

# Trend Analysis of Nutrients

Dave Lorenz

July 26, 2017

## Abstract

This example illustrates the data manipulations for the seasonal Kendall analysis using the monthly option rather than the seasonal option for defining seasons. Note that these data were actually sampled on a periodic basis and would be more appropriate for the seasonal option, but the monthly option is used for illustration purposes. The example uses nutrient data. This example uses a common time frame for all of the trend tests. The common time frame facilitates comparing trends among the stations and constituents. Most often users will want to divide trend analyses into similar groups of analytes like major ions, nutrients and so forth because they will be analyzed in similar ways and will have common sampling time frames.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Summarize the Sample Data</b>	<b>3</b>
<b>3</b>	<b>Set up the Project</b>	<b>6</b>
<b>4</b>	<b>Seasonal Kendall Trend Test</b>	<b>8</b>
<b>5</b>	<b>Trend Results</b>	<b>10</b>
<b>6</b>	<b>Further Remarks</b>	<b>11</b>

# 1 Introduction

Need to add blurb about censoring and no log transform.

The data used in this application are a small subset of the data used by Schertz and others (1991). The data are samples taken from water year 1969 (October, 1968) through water year 1989 (September, 1989). Nineteen stations were selected and four nutrients were selected from the original data.

```
> # Load restrend and other packages and the data
> library(restrend)
> library(smwrBase)
> library(smwrQW)
> data(EstrendSub)
> head(EstrendSub)
```

	STAID	DATES	QI	QD	RN.organic	PN.organic	RAmmonia
1	07227500	1968-10-01	7.6	NA			NA
2	07227500	1968-10-03	5.3	NA			NA
3	07227500	1968-10-16	532.0	NA			NA
4	07227500	1968-10-19	17.0	NA			NA
5	07227500	1968-11-01	17.0	NA			NA
6	07227500	1968-12-01	6.6	NA			NA
	PAmonia	RKjeldahl	PKjeldahl	RTotal.P	PTotal.P	RCopper	PCopper
1	NA		NA		NA		NA
2	NA		NA		NA		NA
3	NA		NA		NA		NA
4	NA		NA		NA		NA
5	NA		NA		NA		NA
6	NA		NA		NA		NA
	RIron	PIron	Calcium	Chloride			
1		NA	95	280			
2		NA	NA	NA			
3		NA	42	106			
4		NA	121	435			
5		NA	150	512			
6		NA	138	510			

## 2 Summarize the Sample Data

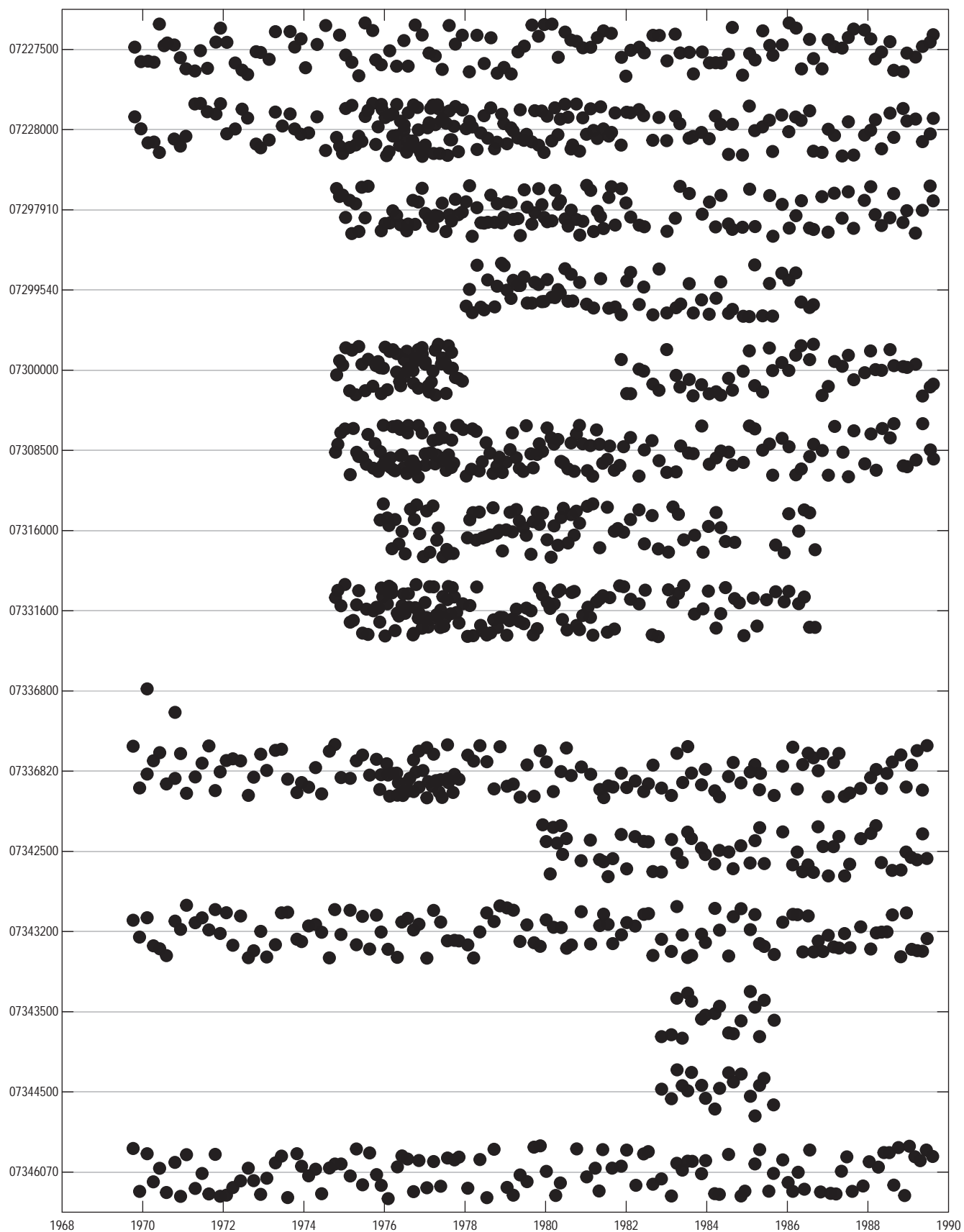
In general, it is desirable, but not necessary, to subset the data before proceeding with the analysis of a subset of the constituents. Nutrient concentrations can vary widely in natural waters and can range from completely uncensored to highly censored. No flow adjustment will be done for these data because of the potential high variability in censoring—for these data, about 10 percent of ammonia samples are censored, but the other constituents have much lower percentages.

```
> # Convert to class qw
> EstrendSub.qw <- convert2qw(EstrendSub)
> # Create the subset
> Nuts <- subset(EstrendSub.qw, select=c("STAID", "DATES",
+   "PN.organic", "PAmmonia", "PKjeldahl", "PTotal.P"))
> # Rename to remove leading P, not required--just pretty
> constituents <- c("N.organic", "Ammonia", "Kjeldahl", "Total.P")
> names(Nuts)[3:6] <- constituents
> # The sampling for nutrients started later, so remove the samples that
> # have no nutrient data.
> Nuts <- dropAllMissing(Nuts, constituents)
```

The `sampReport` function creates a simple PDF file that contains a report of the sample date ranges and graph of samples for each station. It can be used to help define the starting and ending date ranges for the trend tests as well as identifying sample gaps and other sampling issues.

```
> # Create the report
> sampReport(Nuts, DATES="DATES", STAID="STAID", file="NutrientSampling")
```

	STAID	FirstSamp	LastSamp	NumSamp
1	07227500	1969-10-21	1989-08-15	119
2	07228000	1969-10-21	1989-08-16	182
3	07297910	1974-10-24	1989-08-15	123
4	07299540	1978-01-11	1986-08-25	68
5	07300000	1974-10-25	1989-08-16	106
6	07308500	1974-10-15	1989-08-18	153
7	07316000	1975-11-24	1986-09-09	88
8	07331600	1974-10-15	1986-09-09	136
9	07336800	1970-02-10	1970-10-21	2
10	07336820	1969-10-06	1989-06-20	144
11	07342500	1979-12-06	1989-06-19	61
12	07343200	1969-10-06	1989-06-21	122
13	07343500	1982-11-17	1985-09-04	18
14	07344500	1982-11-18	1985-08-29	18
15	07346070	1969-10-06	1989-08-10	117



The call to `sampReport` returns the file name invisibly (`NutrientSampling.pdf`). Because it is a full-size portrait PDF file, it is inserted here with compressed pages. The report gives the actual begin and end dates for sampling and the graph shows the sampling dates for each station. It is easy to see that only 5 stations (07227500, 07228000, 07336820, 07343200, and 07346070) were sampled for the entire retrieval period. Note that the report only shows when any sample was taken, the ranges for individual constituents can differ from the pattern shown. The code immediately following this paragraph demonstrates how to show the sampling pattern for a single constituent, but is not executed. The report would show that Kjeldahl sampling did not start at any station until 1974.

```
# The sampling pattern for Kjeldahl
sampReport(na.omit(Nuts[c("STAID", "DATES", "Kjeldahl")]),
  DATES="DATES", STAID="STAID", file="KjeldahlSampling")
```

### 3 Set up the Project

The user must balance the need to include as many stations as possible and the targeted time frame for the trend estimation. For these data, 5 stations have a reasonably complete record, but to include all of those stations, the analysis period would need to be much shorter. This example will use the full retrieval period and include only the 5 stations with reasonably complete record.

The `(setProj)` function sets up the trend estimation project. There are many arguments to `(setProj)`, see the documentation for details. The constituent names or response variable names are referred to as **Snames** in keeping with the names used in the original ESTREND.

After projects have been set up, the user can get a list of the projects by using `lsProj` or can specify a project to use with `useProj`. The function `useProj` must be used to continue working on a project after the user quits from the R session.

```
> # subset to a few selected stations:
> Nuts <- subset(Nuts, STAID %in% c("07227500", "07228000", "07336820", "07343200",
+   "07346070"))
> # Set up the project
> setProj("nutrients", Nuts, STAID="STAID", DATES="DATES",
+   Snames=constituents,
+   type="monthly", Start="1969-10-01", End="1989-10-01")
```

```
[1] "nutrients"
```

The `(setProj)` function creates a folder in the users workspace with that name. That folder contains R data that are updated after each successful call to an analysis function in **restrend**. Table 1 describes the data created in this example's call to `(setProj)`. Any object of class "matrix" or "by" are indexed by station and **sname**.

**Table 1.** The data created by `(setProj)`.

Name	Class	Description
estrend.cl	list	A record of the calls to analysis functions.
estrend.cn	matrix	A description of the censoring. May be "none," "left," or "multiple."
estrend.cp	matrix	The percent of observations that are left-censored.
estrend.df	by	The dataset, contains STAID, DATES, and the response variable.
estrend.in	list	Information about the project, such as the start and end dates and the names of columns in each dataset.
estrend.ml	by	Details from the monthly selection process. Each is a list from the potential comparisons from each month of the year, the selected months, and the number of months . See Lorenz (2016) for details.
estrend.ms	matrix	The number of months or seasons from the analysis recorded in <b>estrend.ml</b> .
estrend.st	matrix	The status for each station and sname. Must be "OK" to continue with the trend analysis.

It is useful to verify which stations and snames will be analyzed and what the seasonal definitions are. The user need only enter the name of the R data object in the console. For these data, the seasonal definition is 0 in all cases where the status is not "OK."

```
> # Which are OK?
> estrend.st
```

	snames			
stations	N.organic	Ammonia	Kjeldahl	Total.P
07227500	"short record"	"OK"	"short record"	"OK"
07228000	"short record"	"OK"	"short record"	"OK"
07336820	"short record"	"OK"	"short record"	"OK"
07343200	"short record"	"OK"	"short record"	"OK"
07346070	"short record"	"OK"	"short record"	"OK"

```
> # What seasonal definition?
> estrend.ms
```

	snames			
stations	N.organic	Ammonia	Kjeldahl	Total.P
07227500	0	2	0	2
07228000	0	6	0	6
07336820	0	4	0	5
07343200	0	5	0	5
07346070	0	4	0	4

## 4 Seasonal Kendall Trend Test

These data are ready for the seasonal Kendall trend test. The function **SKTrends** executes the trend test on all valid combinations of stations and snames. It can also execute the test on subsets if some changes need to be made.

The **SKTrends** function also creates a PDF file that contains the result of the analysis and a series graph on each page. See the documentation for **seriesPlot** for information about that graph. The file reports the results for each sname by station with the flow-adjusted results following the untransformed results. Most trends are very small for these data; only the reports for Calcium at 07228000 is shown.

```
> # Trend tests, accepting default seasons
> SKTrends()

[1] "nutrients_sk.pdf"
```



07336820 Ammonia

Seasonal Kendall's tau with the Turnbull slope estimator

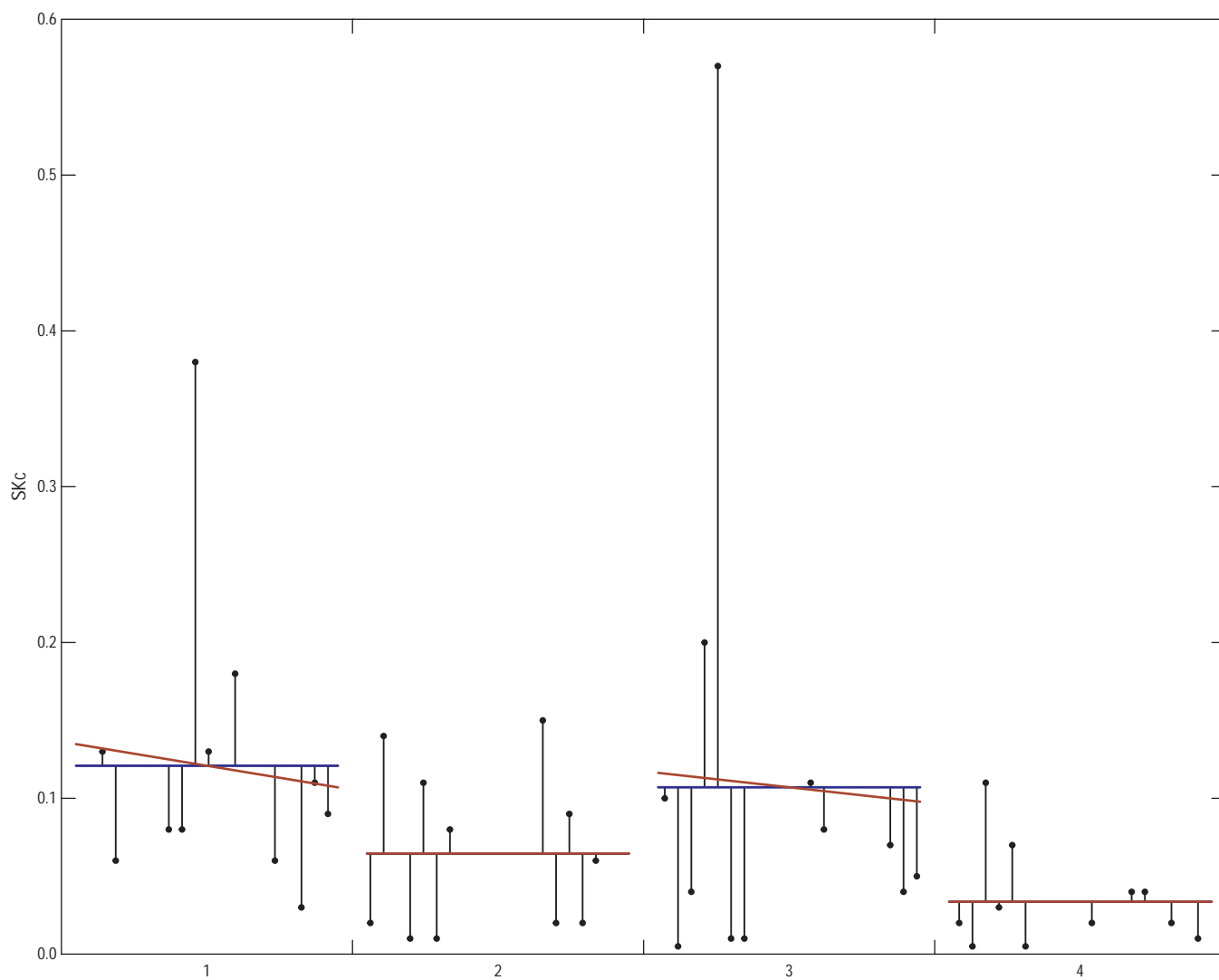
data: Ammonia (20 years and 4 seasons)

tau = -0.038961, p-value = 0.785

alternative hypothesis: true slope is not equal to 0

sample estimates:

slope	median data	median time
0.00	0.06	10.00



## 5 Trend Results

When completed, or to check on intermediate results, the estimated trends can be extracted using the `getTrends` function. By default, all stations and snames are extracted. The output dataset is explained in the documentation for `getTrends`. The user has the option to set a significance level to determine whether there is a significant trend, the default level is 0.05.

```
> # get the trends
> nutrients.tnd <- getTrends()
> print(nutrients.tnd)
```

	Station	Response	Type	NumYears
1	07227500	Ammonia censored seasonal	Kendall	20
2	07227500	Total.P uncensored seasonal	Kendall	20
3	07228000	Ammonia censored seasonal	Kendall	20
4	07228000	Total.P uncensored seasonal	Kendall	20
5	07336820	Ammonia censored seasonal	Kendall	20
6	07336820	Total.P uncensored seasonal	Kendall	20
7	07343200	Ammonia censored seasonal	Kendall	20
8	07343200	Total.P uncensored seasonal	Kendall	20
9	07346070	Ammonia censored seasonal	Kendall	20
10	07346070	Total.P uncensored seasonal	Kendall	20

	NumSeas	Nobs	RepValue	Trend	Trend.pct	P.value
1	2	30	0.1400000	-0.0121428571	-8.6734694	0.003922263
2	2	30	0.8746428	-0.1168781209	-13.3629550	0.056161419
3	6	87	0.0500000	0.0000000000	0.0000000	0.593257540
4	6	88	0.0900000	-0.0036574930	-4.0638812	0.013454203
5	4	45	0.0600000	0.0000000000	0.0000000	0.785049330
6	5	61	0.1100000	-0.0037587682	-3.4170620	0.063969910
7	5	56	0.0800000	0.0000000000	0.0000000	0.878997150
8	5	56	0.1349074	-0.0027091304	-2.0081407	0.233828798
9	4	50	0.0650000	0.0000000000	0.0000000	0.978917699
10	4	51	0.0900000	-0.0004130066	-0.4588962	0.303643316

	Trend.dir
1	down
2	none
3	*
4	down
5	*
6	none
7	*
8	none
9	*
10	none

## 6 Further Remarks

Because trend analysis is not necessarily a straightforward process, but requires user assessments at several points in the process, it is not necessarily a good idea to simply create scripts and run them without any user review and interaction. To overcome recording the steps in a script, the functions in `restrend` record all changes to the projects database in a list called `estrend.cl`. It can be viewed at any time simply by entering `estrend.cl` in the console window. It can be saved with the data to ensure that the trend analysis is reproducible.

```
> # get the history
> estrend.cl
```

```
[[1]]
setProj(project = "nutrients", data = Nuts, STAID = "STAID",
        DATES = "DATES", Snames = constituents, type = "monthly",
        Start = "1969-10-01", End = "1989-10-01")
```

```
[[2]]
SKTrends()
```

## References

- [1] Lorenz, D.L., in preparation, `restrend`: an R package for EStimate TRENDS: U.S. Geological Survey Open File Report, ? p.
- [2] Schertz, T.L., Alexander, R.B., and Ohe, D.J., 1991, The computer program EStimate TREND (ESTREND), a system for the detection of trends in water-quality data: U.S. Geological Survey Water Resources Investigations Report 91-4040, 72 p.