Correlation Analysis of Global Vegetation Coverage Calculated from EVI and NDVI Datasets with Rising Atmospheric CO_2 Concentrations

Background

Plants have many factors that influence growth. Carbon dioxide, essential for photosynthesis, is one such limiting factor. In recent years, global carbon dioxide concentrations have been steadily increasing (Dlugokencky, 2022). If these rising carbon dioxide levels are correlated with changes in global vegetation biomass, the ecological impact would be significant to the biosphere as a whole. This paper discusses the question: Do rising atmospheric CO_2 concentrations affect global vegetation coverage, calculated from EVI and NDVI?

There are two vegetation indices used in this exploration: Normalized difference vegetation index (NDVI), and Enhanced Vegetation Index (EVI).

NDVI is a dimensionless, numerical index from -1 to 1 that quantifies vegetation by calculating the ratio between near-infrared light (NIR) and visible red-wavelength light (R). Because plants reflect most near-infrared light, but largely absorb red light due to photosynthesis, finding the difference with remote sensing techniques can be useful for indicating the lushness of vegetation.

$$NDVI = (NIR - R)/(NIR + R)$$

A negative return value typically indicates water, a value near 0 means there is sparse or no vegetation, and a value near 1 indicates denser and healthier vegetation. Practical uses of NDVI includes determining the likeliness of a wildfire (Warner, 2017).

EVI is another dimensionless index that is used to calculate vegetation levels. It is often used in tandem with *NDVI* in geospatial studies, because using both vegetation indices can both offer insight into vegetation composition. *EVI* is calculated with consideration to atmospheric interference and resistance, such as cloud cover (Didan, K., Munoz, A. B., Solano, R., Huete, A. , 2019). The formulas for *EVI* is as follows:

$$EVI = G\,(NIR - R)/(NIR + C1 \times R - C2 \times Blue + L)$$

NIR and *R*, like in the *NDVI* formula, refer to near-infrared light and visible red-wavelength light respectively. *Blue* indicates the detected reflectance for blue-wavelength light. *C*1 and *C2* are coefficients calculated from *Blue* values that account for aerosol factors. *L* is used to factor in canopy background noise, since canopy characteristics such as leaf morphology and leaf area index can interfere with *NIR* and *R* sensing. Finally, *G* is a gain factor to scale the output value. Like *NDVI*, *EVI* values are standardized between -1 and 1, and healthy vegetation is normally between 0.2 and 0.8. In the MOD13C2 EVI algorithm, L=1, C1=6, C2=7.5, and G=2.5. Because of EVI's optimized formula, it is often used to calculate Gross Primary Productivity (Sims, 2006).

In this experiment, NDVI and EVI image datasets from the Terra Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices Monthly (MOD13C2) Version 6.1 product are used to estimate global vegetation quantity. The hypothesis is that global vegetation will be significantly correlated with the rising atmospheric carbon dioxide levels from 2000-2020. The null hypothesis states global vegetation indices are not significantly correlated with atmospheric carbon dioxide levels.

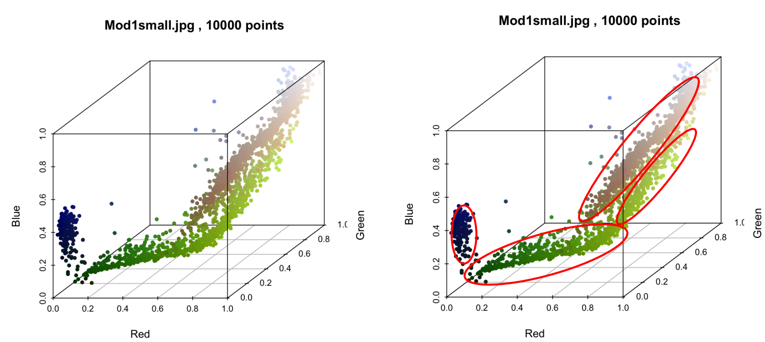
This exploration assumes that factors such as large scale wildfires, which would presumably lower the vegetation index of a region and perhaps even increase carbon dioxide levels, does not significantly impact the results of the exploration.

Materials

- Terra Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices Monthly (MOD13C2) Version 6.1
- R programming language
- RStudio
- countcolors R package
- NOAA Global Monitoring Laboratory Global averaged CO_2 records

Methods

- Download Terra Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices Monthly (MOD13C2) Version 6.1 EVI and NDVI layers from 2000 to 2020
- Identify color clusters after plotting pixels on an RGB coordinate system



○

Fig. 1, (Yuan, 2022)

- Utilize K-Means clustering with 4 identified clusters in RStudio IDE with an EVI 7300 x 3700 JPEG to clarify spherical color spaces (RGB triplets and approximate radius) relevant to the exploration

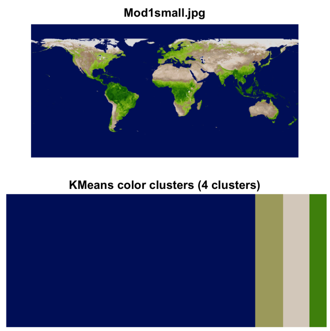
- 

Fig. 2, (Yuan, 2022)

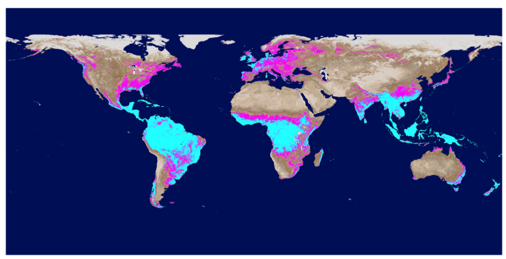
- Call countColors function from countcolors package with parameters containing spherical color spaces, and excluding the blue color range to ignore water
- Optimize the radii of spherical color spaces by observation and adjustments to countColors parameters
  - 

Fig. 3, (Yuan, 2022)

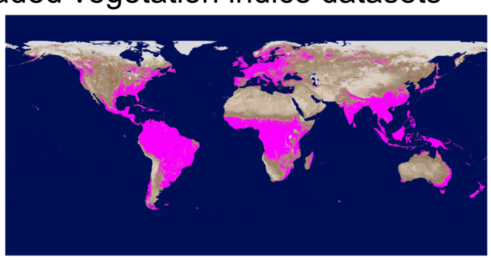
- Use countColorsInDirectory to apply countColors function to all downloaded vegetation indice datasets
  - 

Fig. 4, (Yuan, 2022)

- Record pixel.fraction values ($\frac{pixels\ identified\ in\ specified\ color\ range}{total\ pixels - water\ pixels}$) for EVI and NDVI as data for analysis
  - Ratio is a representation of global vegetation with a possible range from 0 to 1
- Access NOAA Global Monitoring Laboratory Global averaged CO2 records (yearly micromol/mol (ppm) readings with 0.13 ppm uncertainty) from 2000 to 2020 for analysis

RStudio Script

```
#imports
library(countcolors)
library(scatterplot3d)
library(imager)

#defines path to image
#typically, this will take the format of "[path/to/directory]/image_name.jpeg"
image.name <- 'PATH'

#stores and loads image into a variable
file <- load.image(image.name)
#plots image on a Cartesian plane
plot(file)

#specifies RGB triplets to ignore when counting colored pixels in an image
#here, a primarily blue color space is enclosed, to factor out water area
a <- c(0.2, 0.2, 0.45)
b <- c(0, 0, 0)

#plots pixels with RBG coordinates
#n indicates amount of points created on graph
colordistance::plotPixels('PATH', lower = NULL, upper = NULL, n = 10000,
color.space = "rgb")

#uses K-means clustering to find n number of color clusters
kmeans.clusters <- colordistance::getKMeanColors('PATH', n = 4, plotting = TRUE)

#prints rgb triplets and sizes representing clusters
colordistance::extractClusters(kmeans.clusters, ordering = TRUE)

# typical RGB values range between 0 and 255, but R scales them to range between
# 0 and 1, where 1 is maximum brightness

#defining RGB boundaries by creating a center
#later, the radii from the RGB points will be defined to create a spherical color space
#two spheres are created to better capture the pixels indicating vegetation
center.spherical <- c(0.5581260960, 0.68501913, 0.17267655)
center1.spherical <- c(0.1947635184, 0.50021722, 0.01559939)

#reads image into environment
map <- jpeg::readJPEG('PATH')

#identifies and changes target pixels to magenta
#magenta is used in this exploration, since it would not appear on the map prior to the
countColors function
countcolors::changePixelColor(map, map.spherical$pixel.idx, target.color="magenta")

#counts pixels in color range(s) within defined radii, excluding a range of color to be
ignored
```

```
#and changes counted pixels to a color(s)
#plots the map with the color change
#rgb radius extended from spherical center
two.colors <- countcolors::countColors('PATH', color.range="spherical",
                               center = c(center.spherical, center1.spherical), radius = c(0.11,
                               0.175), bg.lower=b, bg.upper=a, plotting = TRUE,
                               target.color=c("magenta", "cyan"))

#prints the fraction of the image with the selected coloration
#excluding the color space defined by bg.lower and bg.upper
two.colors$pixel.fraction

#countColorsInDirectory is a wrapper for countColors
#it returns both images with modified pixels and pixel.fraction values
#radius was approximated with trial and error
final <- countColorsInDirectory('FOLDERPATH', color.range = "spherical",
                              center = c(center.spherical, center1.spherical), radius =
                              c(0.11, 0.175), bg.lower=b, bg.upper=a,
                              target.color=c("magenta"), plotting = FALSE,
                              save.indicator = TRUE, dpi = 72,
                              return.indicator = FALSE)

#loops through and prints elements of the final array
for (val in 1:21){
  print(final[val])
}
```

Ethical considerations

All organizations providing the data used in this exploration allow public use of said data in a research context. This exploration stays within the boundaries and restrictions set by said organizations.

Analysis

Collected Data

Years after 2000	EVI vegetation coverage	NDVI vegetation coverage	CO_2 (ppm)
0	0.251172	0.3389886	370.10
1	0.2496563	0.3485317	371.96
2	0.2465959	0.3330397	374.26
3	0.2469661	0.3355892	376.53
4	0.2505242	0.3499112	378.17
5	0.2453358	0.339129	380.53
6	0.2568633	0.351698	382.39
7	0.2508958	0.3371313	384.46
8	0.2504311	0.3428987	386.27
9	0.2532653	0.3501922	387.84
10	0.2607038	0.3394194	390.36
11	0.2670748	0.3536349	392.04
12	0.2631762	0.3503246	394.39
13	0.2703455	0.3610156	396.88
14	0.2615755	0.3480813	398.84
Years after 2000	**EVI vegetation coverage**	**NDVI vegetation coverage**	**CO2 (ppm)**
15	0.2660175	0.3539659	401.66

16	0.2579623	0.3476244	404.67
17	0.2608254	0.3450579	406.75
18	0.2610538	0.3438963	409.19
19	0.2682372	0.3514652	411.76
20	0.2772413	0.3561561	414.14

Data Overview

R-squared values are calculated between vegetation coverage calculated from EVI and NDVI, and atmospheric CO_2 concentrations in order to determine whether a significant correlation exists. In addition, R-squared values were calculated for both vegetation coverage values and CO_2 concentrations with time.

Processed Data

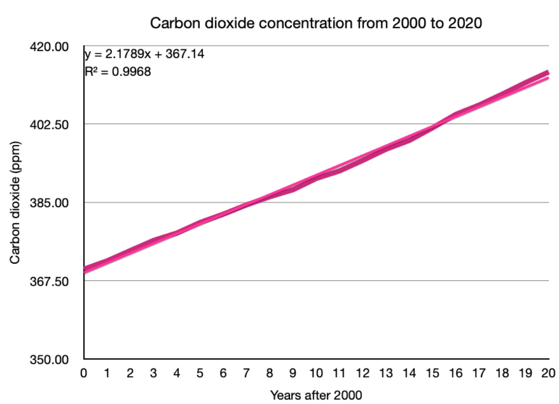


Fig. 5, (Yuan, 2022)

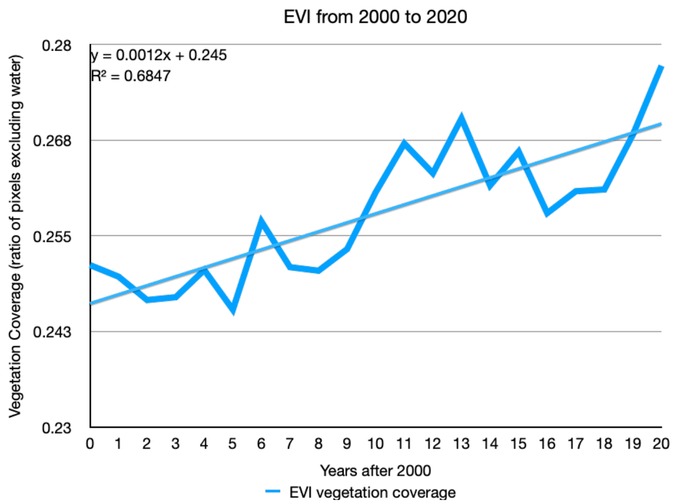


Fig. 6, (Yuan, 2022)

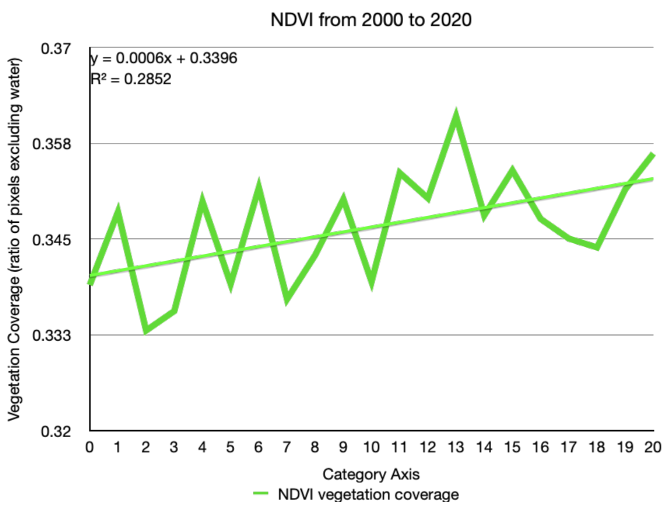


Fig. 7, (Yuan, 2022)

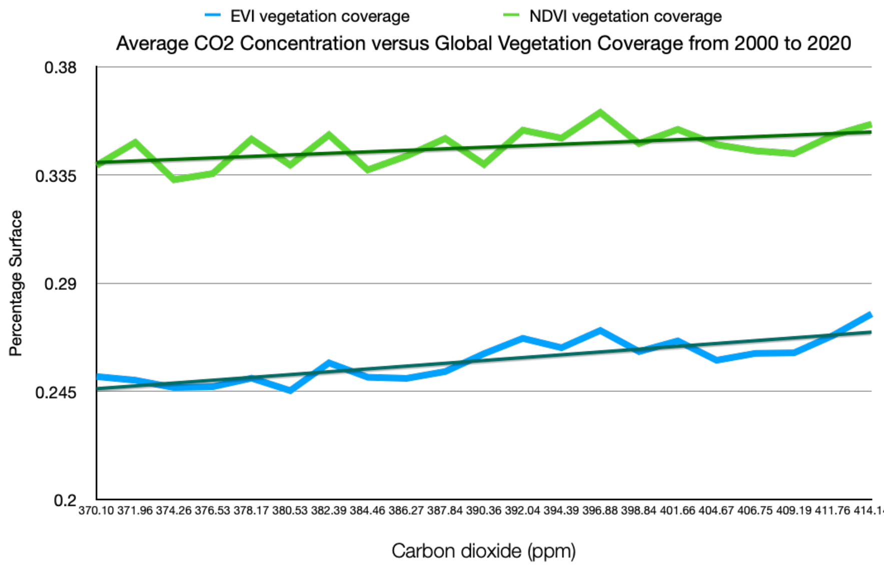


Fig. 8, (Yuan, 2022)

EVI:

$$R^2 = 1 - \frac{Sum\ of\ the\ residuals\ squared}{Total\ sum\ of\ squares}$$

R-squared value = 0.672

NDVI:

R-squared value = 0.268

Results

Carbon dioxide concentration is highly correlated with time, with an R-squared value very close to 1. On the other hand, EVI-calculated vegetation coverage is only moderately correlated with time, while NDVI-calculated coverage is even less correlated. The R-squared value of EVI-calculated global vegetation coverage with CO_2 concentration is 0.672, implying a moderate correlation. The R-squared value of NDVI-calculated global vegetation coverage with CO_2 concentration is 0.268, implying a lesser correlation. It should be noted that, with the current MODIS spatial resolution, a low R-squared value is not unexpected in the context of other NDVI/EVI studies, even one where NDVI before and after forest fires were examined (Warner, 2017). This implies that even a small R-squared value may indicate a greater change in vegetation quality or quantity. The variation within the data, especially the 0.133 ppm uncertainty for CO_2 measurements, is unlikely to greatly affect the statistical significance of the analysis.

Conclusion

Even though correlation between vegetation coverage and CO_2 measurements was not extremely significant, according to prior calculations (R-squared values of 0.672 and 0.268), the possibility of a strong correlation still exists. The timeframe of this experiement's scope is twenty years, and already EVI-calculated vegetation coverage displays moderate correlation with CO_2 levels. NDVI-calculated correlation displayed less correlation, but it should be noted that the underlying NDVI algorithm is less accurate in most cases than EVI. It should be noted that if the period of investigation were expanded, the results may be more conclusive, as all processes involved in experimentation are relatively slow ones.

Extension

In future explorations, a perspective from a different temporal standpoint, as well as calculations with more precise vegetation indices such as Normalized Difference Red Edge Vegetation Index (NDRE) or Modified Soil-Adjusted Vegetation Index (MSAVI) may be utilized in conjunction with the standard NDVI and EVI indices. In addition, rather than observing global vegetation, observing regional vegetation and CO_2 changes more accurately over a period of time may yield more conclusive results. Because regions can be subject to different climatic effects, analyzing smaller areas within their own contexts would provide a more comprehensive set of data. Ultimately, this exploration displayed a possibility for a statistically significant correlation between global vegetation coverage and atmospheric CO_2 levels, though not conclusively.

References

Didan, K., Munoz, A. B., Solano, R., Huete, A. (2019). *MODIS Vegetation Index User's Guide*. Tucson, AZ: The University of Arizona.

Didan, K. (2021). *MODIS/Terra Vegetation Indices Monthly L3 Global 0.05Deg CMG V061*. NASA EOSDIS Land Processes DAAC. Accessed 2022-01-14 from https://doi.org/10.5067/MODIS/MOD13C2.061

Ed Dlugokencky and Pieter Tans. (2022). *NOAA/GML* (gml.noaa.gov/ccgg/trends/)

Hannah Weller (2019). *countcolors: Locates and Counts Pixels Within Color Range(s) in Images*. R package version 0.9.1. https://CRAN.R-project.org/package=countcolors

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RStudio Team (2021). *RStudio: Integrated Development Environment for R*. RStudio, PBC, Boston, MA. http://www.rstudio.com/.

Sims, D. A., et al. (2006), *On the use of MODIS EVI to assess gross primary productivity of North American ecosystems*, J. Geophys. Res., 111, G04015, doi:10.1029/2006JG000162.

Timothy A. Warner, Nicholas S. Skowronski & Michael R. Gallagher (2017) *High spatial resolution burn severity mapping of the New Jersey Pine Barrens with WorldView-3 near-infrared and shortwave infrared imagery*, International Journal of Remote Sensing, 38:2, 598-616, DOI: 10.1080/01431161.2016.1268739

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). *Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product*. Remote Sensing of Environment, 185, 46-56.