NERRS / SWMP

Data Analysis Workshop: *Time Series*

November 17, 2014

SWMP data retrieval and preparation

Marcus W. Beck[1]   Todd D. O'Brien[2]

[1]ORISE, USEPA NHEERL Gulf Ecology Division
Email: beck.marcus@epa.gov

[2]NOAA/NMFS Copepod Project
Email: todd.obrien@noaa.gov

Objectives and agenda

- Objectives

  - ▸ What are the various ways data are obtained from SWMP?

  - ▸ What are some issues that need to be addressed before importing into a statistical program to conduct a time series analysis?

Objectives and agenda

- Objectives

  - ▶ What are the various ways data are obtained from SWMP?

  - ▶ What are some issues that need to be addressed before importing into a statistical program to conduct a time series analysis?

- Agenda

  - ▶ Brief overview of SWMP network and available data

  - ▶ Format and potential issues with output data

  - ▶ Retrieving and importing the data
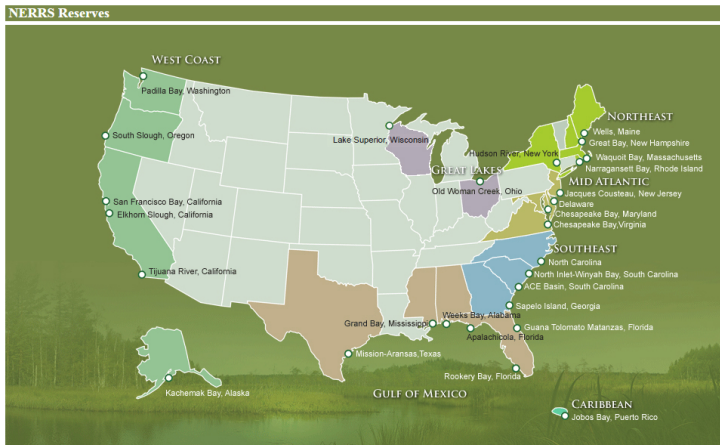
Interactive portion

You can follow along later in this module:

- dataset1

- script1

Interactive! Interrupt me!

Overview of SWMP and available data

SWMP - System Wide Monitoring Program, initiated in 1995 to provide continuous monitoring data at over 300 stations in 28 US estuaries

Overview of SWMP and available data

CDMO is your one-stop shop for retrieving SWMP data

Overview of SWMP and available data

Data can be exported from CDMO several ways:

Data Export System

The DES was developed to provide the majority of users with quick and easy access to SWMP data. The DES utilizes a map-based interface and offers single station exports, yearly authenticated file downloads (these may include non-standard nutrient parameters), charting, and a current conditions display for real-time stations.

To launch the Data Export System, click here.

Advanced Query System

The AQS was developed to specifically address the data delivery needs of those end-users looking for large amounts of data exported in a format that can be easily imported and manipulated for data analysis. The AQS offers three different query options allowing for mass downloads of annual files, customized queries for specific parameters and multiple stations in the same file, and an option to merge water quality, meteorological and nutrient datasets.

To launch the Advanced Query System, click here.

Real Time Data Application

The Real Time Application allows users to view near real time data, real time gauges, and 24 hour graphs with multiple parameters. You may use a bookmarked link to directly access the station of interest, or browse and select your station. The display will update automatically with the latest information as it comes in.

To launch the Real Time Data Application, click here.
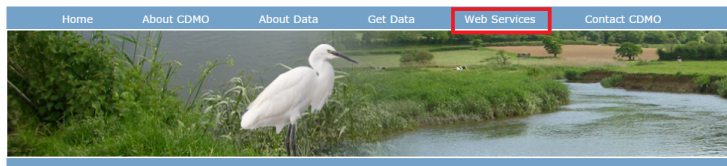

GIS Application

The GIS Application gives users access to Reserve boundary, watershed boundary, and high resolution reserve habitat maps. In addition, Google Earth KML files are available for the Reserve boundaries, watershed boundaries, and monitoring station locations.

To launch the GIS Application, click here.

Overview of SWMP and avialable data

You can also use the SWMPr package to retrieve data

Data retrieval functions connect to the CDMO web services, more about this later

Overview of SWMP and available data

A wide range of data can be requested... a few records for one site to all records for multiple sites

Requests can return a lot of data so make sure you have clear objectives

Check the available data before making a request!

- station names

- data types

- date ranges

- parameters

Overview of SWMP and available data

Available data: http://cdmo.baruch.sc.edu/data/availableOne.cfm

Overview of SWMP and available data

How to view available data:

- Trial-and-error (not recommended)

- View online: http://cdmo.baruch.sc.edu/data/availableOne.cfm

- View after request: 'sampling˙stations.csv'

- View after request: year and station specific .doc file

- Retrieve from within R (will cover later)

Overview of SWMP and available data

How to view available data:

- Trial-and-error (not recommended)

- View online: http://cdmo.baruch.sc.edu/data/availableOne.cfm

- View after request: 'sampling˙stations.csv'

- View after request: year and station specific .doc file

- Retrieve from within R (will cover later)

Now that you have the data, what do they look like?

Format and potential issues with output data

To orient yourself, understand the NERRS/SWMP naming convention

Site (reserve), ***station***, and ***parameter type*** are identified by a 7 or 8 character name

Format and potential issues with output data

To orient yourself, understand the NERRS/SWMP naming convention

Site (reserve), **station**, and **parameter type** are identified by a 7 or 8 character name

E.g., elkcwmet

Format and potential issues with output data

To orient yourself, understand the NERRS/SWMP naming convention

Site (reserve), **station**, and **parameter type** are identified by a 7 or 8 character name

E.g., elkcwmet

- elk: site, Elkhorn Slough

- cw: station, Caspian Weather Station

- met: parameter type (weather)

Format and potential issues with output data

The fundamental unit of data is the 'station' defined by a parameter type

The parameters for a station are specific to the parameter type

Format and potential issues with output data

The fundamental unit of data is the 'station' defined by a parameter type

The parameters for a station are specific to the parameter type

Nutrients	*Water quality*	*Meteorology*
po4f, chla_n, no3f, no2f, nh4f, no23f, ke_n, urea	temp, spcond, sal, do_pct, do_mgl, depth, cdepth, level, clevel, ph, turb, chlfluor	atemp, rh, bp, wspd, maxwspd, wdir, sdwdir, totpar, totprcp, cumprcp, totsorad

Format and potential issues with output data

Each parameter will also have a QAQC column, with the prefix 'f_'

E.g., 'atemp' and 'f_atemp'

Values in these columns describe whether the data passed automated QAQC checks

-5 Outside high sensor range
-4 Outside low sensor range
-3 Data rejected due to QAQC
-2 Missing data
-1 Optional SWMP supported parameter
0 Passed initial QAQC checks
1 Suspect data
2 Open - reserved for later flag
3 Calculated data: non-vented depth/level sensor correction for changes in barometric pressure
4 Historical data: Pre-auto QAQC
5 Corrected data

Format and potential issues with output data

You will have to decide how to handle QAQC...

CDMO has limited export options for dealing with bad QAQC flags

We will learn how to handle QAQC flags in R

Format and potential issues with output data

The final piece of of the puzzle is the DateTimeStamp

Format is month/day/year hours:minutes, based on UTC offset and no daylight savings!

Varies by station and parameter type

	A	B	C
1	StationCo	isSWMP	DateTimeStamp
2	apacpnut	P	1/10/2012 10:20
3	apacpnut	P	2/7/2012 11:41
4	apacpnut	P	3/5/2012 11:51
5	apacpnut	P	4/4/2012 10:30
6	apacpnut	P	5/9/2012 10:12
7	apacpnut	P	5/9/2012 10:15
8	apacpnut	P	5/9/2012 10:20
9	apacpnut	P	6/5/2012 8:30
10	apacpnut	P	7/3/2012 9:58

	A	B	C
1	StationCo	isSWMP	DateTimeStamp
2	apacpwq	P	1/1/2012 0:00
3	apacpwq	P	1/1/2012 0:15
4	apacpwq	P	1/1/2012 0:30
5	apacpwq	P	1/1/2012 0:45
6	apacpwq	P	1/1/2012 1:00
7	apacpwq	P	1/1/2012 1:15
8	apacpwq	P	1/1/2012 1:30
9	apacpwq	P	1/1/2012 1:45
10	apacpwq	P	1/1/2012 2:00

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)

- Dealing with QAQC columns and removing 'bad' observations

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)

- Dealing with QAQC columns and removing 'bad' observations

- Data we don't want... extra columns or irrelevant parameters

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)

- Dealing with QAQC columns and removing 'bad' observations

- Data we don't want... extra columns or irrelevant parameters

- Combining data for comparison

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)

- Dealing with QAQC columns and removing 'bad' observations

- Data we don't want... extra columns or irrelevant parameters

- Combining data for comparison
- Issues inherent with time series, e.g., missing data

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)

- Dealing with QAQC columns and removing 'bad' observations

- Data we don't want... extra columns or irrelevant parameters

- Combining data for comparison
- Issues inherent with time series, e.g., missing data

    We will learn how to handle most of these challenges!

Overview of the SWMPr package

What: An R package for retrieving, organizing and analyzing SWMP data

Overview of the SWMPr package

What: An R package for retrieving, organizing and analyzing SWMP data

Why: There are many challenges for working with SWMP data... a toolkit for addressing these challenges using an open-source format will be useful (I hope!)

Overview of the SWMPr package

What: An R package for retrieving, organizing and analyzing SWMP data

Why: There are many challenges for working with SWMP data... a toolkit for addressing these challenges using an open-source format will be useful (I hope!)

How:

- Install R/RStudio on your computer (done already?)

- Install the SWMPr package

- Use the SWMPr functions to *retrieve*, *organize*, and *analyze* SWMP data

Overview of the SWMPr package

This is where SWMPr lives - https://github.com/fawda123/SWMPr

Overview of the SWMPr package

Scroll down the page to view the README file, all instructions here...

Overview of the SWMPr package

Installation instructions are in the README

Run these four lines to install the package

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

What is it doing?

Packages can be installed from Github using the 'install_github' function from the devtools package

Overview of the SWMPr package

Your R console should look like this...

Overview of the SWMPr package

What is provided in the SWMPr package?

Retrieve

```
all_params
all_params_dtrng
single_param
import_local
```

Organize

```
qaqc.swmpr
subset.swmpr
setstep.swmpr
comb.swmpr
```

Analyze

```
aggregate.swmpr
smoother.swmpr
na.approx.swmpr
plot.swmpr
hist.swmpr
lines.swmpr
decomp.swmpr
```

Built around the concept of ***object-oriented programming*** - retrieval functions return a data type with specific methods to organize and analyze

Overview of the SWMPr package

To view the help file for any function (including examples):

```
?all_params
```

all_params {SWMPr}   R Documentation

Import current station records from the CDMO

Description

Import current station records from the CDMO starting with the most current date, CDMO equivalent of
`exportAllParamsXMLNew`

Usage

`all_params(station_code, Max = 100)`

Arguments

`station_code` chr string of station, 7 or 8 characters

`Max`   numeric value for number of records to obtain from the current date, maximum of 100

Value

Returns a swmpr object, all available parameters including QAQC columns

Overview of the SWMPr package

Let's get some data into R!

The **retrieval** functions do two things:

Import data directly from the CDMO:

```
all_params
all_params_dtrng
single_param
```

These functions require registering your IP address with CDMO, data are also rate-limited

Import data from a local path:

```
import_local
```

Allows import of data obtained from the zip downloads feature

Overview of the SWMPr package

After unzipping, data from zip downloads will have separate .csv files for each station and year

Name	Date modified	Type	Size
apacpnut2011.csv	9/19/2014 7:04 AM	Microsoft Excel C...	3 KB
apacpnut2012.csv	9/19/2014 7:04 AM	Microsoft Excel C...	3 KB
apacpnut2013.csv	9/19/2014 7:04 AM	Microsoft Excel C...	3 KB
apacpwq2011.csv	9/19/2014 7:06 AM	Microsoft Excel C...	5,481 KB
apacpwq2012.csv	9/19/2014 7:06 AM	Microsoft Excel C...	5,472 KB
apacpwq2013.csv	9/19/2014 7:06 AM	Microsoft Excel C...	5,567 KB
apadbnut2011.csv	9/19/2014 7:06 AM	Microsoft Excel C...	3 KB
apadbnut2012.csv	9/19/2014 7:06 AM	Microsoft Excel C...	3 KB
apadbnut2013.csv	9/19/2014 7:06 AM	Microsoft Excel C...	3 KB
apadbwq2011.csv	9/19/2014 7:08 AM	Microsoft Excel C...	5,407 KB
apadbwq2012.csv	9/19/2014 7:08 AM	Microsoft Excel C...	5,483 KB
apadbwq2013.csv	9/19/2014 7:08 AM	Microsoft Excel C...	5,337 KB
apaebmet2011.csv	9/19/2014 7:10 AM	Microsoft Excel C...	5,453 KB
apaebmet2012.csv	9/19/2014 7:10 AM	Microsoft Excel C...	5,401 KB
apaebmet2013.csv	9/19/2014 7:11 AM	Microsoft Excel C...	5,669 KB

Overview of the SWMPr package

Use the following to import data into R:

```
# get data for apacpwq, all years

# location of data
# mypath <- 'C:/data/dataset1'
mypath <- system.file('zip_ex', package = 'SWMPr')

# import and assign to 'dat'
dat <- import_local(mypath, 'apacpwq', trace = T)
```

Overview of the SWMPr package

Your console should look something like this:

```
# get data for apacpwq, all years

# location of data
# mypath <- 'C:/data/dataset1'
mypath <- system.file('zip_ex', package = 'SWMPr')

# import and assign to 'dat'
dat <- import_local(mypath, 'apacpwq', trace = T)

## Loading files...
##
## apacpwq2011.csv   apacpwq2012.csv   apacpwq2013.csv
##
## Combining data...
##
```

Overview of the SWMPr package

Now we have data in our 'workspace' that we can organize/analyze

```
head(dat)

##          datetimestamp temp f_temp spcond f_spcond sal f_sal do_pct f_do_pct
## 1 2011-01-01 00:00:00   11   <0>      44     <0>    28  <0>     68    <0>
## 2 2011-01-01 00:15:00   11   <0>      44     <0>    28  <0>     68    <0>
## 3 2011-01-01 00:30:00   11   <0>      44     <0>    28  <0>     68    <0>
## 4 2011-01-01 00:45:00   11   <0>      44     <0>    28  <0>     68    <0>
## 5 2011-01-01 01:00:00   11   <0>      44     <0>    29  <0>     68    <0>
## 6 2011-01-01 01:15:00   11   <0>      44     <0>    29  <0>     67    <0>
##   do_mgl f_do_mgl depth f_depth cdepth f_cdepth level f_level clevel f_clevel
## 1      6   <0>      2    <0>       2    <3>       NA   <-1>     NA      NA
## 2      6   <0>      2    <0>       2    <3>       NA   <-1>     NA      NA
## 3      6   <0>      2    <0>       2    <3>       NA   <-1>     NA      NA
## 4      6   <0>      2    <0>       2    <3>       NA   <-1>     NA      NA
## 5      6   <0>      2    <0>       2    <3>       NA   <-1>     NA      NA
## 6      6   <0>      2    <0>       2    <3>       NA   <-1>     NA      NA
##   ph f_ph turb f_turb chlfluor f_chlfluor
## 1  8 <0>     3   <0>       NA      <-1>
## 2  8 <0>     3   <0>       NA      <-1>
## 3  8 <0>     2   <0>       NA      <-1>
## 4  8 <0>     1   <0>       NA      <-1>
## 5  8 <0>     2   <0>       NA      <-1>
## 6  8 <0>     1   <0>       NA      <-1>
```

NERRS / SWMP
Data Analysis Workshop: *Time Series*
November 17, 2014

Questions??