Using R to Analyze Water-Quality Trends in Estuarine Areas

Wendy Wright [1,2], Todd C. Rasmussen [1] and M. Brian Gregory [2]

[1] University of Georgia, Warnell School of Forestry and Natural Resources [2] National Park Service, Southeast Coast Network Inventory & Monitoring

Timucuan Ecological and Historic Preserve

Timucuan Ecological and Historic Preserve (TIMU) encompasses 18,600 hectares (46,000 acres) of salt marsh and coastal hammock habitat in addition to estuarine, marine, and brackish open waters. TIMU contains the seaward confluence of the Nassau and St. Johns Rivers and is along the northeastern coast of Florida in Duval County and entirely within the city limits of Jacksonville. TIMU includes several rare and vulnerable natural communities including coastal strand, maritime hammock, scrub, and shell mound habitat. National Park Service (NPS) facilities within TIMU include Fort Caroline visitor center and maintenance area, the Theodore Roosevelt area with park headquarters, the Kingsley Plantation, the Ribault Club, and the historic Broward house. Other state and city parks in the area include Big and Little Talbot Island State Parks, Fort George Island Cultural State Park, Little Jetties Park, Huguenot Memorial Park, and the Sisters Creek Park and boat ramp.

Water resources are an integral part of TIMU with approximately 75% of the protected area composed of wetlands and open water. These resources include numerous tidal creeks, portions of the Nassau and St. Johns Rivers, Sisters Creek/Intracoastal Waterway (ICWW), and the Fort George River. TIMU's estuarine setting serves as a vital ecological link between freshwater habitats and the ocean. The waters of TIMU have been impacted by land use in the surrounding watersheds. Examples of water-quality impacts applicable to TIMU include nonpoint source pollution from urban and agricultural areas, elevated metal concentrations in the sediments of the St. Johns River, impacts of several Superfund sites and landfills, and water pollution from malfunctioning septic systems within and adjacent to TIMU.[3]

Because of the importance of water resources to park management from ecological, regulatory, and visitor experience perspectives, marine water quality in and around parks was selected as one of several Vital Signs to be monitored by the NPS Southeast Coast Network Inventory and Monitoring Program (SECN I&M). Data collected as a part of this effort are intended to foster a better understanding of water resources in and around parks, and park ecological processes and physical systems.
To effectively monitor the diverse and dynamic set of estuarine water resources, the SECN has implemented a two-pronged monitoring approach. The first aspect consists of collecting data from continuously recording fixed-stations located in each park unit and collecting discrete nutrient samples during monthly site visits. The second aspect consists of conducting periodic (every five years) synoptic assessments of water quality at a larger spatial scale in the waters surrounded coastal park areas. When combined, these two approaches give a more complete picture of physical processes, diurnal and seasonal variations, and spatial patterns during critical portions of the year than either could if implemented singularly. [4]
Data for this trend analysis was collected through a partnership between the NPS SECN I&M and the City of Jacksonville, Florida, Environmental Quality Division, with special thanks to Betsy Deuerling of City of Jacksonville EQD.

Methods

R is an open-source language and environment for statistical computing, and it is freely available at http://www.r-project.org/.[5] Several packages are available for R, and are written to target specific issues or problems. "wq" is an R package designed for water quality analysis. It is targeted at monthly discrete water-quality sampling programs, and contains a wide array of functions to analyze time-series datasets.[6]

Analysis with wq begins with a dataset in either long or wide format. Data is imported, and columns are given new row names that are compliant with wq requirements.

```
TIMU <- read.csv ("TIMU.nut.trend.csv", header=FALSE, as.is=TRUE, skip = 1)
names(TIMU) <- c('depth', 'site', datetime', 'TP', 'TN', 'Chla')
```

Data is then converted to a wq object:

```
TIMUwq <- wqData(TIMU, c(3,2,1),4:6, site.order = TRUE, type = "wide", time.format = "%m/%d/%Y")
```

At this point, it is very easy to obtain summaries of the dataset.

```
TIMUwq <- wqData(TIMU, c(3,2,1),4:6, site.order = TRUE, type = "wide", time.format = "%m/%d/%Y")
write.csv(summary$observations, file = "summary.nut.csv")
```
(Table 1.)

```
write.csv(summary$quartiles, file = "summary.quartiles.csv"
```
(Table 2.)

Summary Graphs can also be run, for individual sites, or all sites.

```
plot(TIMUwq, vars = c('TP'), num.col=0.5)
```
(Figure 1.)

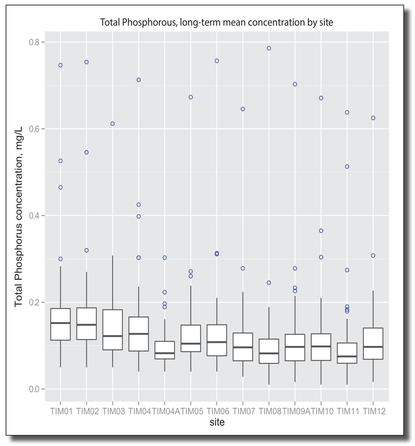
Table 1. Number of observations at each site in TIMU dataset.

	TP	TN	Chla
TIM01	71	69	34
TIM02	71	69	33
TIM03	70	68	33
TIM04	71	69	35
TIM04A	22	22	19
TIM05	72	70	32
TIM06	71	69	31
TIM07	69	67	30
TIM08	73	71	33
TIM09A	69	67	31
TIM10	66	64	29
TIM11	88	87	54
TIM12	66	62	25
TIM13A	0	0	38

Table 2. Minimum, maximum, mean, median and 1st and 3rd quartiles across all sites for TP, TN and Chlorophyll *a*.

	Min.	Q1	Median	Mean	Q3	Max.
TP (mg/L)	0.01	0.0735	0.106	0.1255	0.152	0.786
TN (mg/L)	0.041	0.4882	0.708	0.7705	0.9805	3.2
Chla (µg/L)	0.63	4.58	7.11	8.682	10.58	68.53

Figure 1. Phosphorus concentrations: long-term mean with error bars, for all sites.



The function 'tsMake' converts the wq dataset into a time series. Once the dataset is converted, many time-series graphs and tables can be produced. In these samples, phosphorus is graphed for all sites across the period of record (Figure 2). Data can also be filtered for a single site and graphed (Figure 3).

```
TIMUtsTP <- tsMake(TIMUwq, focus = "TP", layer = c(0,5)) #, type = c("ts.mon"), qprob = NULL) plotTs(TIMUtsTP, ylab = "Total Phosphorus, mg/L", ncol = 2

TIM09A <- TIMU[which(TIMU$site=='TIM09A'),]

plotTs(TIMU09AtsTP, ylab = "Total Phosphorus, mg/L", ncol = 1)
```

Figure 2. Phosphorus concentrations across all sites for the time period 2000-2013.

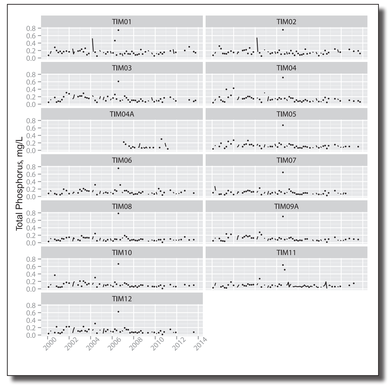
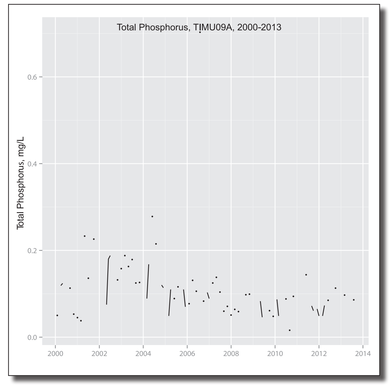


Figure 3. Phosphorus concentrations for site TIMU09A for the time period 2000-2013.



Trends over time are computed in the wq package using a non-parametric Mann-Kendall test, which tests for a change in the median values over time. Seasonal Kendall tests are also available.

Annual Mann-Kendall tests are simple to perform with a time-series dataset, and can be output in tabular format (Table 3) or graphical format (Figure 4). Tables can be output to .csv files if further manipulation in Excel is desired. Results include the slope, slope percent, a p-value giving the significance of the slope, and Kendall's Tau. There is also an indication if the dataset is missing more than 50% of the records in the dataset.

```
TIMU.TP.G <- mannKen(TIMUtsTP, plot=TRUE, type = c("slope", "pct", "tau"), order=FALSE)
TIMU.TP.G

TIMU.TP <- mannKen(TIMUtsTP, plot=FALSE, type = c("slope", "pct", "tau"), order=TRUE)
TIMU.TP

write.csv(TIMU.TP, file="mann.ken.TP.TIMU.csv")
```

For site TIM09A, eight sites are showing a statistically significant downward trend for Total Phosphorus concentrations.

Figure 4. Kendall's Tau and significance of trends for phosphorus during the period of record (2000-2013).

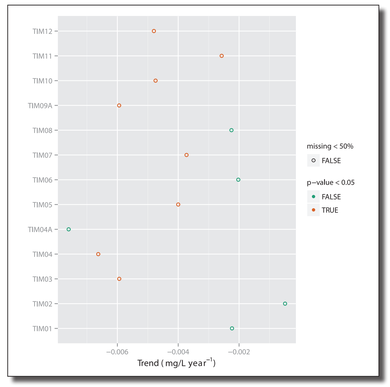


Table 3. Kendall's Tau and significance of trends for phosphorus during the period of record (2000-2014) in tabular format.

	sen. slope	sen. slope.pct	p.value	tau
TIM01	-0.0022	-1.3226	0.2202	-0.0998
TIM02	-0.0005	-0.2958	0.7170	-0.0298
TIM03	-0.0059	-4.1098	0.0058	-0.2257
TIM04	-0.0066	-4.5897	0.0010	-0.2672
TIM04A	-0.0076	-7.0938	0.3231	-0.1558
TIM05	-0.0040	-3.2283	0.0077	-0.2148
TIM06	-0.0020	-1.6190	0.1556	-0.1155
TIM07	-0.0037	-3.4191	0.0097	-0.2131
TIM08	-0.0023	-2.2661	0.0725	-0.1438
TIM09A	-0.0059	-5.2521	0.0002	-0.3090
TIM10	-0.0047	-4.1478	0.0160	-0.2033
TIM11	-0.0026	-2.7332	0.0478	-0.1437
TIM12	-0.0048	-4.1651	0.0114	-0.2135

Time-series can also be graphed by era (Figure 5) or by month (not shown). The data in Figure 5 was divided into two eras, and the graph shows a very distinct visual indication of the downward trends of phosphorus levels calculated in the Mann-Kendall analysis.

```
TPseas <-plotSeason(TIMU09AtsTP,
type=c("by.era"), num.era=2, same.
plot=TRUE, ylab="Phosphorus Levels by
Era, mg/L, TIMU 9A", num.col=3)
```

Site coordinates can be mapped and trends displayed spatially (Figure 6).

In addition to the methods outlined here, wq has capabilities to do time-series decompostion, Kendall seasonal trends, data reshaping, regional Kendall tests, and empirical orthogonal functions.

R, with its strong base functions, and dozens of custom packages available, is a powerful tool for data manipulation and statistical analysis for all fields.

Figure 5. Phosphorus levels at site TIMU09A divided into two eras, (2000-2006) and (2006-2013)

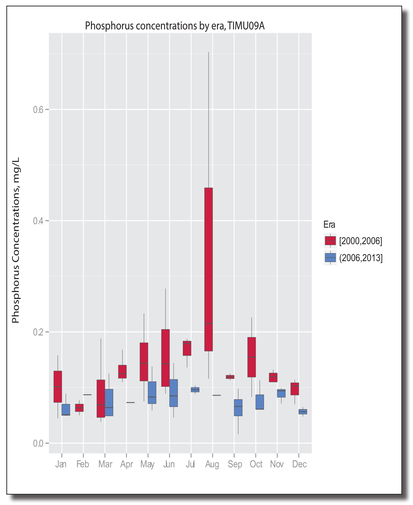
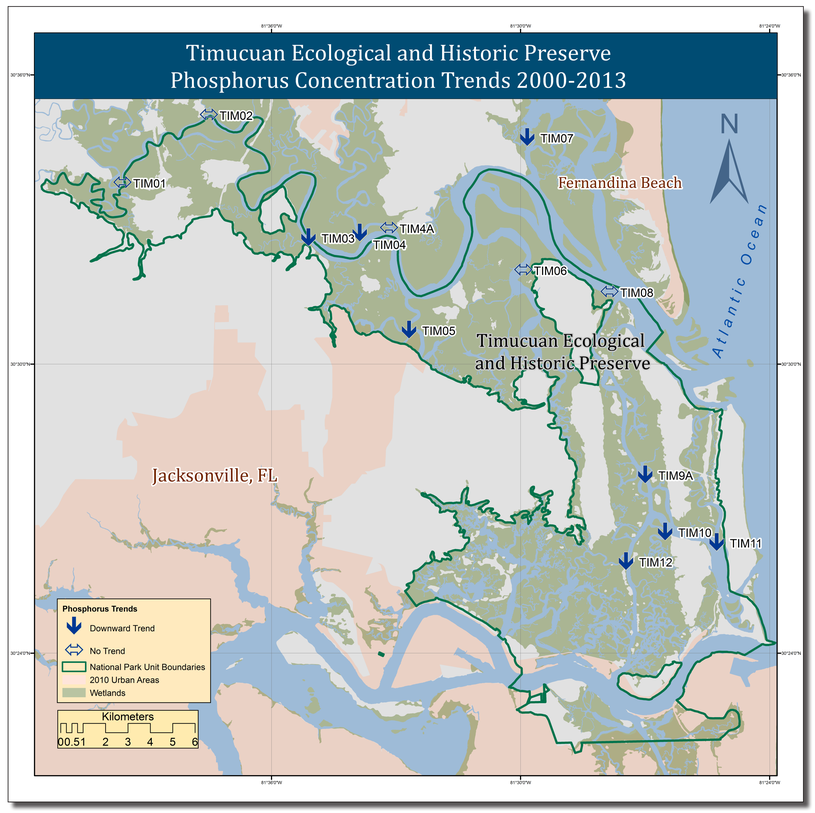


Figure 6. Map of Timucuan Ecological and Historic preserve and vicinity, showing spatial distribution of phosphorus concentration trends across all sites sampled in 2000-2013.



[3] Wright, W., M. B. Gregory and J. Asper. 2013. Assessment of estuarine water quality at Timucuan Ecological and Historic Preserve, 2013. Natural Resource Data Series NPS/SECN/NRDS—2013/598 National Park Service, Fort Collins, Colorado.
[4] Gregory, M. B., J. C. DeVivo, E. M. DiDonato, C. J. Wright, and E. Thompson. 2013. Protocol for monitoring estuarine water and sediment quality in selected Southeast Coast Network parks. Natural Resource Report NPS/SECN/NRR—2013/644. National Park Service, Fort Collins, Colorado.
[5] The R Project for Statistical Computing . 2013. http://www.r-project.org/.
[6] Jassby, Alan D. and James E. Cloern. 2013. Package 'wq'. http://cran.r-project.org/web/packages/wq/wq.pdf