# Workshop 3_Timeseries Models

Donna Marain

1/24/2020

## Objectives and Site Information

The primary objective of this analysis is to determine which environmental variable best correlates with the patterns seen in net ecosystem exchange (NEE) at the TS/Ph-7 eddy covariance flux tower located at 25.19080491, -80.63910514 in Everglades National Park. This investigation is important because this flux tower was the first such tower to be installed in a mangrove forest back in 2003.[1] This tower therefore provides the longest dataset recording mangrove forest productivity.

## Methods

I fit autoregressive integrated moving average (ARIMA) models from R package forecast to the NEE data from calendar year 2018 at TS/Ph-7. I only tested six models, four of which were provided in class and two of which were identified by Barr et al (2010) as the most important environmental drivers in this system: air temperature and phosynthetically active radition (PAR).

### Statistical Analysis

Environmental variables were cleaned of outliers using the tsclean() function. Cleaned variables were then tested for stationarity with an augmented Dickey-Fuller test. If variables failed this test, cleaned variables were differenced and the ADF test re-run to see if the differenced clean variables were stationary. Stationary variables were then added to an auto-ARIMA function as an explanatory variable for nee.

The models produced by the auto-ARIMA function were evaluated by their Akaike's Information Criterion (AIC). The model with the smallest AIC is considered the most parsimonious, meaning it fits the data best with the fewest parameters. By convention, models with AICs that differ by less than 2 are considered the most supported by the data, those that differ between 2 and 7 are considered to be somewhat supported by the data, and those that differ by more than 7 are considered to have little support in the data.

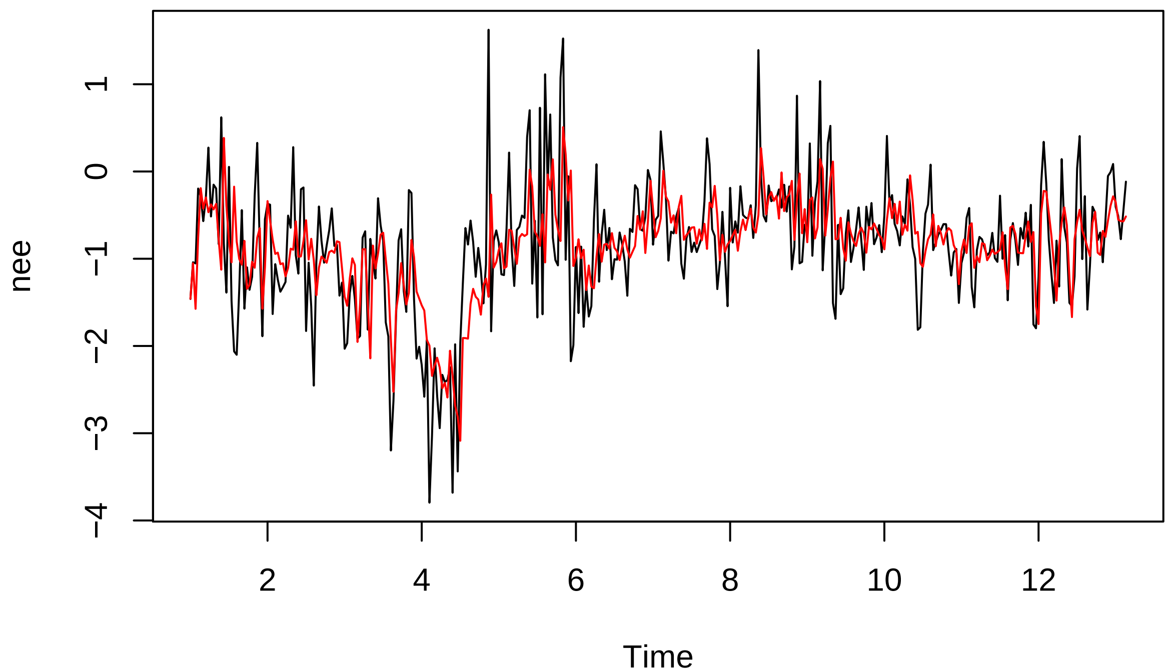
## Results (minimum of 1 plot and one table)

The auto-ARIMA model which used air temperature as its explanatory variable had the best fit to the data (see Table 1).

| Model Name | Model parameters | df | AIC |
|---|---|---|---|
| arima.nee1 | y=nee, no x | 9 | 703.09 |
| arima.nee2 | y=nee, no x, lag=10 | 18 | 704.77 |
| arima.nee3 | y=nee, x=salinity | 12 | 706.19 |

| Model Name | Model parameters | df | AIC |
|---|---|---|---|
| arima.nee4 | y=nee, x= extreme salinity | 9 | 700.57 |
| arima.nee5 | y=nee, x= par | 9 | 683.68 |
| arima.nee6 | y=nee, x= air temperature | 8 | 660.8 |

Table 1. Model results.




Plot 1: Best model (arima.nee6) predictions in red plotted against the raw nee data in black.

## Discussion (1 paragrapgh)

The discrepancies between the red and black lines in Plot 1 shows that my best model is not a perfect fit to the raw data, but the AIC values show it is vastly superior to any other model I ran. Models that include multiple explanatory variables might provide a better fit to the data, but at the same time a perfect fit is not desirable because then the model would not be generalizable to other times and places.

#References 1. https://fcelter.fiu.edu/research/key-findings-productivity-gradients-in-mangroves/index.html

Barr, J.G., V. Engel, J.D. Fuentes, J.C. Zieman, T.L. O'Halloran, T.J. Smith, G. Anderson. 2010. Controls on mangrove forest-atmosphere carbon dioxide exchanges in western Everglades National Park. Journal of Geophysical Research 115: G02020.