# SWMP data retrieval and preparation

Marcus W. Beck

USEPA NHEERL Gulf Ecology Division, Gulf Breeze, FL
Email: beck.marcus@epa.gov, Phone: 850 934 2480

# Objectives and agenda

- Objectives
  - What are the various ways data are obtained from SWMP?
  - What needs to be done to the SWMP data to get it into a format to enter 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 needs to be done to the SWMP data to get it into a format to enter 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

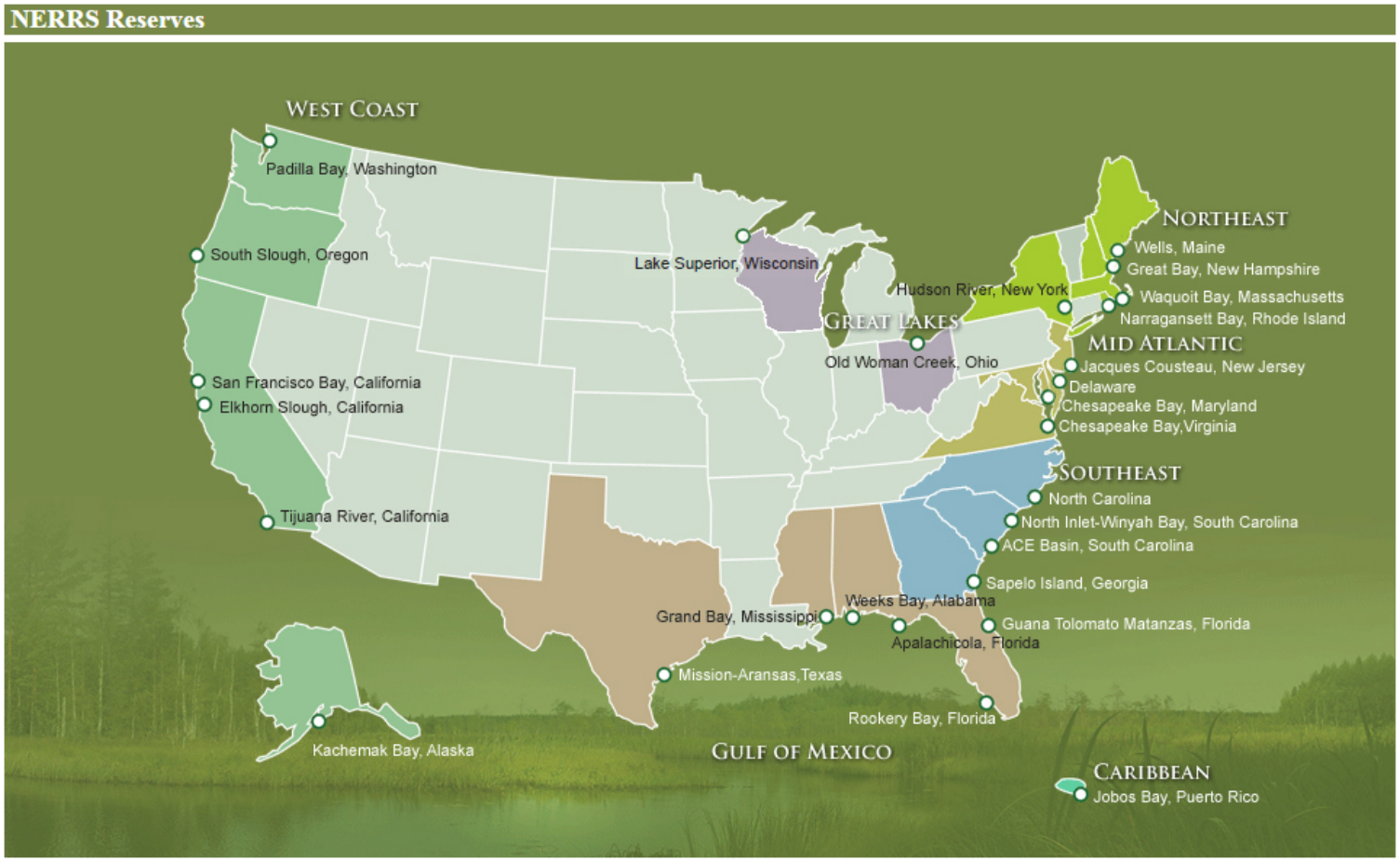
You can follow along later in this module:

- 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




http://nerrs.noaa.gov/ReservesMap.aspx

# Overview of SWMP and available data

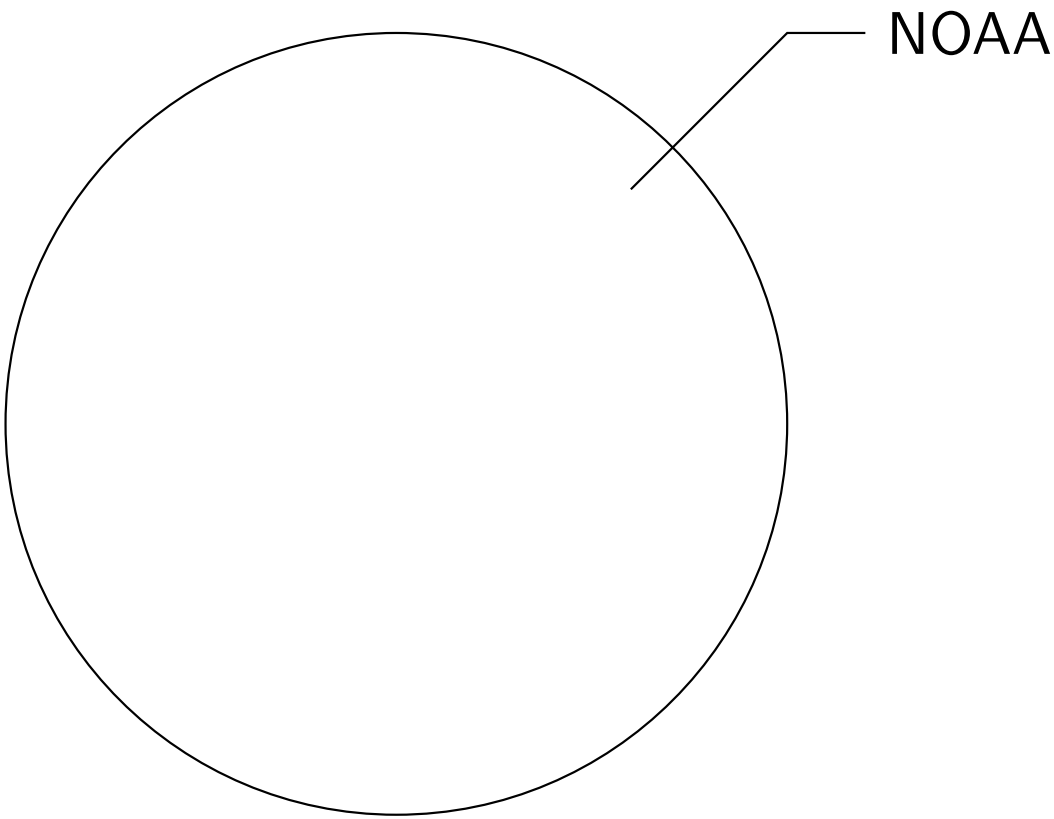
The first challenge in analyzing time series is obtaining the data

SWMP data are available through the Centralized Data Management Office (CDMO)

# Overview of SWMP and available data

The first challenge in analyzing time series is obtaining the data

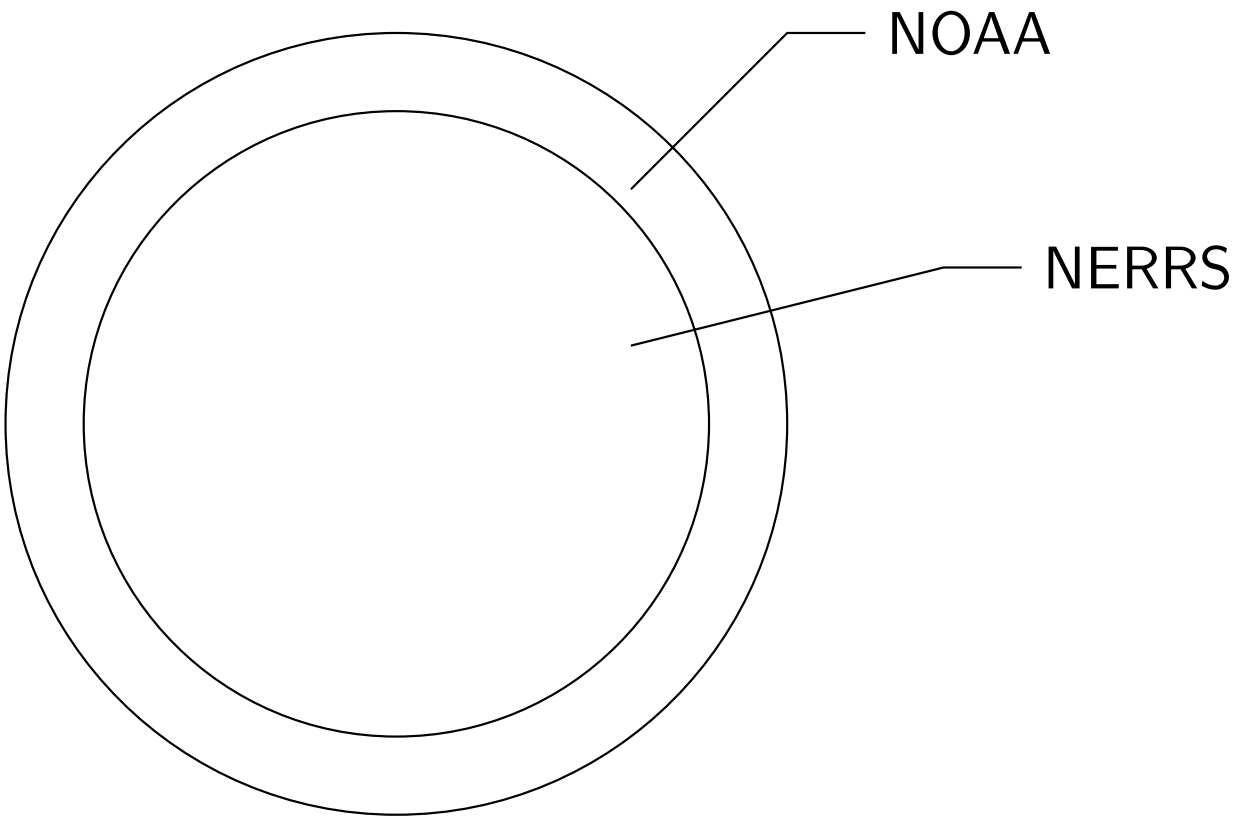
SWMP data are available through the Centralized Data Management Office (CDMO)



# Overview of SWMP and available data

The first challenge in analyzing time series is obtaining the data

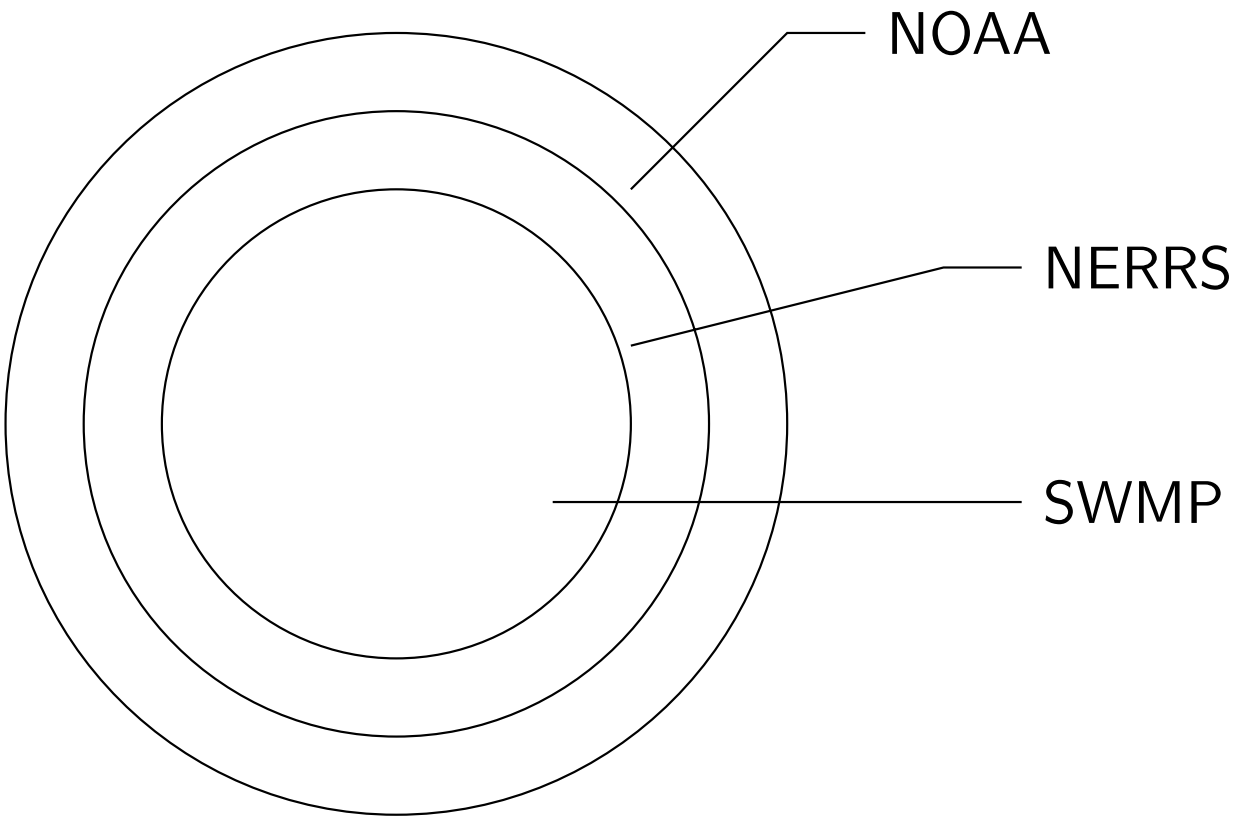
SWMP data are available through the Centralized Data Management Office (CDMO)



# Overview of SWMP and available data

The first challenge in analyzing time series is obtaining the data

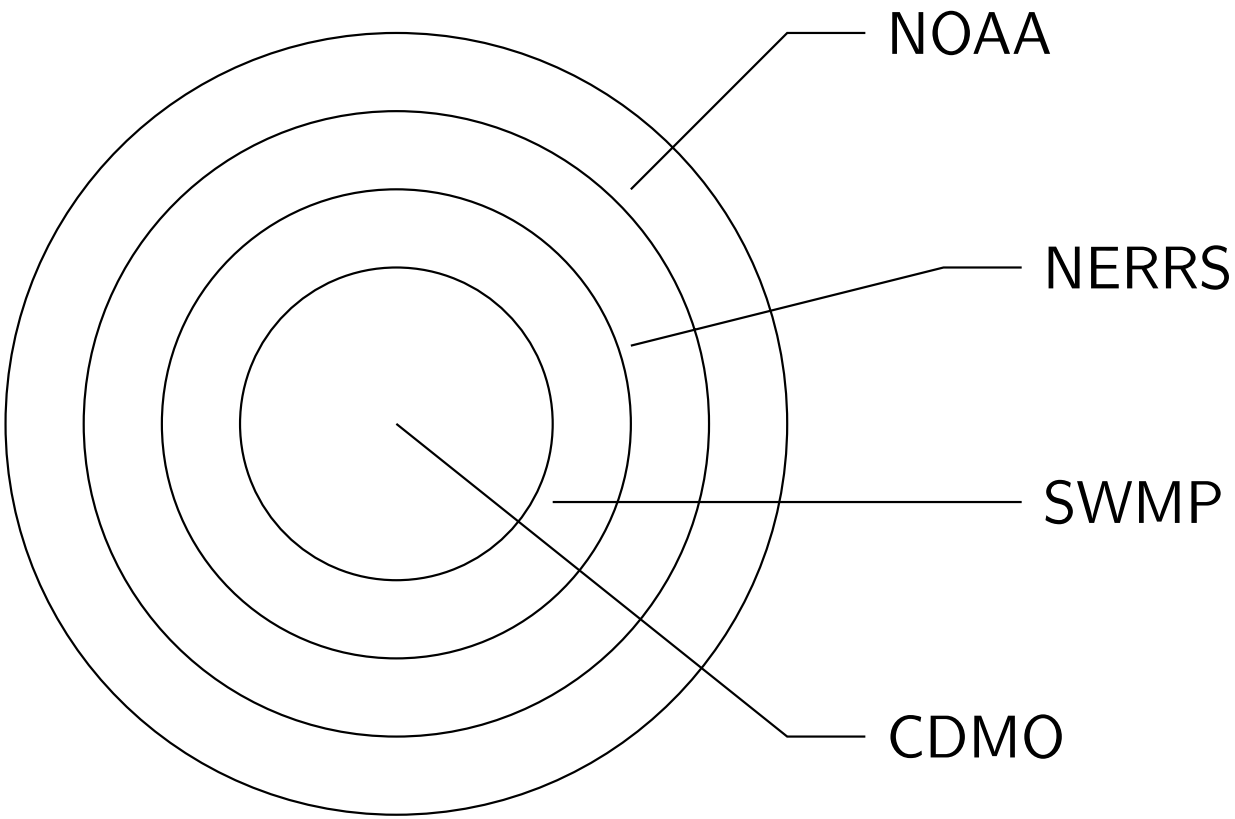
SWMP data are available through the Centralized Data Management Office (CDMO)



# Overview of SWMP and available data

The first challenge in analyzing time series is obtaining the data

SWMP data are available through the Centralized Data Management Office (CDMO)



# 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*

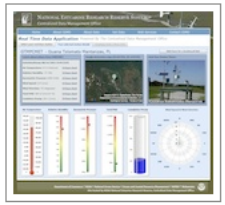
The DES was developed to provide the majority of users with quick and easy access to SWMP data. The DES utilizes a map-based interface and offers single station exports, yearly authenticated file downloads (these may include non-standard nutrient parameters), charting, and a current conditions display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

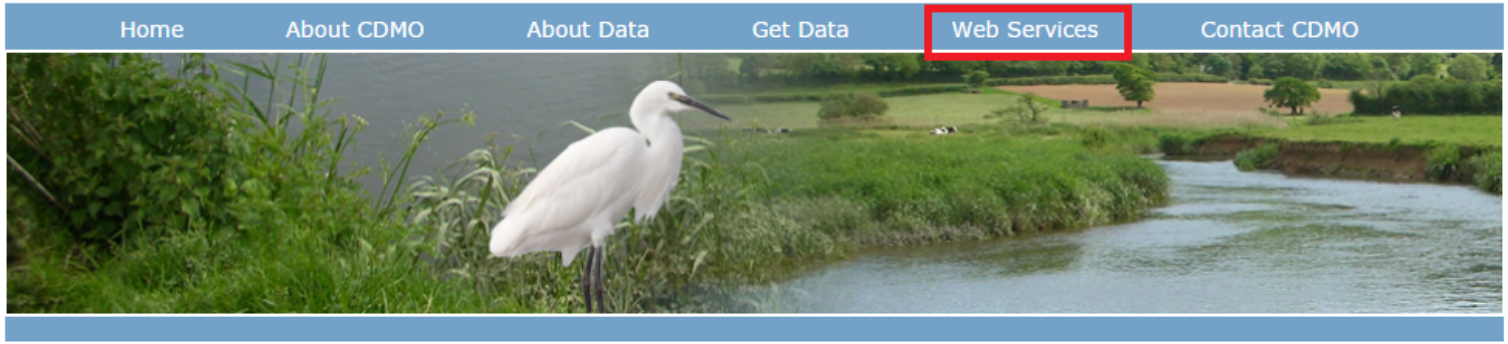
The GIS Application gives users access to Reserve boundary, watershed boundary, and high resolution reserve habitat maps. In addition, Google Earth KML files are available for the Reserve boundaries, watershed boundaries, and monitoring station locations.

*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




## Web Services

### Web Services

In an effort to increase distribution and use of the data collected by the NERRS, the Centralized Data Management Office has created several web service products for this purpose. These services can be used to pull real-time data from our databases for use by other individuals and organizations.

***You must contact cdmowebmaster@belle.baruch.sc.edu for authorization before pulling data from the CDMO.***

**Transmission Time Schedule**

- Realtime satellite transmissions are received hourly based upon the diagram shown below. Decoding occurs at 11 minutes past the hour and 41 minutes past the hour. This data becomes available on the CDMO website at 15 and 45 minutes past the hour. Each transmission will have 4 15-minute records in it.
- What does all this mean? Let's take Block #5 as an example. Three stations are transmitted in that block from 47:00 - 47:30 past the hour. The data from those stations will be decoded at 11 minutes past the hour and will be available on this website at 15 minutes past the hour.

# 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

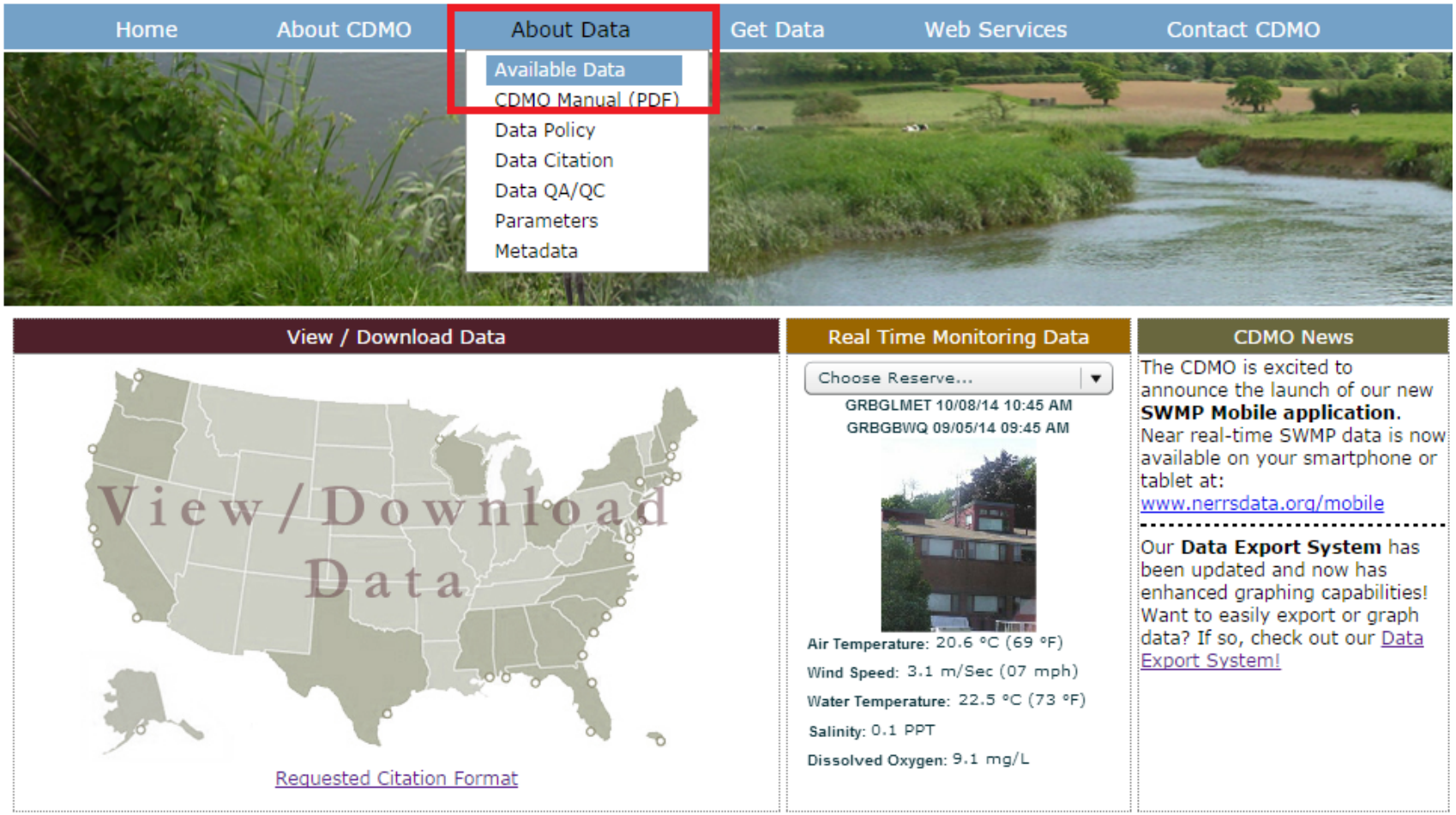
Requests can return a lot of data so make sure you have clear objectives

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

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

## Available Data

A total of **56,247,777** NERR SWMP data records are currently available from the CDMO as of 08-Oct-14 11:31 AM.

| Weather Data Records | Water Data Records | Nutrient Data Records |
|---|---|---|
| 13,066,086 | 43,088,226 | 93,465 |

---

The following table shows the type of data available at each reserve.

Data Availability Summary

**ACE Basin , SC**

Water Quality Data
1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

Weather Quality Data
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

Nutrient Data
2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Nutrient Parameters Available
NO23F, PO4F, CHLA_N, NO3F, NO2F, DIN, NH4F

---

**Apalachicola , FL**

Water Quality Data
1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

Weather Quality Data
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

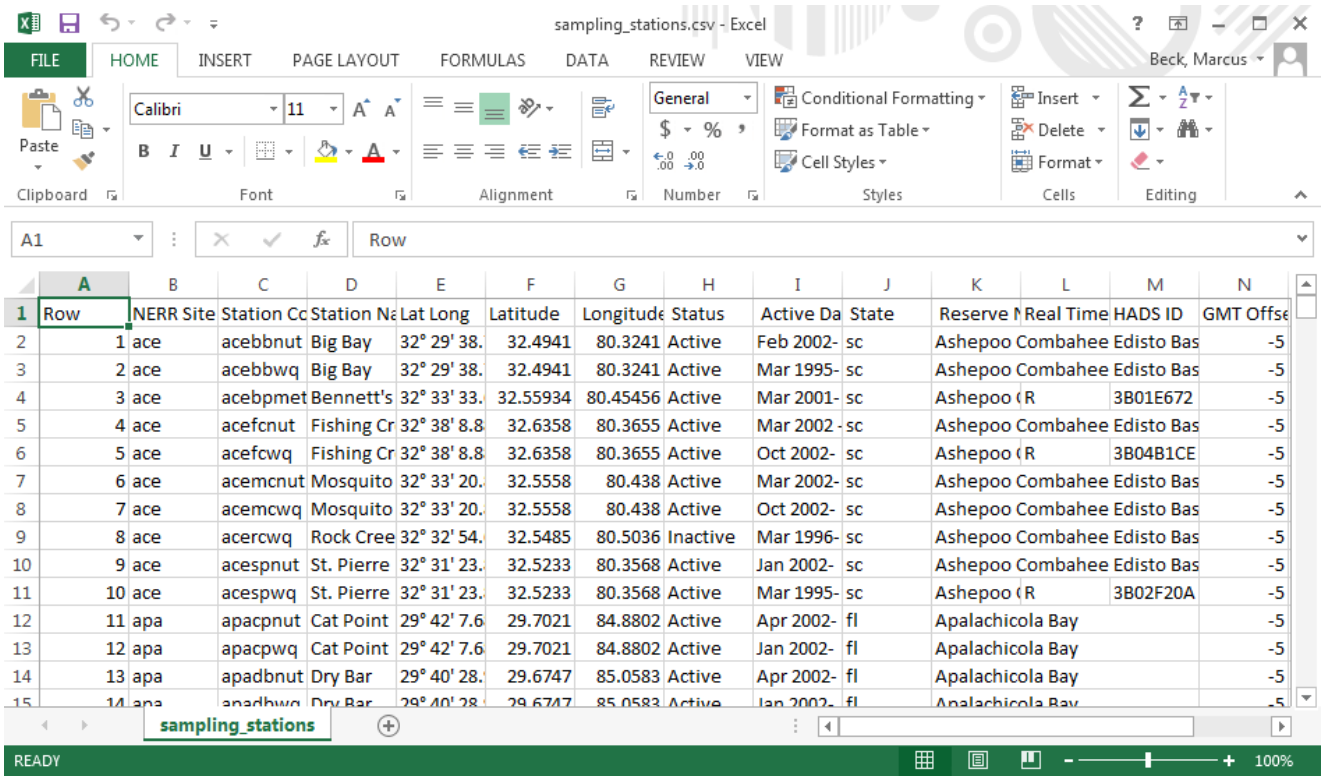
Nutrient Data
2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Nutrient Parameters Available
PO4F, NH4F, NO2F, NO3F, NO23F, DIN, CHLA_N, WTEM_N, SALT_N, DO_N, DO_S_N, TURB_N, PHEA

# Overview of SWMP and available data

Metadata are also returned with any data request... a summary of availability

As 'sampling_stations.csv':

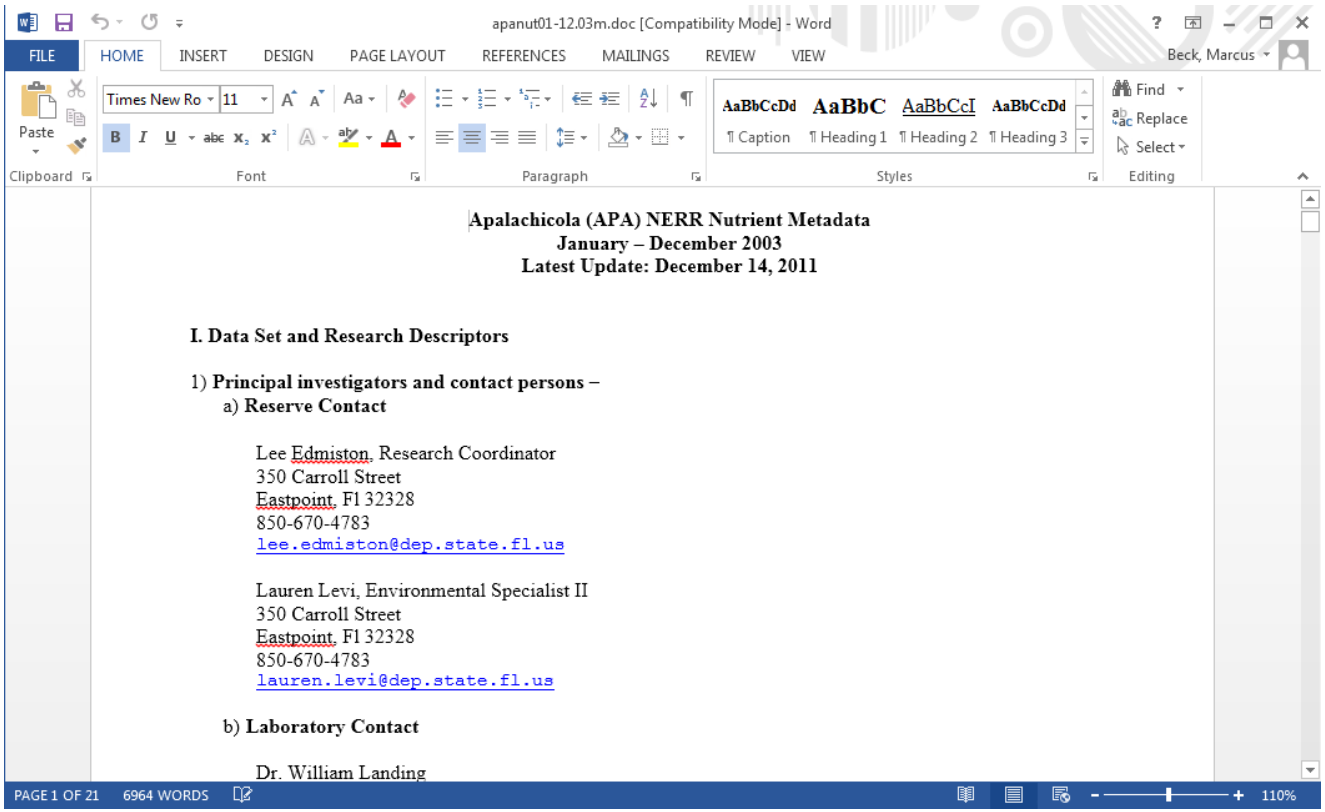



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Row | NERR Site | Station Cc | Station Na | Lat Long | Latitude | Longitude | Status | Active Da | State | Reserve N | Real Time | HADS ID | GMT Offse |
| 2 | 1 | ace | acebbnut | Big Bay | 32° 29' 38. | 32.4941 | 80.3241 | Active | Feb 2002- | sc | Ashepoo | Combahee | Edisto Bas | -5 |
| 3 | 2 | ace | acebbwq | Big Bay | 32° 29' 38. | 32.4941 | 80.3241 | Active | Mar 1995- | sc | Ashepoo | Combahee | Edisto Bas | -5 |
| 4 | 3 | ace | acebpmet | Bennett's | 32° 33' 33. | 32.55934 | 80.45456 | Active | Mar 2001- | sc | Ashepoo | R | 3B01E672 | -5 |
| 5 | 4 | ace | acefcnut | Fishing Cr | 32° 38' 8.8 | 32.6358 | 80.3655 | Active | Mar 2002 - | sc | Ashepoo | Combahee | Edisto Bas | -5 |
| 6 | 5 | ace | acefcwq | Fishing Cr | 32° 38' 8.8 | 32.6358 | 80.3655 | Active | Oct 2002- | sc | Ashepoo | R | 3B04B1CE | -5 |
| 7 | 6 | ace | acemcnut | Mosquito | 32° 33' 20. | 32.5558 | 80.438 | Active | Mar 2002- | sc | Ashepoo | Combahee | Edisto Bas | -5 |
| 8 | 7 | ace | acemcwq | Mosquito | 32° 33' 20. | 32.5558 | 80.438 | Active | Oct 2002- | sc | Ashepoo | Combahee | Edisto Bas | -5 |
| 9 | 8 | ace | acercwq | Rock Cree | 32° 32' 54. | 32.5485 | 80.5036 | Inactive | Mar 1996- | sc | Ashepoo | Combahee | Edisto Bas | -5 |
| 10 | 9 | ace | acespnut | St. Pierre | 32° 31' 23. | 32.5233 | 80.3568 | Active | Jan 2002- | sc | Ashepoo | Combahee | Edisto Bas | -5 |
| 11 | 10 | ace | acespwq | St. Pierre | 32° 31' 23. | 32.5233 | 80.3568 | Active | Mar 1995- | sc | Ashepoo | R | 3B02F20A | -5 |
| 12 | 11 | apa | apacpnut | Cat Point | 29° 42' 7.6 | 29.7021 | 84.8802 | Active | Apr 2002- | fl | Apalachicola Bay | | | -5 |
| 13 | 12 | apa | apacpwq | Cat Point | 29° 42' 7.6 | 29.7021 | 84.8802 | Active | Jan 2002- | fl | Apalachicola Bay | | | -5 |
| 14 | 13 | apa | apadbnut | Dry Bar | 29° 40' 28. | 29.6747 | 85.0583 | Active | Apr 2002- | fl | Apalachicola Bay | | | -5 |
| 15 | 14 | apa | apadbwq | Dry Bar | 29° 40' 28. | 29.6747 | 85.0583 | Active | Jan 2002- | fl | Apalachicola Bay | | | -5 |

# Overview of SWMP and available data

Metadata are also returned with any data request... a summary of availability

As Word document (e.g., 'apanut01-12.03m.doc') :



# 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

| **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

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 |

Note that these are lower case, same as the data you will work with in R but not the same as data from the CDMO

# 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...

| | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | StationCo | isSWMP | DateTimeStamp | Historical | Provisiona | CollMetho | REP | F_Record | PO4F | F_PO4F | NH4F | F_NH4F |
| 2 | apacpnut | P | 1/10/2012 10:20 | 0 | 1 | 1 | 1 | | 0.003 | <-4> [SBL] | 0.03 | <0> |
| 3 | apacpnut | P | 2/7/2012 11:41 | 0 | 1 | 1 | 1 | | 0.005 | <0> | 0.019 | <0> |
| 4 | apacpnut | P | 3/5/2012 11:51 | 0 | 1 | 1 | 1 | | 0.003 | <-4> [SBL] | 0.041 | <0> |
| 5 | apacpnut | P | 4/4/2012 10:30 | 0 | 1 | 1 | 1 | | 0.003 | <-4> [SBL] | 0.043 | <0> |
| 6 | apacpnut | P | 5/9/2012 10:12 | 0 | 1 | 1 | 1 | | 0.003 | <0> | 0.053 | <0> |
| 7 | apacpnut | P | 5/9/2012 10:15 | 0 | 1 | 1 | 2 | | 0.003 | <-4> [SBL] | 0.022 | <0> |
| 8 | apacpnut | P | 5/9/2012 10:20 | 0 | 1 | 1 | 3 | | 0.003 | <0> | 0.016 | <0> |
| 9 | apacpnut | P | 6/5/2012 8:30 | 0 | 1 | 1 | 1 | | 0.003 | <-4> [SBL] | 0.04 | <0> |
| 10 | apacpnut | P | 7/3/2012 9:58 | 0 | 1 | 1 | 1 | {CSM} | 0.004 | <0> | 0.094 | <0> |
| 11 | apacpnut | P | 7/3/2012 9:59 | 0 | 1 | 1 | 2 | {CSM} | 0.004 | <0> | 0.066 | <0> |
| 12 | apacpnut | P | 7/3/2012 10:01 | 0 | 1 | 1 | 3 | {CSM} | 0.005 | <0> | 0.069 | <0> |
| 13 | apacpnut | P | 8/7/2012 9:53 | 0 | 1 | 1 | 1 | {CSM} | 0.003 | <-4> [SBL] | 0.05 | <0> |
| 14 | apacpnut | P | 9/5/2012 10:56 | 0 | 1 | 1 | 1 | | 0.003 | <-4> [SBL] | 0.026 | <0> |
| 15 | apacpnut | P | 10/2/2012 9:22 | 0 | 1 | 1 | 1 | | 0.003 | <-4> [SBL] | 0.042 | <0> |
| 16 | apacpnut | P | 10/2/2012 9:27 | 0 | 1 | 1 | 2 | | 0.003 | <-4> [SBL] | 0.024 | <0> |
| 17 | apacpnut | P | 10/2/2012 9:32 | 0 | 1 | 1 | 3 | | 0.003 | <0> | 0.042 | <0> |
| 18 | apacpnut | P | 11/6/2012 10:30 | 0 | 1 | 1 | 1 | | 0.003 | <-4> [SBL] | 0.07 | <0> |
| 19 | apacpnut | P | 11/26/2012 11:39 | 0 | 1 | 1 | 1 | | 0.003 | <-4> [SBL] | 0.041 | <0> |

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!

Time step also varies by station, parameter type, and when the data were obtained

| | 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)
- Dealing with QAQC columns and removing 'bad' observations
- Comparing data of different parameter types
- Comparing data with different time steps
- Comparing data between stations
- Data we don't want... extra columns or irrelevant parameters

# Format and potential issues with output data

Not to mention the inherent issues with time series...

- Missing data
- Noise vs signal
- Drift or instrument malfunction
- Others?

# Format and potential issues with output data

Not to mention the inherent issues with time series...

- Missing data
- Noise vs signal
- Drift or instrument malfunction
- Others?

*We will learn how to handle most of these challenges!*

# Overview of the SWMPr package