Communication

An analysis of the Reproducibility of the Methods found in the Research Article "Where does your guacamole come from? Detecting deforestation associated with the export of avocados from Mexico to the United States"

Kristen Gollwitzer

[1] Kristen.Gollwitzer@slu.edu

Abstract: This study attempts to reproduce the results of the research endeavors undertaken by Cho in [11]. In carefully attempting to recreate the methods of the remote sensing piece to that study, this remote sensing study takes a turn, and opts for a less complicated and involved methodology than Cho, yet still an interesting take. This remote sensing study utilizes five different indices (EVI, NBR, NDVI, SAVI, and WDVI) to understand vegetal patterns in the Michoacan region of Mexico. It utilizes google earth engine and the associated python API to glean this information from Landsat imagery in the years between 2016 and 2020. The study concludes that it can accurately track patterns in vegetal health given the consistency of the indices it finds, but that it can say nothing about that nature of the avocado industry and its impact on deforestation in the Michoacan region of Mexico.

Scientific Question: Are the methods of Cho [11] reproducible given only the information in their paper?

Citation: Gollwitzer, K. Title. *An analysis of the reproducibility*, **2022**, *22*, x. https://doi.org/10.3390/xxxxx

1. Introduction

This study seeks to understand remote sensing methodologies to conduct environmental economics research [benefits]. The study utilizes the paper [] as a template for an introductory remote sensing study on the avocado industry in the Michoacan region of Mexico, and its effects on the region's vegetative growth. This study utilizes primarily vegetal indices to understand change over time in the quantity and quality of vegetation in this particular region of Mexico. There is some discussion about what inferences can be made from fluctuations in vegetation onto industries around the areas being analyzed [7]. The paper goes to great lengths to track the exact extent that any remote sensing study can incriminate an industry in the way of deforestation. This communication is primarily meant to use the background information found in [4] to justify an application of vegetal indices for understanding the avocado industry and its impact on the Michoacan region. In this way, the justification is prepared by another scientist, and this communication serves to replicate the methods of the previous study with alterations allowing for a full application in python.

This study utilized five indices to understand the vegetal health and extent in the Michoacan region. The first of the five indices is the Enhanced Vegetation Index (EVI) which uses the NIR, red, and blue bands to measure healthy vegetation. This index is most appropriate for densely forested areas. When an EVI is calculated healthy vegetation usually falls between values 0.2 and 0.8.

The second index utilized was the Normalized Burn Ratio (NBR). The NBR uses the NIR and SWIR bands to understand the severity of a burn area. In this study, this index could be useful to understand the areas where the avocado industry has burned forest to clear it for agriculture use. The ratio ranges with result from -1 to 1. A result of -0.1 relates to enhanced regrowth after a fire, and a result of 0.66 relates to a high severity burn in the area. The downside to this index for use in Mexico is that it is most usable in areas where regrowth after a fire is very slow (like the temporal United States) whereas Mexico falls close to the tropics and regrowth would be relatively quick.

The third Index used was the NDVI, which is a very popular marker for vegetal health and abundance in remote sensing studies[10]. This index utilizes the NIR and Red bands to understand the location and relative abundance of vegetation. With a result value close to 1, dense vegetation is likely.

The fourth index used was the Soil Adjusted Vegetation Index. This index is used to minimize the impact that soil has on the classification and measurement of vegetation. The index utilizes the NIR and Red bands along with 'L' which in this study was equated to the calculated NDVI for the image at hand. In areas with bare soil the result of a SAVI will be -1 to 0.2. In areas where plants are putting out leaves the result will be from 0.4 to 0.6.

The fifth and final index used in this study was the weighted difference vegetation index. This index is a more simplified version of other vegetal indices. This is interpreted in the same bounds as the NDVI.

Calculating and analyzing the results of the five indices listed above is meant to help gain a more thorough understanding of the vegetal growth within the area of interest. For instance, the NDVI without the presence of the SAVI does not make any efforts to account for bare soil brightness which could be a major skew in the data. Any information that can be gleaned about the nature of the vegetation in the area of interest is of great value in this study. The very nature of comparing forests to avocado farms is a challenging feat without a plethora of ground truthing data to work from. Even then the controversy of using this information to directly incriminate the avocado industry with the deforestation of this region in Mexico is still present. That is why the original Cho [11] paper went to great lengths to get ground truthing data to validate their findings and to understand exactly where the avocado industry put its fingers in the Michoacan region of Mexico. The main information that was utilized in the study was trade data that I did not have access to, interviews that I do not have access to and a remote sensing study that I do have access to. That is why this study became hyper focused on indices and strayed away from the random forest [3,4,6] classification that was used in the original paper. Rather than focusing on identifying avocado plantations and differentiating these from forest this study assumes that forest would indicate the highest levels of vegetation compared with an avocado plantation [3]. Therefore, high levels of vegetation will be taken as good news and low levels of vegetation will be taken as bad news.

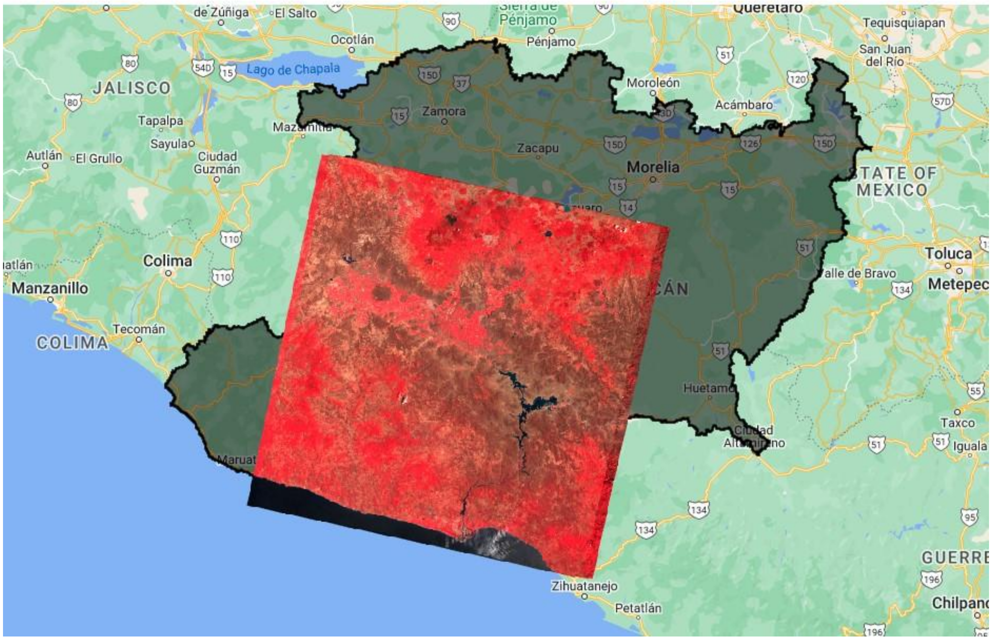
Furthermore, avocado plantations would reveal more bare and bright soil than a forest would, and this is accounted for by calculating a SAVI. Therefore, a high level of vegetal growth means a low level of deforestation for the sake of my study. I am working with far less ground truthing, and protected industry data than the original researchers in this study and therefore I am forced to make larger assumptions. This goes to show the level of comprehensive research that must go into making statements about and industry's environmental activities.

While this study does not contain a successful Random Forest classification this would have been very helpful in understanding whether vegetation was a related to a forest or an avocado plantation [][].

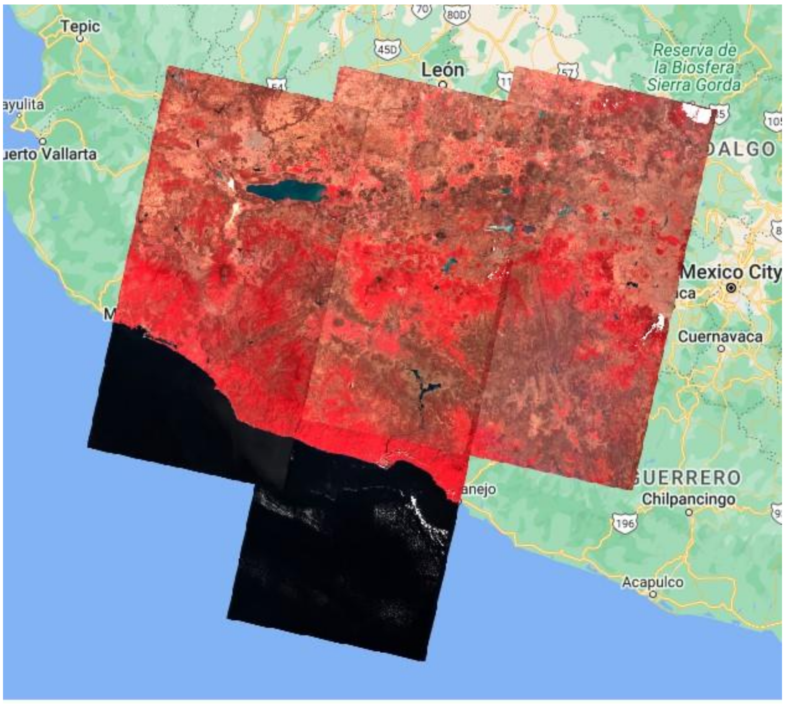
2. Materials and Methods

2.1 Graphic Workflow

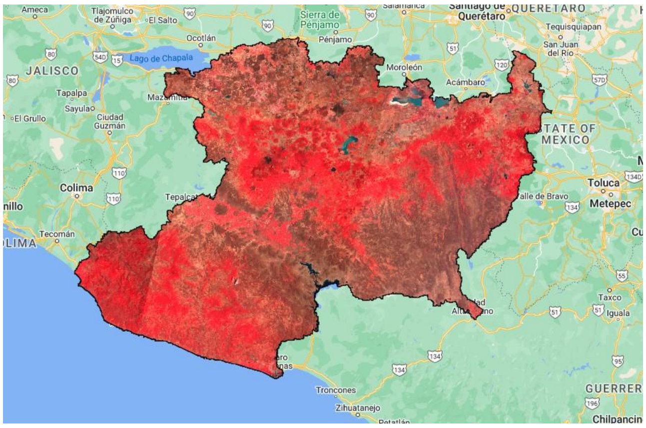
Load the imagery collection onto the map….



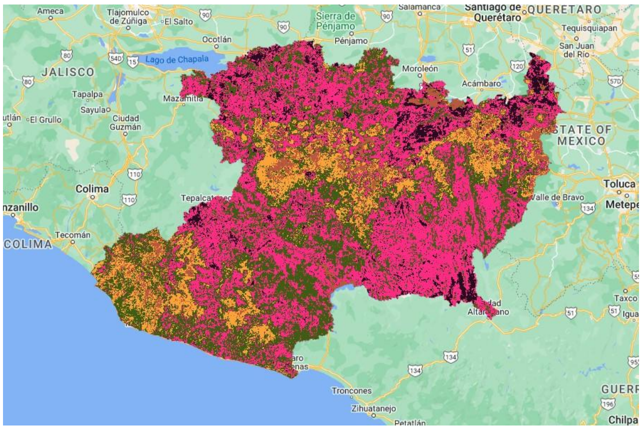
Load a second imagery collection onto the map to cover the entire region…..



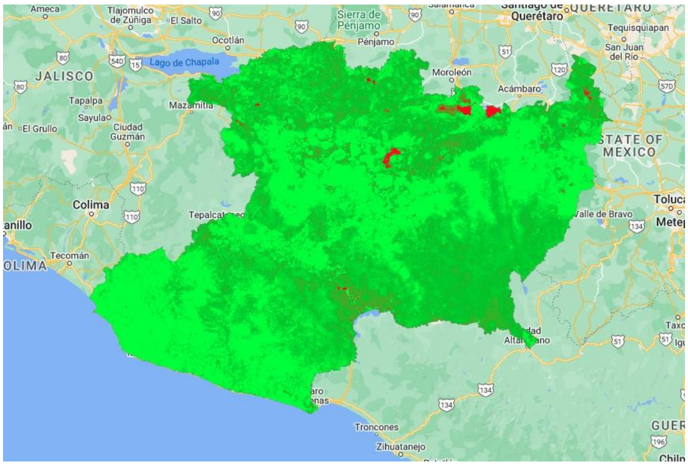
Merge, Mosaic and Clip Raster From Imagery Collection……



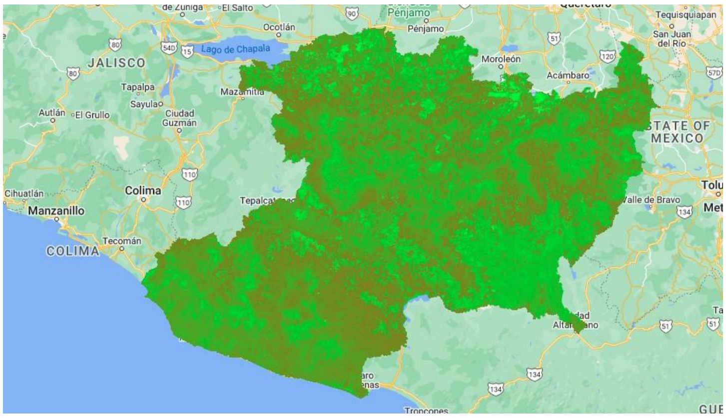
Perform Unsupervised Classification…….



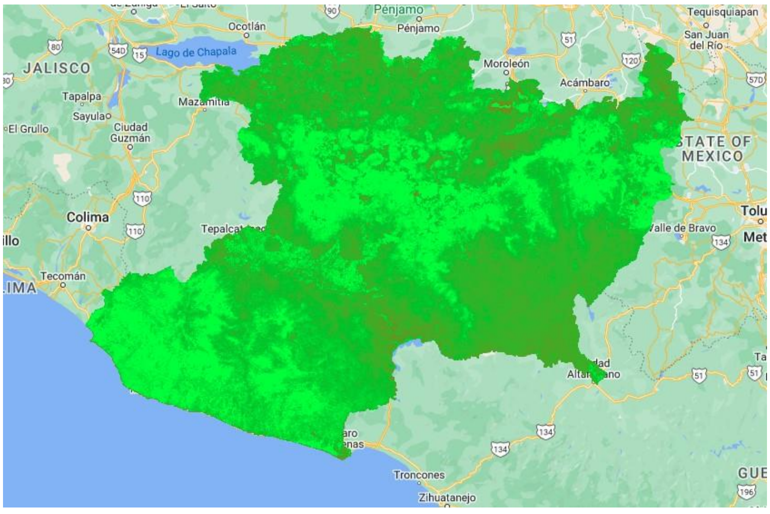
Run EVI on the Image……



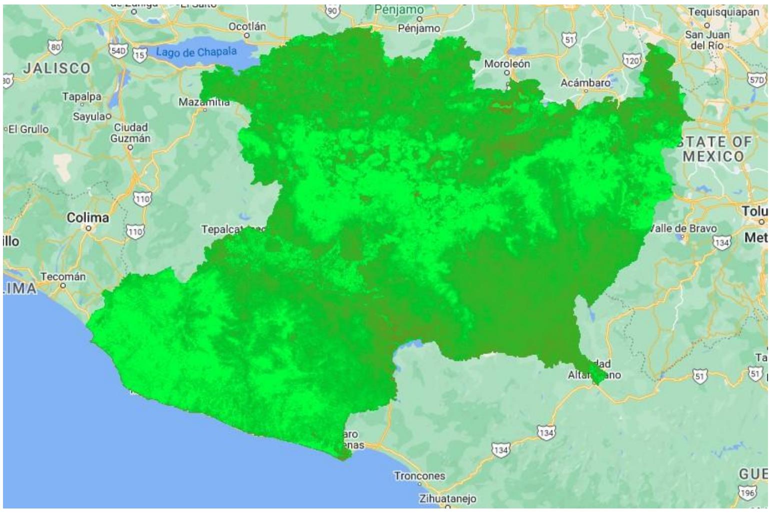
Run NBR on the Image ……



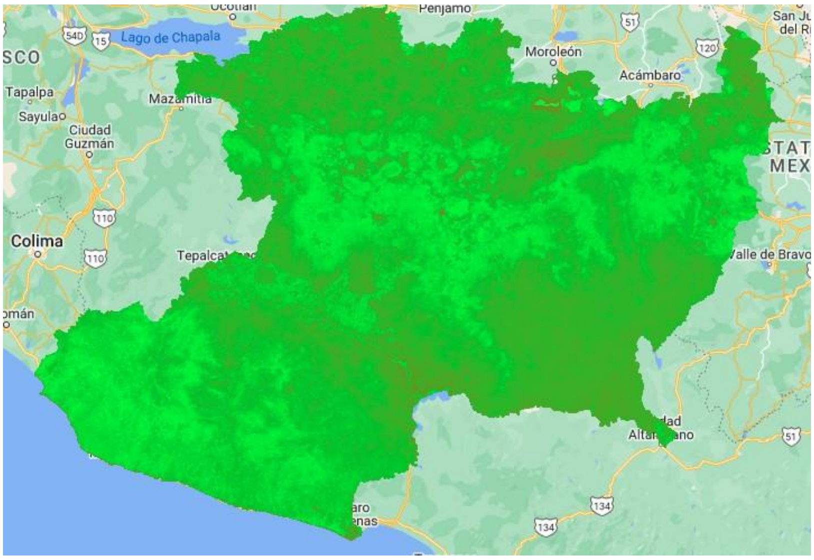
Run NDVI on the Image……..



Run SAVI on the Image…….



Run WDVI on the Image ………



Repeat all the above steps on four more image collections each of different time periods.

2.2 Methodology Explanation

The first step to this remote sensing study was obtaining an accurate boundary for the Michoacan region of Mexico. This was downloaded off an academic website. The original boundary was split into many different polygons that were meant to represent political boundaries. This dataset was then downloaded and merged into a single polygon in ArcGIS. Then, this polygon was exported as a shapefile. The next step of the process was to read this shape file into a google collab notebook using python. In order to do this the pyshp, geemap, pycrs libraries were pip installed. Then these same libraries were imported. In order to read into the shapefile both the Michoacan.shp and Michocacan.dbf file needed to be uploaded. These were then converted to a single Earth Engine object. After that they were converted to an Earth Engine feature collection and read into the map as a layer.

The next step was to identify five different time periods to run analyses on. This was particularly challenging because the time periods that would be most meaningful in the context of the original study are those between 2001 and 2016. Landsat 8 data which was the data used for my study was not collected before 2014. Therefore, the time periods had to be amended to continue the work of the original study with imagery spanning from 2016 to 2020. The final time periods were the months of April and May in the years 2016, 2017, 2018, 2019, and 2020. This mimics the original study in the way that it prioritizes imagery from the prime avocado growth months of April and May that has very little cloud cover.

The parameters for each of the five image collections were all the same aside from the date. The cloud cover was set to be less than a value of two. The most notable thing about the time steps in this study is that there needed to be two separate image collections to cover the entirety of the Michoacan region. This eventually compromised the use of only imagery from April and May. The reason for this was that particular swaths of the region were only taken during months of the year other than April and May. Therefore the time-stamp had to be widened for only one of the collections in order to encompass the entire Michoacan region.

After this was done, the two collections were merged, then mosaiced, and finally they were clipped to the shape of the Michoacan region. This data was then transformed into a single image in order to run an unsupervised classification. This was performed by using the .first() function of an image collection.

Attempts were made at performing a supervised classification using the random forest classifier. After many different scripts, and lots of troubleshooting there was an error that I just could not get past. There were no answers online after many hours of searching so sadly I said goodbye to the idea of a random forest classification.

After performing an unsupervised classification, it was time to run the indices on each of the five image collections. This step came with relative ease.

To start a function was created to calculate the EVI by utilizing a image expression. The NIR, Red, and Blue bands were put into a calculation as follows:

$$2.5 * ((nir- red) / (nir + 6 * red – 7.5 * blue +1))$$

This was used to render an image that visualized the EVI calculation across the Michoacan region. The most effective step here would have been to then create areas of interest around those that the original study found to be avocado plantaions, and those that they found to be forest in order to derive EVI values at each of these locations. This step did not happen as it would have been quite a challenge to convert the image back to an image collection in order to put it through the raster clipping tool that was used earlier. The statistics were derived from this EVI by using the EE Reducer function to reduce the region to a mean and standard deviation.

The next step was to run a normalized burn ratio on the image at hand. This was done by utilizing the normalized difference Image function in google earth engine and inputting band five and seven of the Landsat imagery. The image was then visualized on the same color palette as the EVI in order to glean visual differences from the two indices for pure intrigue.

After that the NDVI was performed by utilizing the normalized difference function and inputting bands five and four. The statistics for the NDVI were derived by using the reducer function in google earth engine to derive a mean and a standard deviation for the entire region of Michoacan.

Furthermore a SAVI was performed by utilizing a an image expression and some band math. The band math is as follows:

(1 + L) * float(nir – red) / (nir + red + L)

In this soil adjusted vegetation index the L represented the NDVI that had already been calculated for that image. This way, the soil adjustment would take into account the actual level of vegetation calculated within the ROI.

Lastly the WDVI was performed by also utilizing band math which goes as follows:

Float(nir – red) / (nir + (a * red) + b)

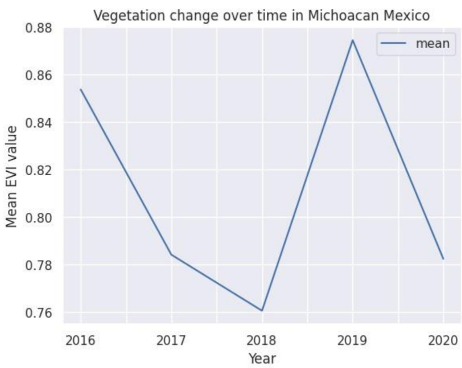
This incorporated fixed values at a and b to imitate those values that were actual in the ROI.

3. Results

The most important results of this study were the statistics derived from the indices. Those are interpreted below in the order that they were calculated.

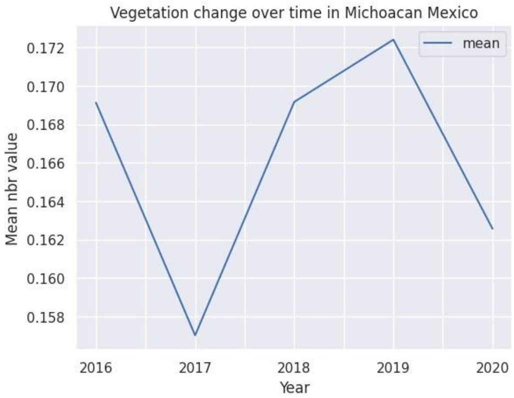
3.1. EVI Results

The EVI value for 2017 was 0.85 which indicates that there was a high level of vegetative growth in the Michoacan region during this time period. The next year it dipped slightly to 0.78 which indicates there is still a high level of vegetative growth in the area. In 2018, the number stays nearly steady with the previous year at around 0.76 still indicating lots of vegetal growth. The following year (2019) the EVI increases to 0.87 and finally in 2020 dips back to 0.78. This indicates that during the years between 2017 and 2020 there was a relatively high level of vegetal growth



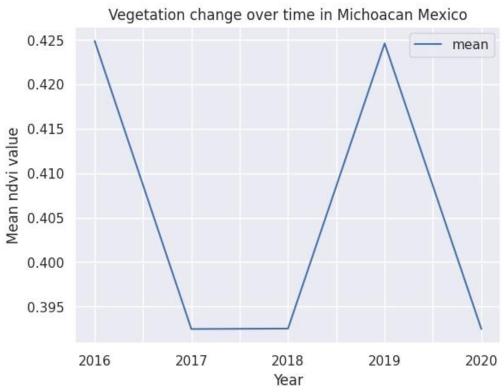
in the Michoacan region overall. These numbers are means of the entire region, and give no indication into whether avocado plantations or forests are thriving during this time.

3.2 NBR Results



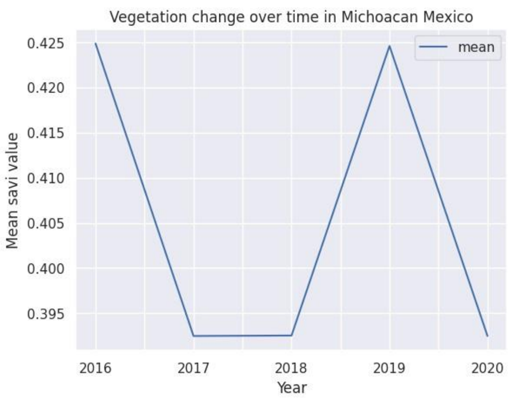
The NBR for the first year of the time period observed sat at 0.1. Through the next five year it stayed within 0.1 varying only a little. This leads one to believe that there were very few burn areas in the Michoacan region during this period of time, or that vegetal regrowth in these burn areas was very quick, and was able to disguise them quickly which would make sense for a semi-tropical area.

3.3 NDVI Results



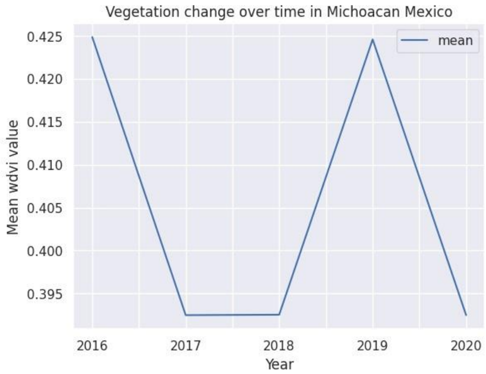
Just like the EVI, the NDVI had a slow and steady dip from 0.4 to 0.39 from the years 2016 to 2018 and then had an uptake in growth during the year 2019 only to dip back down again in 2020.

3.4 SAVI Results



The SAVI also reflected the EVI and the NDVI in that it had a slow dip from 2016 to 2018 only to have an uptake in vegetal growth and abundance in 2019 and end on a dip in 2020.

3.5 WDVI Results



The WDVI stayed very steady over the time period similar to the NBR.

4. Discussion

This research endeavor started out as an attempt to directly copy the work of [11] as much as I was able to do with my skills and knowledge. I always underestimate just how much troubleshooting needs to happen at every step of the process. Though I was unable to accomplish even close to what Cho did, I am still very proud for what I was able to glean from this remote sensing study.

Given that all of the indices that directly related to vegetation followed the same patterns it is fair to say that these were mostly accurate. The EVI, NDVI and the SAVI all decreased from the years 2016 to 2019 which leads me to believe that there was some level of deforestation in this time given that most of the Michoacan area is forest [11]. However, it does not directly speak to the level of deforestation that was caused by the avocado plantations. The NBR and the WDVI stayed consistent leading me to believe that there were no large burn areas at the time.

5.) Conclusion

While it is fair to say given the results of my indices that there was some vegetative decline from 2016 to 2018, the study was not comprehensive enough to give definitive results that the avocado industry was a main cause of notable deforestation during that time. In this way, the results of my study vary from that of [11] given that Cho was able to definitively say that the avocado industries activities during the time period that they studied were impactful on the rates of deforestation.

References

1. V.F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J.P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification", ISPRS Journal of Photogrammetry and Remote Sensing, Volume 67,2012,Pages 93-104,ISSN 0924-2716,Author 1, A.; Author 2, B. Title of the chapter. In *Book Title*, 2nd ed.; Editor 1, A., Editor 2, B., Eds.; Publisher: Publisher Location, Country, 2007; Volume 3, pp. 154–196.

2. Barsimantov and Antezana, 2012,“Forest cover change and land tenure change in Mexico's avocado region: is community forestry related to reduced deforestation for high value crops?”Appl. Geogr., 32 (2) (2012), pp. 844-853
3. Belgiu and Drăguţ, 2016, “Random forest in remote sensing: a review of applications and future directions”, ISPRS J. Photogrammetry Remote Sens., 114 (2016), pp. 24-31,
4. Breiman, 2001, L. Breiman, “Random forests”, Mach. Learn., 45 (1) (2001), pp. 5-32
5. J.G.P.W. Clevers, “Application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture”, Rem Sens Environ., 29 (1) (1989), pp. 25-37
6. P.O. Gislason, J.A. Benediktsson, J.R. Sveinsson, “Random forests for land cover classification”, Pattern Recogn. Lett., 27 (4) (2006), pp. 294-300,
7. M. Jain, “The benefits and pitfalls of using satellite data for causal inference”, Rev. Environ. Econ. Pol., 14 (1) (2020), pp. 157-169
8. Key, C. H., & Benson, N. C. (1999). “The Normalized Burn Ratio (NBR): A Landsat TM Radiometric Measure of Burn Severity. United States Geological Survey, Northern Rocky Mountain Science Center”, (Bozeman, MT).
9. A. Liaw, M. Wiener, “Classification and regression by randomForest”, R. News, 2 (3) (2002), pp. 18-22
10. J.G. Lyon, D. Yuan, R.S. Lunetta, C.D. Elvidge, “A change detection experiment using vegetation indices”, Photogramm. Eng. Rem. Sens., 64 (2) (1998), pp. 143-150
11. Kimin Cho, Benjamin Goldstein, Dimitrios Gounaridis, Joshua P. Newell,"Where does your guacamole come from? Detecting deforestation associated with the export of avocados from Mexico to the United States", Journal of Environmental Management,Volume 278, Part 1,2021,111482,ISSN 0301-4797,https://doi.org/10.1016/j.jenvman.2020.111482.