Marcus W. Beck

ORISE, USEPA NHEERL Gulf Ecology Division
Phone: 8509342480, Email: beck.marcus@epa.gov

May 8, 2015

What is NERRS/SWMP?

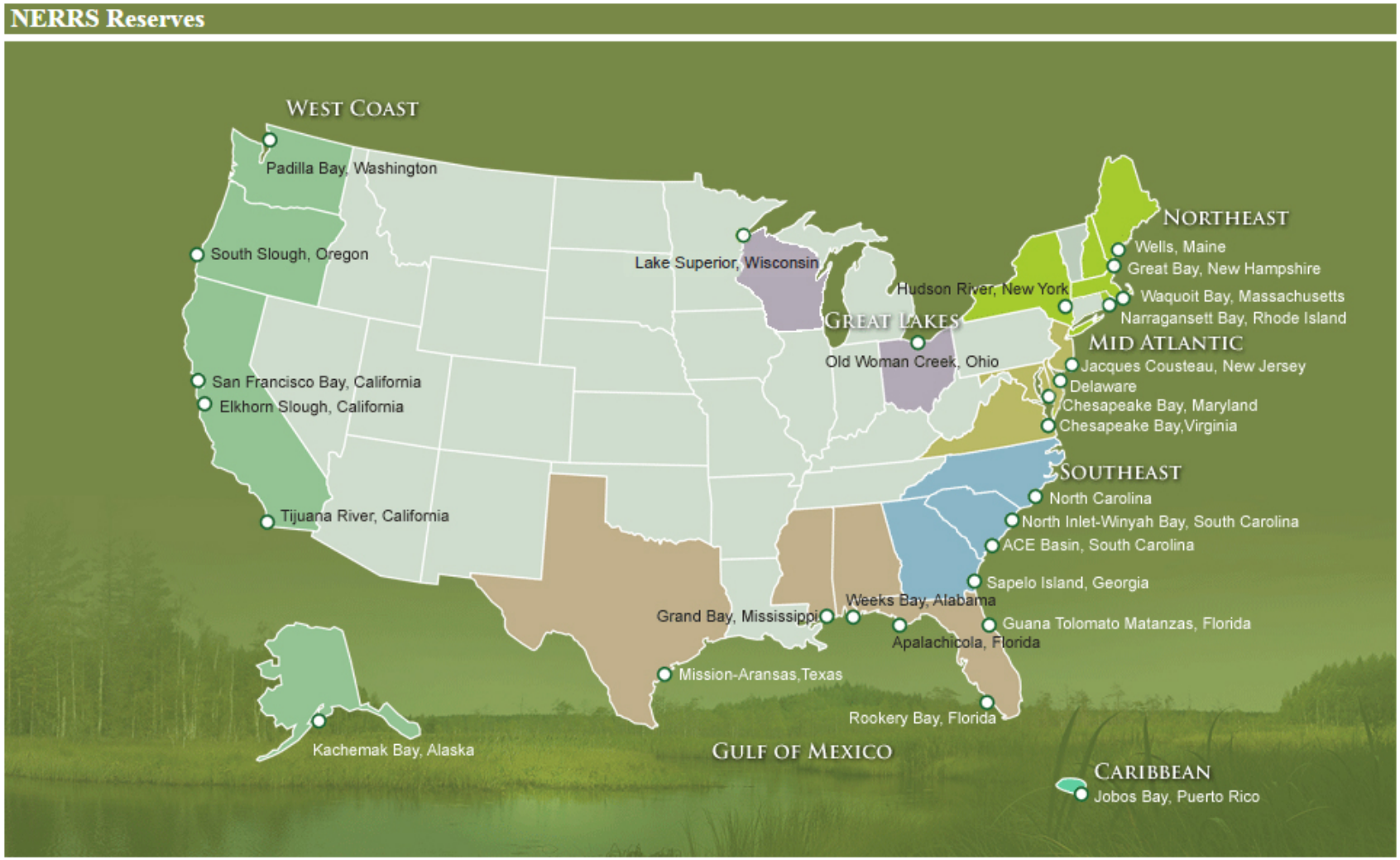
NERRS

National Estuarine Research Reserve System, established by Coastal Zone Management Act of 1972. Focus on ***long-term research***, ***monitoring***, ***education***, and ***stewardship*** for more effective coastal management.

SWMP

System Wide Monitoring Program, initiated in 1995 to provide ***continuous monitoring*** data at over 140 stations in each of the 28 NERRS reserves

What is NERRS/SWMP?



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

What is NERRS/SWMP?

Data maintained by the Centralized Data Management Office (CDMO)

What is NERRS/SWMP?

Each reserve has fixed, continuous monitoring stations for ***water quality*** (15 min), ***meteorology*** (15 min), and ***nutrients*** (monthly)

CDMO is an existing data management infrastructure for SWMP:

- Automated QAQC
- Numerous data download options
- Web services/API for remote data upload/retrieval
- Simple viz tools

What is NERRS/SWMP?

As of May 1, $>$ 58 million SWMP data records available from CDMO

Raw data will look like this...

	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>

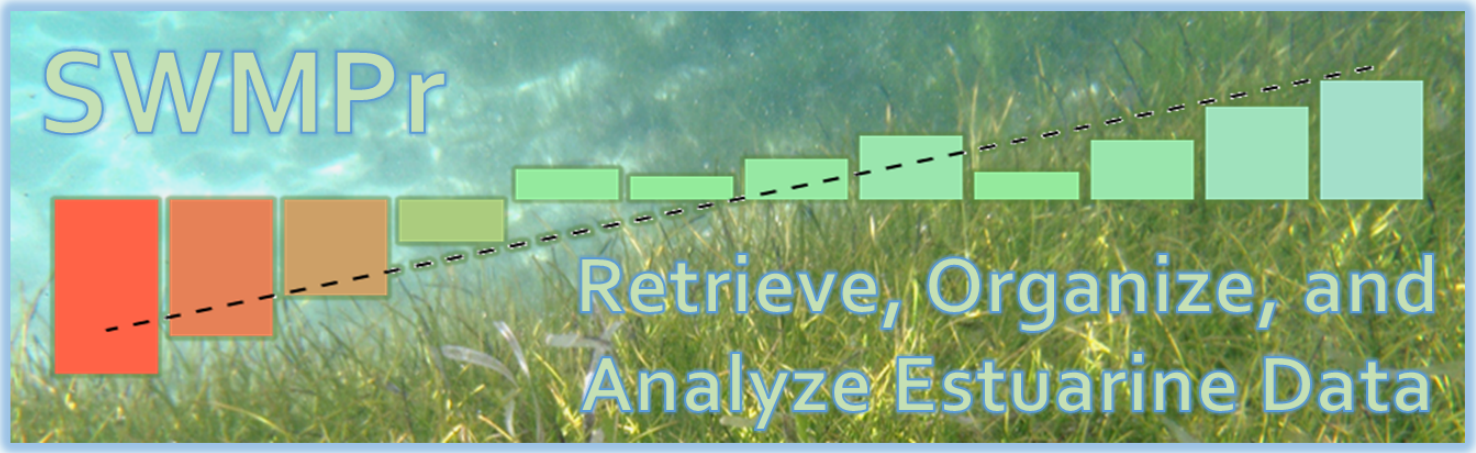
What is the problem?

An invaluable data source but no recent comparative analyses, need for tools for simple trend analysis

Some specific issues:

- Getting the data into an analysis program
- Dealing with QAQC columns or removing 'bad' observations
- Combining data for comparison
- Issues inherent with time series, e.g., signal vs. noise, data quantity
- ...and analysis

What is the (potential) solution?



An R package for SWMP data, SWMPr v2.0.0 is officially released!

```
> install.packages('SWMPr')
> library(SWMPR)
```

Still in development, currently on v2.0.5

SWMPr is fully documented

Package 'SWMPr'

April 1, 2015

Type Package

Title Retrieving, Organizing, and Analyzing Estuary Monitoring Data

Version 2.0.0

Date 2015-4-1

Author Marcus W. Beck [aut, cre]

Maintainer Marcus W. Beck <mbafs2012@gmail.com>

Description Tools for retrieving, organizing, and analyzing environmental data from the System Wide Monitoring Program of the National Estuarine Research Reserve System. These tools address common challenges associated with continuous time series data for environmental decision making.

BugReports https://github.com/fawda123/SWMPr/issues

License CC0

Imports data.table, httr, ggmap, gridExtra, maptools, oce, dplyr, reshape2, tictoc, tidyr, wq, XML

LazyData true

Depends R (>= 3.1.1), ggplot2, zoo

NeedsCompilation no

Repository CRAN

Date/Publication 2015-04-01 23:58:53

R topics documented:

SWMPr is fully documented

qaqc *QAQC filtering for SWMP data*

Description

QAQC filtering for SWMP data obtained from retrieval functions, local and remote

Usage

```
qaqc(swmpr_in, ...)

## S3 method for class 'swmpr'
qaqc(swmpr_in, qaqc_keep = 0, trace = FALSE, ...)
```

Arguments

swmpr_in	input swmpr object
...	arguments passed to or from other methods
qaqc_keep	numeric vector of qaqc flags to keep, default 0
trace	logical for progress output on console, default FALSE

Details

The qaqc function is a simple screen to retain values from the data with specified QAQC flags, described online: http://cdmo.baruch.sc.edu/data/qaqc.cfm. Each parameter in the swmpr data typically has a corresponding QAQC column of the same name with the added prefix 'f_'. Values in the QAQC column specify a flag from -5 to 5. Generally, only data with the '0' QAQC flag should be used, which is the default option for the function. Data that do not satisfy QAQC criteria are converted to NA values. Additionally, simple filters are used to remove obviously bad values, e.g., wind speed values less than zero or pH values greater than 12. Erroneous data entered as -99 are also removed. Processed data will have QAQC columns removed, in addition to removal of values in the actual parameter columns that do not meet the criteria.

Value

Returns a swmpr object with NA values for records that did not match qaqc_keep. QAQC columns are also removed.

What can SWMPr do?

SWMPr functions are grouped into three categories that describe their use in the 'data workflow'

1) Retrieve → ***2) Organize*** → ***3) Analyze***

- Retrieve metadata
- Import from CDMO into R

- Manipulate data for analysis
- Functions to clean, combine, change time step, etc.

- Generic to specific applications
- Visualization and graphics

What can SWMPr do?

Function types are searchable in R:

```
> help.search('analyze', package = 'SWMPr')
```

Search Results

Help pages:

SWMPr::aggremetab	Aggregate metabolism data
SWMPr::aggreswmp	Aggregate swmpr data
SWMPr::decomp	Simple trend decomposition of swmpr data
SWMPr::decomp_cj	Simple trend decomposition of monthly swmpr data
SWMPr::ecometab	Ecosystem metabolism
SWMPr::hist.swmpr	Plot swmpr using a histogram
SWMPr::map_reserve	Map a reserve
SWMPr::na.approx.swmpr	Linearly interpolate gaps
SWMPr::lines.swmpr	Plot swmpr data
SWMPr::plot_metab	Plot ecosystem metabolism for a swmpr object
SWMPr::plot_summary	Plot graphical summaries of SWMP data
SWMPr::smoother	Smooth swmpr data

How are data ***retrieved***?

SWMPr functions can be used to import data into R three ways

1. Import from a local path
2. Retrieve SWMP data from a third-party server
3. Call the existing CDMO web services to import directly

Multiple options to accommodate different types of users

Bridges the gap from CDMO to analysis software

How are data *retrieved*?

The end result is the same - data are imported as a **swmpr** data object

```
> dat <- import_remote('kacsswq')
> class(dat)

## [1] "swmpr"      "data.frame"

> head(dat, 1)

##   datetimestamp temp spcond sal do_pct do_mgl depth cdepth
## 1    2004-01-01    2     42  26    101     12   0.7     NA
##   level clevel ph turb chlfluor
## 1    NA     NA  8    6       NA

> names(attributes(dat))

## [1] "names"       "row.names"   "class"       "station"
## [5] "parameters"  "qaqc_cols"   "date_rng"    "timezone"
## [9] "stamp_class"
```

How are data *retrieved*?

The remaining functions have `swmpr` methods

```
> methods(class = 'swmpr')
```

```
##  [1] aggremetab   aggreswmp    comb         decomp
##  [5] decomp_cj    ecometab     hist         lines
##  [9] na.approx    plot         plot_metab   plot_summary
## [13] qaqc         qaqcchk      rem_reps     setstep
## [17] smoother     subset
## see '?methods' for accessing help and source code
```

These are functions that were written for, and work specifically, with `swmpr` objects

`swmpr` objects can also use methods from the basic data frame class, i.e., you can exit the SWMPr workflow at any time

How are data ***organized***?

Data organization depends on the analysis needs - it is usually tedious

Example: Filter by QAQC flags

- Remove observations with a specified QAQC flag value
- Remove QAQC columns: Link to QAQC codes

-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

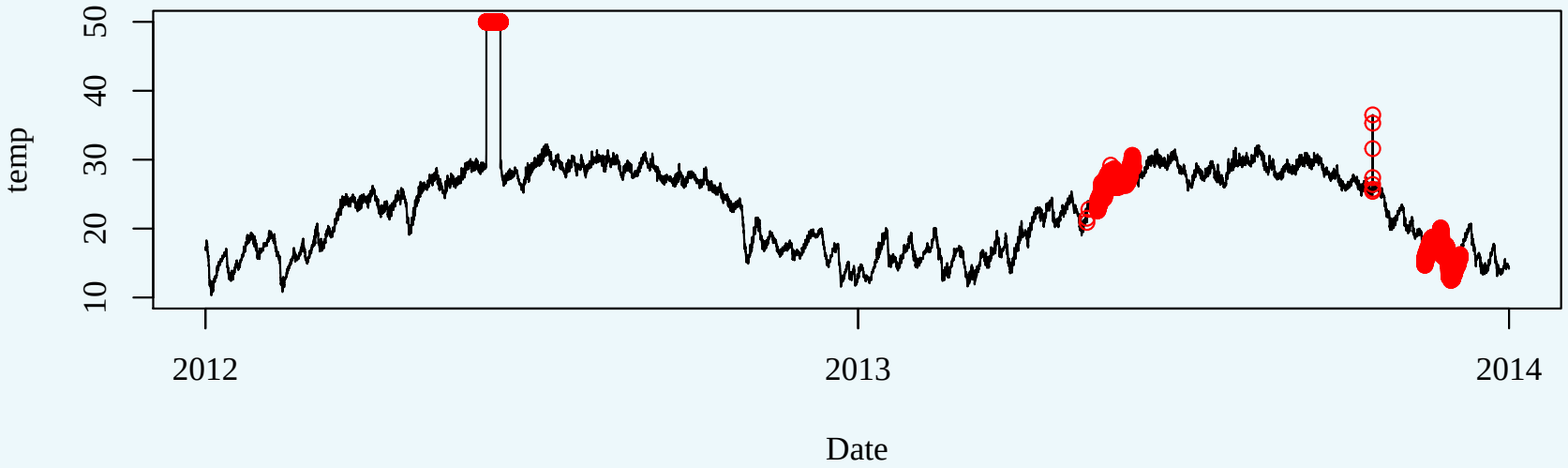
Retrieve SWMP data

Raw data with QAQC columns

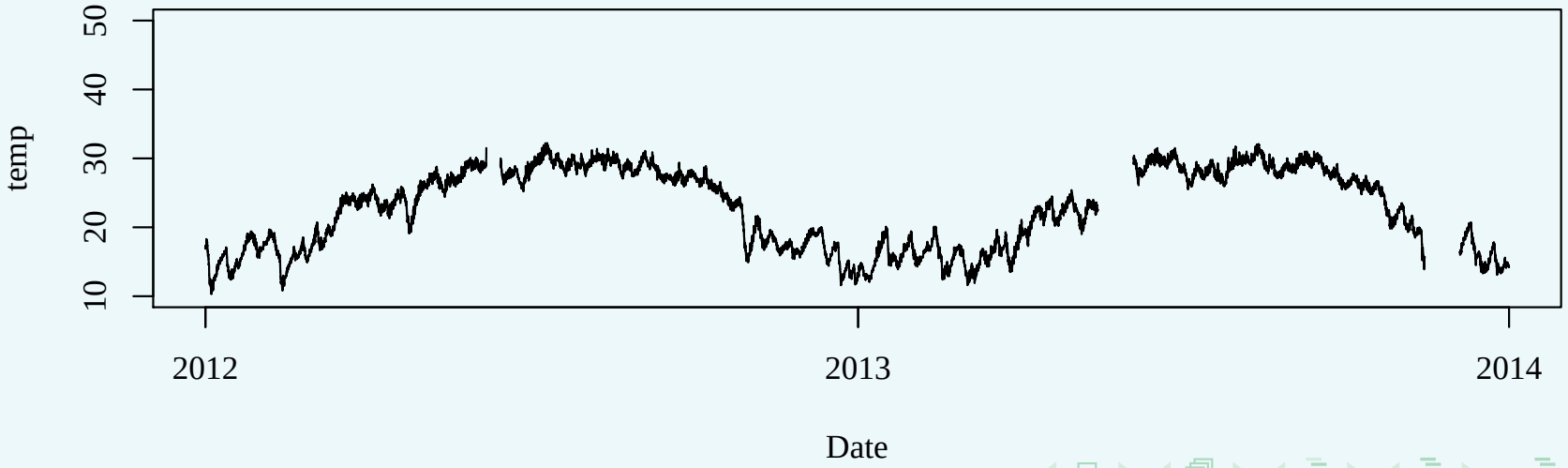
```
##          datetimestamp temp f_temp spcond f_spcond sal f_sal
## 1 2012-01-01 00:00:00   17    <0>      46      <0>   30  <0>
## 2 2012-01-01 00:15:00   17    <0>      46      <0>   30  <0>
## 3 2012-01-01 00:30:00   17    <0>      45      <0>   29  <0>
##   do_pct f_do_pct do_mgl f_do_mgl depth f_depth cdepth
## 1     89      <0>      7      <0>     1    <0>      1
## 2     88      <0>      7      <0>     1    <0>      1
## 3     89      <0>      7      <0>     1    <0>      1
##   f_cdepth level f_level clevel f_clevel ph f_ph turb
## 1      <3>    NA    <-1>      NA       NA  8 <0>    2
## 2      <3>    NA    <-1>      NA       NA  8 <0>    4
## 3      <3>    NA    <-1>      NA       NA  8 <0>    4
##   f_turb chlfluor f_chlfluor
## 1    <0>       NA      <-1>
## 2    <0>       NA      <-1>
## 3    <0>       NA      <-1>
```

Organize SWMP data

Data in red are 'bad' QAQC flags



After using `qaqc` function



How are data ***organized***?

Example: Combine data to compare time series from different sites

- Data may have arbitrary time steps that do not match between sites
- Date ranges may also differ

The `comb` function addresses these issues!

```
> # import all weather and wq data for Apalachicola
> met <- import_remote('apaebmet')
> wq <- import_remote('apacpwq')
```

How are data ***organized***?

```
> dim(met)

## [1] 490847     11

> dim(wq)

## [1] 455808     13

> # standardize time step to two hours
> # combine only overlapping time ranges
> dat <- comb(wq, met, timestep = 120, method = 'intersect')
> dim(dat)

## [1] 56977    23
```

How are data ***organized***?

The combined dataset

```
##          datetimestamp atemp rh   bp wspd maxwspd wdir
## 1 2001-12-31 23:00:00     4 69 1017    4       NA  347
## 2 2002-01-01 01:00:00     3 75 1017    3       NA    9
## 3 2002-01-01 03:00:00     2 77 1018    3       NA  331
## 4 2002-01-01 05:00:00     1 82 1019    4       NA    0
##   sdwdir totpar totprcp totsorad temp spcond sal do_pct
## 1     NA      0      NA        NA   NA     NA  NA     NA
## 2     NA      0      NA        NA   12     37  24    104
## 3     NA      0      NA        NA   12     40  26     99
## 4     NA      0      NA        NA   11     42  26     98
##   do_mgl depth cdepth level clevel ph turb chlfluor
## 1     NA    NA     NA    NA     NA NA   NA       NA
## 2     10     2     NA    NA     NA NA    3       NA
## 3      9     2     NA    NA     NA NA    4       NA
## 4      9     2     NA    NA     NA NA    5       NA
```

How are data ***analyzed***?

Time series analysis can range from very general to very specific

SWMPr functions include...

General

- Approximate missing data
- Smoothing with moving windows
- Aggregate by time periods
- Basic plots and histograms

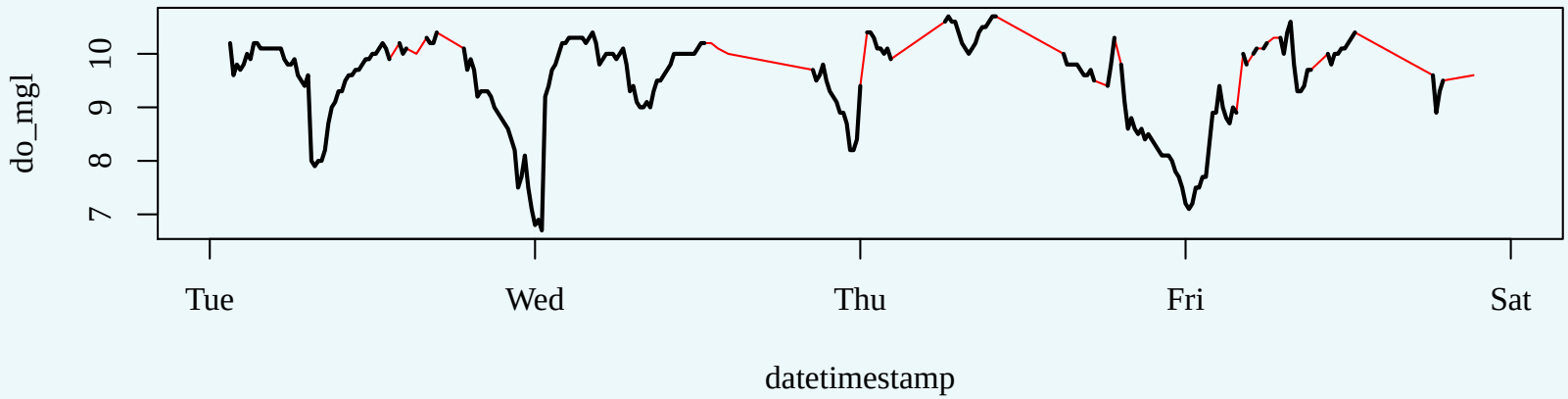
Specific

- Time series decomposition
- Estimate net ecosystem metabolism
- Aggregate metabolism
- Summary plots of raw data

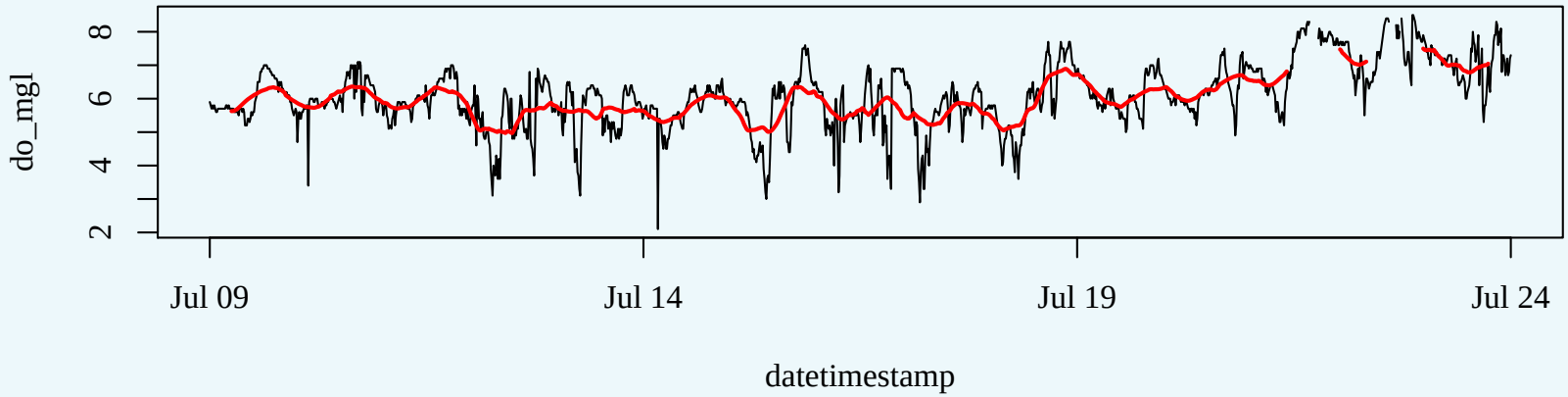
...or exit the SWMPr workflow and evaluate with other R packages

How are data ***analyzed***?

Example: fill missing data

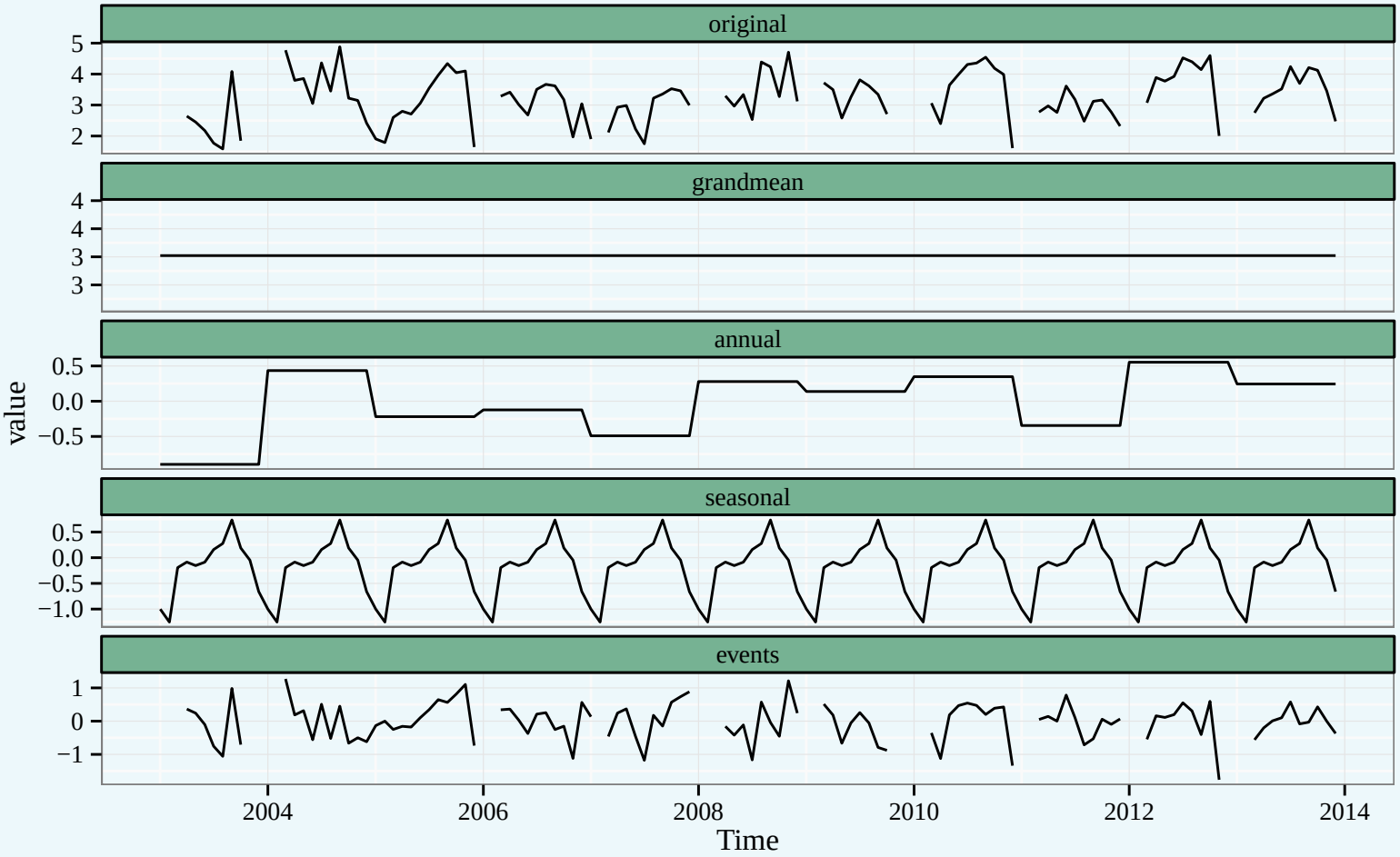


Example: smooth data



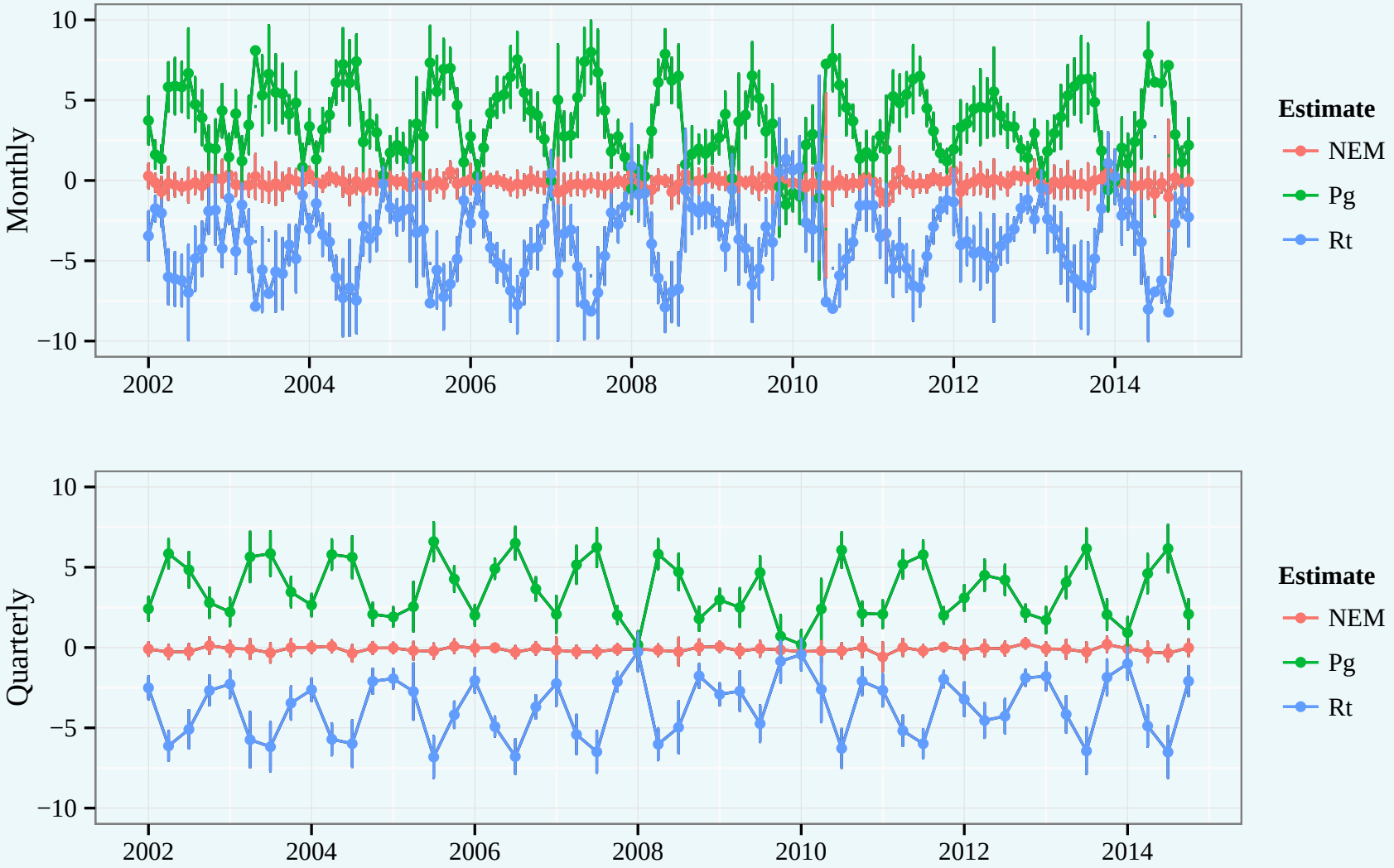
How are data ***analyzed***?

Example: time series decomposition (chl-a at cbmocnut)



How are data *analyzed*?

Example: estimate ecosystem metabolism



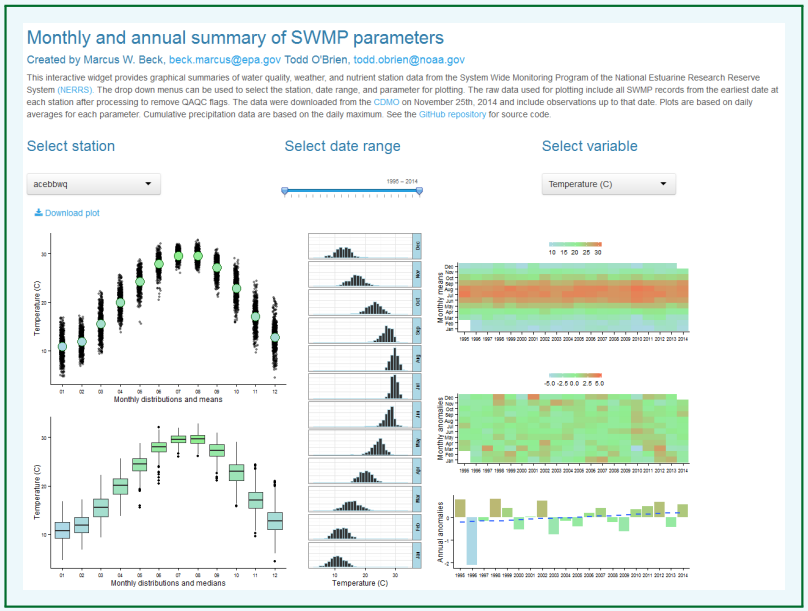
SWMPr applications

The most common question - what is the change over time at my site?

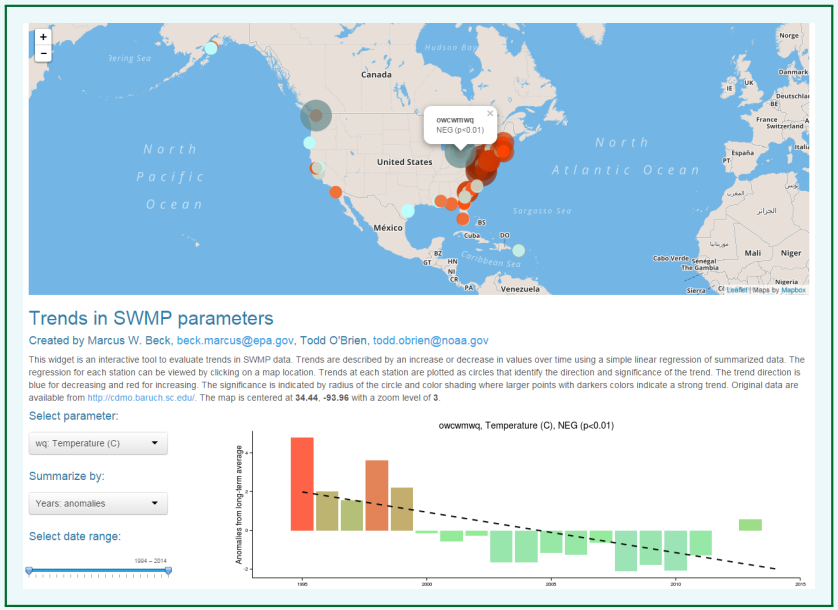
The functions in SWMPr can help, but it's easier to interact!

Two online applications can help visualize trends

Summary plots

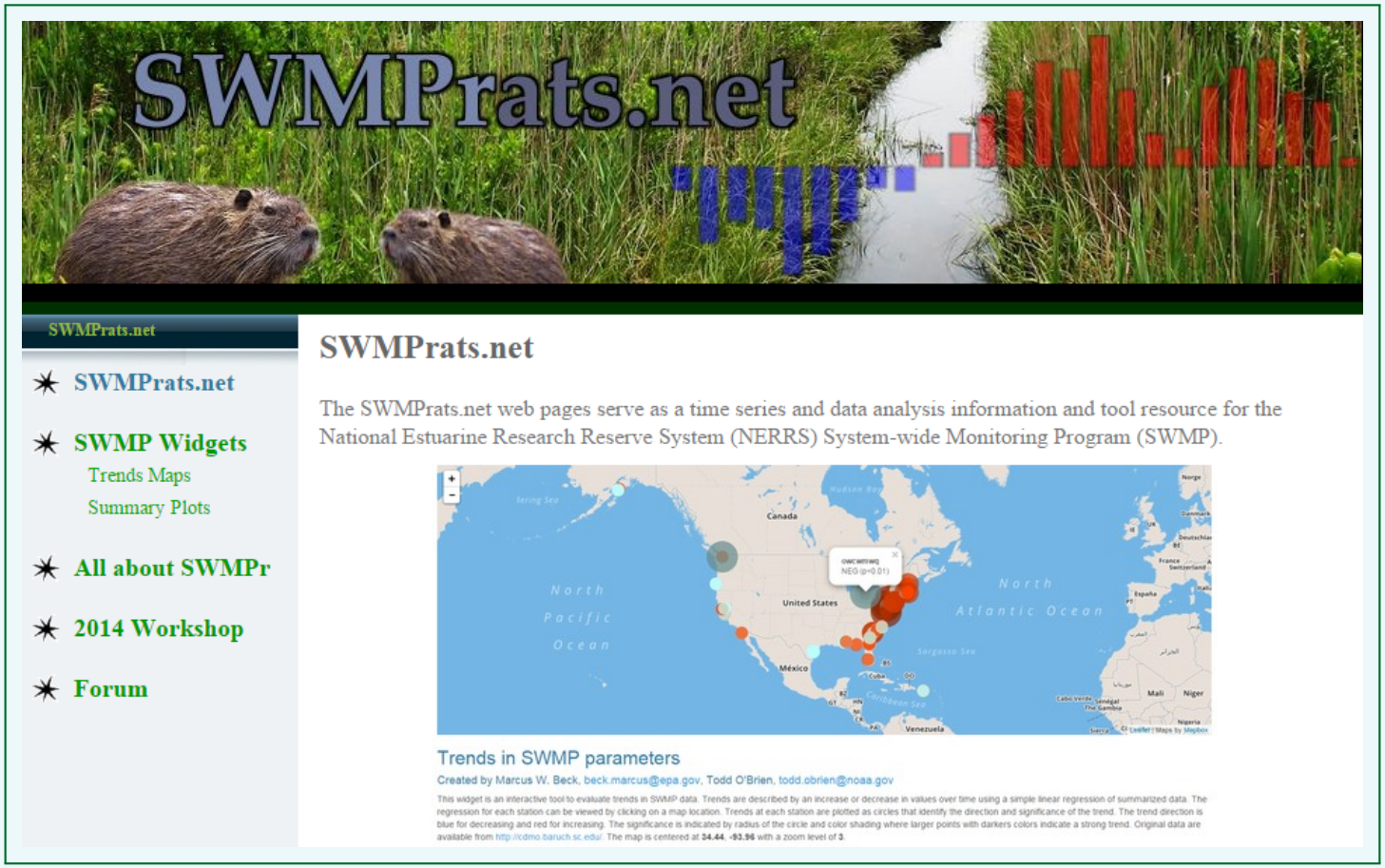


Trends map



SWMPr applications

SWMPrats.net: ***S***ystem-***W***ide ***M***onitoring ***P***rogram ***R***esources for the ***A***nalysis of ***T***ime ***S***eries



SWMPr applications

The SWMPr package provides an R-centric approach to ***retrieve***, ***organize***, and ***analyze*** estuary data

A new program, but already seeing heavy use:

- SWMPr downloaded 359 times from R network (as of May 7)
- Apps used 347 hours in April

SWMPr is meant to ***augment***, not replace, existing data management programs (i.e., CDMO web services)

Deals with lots of the heavy lifting with large, unrefined datasets - greatly improves accessibility for analysis