Time Series analysis of Rivers parameters

Iuri Lira Santos

Introduction

Dissolved Oxygen (DO) concentration in rivers is an important parameter to understand the health of the river. Dissolved Oxygen concentration determines the amount of aquatic species living in the river. Excessive amount or lack of DO in the water can harm the species since oxygen is responsible for the respiration of several species.

In order to understand the variability of DO during time, a time series data will be evaluated. Two different locations with two different sizes were selected: one in the northern US with high flow and one in southern US with low flow. By that, is expected to understand the influence of temperature and flow in DO concentrations in Rivers.

Data description

The data being analyzed in this project consists in 7 years of registered flow from the Genesee River in NY. The collection station is the USGS station: "04231600 - Genesee River at Ford Street Bridge, Rochester, NY" The data also contains information about temperature, conductance, pH, turbidity and dissolved oxygen. Data was collected daily in a frequency of 15 minutes from 12/12/2010 to 10/5/2017.

The second set of data is from the Pecos River in Texas. The collection station is the USGS station: "08446500 – Pecos River near Girvin, TX". The data is collected in a frequency of 15 minutes from 10/01/2011 to 8/18/2015.

Objective

The main objective of this project will be to use the time series analysis to understand the relationship between dissolved oxygen and flow and temperature. To achieve the main goal, a study of the data will be performed using the RStudio statistical package. To study the data variability, the autocorrelation functions and the decomposition will be performed on each set of data. To analyze the relationship between DO and flow and temperature, autocorrelation function between the two time series will be performed.

Methods

To analyze the data as a time series, the gaps in the data had to be interpolated. The data was obtained at the USGS website as a text file and exported to excel, to generate a csv file. The csv was imported to the R environment using the read_csv function and the columns with the discharge, temperature and dissolved oxygen were converted to a time series by the "TS" function. The gaps in the data was interpolated by a linear interpolation using the "na.interpolation" function since the "spline" and the "stine" produced a peak in the time series due to the high slope that the previous point produced. Once the gaps in the data were interpolated, the analysis functions for time series could be applied.

The tree time series for the two evaluated rivers were tested on autocorrelation, partial autocorrelation, decomposed and then the cross correlation between dissolved oxygen and temperature and oxygen and discharge was studied. The evaluation of autocorrelation was performed utilizing the "stat" library using the "acf" function and the partial autocorrelation the "pacf" function, decomposition was studied by the "decompose" function and the cross correlation by the "ccf" function. Smoothing of the time series was performed in an attempt to reduce the error in the decomposition since the in first analysis of the data no trend was identified, therefore a daily, weekly and 2 weeks moving average smoothing was used applying the function "ma" from the "FPP" package (Hyndman and Anthanasopoulos, 2012.

Results

Interpolation of the gaps in the data generated the time series for the discharge, temperature and dissolved oxygen for the Genesee River (Figures 1 to 3).

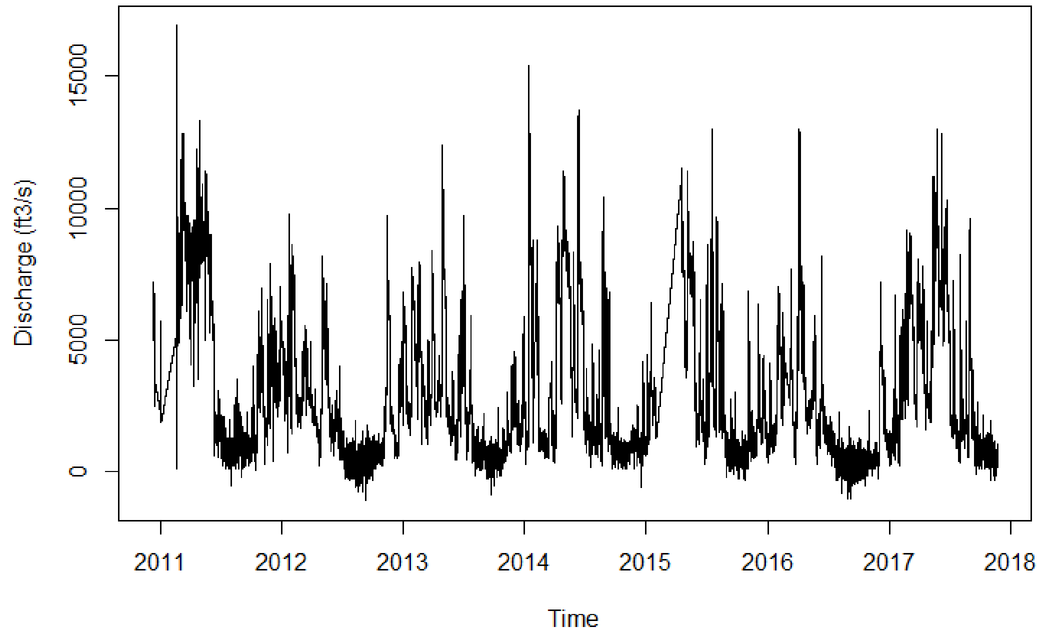


Figure 1 – Genesee River discharge time series with gaps filled by linear interpolation.

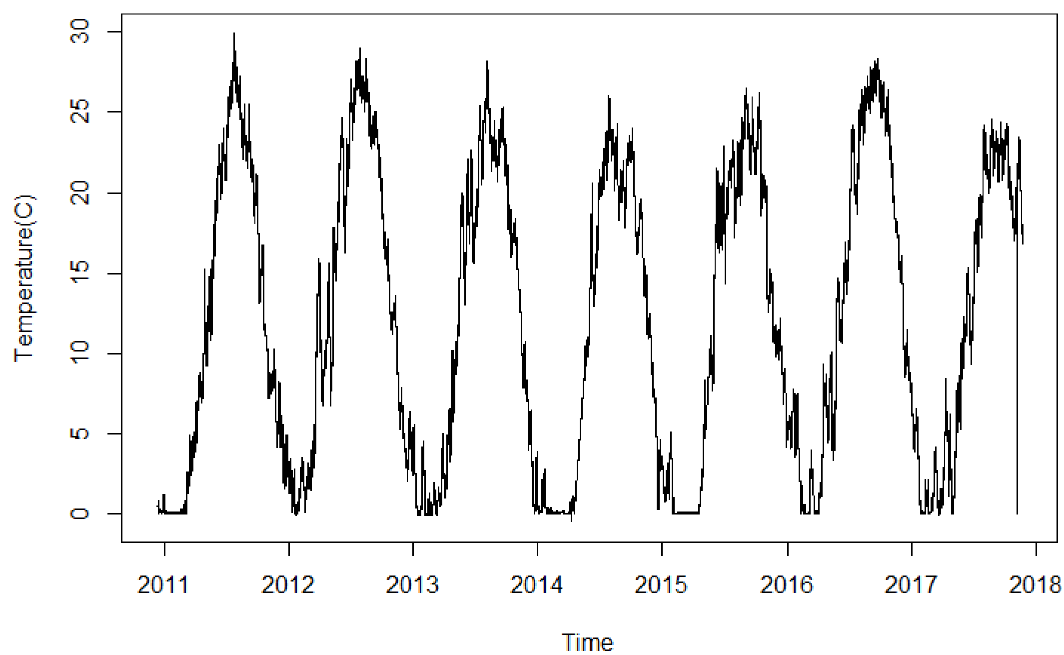


Figure 2 – Genesee River temperature time series with gaps filled by linear interpolation.

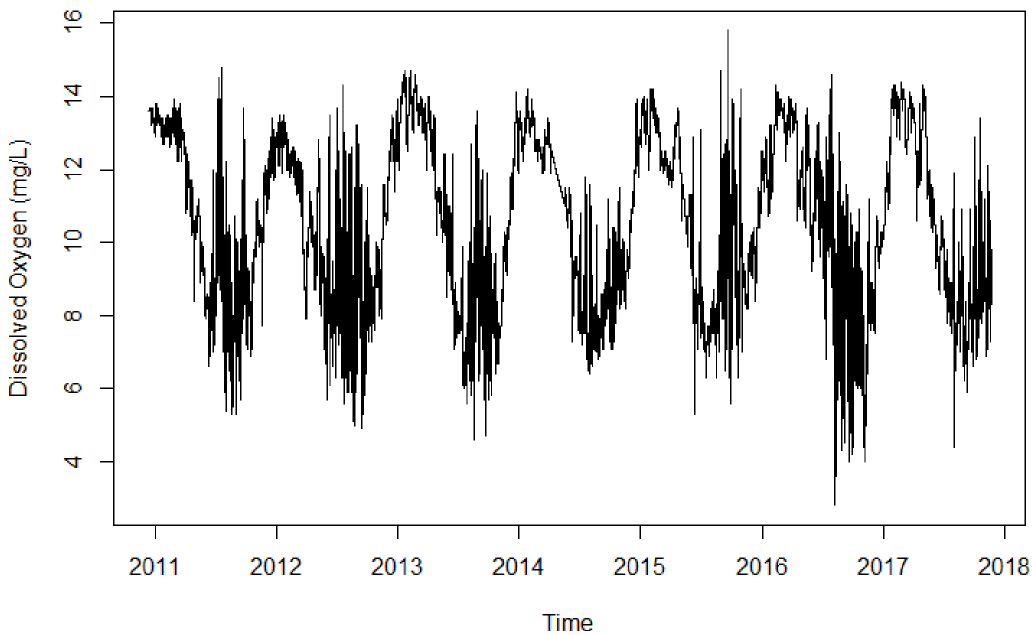


Figure 3 – Genesee River dissolved oxygen time series with gaps filled by linear interpolation.

Analysis of the autocorrelation and partial correlation are shown in figures 4 to 6 for the interpolated time series of the Genesee River.

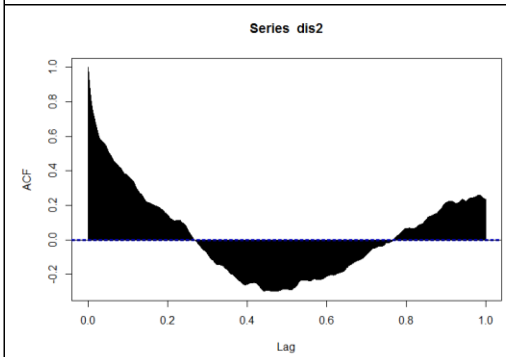
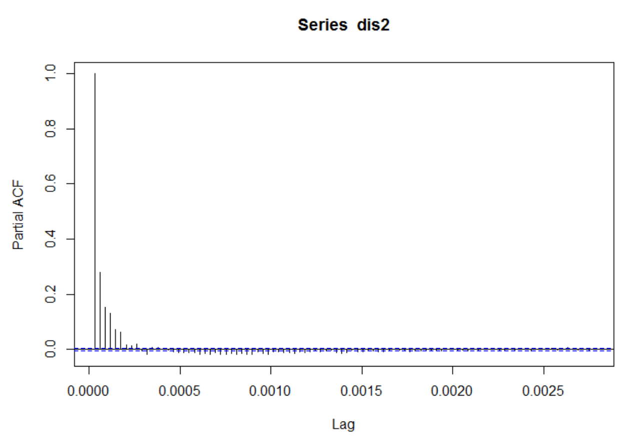
a)	b)
	

Figure 4 - a) autocorrelation function for the discharge time series of the Genesee River for one year lag and b) partial autocorrelation function for the discharge time series of the Genesee River one day lag.

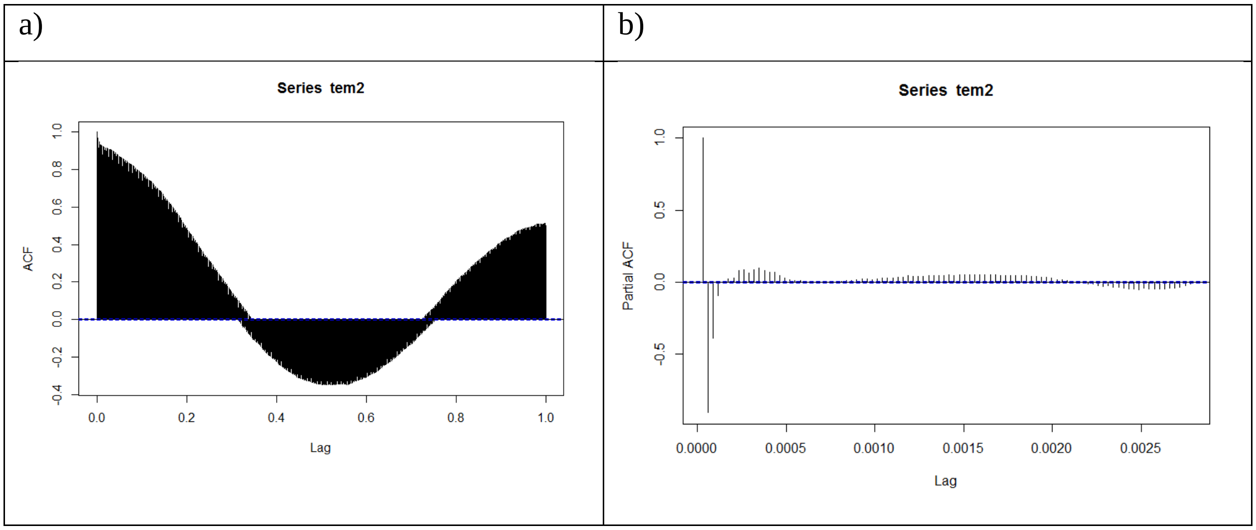


Figure 5 - a) autocorrelation function for the temperature time series of the Genesee River and b) partial autocorrelation function for the temperature time series of the Genesee River.

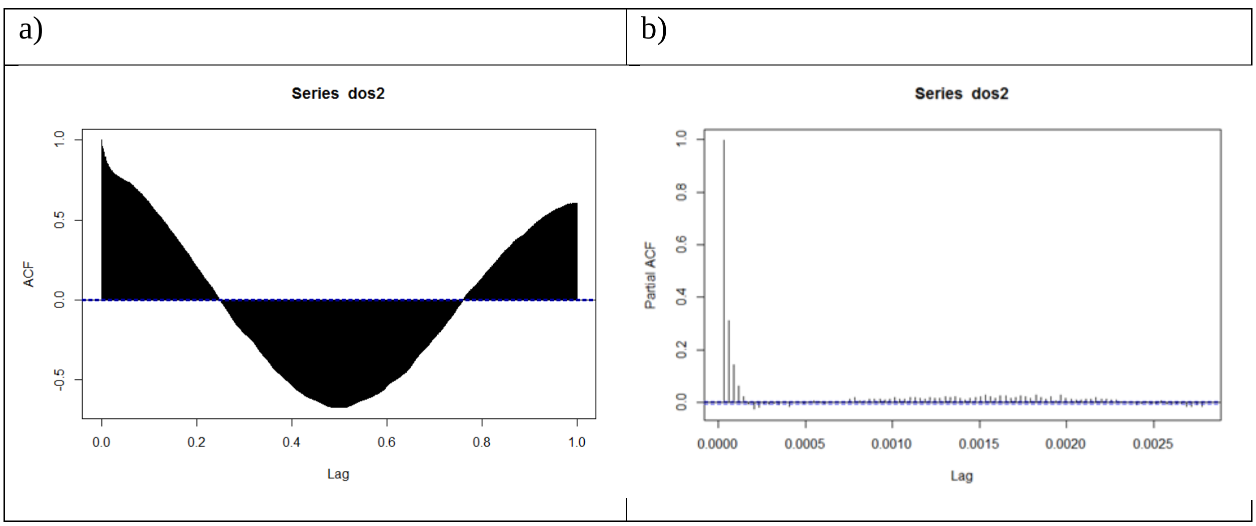
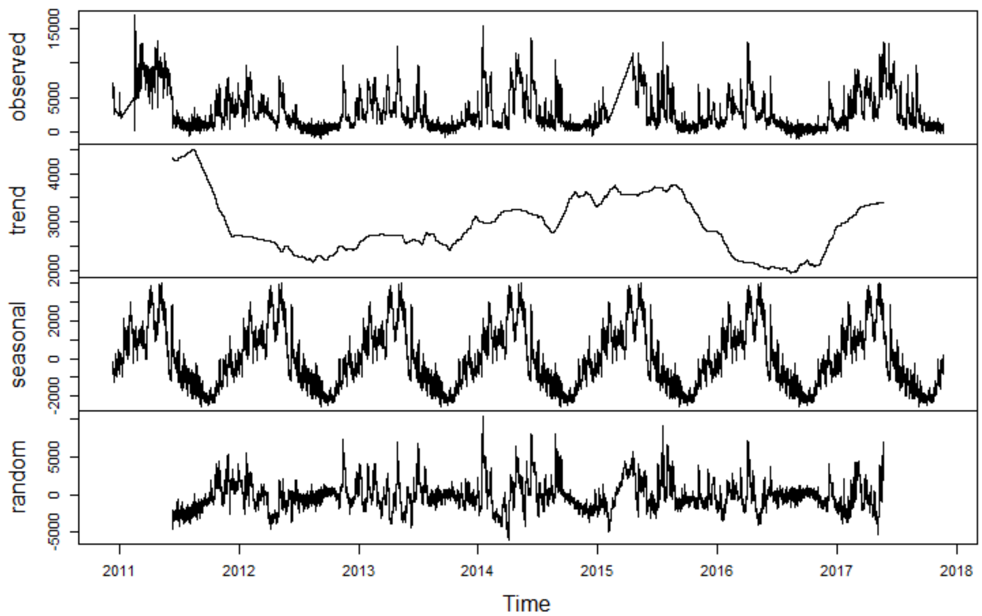
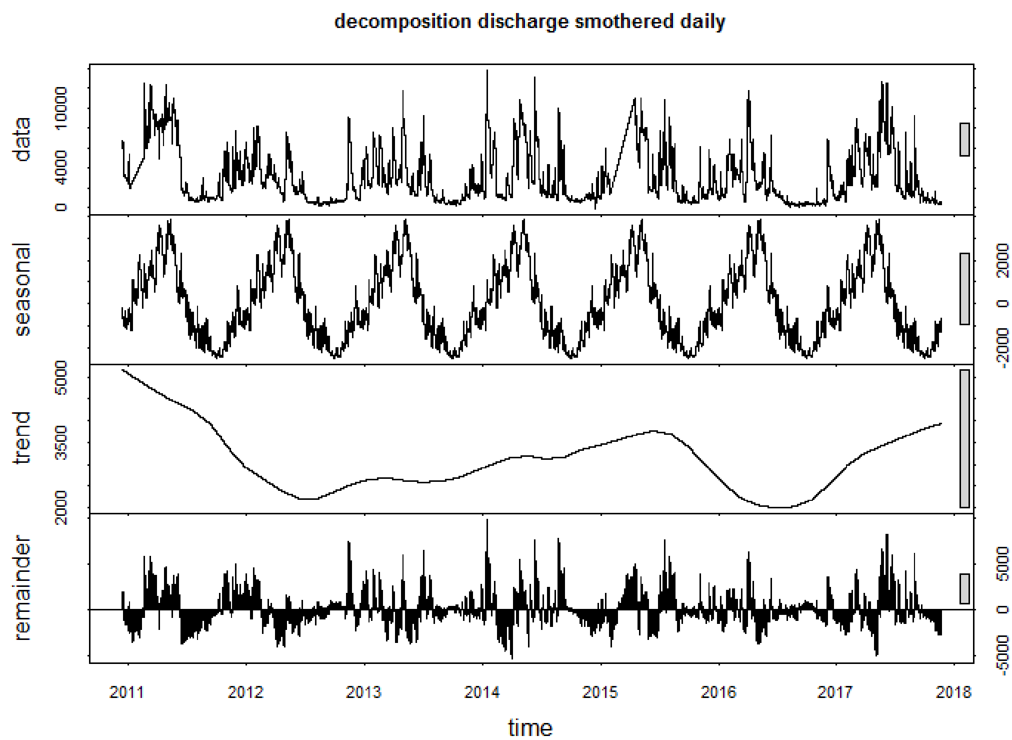
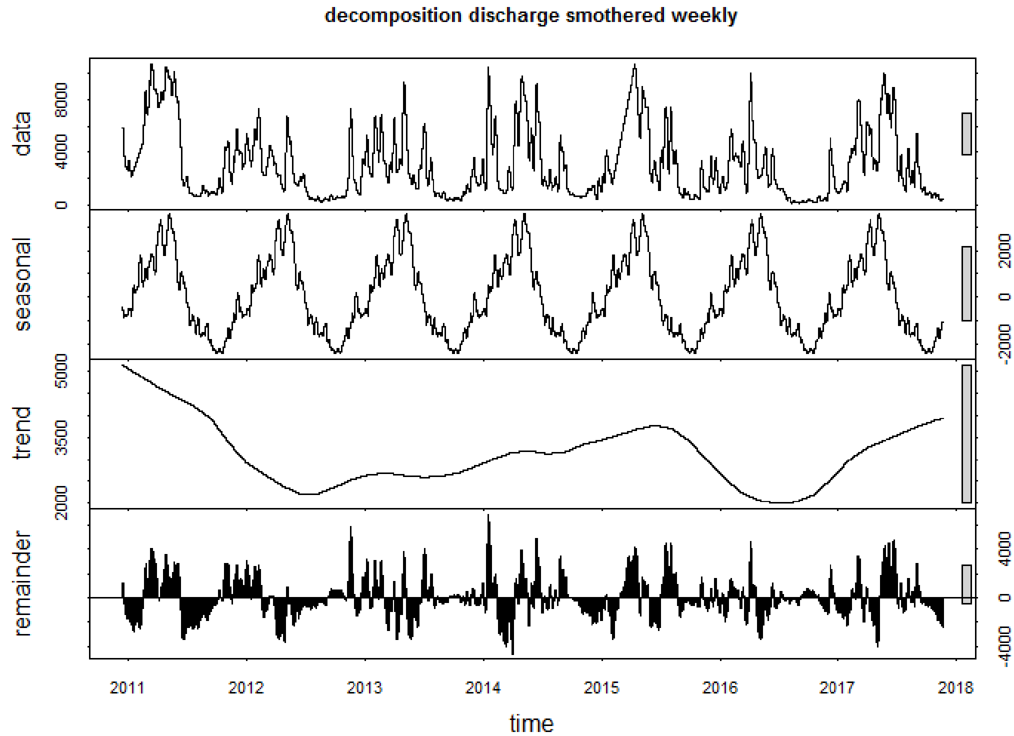


Figure 6 - a) autocorrelation function for the dissolved oxygen time series of the Genesee River and b) partial autocorrelation function for the dissolved oxygen time series of the Genesee River.

Decomposition of the 3 studied time series for the Genesee River is plotted in figures 7 to 9.







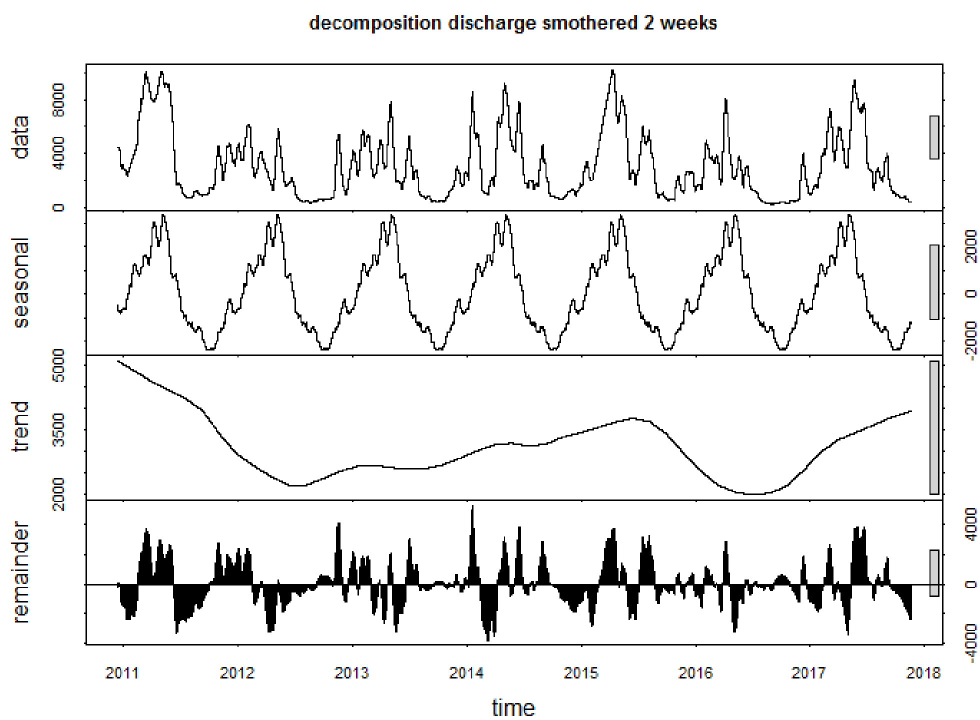
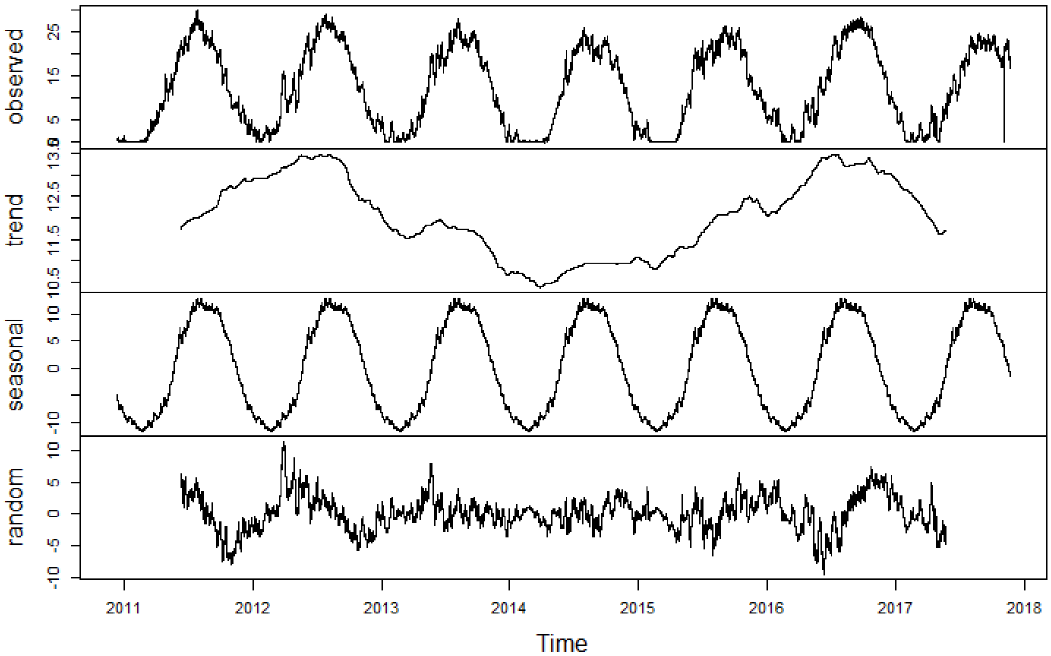
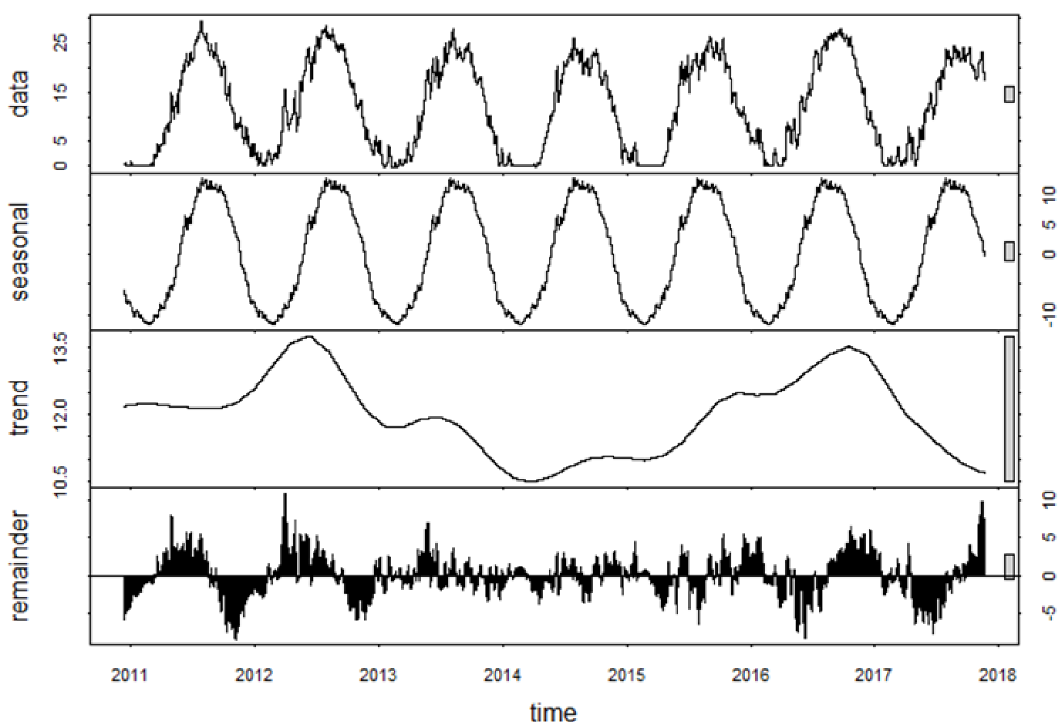


Figure 7 – Genesee River discharge time series decomposition



decomposition temperature smothered daily



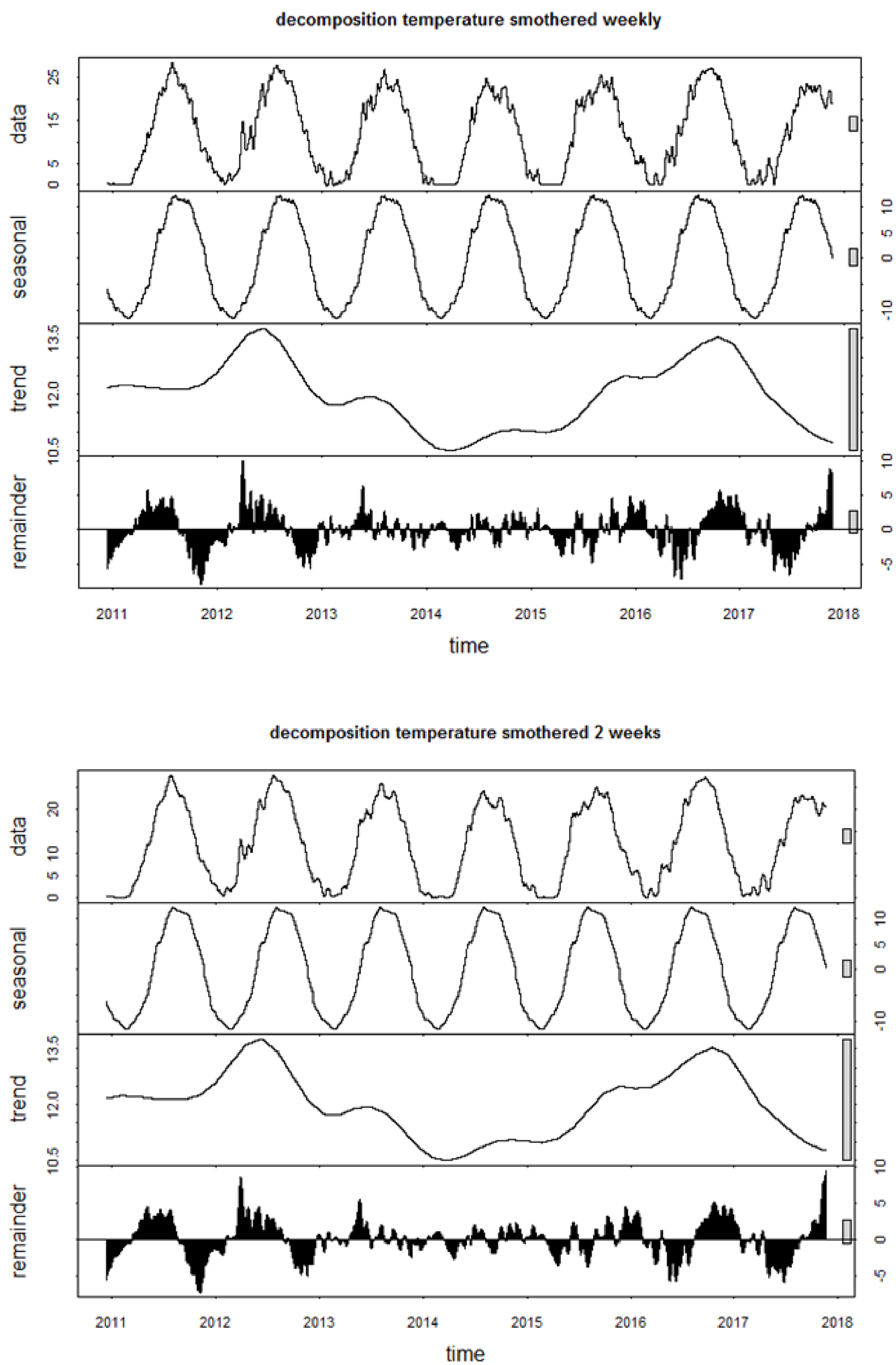
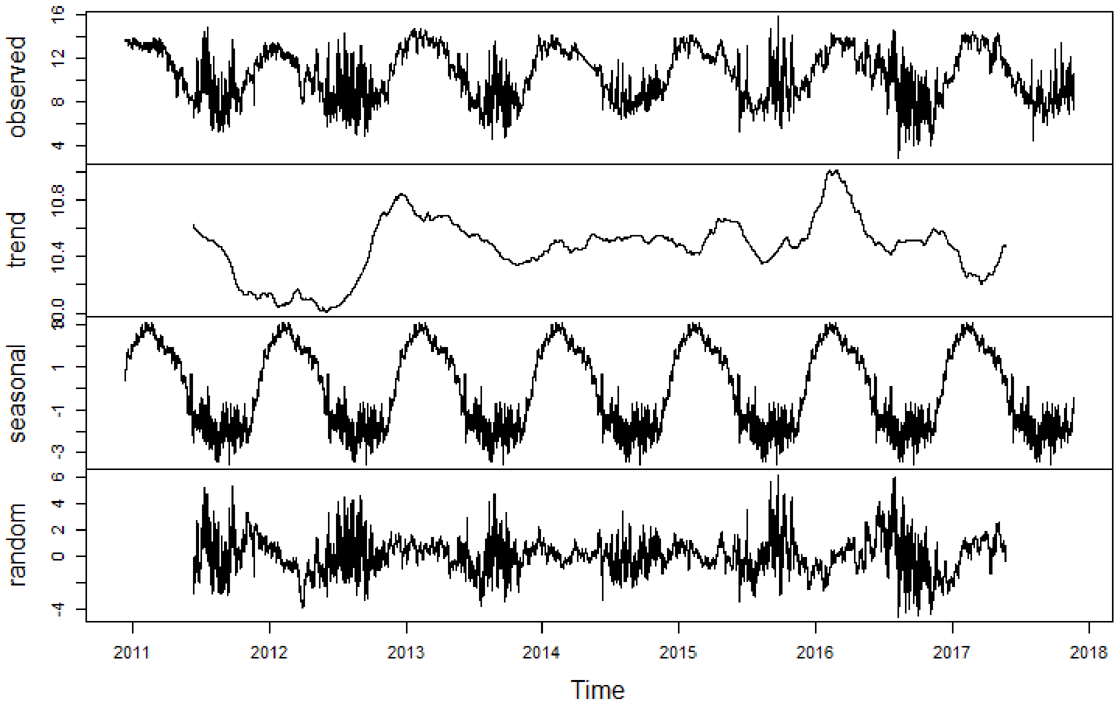
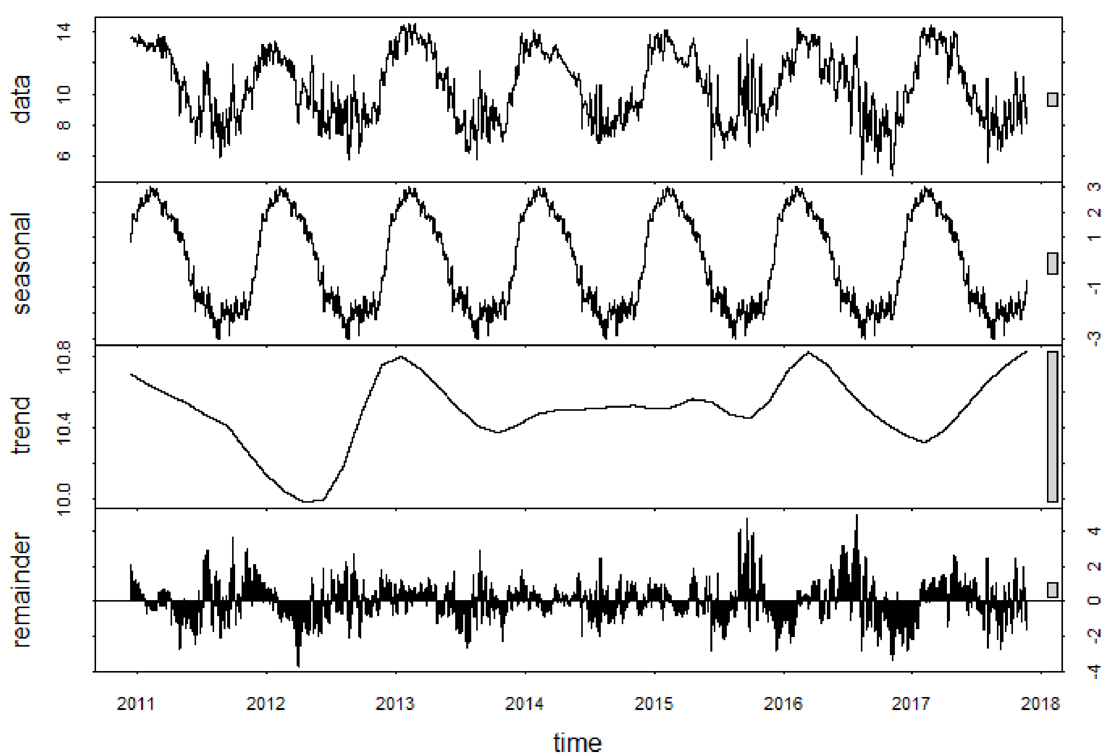
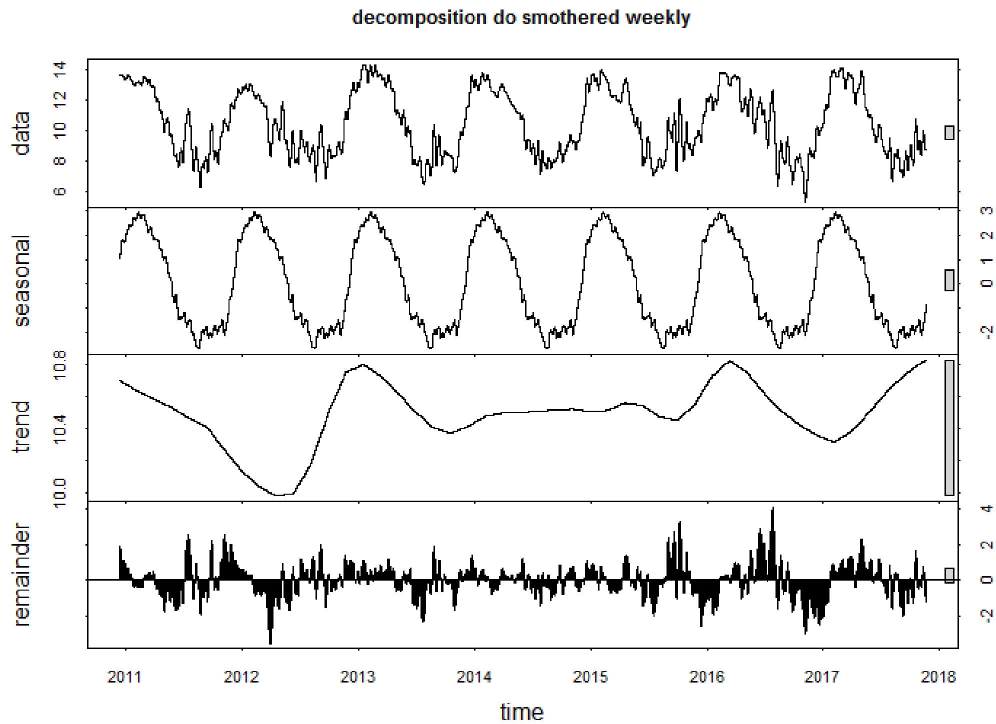


Figure 8 - Genesee River temperature time series decomposition



decomposition DO smothered daily





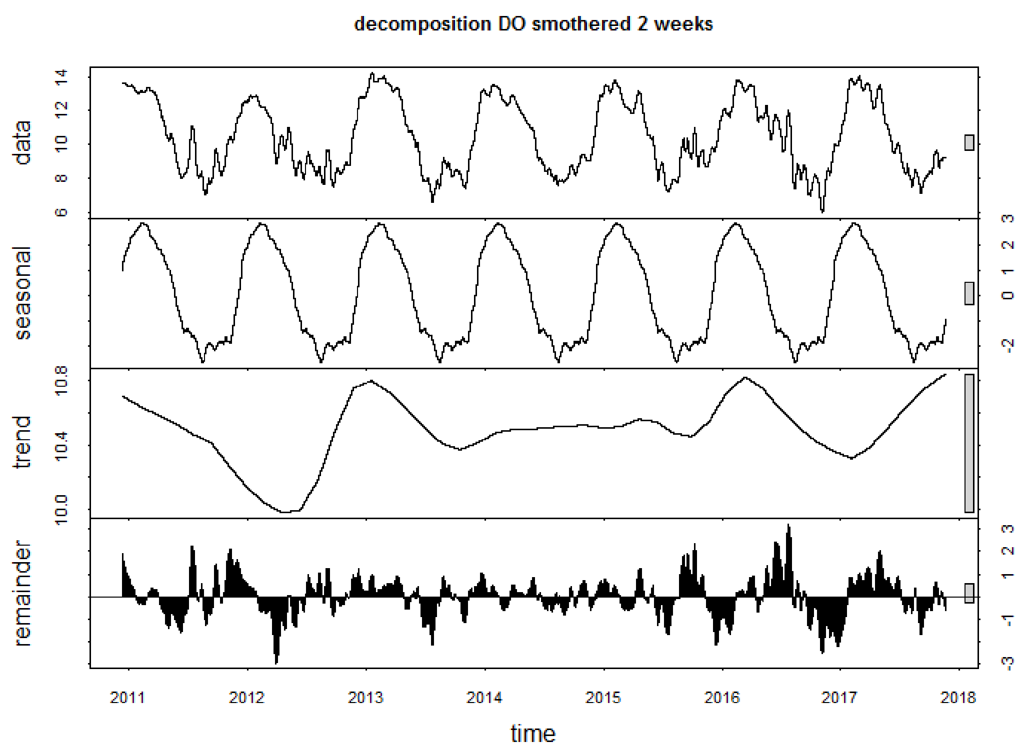


Figure 9 - Genesee River dissolved oxygen time series decomposition

Cross autocorrelation between the DO and temperature and discharge is plotted on Figures 10 and 11.

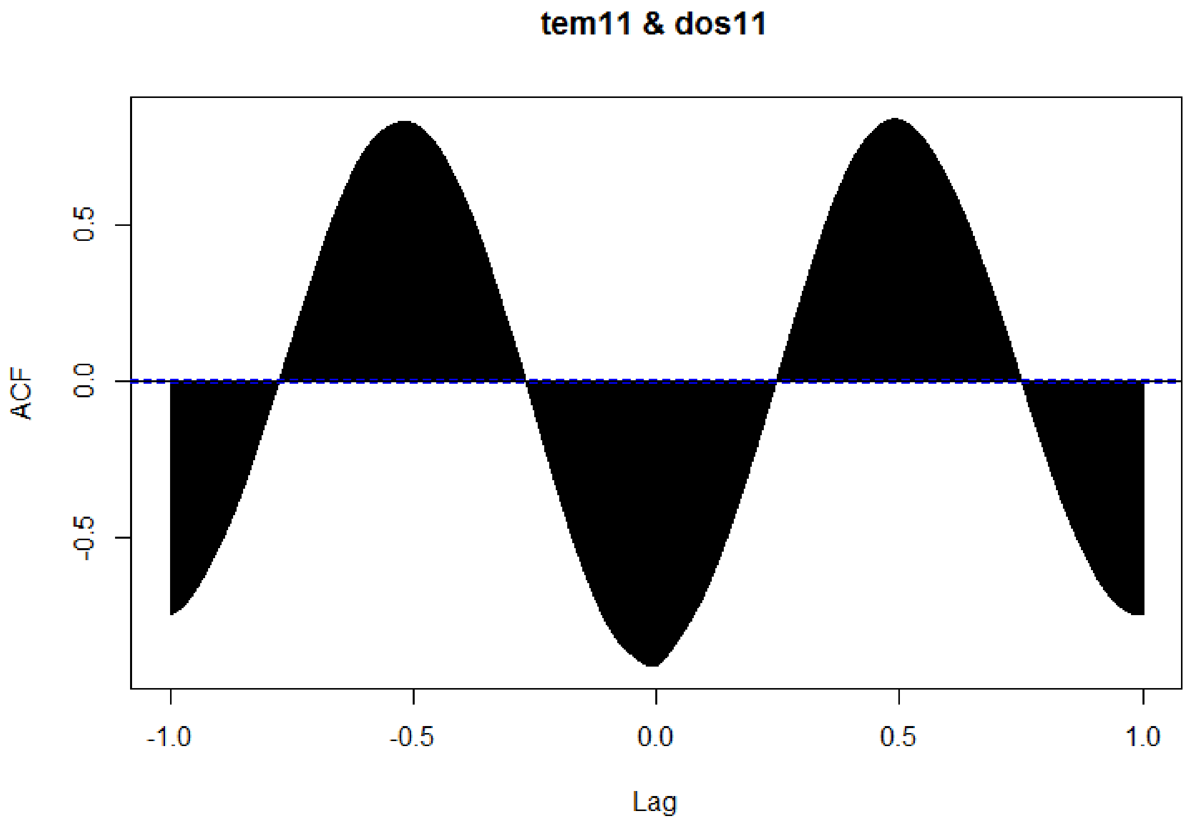


Figure 10 - Genesee River cross correlation between temperature and dissolved oxygen

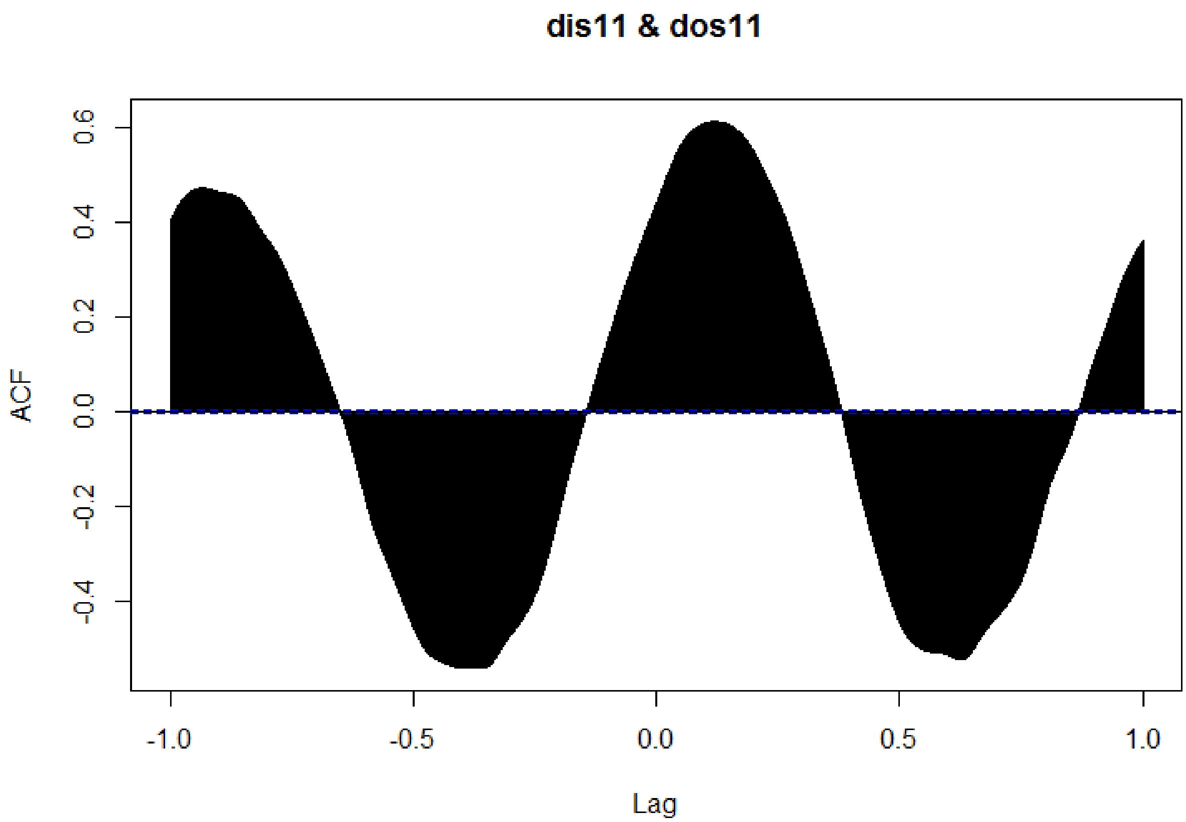


Figure 11 - Genesee River cross correlation between discharge and dissolved oxygen

Proceeding with the analysis, the same steps used for the Genesee River will be performed for the Pecos River. The linear interpolation of the discharge, temperature and DO time series are plotted in figures 12 to 14.

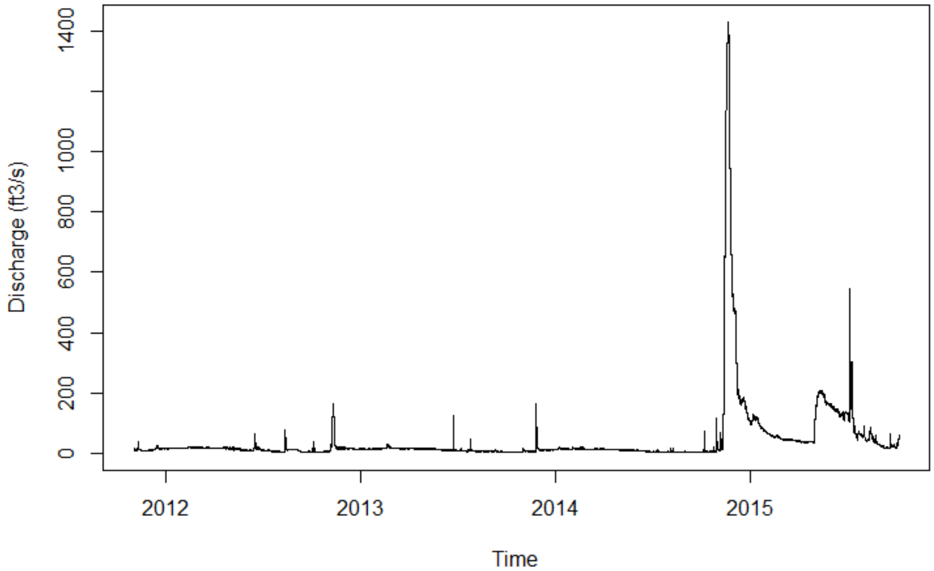


Figure 12 – Pecos River discharge time series with gaps filled by linear interpolation.

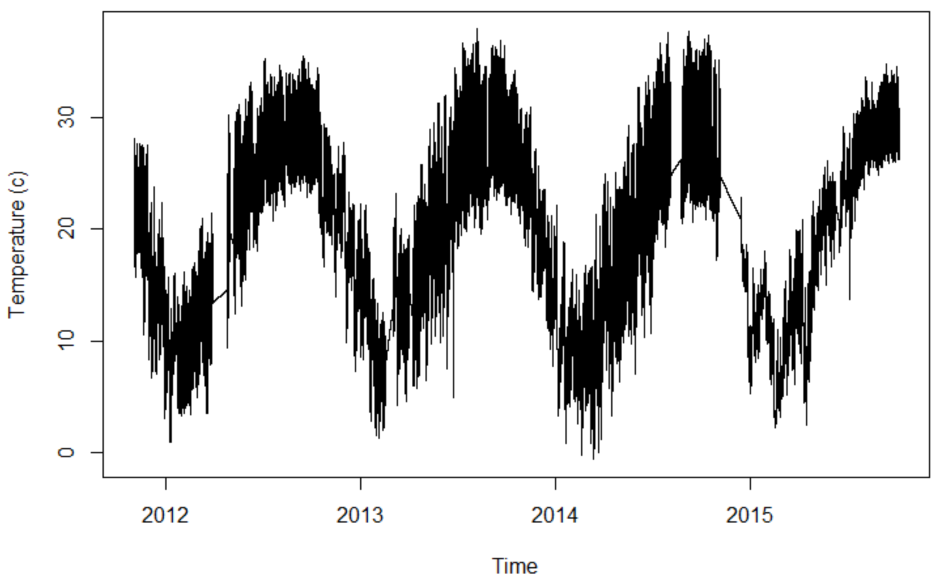


Figure 13 - Pecos River temperature time series with gaps filled by linear interpolation.

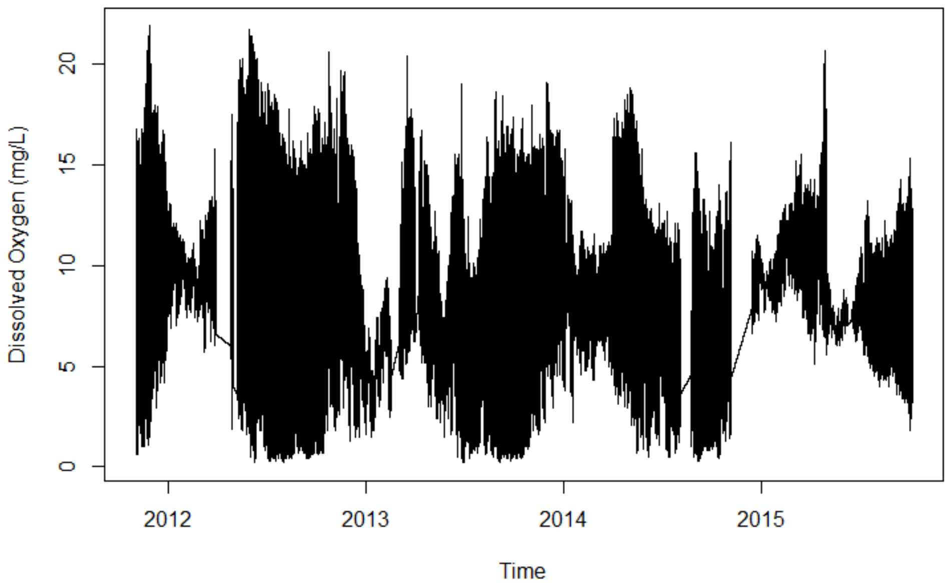


Figure 14 - Pecos River dissolved oxygen time series with gaps filled by linear interpolation.

Analysis of the autocorrelation and partial correlation are shown in figures 4 to 6 for the interpolated time series of the Pecos River.

a)	b)

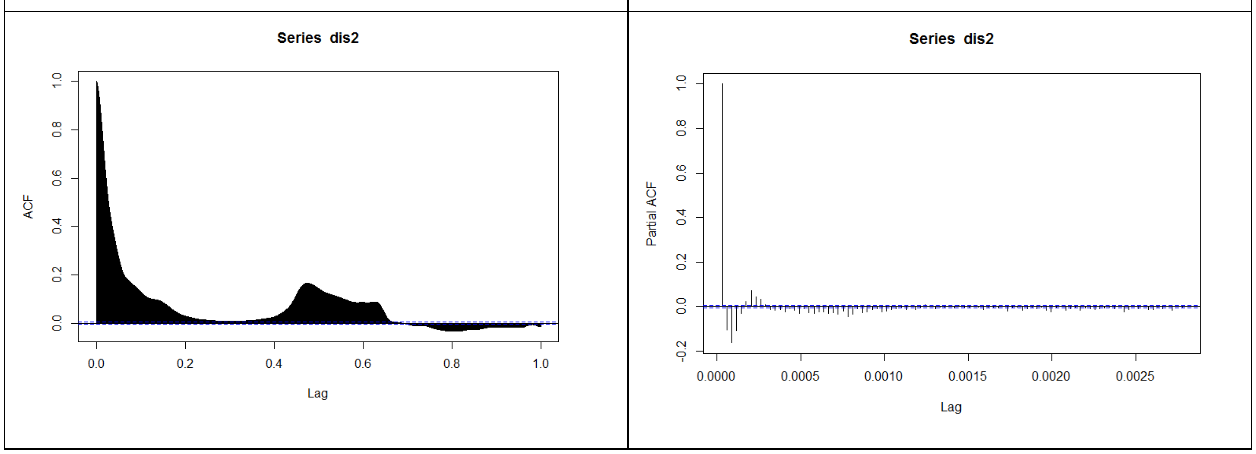


Figure 15 - a) autocorrelation function for the discharge time series of the Pecos River and b) partial autocorrelation function for the discharge time series of the Pecos River.

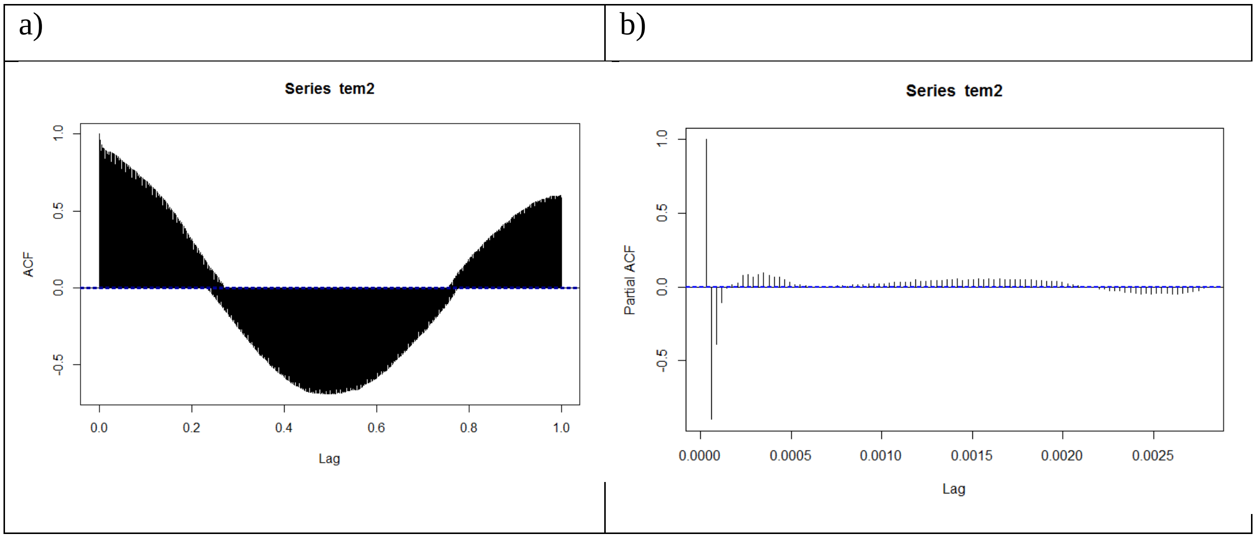


Figure 16 - a) autocorrelation function for the temperature time series of the Pecos River and b) partial autocorrelation function for the temperature time series of the Pecos River.

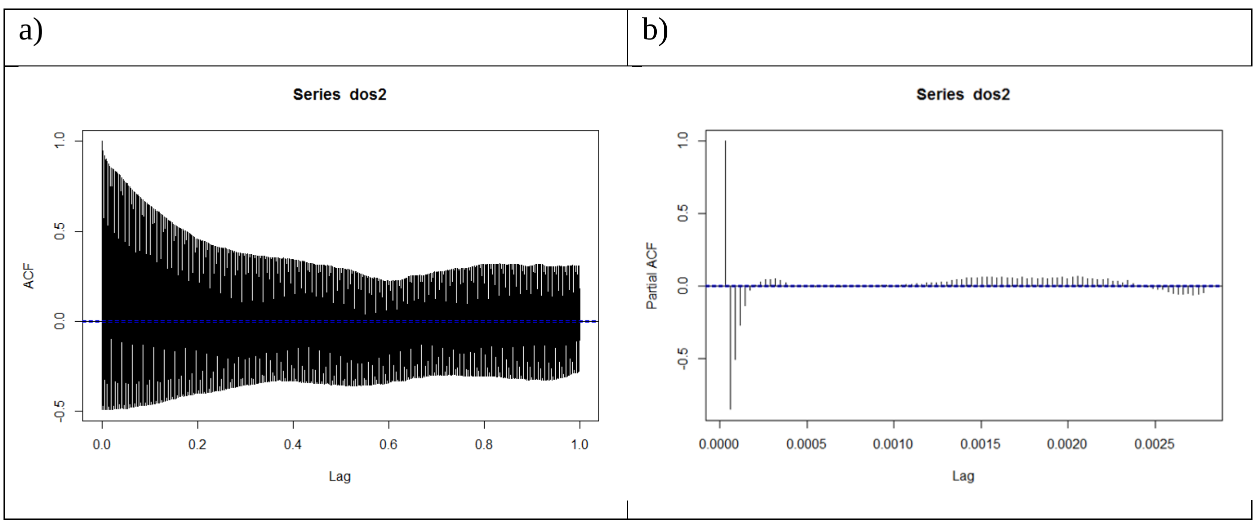
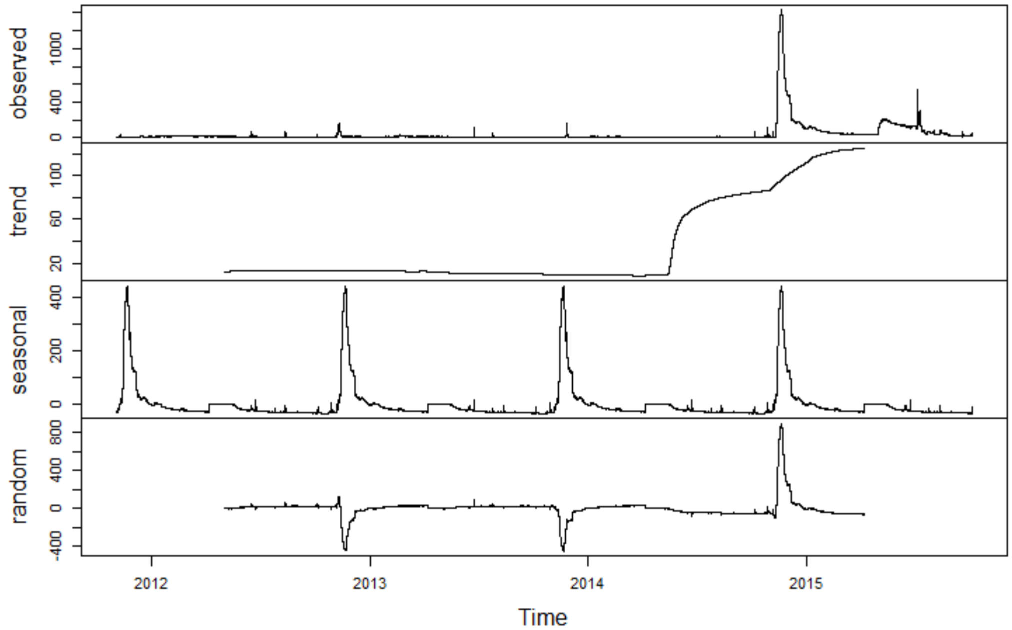
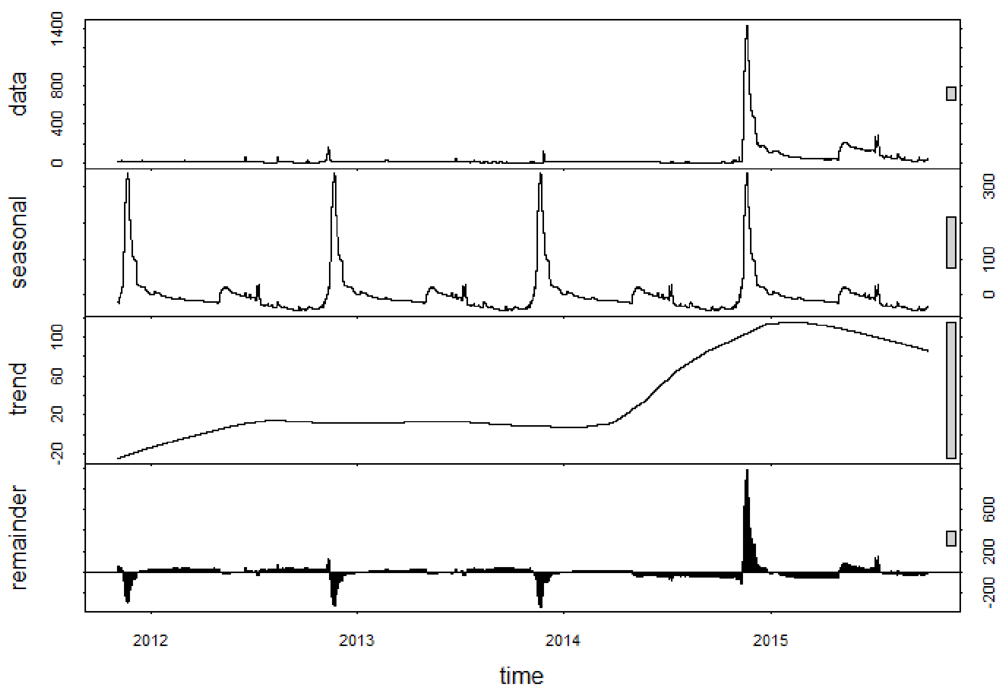


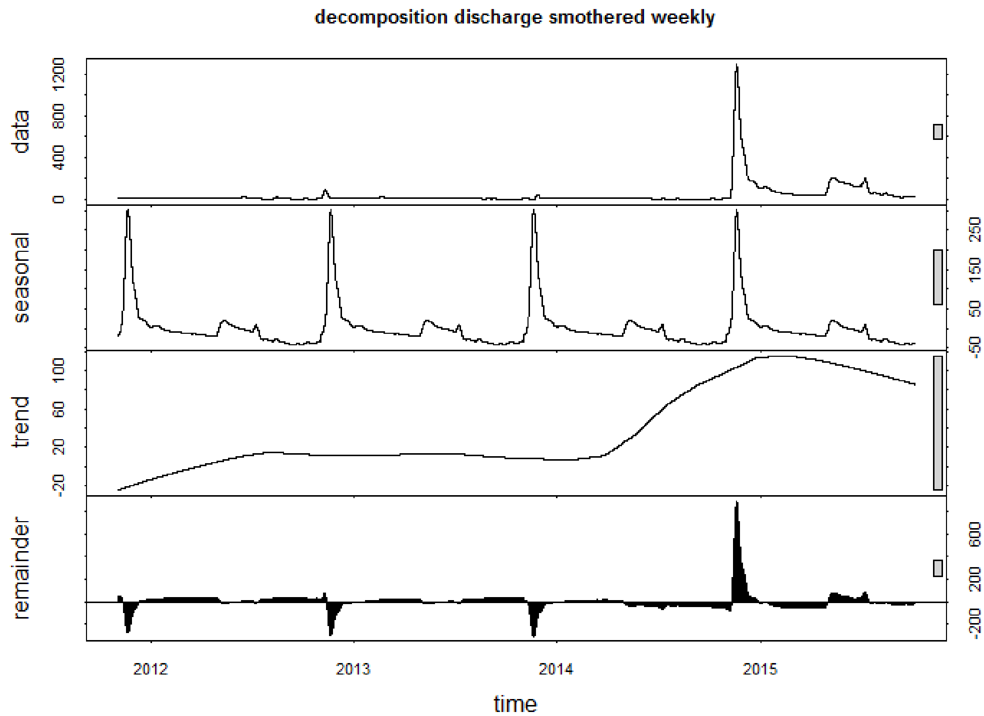
Figure 17 - a) autocorrelation function for the dissolved oxygen time series of the Pecos River and b) partial autocorrelation function for the dissolved oxygen time series of the Pecos River.

Time series decomposition for the discharge, temperature and dissolved oxygen is shown in Figures 18 to 20.



decomposition discharge smothered daily





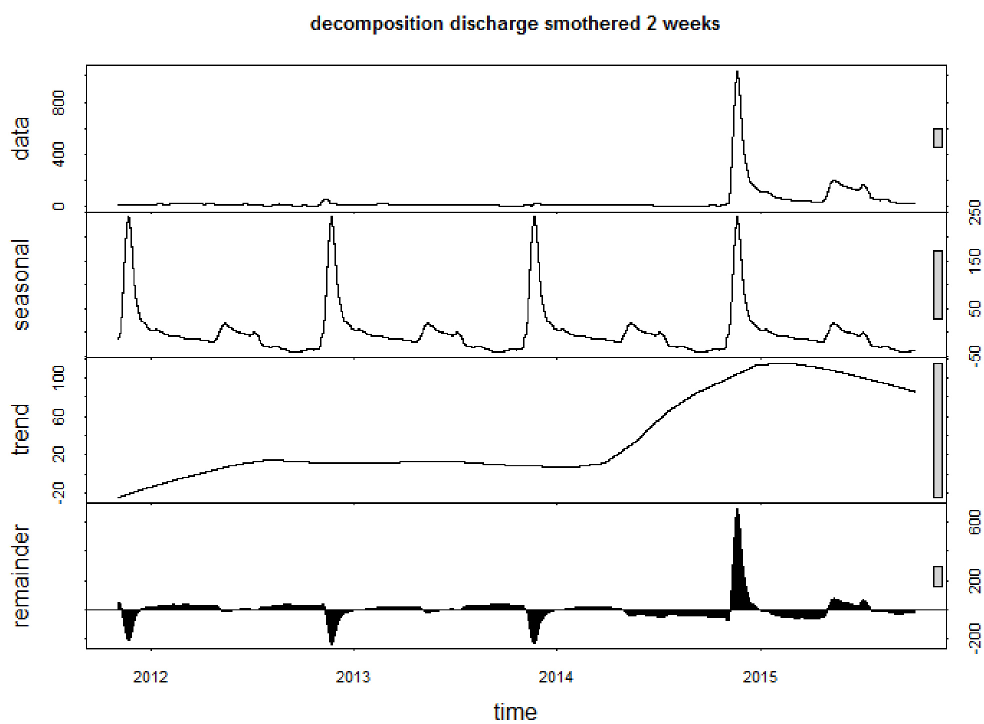
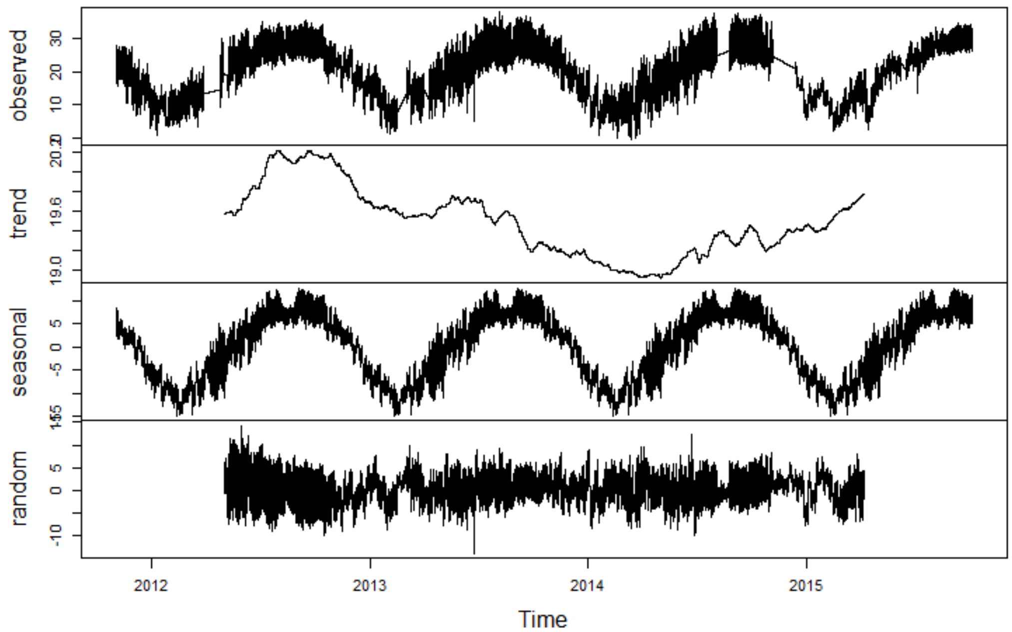
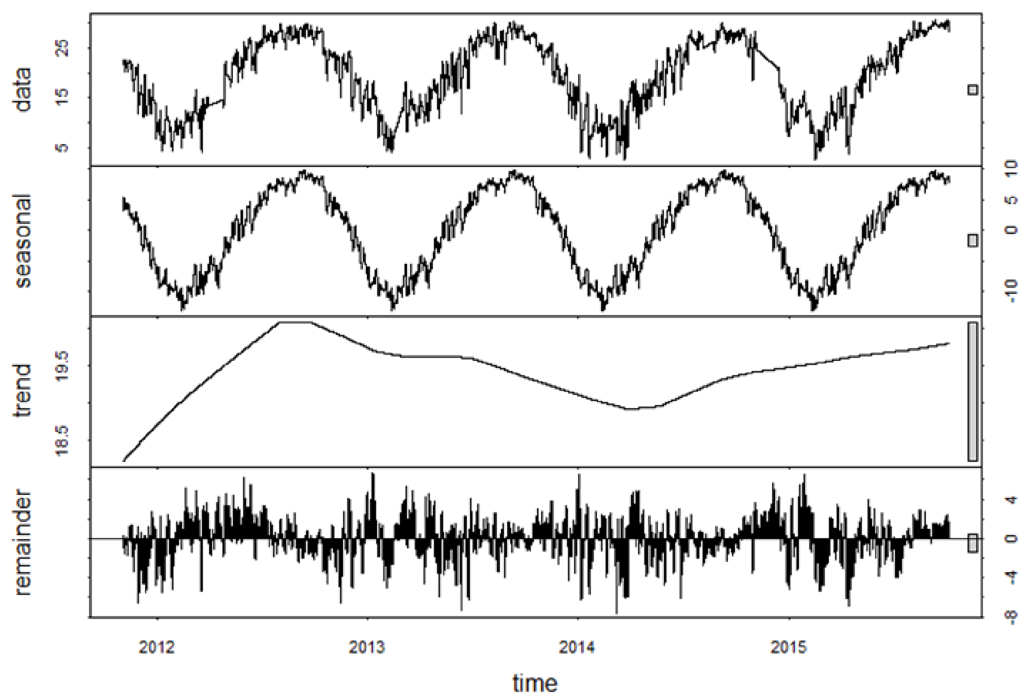
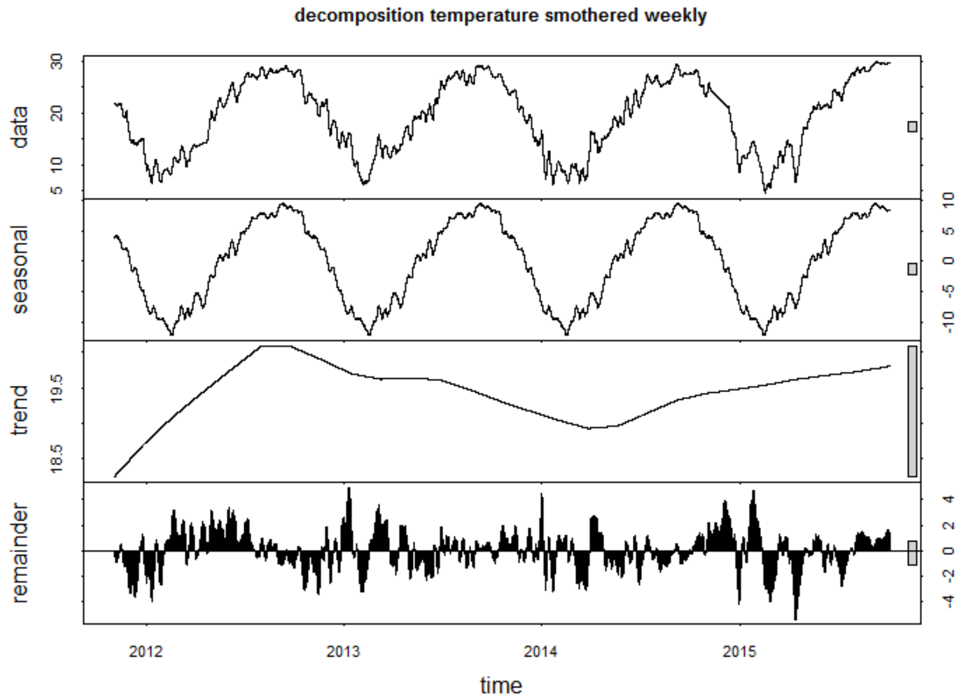


Figure 18 - Pecos River discharge time series decomposition



decomposition temperature smothered daily





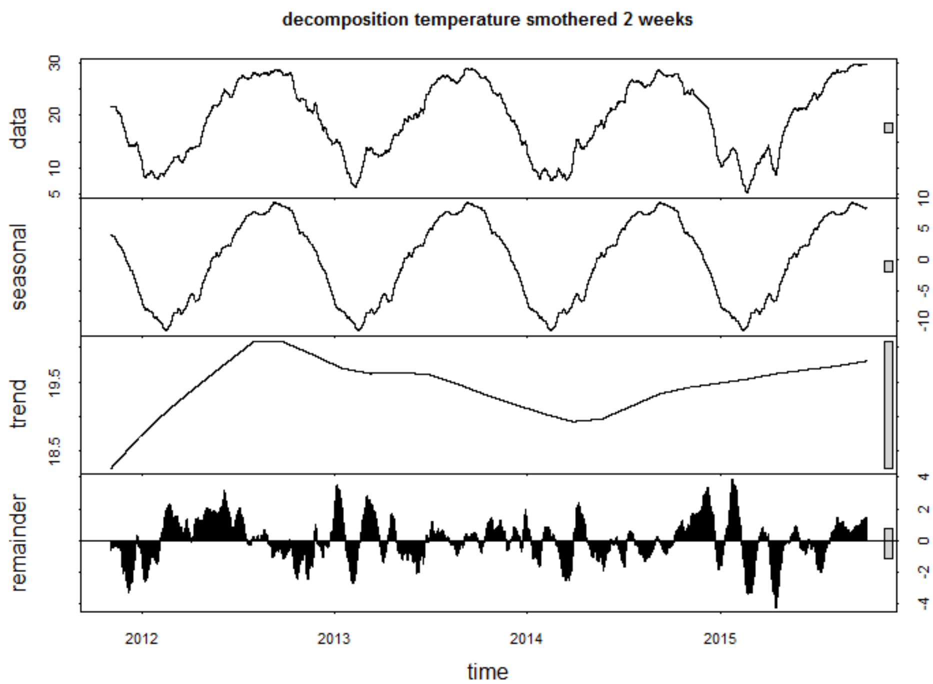
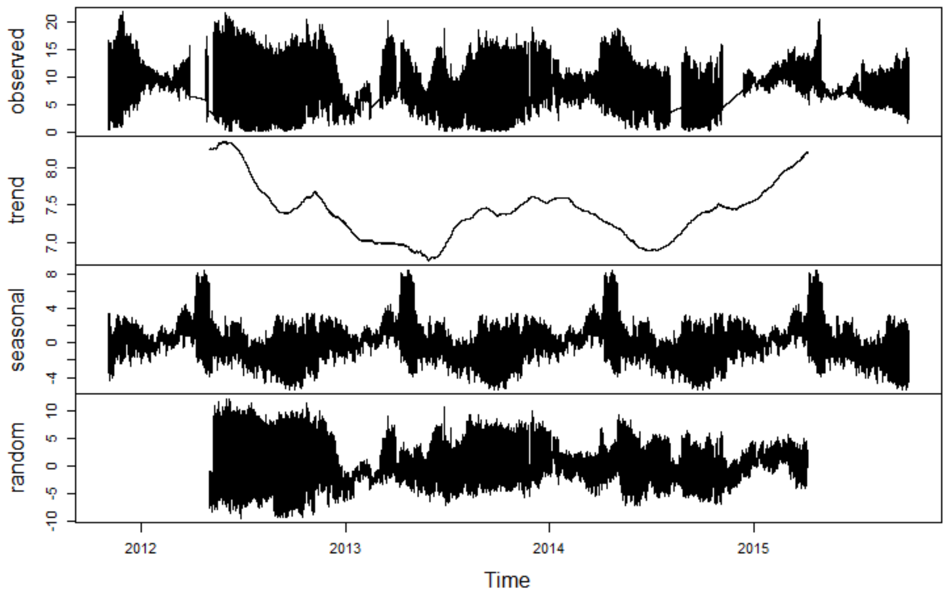
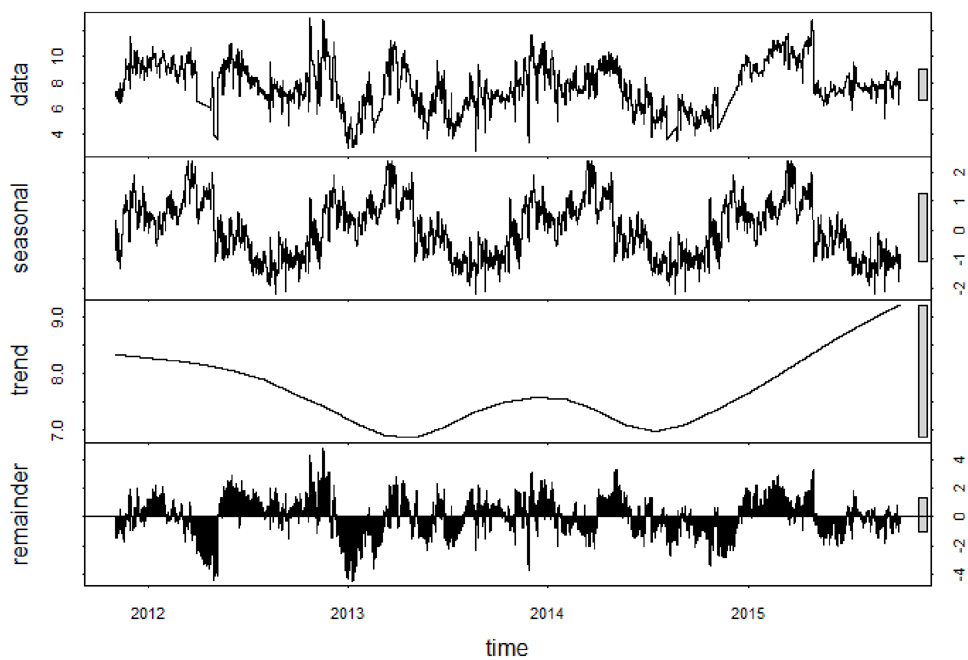
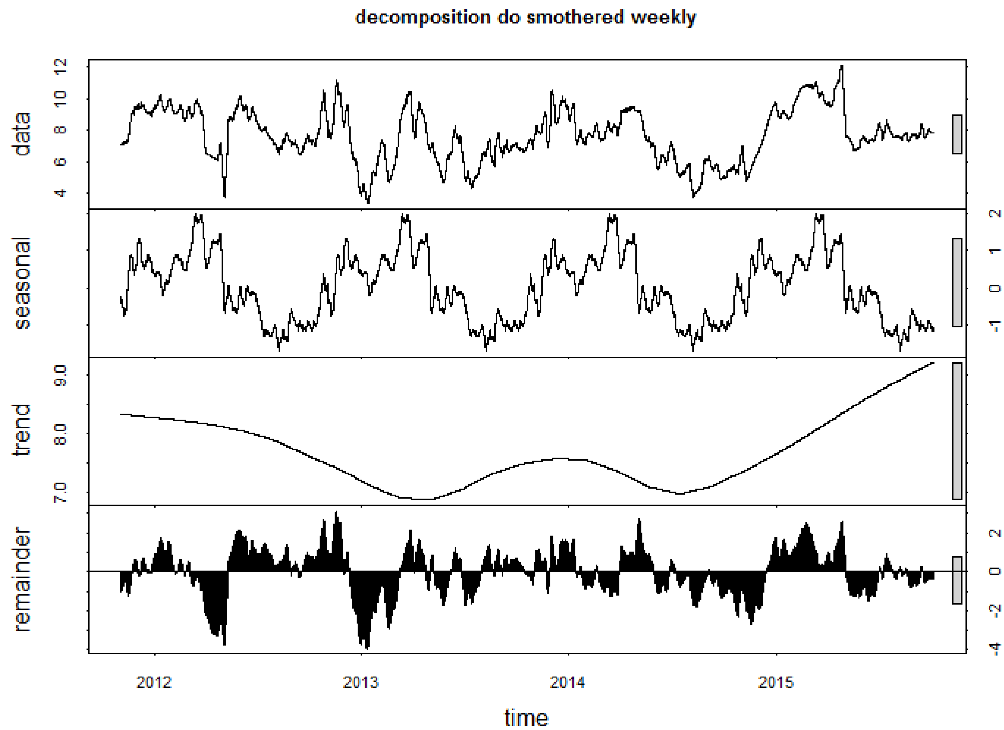


Figure 19 - Pecos River temperature time series decomposition



decomposition DO smothered daily





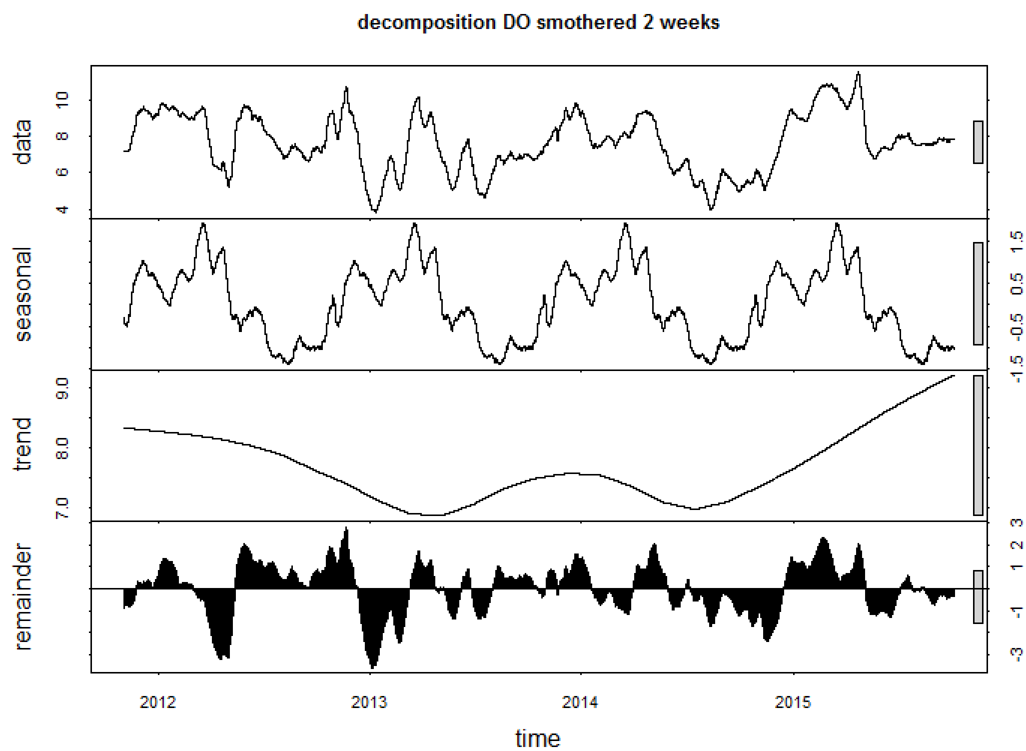


Figure 20 - Pecos River dissolved oxygen time series decomposition

Cross correlation between dissolved oxygen and discharge and dissolved oxygen and temperature is show in Figures 21 and 22.

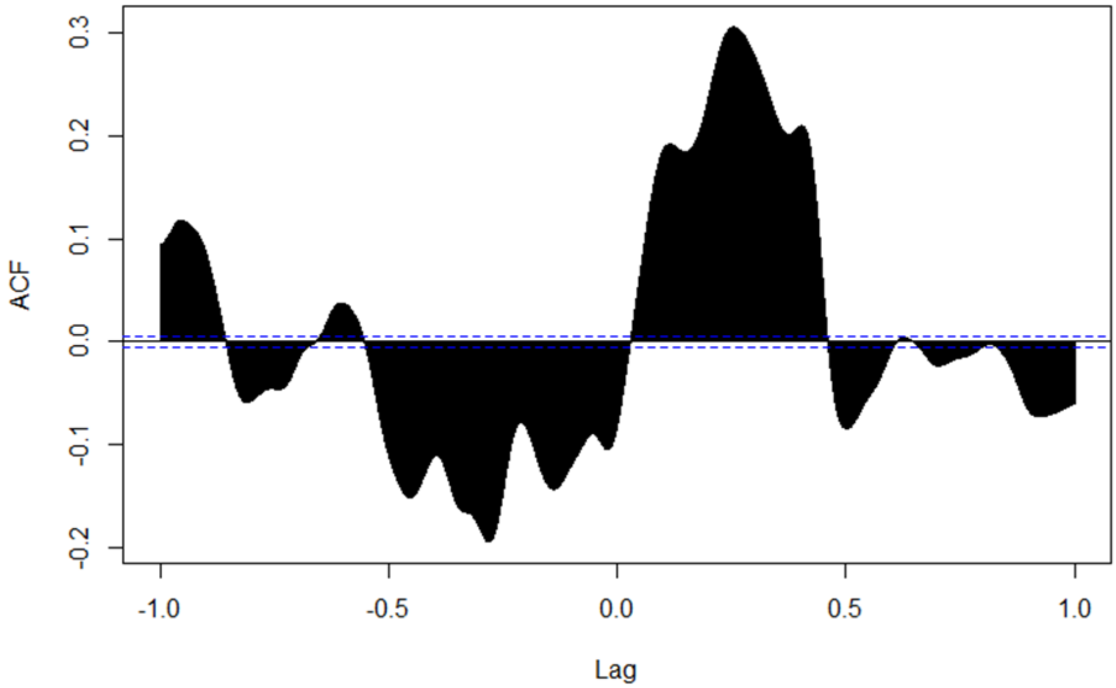


Figure 21 - Pecos River cross correlation between discharge and dissolved oxygen

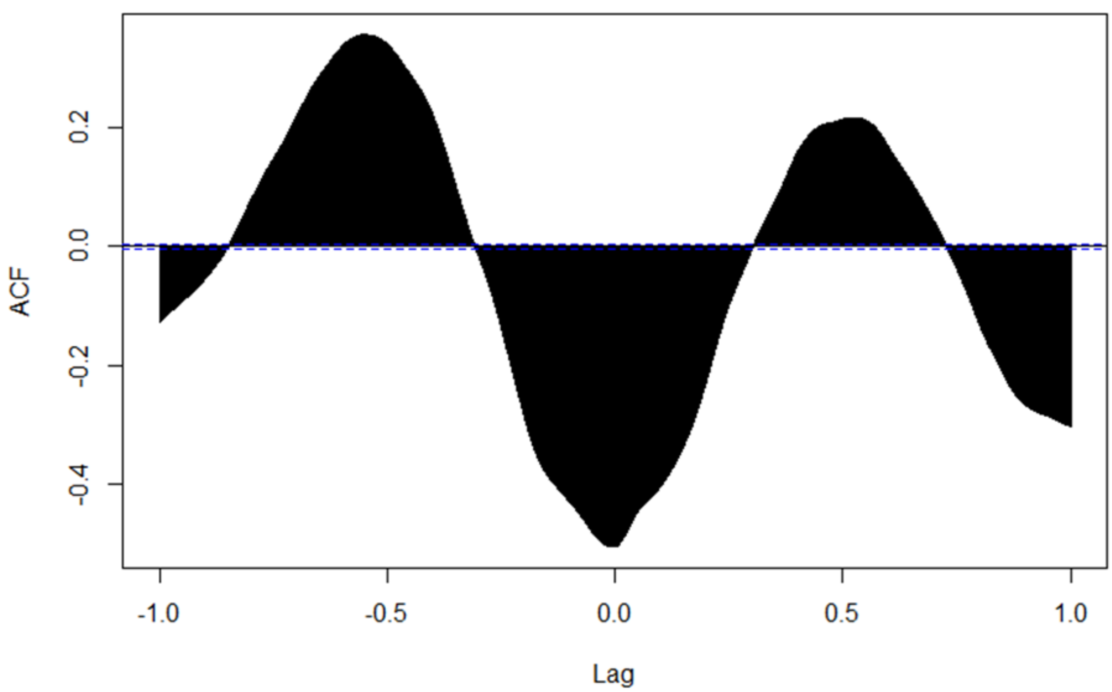


Figure 22 - Pecos River cross correlation between temperature and dissolved oxygen

Discussion and conclusions

Autocorrelation analysis proved that the seasonality in the data is present. Interpretation of the Genesee autocorrelation plots shows the variability of the autocorrelation according to the seasons varying the correlation from positive to negative, with the winter correlation being high than the summer. The same behavior was noticed on the Pecos River data except for the discharge in which presented low flow for the first 4 years and then a spike in the data due to a flooding in the region, interfering the correlation analysis. The daily variation in the DO can be noticed in the autocorrelation plots especially for the Pecos data.

Plots of the partial autocorrelation for a daily analysis showed that the points correlates to the closest 8 measurements (2 hours), resulting of the precision of the data (15 minutes). However, statistically meaningful data is identified along the time, suggesting that the correlation of the data extends beyond a day.

Temperature and dissolved oxygen for both rivers presented a seasonality component in the decomposition (raw data and smothered) in which the dissolved oxygen was inverse correlated to temperature. The discharge factor presented to be a function of precipitation and snow melting since the high discharge values match with those time of the year (summer and spring). On all dissolved oxygen decompositions the remainder/error component of the data presented to be in the same order of magnitude of the seasonality, indicating the variability of the data along time while the trending component did not present relevant in any of the analysis.

Algae bloom is common to be seen in the Genesee River basin. Phosphorous concentration coming from non-point sources in the basin ending up in the lake Ontario. Lake Ontario Biodiversity Conservation Strategy classifies many tributaries in the basin as “priority action” for controlling pollutants to the river basin (Genesee River Watch, assessed in 12/07/2017). The algae bloom in this river can be identified in the DO time series during the summer periods, presenting a lot of variation during the day/night time, while the algae are doing photosynthesis during the day and reduce the oxygen concentration during night when it stops.

As presented in Israel et al. (2014) presence of algae in the Pecos River is disturbing the natural state of the waters. Golden Algae was identified to change the salinity of the river (measured by specific conductance between 4,408 to 73,786 μS/cm from January 2012 to July 2013. The presence of algae can explain the high variability of the dissolved oxygen concentration during a short period.

Dissolved oxygen presented a correlation with temperature of 70 % at the Genesee River and up to 45% for Pecos River. Correlation between discharge and dissolved oxygen at the Genesee River reaches 60% and 30% for Pecos River, however the flooding event (peak in discharge) compromised the analysis.

Variation of the dissolved oxygen was greater in the Pecos River when compared to the Genesee. Presence of algae bloom in a small flow (less than 100 ft^3/s) river can influence much more than a high flow river (10000 ft^3/s) since algae tends to be on the water surface, the deeper the river DO can be dissolved.

It can be conclude that the main driver for DO increase is the temperature (weather). Either controlling the water oxygen capacity when low temperatures or allowing the algae bloom during warm temperatures.

Limitations

1. The 2 data used in this work presented high precision measurement (every 15 minutes) but they were not long enough (comparing to a river change time scale) to understand any trending as presented in the data.
2. The flooding event compromised the discharge analysis for the Pecos River as noticed in the seasonality part of the decomposition
3. Presence of NAs on the data that had to be interpolated linearly created errors in the small scale analysis.

References

Israel, N.M.D., VanLandeghem, M.M., Denny, S., Ingle J., Patino R. (2014). “Golden Alga presence and abundance are inversely related to salinity in a high-salinity river ecosystem, Pecos River, USA”. Harmful Algae 39: 81-91.

Genesee River Watch - https://geneseeriverwatch.org/index.php/genesee-river/the-genesee-river-watershed, assessed in 12/07/2017.

Hyndman, R.J., Athanasopoulos, G., (2012). “Forecast Principles and Practice” available at: https://www.otexts.org/fpp/

Rcode

As the CSV file imported to R environment had the columns named equally, the same code could be used for both datasets.

```
pecos <- read_csv("C:/Users/geol659-2/Downloads/pecos.csv")
library(imputeTS)
dis=ts(pecos$discharge,start=c(2011,28800),frequency=34560)
dos=ts(pecos$do,start=c(2011,28800),frequency=34560)
tem=ts(pecos$temperature,start=c(2011,28800),frequency=34560)
dis2=na.interpolation(dis,option ="linear")
dos2=na.interpolation(dos,option ="linear")
tem2=na.interpolation(tem,option ="linear")
plot.ts(dis2,ylab="Discharge (ft3/s)")
plot.ts(dos2,ylab="Dissolved Oxygen (mg/L)")
plot.ts(tem2,ylab="Temperature (c)")
acf(dis2)
acf(dos2)
acf(tem2)
acf(dis2,34560)
acf(dos2,34560)
acf(tem2,34560)
pacf(dis2,96)
pacf(dos2,96)
pacf(tem2,96)
pacf(dis2,34560)
pacf(dos2,34560)
pacf(tem2,34560)
library("TTR")

dis2ema50=EMA(dis2,n=50)
dos2ema50=EMA(dos2,n=50)
```

```
tem2ema50=EMA(tem2,n=50)
plot(dis2ema50)
plot(dos2ema50)
plot(tem2ema50)
library(xts)
ccf(dis2,dos2)
ccf(tem2,dos2)
ccf(dis2,dos2,20000)
ccf(tem2,dos2,20000)
spectrum(dis2,method="ar")
spectrum(dos2,method="ar")
spectrum(tem2,method="ar")
library(imputeTS)
x=decompose(dis2)
plot(x)
x1=decompose(dos2)
plot(x1)
x3=decompose(tem2)
plot(x3)
install.packages("fpp")
require (fpp)

dis3=ma(dis2,order=672)
dos3=ma(dos2,order=672)
tem3=ma(tem2,order=672)

plot(dis3)
plot(dos3)
plot(tem3)
x4=decompose(dis3)
plot(x4)
```

```
x5=decompose(dos3)
plot(x5)
x6=decompose(tem3)
plot(x6)
dis4=na.interpolation(dis3,option ="linear")
dos4=na.interpolation(dos3,option ="linear")
tem4=na.interpolation(tem3,option ="linear")
x7=stl(dis4,"per")
x8=stl(dos4,"per")
x9=stl(tem4,"per")
plot(x7,main ="decomposition discharge smothered weekly")
plot(x8,main ="decomposition do smothered weekly")
plot(x9,main ="decomposition temperature smothered weekly")
tem5=ma(tem2,order=96)
tem6=na.interpolation(tem5,option ="linear")
x10=stl(tem6,"per")
plot(x10,main ="decomposition temperature smothered daily")
tem10=ma(tem2,order=1344)
tem11=na.interpolation(tem10,option ="linear")
x15=stl(tem11,"per")
plot(x15,main ="decomposition temperature smothered 2 weeks")

dis5=ma(dis2,order=96)
dis6=na.interpolation(dis5,option ="linear")
x20=stl(dis6,"per")
plot(x20,main ="decomposition discharge smothered daily")
dis10=ma(dis2,order=1344)
dis11=na.interpolation(dis10,option ="linear")
x21=stl(dis11,"per")
plot(x21,main ="decomposition discharge smothered 2 weeks")
```

```
dos5=ma(dos2,order=96)
dos6=na.interpolation(dos5,option ="linear")
x30=stl(dos6,"per")
plot(x30,main ="decomposition DO smothered daily")
dos10=ma(dos2,order=1344)
dos11=na.interpolation(dos10,option ="linear")
x31=stl(dos11,"per")
plot(x31,main ="decomposition DO smothered 2 weeks")

ccf(dos11,tem11,34560)
ccf(dos11,dis11,34560)
```