Analyzing the Relationship between Land Surface Temperature and Vegetation Indices

Joe Fragala fragaj3@rpi.edu

Rensselaer Polytechnic Institute, Troy, NY, United States,

Abstract

This poster focuses on the crucial connection between the heat emitted from the Earth's surface and the health of its plant life. Using NASA's satellite data, we examine how LST interact with vegetation through two specific measures: NDVI and EVI.

As these terms will reoccur throughout the poster:
LST - Land Surface Temperature
NDVI - Normalized Difference Vegetation Index
EVI - Enhanced Vegetation Index

These indices help us understand the vitality of plants in relation to their thermal environment. The goal is to use this analysis to contribute to environmental monitoring and sustainability efforts.

The methodology involves preprocessing data to handle missing values, visualizing distributions via heatmaps and scatter plots, and employing random sampling for tractable data analysis. We use linear regression and decision tree models to predict vegetation indices from LST, assessing model performance with standard error metrics and visualization of predicted versus actual values. This streamlined approach facilitates a comprehensive yet efficient exploration of the temperature-vegetation relationship.

Hypothesis

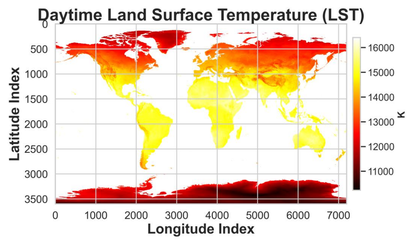
There is a positive correlation between land surface temperature (LST) and vegetation indices (NDVI and EVI). This suggests that as temperatures rise, there may be an increase in vegetation and health, possibly due to longer growing seasons or other temperature-related growth factors.

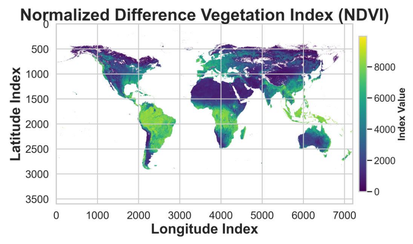
The Data

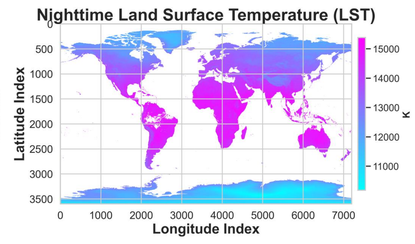
MOD11C3 v061 - This is the MODIS/Terra Land Surface Temperature/Emissivity (LST/E) Monthly L3 Global 0.05 Deg CMG product. It provides monthly average land surface temperature data over the Earth's surface in a gridded manner. The 'LST_Day_CMG' and 'LST_Night_CMG' fields represent the daytime and nighttime land surface temperatures, respectively, expressed in Kelvin. The data type for these temperatures is a 16-bit unsigned integer with a range of 7500 to 65535. A scaling factor is applied to these values, and certain values are used to indicate no data or fill values. This product helps in understanding surface heating dynamics and can be used to study the effects of land cover changes on surface temperatures.

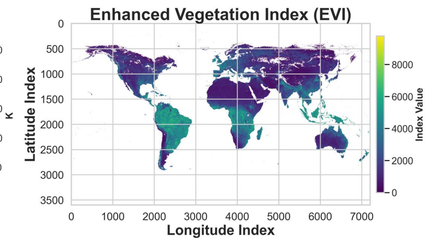
MOD13C2 v061 - This dataset, MODIS/Terra Vegetation Indices Monthly L3 Global 0.05 Deg CMG, includes vegetation indices that are indicators of plant health, vigor, and biomass. The 'CMG 0.05 Deg Monthly NDVI' provides the Normalized Difference Vegetation Index (NDVI), and the 'CMG 0.05 Deg Monthly EVI' offers the Enhanced Vegetation Index (EVI). Both indices use 16-bit signed integer data types and are scaled by a factor. NDVI values typically range from -2000 to 10000, and EVI values from -3000 to 10000, with negative values indicating no data or water bodies. These indices are essential for monitoring vegetation dynamics, assessing plant productivity, and estimating land cover conversion rates.

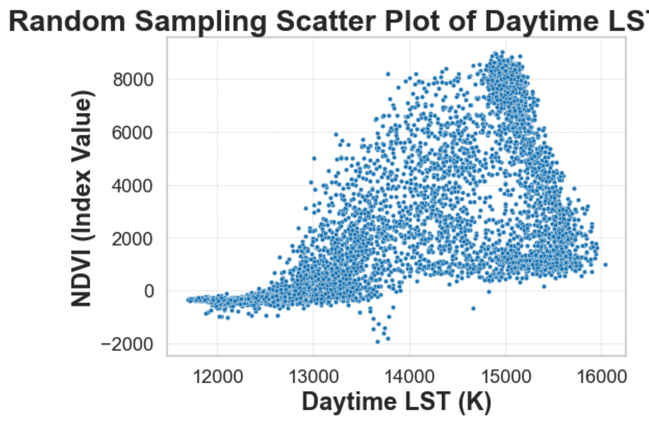
Analysis

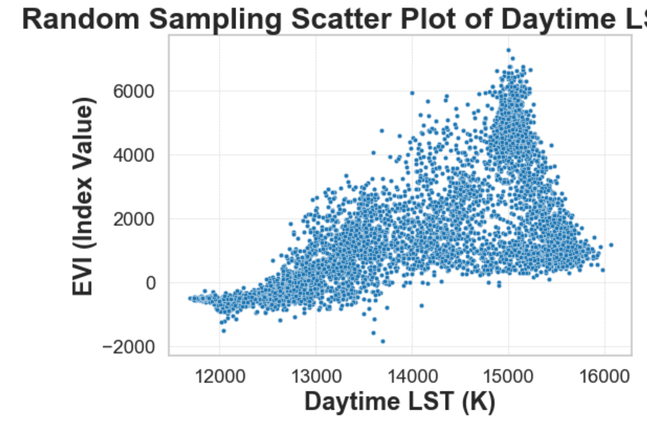


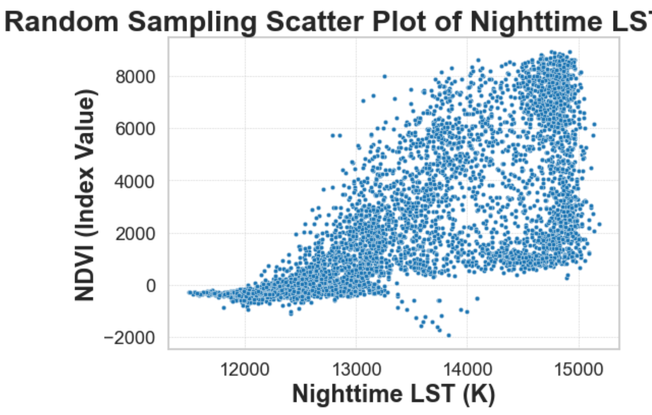


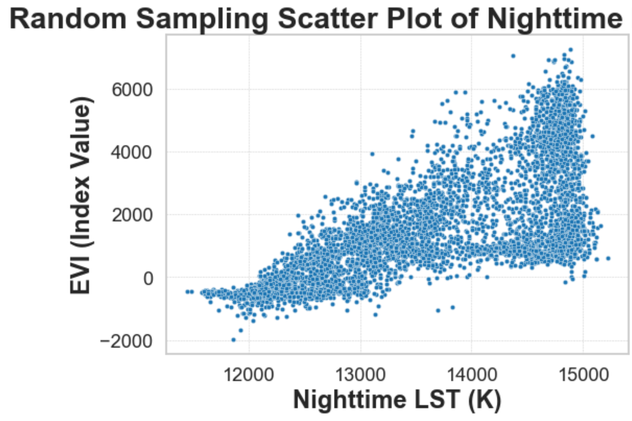






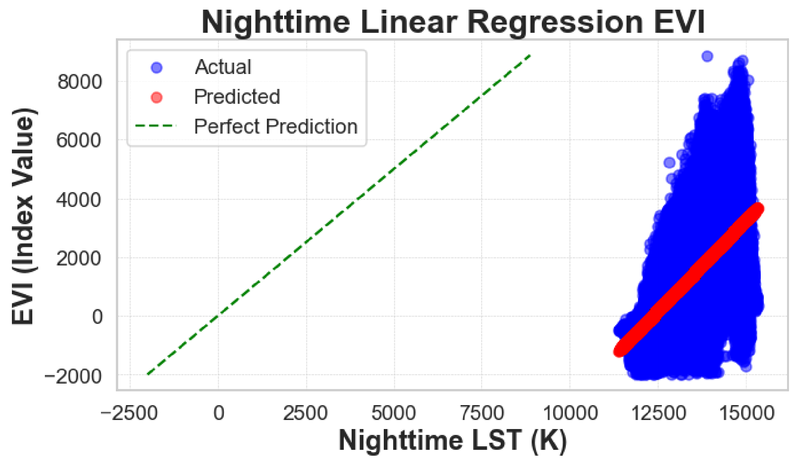


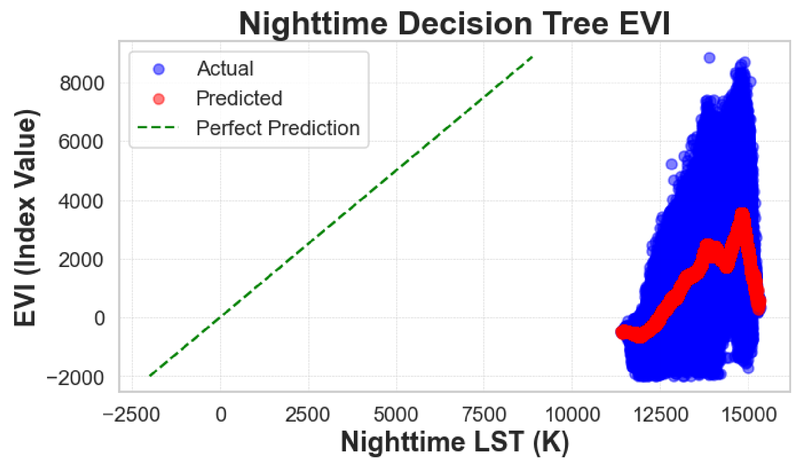




Daytime LST vs. NDVI: There is a moderate positive correlation (0.590) between daytime LST and NDVI. This suggests that as the daytime land surface temperature increases, the vegetation health (as indicated by NDVI) also tends to increase to some extent.
Nighttime LST vs. NDVI: The positive correlation is stronger (0.695) for nighttime LST and NDVI compared to the daytime. This might imply that nighttime temperatures are a more significant predictor of vegetation health or that the health of vegetation influences the surface temperature more at night.
Daytime LST vs. EVI: There is a slightly stronger correlation (0.614) between daytime LST and EVI than between daytime LST and NDVI. This can mean that the EVI, which corrects for some atmospheric conditions and canopy background signals, might be a more sensitive indicator of vegetation health in relation to temperature.
Nighttime LST vs. EVI: The correlation is highest in this pairing (0.702), indicating a strong positive relationship between nighttime LST and EVI. This again suggests that nighttime land surface temperatures might be more closely linked to vegetation health, or vice versa, as measured by EVI.

Based on the above, the following data was decided to be the most appropriate Linear Regression and Decision Tree to display as it showed the strongest results of the observed trend:





The plots display the relationship between Nighttime Land Surface Temperature (LST) in Kelvin and the Enhanced Vegetation Index (EVI) using linear regression and decision tree models. The blue points represent the actual data points, and the red points are the predicted values from the models. The green dashed line represents the line of perfect prediction where the predicted values perfectly match the actual values.
Linear Regression: The mean squared error (MSE) for the linear regression is approximately 1,568,311, indicating the average squared difference between the estimated values and the actual value. The R^2 value is 0.4907, meaning that approximately 49% of the variance in EVI is predictable from LST using this model.
Decision Tree Regression: The decision tree model has a slightly lower MSE of approximately 1,463,591, suggesting better predictive performance than the linear regression model. The R^2 value is higher at 0.5247, indicating that around 52% of the variance in EVI is predictable from LST using the decision tree model.
Based on the data observed, there is a positive correlation between land surface temperature and the vegetation indices. This suggests there may be an increase in vegetation due to longer growing seasons.

Glossary:
HDF (Hierarchical Data Format): File format designed to store and organize large amounts of data.
Linear Regression: A statistical method for modeling the relationship between a dependent variable and independent variables, assuming a linear relationship.
Decision Tree Regressor: A machine learning algorithm that models decisions and possible consequences.
Mean Squared Error (MSE): A measure of the quality of an estimator.
R2 Score: The proportion of variance in the dependent variable that is predicted from the independent variable(s).
Random Sampling: A statistical method to select a subset of data from within a dataset to estimate population.

Resources:
Read hfd files: https://www.hdfgroup.org/downloads/hdfview/
pyhdf package in python: https://pypi.org/project/pyhdf/
numpy package in python: https://numpy.org/
matplotlib library in python: https://matplotlib.org/
seaborn library in python: https://seaborn.pydata.org/
random library in python: https://docs.python.org/3/library/random.html
sklearn package in python: https://scikit-learn.org/stable/