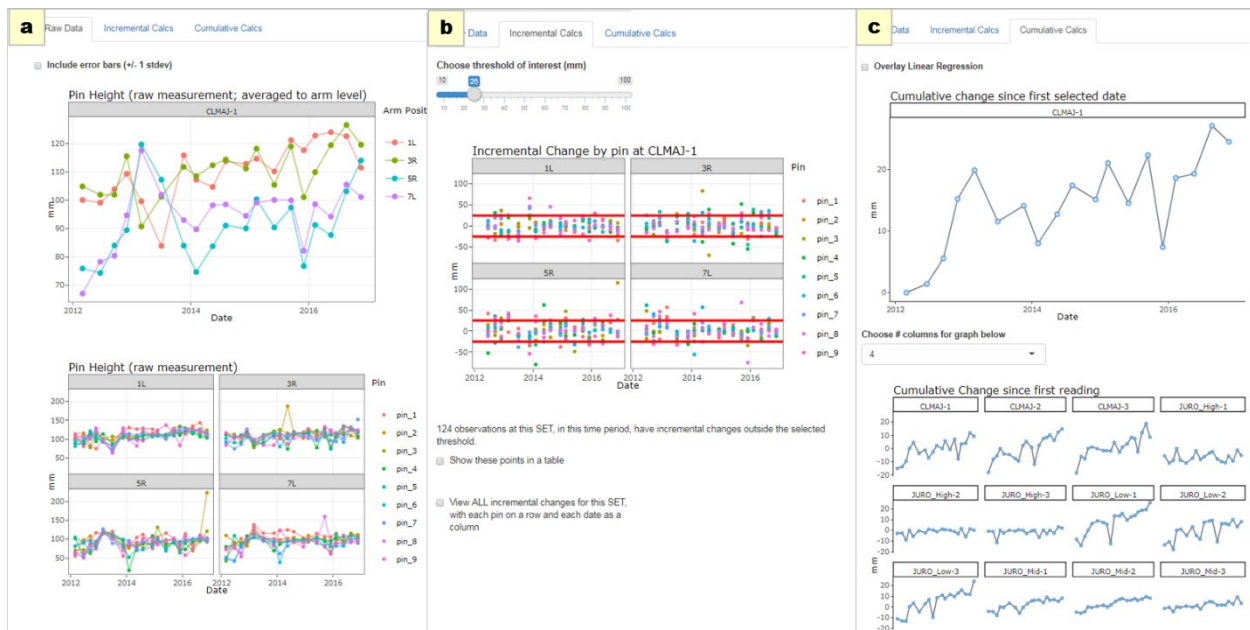


Figure 3: Interactive QA/QC app tabs. a) raw data; b) incremental calculations; c) cumulative change

V. Fix/flag any problems

Make any changes in the raw data spreadsheet

1. **data/raw/gndset.xlsx**
2. Refer to **metadata/user_defined_inputs.xlsx**, qaqc_codes tab to add codes to your data file

VI. Update metadata

Optional; if sites have been added or changed or if rate of long-term SLR should be updated.

1. **metadata/gndset_metadata.xlsx**
2. A detailed breakdown of the metadata template is in [section 4](#) of this document
3. Ideally, in the future there will also be a narrative form of metadata that can include longer comments and explanations

VII. Re-run processing script

Required if you have modified your raw data file

1. Open the RStudio Project file: **Reserve_Template.Rproj**
2. Open and run: **R_scripts/01_process_raw_data.R**
 - a. script purpose:
 - i. pull in your **updated** raw data and reformat it into a computer-friendly version
 - ii. the file produced by this script will be used in all further analyses
 - b. files used by script:
 - i. data/raw/gndset.xlsx (**updated** in step 5)
 - ii. R_scripts/sourced/000_functions.R
 - c. files produced by script:
 - i. data/processed/gndset_processed.csv (**updated** version; will overwrite previous version)

VIII. Specify output options

You can come back to this step after running scripts and seeing outputs, if you want to change any options

1. **metadata/user_defined_inputs.xlsx**
2. This file includes aesthetic options for graphs (see [section 4](#)), such as how many columns you want in multi-panel graphs; dimensions you'd like for saved graphics; and which QA/QC codes (see section 4: [narrative](#) or [table](#)) you want to use to indicate data that should be ignored in these analyses.

IX. Run analysis scripts

1. Open the RStudio Project file: **Reserve_Template.Rproj**
2. Open and run: **R_scripts/03_analyze_word.R**
 - a. script purpose:
 - i. generate two reports for your reserve, one for technical audiences (saved in the **R_output/analysis** folder) and one for NERR outreach staff to translate to their audiences (saved in the **R_output/outreach_doc** folder). Each will have versions of:
 1. SET elevation change rates at your reserve
 2. comparisons to 0 and to sea level rise
 3. tables, graphs, and maps illustrating these results

- ii. generate a summary file of rates, comparisons, and SET locations that can be used later in the interactive mapping script
 - b. files used by script:
 - i. data/processed/gndset_processed.csv
 - ii. metadata/gndset_metadata.csv
 - iii. metadata/slr_rates.csv
 - iv. metadata/user_defined_inputs.xlsx
 - v. R_scripts/sourced/000_functions.R
 - vi. R_scripts/sourced/001_user_options.R
 - vii. R_scripts/sourced/002_rate_calculations.Rmd
 - viii. R_scripts/sourced/003_outreach_graphics.Rmd
 - ix. R_scripts/sourced/004_map_making_static.Rmd
 - x. img/[all files in directory].png
 - c. files produced by script:
 - i. data/intermediate/rate_summary.csv
 - ii. R_output/analysis/SET_Analyses_yyyy-mm-dd.docx
 - iii. R_output/log_files/yyyy-mm-dd_002ratecalculations_logfile.txt
 - iv. R_output/log_files/yyyy-mm-dd_003outreach_logfile.txt
 - v. R_output/log_files/yyyy-mm-dd_03analyzewood_logfile.txt
 - vi. R_output/log_files/yyyy-mm-dd_004mapsstatic_logfile.txt
 - vii. R_output/outreach_doc/SET_Outreach_yyyy-mm-dd.docx
 - viii. R_output/figures/cumu_change_plots/[folder with change-over-time graphs]
 - ix. R_output/figures/maps/[folder with maps]
 - x. R_output/figures/summary_plots/[folder with dot-and-whisker plots]
3. Open and run: ***R_scripts/04_interact_maps.R***
- a. script purpose:
 - i. Make maps that you can interact with to see your results spatially
 - ii. Options in the maps:
 - 1. Background; can be toggled between:
 - a. ESRI World Gray Canvas (default)
 - b. ESRI World Top Map
 - c. ESRI default
 - 2. What to compare the SET elevation change rates to; can be toggled but you need to make sure only one is selected at a time:
 - a. 0 (is elevation changing?)
 - b. long-term sea level rise
 - c. 19-year water level change
 - 3. Zooming and scrolling – to allow more detailed examination of the sites.
 - iii. The maps can be saved one of two ways:
 - 1. Clicking “Export” in the Viewer Pane (next to “Zoom”) (this may require re-sizing the Viewer Pane to get the proportions right)
 - 2. A normal screen shot
 - b. files used by script:
 - i. data/intermediate/rate_summary.csv
 - ii. R_scripts/sourced/005_map_making.R
 - iii. img/[all files in directory].png

- c. files produced by script:
 - i. none unless you save things interactively from RStudio's Viewer pane

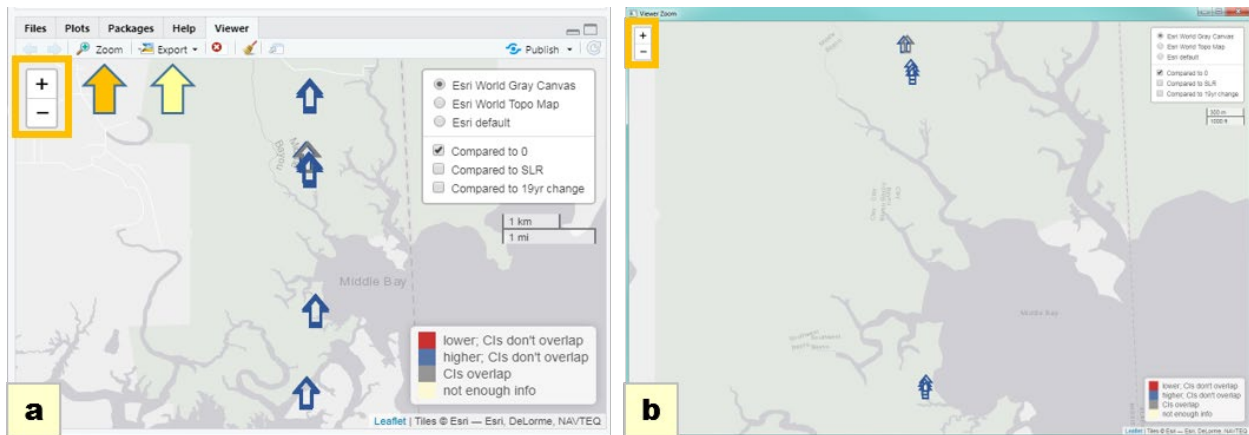


Figure 4: Interactive maps. a) in the viewer pane, bottom right panel of RStudio. Click on “Zoom” (orange arrow) to pop it out into its own window, or “Export” (yellow arrow) to save. b) the pop-up window produced by clicking “zoom”. In both map versions, you can zoom in and out with the + and – signs in the upper left corner (orange box), and toggle other options in the upper right corner.

X. Submit processed data and metadata to the CDMO for flat file hosting

Not yet possible, but recommended annually (optional)

once approved by all relevant workgroups (SET working groups, DMC) and CDMO

1. **data/processed/gndset_processed.csv**
2. **metadata/gndset_metadata.xlsx**
3. narrative metadata, if developed

This page intentionally left empty for clean double-sided printing

4. Details on data & metadata files and QA/QC codes

data/raw folder

Guidelines for use of the raw file:

- **DO NOT CHANGE** column names
- the file name **must** end with “set.xls” or “set.xlsx”, with your 3-letter reserve code before “set”.
- you *can* change colors and column widths
- If you enter clock times, make sure the column is text – Excel does weird conversions of times that aren’t associated with dates

How the Excel “raw” files are organized (Figure 5):

- one worksheet per SET
- one row per arm per date = 4 rows per date for most reserves; more if you measure more arms
- columns: set_id, date, arm_position, arm_qaqc_code, pin height and qaqc_code columns (see below), any additional things you've recorded (reserve; reader; weather)
- 18 pin columns: each of the 9 pins gets 2 columns: one for height (mm or cm; incorporated in column name based on what the Reserve was already recording – it is important to be consistent and stay consistent) and one for a qaqc code
- pin qaqc_code columns are to the right of the pin measurement columns, because in theory they'll be used less often; so they're out of the way
- all other columns in the original reserve file have been preserved in this file, on the far right

“Raw” Data Format: data/raw/gndset.xlsx

one row per arm, per date
= 4 rows per SET, per date.
shading keeps dates
visually separate

9 columns for readings
(pins 1-9; can be mm or cm)

columns for QC flags/codes
off to the right to ease data
entry

set_id	year	month	day	reserve	arm_position	arm_qaqc_code	pin_1_height mm	pin_2_height mm	pin_3_height mm	pin_4_height mm	pin_1_qaqc_code	pin_2_qaqc_code	pin_3_qaqc_code	pin_4_qaqc_code
JURO_High-2	2012	2	28	GND	1L		180	180.5	181.5	185				
JURO_High-2	2012	2	28	GND	3R		165	161.5	173	188.5				
JURO_High-2	2012	2	28	GND	5R		126	174	174	172				
JURO_High-2	2012	2	28	GND	7L									
JURO_High-2	2012	6	14	GND	1L		195	192.5	191.5	189				
JURO_High-2	2012	6	14	GND	3R		184.5	180	157	183				
JURO_High-2	2012	6	14	GND	5R		163	162.5	174	186				
JURO_High-2	2012	6	14	GND	7L		149.5	169	167	146				
JURO_High-2	2012	9	13	GND	1L		172.5	188	180	179.5				
JURO_High-2	2012	9	13	GND	3R		181	174.5	172	176.5				
JURO_High-2	2012	9	13	GND	5R		151	152	156	162.5				
JURO_High-2	2012	9	13	GND	7L		180	177.5	157	165.5				
JURO_High-2	2012	11	28	GND	1L		189	192.5	191	190				
JURO_High-2	2012	11	28	GND	3R		194.5	192	194.5	201.5				
JURO_High-2	2012	11	28	GND	5R		163.5	168	173.5	174				
JURO_High-2	2012	11	28	GND	7L		170.5	169	171.5	159.5				
JURO_High-2	2013	2	19	GND	1L		186.5	185.5	184	182				
JURO_High-2	2013	2	19	GND	3R		186	191	191.5	192				
JURO_High-2	2013	2	19	GND	5R		151	158.5	167	165				
JURO_High-2	2013	2	19	GND	7L		159.5	161	176	163				

.xlsx file

Any extra
columns
included in
reserves'
original files(e.g.
Observer) were
preserved

Each SET on its own worksheet in the same file

Figure 5: Annotated example of raw data file

data/processed folder

DO NOT MODIFY THIS FILE

This is only an explanation of what it contains

- columns are: (see Figure 6 below)
 - set_id
 - year, month, day – for measurement date; kept as separate columns for formatting
 - arm_position
 - arm_qa_qc_code
 - pin_number
 - height_mm or height_cm
 - qaqc_code – pin-level documentation
 - reserve
 - any other columns included in original data file
- generated by script R_scripts/01_process_raw_data.R; used by all subsequent scripts and analyses

“Processed” Data: data/processed/gndset_processed.csv

	A	B	C	D	E	F	G	H	I	J
1	set_id	year	month	day	arm_position	arm_qa_qc_code	pin_number	height_mm	qaqc_code	reserve
2834	JURO_High-2	2012	9	13	5R		pin_7	166		GND
2835	JURO_High-2	2012	9	13	5R		pin_8	160.5		GND
2836	JURO_High-2	2012	9	13	5R		pin_9	165.5		GND
2837	JURO_High-2	2012	9	13	7L		pin_1	180	LHE CB	GND
2838	JURO_High-2	2012	9	13	7L		pin_2	177.5		GND
2839	JURO_High-2	2012	9	13	7L		pin_3	157		GND
2840	JURO_High-2	2012	9	13	7L		pin_4			
2841	JURO_High-2	2012	9	13	7L		pin_5			
2842	JURO_High-2	2012	9	13	7L		pin_6			
2843	JURO_High-2	2012	9	13	7L		pin_7			
2844	JURO_High-2	2012	9	13	7L		pin_8			
2845	JURO_High-2	2012	9	13	7L		pin_9			
2846	JURO_High-2	2012	11	28	1L		pin_1			
2847	JURO_High-2	2012	11	28	1L		pin_2			
.... all SETs, all readings in one file										
10252	SPALT-3	2016	11	23	5R		pin_9			
10253	SPALT-3	2016	11	23	7L		pin_1			
10254	SPALT-3	2016	11	23	7L		pin_2			
10255	SPALT-3	2016	11	23	7L		pin_3			
10256	SPALT-3	2016	11	23	7L		pin_4			
10257	SPALT-3	2016	11	23	7L		pin_5			
10258	SPALT-3	2016	11	23	7L		pin_6			
10259	SPALT-3	2016	11	23	7L		pin_7			
10260	SPALT-3	2016	11	23	7L		pin_8			
10261	SPALT-3	2016	11	23	7L		pin_9			
10262										

Annotations:

- .csv file - should not be modified by hand
- “long” format: one row per pin = 36 rows per SET per date
- this file is generated through R_scripts/01_process_raw_data.R and used in all other scripts and analyses
- This format, even for non-reserve data, works with the scripts from this project

Figure 6: Annotated example of processed data file

metadata folder

Main metadata file, gndset_metadata.xlsx:

This file was built off of a template under development by the NERRS SET working group. Most columns are in the same order even if they were not used in this project. The following columns are those that have been used in SETr code, along with a parenthetical notation about whether it was in the original SET working group template (*existing*), modified from that template (*modified*), or completely new (*created*) for this project.

- Unique SET ID (*existing*)
- User-friendly SET name (*created*)
- Numerical Order
- Reserve
- CoDominant Species1 (*modified*): originally “Dominant Vegetation”; end-users on SETr requested multiple columns as some sites have co-dominant vegetation.
- CoDominant Species2, 3, 4 (*created*): to address the above.
- Latitude (Dec. Deg.) (*existing*)
- Longitude (Dec. Deg.) (*existing*): use negative numbers for “west” longitude
- Set Type (*existing*): Generally deep rod, shallow, or original
- Surface Elevation Ground (NAVD88) (*existing*)
- Surface Elevation Receiver (NAVD88) (*existing*)
- Notes (*created*)

Information on Sea Level Rise rates, slr_rates.csv:

This file originated as a google doc in which we compiled “local” sea level rise information for participating reserves. Each reserve was asked to provide information from the nearest NWLON station, including the NOAA COOPS-generated calculation and 95% confidence interval for sea level rise. Some of these NWLON stations are more than 30 km away from the reserves, which is why “local” is in quotes. Information pulled from an NWLON’s station webpage is highlighted in figure 7. Columns in this file are:

- Reserve – three-letter reserve code
- Nearest NWLON Station – station name of NWLON gauge closest to the reserve and its sites
- Distance to NWLON Station – approximate and somewhat subjective, as provided by reserves
- NWLON Station Number – the identifying number for the NWLON gauge
- SLR rate (mm/yr) – sea level rise rate provided on the graph for the NWLON station’s “sltrends” web page
- +/- (95% CI) – confidence interval half-width provided by NOAA COOPS for the rate of sea level change at this NWLON station (e.g., SLR rate +/- this number gives the full 95% confidence interval for the SLR estimate)
- data start – year that the time series used in calculating sea level rise starts at this station; provided from the NWLON station site
- data end – year that the time series used in calculating sea level rise ends at this station; provided from the NWLON station site
- set start – year that SET data collection started at a reserve in the file submitted for SETr
- slr_19-yr – 19-year water level change, calculated as part of the SETr project, using the same deseasonalized monthly average sea level data and same method (ARIMA 1,0,0) used in the long-term sea level rate calculations. This interval was chosen at a mid-point project workshop

to provide a shorter-term estimate of sea level change than the entire period of record, and yet capture an entire period of the metonic cycle. We verified that SETr calculations of long-term rates and confidence intervals matched those on the website for each NWLON station, then applied the same calculations to a 19-year period and entered these numbers into the *slr_rates.csv* spreadsheet.

- ci95_19yr – confidence interval half-width generated by SETr code for 19-year sea level change rate (e.g., slr_19yr +/- this number gives the full 95% confidence interval for the 19-year sea level change rate estimate)
- yr19_start – the start of the period for 19-yr calculations; based on the given end-date for the long-term SLR estimate, minus 19 years.
- yr19_end – the same end date as provided by the reserve for long-term SLR rate
- link – link to NWLON station page, provided by reserve

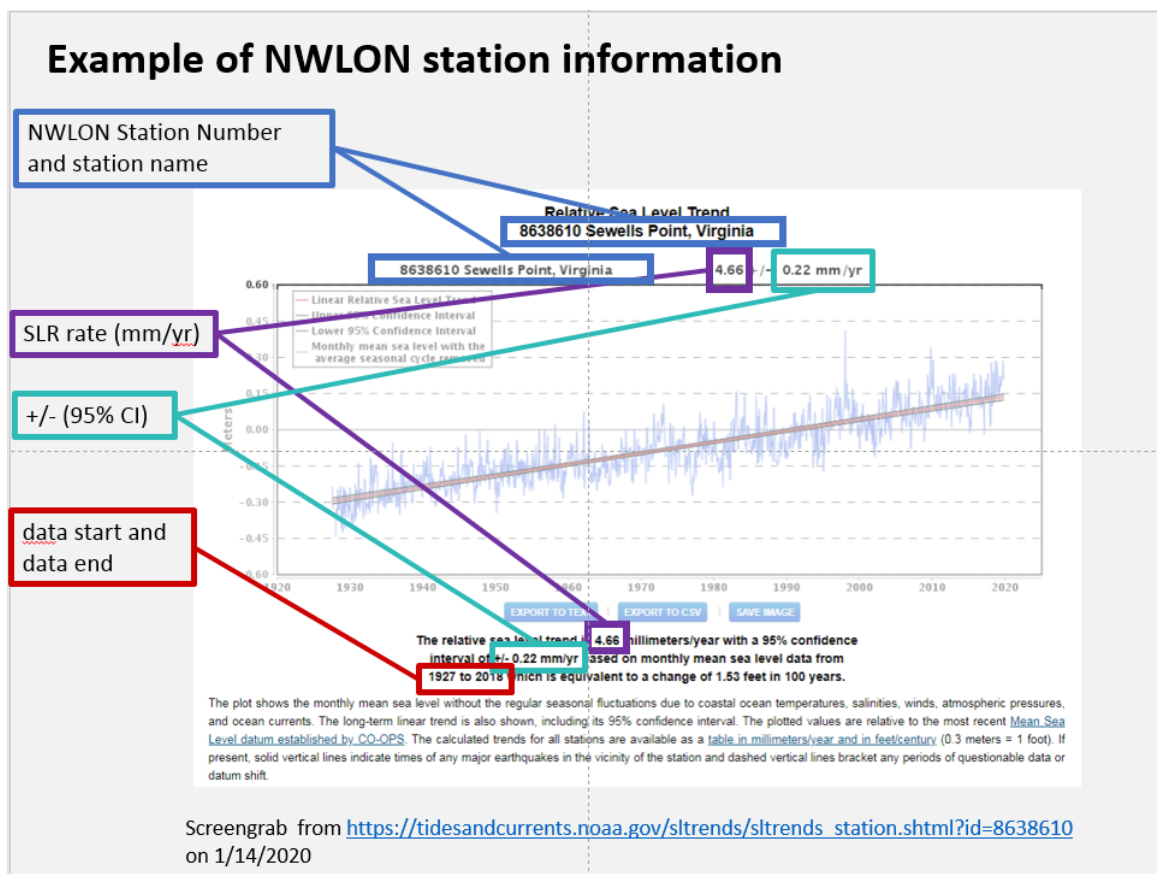


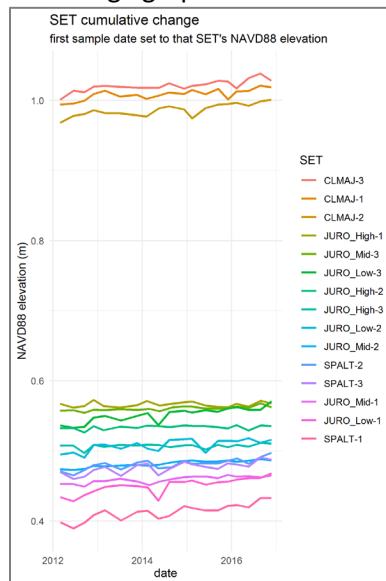
Figure 7: Example of NWLON station website and information from it that was provided by reserves in the *slr_rates.csv* spreadsheet. Most pieces of information can be found in multiple places on this page.

user_defined_inputs file:

- qaqc_codes tab
 - includes the list of codes in Table 1
 - to exclude data points associated with certain codes (full codes and combinations in column J) from analysis, type -3 (no quotes) into column A
- general tab
 - The only changes you should make in this sheet are in the Value column. DO NOT CHANGE anything in the R_param column. DO NOT CHANGE column names.
 - Which graphs are controlled by which dimensions?

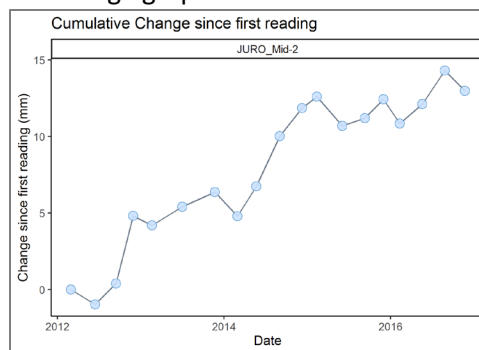
▪ **“tall, skinny graphs” – rows 5 and 6:**

- cumulative change graph with NAVD88 y-axis:

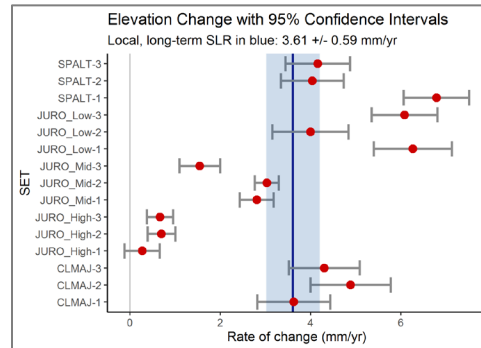


▪ **“wider than tall graphs” – rows 7 and 8**

- cumulative change graphs for individual SETs

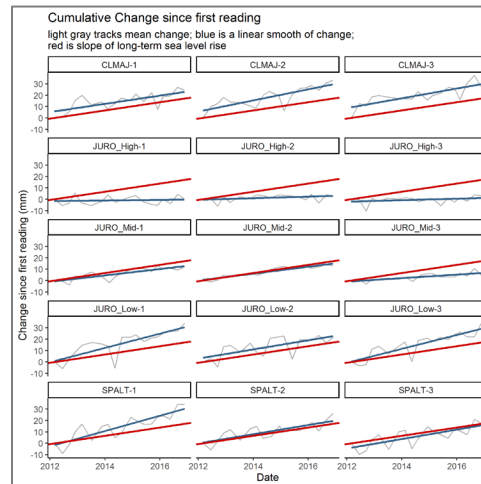


- summary plots



▪ “square graphs” – rows 9 and 10

- cumulative change graphs with multiple panels (one panel per SET)



QA/QC codes (Table 1):

These QA/QC codes were iteratively developed over the course of the SETr project. We see them as a starting point as the NERR system moves forward with standardized data management practices. Through many discussions of things that can go wrong during measurement of SET pins, we first developed a list of common problems. Then we developed 3-letter codes, to be consistent with current QA/QC coding for SWMP water quality/weather/nutrient data. These were broken into three main categories: a pin landed higher than expected (codes start with H), lower than expected (codes start with L), or a general comment (codes start with C). After more discussion of specific situations encompassed by each 3-letter code, we also developed suggested secondary codes to concisely capture more detail about common problems. These do not follow a naming pattern like the primary 3-letter codes do, and should only be used along with a primary, 3-letter code. We did not develop SWMP-like numeric flags as we felt judgements of what should in general be marked suspect or rejected should be reserved for future practitioners. For now, decisions about what to exclude from analysis are left up to reserves, and can be specified in the spreadsheet metadata/user_defined_inputs.xlsx, on the qaqc_codes tab by typing “-3” (no quotes) in column A; -3 is the general SWMP flag for rejected data.

Use of QA/QC codes in the raw data file:

- arm_qaqc_code column exists in case there's some landscape-level issue affecting an entire arm.
- pin_#_qaqc_code columns are off to the right. There is one qaqc_code column for each pin.
- use 3 and 2 letter codes from the qaqc_codes worksheet in those columns.
- you can put multiple codes in a column; separate them by a space (e.g. "HSM LS")
- don't use a 2-letter code without the corresponding 3-letter code (e.g. NOT just "LS" – use "HSM LS")
- codes are roughly divided into: C-- (general comments); H-- (pin landed higher than "expected"); L-- (pin landed lower than "expected")

Table 1. Suggested QA/QC codes for use with SET data.

SET QA/QC Codes

QAQC Codes	1° code	2° code; use only with accompanying primary code
Higher than "expected"	HMD	High - Mound
	CH	Chimneys (e.g. crab chimneys or worm mound/casts)
	AL	Algae
	HPD	High - Plant, Dead
	ST	Stems
	RT	Roots
	WR	Wrack
	AL	Algae
	VV	Vascular vegetation
	VH	Vegetation Hummock
	HPL	High - Plant, Live
	ST	Stems
	RT	Roots
	AL	Algae
	VV	Vascular vegetation
	VH	Vegetation Hummock
	HSM	High - Solid Matter
	LS	Live Shellfish
	SH	Shell (dead shellfish)
	RK	Rock(s)
	SK	Stick(s)
	HDI	High - deposition derived from ice rafting
Lower than "expected"	LDP	Low - depression
	AP	Animal Prints (ephemeral)
	HP	Human Prints (footprints)
	TR	Trough, Shallow

	MB	Mammal Burrow (e.g. muskrats)
	WR	Wrack, removed
	SR	Shell, removed
LHE	Low - hole	
	CB	Crab Burrow
LSC	Low - Scour/trough	
LSM	Low - soft mud	
LPS	Low - pin slipped/dropped (not lowered slowly, may be "too far down")	
General Comments	CRM	Comment - reading missing
	PM	Pin missing/ broken
	SI	Sudden inundation
	CUW	Comment - pin under water
	SW	Standing Water
	SI	Sudden Inundation
	IT	Incoming Tide
	PA	Paper used for resistance
	CUS	Comment - Uncertain Surface (e.g. post-hurricane debris, wrack, pine needles, algae mats)
	COS	Change in Observer - short term (< 1 seasons' worth)
	COL	Change in Observer - long term
	CEP	Change in Equipment/Protocol
	AH	new arm hardware
	AP	new arm position
	NP	new pins
	SM	SET modification, e.g. raised receiver due to sedimentation
	CCF	Comment - Correction Factor Needed
	MT	Measurement technique has changed
	NP	New pins; when combined with CCF, different length than previous
	SM	SET modification, e.g. raised receiver due to sedimentation
	CSI	Comment - Structural Instability
	RU	Rod not stable
	WE	Wind affecting arms during measurements
	CLY	Comment - Measured less than a year post-installation
	CHL	Comment - Hummocky Landscape (whole arm)
	COT	Comment - Other

5. Contents of the entire SETr folder

details on each folder and subfolder, sorted alphabetically

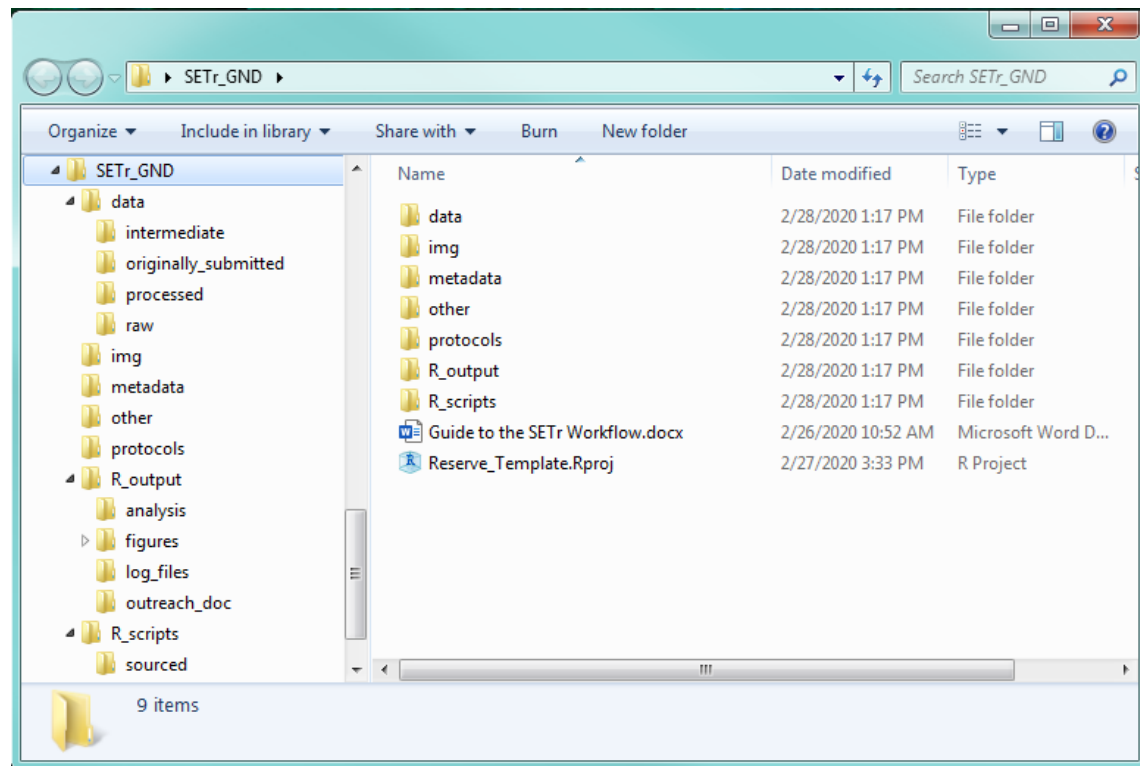


Figure 8: Overview of the entire directory. On left, the directory “tree”, showing the folders within each folder. On right, the main directory is open, showing all files and folders in the “top level”.

1. data

- a. intermediate: *do not modify this folder on your own*. It contains automatically generated files from the analysis. Currently the only file is:
 - i. rate_summary.csv - a table that contains a single row for each SET, and columns including the rate of change in mm/yr and the 95% CI bounds
- b. processed: *do not modify this folder on your own*. It contains an automatically generated file containing your SET data in a long format: one row per pin. The file contains all SETs, all dates, and all parameters you included in your raw file. This file is generated by the R script ‘01_process_data.R’ and is used in all subsequent scripts. The name format of the file is, with your reserve abbreviation in place of ‘GND’:
 - i. gndset_processed.csv
- c. raw: **DO modify this file**. This is the raw data file. It is an Excel file, with one worksheet per SET. Within a worksheet, there is one row per arm position per date, with separate columns for each pin measurement on that arm, columns for qa/qc codes, and columns for anything else your reserve has been recording. The name format is:
 - i. gndset.xlsx

2. **img** *do not modify this folder on your own*. These files must be present for the scripts to work properly.

- a. arrows.pptx – has all of the map icons that generated the pngs below, along with their hex and RGB colors. This is the only file that can be modified or removed without a problem.
 - b. blue_up_arrow.png
 - c. gray_down_arrow.png
 - d. gray_up_arrow.png
 - e. red_down_arrow.png
 - f. SET_diagr2_NPSprotocols.png
 - g. SET_measurement_creditHudsonRiverNERR.jpg
 - h. yel_not_enough_info.png
3. **metadata**
 - a. gndset_metadata.xlsx **DO modify this file** as needed. Do NOT change existing column names, but feel free to add new columns.
 - b. slr_rates.csv **DO modify this file** as needed. Do NOT change existing column names, and make sure there is only a single row for your reserve. This spreadsheet is where the scripts find your local rate of sea level rise for comparisons to SET elevation changes.
 - c. user_defined_inputs.xlsx **DO modify this file!** There are options for customizing your graphs and plots, as well as excluding data associated with certain QA/QC codes.
4. **other** **DO modify this folder.** There is nothing related to these R scripts; it is simply for you to include any other relevant files you want in the same directory.
5. **protocols** **DO modify this folder.** This is where you can include your Sentinel Site Plan and other reserve-level documentation.
6. **R_output** *do not modify this folder on your own* unless it's to change the name of an output file. Anything in here is automatically generated from the scripts.
 - a. analysis This is where the technical Word report is saved.
 - b. figures This is where plots and graphs will be saved.
 - c. log_files The main scripts generate .txt files with the output of R's sessionInfo() function. These will only be used for troubleshooting and hopefully will never be needed.
 - d. outreach_doc This is where the outreach-oriented Word report is saved.
7. **R_scripts** *don't mess with these unless you really know what you're doing*

Appendix I: Troubleshooting tips

General best practices:

- Make sure you have an up-to-date version of **R**; **3.5+ is required**. You can check this by opening RStudio; the version of R is the top line in the startup message of the Console (lower left pane).
- Make sure you have an up-to-date version of **RStudio**; **1.2+ is required**. In RStudio's toolbar, choose 'Help' – 'About RStudio', and the version is at the top of the pop-up window.
- Run ***R_scripts/00b_check_packages.R*** to make sure all package versions are up-to-date. If they are not, the script output should help you fix this.
- Make sure to restart R (*Session* from RStudio's top menu bar → *Restart R*) before running each script, especially if you have been running others already
- Make sure your SETr project folder is on your C: drive, not a network drive.
- Make sure you are running the scripts inside the RStudio Project. You can check this by looking in the upper right corner of RStudio; there is a little blue box with an R in it. Next to it, you should see something like "SETr_GND" (and not "Project: (None)").

Updating outputs each year – if you enter new data into the raw .CSV, you need to run the scripts from the start (i.e., start with 01_process_raw) to regenerate the updated products.

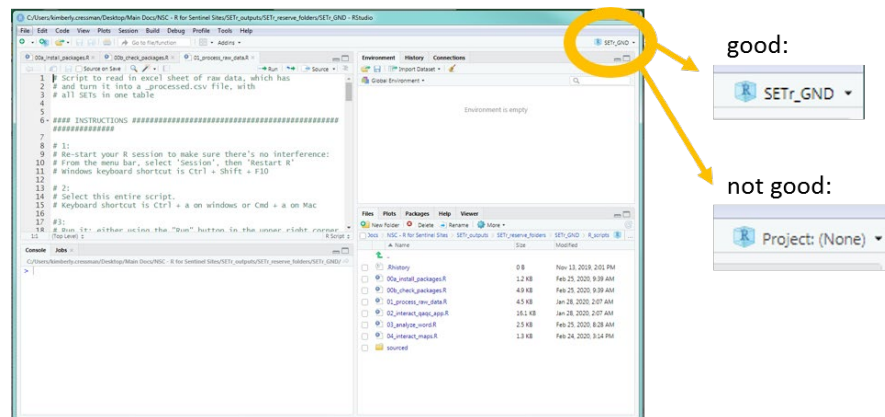
Multiple raw data files – you may have multiple files from multiple years or sites/components. If you have more than one .csv file in the raw data folder, be aware it will take the first one listed alphabetically. Each reserve component should have its own SETr directory.

Warnings and Errors:

Warnings are usually ok and can be ignored, but Errors mean that the script did not run correctly and something needs to be fixed

Errors while running a script – if something goes wrong before outputs are created, try these steps:

- Make sure you have the .Rproj open by checking the upper right corner of RStudio (see below). If you see "Project: (None)", see the [Figure 1 in the workflow instructions](#) for how to open up the proper instance of RStudio.



- Make sure the folder you're running everything from has been unzipped!
- Make sure your data and metadata spreadsheets are *not* open.

- Run the script line by line - If you're getting errors, run the script line by line to figure out where exactly the initial error is occurring. Usually the first error will trigger all the other errors.
- If you get an Error from a line that reads `source(...)`, open the script in the parentheses and try running that script line by line. This will allow you to locate where the initial error is occurring within that script. (Note: `source(...)` means that R is reading from another script.)
- If it's a problem with a package, try re-installing that package (just installing it will automatically uninstall the old version and reinstall the new version). If R refuses to uninstall it, you can go into your R file folders and delete the folder with that name. But be very careful when deleting files in here!!! If you mess up though, you can always reinstall R.
 - If you get an error about needing "RTools" – this is not a typical R package; you need to go here to download the latest version for your system (the one that is highlighted in green):
- Check the raw data file - remember that spacing and capitalization matter, a lot:
 - Did you modify file names?
 - Did you modify column names?
 - Did you add new columns to the data file?
 - Has it been added to ALL worksheets?
 - Is it named the same way in all worksheets?

Warnings in outputs – Occasionally you will see warnings in your report documents; these are general notifications.

Examples:

- **"warning, you have SET IDs in the metadata that aren't in the data"** or vice versa – the report will non-matching SET IDs and you can make sure there aren't spelling or spacing differences (SET1 is NOT the same as SET 1 or SET-1)
- **"warning: 12 failed to parse"** means that while data files were being read in, R didn't know what kind of column to turn it into. This should not cause problems, but please let Kim know if you have issues.

Problems with maps in Word documents – It is still surprisingly complicated to use the same R package for both interactive maps and static versions of those same maps. For the R package used for maps in this project (*leaflet*), the best way to capture static versions is through a combination of the packages *mapview* and *webshot*, along with a piece of software called **PhantomJS**. **PhantomJS** is *not* an R package. The *webshot* package has a function, `"install_phantomjs()"`, which is meant to easily install **PhantomJS**. This function was included in the SETr script *00a_install_packages.R*. Sometimes it does not work, and you may need to download the software directly from the website, <https://phantomjs.org/download.html>. The help file for `"install_phantomjs()"` says "Make sure the executable can be found via the PATH variable". You may need to ask your IT department for help with this. The good news is, even if you can't make this work, you can still make and view maps interactively with *04_interact_maps.R*, then save and paste those in wherever you need them.

Appendix II: Suggested best practices for SET data collection

Basic Terminology (see Figure 1):

- SET instrument – a portable mechanical leveling device for measuring the relative elevation change of wetland sediments
- Arm – the part of the instrument where the pins are inserted
- Level bubble – a way to ensure the instrument is leveled in x, y, z planes
- Measuring pins – inserted into holes in the arm, the pin sits atop the surface of the wetland. The SET measurement is made from the top of the arm to the top of each pin.
- Receiver – where the SET instrument attaches to the deep rod in only one configuration. The receiver is attached and concrete-sealed to the deep rod. The top of the receiver is typically the vertical control point for the SET site.
- Eight-hole collar – The top of the SET instrument sits on the collar in one of eight orientations (which are numbered/etched every other hole as 1, 3, 5, 7). Arm direction is directly related to orientation of top of SET instrument on the collar.

Consistency of measurements:

- Use a light touch when measuring – particularly as you move out the arm towards pin 9; putting weight on the arm as you take the measurement can affect accuracy of the data.
- Keep pins as clean as possible – challenging, but helps with insertion of pins into arms
- Check bubble level in between every reading to make sure jostling as you measure hasn't thrown it out of level.
- Measure each pin; label pins 1-9 and use each pin with its corresponding hole (pins won't be exactly same length so keep any variability consistent across measurements/years). Record these pins measurements in the metadata.
- As a QA/QC technique, have data recorder repeat the value that they heard and are writing down to insure it is correct.
- Be aware of timing of measurements and environmental variability
 - Seasonality of readings – it is best to keep a consistent time frame to reduce effects of seasonal oscillations having unknown effects on trends (e.g., seasonal differences in plant production).
 - Tides – Similarly, stay as consistent as possible in relation to tides (i.e., inundation). Record time difference from nearest predicted tide, the predicted water level, and the station identification. Try to avoid taking readings in standing water; if unavoidable record there is standing water for every pin. If your protocol calls for using paper under each pin to better visualize when the sediment surface is contacted, record that paper was used with each observation.
 - Storms – note whether there was a recent large storm event or other abnormal inundation event
- Limit observers – if possible, have only permanent staff (not seasonal) read SETs, ideally one person with another as backup. If there is a change in reader, note that as a QAQC code and preferably have a crossover period of double readings, to be able to standardize readings.

- Add observer judgement calls to the Reserve protocol – examples of this include determining what to consider the surface if there is thick wrack or duff layer, deciding to remove a small log from plot or leave there, etc.
- Review [QAQC codes](#) (Table 1 in this workflow guide; tab in *metadata/user_defined_inputs.xlsx*) prior to taking measurements. This will help observer keep in mind potential variabilities in readings that should be recorded.

NAVD88:

- Measure elevation height of SET receiver (see Figure II-1)— collect baseline measurement, and complete annual checks based on NERRS SSAM1 elevation protocols (e.g. 2019. NERRS SWMP Vegetation Monitoring Protocol: Long-term Monitoring of Estuarine Vegetation Communities. National Estuarine Research Reserve System Technical Report. 36pp.);
- Document how you obtained your NAVD88 elevation data (e.g., RTK, laser level, static occupation, statewide local reference network [SRN]). Also document where your vertical control mark is (e.g., typically the SET receiver).
- Record the vertical offset of each SET (annually is ideal) - use a laser level to obtain the vertical difference between the top of the arm and the top of receiver. If arm is double-sided, do this on both sides. Move receiver to each position where you take readings and record offset for each position. Tie it to NAVD88 and document the methods you used to do so.
- Converting from NAVD88 to another datum (e.g., MLLW) – Document how you’re converting between datums (e.g., vDatum or [CO-OPS tidal analysis datum calculator](#)) in your metadata.
- Also note that in 2020 a shift in the Geoid model for calculating ortho heights (i.e., NAVD88) will occur

New SETs:

- Due to settling time after SET installation, some analysts recommend waiting at least one year after installation before taking your initial baseline measurement.
- If measurements are taken less than one year post-installation, document it in the metadata and use the QAQC code designed for this purpose.
- Measure the compass direction of the arm at the four positions for your SET (e.g., positions 1, 3, 5, 7) and document in the metadata (in case different arms in future are slightly offset)

Also:

- Make provisions for adding arm positions – when a certain arm position has to be changed or temporarily abandoned (e.g. muskrat burrows; log in plot, etc), make sure to document this shift in your metadata
- When you update your SET data (annually) make sure to go back and update long-term SLR rate from relevant NWLON station

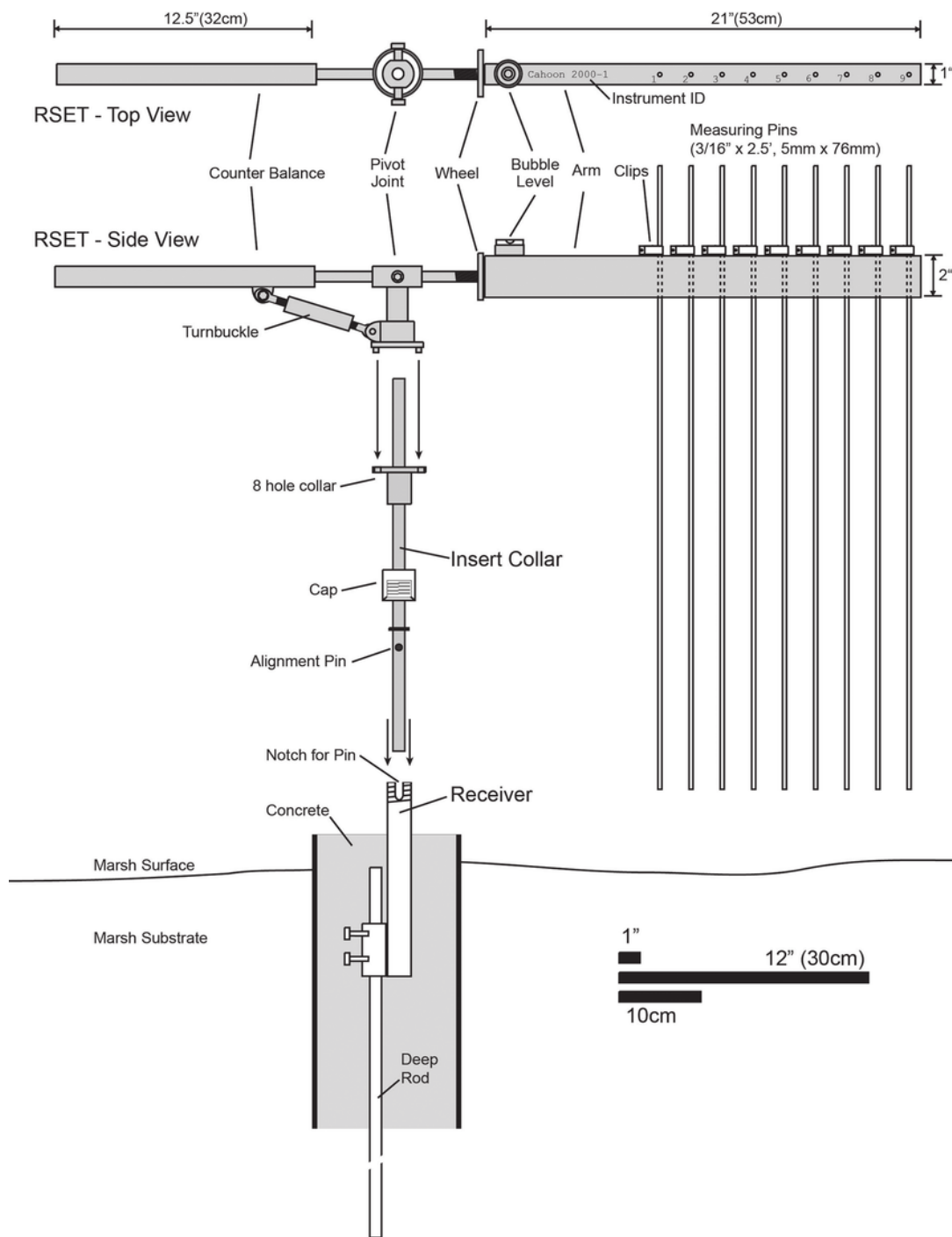


Figure II-1: Diagram of SET components including deep rod, receiver and removable SET instrument. From: Callaway, J.C., Cahoon, D.R. and Lynch, J.C.. 2013. "The surface elevation table-marker horizon method for measuring wetland accretion and elevation dynamics." *Methods in Biogeochemistry of Wetlands*, 10, pp.901-917.

This page intentionally left empty for clean double-sided printing

Appendix III: Statistical Considerations

In this project, rates of elevation change at each SET were generated using random-intercept linear mixed models. See Zuur et al. (2009) and Cahoon et al. (2019) for details. Protocols drawn up for the National Park Service by Lynch et al. (2015) describe several potential analyses for SET data. A complete chapter is dedicated to sampling design, and SOP 8 is dedicated to analysis of SET and marker horizon data. Analysts should consider the reasoning behind SET placement – are there multiple replicates within a marsh zone? are deep SETs paired with shallow SETs? – when deciding on the best methods for an analysis, especially when the questions are at a landscape-scale.

Linear mixed models extend traditional linear regression models by allowing for the inclusion of both fixed and random effects. *These types of models are particularly useful when the data are structured hierarchically*, as with SET data. Data for each SET is analyzed separately using pin height as the response variable; arm and pin (nested in arm) are treated as random effects; and date is considered a numeric covariate. As both the intercept and slope include random effects, methods other than least squares must be employed.

For each SET, we initially considered two LMMs, as in Cahoon et al. (2019): a LMM that includes a random intercept (with a random effects for arm and for pin nested in arm) and a LMM that includes both a random slope and a random intercept (with random effects for arm and for pin nested in arm). For many SETs, we observed that the random intercept model fit better. At other SETs, the random slope and intercept model produced better fitting models (based on AIC). However, the *resulting point estimates showed only small differences between the two approaches*. As the random intercept models did not require the same degree of scrutiny when model fitting, and did not cause as many script-running errors, we exclusively employed random intercept models in these automated R scripts. **For more detailed analyses at a smaller level, we recommend consideration of both models.**

In this project, rates of elevation change at each SET were compared to rates of long-term SLR and 19-year water level change by *investigating whether confidence intervals overlapped*. This method of comparison was chosen because different methods were used to calculate rates for sea level rise (ARIMA) and SET elevation change (LMMs), using data from different sources. We note that each individual interval has 95% confidence associated with it, and conclusions that are made based on pairwise comparison of these intervals will not necessarily be equivalent to conducting a formal hypothesis test for a difference at the 5% level (Schenker and Gentleman, 2001).

General recommendations on analysis of SET data

- Before analysis, the analyst should carefully examine plots of the data to ensure that a linear model is appropriate. *If the points exhibit curvature or if there is some sort of a changepoint, this type of analysis may not be appropriate*. Graphs are provided in this report to help with this determination.
- The analyst should also look for highly influential observations. One way that this could occur is when there is a *large temporal gap* in the data.
- Be aware that the LMMs used in this project do not account for *temporal dependence* in the data. This could mean that confidence intervals are narrower than they should be and may be something the analyst wishes to address in future analyses.

- The analyst should employ a statistically valid model fitting strategy. When using Linear Mixed Models (LMMs), this should also include addressing issues such as:
 - ensuring convergence of the numerical optimization
 - exploring sensitivity to starting values in the optimization procedure
 - determining whether a random slope model or a random slope/random intercept model is most appropriate, and
 - verifying model assumptions.

References

- Cahoon, D.R., J.C. Lynch, C.T. Roman, J.P. Schmit, and D.E. Skidds. 2019. Evaluating the Relationship Among Wetland Vertical Development, Elevation Capital, Sea-Level Rise, and Tidal Marsh Sustainability. *Estuaries and Coasts* 42:1-15. <https://doi.org/10.1007/s12237-018-0448-x>
- Lynch, J.C., P. Hensel, and D.R. Cahoon. 2015. The surface elevation table and marker horizon technique: A protocol for monitoring wetland elevation dynamics. Natural Resource Report NPS/NCBN/NRR—2015/1078. National Park Service, Fort Collins, CO.
- Schenker, N. and J.F. Gentleman. 2001. On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals. *The American Statistician* 55(3):182-186. <https://doi.org/10.1198/000313001317097960>
- Zuur, A.F., E.N. Ieno, N.J. Walker, A.A. Saveliev, and G.M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer.