Assessment of Spatial Variability of Large-Scale Fields During the Crop Growing Season Based on Vegetative Indices Derived From Sentinel-2 Images

Authors: Ana Morales, Davide Cammarano, Francelino Rodriguez, Keith Cherkauer

Date of document competition: Friday, May 7, 2021

Publication synopsis: This project assesses spatial variability of large-scale fields (5 to over 40 ha) under three scenarios (bare soil, vegetative period, and reproductive period) using vegetative indices (NDVI, GNDVI, PPRB) derived from satellite imagery (Sentinel 2).

Publication description: Crop growth models are process-based models that simulate the daily interaction between plant, weather, soil and management practices. They can be used to extrapolate, over time, the impact of climate variability on different agronomic practices (Basso et al., 2001, 2011; Cammarano et al., 2020). Field data is required to calibrate, validate, and improve crop growth models to enhance their reliability for regional impact assessment. For large-scale fields (5 to over 40 ha), the establishment of sampling locations is a key step in order to account for the spatial variability that characterize these fields. Soil and terrain elevation maps can be useful tools to identify spatial variability. However, soil and elevation are characteristics that do not change over time, while vegetation does. Characterizing spatial variability of a field during the crop growing can help to define sampling locations that take in account that variability. Vegetative indices, derived from satellite imagery, are crop status indicators that can be used to accomplish that. The traditional approach for calculating vegetation indices has been to download the satellite image corresponding to the date of interest, use ArcGIS to clip the area corresponding to the field from which the data will be collected, and calculate manually the vegetative indices of interest. When working with more than one field and several dates, it becomes a challenge to handle a big amount of satellite images and the outputs generated. This project aims to: 1) asses spatial variability under three key scenarios (bare soil, vegetative period, and reproductive period) during the crop growing season using three vegetative indices (NDVI, GNDVI, PPRB) derived from satellite imagery (Sentinel 2), and 2) Create a Python program that handle, process, and organize inputs and outputs involved in the calculation of vegetative indices for the assessment of spatial variability. Results from this project will be important for the establishment of sampling locations that take in account spatial variability of fields over time.

Tags: spatial variability, remote sensing, Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation index (GNDVI), and Plant Pigment Ratio (PPRB)

Source of data overview

Data	Type	Provenance
Sentinel-2 images • Years: 2016-2020 • Months: April, May, June, July, August, and September.	Raster • Format: JP2	United States Geological Survey (USGS) https://earthexplorer.usgs.gov/
Perimeter of the fields of interest	Vector • Format: Shapefile (polygon)	Two sources: Provided by farmer, if available, or generated by the project.

Methods

The project involves 5 large-scale fields located around Indiana. The shapefile corresponding to the boundary of each field was stored in the folder “Field boundaries” in the main directory. For the purposes of this report, only the results from one field will be presented.

Spatial variability under three field scenarios was evaluated: 1) field with bare soil, 2) field with corn during vegetative period when plants are small and soil is still visible from above, and 3) field with corn during reproductive period when vegetation cover most of the surface. For this, selection of the Sentinel 2 satellite imagery was predicated on imagery available primarily in April, May, June, July, August, and September. Imagery from 2016 to 2020 was downloaded for all locations. A cloud cover “lower than 10%” was selected as part of the search criteria available in the United States Geological Survey website (https://earthexplorer.usgs.gov/). As a result, per each date of interest, a zipped folder containing 14 bands (blue, green, red, etc) and additional data was downloaded to a folder named “Sentinel_Images” in the main directory.

Data used in this project was handled and processed using Spyder, which is an open-source cross-platform for scientific programming in the Python language. The program created was saved under the name: “ABE651_SpatialAnalysisUsingSentinelData.py”.

Per each date of interest, the first step was to stack the layers that will be used for the calculation of the vegetative indices evaluated (NDVI, GNDVI, and PPRB). Layers corresponding to the blue (B), green (G), red (R), and near-infrared (NIR) bands were stacked in a multiband raster, using the function *“stack_bands”*. Then, information from the multiband raster and the field boundary shapefile were retrieved using the functions *“info_raster”* and *“info_vector”* respectively to identify if both files had the same coordinate reference system (CRS). In case the CRS was different, the multiband raster was reprojected using the function *“raster_change_crs”* to have the same CRS as the field boundary shapefile. Finally, the field boundary was clipped from the reprojected raster using the function *“clip_fromRaster”*. During the clipping process, the “no data” values were masked out.

Prior to calculating the three vegetation indices of interest, only NDVI (NIR-R / NIR+R) was calculated using the function *“calculate_NDVI”* and plotted for a visual check of cloud presence using the function *“raster_individual_plot”*. In addition, stats (minimum, maximum, mean, and standard deviation) of the NDVI layer were calculated using the function *“raster_min_max_mean_stdev”* for a quantitative check.

For images with no clouds within the field boundary of interest, NDVI, GNDVI (NIR-G / NIR+G), and PPRB (G-B/G+B) were calculated using the functions *“calculate_NDVI”*, *“calculate_GNDVI”*, and *“calculate_PPRB”* respectively. These three vegetation indices have been proved to be good indicators of vegetation status. They differ in the spectral information used for their calculation. While NDVI and GNDVI used near-infrared (NIR) data, PPRB uses data from the visible spectrum only (R, G, and B).

As future steps, historical NDVI, GNDVI, and PPRB will be calculated for the three scenarios: 1) field with bare soil, 2) field with corn during vegetative period, and 3) field with corn during reproductive period. The resulting historical NDVI, GNDVI, and PPRB will be correlated with ground truth data (corn biomass) collected at growth stage V8 and R2 to identify which one has the greatest correlation.

Graphical data analysis

For each vegetative index calculated, a plot was generated for visual check, stats (minimum, maximum, mean, and standard deviation) were calculated for a quantitative check, and box whisker plots were generated for a visual assessment of the values distribution. Before the creation of the whisker plots, NDVI, GNDVI, and PPRB were converted from raster to array using the function "*rasterToArray*".

The specific questions to address in this project were: 1) are visual differences between the three scenarios? 2) are visual differences between the three scenarios enhanced by the type of vegetative index evaluated (NDVI, GNDVI, and PPRB)? 3) how NDVI, GNDVI, and PPRB data distribution vary depending on the scenario evaluated? and 4) what index reflect the greatest field variability per each scenario evaluated? To answer these questions, Figure 1 shows the results obtained from the three vegetative indices under the three scenarios.

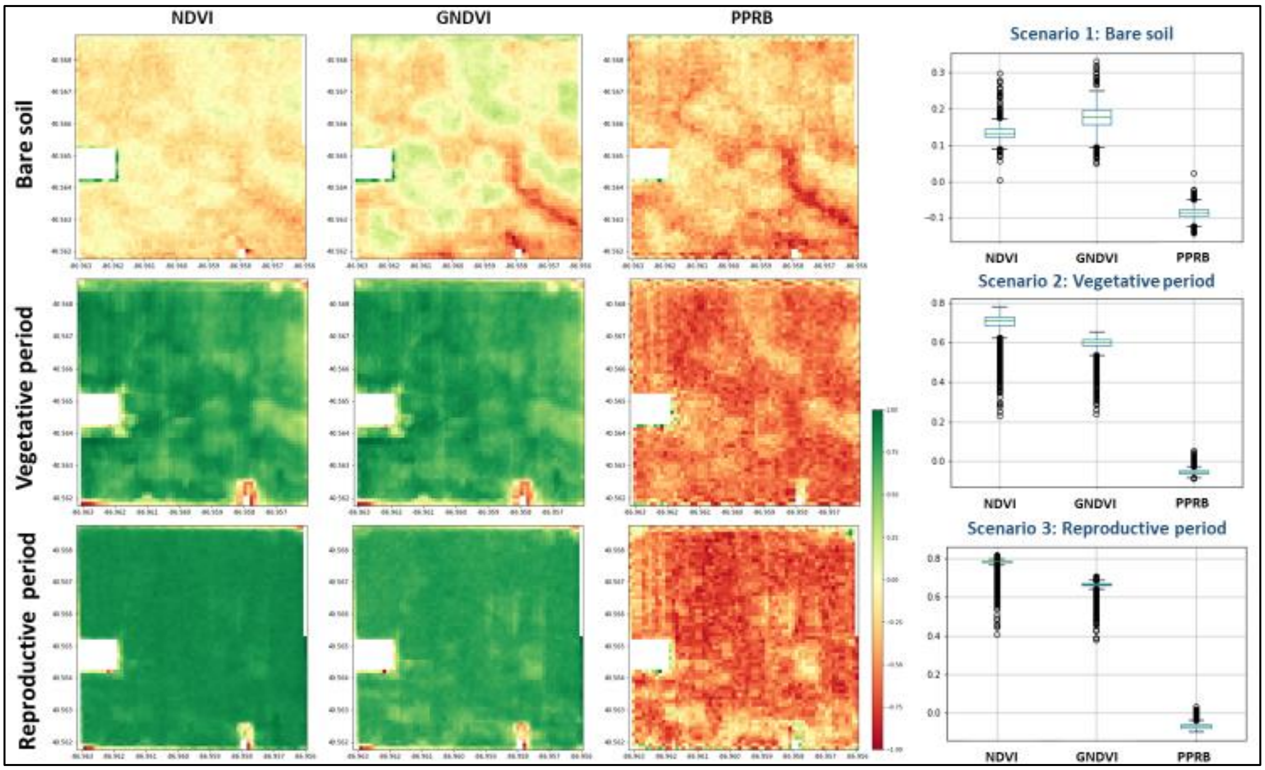

Figure 1. NDVI, GNDVI, and PPRB plots and box-whisker figures under three scenarios: 1) bare soil, 2) vegetative period, and 3) reproductive period.

Overall, there are visual differences between the three scenarios. Regardless of the index evaluated, the greatest variability of the field is observed under scenario 1 (bare soil) and the least variability under scenario 3 (reproductive period). This may be explained by the amount of soil background visible in the images from which the vegetative indices were derived. The visual difference between the three scenarios is enhanced by the vegetative index evaluated. For example, under scenario 3, GNDVI and PPRB show more variability compared to NDVI. As regards of data distribution, the box-whisker figures show that most the values are accumulated in a short range, particularly under scenario 3. In general, there was not an index that showed more variability for the three scenarios consistently. Outputs from the program obtained for this section are stored in the folder "Doug", which is within the folder "Outputs" located in the main directory.

Data quality checking

A key step when handling spatial data is to check that the coordinate reference system (CRS) is the same for the inputs to be processed. If CRS is not the same, then it is necessary to change it to a consistent CRS for all the inputs. The coordinate reference system for the sentinel images was EPSG:32616, while for the field boundary (shapefile) of the field of interest was EPSG:4326. Prior to conducting any further process, the CRS was set consistent for both inputs (images and field boundary). Another important step is to mask out "no data" values, since they can lead to parameters (e.g. NDVI) values that are out of range. This was achieved during the clipping process, in which the "no data" values were handled.

Although the search criteria available in the United States Geological Survey website allow users to define a cloud cover threshold, the lowest value is 10%. Therefore, it is important to verify that the image corresponding to the field of interest does not have clouds.

To achieve this, once the parameter of interest was calculated (NDVI), it was plotted for a visual assessment. For the purposes of this report, an example of a NDVI map derived from an image with no clouds versus a NDVI derived from an image with clouds are showed below for each of the three scenarios evaluated (bare soil, vegetative period, and reproductive period) (Figure 2), and the metrics (min, max, range, mean, and standard deviation) corresponding to each are summarized in Table 1. Outputs from the program obtained for this section are stored in the folder "Cloud_check" in the main directory.

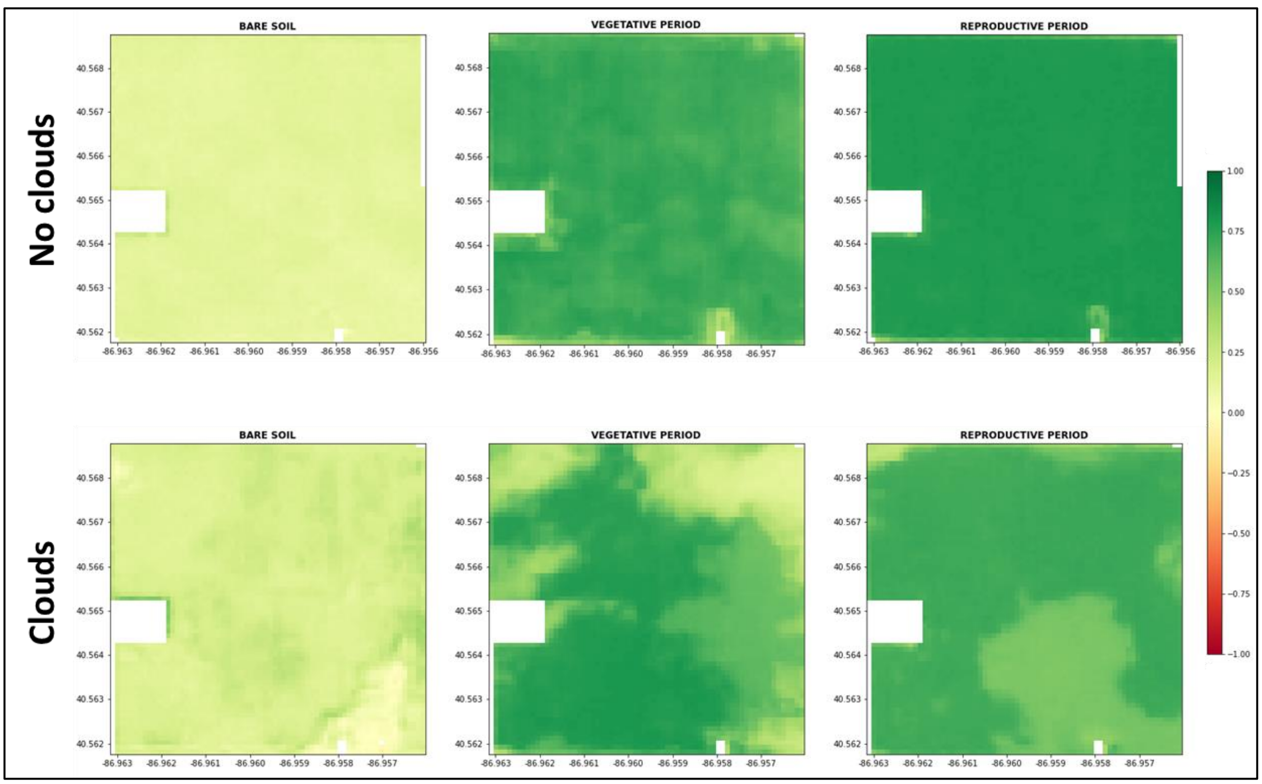

Figure 2. NDVI plots under three scenarios: 1) bare soil, 2) vegetative period, and 3) reproductive period. Top include NDVI maps derived from images without clouds, and bottom includes NDVI maps derived from images with cloud presence.

Table 1. Min, max, range, mean, and standard deviation of NDVI maps derived from images without and with clouds under three scenarios.

Metrics	Bare soil		Vegetative period		Reproductive period	
	No clouds	Clouds	No clouds	Clouds	No clouds	Clouds
Mean	0.13	0.19	0.70	0.61	0.78	0.65
Range	0.29	0.62	0.55	0.68	0.41	0.47
Min	0.00	-0.08	0.23	0.12	0.41	0.26
Max	0.30	0.54	0.78	0.80	0.82	0.73
StnDev	0.02	0.06	0.06	0.19	0.03	0.08

Summary of statistics and metrics

For each vegetative index calculated, stats (min, max, mean, and standard deviation) were calculated for a quantitative assessment. While the graphical data analysis showed the visual differences in spatial variability between the three scenarios based on NDVI, GNDVI, and PPRB, as well as their data distribution (box-whisker figures), this section is focused on the quantitative results obtained.

Table 2. Min, max, range, mean, and standard deviation of NDVI, GNDVI, and PPRB under three scenarios.

Metrics	Bare soil			Vegetative period			Reproductive period		
	NDVI	GNDVI	PPRB	NDVI	GNDVI	PPRB	NDVI	GNDVI	PPRB
Mean	0.13	0.18	-0.09	0.70	0.59	-0.05	0.78	0.66	-0.07
Range	0.29	0.28	0.17	0.55	0.42	0.14	0.41	0.33	0.13
Min	0.00	0.05	-0.14	0.23	0.24	-0.09	0.41	0.38	-0.10
Max	0.30	0.33	0.02	0.78	0.66	0.05	0.82	0.71	0.03
StnDev	0.02	0.03	0.02	0.06	0.04	0.01	0.03	0.03	0.01

Overall, regardless of the vegetative index evaluated, the lowest mean values correspond to scenario 1 (bare soil). This was expected since NDVI, GNDVI, and PPRB are plant status indicators. In absent of plants, it was expected that values were low since they are derived mostly from soil and plant residue. The amount of soil had an impact on the visual differences between the three scenarios, and it did too on the quantitative differences. The lowest vegetative indices values correspond to the scenario 1 and the greatest to scenario 3. Stats of NDVI and GNDVI results were closer to each other, in comparison with PPRB stats, which were consistently lower among the indices evaluated.

Conclusions

This project assessed spatial variability of large-scale fields (5 to over 40 ha) under three scenarios (bare soil, vegetative period, and reproductive period) using vegetative indices (NDVI, GNDVI, PPRB) derived from satellite imagery (Sentinel 2). Data quality check was important to make sure that all spatial data was on the same coordinate reference system, and that only satellite images without clouds were used in the analysis.

Plots of the three indices showed visual differences in spatial variability between the three scenarios evaluated, which were supported by the quantitative metrics of the indices. The results showed that most spatial variability was attributed to soil background. From this perspective, using a soil map for defining sampling locations would be enough given that most spatial variability showed on the NDVI, GNDVI, and PPRB derived from Sentinel data were attributed to soil differences. Here, it is important to keep in mind that due to the coarse spatial resolution of Sentinel images (10 m), it is mandatory to take in account the soil background in the calculation of vegetative indices.

It would be interesting to calculate the same indices from images with a higher resolution (e.g. 2 inches) and mask out the soil background to identify if the results are the same as using indices derived from satellite images with coarser resolution.

An important product from this project was the generation of a python program to handle, process, and organize inputs and outputs involved in the calculation of vegetative indices for the assessment of spatial variability. This program will become an important foundation for future programs created for the assessment of spatial and temporal variability of crops using remote sensing and crop simulation models.

Repository

https://github.com/amoralesona/ABE651-final-project-amoralesona.git