SETr Workflow Guide

Is marsh surface tracking sea level change?
Developing tools and visualizations for NERRS Sentinel Site Data

In this guide:

This work is 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:

Workflow Instructions

1. **Enter new data:**
    a. into raw data spreadsheet: data/raw/GNDset.xlsx
    b. a detailed breakdown of this spreadsheet follows these workflow instructions

2. **Update your metadata**, if necessary.
    a. metadata/GNDset_metadata.xlsx

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

4. **Run interactive QC app**.
    a. First open the RStudio Project file: Reserve_Template.Rproj
    b. Open and run: R_scripts/02_interact_qaqc_app.R
        i. script purpose:
            1. allow you to interact with data in various ways
        ii. files used by script:
            1. data/processed/GNDset_processed.csv
            2. R_scripts/sourced/000_functions.R
        iii. files produced by script:
            1. none

5. **Fix/flag any problems** *in the raw data spreadsheet*
    a. data/raw/GNDset.xlsx
    b. Refer to metadata/qaqc_codes.xlsx to add QA/QC codes to your data file

6. **Update metadata** if necessary to address any data quality issues
    a. metadata/GNDset_metadata.xlsx
    b. ideally there will also be a narrative form of metadata at some point

7. **Re-run processing script** from step 3, to generate an updated processed file.
    a. First open the RStudio Project file: Reserve_Template.Rproj
    b. Open and run: R_scripts/01_process_raw_data.R
        i. script purpose:
            1. pull in your ***updated*** raw data and reformat it into a computer-friendly version

2. the file produced by this script will be used in all further analyses

ii. files used by script:
1. data/raw/GNDset.xlsx (***updated*** in step 5)
2. R_scripts/sourced/000_functions.R

iii. files produced by script:
1. data/processed/GNDset_processed.csv (***updated*** version; will overwrite previous version)

8. **Specify output options**
   a. metadata/user_defined_inputs.xlsx

9. **Run analysis scripts**
   a. First open the RStudio Project file: Reserve_Template.Rproj
   b. Open and run: R_scripts/03_analyze_word.R
      i. script purpose:
         1. generate a report for your reserve with:
            a. SET elevation change rates at your reserve
            b. comparisons to 0 and to sea level rise
            c. tables, graphs, and maps illustrating these results
         2. generate a summary file of rates, comparisons, and SET locations that can be used later in the interactive mapping script
      ii. files used by script:
         1. data/processed/GNDset_processed.csv
         2. metadata/GNDset_metadata.csv
         3. metadata/slr_rates.csv
         4. metadata/user_defined_inputs.xlsx
         5. R_scripts/sourced/000_functions.R
         6. R_scripts/sourced/005_rate_calculations.Rmd
         7. R_scripts/sourced/006_map_making_static.Rmd
         8. img/[all files in directory].png
      iii. files produced by script:
         1. data/intermediate/rate_summary.csv
         2. R_output/analysis/SET_Analyses_yyyy-mm-dd.docx
   c. Open and run: R_scripts/04_interact_maps.R
      i. script purpose:
         1. Make maps that you can interact with to see your results at a glance
      ii. files used by script:
         1. data/intermediate/rate_summary.csv
         2. R_scripts/sourced/007_map_making.R
         3. img/[all files in directory].png
      iii. files produced by script:
         1. none unless you save things interactively from the Viewer

10. Suggestion only: **Submit processed data and metadata** to the CDMO annually for flat file hosting
    a. data/processed/GNDset_processed.csv
    b. metadata/GNDset_metadata.xlsx
    c. any narrative metadata that's developed

Details on the raw data file and QA/QC codes

Guidelines for use of the raw file:

- **DO NOT CHANGE** column names
- 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:

- one worksheet per SET
- one row per arm per date = 4 rows per date
- columns: set_id, date, arm_position, arm_qaqc_code, pin information (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) 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 and are to the far right

QA/QC codes:

- 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 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
- don't use a 2-letter code without the corresponding 3-letter code
- codes are roughly divided into: C-- (general comments); H-- (Higher pin than "expected"); L-- (Lower pin than "expected")

Where to find important files

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/qaqc_codes.xlsx

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

Contents of the Directory

details on each folder and subfolder, sorted alphabetically

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.* It contains image files that are pulled into R and used as icons on the maps.

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. qaqc_codes.xlsx *this file should not need modification*; it is a list and explanations of the QA/QC codes that have been generated for use across reserves.
    c. 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.
    d. 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 Word report is saved.
    b. figures This is where plots and graphs will be saved.
    c. processing_and_qc If we generate static QA/QC reports, they will go here.

7. **R_scripts** *don't mess with these unless you really know what you're doing*