# Automatic Pipeline for Estimating Normalized Difference Vegetation Index

Victor Feagins[1] Advisers: Kathryn Wheeler[2] & Dr. Michael Dietze[2]

*The University of Texas at San Antonio[1], Boston University (BU) Bioinformatics BRITE REU Program, Summer 2021[1], BU Earth & Environment[2]*

## Background

**What is Normalized Difference Vegetation Index**

Normalized Difference Vegetation Index (NDVI) is a widely used remote sensing metric for the density and greenness in plants. Calculated with the infrared and visible red light $\text{NDVI} = \frac{\rho\text{NIR} - \rho\text{Red}}{\rho\text{NIR} + \rho\text{Red}}$

Forest Canopy

High Infrared
Low Visible Red
High NDVI Value

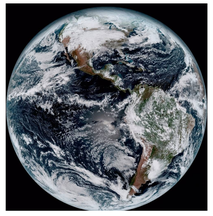
Lawn Grass

Less Infrared
Low Visible Red
Smaller NDVI Value

**New Data Source: Satellite Images**

The launch of National Oceanic and Atmospheric Administration(NOAA)'s Geostationary Operational Environmental Satellite (GOES) 16 has led to a tremendous amount of data that has high spatial and temporal coverage that can be used to estimate NDVI values over time daily.

GOES Spatial coverage of North and South America



## Motivation

**Noisy Data from Satellites**

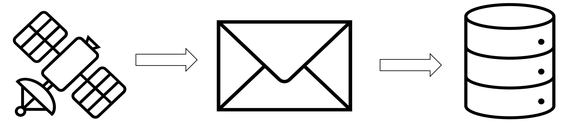
Data collected from GOES 16 is more noisy and negatively bias due to atmospheric interference compared to the forest camera method of acquiring data.

**Pipeline Required to Process Data**

Due to the noise, bias, volume and velocity of the data a data pipeline involving acquiring, cleaning, and statistical modeling to produce valid NDVI estimates is needed.

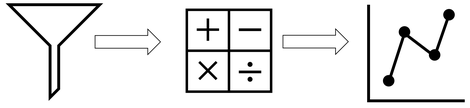
## Methodology

**Acquiring Data : Every 5 minutes**



An email is sent every 5 minutes from NOAA containing File Transfer Protocol addresses. Python and R is utilized to scan the text of these emails and download the raw infrared and visible red data in parallel. This results in 17 GB of raster data daily

**Cleaning and Processing**



Observations are filtered out based on the location, NOAA's quality flags and there is not cloud overcast. We then calculate our Raw NDVI Time Series daily.

**Modeling**

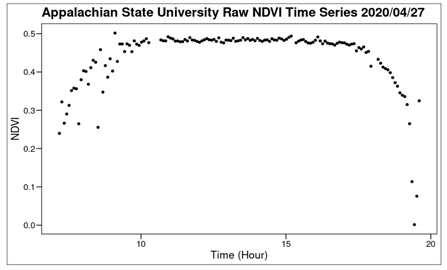
To estimate a less bias and noisy daily NDVI value we fit the data with a double exponential curve. This is done using JAGS and can be very time consuming but allows us to have a credible interval for the NDVI estimate. For more information see (Wheeler and Dietze 2019)

## Results

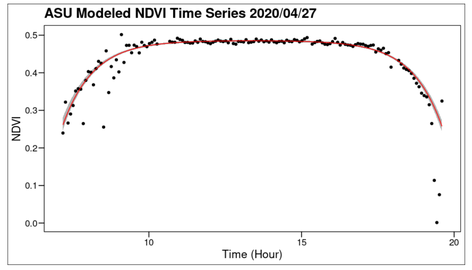
The result of the pipeline is the daily less biased NDVI time series. This can be used in doing research about forests for example looking at seasonal changes and measuring biomass. An exploratory Shiny app was also created to look at various model fits.

For code see:
https://github.com/victorfeagins/NDVIModeling

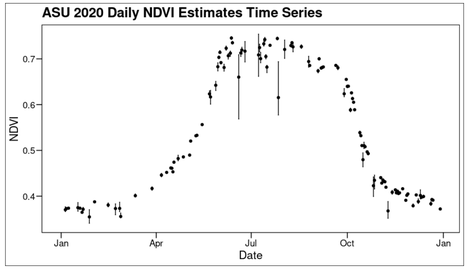
**Modeling Process Sequence**




1. Retrieve Raw Time Series for day

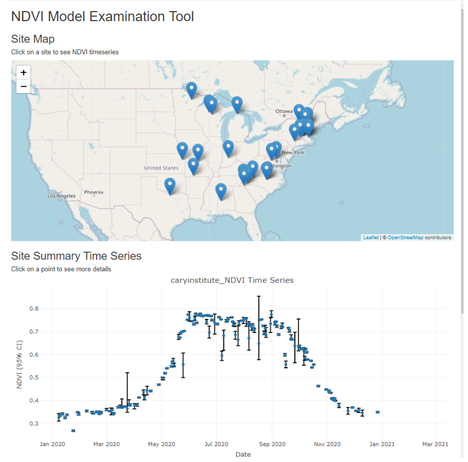



2. Fit Double Exponential Curve




3. Each day has an NDVI value which can be examined for seasonal changes

**Shiny App to Explore NDVI Fits**




## Citations

Wheeler, K.I.; Dietze, M.C. A Statistical Model for Estimating Midday NDVI from the Geostationary Operational Environmental Satellite (GOES) 16 and 17. Remote Sens. 2019, 11, 2507. https://doi.org/10.3390/rs11212507

Huang, S., Tang, L., Hupy, J.P. *et al.* A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J. For. Res.* **32,** 1–6 (2021). https://doi.org/10.1007/s11676-020-01155-1

## Acknowledgements

This work was funded, in part, by NSF grant DBI-1949968, awarded to the Boston University Bioinformatics BRITE REU program and MSB Grant 1638577 to The Near-term Ecological Forecasting Initiative