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:

```r
install.packages('EnvStats')
library(EnvStats)
```
Is everything installed?

We will use functions in the EnvStats package.

Option 1, from the R Console prompt:

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

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

```r
# 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:

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

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

```r
# 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)
```
Is everything installed?

We will use functions in the lubridate package

Option 1, from the R Console prompt:

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

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

```r
# 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 $\tau$ (‘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 $\tau$ (‘tau’)
- Provide a slope as the rate of change
Theory and background

The **Kendall test** for time series:

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

\[
\hat{\tau} = \frac{2S}{n (n - 1)}
\]

\(\hat{\tau}\) 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:

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

\( \hat{\beta}_1 \) 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:

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

\(\hat{\beta}_1\) 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

More info in help documentation for {	t kendallTrendTest}, {	t kendallSeasonalTrendTest} in EnvStats, [Hirsch et al., 1982, Millard, 2013]
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 $\hat{r}$ 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 lighter points with darker colors indicate a strong trend. Original data are available from [http://cdmo.baruch.sc.edu/](http://cdmo.baruch.sc.edu/). See the GitHub repository for source code. The data include observations through December 2016 (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, -77.06 with a zoom level of 10.

Select parameter:

- nut: Chlorophyll-a (ug/L)

Summarize by:

- Years: anomalies

Select data range:

2002 - 2015

noczbrut, Chlorophyll (ug/L), POS (p<0.01)
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.o'brien@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

Date range

Aesthetics

Y-axis

Lines

Common

Points

Plots Tables Downloads

Parameter type

Nutrients

Chlorophyll (ug/L)

Parameter

Reserve

Stations

noc

nocenut

cnocnut

nocznut

nocenut

nocznut

2005

2010

2015

Chlorophyll (ug/L)

0

5

10

15

20

M. Beck

Seasonal Kendall
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 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

```r
# load SWMPPr, 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)
```

```r
## 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:

```r
plot(nut)
```
Using `kendallTrendTest` with NERRS data

2 Test with `kendallTrendTest`  

Run the test:

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

# run test  
eststs_k1 <- kendallTrendTest(chla_n ~ dec_yr, nut)  
eststs_k1$estimate

## tau slope intercept  
## 0.3146341 0.8865779 -1772.6181270

eststs_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

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

# convert datetime 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`

```r
# 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
```
### Kendall

<table>
<thead>
<tr>
<th>tau</th>
<th>slope</th>
<th>intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.315</td>
<td>0.887</td>
<td>-1772.618</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1e-11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LCL</th>
<th>UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.606</td>
<td>1.185</td>
</tr>
</tbody>
</table>

### Kendall by yr

<table>
<thead>
<tr>
<th>tau</th>
<th>slope</th>
<th>intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.758</td>
<td>1.394</td>
<td>-2789.736</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000197</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LCL</th>
<th>UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.709</td>
<td>2.014</td>
</tr>
</tbody>
</table>

### Seasonal Kendall

<table>
<thead>
<tr>
<th>tau</th>
<th>slope</th>
<th>intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.424</td>
<td>0.907</td>
<td>-1354.312</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chi-Square (Het)</th>
<th>z (Trend)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.35e-02</td>
<td>9.87e-17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LCL</th>
<th>UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.667</td>
<td>1.209</td>
</tr>
</tbody>
</table>
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
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

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.