SWMPr: An R package for retrieving, organizing, and analyzing environmental data for estuaries

Marcus William Beck[1, *]

1 ORISE Research Participation Program, USEPA National Health and Environmental Effects Research Laboratory, Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, FL 32651, USA

*** beck.marcus@epa.gov**

Abstract

Standardized monitoring programs have vastly improved the quantity and quality of data that form the basis of environmental decision-making. One example is the System Wide Monitoring Program (SWMP) that was implemented in 1995 by the federally-funded National Estuarine Research Reserve System (NERRS) . This program has provided two decades of continuous monitoring data at over 300 fixed stations in 28 estuaries across the United States. SWMP data have been used in a variety applications with the general objective of describing dynamics of estuarine ecosystems to better inform effective coastal management. However, simple tools for processing and evaluating the large and increasing quantity of data provided by the monitoring network have prevented large-scale comparisons between systems and, in some cases, simple trend analysis of water quality parameters at individual sites. We describe a new open-source software package, SWMPr, developed in program R for use with SWMP environmental data. The package provides several functions that facilitate data retrieval, organization, and analysis of time series data to describe water quality, weather, and nutrient dynamics in the reserve estuaries. Previously unavailable functions for estuaries are also provided to estimate rates of ecosystem metabolism using the open-water method. Tools included with the SWMPr package have facilitated a cross-reserve comparison of trends, including simple evaluation of changes over time and comparisons of patterns in primary productivity. Overall, the package provides an effectives approach to link quantitative information with analysis tools that will greatly inform management programs aimed at coastal protection and restoration.

Introduction

Info about SWMP, etc.

Materials and Methods

SWMPr is an R package that contains functions for retrieving, organizing, and analyzing estuary monitoring data from the System Wide Monitoring Program (SWMP). SWMP was implemented by the National Estuarine Research Reserve System (NERRS) in 1995 to provide continuous monitoring data at over 300 stations in 28 estuaries across the United States. SWMP data are maintained online by the

Centralized Data Management Office (CDMO). This R package provides several functions to retrieve, organize, and analyze SWMP data from the CDMO. Information on the CDMO web services are available here. Data can be downloaded directly from the CDMO using functions in this package, although it is easier to first download data outside of R and then use the `import_local` function. Your computer's IP address must be registered by CDMO staff for direct downloads in R. Detailed methods for importing SWMP data in R are described below.

All data obtained from the CDMO should be cited using the format:

National Estuarine Research Reserve System (NERRS). 2012. System-wide Monitoring Program. Data accessed from the NOAA NERRS Centralized Data Management Office website: http://cdmo.baruch.sc.edu/; accessed 12 October 2012.

To cite this package:

Beck MW. 2015. SWMPr: An R package for the National Estuarine Research Reserve System. Version 1.6.0. https://github.com/fawda123/SWMPr

Installing the package

This package is currently under development and has not been extensively tested. The development version can be installed from Github:

```r
install.packages('devtools')
library(devtools)
install_github('fawda123/SWMPr')
library(SWMPr)
```

Note that the current version of devtools (v1.6.1) was built under R version 3.1.1. The SWMPr package may not install correctly with older versions of R.

Data retrieval

SWMP data can be used in R after they are obtained directly from the CDMO through an online query or by using the retrieval functions provided in this package. In the latter case, the IP address for the computer making the request must be registered with CDMO. This can be done by following instructions here. The metadata should also be consulted for available data, including the parameters and date ranges for each monitoring station. Metadata are included as a .csv file with data requested from the CDMO and can also be obtained using the `site_codes` (all sites) or `site_codes_ind` (individual site) functions. Again, these functions will only work if the computer's IP address is registered with CDMO.

```r
# retrieve metadata for all sites
site_codes()

# retrieve metadata for a single site
site_codes_ind('apa')
```

Due to rate limitations on the server, the retrieval functions in this package return a limited number of records. The functions are more useful for evaluating short time periods, although these functions could be used iteratively (i.e., with `for` or `while` loops) to obtain longer time series. Data retrieval functions to access the CDMO include `all_params`, `all_params_dtrng`, and `single_param`. These are functions that call the existing methods on the CDMO web services. `all_params` returns the most recent 100 records of all parameters at a station, `all_params_dtrng` returns all records within a date range for all parameters or a single parameter, and single_param is identical to all_params except that a single parameter is requested.

```
# all parameters for a station, most recent
all_params('hudscwq')

# get all parameters within a date range
all_params_dtrng('hudscwq', c('09/10/2012', '02/8/2013'))

# get single parameter within a date range
all_params_dtrng('hudscwq', c('09/10/2012', '02/8/2013'), param = 'do_mgl')

# single parameter for a station, most recent
single_param('hudscwq', 'do_mgl')
```

For larger requests, it's easier to obtain data outside of R using the CDMO query system and then importing within R using the import_local function. Data can be retrieved from the CDMO several ways. The import_local function is designed for data from the zip downloads feature in the advanced query section of the CDMO. The function may also work using data from the data export system, but this feature has not been extensively tested (expect bugs). The zip downloads feature is an easy way to obtain data from multiple stations in one request. The downloaded data will be in a compressed folder that includes multiple .csv files by year for a given data type (e.g., apacpwq2002.csv, apacpwq2003.csv, apacpnut2002.csv, etc.). It is recommended that all stations at a site and the complete date ranges are requested to avoid repeated requests to CDMO. The import_local function can be used after the folder is decompressed.

Occasionally, duplicate time stamps are present in the raw data. The function handles duplicate entries differently depending on the data type (water quality, weather, or nutrients). For water quality and nutrient data, duplicate time stamps are simply removed. Note that nutrient data often contain replicate samples with similar but not duplicated time stamps within a few minutes of each other. Replicates with unique time stamps are not removed but can be further processed using rem_reps. Weather data prior to 2007 may contain duplicate time stamps at frequencies for 60 (hourly) and 144 (daily) averages, in addition to 15 minute frequencies. Duplicate values that correspond to the smallest value in the frequency column (15 minutes) are retained.

```
# import data for apaebmet that you downloaded

# this is an example path with the csv files, change as needed
path <- 'C:/my_path/'

# import, do not include file extension
import_local(path, 'apaebmet')
```

Raw csv data have not been included in the package due to size limitations. However, a sample dataset can be downloaded for use with the examples below. This dataset has an identical format as the data returned from the zip downloads feature of the CDMO. However, import time of the raw data may slow down use of the examples below and I have included binary data files (.RData) that are processed versions of the raw data.

Online widget

Results

Discussion

Supporting Information

Online widget

Online widget. Description of online widget.

Acknowledgments

Text

References