Workshop 3: Timeseries Models

Rachel Prokopius

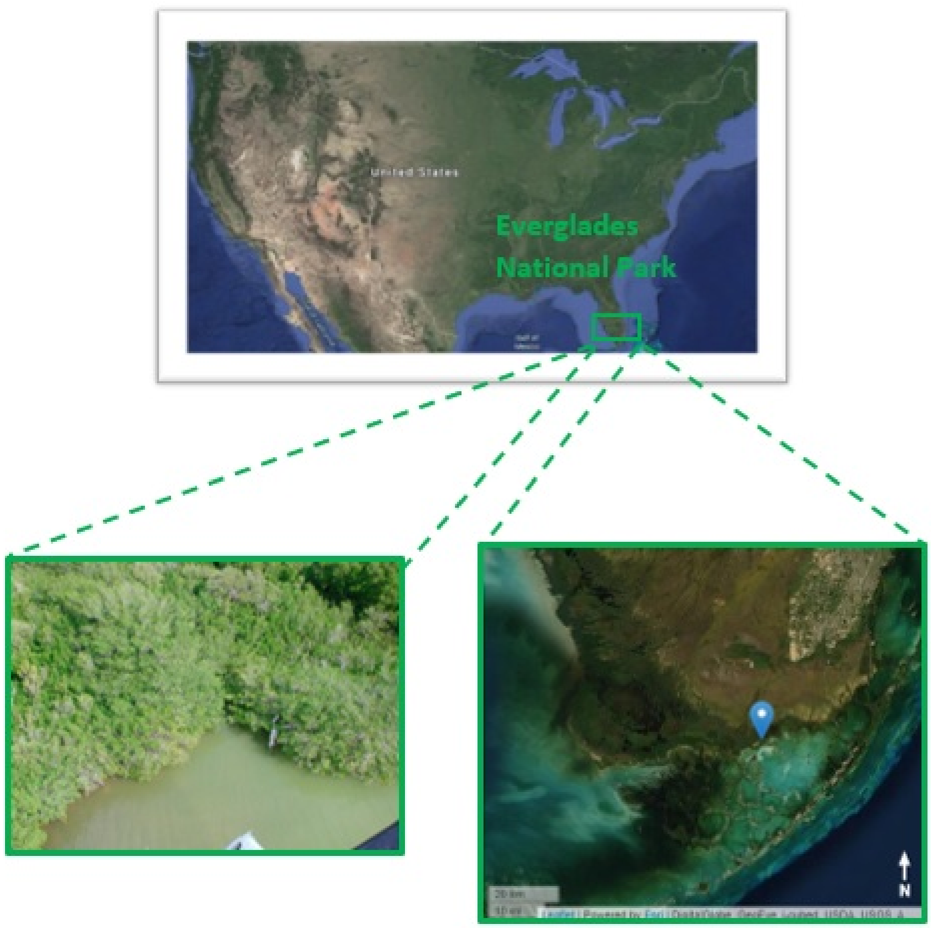
1/31/2020

Objectives

The primary objectives of this analysis is to determine and account for the temporal variation and seasonality in a data set that violates the assumed independence of the data set. Once this variation is accounted for, it can be determined which independent variables best explain and can predict future values of the dependent variable.

Methods

Site Information



Site images and map locations of TS/Ph-7a

The research area is the TS/Ph-7a site, one of the 23 research sites for the Florida Coastal Everglades Long-term Ecological Research program. The site sits at the mouth of Taylor Slough and the GPS coordinates for the site are:

Latitude: 25.19080491 Longitude: -80.63910514

The site is comprised of limestone bedrock with mangrove forests serving as the primary vegetation. The site experiences distinct wet (June-November) and dry (December-May) seasons. An Environmental Measurement Station Eddy Flux Tower (EMS) is located at the site and recorded the data set for this analysis, which includes NEE (g C m2 day−1), total daily photosynthetically active radiation (W m−2 day−1), air temperature (°C), water temperature minimum and maximum (°C), and water salinity minimum, maximum, and mean from January 1st through December 31st of 2018.

Site data and images obtained from https://fce-lter.fiu.edu/research/sites/index.php (https://fce-lter.fiu.edu/research/sites/index.php) and Google Maps

Statistical Analysis

Four R packages were loaded and used to analyse the data: zoo, tsclean, forecast and xts.

The dependent variable (Net Ecosystem Exchange, or NEE) was graphed as a time series to determine any outliers, and if outliers were present the tsclean package was run to remove them. Once cleaned, the dependent variable was decomposed and tests for stationarity ($p<0.05$) and autocorrelation were performed and plotted for visual analysis. If the data was not stationary, the function diff() was used to difference the time series and the test for stationarity was rerun. An ARIMA model was fit to the data set and, if autocorrelations were present, the model was rerun with altered p redisual values that corresponded to spikes in the autocorrelation plots until the resulting plots no longer showed autocorrelations. Akaike Information Criteria (AIC) values of the models were compared using the AIC function, and Ljung-Box tests of independence were performed to ensure the residuals of the resulting ARIMA model were not autocorrelated ($P>0.05$). The forcast function was used to determine projections of NEE with the ARIMA models created from the data set.

The above steps were repeated with the independent variables of maximum salinity (Max Salinity) and air temperature (Air Temperature °C).

Results

Table 1

Data series	Dickey-Fuller Value	Stationary P-value	ARIMA df	ARIMA AIC value	Ljung-Box test Q* value	df	P-value
NEE	-4.4703	<0.01	18	704.7739	23.109	19	0.2326
diff(Max Salinity)	-8.9764	<0.01	9	700.5729	37.929	28	0.09976
diff(Air Temperature °C)	-9.979	<0.01	8	660.8001	37.296	29	0.1388

Output values for tests of stationarity (Dickey-Fuller Value), ARIMA model AIC value and tests of independence (Ljung-Box Test) for dependent variable and independent variable parameters on dependent variable

Table 1

All data series in Table 1 had outliers present which were removed through the tsclean package. The test of stationarity for the NEE model without independent variable addition (Model NEE) had a signifiant p-value, and could therefore be used without differencing the series.The maximum salinity and air temperatures were originally determined not to be stationary (Dickey-Fuller = -2.1936, p-value = 0.4952 and Dickey-Fuller = -2.1179, p-value = 0.5272 respectively) and were differenced in order to obtain stationary data series (Table 1). All tested data series showed independence through the Ljung-Box test with p-values greater than 0.05 (Table 1).

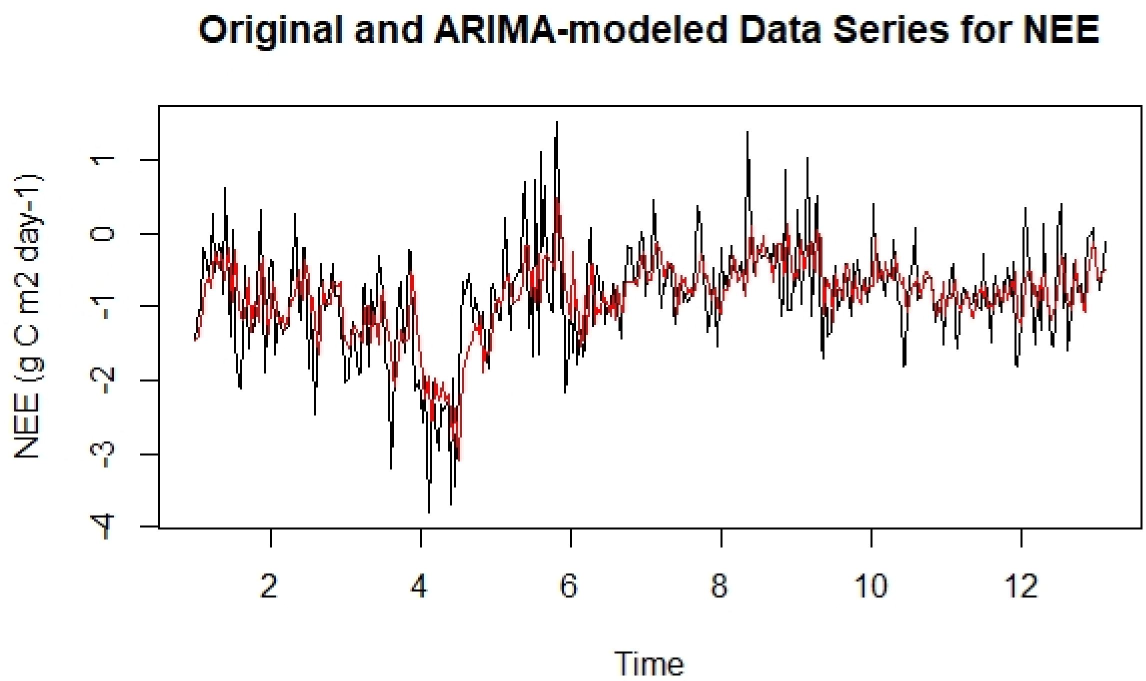


Figure 1. Time series of NEE plotted with original data series (black) and ARIMA-modeled data series (red)

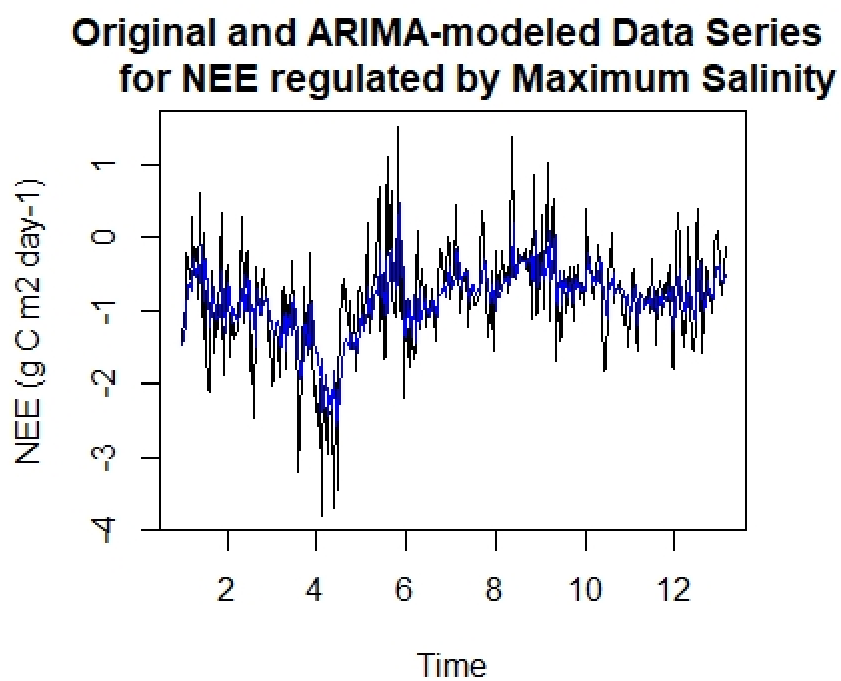


Figure 2. Time series of NEE plotted with original data series (black) and ARIMA-modeled data series regulated by maximum salinity (blue)

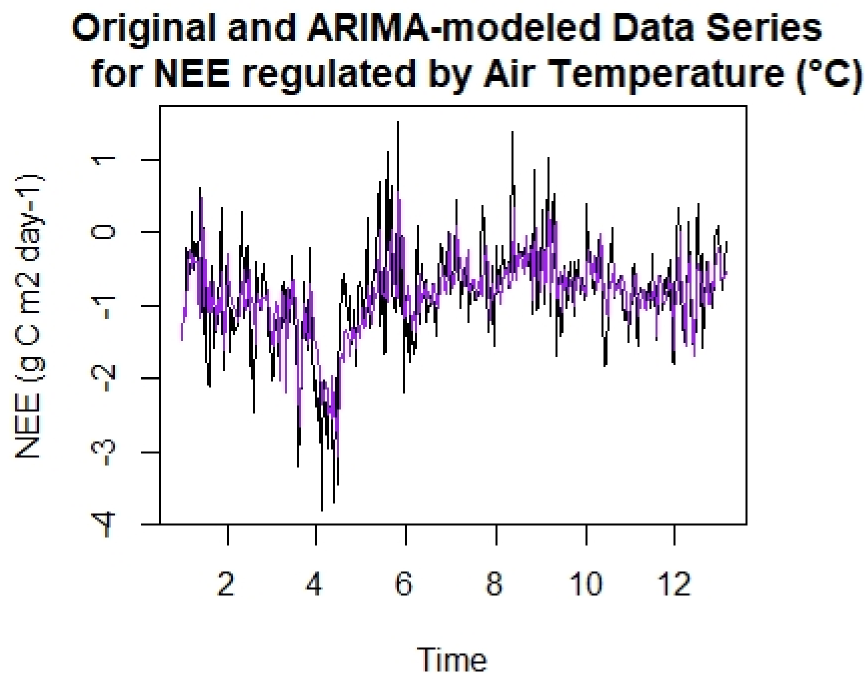


Figure 3. Time series of NEE plotted with original data series (black) and ARIMA-modeled data series regulated by air temperature (purple)

ARIMA models were run for the dependent variable NEE and the independent variables maximum salinity and air temperature to account for temporal and seasonal variation in the data series. The resulting models were plotted against the original data series (Figures 1-3).

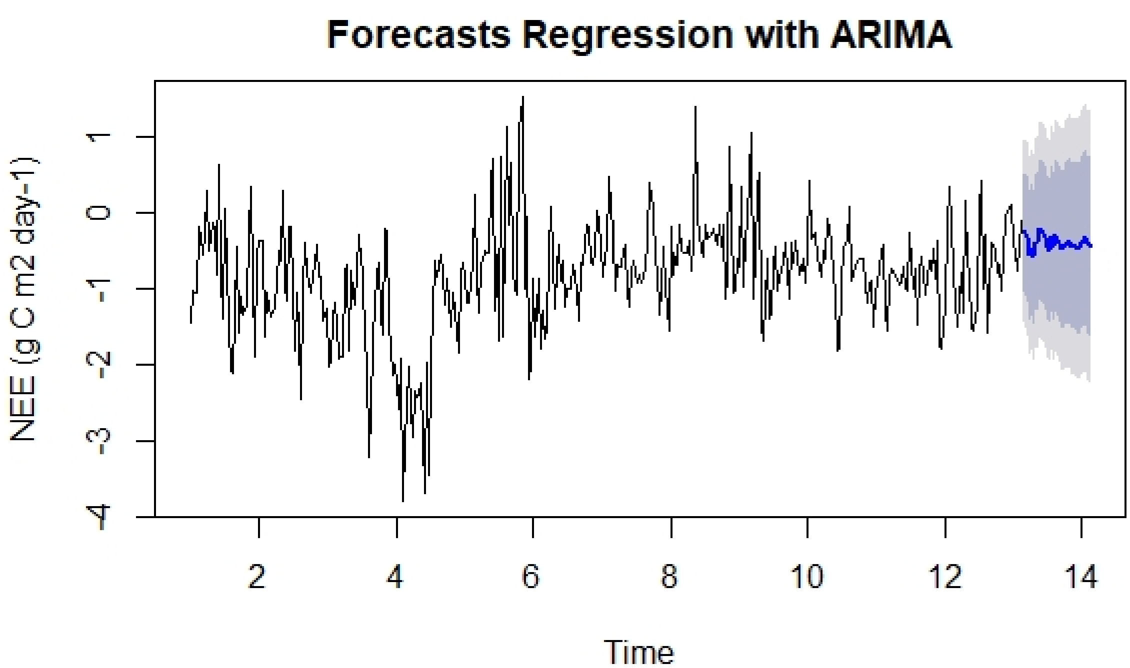


Figure 4. Forecast for future NEE values. ARIMA(10,1,3)(2,0,2)30

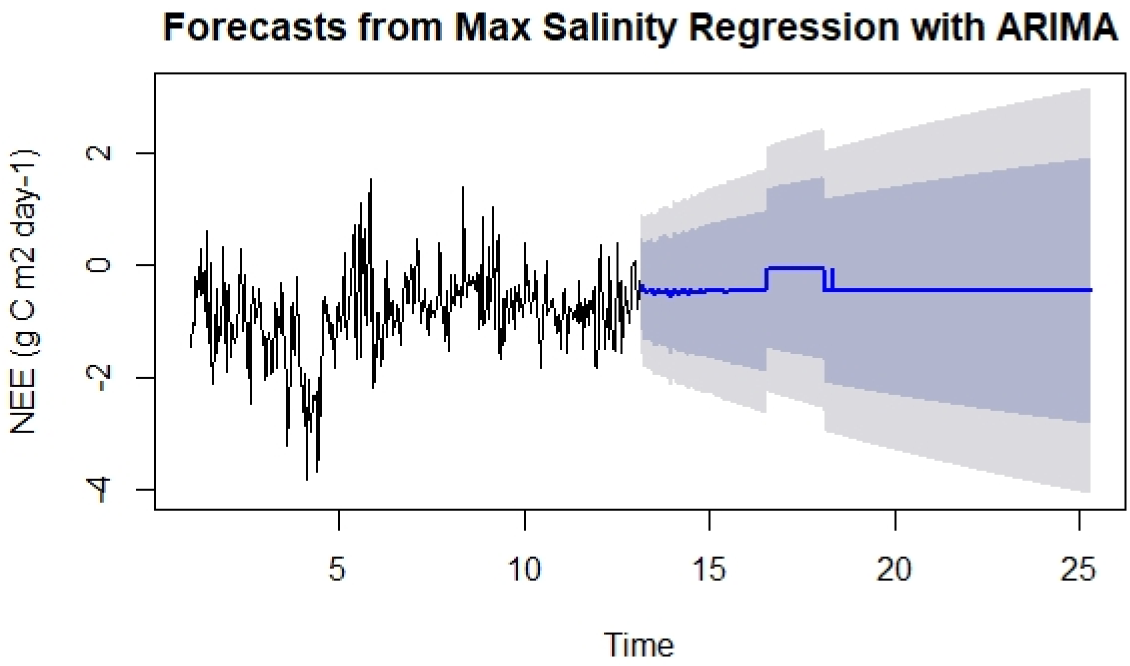


Figure 5. Forecast for future NEE values using maximum salinity as the regressor. ARIMA(3,1,1)(2,0,2)30

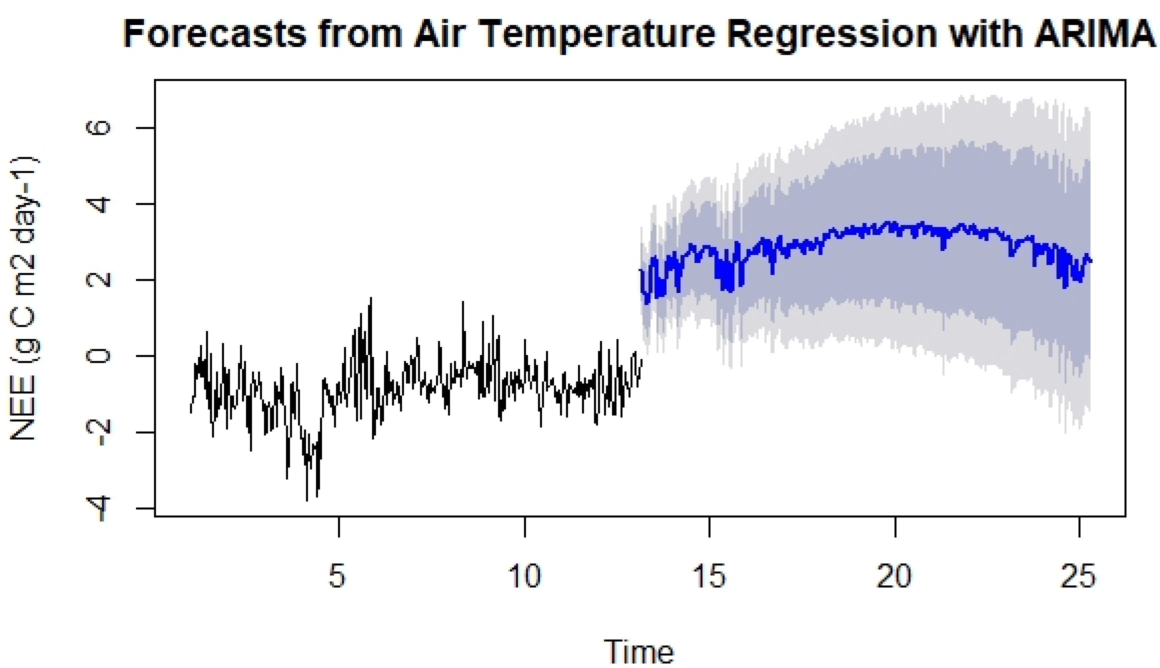


Figure 6. Forecast for future NEE values using air temperature as the regressor. ARIMA(0,1,3)(2,0,1)30

ARIMA forecast projections were run for the dependent variable NEE and the the effect of independent variables maximum salinity and air temperature on the dependent variable (Figures 4-6).

Discussion

When analyzing ARIMA models, the AIC value tests the goodness of fit of the model to the data set. The lower the AIC value is, the better the model fits the data set. While maximum salinity of water at the TS/Ph-7a site in the Florida Everglades generates a model of NEE with an AIC value only slightly lower than the the model generated by NEE alone, air temperature greatly decreases the AIC value of the resulting model (Table 1). This result suggests that air temperature is a better estimator of NEE than maximum salinity is for the TS/Ph-7a site of the Florida Everglades, and would therefore be more useful in creating future projections of NEE (Figure 6). Net Ecosystem Exchange (NEE) is defined as the overall exchange of carbon dioxide in an ecosystem (Chapin et al. 2011). Much of the carbon dioxide exchange in plant ecosystems is due to plant photosynthesis and respiration, both of which are affected by air temperature (Hofstra and Hesketh 1969). Air temperature is also known to fluctuate seasonally, so data sets with outputs such as NEE that are affected by air temperature are ideal for the use of ARIMA models to account for the effect of seasonality on the dependent variable.

Literature Cited

1.Chapin III, S. F., Matson, P. A. & Vitousek, P. Principles of Terrestrial Ecosystem Ecology. (Springer-Verlag, 2011).

2.Hofstra, G. & Hesketh, J. D. Effects of temperature on the gas exchange of leaves in the light and dark. *Planta* 85, 228–237 (1969).