Time series topic 3: Seasonal Kendall Tests

Marcus W. Beck[1]

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

Objectives for the session (4:15 - 5:00)

- What is and why would we use a Kendall test
- What is a why would we use a *Seasonal* Kendall test
- Application with NERRS data
  - Data prep
  - Execution
  - Interpretation

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 functions in the EnvStats package

Option 1, from the R Console prompt:

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

Is everything installed?

We will use functions in the EnvStats package

Option 1, from the R Console prompt:

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

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

```
# change as needed
path_to_file <- 'C:/Users/mbeck/Desktop/EnvStats_2.1.1.tar.gz'

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

Is everything installed?

We will use functions in the lubridate package

Option 1, from the R Console prompt:

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

Is everything installed?

We will use functions in the lubridate package

Option 1, from the R Console prompt:

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

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

```
# change as needed
path_to_file <- 'C:/Users/mbeck/Desktop/lubridate_1.6.0.tar.gz'

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

Theory and background

Use these tests to answer the question:

Is there a ***monotonic trend*** and what is the ***significance***?

Theory and background

Use these tests to answer the question:

Is there a ***monotonic trend*** and what is the ***significance***?

Both tests:

- Are non-parametric (no distribution assumptions, based on signs)

Theory and background

Use these tests to answer the question:

Is there a ***monotonic trend*** and what is the ***significance***?

Both tests:

- Are non-parametric (no distribution assumptions, based on signs)
- Work with a user-defined time interval

Theory and background

Use these tests to answer the question:

Is there a ***monotonic trend*** and what is the ***significance***?

Both tests:

- Are non-parametric (no distribution assumptions, based on signs)
- Work with a user-defined time interval
- Provide a p-value indicating the probability due to chance alone

Theory and background

Use these tests to answer the question:

Is there a ***monotonic trend*** and what is the ***significance***?

Both tests:

- Are non-parametric (no distribution assumptions, based on signs)
- Work with a user-defined time interval
- Provide a p-value indicating the probability due to chance alone
- Provide a direction of the trend as τ ('tau')

Theory and background

Use these tests to answer the question:

Is there a ***monotonic trend*** and what is the ***significance***?

Both tests:

- Are non-parametric (no distribution assumptions, based on signs)
- Work with a user-defined time interval
- Provide a p-value indicating the probability due to chance alone
- Provide a direction of the trend as τ ('tau')
- Provide a slope as the rate of change

Theory and background

The ***Kendall test*** for time series:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign\left[\left(X_j - X_i\right)\left(Y_j - Y_i\right)\right]$$

$$\hat{\tau} = \frac{2S}{n\left(n-1\right)}$$

$\hat{\tau}$ will vary from -1 to 1 similar to a correlation coefficient, follows an approximate normal-distribution for hypothesis-testing

Theory and background

The ***Kendall test*** for time series:

$$\hat{\beta}_1 = Median\left(\frac{Y_j - Y_i}{X_j - X_i}\right), i < j$$

$\hat{\beta}_1$ is the Theil (Sen) non-parametric estimate of slope or the rate of change in the interval

Theory and background

The ***Kendall test*** for time series:

$$\hat{\beta}_1 = Median\left(\frac{Y_j - Y_i}{X_j - X_i}\right), i < j$$

$\hat{\beta}_1$ is the Theil (Sen) non-parametric estimate of slope or the rate of change in the interval

All you need to know:

- $\hat{\tau}$ is direction and magnitude of trend
- $\hat{\beta}_1$ is the linear rate of change

Theory and background

The ***Seasonal Kendall test*** is exactly the same...
...except separate tests by month across years (January 1981, 1982, ..., February 1981, 1982, ...), results are pooled.

Theory and background

The ***Seasonal Kendall test*** is exactly the same...
...except separate tests by month across years (January 1981, 1982, ..., February 1981, 1982, ...), results are pooled.

- Overall $\hat{\tau}$ is the weighted average of the seasonal estimates

Theory and background

The ***Seasonal Kendall test*** is exactly the same...
...except separate tests by month across years (January 1981, 1982, ..., February 1981, 1982, ...), results are pooled.

- Overall $\hat{\tau}$ is the weighted average of the seasonal estimates
- Overall $\hat{\beta}_1$ is the median of all two-point slope estimates within each season

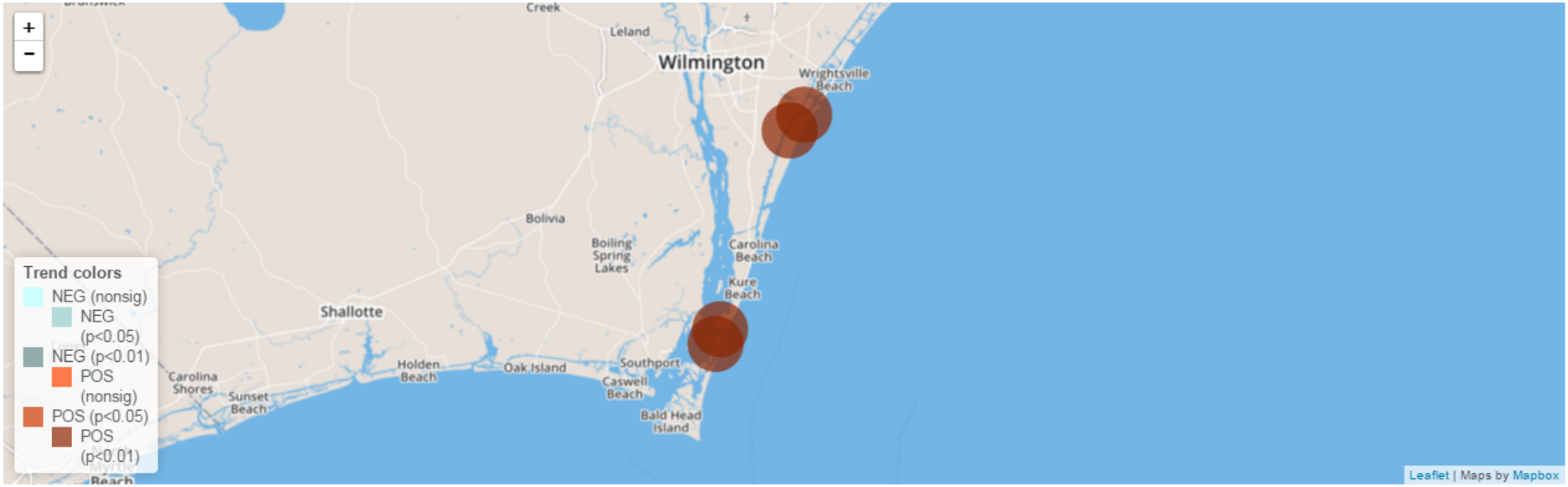
Theory and background

The ***Seasonal Kendall test*** is exactly the same...
...except separate tests by month across years (January 1981, 1982, ..., February 1981, 1982, ...), results are pooled.

- Overall $\hat{\tau}$ is the weighted average of the seasonal estimates
- Overall $\hat{\beta}_1$ is the median of all two-point slope estimates within each season

More info in help documentation for `kendallTrendTest`, `kendallSeasonalTrendTest` in EnvStats,
[Hirsch et al., 1982, Millard, 2013]

Using `kendallTrendTest` 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:

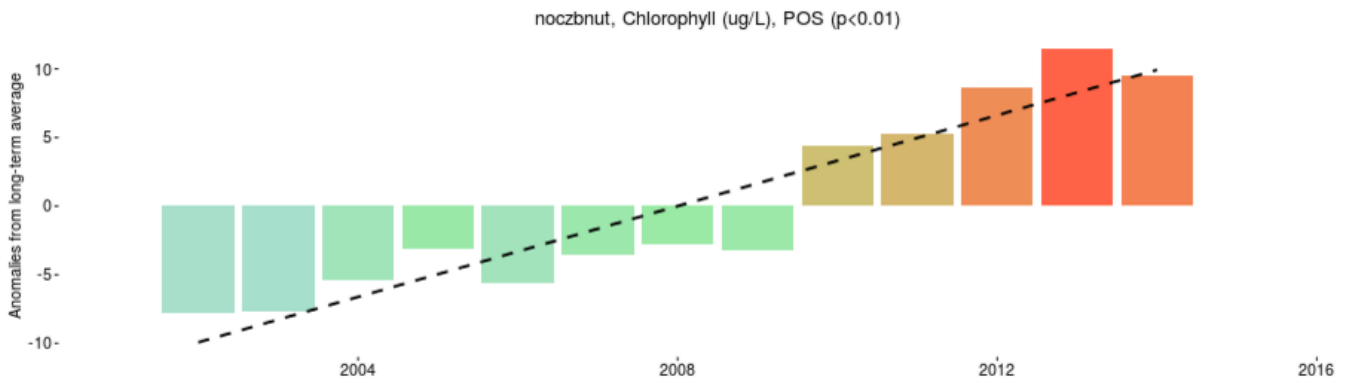
nut: Chlorophyll-a (ug/L)

Summarize by:

Years: anomalies

Select date range:

2002 — 2015



Using `kendallTrendTest` 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 2014 (1995 1997 1999 2001 2003 2005 2007 2009 2011 2013 2015)

Aesthetics: Lines, Points

Y-axis: Common

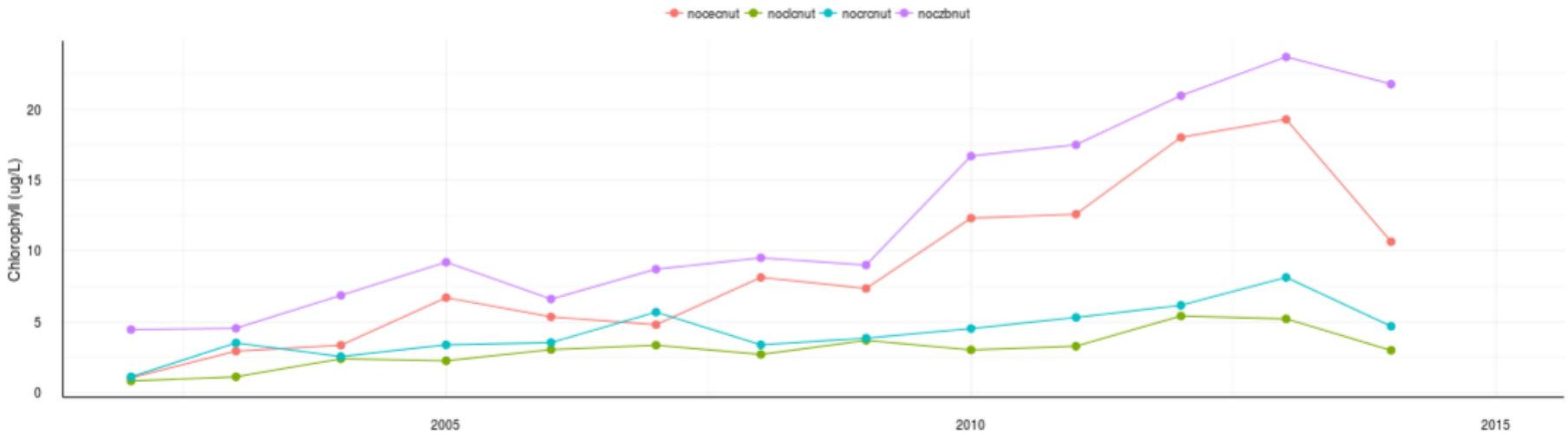
Plots | Tables | Downloads

Parameter type: Nutrients

Parameter: Chlorophyll (ug/L)

Reserve: noc

Stations: nocecnut, noclcnut, nocrcnut, noczbnut



Using `kendallTrendTest` with NERRS data

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

1. Import nutrient data, prep

Using `kendallTrendTest` with NERRS data

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

1. Import nutrient data, prep
2. Test with `kendallTrendTest`

Using `kendallTrendTest` with NERRS data

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

1. Import nutrient data, prep
2. Test with `kendallTrendTest`
3. Interpret results

Using `kendallTrendTest` with NERRS data

1. Import nutrient data, prep

Using kendallTrendTest with NERRS data

1. Import nutrient data, prep

```
# 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')
nut <- na.omit(nut)
head(nut)
```

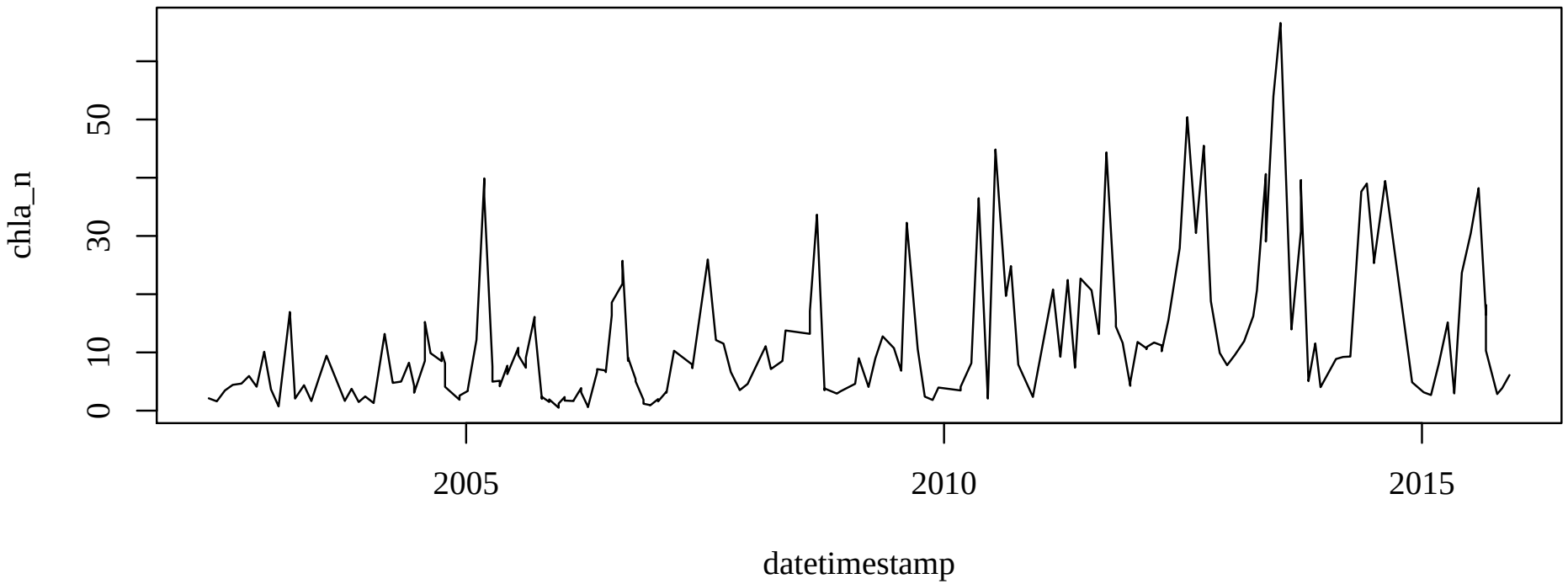
```
##          datetimestamp chla_n
## 1 2002-04-23 15:35:00   2.12
## 2 2002-05-24 09:20:00   1.60
## 3 2002-06-24 10:35:00   3.47
## 4 2002-07-24 09:40:00   4.43
## 5 2002-08-26 11:31:00   4.65
## 6 2002-09-24 10:40:00   5.95
```

Using `kendallTrendTest` with NERRS data

1. Import nutrient data, prep

First look at a plot:

```
plot(nut)
```



Using kendallTrendTest with NERRS data

2. Test with `kendallTrendTest`

Run the test:

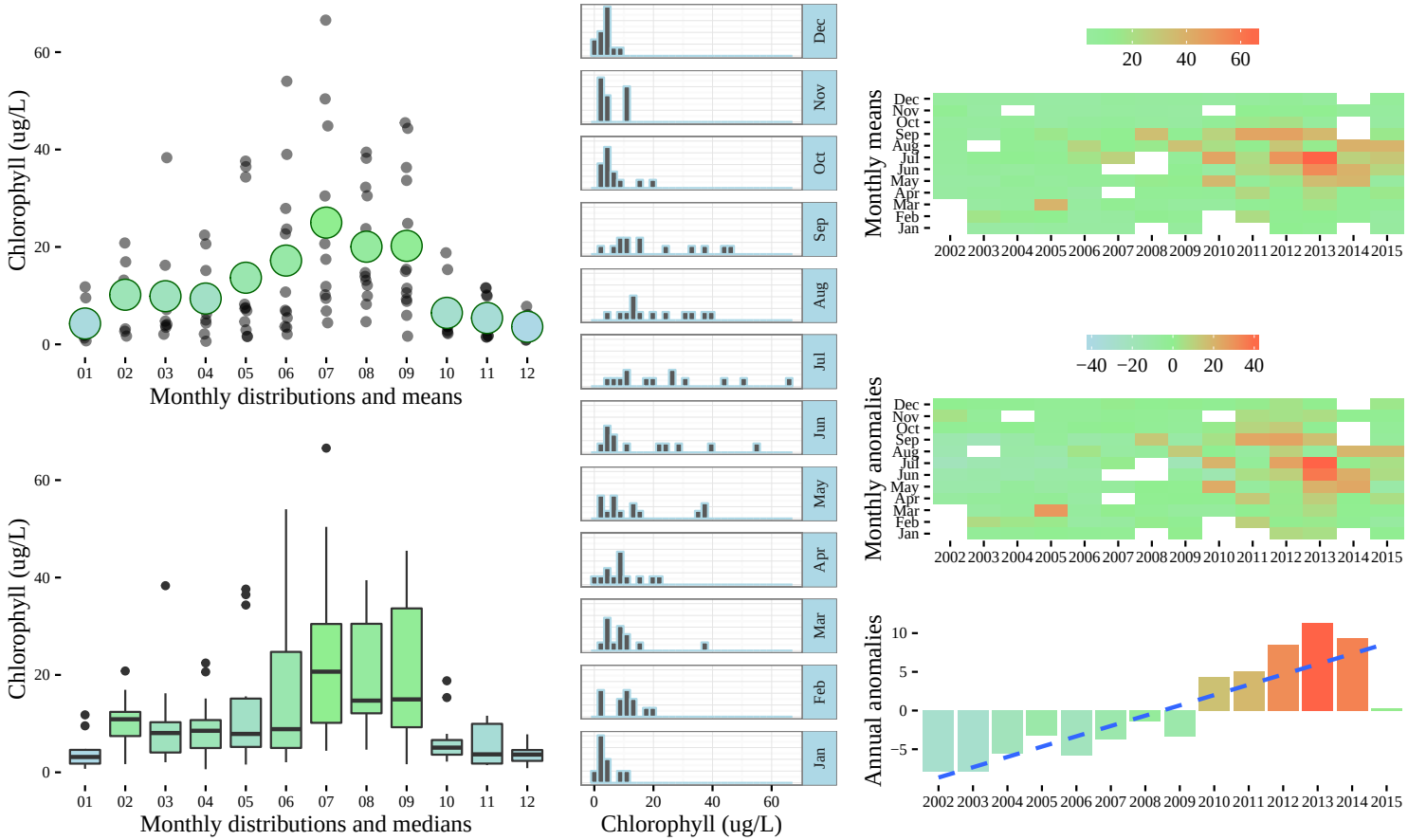
```
# load libraries, add decimal date
library(EnvStats)
library(lubridate)
nut$dec_yr <- decimal_date(nut$datetimestamp)

# run test
ests_k1 <- kendallTrendTest(chla_n ~ dec_yr, nut)
ests_k1$estimate

##           tau         slope     intercept
##     0.3146341     0.8865779 -1772.6181270

ests_k1$p.value

##             z
## 2.097253e-11
```

Using `kendallTrendTest` with NERRS data

3. Interpret results

Do they make sense? Check seasonality in observed data:

```
plot_summary(nut, param = 'chla_n')
```



Using kendallTrendTest with NERRS data

Option 1: aggregate by years

```
# get annual averages
nutyr <- aggreswmp(nut, by = 'years')

# convert datetimestamp to numeric for year
nutyr$yr <- year(nutyr$datetimestamp)

# run test
ests_k2 <- kendallTrendTest(chla_n ~ yr, nutyr)
ests_k2$estimate

##           tau         slope     intercept
##     0.7582418     1.3938194 -2789.7357522

ests_k2$p.value

##             z
## 0.0001971407
```

Using kendallSeasonalTrendTest with NERRS data

Option 2: use `kendallSeasonalTrendTest`

```
# some additional prep for season and year columns
nut$season <- month(nut$datetimestamp)
nut$yr <- year(nut$datetimestamp)

# run test
ests_sk <- kendallSeasonalTrendTest(chla_n ~ season + yr, data = nut)
ests_sk$estimate

##          tau        slope    intercept
##     0.423886     0.907000 -1354.312500

ests_sk$p.value

## Chi-Square (Het)        z (Trend)
##     7.346318e-02     9.868358e-17
```

Summary

Kendall

```
##       tau     slope intercept
##     0.315     0.887 -1772.618
##         z
## 2.1e-11
##   LCL   UCL
## 0.606 1.185
```

Kendall by yr

```
##       tau     slope intercept
##     0.758     1.394 -2789.736
##         z
## 0.000197
##   LCL   UCL
## 0.709 2.014
```

Seasonal Kendall

```
##       tau     slope intercept
##     0.424     0.907 -1354.312
## Chi-Square (Het)         z (Trend)
##         7.35e-02          9.87e-17
##   LCL   UCL
## 0.667 1.209
```

Summary

Final comments:

- If you expect cyclical variation - aggregate to remove or use Seasonal Kendall

Summary

Final comments:

- If you expect cyclical variation - aggregate to remove or use Seasonal Kendall
- Seasonal Kendall requires more work but has more power

Summary

Final comments:

- If you expect cyclical variation - aggregate to remove or use Seasonal Kendall
- Seasonal Kendall requires more work but has more power
- All of the above methods only detect a monotonic trend, do not account for other variables

Summary

Final comments:

- If you expect cyclical variation - aggregate to remove or use Seasonal Kendall
- Seasonal Kendall requires more work but has more power
- All of the above methods only detect a monotonic trend, do not account for other variables
- You can pick your time interval or use an alternative approach (e.g., WRTDS)

Up next... nothing!

Questions??

References

Hirsch RM, Slack JR, Smith RA. 1982.
Techniques of trend analysis for monthly water quality data.
Water Resources Research, 18:107–121.

Millard SP. 2013.
EnvStats: An R Package for Environmental Statistics.
Springer, New York.