Time series topic 1: Weighted regression

Marcus W. Beck[1]

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

Objectives for the session (2:00 - 3:00)

- What is weighted regression
- The WRTDStidal package
- Application to NERRS data
  - Fitting a model
  - Evaluating a model
  - Viewing a model

Interactive portion

Follow along as we go:

- flash drive
- online: swmprats.net 2016 workshop tab

Interactive portion

Follow along as we go:

- flash drive
- online: swmprats.net 2016 workshop tab

You will run examples whenever you see this guy:

Is everything installed?

We will use the WRTDStidal package

Option 1, from the R Console prompt:

```
install.packages('WRTDStidal')
library(WRTDStidal)
```

Is everything installed?

We will use the WRTDStidal package

Option 1, from the R Console prompt:

```
install.packages('WRTDStidal')
library(WRTDStidal)
```

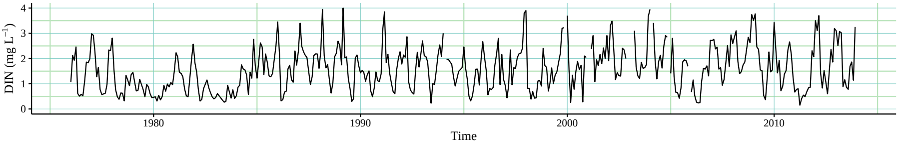
Option 2, install the source file from the flash drive:

```
# change as needed
path_to_file <- 'C:/Users/mbeck/Desktop/WRTDStidal_1.0.1.tar.gz'

# install, load
install.packages(path_to_file, repos = NULL, type="source")
library(WRTDStidal)
```

Model theory and background - what is WRTDS

Observed data represents effects of many processes



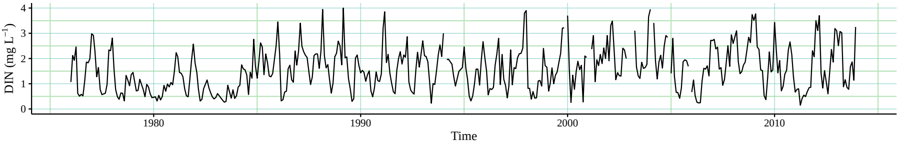
Climate
precipitation
temperature
wind events
ENSO effects

Local
light/turbidity
residence time
invasive species
trophic effects

Regional/historical
watershed inputs
point sources
management actions
flow changes

Model theory and background - what is WRTDS

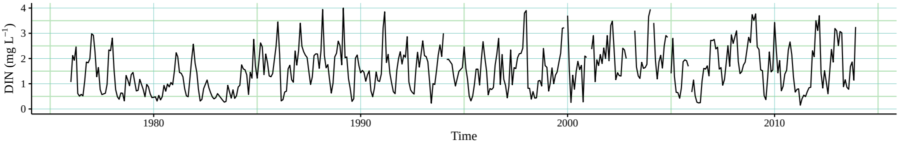
Observed data represents effects of many processes



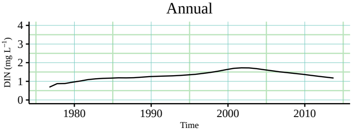
Models should describe components to evaluate effects

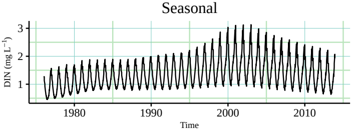
Model theory and background - what is WRTDS

Observed data represents effects of many processes



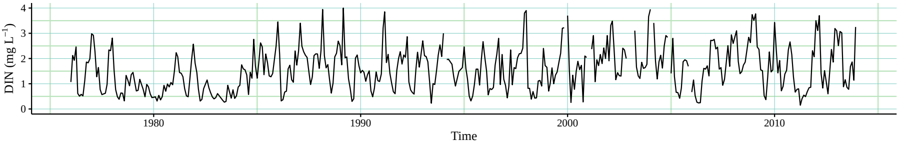
Models should describe components to evaluate effects



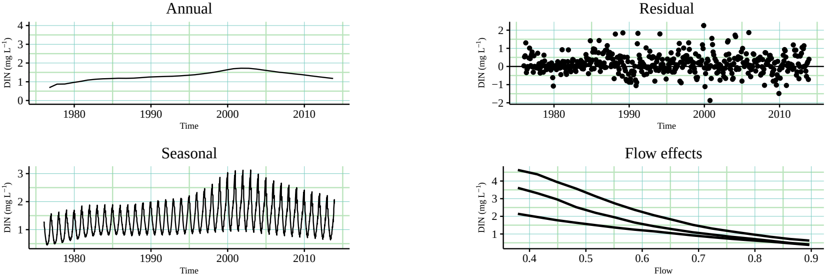


Model theory and background - what is WRTDS

Observed data represents effects of many processes



Models should describe components to evaluate effects



Model theory and background - what is WRTDS

Weighted **R**egression on **T**ime, **D**ischarge, and **S**eason

- Describes a time series in the context of these parameters, locally fitted
- Useful to describe long-term trends, ie., multi-decadal time series
- Evaluation of flow-normalized trends, hypothesis generation

Model theory and background - what is WRTDS

Weighted **R**egression on **T**ime, **D**ischarge, and **S**eason

- Describes a time series in the context of these parameters, locally fitted
- Useful to describe long-term trends, ie., multi-decadal time series
- Evaluation of flow-normalized trends, hypothesis generation

Developed by [Hirsch et al., 2010] for pollutants in stream/rivers

Adapted for tidal waters by [Beck and Hagy III, 2015]

Model theory and background - what is WRTDS

How does it work?

$$\ln(N) = \beta_0 + \beta_1 t + \beta_2 Sal + \beta_3 \sin(2\pi t) + \beta_4 \cos(2\pi t)$$

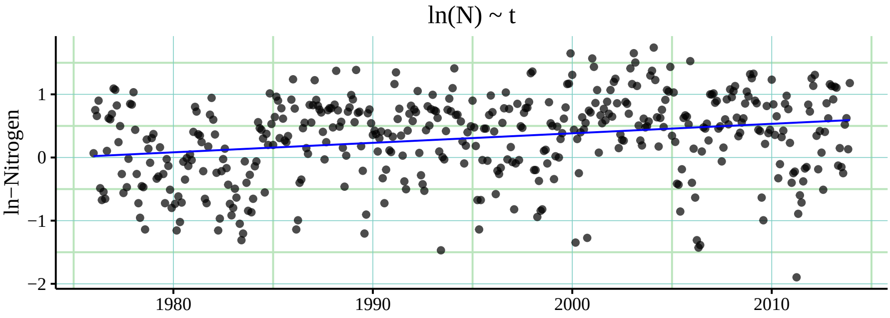
N: nitrogen (or other response endpoint)
t: time
Sal: Salinity (or other flow-related variable)

Model theory and background - what is WRTDS

How does it work?

$$\ln(N) = \beta_0 + \beta_1 t + \beta_2 Sal + \beta_3 \sin(2\pi t) + \beta_4 \cos(2\pi t)$$

N: nitrogen (or other response endpoint)
t: time
Sal: Salinity (or other flow-related variable)

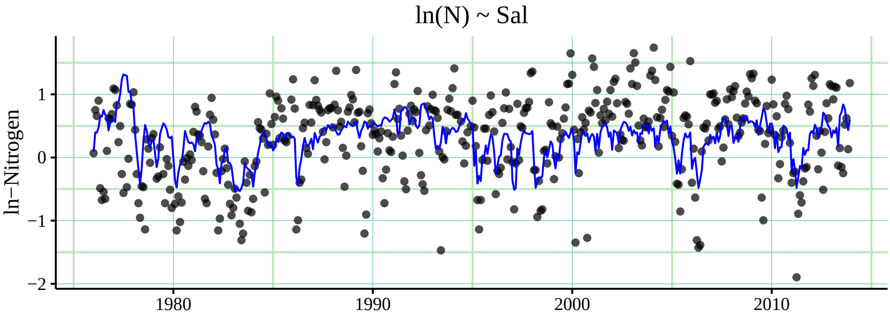


Model theory and background - what is WRTDS

How does it work?

$$\ln(N) = \beta_0 + \beta_1 t + \beta_2 Sal + \beta_3 \sin(2\pi t) + \beta_4 \cos(2\pi t)$$

N: nitrogen (or other response endpoint)
t: time
Sal: Salinity (or other flow-related variable)

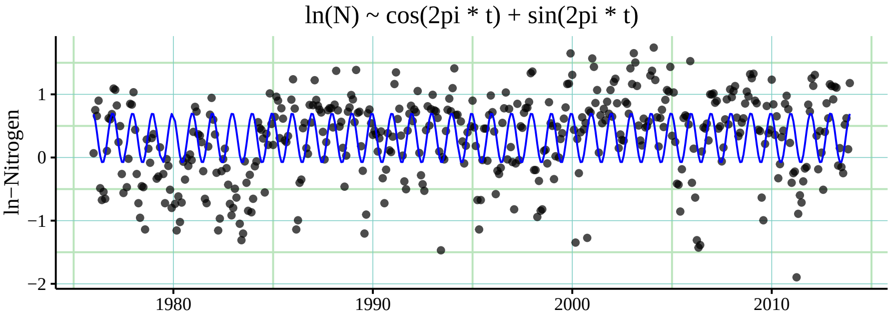


Model theory and background - what is WRTDS

How does it work?

$$\ln(N) = \beta_0 + \beta_1 t + \beta_2 Sal + \beta_3 \sin(2\pi t) + \beta_4 \cos(2\pi t)$$

N: nitrogen (or other response endpoint)
t: time
Sal: Salinity (or other flow-related variable)

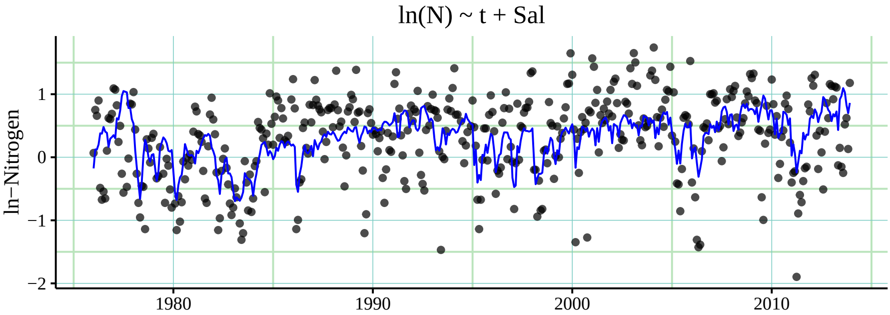


Model theory and background - what is WRTDS

How does it work?

$$\ln(N) = \beta_0 + \beta_1 t + \beta_2 Sal + \beta_3 \sin(2\pi t) + \beta_4 \cos(2\pi t)$$

N: nitrogen (or other response endpoint)
t: time
Sal: Salinity (or other flow-related variable)

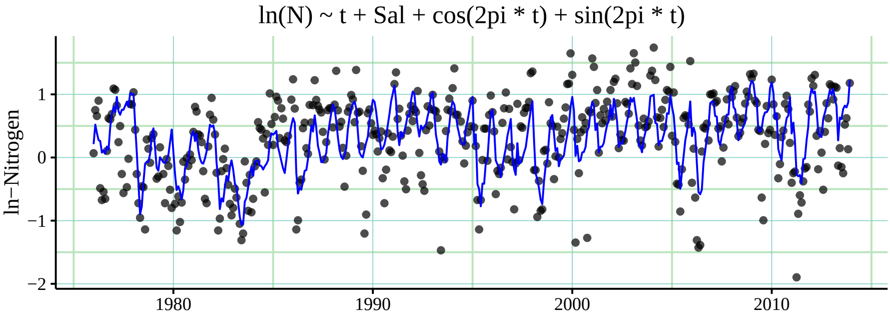


Model theory and background - what is WRTDS

How does it work?

$$\ln(N) = \beta_0 + \beta_1 t + \beta_2 Sal + \beta_3 \sin(2\pi t) + \beta_4 \cos(2\pi t)$$

N: nitrogen (or other response endpoint)
t: time
Sal: Salinity (or other flow-related variable)



Model theory and background - what is WRTDS

This is not the whole story...

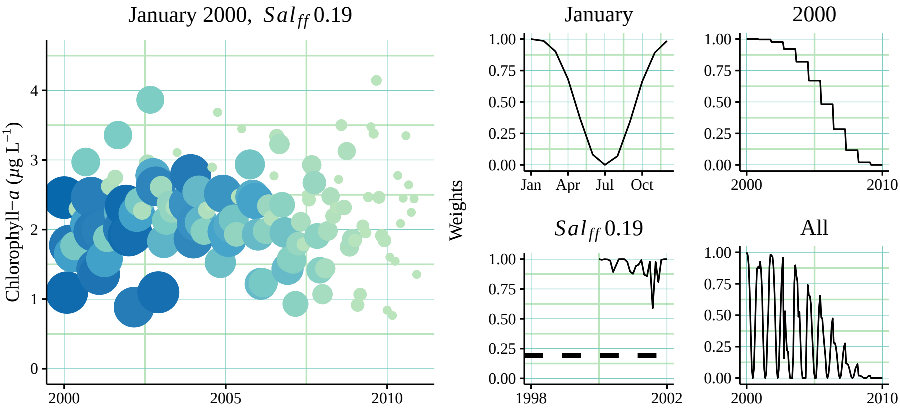
$$\ln(N) = \beta_0 + \beta_1 t + \beta_2 Sal + \beta_3 \sin(2\pi t) + \beta_4 \cos(2\pi t)$$

One parameter set to many parameter sets - a moving window regression

Within each window, a unique regression is fit, weighted by the local salinity, time, and season

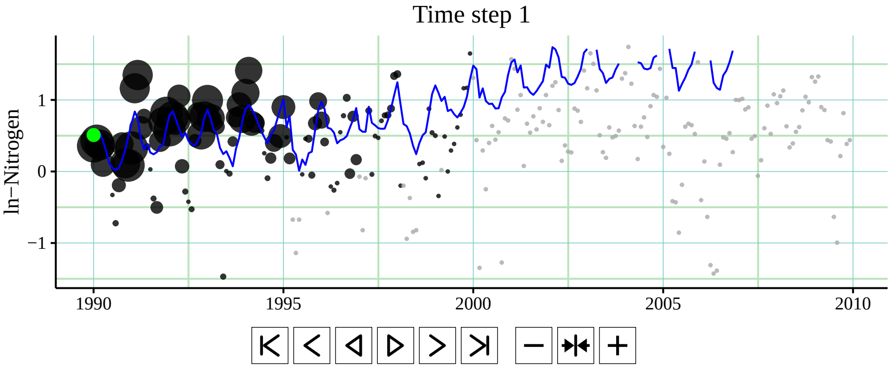
Similar to a loess/spline smooth but specific to the effects of these three variables on the response

Model theory and background - what is WRTDS



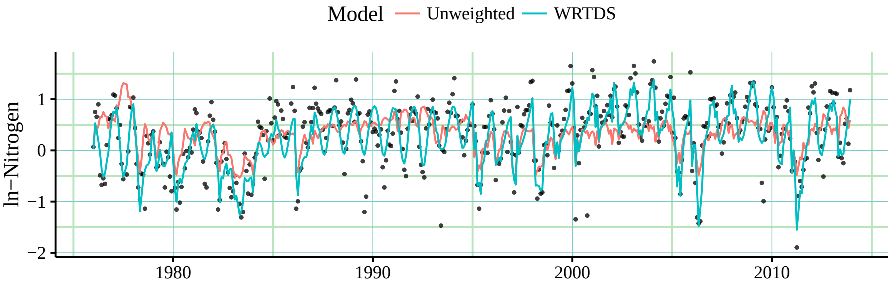
Model theory and background - what is WRTDS

Points: observed time series (black are weighted, grey is zero weight)
Green point: observation at the center of the regression
Blue line: Global model with weights specific to the window
Red line: Accumulated WRTDS model



Model theory and background - what is WRTDS

RMSE fit for unweighted = 0.58, WRTDS = 0.36



Model theory and background - what is WRTDS

All you need to know:

- Describe a response variable in relation to time, salinity (discharge), and season

Model theory and background - what is WRTDS

All you need to know:

- Describe a response variable in relation to time, salinity (discharge), and season
- A ‘moving window, weighted regression’, like a loess/spline smooth

Model theory and background - what is WRTDS

All you need to know:

- Describe a response variable in relation to time, salinity (discharge), and season
- A ‘moving window, weighted regression’, like a loess/spline smooth
- WRTDStidal package does all of it for you - including viz

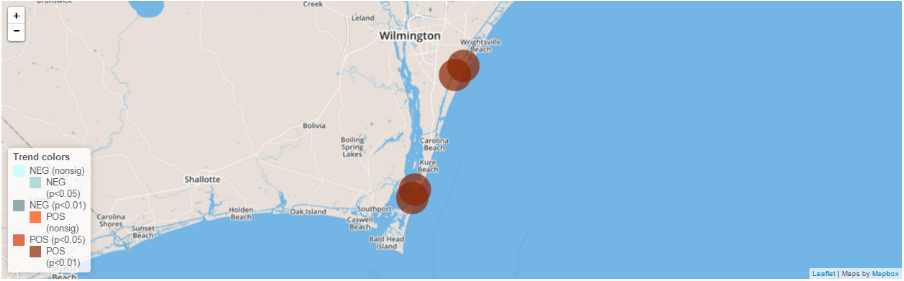
Model theory and background - what is WRTDS

All you need to know:

- Describe a response variable in relation to time, salinity (discharge), and season
- A 'moving window, weighted regression', like a loess/spline smooth
- WRTDStidal package does all of it for you - including viz

...let's not forget about flow-normalization, more about this later

Using WRTDS with NERRS data



Trends in SWMP parameters

Created by Marcus W. Beck, beck.marcus@epa.gov, Todd O'Brien, todd.obrien@noaa.gov

This widget is an interactive tool to explore trends in SWMP data. Trends are described by an increase or decrease in values over time using a simple linear regression of summarized data. The regression for each station can be viewed by clicking on a map location. Trends at each station are plotted as circles that identify the direction and significance of the trend. The trend direction is blue for decreasing and red for increasing. The significance is indicated by radius of the circle and color shading where larger points with darkers colors indicate a strong trend. Original data are available from http://cdmo.baruch.sc.edu/. See the GitHub repository for source code. The data include observations through December 2015 (if available) and are current as of May 31, 2016. Please note that the use of simple regression to identify trends is for exploratory purposes only and may not be appropriate for all datasets. The map is centered at **34.04**, **-77.86** with a zoom level of **10**.

Select parameter:

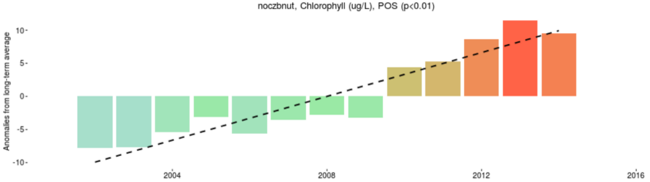
nut: Chlorophyll-a (ug/L)

Summarize by:

Years: anomalies

Select date range:

2002 – 2015



Using WRTDS with NERRS data

Aggregation of SWMP parameters within/between reserves

Created by Marcus W. Beck, beck.marcus@epa.gov Todd O'Brien, todd.obrien@noaa.gov

This interactive widget can be used to compare time series of site data within and between reserves from the System Wide Monitoring Program of the National Estuarine Research Reserve System (NERRS). Data are based on monthly averages of raw observations through December 2015 and are current as of May 31, 2016. Two plots are shown for selected parameters and reserves that include time series of all sites at each location. The monthly averages are shown by default. Data can also be viewed as quarterly (every three months) or annual aggregations based on averages of the monthly summaries. Tabular data for each plot can be viewed on the tables tab and downloads of the plots and tables are available on the downloads tab. See the GitHub repository for source code or to post issues if problems occur.

Aggregate by: year

Date range: 1995 — 2002 to 2014 — 2015 (1995 1997 1999 2001 2003 2005 2007 2009 2011 2013 2015)

Aesthetics: ☑ Lines ☑ Points

Y-axis: ☐ Common

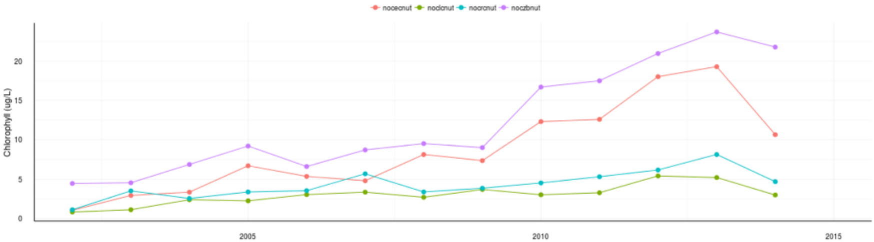
Plots | Tables | Downloads

Parameter type: Nutrients

Parameter: Chlorophyll (ug/L)

Reserve: noc

Stations: ☑ nocecnut ☑ noclcnut ☑ nocrcnut ☑ noczbnut



Using WRTDS with NERRS data

Using nutrient data from North Carolina NERR, Zeke's Basin site:

1. Import nutrient data, organize

Using WRTDS with NERRS data

Using nutrient data from North Carolina NERR, Zeke's Basin site:

1. Import nutrient data, organize
2. Import wq data, organize

Using WRTDS with NERRS data

Using nutrient data from North Carolina NERR, Zeke's Basin site:

1. Import nutrient data, organize
2. Import wq data, organize
3. Combine chlorophyll and salinity time series

Using WRTDS with NERRS data

Using nutrient data from North Carolina NERR, Zeke's Basin site:

1. Import nutrient data, organize
2. Import wq data, organize
3. Combine chlorophyll and salinity time series
4. Prep for WRTDS

Using WRTDS with NERRS data

Using nutrient data from North Carolina NERR, Zeke's Basin site:

1. Import nutrient data, organize
2. Import wq data, organize
3. Combine chlorophyll and salinity time series
4. Prep for WRTDS
5. Create WRTDS model

Using WRTDS with NERRS data

Using nutrient data from North Carolina NERR, Zeke's Basin site:

1. Import nutrient data, organize
2. Import wq data, organize
3. Combine chlorophyll and salinity time series
4. Prep for WRTDS
5. Create WRTDS model
6. Assess model performance, plot results

Using WRTDS with NERRS data

1. Import nutrient data, organize

Using WRTDS with NERRS data

1. Import nutrient data, organize

```
# load SWMPr, nutrient data
library(SWMPr)
load(file = 'data/noczbnut.RData')

# rename, qaqc clean up, subset
nut <- noczbnut
nut <- qaqc(nut, qaqc_keep = c(0, 4))
nut <- subset(nut, select = 'chla_n')
head(nut)
```

```
##          datetimestamp chla_n
## 1 2002-04-23 15:35:00   2.12
## 2 2002-05-24 09:20:00   1.60
## 3 2002-06-24 10:35:00   3.47
## 4 2002-07-24 09:40:00   4.43
## 5 2002-08-26 11:31:00   4.65
## 6 2002-09-24 10:40:00   5.95
```

Using WRTDS with NERRS data

2. Import wq data, organize

Using WRTDS with NERRS data

2. Import wq data, organize

```
# load wq data
load(file = 'data/noczbwq.RData')

# rename, qaqc clean up, subset
wq <- noczbwq
wq <- qaqc(wq, qaqc_keep = c(0, 4))
wq <- subset(wq, select = 'sal')
head(wq)

##           datetimestamp  sal
## 1 2002-03-01 12:30:00 26.8
## 2 2002-03-01 13:00:00 26.8
## 3 2002-03-01 13:30:00 26.7
## 4 2002-03-01 14:00:00 26.7
## 5 2002-03-01 14:30:00 26.6
## 6 2002-03-01 15:00:00 26.6
```

Using WRTDS with NERRS data

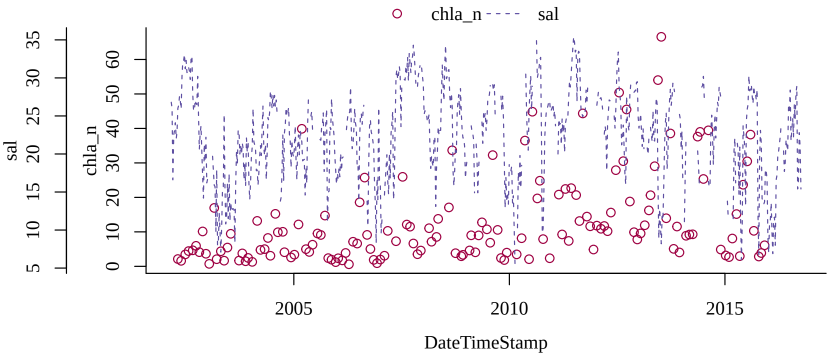
3. Combine chlorophyll and salinity time series

Using WRTDS with NERRS data

3. Combine chlorophyll and salinity time series

```
# combine at weekly time step
tomod <- comb(nut, wq, timestep = 'weeks')

# plot both
overplot(tomod, type = c('p', 'l'))
```



Using WRTDS with NERRS data

4. Prep for WRTDS - requires a tidalmean object with four columns (date, response, salinity/flow, detection limit)

Using WRTDS with NERRS data

4. Prep for WRTDS - requires a tidalmean object with four columns (date, response, salinity/flow, detection limit)

```
library(WRTDStidal)

# add arbitrary limit column, datetimestamp as date
tomod$lim <- -1e6
tomod$datetimestamp <- as.Date(tomod$datetimestamp)

# create tidalmean object, note if response is in log or not
tomod <- tidalmean(tomod, reslog = FALSE)
head(tomod)


##         date res       flo    lim not_cens   day_num month year dec_time
## 1 2002-02-28  NA 0.7123746 -1e+06       NA 0.1643836     2 2002 2002.164
## 2 2002-03-07  NA 0.6856187 -1e+06       NA 0.1835616     3 2002 2002.184
## 3 2002-03-14  NA 0.3712375 -1e+06       NA 0.2027397     3 2002 2002.203
## 4 2002-03-21  NA 0.5685619 -1e+06       NA 0.2219178     3 2002 2002.222
## 5 2002-03-28  NA 0.5886288 -1e+06       NA 0.2410959     3 2002 2002.241
## 6 2002-04-04  NA 0.6220736 -1e+06       NA 0.2602740     4 2002 2002.260
```

Using WRTDS with NERRS data

5. Create WRTDS model

Using WRTDS with NERRS data

5. Create WRTDS model

```
# use modfit function
mod <- modfit(tomod)


##
## Estimating interpolation grid for mean response, % complete...
##
## 5  10  15  20  25  30  35  40  45  50  55  60  65  70  75  80  85  90  95  100
##
## Estimating predictions...
##
##
## Normalizing predictions...
```

Using WRTDS with NERRS data

5. Create WRTDS model

```
# use modfit function
mod <- modfit(tomod)

##
## Estimating interpolation grid for mean response, % complete...
##
## 5  10  15  20  25  30  35  40  45  50  55  60  65  70  75  80  85  90  95  100
##
## Estimating predictions...
##
##
## Normalizing predictions...
```

6. Assess performance

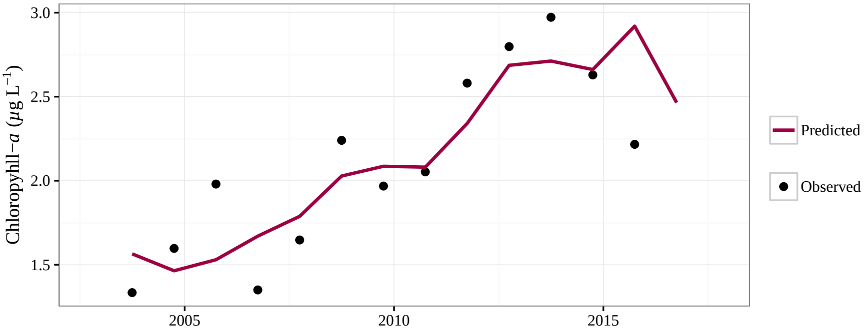
```
wrtdsperf(mod)

##        rmse      nmse
## 1 0.6417556 0.3476415
```

Using WRTDS with NERRS data

6. Plot results: fitplot shows observed, predicted, averaged by water years

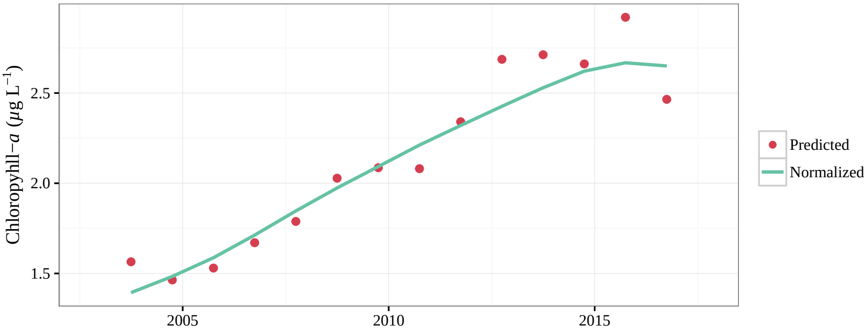
```
fitplot(mod)
```



Using WRTDS with NERRS data

6. Plot results: prdnrmplot shows predicted, flow-normalized predictions, averaged by year

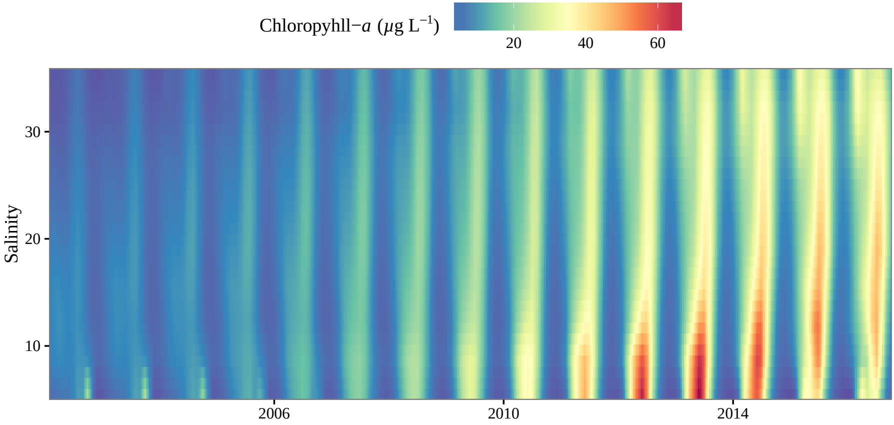
```
prdnrmplot(mod)
```



Using WRTDS with NERRS data

6. Plot results: gridplot shows how the flow, chlorophyll response has changed by season, year

```
gridplot(mod, logspace = F, month = 'all', floscl = F)
```



WRTDS summary

- WRTDS has value for describing decadal trends in the unique context of time, discharge (salinity), and season

WRTDS summary

- WRTDS has value for describing decadal trends in the unique context of time, discharge (salinity), and season
- Flow-normalized results provide a more ‘genuine’ description of the trend

WRTDS summary

- WRTDS has value for describing decadal trends in the unique context of time, discharge (salinity), and season
- Flow-normalized results provide a more ‘genuine’ description of the trend
- Just have a general idea of the theory, WRTDStidal package does the heavy lifting for you

WRTDS summary

- WRTDS has value for describing decadal trends in the unique context of time, discharge (salinity), and season
- Flow-normalized results provide a more ‘genuine’ description of the trend
- Just have a general idea of the theory, WRTDStidal package does the heavy lifting for you
- Lots shown here, the code can be used step-by-step with other stations

WRTDS summary

- WRTDS has value for describing decadal trends in the unique context of time, discharge (salinity), and season
- Flow-normalized results provide a more ‘genuine’ description of the trend
- Just have a general idea of the theory, WRTDStidal package does the heavy lifting for you
- Lots shown here, the code can be used step-by-step with other stations

Checkout the references and the website for more info:
https://github.com/fawda123/WRTDStidal

Up next... Time Series Topic 2: Decomposition

Questions??

References

Beck MW, Hagy III JD. 2015.
Adaptation of a weighted regression approach to evaluate water quality trends in an estuary.
Environmental Modelling and Assessment, 20(6):637–655.

Hirsch RM, Moyer DL, Archfield SA. 2010.
Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs.
Journal of the American Water Resources Association, 46(5):857–880.