Image Processing Analysis Project EXAM

Xavier MARTIN

Promotion 2023-2024

Introduction

Thanks to the 12 band images of multiples dates of the boulinsard field, we want to extract already knowned associations of band that used the wavelength of the sentinel-II satellite. In fact, the 12 band images of Boulinsard are originated from the pictures taken by sentinel-II. These associations of wavelength are known in the modern professional agricultural sector as indices. The most known of these indices is the normalized difference vegetation index (NDVI). Here is a brief overview of the litterature among this subject.

State of art

- In biology of plant as the science of agronomy, numerous studies have shown that there is correlation between a specific index and what should be demonstrate by this index such as biomass, the foliar area, nitrogen content, etc[1,2] …
- The NDVI is the most popular indice to know the global heatlh of the vegetation but there is evidence of use for other indices:
  - In large density vegetation WDRVI can be better than NDVI (proof on winter wheat)[3,4]
  - In earlyer vegetation stage, GNDVI and EVI have better monitoring capacity than NDVI[5].

Introduction Sentinell-II

Data used for Boulinsard field analytics are sattelite-images. There is a huge amount of sattelites that take pictures of the earth surface worldwide. Each sattelite have difference frequencies sensor to take picture of a same space and take different band with different resolution level. For the field we study there, its place is in Normandy and all images used are taken from the sattelite sentinell II. Above you can see the sentinell-II bands that we have access to.

Sentinel-2 Bands	Central Wavelength (µm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

There is 12 (to 13 ?) bands generated in the multi-spectral images taken by the sentinel-II satellite.

Framed, there are the band actualy used by the indices.

We are aloso going to use for machine learning the 3 band of vegetation red edge (5, 6, 7).

Indices

In this diapositive, you can see the different reference use to assess indexes for this study.

All these indices where choosen because they seems to be covering vegetation khey criteria and are sufficiently different from NDVI.

Further in this analysis we will see how these indices evolve with each other.

Here is a website with all the vegetation indexes and all the corresponding formula for the different sattelite existing worldwide : https://www.indexdatabase.de/db/i-single.php?id=104

VI	Abbreviation	Equation	Reference
Green chlorophyll vegetation index	GCVI	$GCVI = \left(\frac{NIR}{G}\right) - 1$	Gitelson et al. (2003)
Green normalised difference vegetation index	GNDVI	$GNDVI = \frac{NIR - G}{NIR + G}$	Gitelson et al. (1996)
Normalised difference vegetation index	NDVI	$NDVI = \frac{NIR - R}{NIR + R}$	Rouse et al. (1973)
Simple ratio	SR	$SR = \frac{NIR}{R}$	Jordan (1969)
Wide dynamic range vegetation index	WDRVI	$WDRVI = \frac{0.2 * NIR - R}{0.2 * NIR + R}$	Gitelson (2004)

$$\text{Enhanced Vegetation Index, EVI} = 2.5\frac{R_{842} - R_{665}}{R_{842} + 6R_{665} - 7.5R_{490} + 1}$$

$$\text{Normalized Multiband Drought Index, NMDI} = \frac{R_{842} - (R_{1610} - R_{2190})}{R_{842} + (R_{1610} - R_{2190})}$$

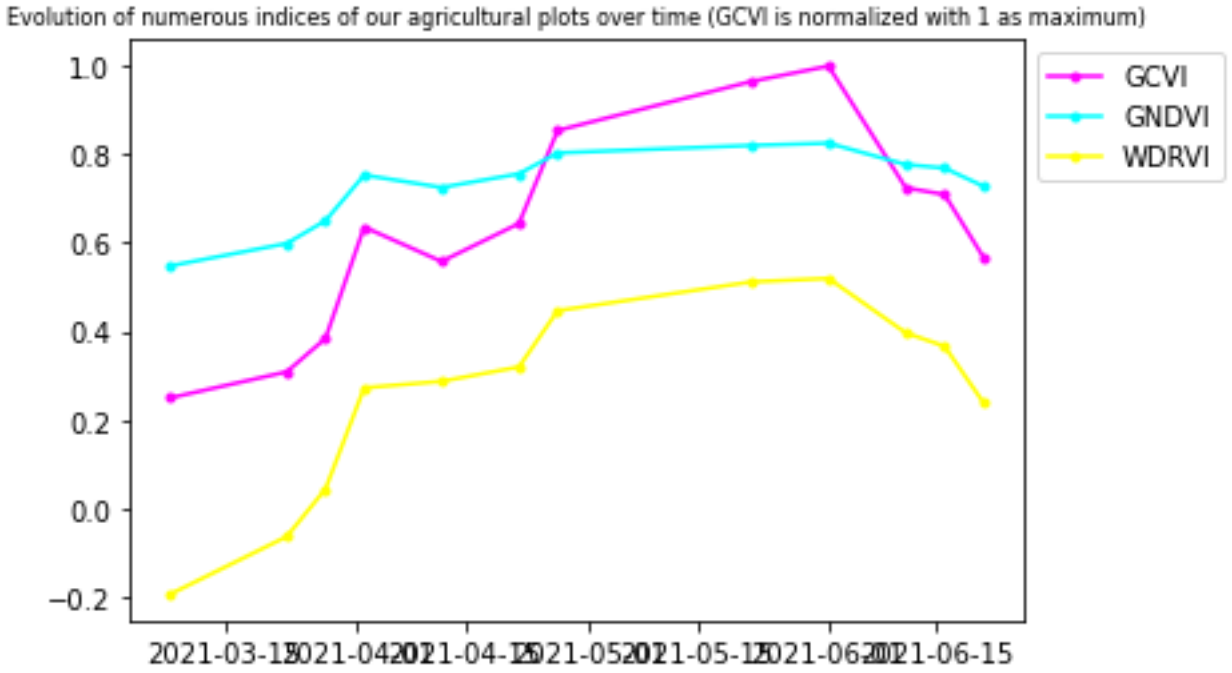
$$ARVI = \frac{B8-(2*B4-B2)}{B8+(2*B4-B2)}, \quad OSAVI = \frac{(1+0.16)B8-B4}{B8+B4+0.16}$$

Indices GCVI, GNDVI et WDRVI

- Green chlorophyl vegetation (GCVI) index try to estimate chlorophyll content in the plant
- Green normalized difference vegetation (GNDVI) index try to estimate the active photosynthetic region of the plant with green band at the place of the red band, with more accuracy than NDVI in early stage, or in contrario late stage with high density but less in middle stage.
- Wide dynamic range vegetation index, (WDRVI) is a modified NDVI indexes with more impact of the NIR band allowing more biophysical caracteristics of the plant such as Leaf Area (monitor by LAI) taken into account than in NDVI.

Temporal evolution for GCVI, GNDVI and WDRVI

We can see for these 3 indexes an increase from the 12 of march to the 1st of april, then a decrease for GNDVI and GCVI



The WDRVI is still increasing after the 12 of april indicating that biophysical factor like leaf area are still increasing after this date but much more slowly.

The GNDVI indicating a slight decrease in photosynthetic activity whereas the GCVI indicating a slightly important decrease in chlrophyl content.

All of this indicating a stress factor that souldn’t be hail because there is no decrease in LAI or biophysical factor (WDRVI).

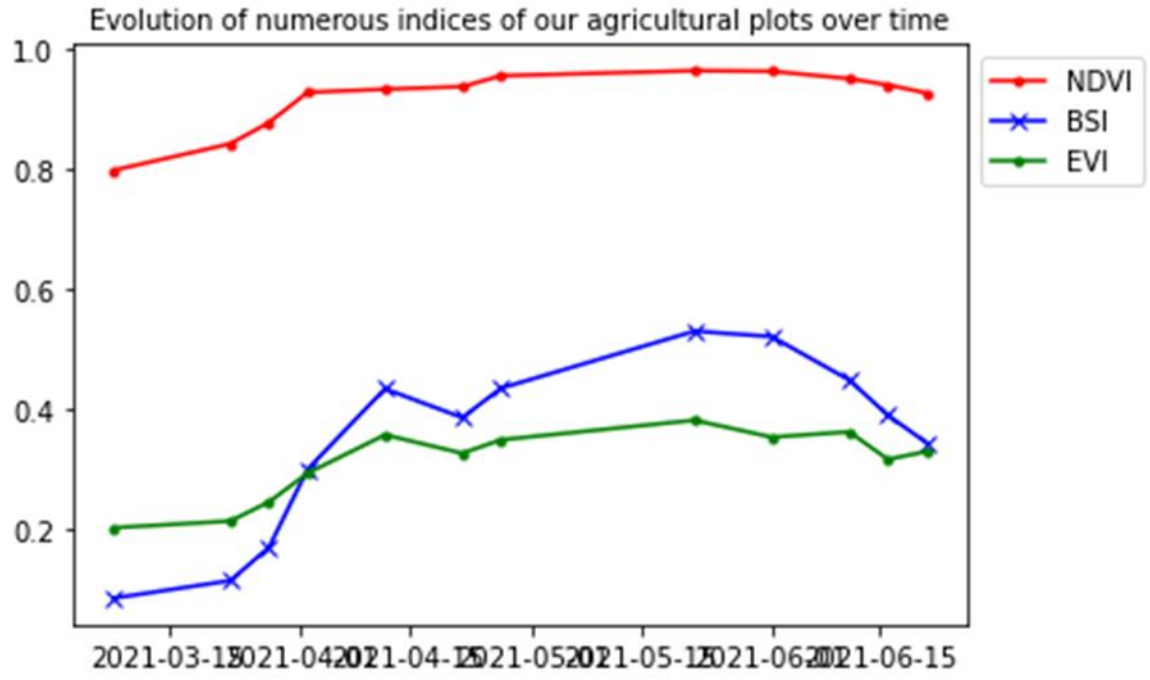
Next evaluation will add more indices of what is this stress

Indices NDVI, BSI, EVI

- Normalized vegetation difference index (NDVI) is the most popular index, it’s help to see stressed area or nutrient defficiency. It’s usefull for middle stage as it’s the more accurate index.
- Bare soil index (BSI) is indincating exposed bare soil area, whereas negative values is indicating vegetation area
- Enhanced vegetation index (EVI) is like the WDRVI with one more band, the blue, in the aim to deacrease the impact of atmospheric/background influences.

Temporal evolution indices NDVI, BSI and EVI

The NDVI is not showing strong evolution modification but we can see a kind of correlation with WDRVI.



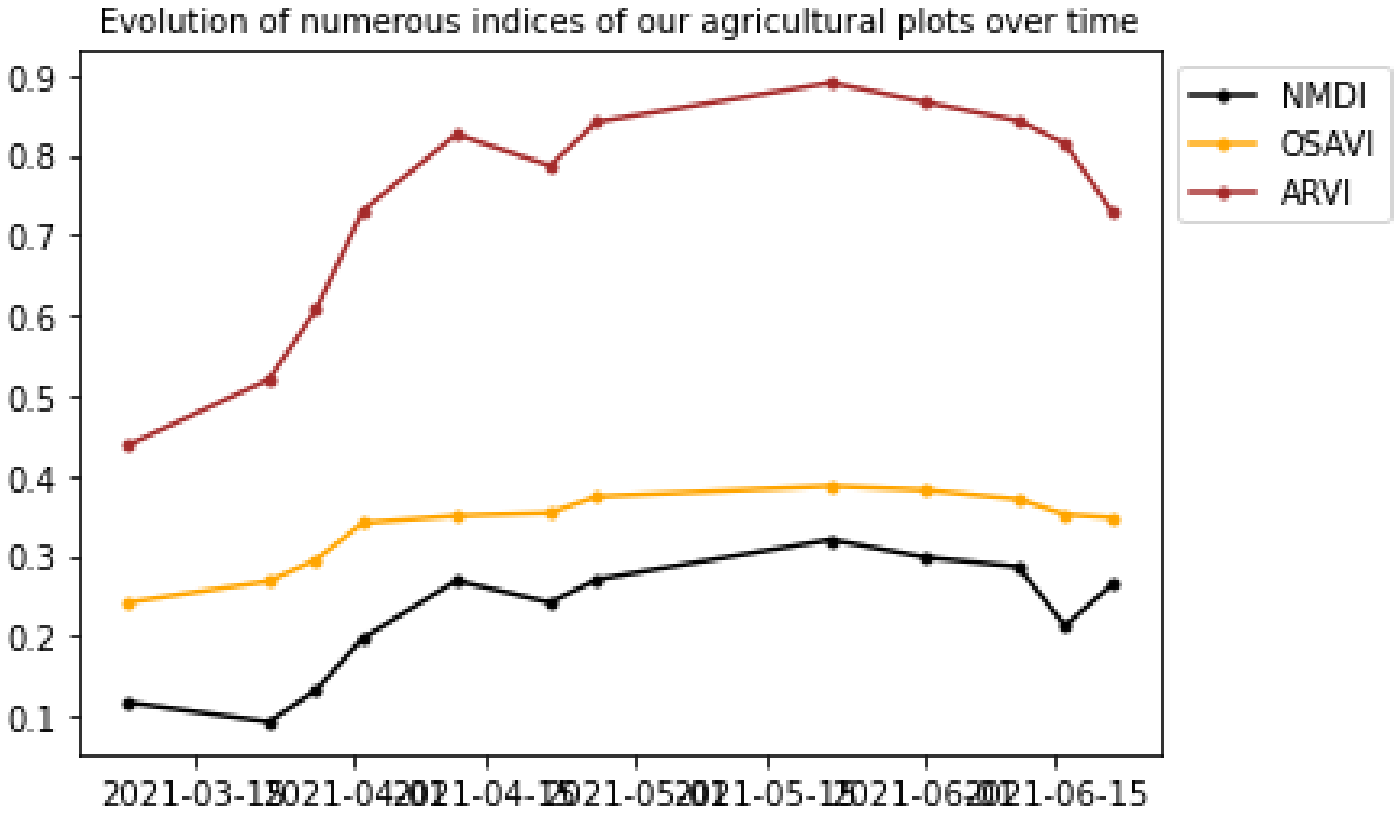
The BSI is the bare soil index. This index is indicating the amount of soil area visible. To show its curve more simply, i add a minus at the begining of the formula. The augmentation in BSI is indicating less soil area visible. Strangely there is a slight decrease, so in consequence an increase in soil detection after (between the 17 april and the 25). The EVI is also decreasing between this two dates, indicating that maybe that there was a decrease in leaf area or biophysical factor, that weren’t shown by the WDRVI because this isn’t taking into account the atmospheric incidence.

Indices evolution NMDI, OSAVI, ARVI

- NMDI is used to monitor drought for example in forest to monitor fire risk
- OSAVI is used to measure green biomass in environnement with low-density vegetation like semi arid region.
- ARVI is like NDVI but with some band to decrease aerosol diffusion effect. It’s widely used in dusty region

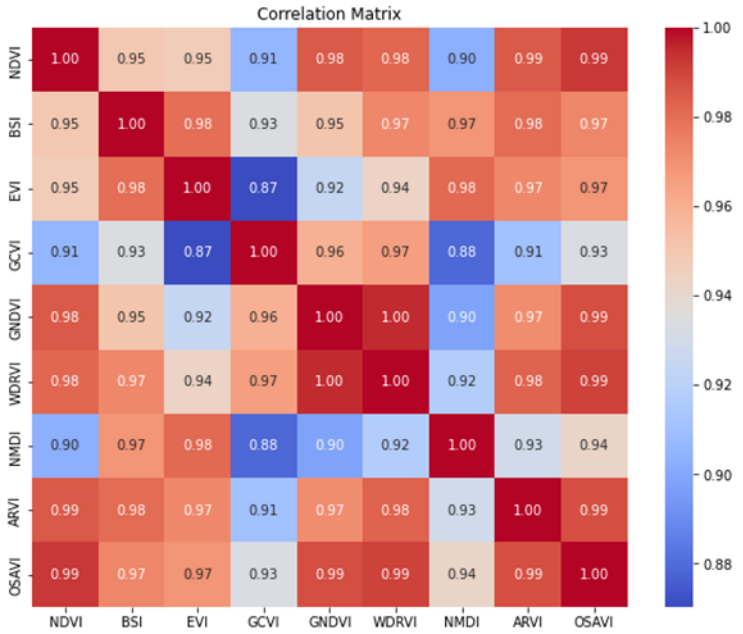
Temporal evolution of NMDI, OSAVI and ARVI

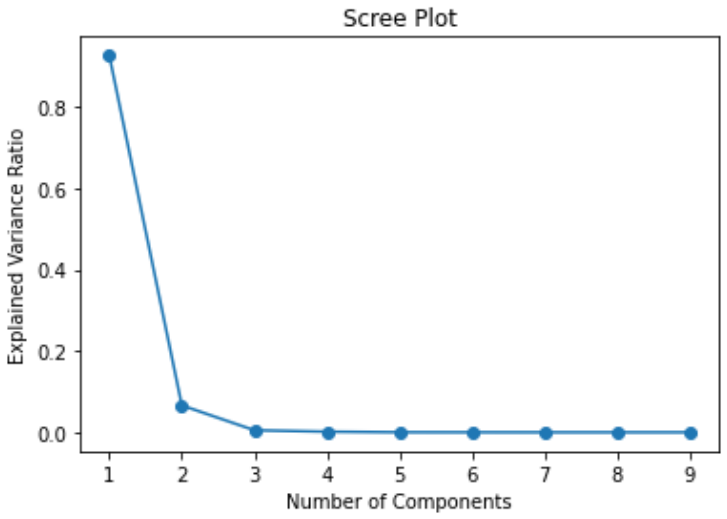
The OSAVI show very few modification in this evolution, while the ARVI show greate evolution modification.

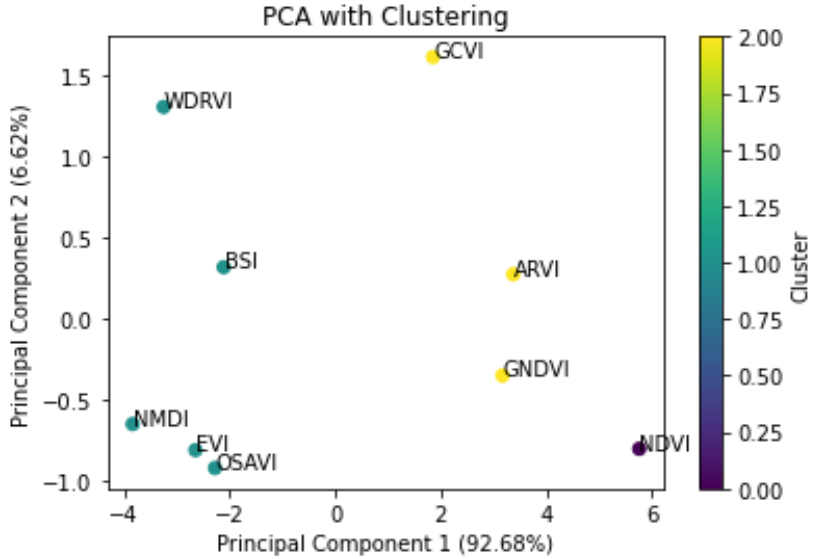


The ARVI show a very important evolution at the begining. This index is like the NDVI while taking into account aerosol. It seems like it will have a lot of aerosol in this regions (that wasn't especialy expected because normandy is not known for that) as it show a lot of more amplitude of variation than for NDVI. It seems like the decrease in the heatlh of plant appearing after the 12 of april is not that important comparing to the previous growth. The slight decrease is NMDI is maybe not the cause of the stress as it appears to be decreasing also in the begining of the growth process.

Correlation matrix and PCA and cluster of indices





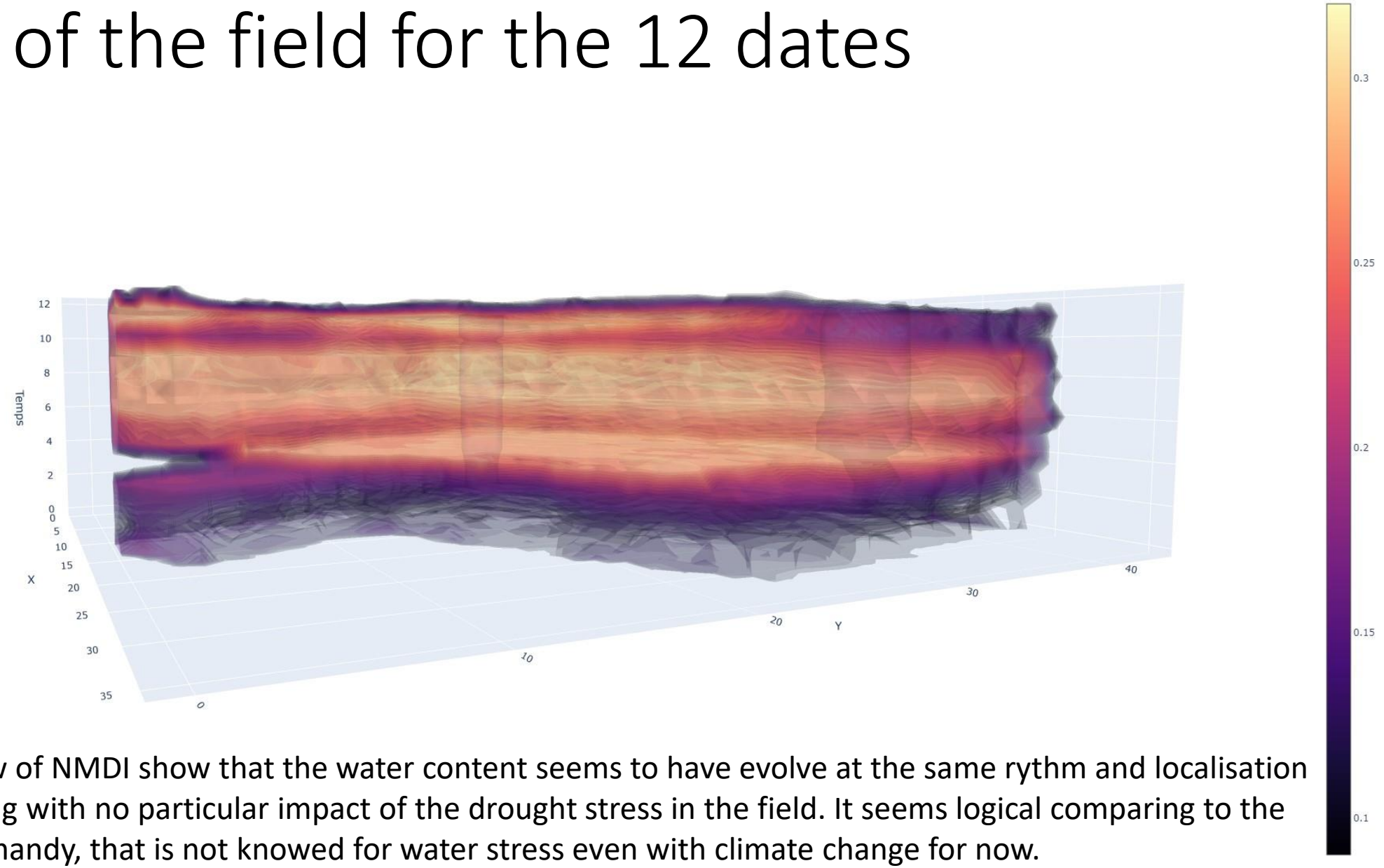


The correlation matrix show very close proximity between GNDVI and WDRVI that are not in the same cluster. It's difficult to interpret this because the GNDVI and the GCVI are in the same cluster but are very slightly correlated 0.92 comparing to other matches between indices. WDRVI and GNDVI show a 1.00 correlation but are largely separated on the PCA both ont first and second components.

Conclusion on the indexes analyses

- The most decreasing index directly assessing plant factors comparing to the global amplitude modification of the curve after the 12 april was the GCVI that is the most correlated studied indexes with chrorphyll and so on, nitrogen.
- The most suceptible stress that should have provoced the deacrease in the growth after the 12 of april seems to be a lack of nitrogen in the soil.
- This was probably during the grain filling as it corresponds the phenological stage in the global indicator evolution of the wheat and as is it the more suceptible stage for lack of photosynthetic activity and chlorophyll content for wheat during its growth stage.

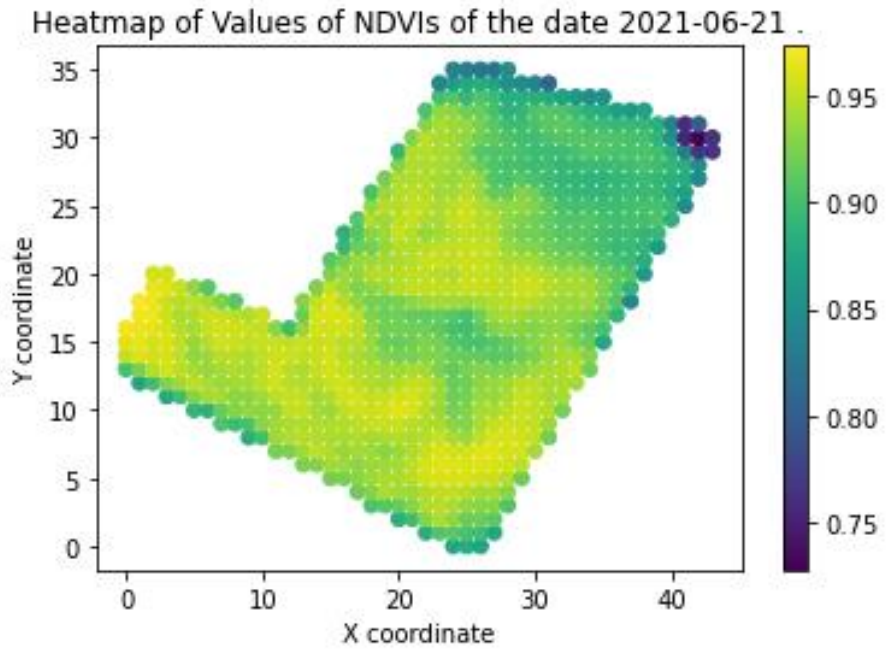
NMDIs of the field for the 12 dates

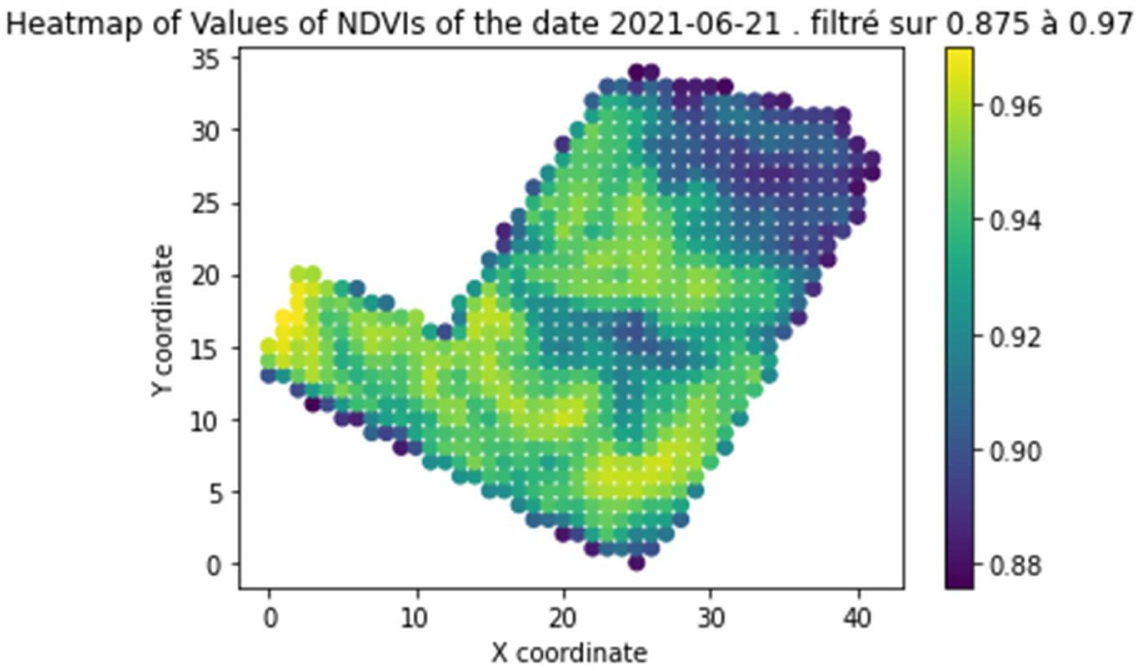


This tempo-regional view of NMDI show that the water content seems to have evolve at the same rythm and localisation as other indices indicating with no particular impact of the drought stress in the field. It seems logical comparing to the region of the field : Normandy, that is not knowed for water stress even with climate change for now.

Geospatial clustering : NDVI

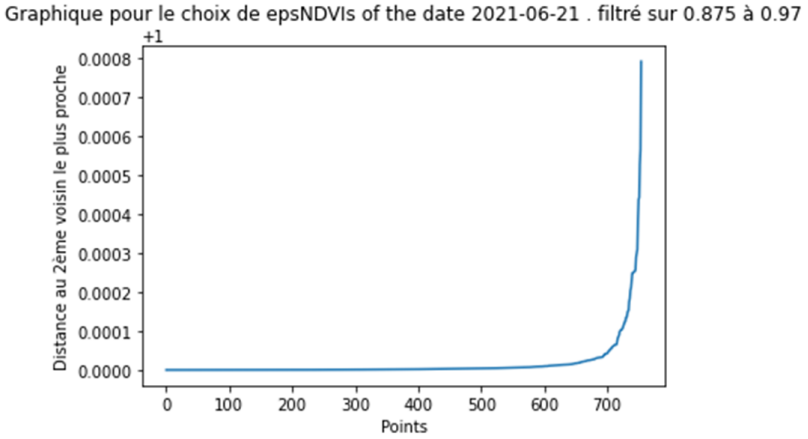
Cleaning of extreme value allow largely better view of regional spatialisation of the NDVI values



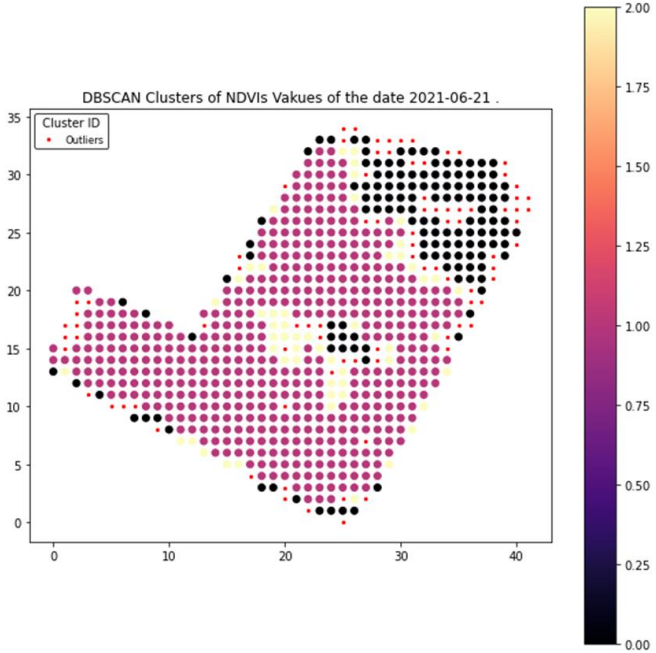


Geospatial clustering with DBSCAN

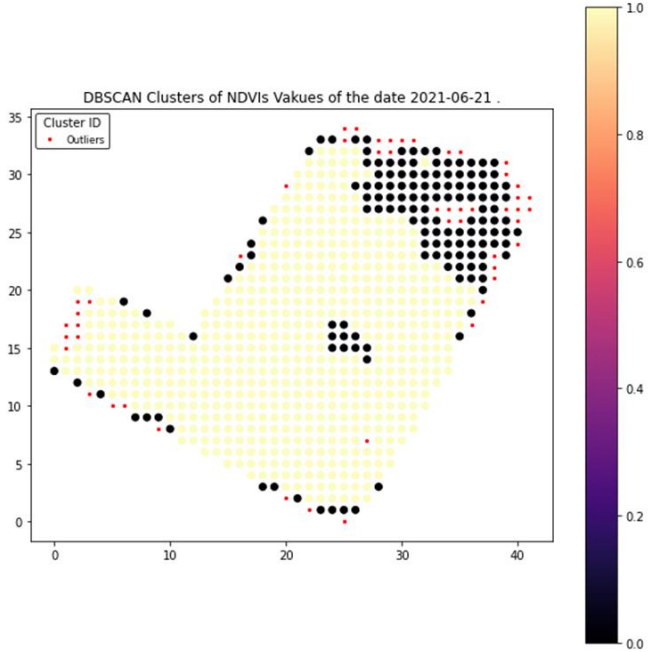
Distance for choice of the eps value for the Debscan and use of different EPS with 20 individuals by samples



DBSCAN with eps=0,0015 and min_samples = 20



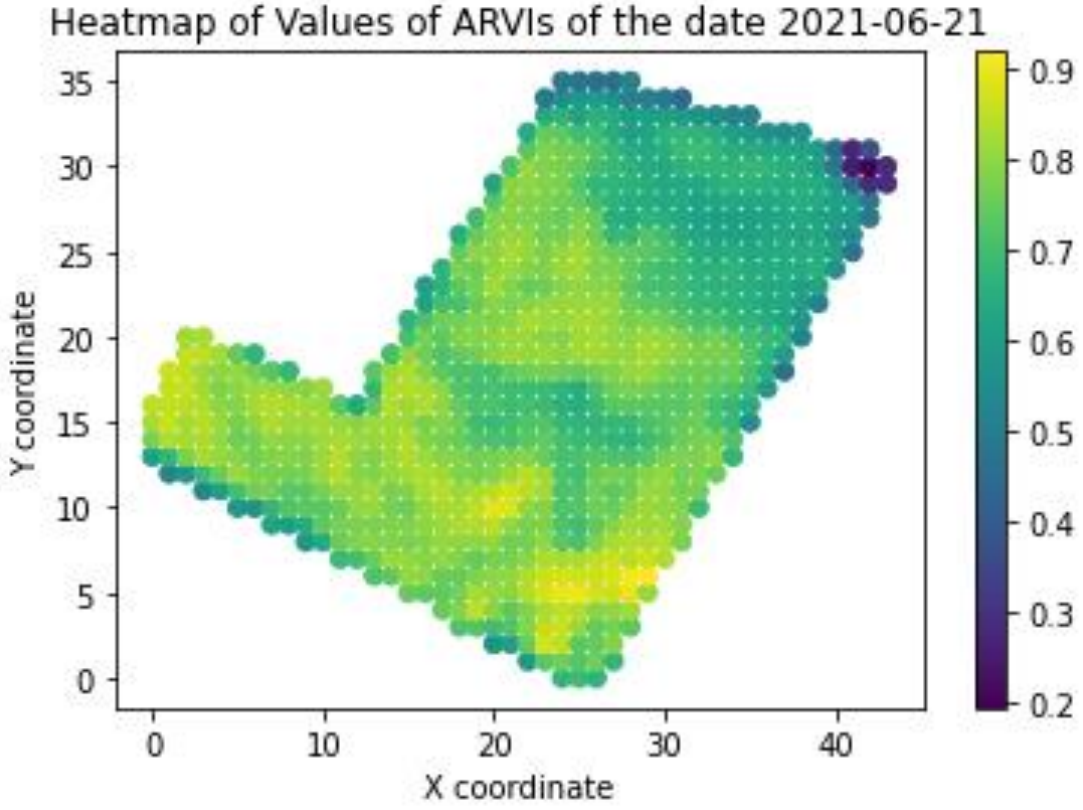
DBSCAN with eps=0,001742 and min_samples = 20

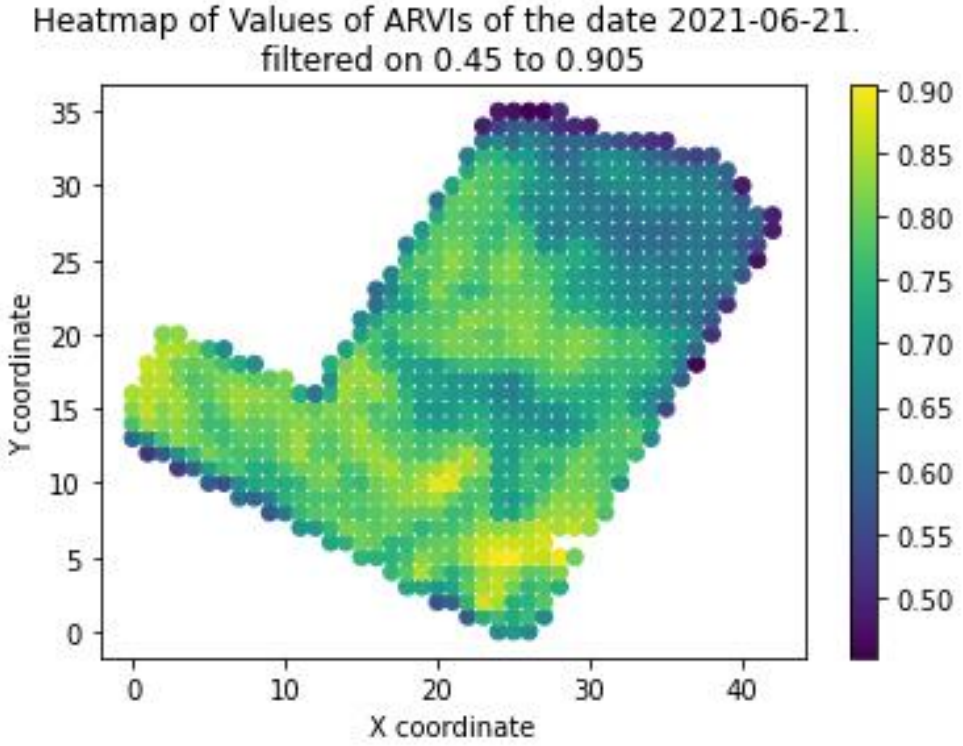


It seems to have a middle and left-right area with less plant general health than in the other part of the field.

Geospatial clustering : ARVI

Cleaning of extreme value for ARVI is also allowing better spatialisation of the field with the same pattern in the first view than which obtain with NDVI values.



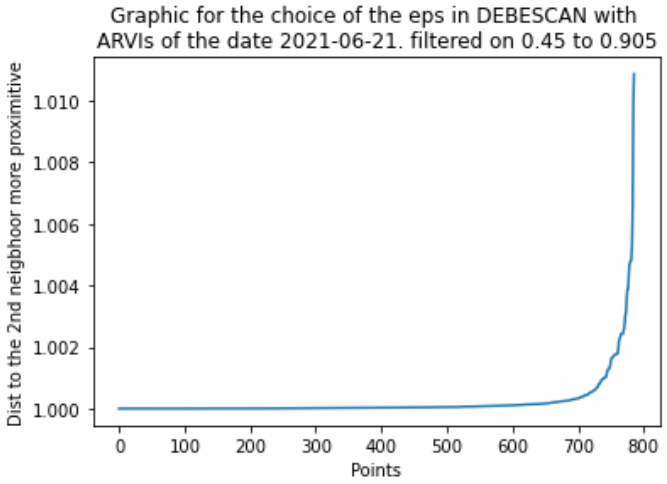


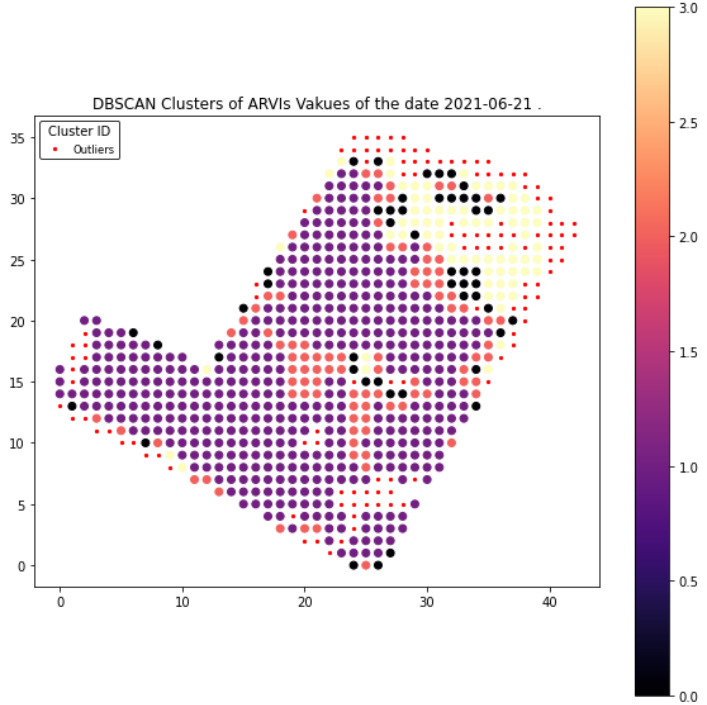
Geospatial clustering DEBSCAN

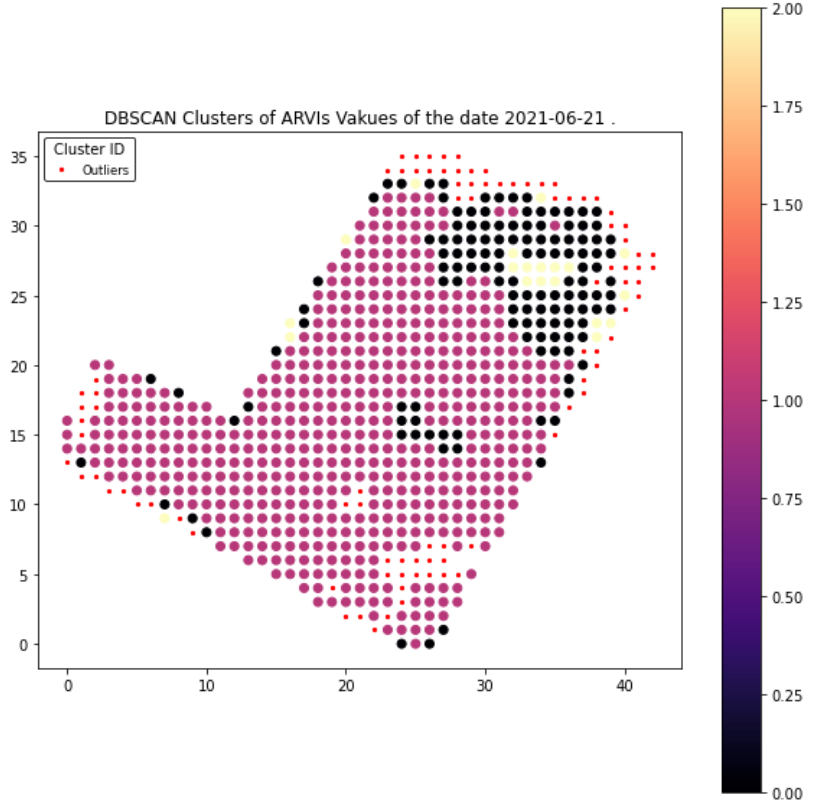
Distance for choice of the eps value for the Debscan and use of different EPS with 16 and 20 min samples

Debescan eps 0,055 min_sample = 20

Debescan eps 0,0413 min_sample = 16

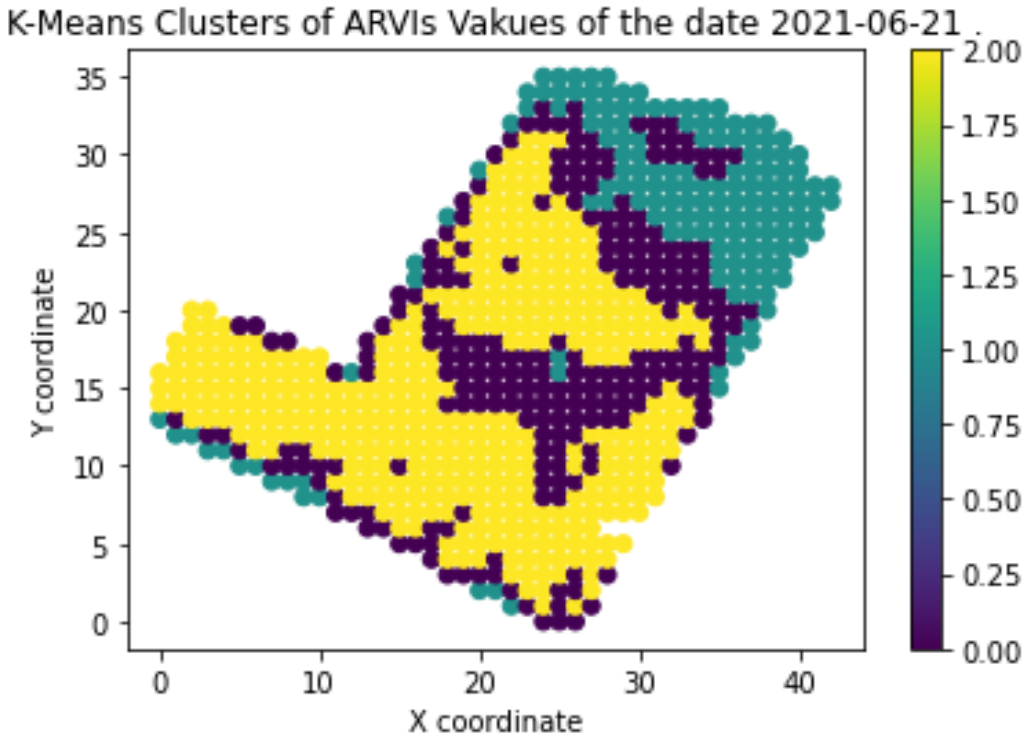






Geospatial clustering K-means

N_clusters = 3, random_state = 0

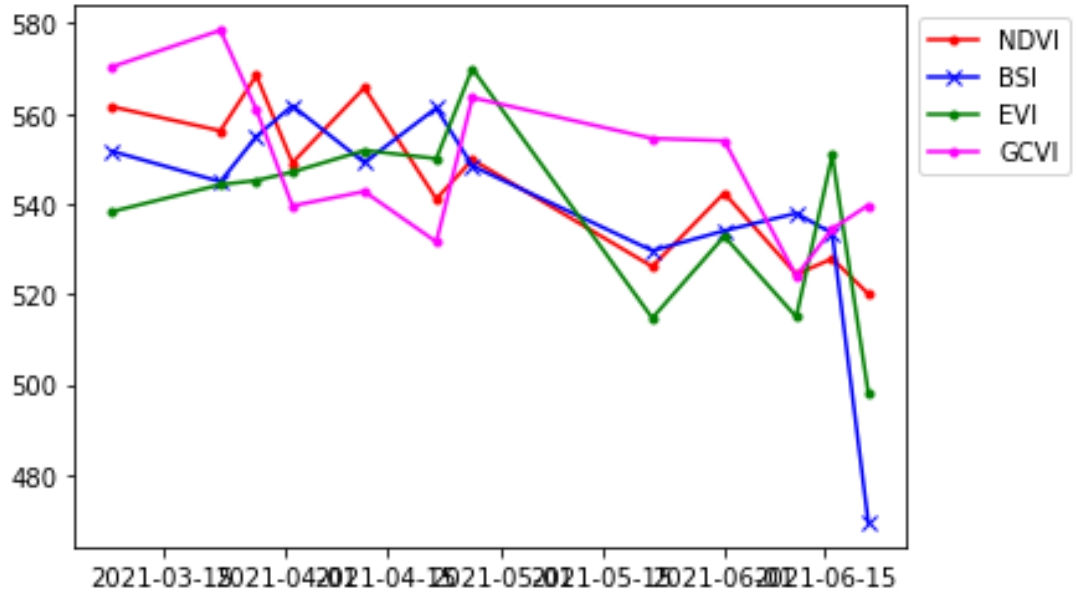


K-means seams to be bettter for clustering as it integrated more points into each cluster. The DBSCAN is still intersting as it allowed more restricted and maybe accurate cluster. A middle decision between this two clusters can be interresting fot the agricultor.

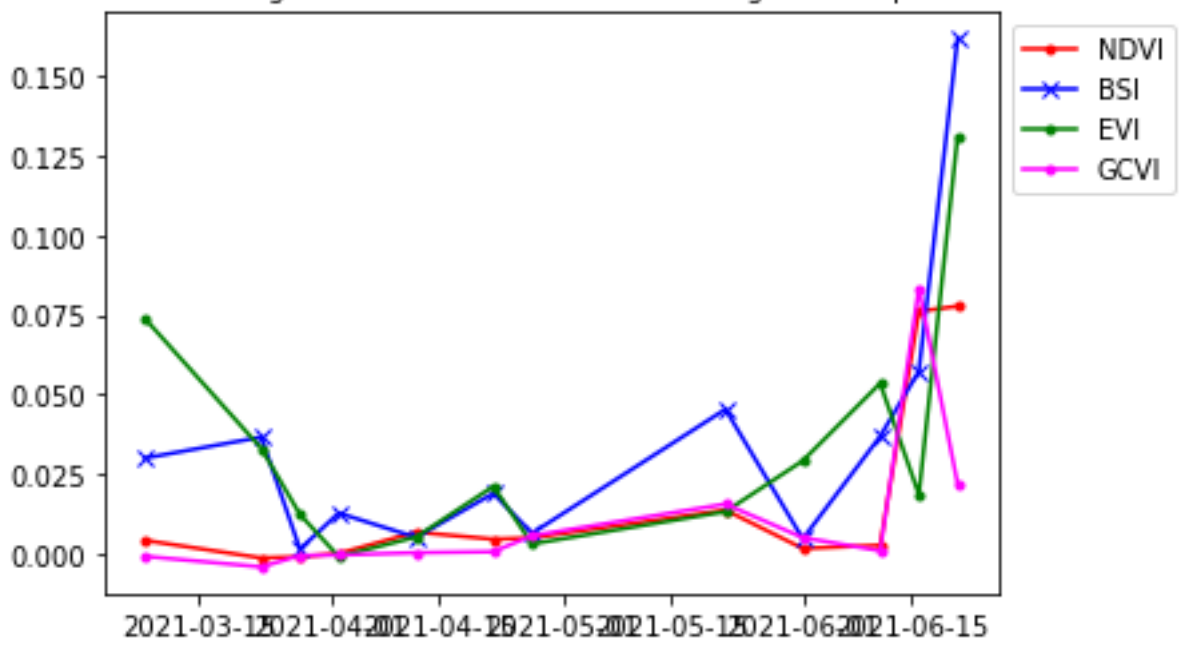
Machine Learning

- In the final stage, we are going to assess the yield on each date with SktLearn
- First we fit the model with linear regression and then we predict the yield with the point of each image for a specific index and a specific date
- Then we assess the error level for each indicator and dates
- Then we create images of the prediction and present some of them

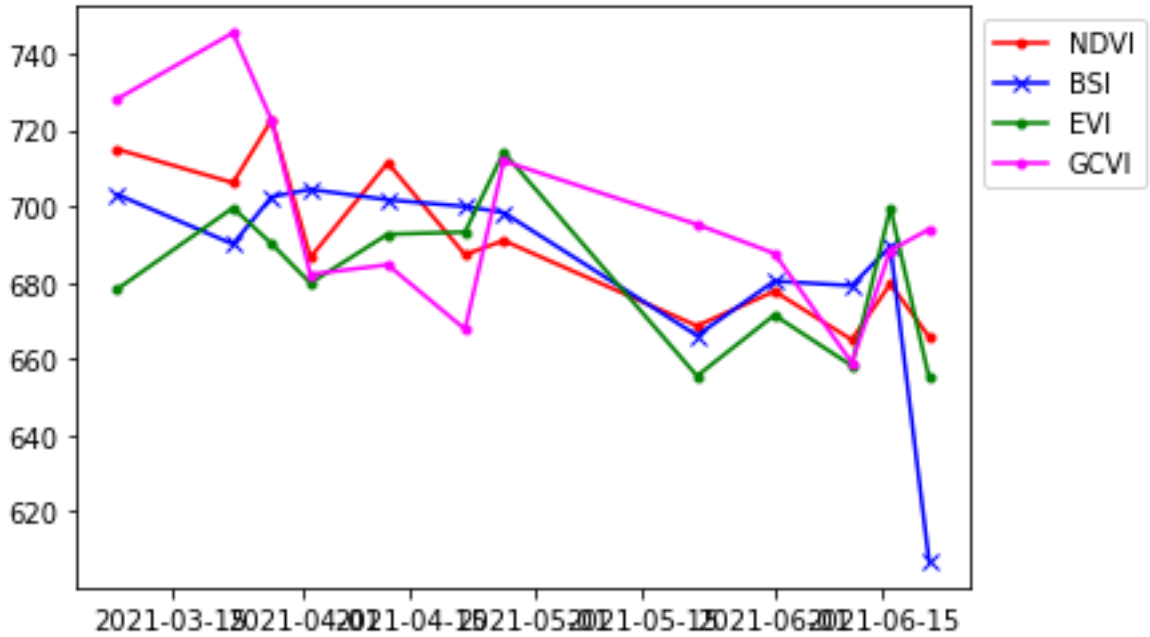
Mean absolute error score of featuring Yield of numerous indices of our agricultural plots over time



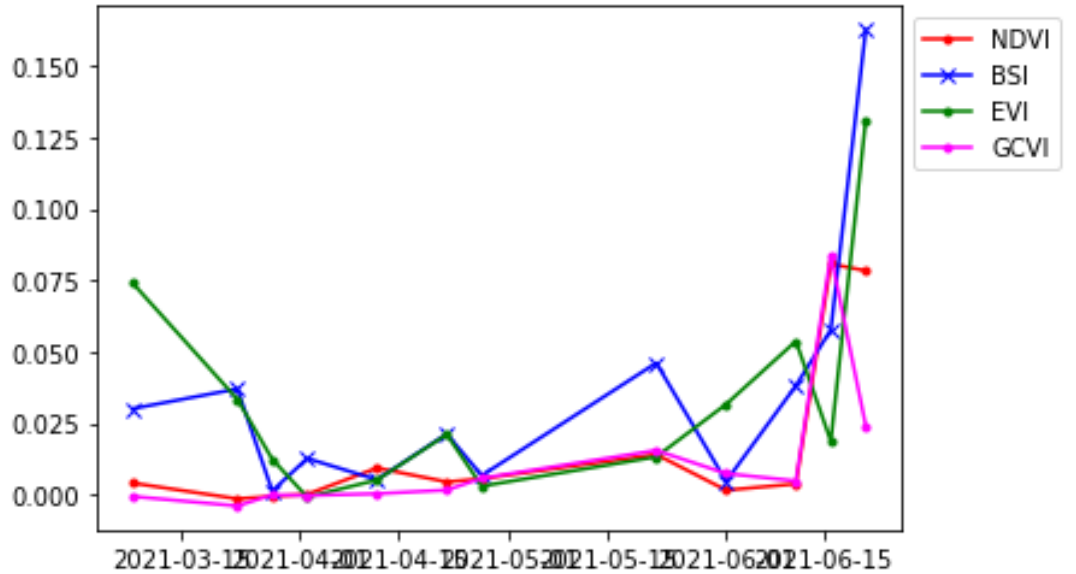
R² score of featuring Yield of numerous indices of our agricultural plots over time



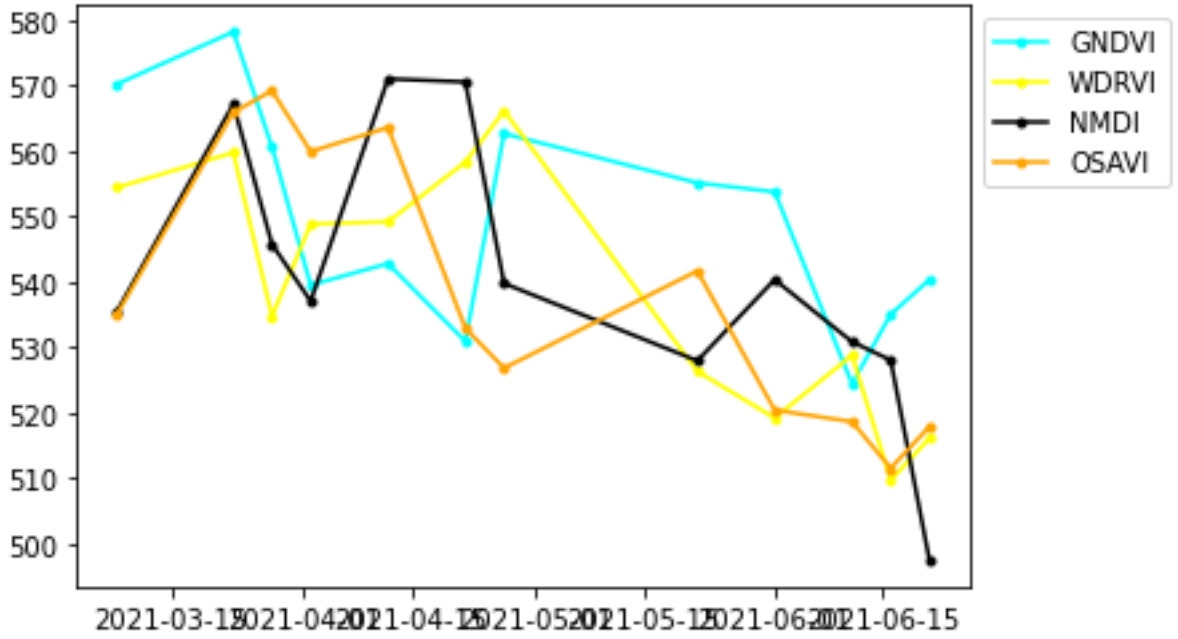
Root mean square error of featuring Yield of numerous indices of our agricultural plots over time



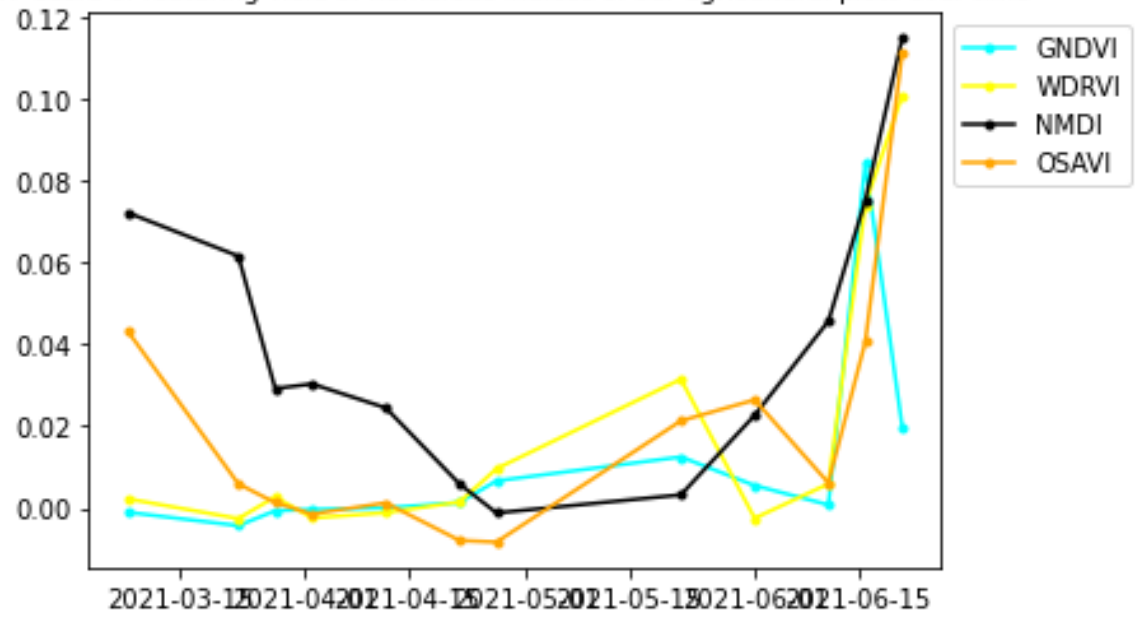
Explained variance score of featuring Yield of numerous indices of our agricultural plots over time



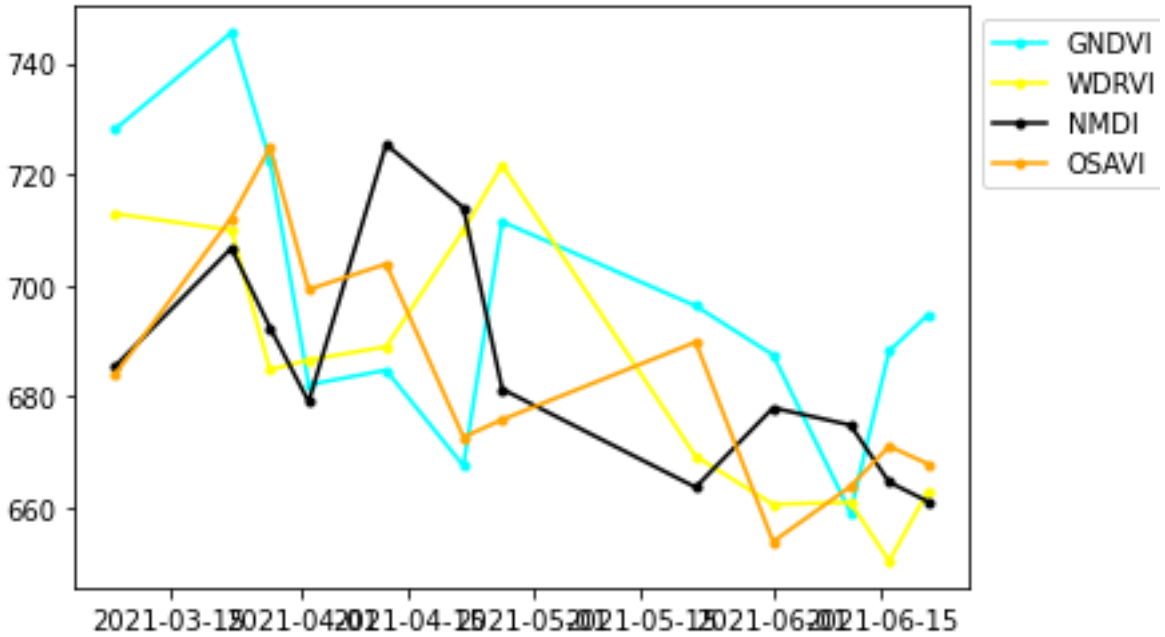
Mean absolute error score of featuring Yield of numerous indices of our agricultural plots over time



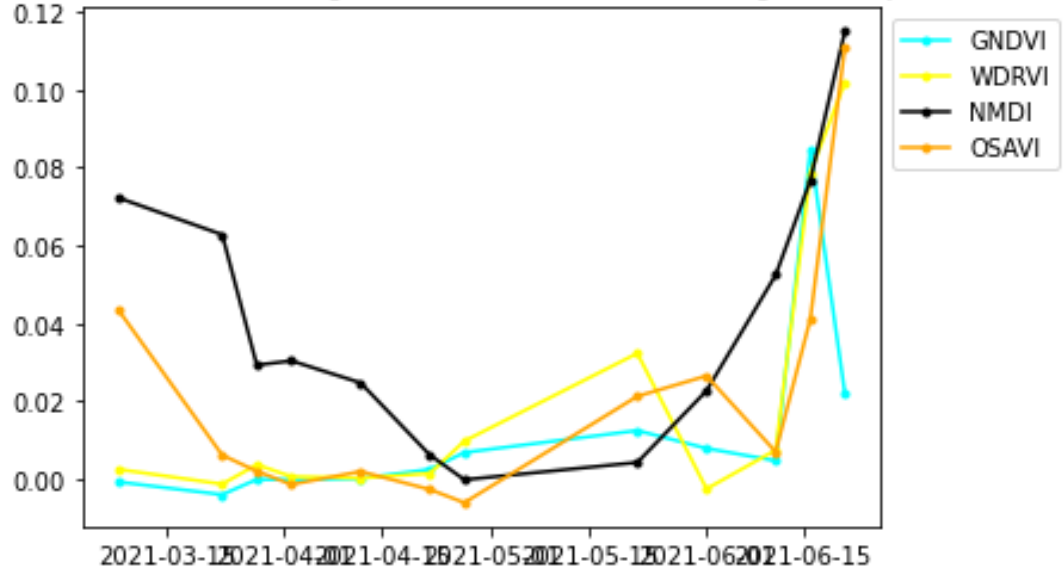
R² score of featuring Yield of numerous indices of our agricultural plots over time



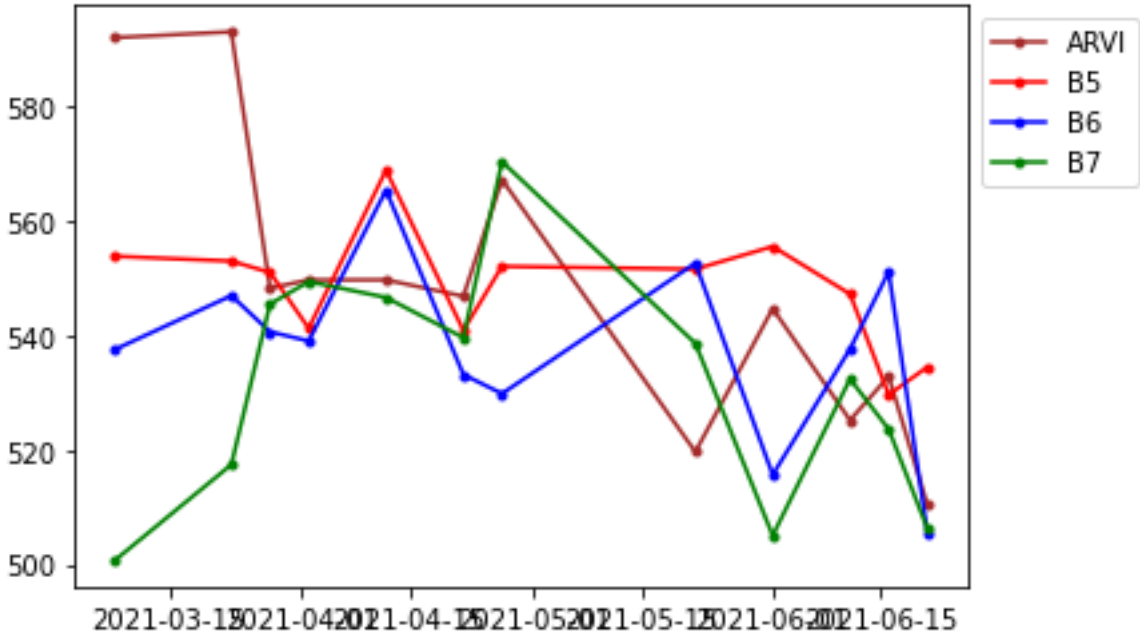
Root mean square error of featuring Yield of numerous indices of our agricultural plots over time



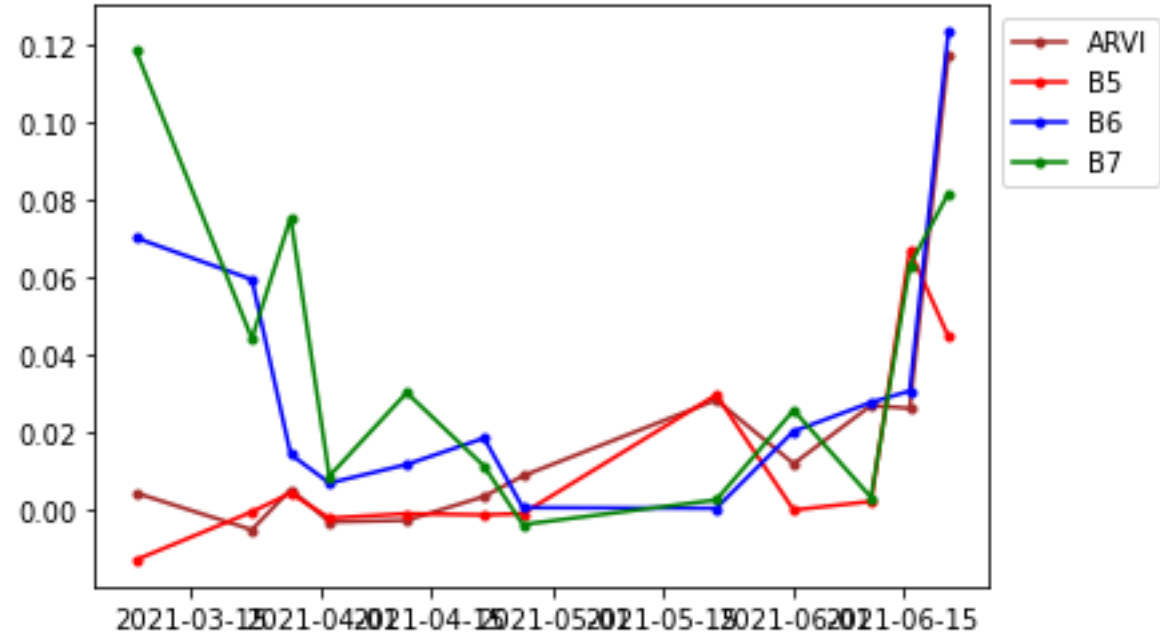
Explained variance score of featuring Yield of numerous indices of our agricultural plots over time



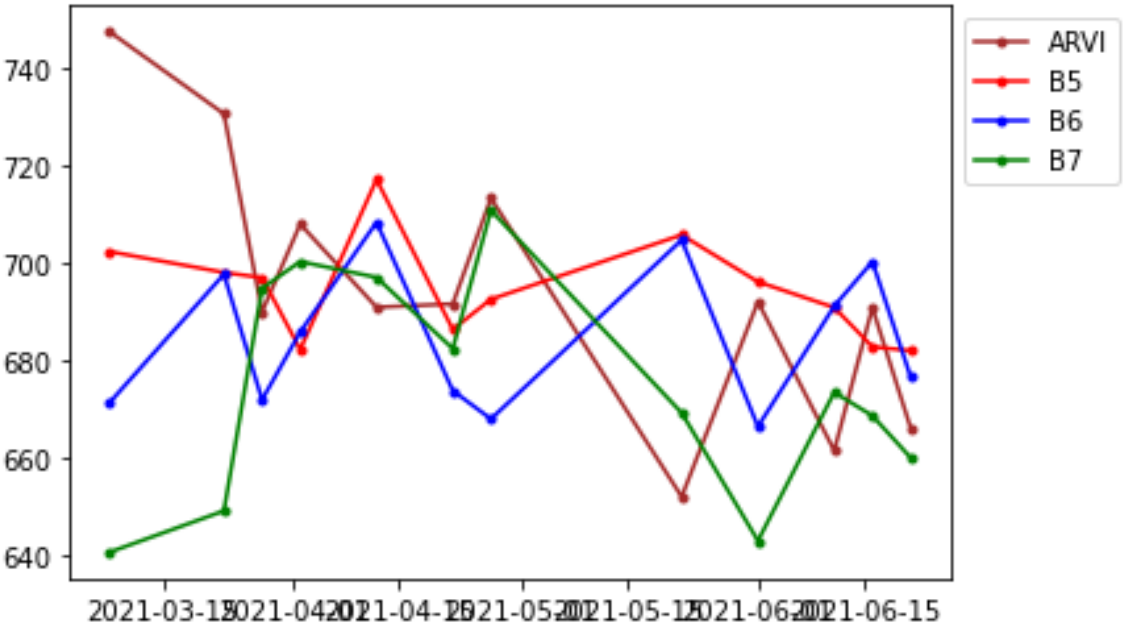
Mean absolute error score of featuring Yield of numerous indices of our agricultural plots over time



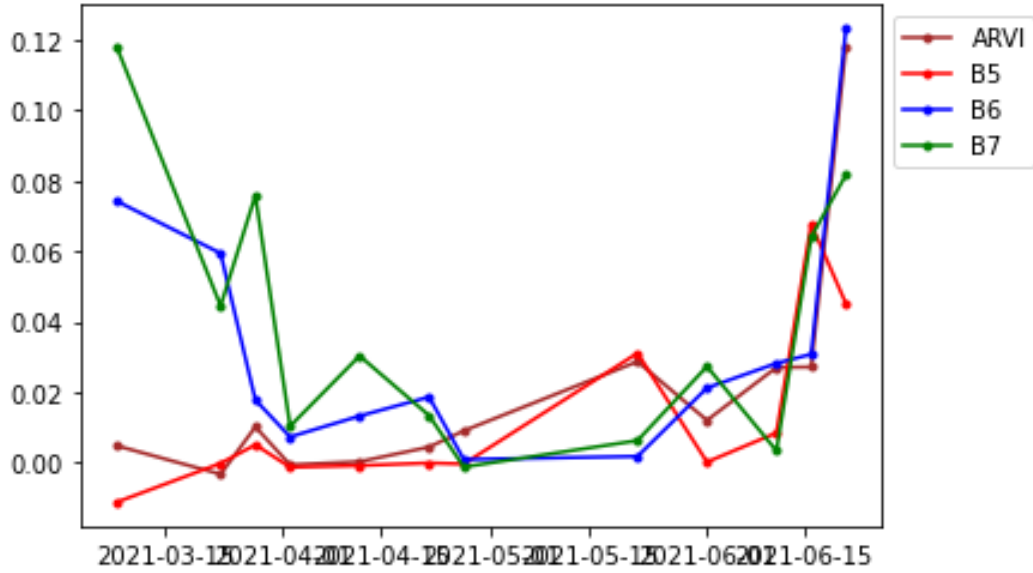
R² score of featuring Yield of numerous indices of our agricultural plots over time



Root mean square error of featuring Yield of numerous indices of our agricultural plots over time



Explained variance score of featuring Yield of numerous indices of our agricultural plots over time

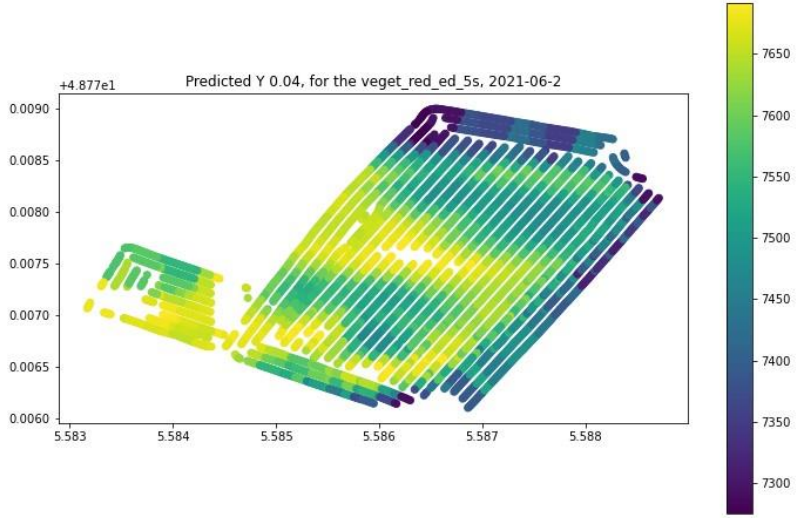


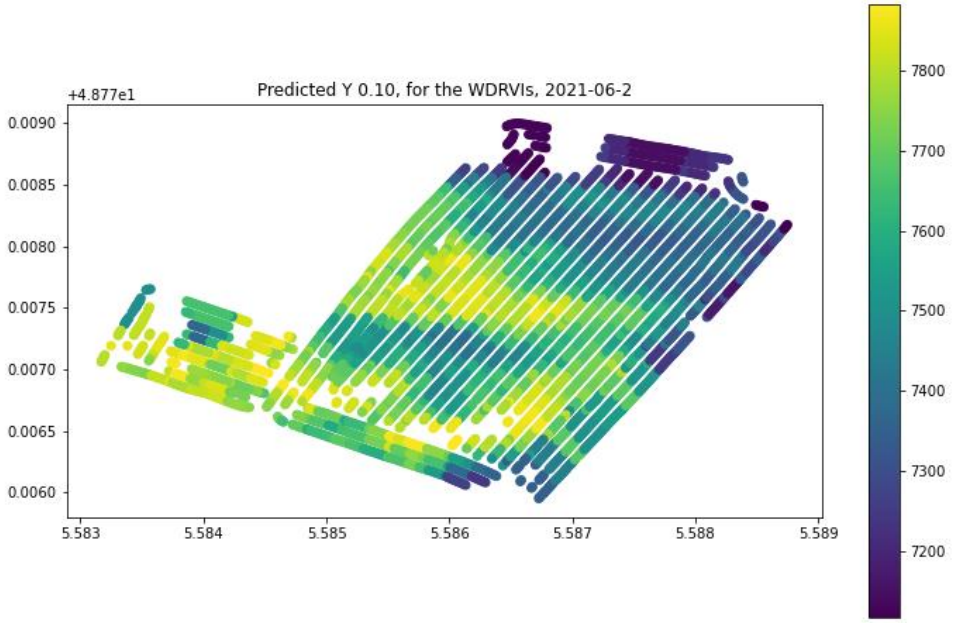
Synthetic view of the result of error and explainded variance for indexes

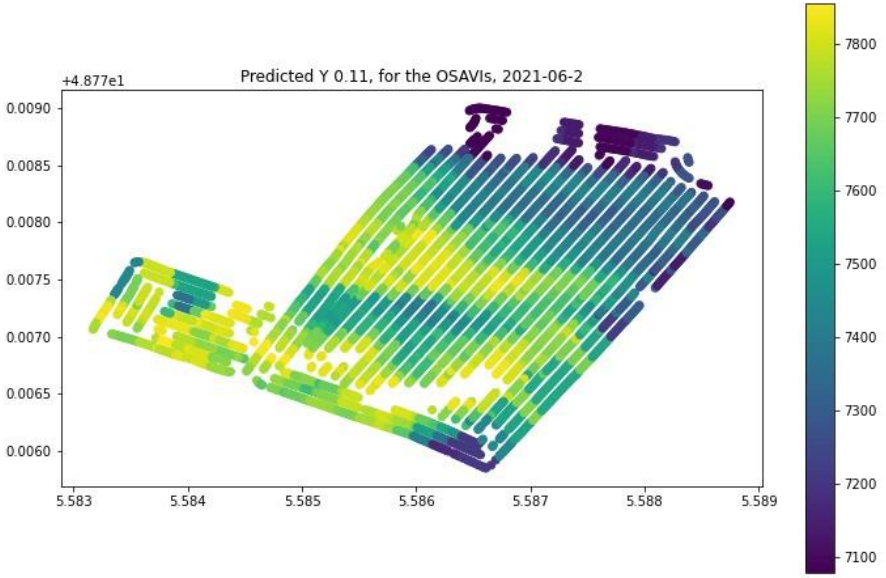
- Band 7 of sentinell-II seems to have a better correlation to the yield the earlyer in the developpment stage.
- Strangely BSI seems to have the better correlation in the late stage with yield
- Even more suprisingly, NDVI seems to be on the badest indicator of the yield
- NMDI seems to have good predictive score for both early and late stage but bad for middle stages.
- Lauched more machine learning prediction maybe can help to differenciate this indices

Best predictions

Here is an overview of the best between 144 (12 dates by 12 indicators) predictive model of the yield using specific date of indocator







Perspective

- Using multivariable area data (like multiple date selection or multiple indexes) can help
- Launch multiple machine learing trial
- Use different machine learning model
- Use deep learning
- NMDI seems to be strongly correlated to the yield obliging us to reconsider the problem of drought or water status in yield
- Use of specific band of the sentinell-II for the vegetation especialy the band 7
- Maybe associate this 7 band with NMDI and WDRVI because it’s the most correlated indexes with GCVI (as GCVI seems to be strongly correlated to the stress moment encounter in this wheat field) but GCVI show poor ability to predict the yield.
- Adding also in that new indexes if it’s possible the inverst of BSI as it whow great ability to predict the yield
- Take into consideration blue band in indexes to decrease atmorspheric/aerosol effect

Conclusion and solution

- The stress seems to be due to the nitrogen content but we shouldn’t totaly exclude the water stress
- Use a new indexes that is a combination of other indexes and use band 7 of sentinell III
- Add more nitrogen in the upper-right and central area of the field before the grain fill if the soil cationic exchange capacity and depth allows to retains it (and maybe less nitrogen in other region)
- Use more machine learning method, and soil potential parameter to better cluster and understand problematic of the field
- Adjust index thanks to the improve analytics
- Finally adjust technical route, especially nitrogen amendment to overome problematic of the increase of nitrogen prices.

Bibilography

1. Assessing Performance of Vegetation Indices to Estimate Nitrogen Nutrition Index in Pepper
2. Performance of Vegetation Indices to Estimate Green Biomass Accumulation in Common Bean
3. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation
4. Comparison of winter wheat NDVI data derived from Landsat 8 and active optical sensor at field scale
5. Evaluation of the Plant Phenology Index (PPI), NDVI and EVI for Start-of-Season Trend Analysis of the Northern Hemisphere Boreal Zone