# ANALYZING THE STATUS OF MANGO TREES IN BRGY. CANTIPAY, CARMEN, CEBU USING NDVI AND TIME SERIES CLUSTERING

R. B. Navaja [1], F. P. Campomanes [1], C.L. Patiño [1], M. J. L. Flores [2]*

[1] Central Visayas Center for Environmental Informatics, University of the Philippines Cebu, Gorordo Ave., Brgy. Lahug, Cebu City, Philippines - (rbnavaja, fpcampomanes, clpatino)@up.edu.ph

[2] Department of Biology and Environmental Science, College of Science, University of the Philippines Cebu, Gorordo Ave., Brgy. Lahug, Cebu City, Philippines - mlflores2@up.edu.ph

**Commission IV**

**KEY WORDS:** Mango, Trees, NDVI, Multi-temporal, Time Series Clustering, Cebu

**ABSTRACT:**

The Department of Agriculture – Region VII reports that many mango orchards in Cebu province are dying because of the absence of required post-harvest attention. Lacklustre yields and erratic pest infestations have driven some farmers and growers to abandon mango orchards. To help revive low-yielding mango orchards, there is a need to distinguish actively bearing mango trees from those that remain dormant throughout the year. Using remote sensing techniques, mango trees from separate orchards in Brgy. Cantipay, Carmen, Cebu were mapped and studied using multi-temporal Sentinel-2 data (from January 2018 through May 2019). Prior to that, a field visit was conducted to survey the area using UAVs and field observation, and in the process, was able to identify an abandoned mango orchard. Pixel-based Normal Difference Vegetation Index (NDVI) values were extracted from each of the 822 geotagged mango trees with an average of 16 trees among 53 divisions. Time series were derived from the average of the NDVI values from each division and plotted per month of extraction from oldest to latest. Clustering was applied to the time series data using Hierarchical Clustering with Ward's Minimum Variance as an algorithm to determine the divisions with the closest time series. Using the resulting dendrogram as basis, two major clusters were selected based on the value of their distances with each other: Cluster 1 containing 29 Divisions, and Cluster 2 containing 24 Divisions. Cluster 1 contains most of the Divisions in and around the biggest active mango orchard. In contrast, Cluster 2 contains most of the Divisions that are in and around the previously identified abandoned mango orchard. An alternative dendrogram was also created by using Complete Linkage algorithm in Hierarchical Clustering, after which 3 relevant clusters were selected. The second dendrogram highlights the stark difference between Division 1, contained in Cluster 3, from the rest of the other clustered divisions at 2.17 units from the next closest one. Notably, Division 1 is located smack in the middle of the abandoned orchard The remaining clusters, Cluster 2 with 21 divisions containing most of the divisions in the abandoned orchard, is 2.46 distance units away from Cluster 1, which has 31 and hosting most of the divisions in the active mango orchards. Two major clusters emerged from using the two algorithms. Divisions with higher and more variant NDVI values seemed to come from the mango trees which were more active during the fruiting cycle. Divisions from the abandoned mango orchards were observed to have lower and less varied NDVI values because of minimal activity in the trees. Other Divisions clustered under the abandoned orchard could have been juveniles based on their size.

## 1. INTRODUCTION

Cebu is known as one of the top mango producing provinces in the Philippines. In 2018, Cebu produced over 43,480.80 metric tons of mangoes or 6.1% of the overall production share in the country. There was a notable increase in the production of mangoes in 2018 compared with the previous two years; still a far cry from the numbers tallied in 2015, where production in the province peaked in recent years at 53,269.89 metric tons (Philippine Statistics Authority, 2018). The Department of Agriculture – Region VII cites the deteriorating health of the mango trees in the province as one of the reasons why the mango industry in the province has not yet reach its full potential. A number of mango orchards have been reported to be dormant or unable to bear fruits because either: they were affected by previous infestation; their health are deteriorating from over inducing and pesticide abuse; or the trees weren't sprayed with flower inducers at all because of farmers demotivated by small to no profits (Fernandez-Stark et al., 2017).

Farmers can avail three types of contracts in the country: leasehold, output-sharing, and broker. Amongst the three, leasehold has been regarded as the most disruptive and most unsustainable (Briones, 2013). When the leaseholders take over the farms, they exhaust the trees by over-spraying inducers to produce as much fruit as possible and using unapproved chemicals to get as much profit as possible within the limited amount of time indicated in the contract. By the end of their contracts, these leaseholders leave the farms in a damaged state. This short-sighted, unsustainable farming practice can lead to the deterioration of the trees' health, which could result to long time dormancy or the imminent death of the trees (Fernandez-Stark et al., 2017).

This study intends to distinguish actively bearing mango trees from mangos that remained dormant throughout a certain period of time using remote sensing techniques by employing multi-temporal multi-spectral Sentinel-2 data and clustering on the extracted NDVI-based phenological time series. Remote sensing provides a way to monitor vegetation using indices derived from multi-spectral maps (Johansen, Raharjo, 2017; Sun et. al., 2018; ; Peña, 2017;). Vegetation index time series are employed to study and track the phenology of various types of vegetation over a period (Sakamoto et. al., 2005; Aktaş, Burak, 2017; Seo et. al., 2019). NDVI are ubiquitous in remote sensing studies for their efficacy in detecting vegetation phenology (Chavez et. al., 2018; Pan et. al., 2014). Using time series clustering, similarities between trends are measured in order to form clusters of the same pattern (Jamali et. al, 2015; Roelofsen, 2018; Kunze, 2018; Gonçalves et. al., 2018).

## 2. REVIEW OF RELATED LITERATURE

### 2.1 Normalized Difference Vegetation Index

Normalized Difference Vegetation Index or NDVI is a known indicator of vegetation where greenness is quantified by subtracting the Red band from the Near-Infrared band of multi-spectral maps. NDVI is used in various purposes such as vegetation changes of an area, following a plant's phenological cycle, estimating yield, and many more (Schinasi et al., 2018). NDVI was used in the study for its ubiquity in various researches in the field of remote sensing and agriculture.

Normalized Difference Vegetation Index (NDVI) is extracted from the bands found in multi-spectral images. Managed by the European Space Agency, Sentinel-2 mission produces free and open multi-spectral maps with spatial resolution of 10m, 20m, and 60m and containing thirteen bands from three parts of the spectrum: visible, near infrared, and short wave. Applications of Sentinel-2 data include monitoring land cover changes, crop monitoring and yield estimation, and vegetation monitoring (European Space Agency, n.d.).

The NDVI values can help determine the green density of an area observing the unique hues of the reflected visible and near-infrared light from plants. The higher the NDVI, the healthier and denser the vegetation while the lower the NDVI, the less amount of vegetation present (Drisya et al., 2018). Bands 8 (Near Infrared) and 4 (Red) from the Sentinel-2 data are the bands used in order to calculate the NDVI values per pixel and create an equivalent NDVI map. To derive the NDVI values, difference between the Near Infrared band and the Red band is calculated. To extract NDVI values from multi-spectral data, the difference of Near Infrared (NIR) and Red (RED) bands is divided by the sum of the two as seen in Equation 1. Values of NDVI range from -1 to +1 (Drisya et al., 2018).

$$NDVI = \left( \frac{NIR-RED}{NIR+RED} \right) \quad (1)$$

where

NDVI = normalized difference vegetation index
NIR = near-infrared band
RED = Red band

### 2.2 Time Series Clustering using Hierarchical Clustering

Time Series is a compilation of discrete data over time that are graphed, indexed, and listed chronologically. The data are equally sequenced in a consecutive manner (Rani, Sikka, 2012). Clustering is the method of looking for concealed patterns in the data or grouping data that exhibits similarity among each other. In practice, cluster analysis is done in three successive steps: choosing a metric that quantifies similarity among data; enlisting a method responsible for implementing the clusters; and manually deciding the number of clusters (Roelofsen, 2018). Time series clustering is an example of unsupervised classification that lumps time series sets with no labels. Clustering the data is based on the level of similarity the data have by looking at the distance of their points from every given time (Rani, Sikka, 2012). One form of time series clustering is hierarchical clustering.
A type of clustering where a hierarchy of clusters is created in order to show the level of similarity of data by their distance is referred to as hierarchical clustering (Roelofsen, 2018). This type of clustering does not need to have defined cluster count in order to function but when large data is fed, clustering can be time consuming as distance matrix of all pairs are extracted. Nested clusters in a tree are visualized as dendrograms as shown in Figure 1.
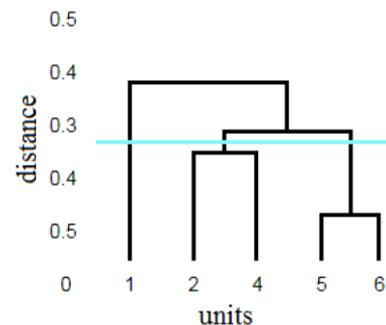


Figure 1. Sample Dendrogram cut into three different clusters

Hierarchical clustering can be classified into two according to their approach, agglomerative and divisive. Agglomerative clustering, the more commonly used among the two, starts from the bottom of all the observations where at each step, observations and clusters can be grouped together. Divisive clustering starts from the top and splits the cluster at each step until they are split individually (Roelofsen, 2018).

## 3 METHODOLOGY

### 3.1 Study Area and Scope



Figure 2. Orthophoto of Brgy. Cantipay, Carmen, Cebu (L). Closer look at the identified dormant mango orchard (R).

Field visits to the different barangays in different municipalities and cities in the province of Cebu were done. Locals were asked if there are abandoned or unproductive mango trees in recent times in their area. Prospected sites include mango orchards in: Brgy. Kalunasan, Brgy. Sapangdaku, and Brgy. Guba in Cebu City; Brgy. Buluang and Brgy. Bagalnga in Compostela, Cebu; and Brgy Cantukong and Brgy. Cantipay in Carmen. In the end, Brgy. Cantipay in Carmen, Cebu was chosen because of the number of mango orchards in the area, area of each orchard, density of mango trees per orchard, and the presence of an abandoned mango orchard. Images of the mango orchards in Brgy. Cantipay and the identified abandoned mango orchard are found in Figure 2 above. During the field visit to a number of orchards in Brgy. Cantipay field observations were conducted and images were taken using Unmanned Aerial Vehicles (UAVs) to map the mango trees. An orchard that has been unproductive for around 3 years was identified. However, most of the surrounding orchards were observed to be well pruned and well maintained with some trees in the flowering and fruiting stages based on visual inspection of the UAV images and from actual field observations.

## 3.2 Google Earth Engine
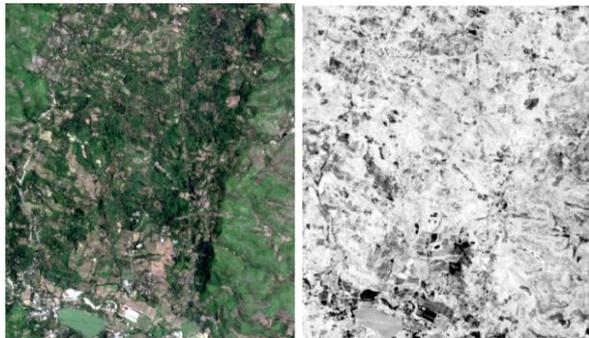


Figure 3. Sample RGB map (L), Sample NDVI map (R)

Using Google Earth Engine (GEE), an online map repository and geospatial analysis provider, images of the area (dating between January 2018 to August 2019) from Sentinel-2 missions were loaded and RGB maps were created and downloaded. The quality of each RGB maps (sample shown in Figure 3) were manually examined and fourteen dates were chosen for their clarity and minimal presence of noisy data caused by clouds and shadows. NDVI maps (sample shown in Figure 3) were calculated using the formula in Equation 0.0 and downloaded using GEE.

## 3.3 Geographical Information System



Figure 4. Clouds and shadows present in maps from each date were masked using polygons and aggregated into one shape file

Before geotagging the mango trees, erratic data such as clouds and shadows were masked using polygons as such in Figure 4. This is to ensure that no point will contain NDVI values of clouds or shadows in any given period.



Figure 5. Geotagged mango trees (L), Divisions were made to resample data (R)

Using orthophotos from the Phil-LiDAR Project and as shown in the left image in Figure 5, mango trees were geotagged manually using points so extracted values would be more accurate per tree, and mango trees that are too close to other types of vegetation will be excluded. The trees were also divided so that there will be less data to be processed as in the right image in Figure 5. In dividing the mango trees using polygons, it was ensured that the trees were close to each other, belong to the same orchard, and no polygons should overlap. After geotagging and dividing the

trees as seen in Figure 0.0, the ids of the polygons were assigned to every tree that was inside them. NDVI values were then extracted using the geotagged points in all the multi-temporal maps. The attribute table of the points was exported, containing the id of each tree, their assigned divisions, and NDVI values of each point found in each column where every column represents a different date.

## 3.4 Creating the Time Series Data

Raw data consists of the id of the trees, corresponding polygon ids per tree, and NDVI values of each date in columns arranged chronologically. Using Excel, NDVI values of each tree belonging to the same division based from the polygon ids were averaged across all dates. Since Sentinel-2 data do not have consistent brightness levels, min-max normalization, as seen in Equation 2 below, was applied to all the average NDVI values per Date. The resulting data contained time series of each division per row of average NDVI values after being normalized.

$$y = (x - min)/(max - min) \tag{2}$$

where    X = dataset containing list of x values
min = minimum value of X
max = maximum value of X
x = value to be normalized

## 3.5 Time Series Clustering

Using scipy library and python, a program was created to apply hierarchical clustering to the time series data. Existing functions from the scipy library were used to calculate the linkage of the time series data. Linkage was done twice using two different methods, the Complete Linkage and Ward's Minimum Variance algorithms, and visualized by dendrograms. The number of clusters per dendrogram were decided based on the density and distance values of each cluster. Z-value results from the linkage function application were clustered again to retrieve the cluster values of each division.

## 4    RESULTS AND RECOMMENDATIONS
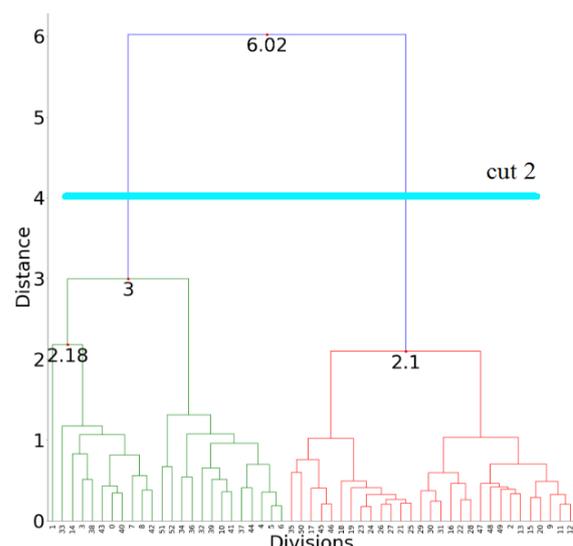


Figure 6. Hierarchical Clustering dendrogram using Ward's Minimum Variance algorithm cut into 2 clusters

Figure 6 shows the dendrogram after using Ward's Minimum Variance algorithm in the hierarchical cluster function. Two major clusters are prominent with a distance of 6.02 units. Cluster 1 (in red) with 29 overall clusters, contains majority of Divisions that represent average values of trees in the identified active mango orchard. Cluster 2 (in green) with 24 overall clusters, contains 8 out of the 9 Divisions (Divisions 0-8) in the previously identified abandoned orchard, excluding Division 2 which is found in Cluster 1. Division 1 appears to be an outlier inside Cluster 2 given its high distance value of 2.18 units.
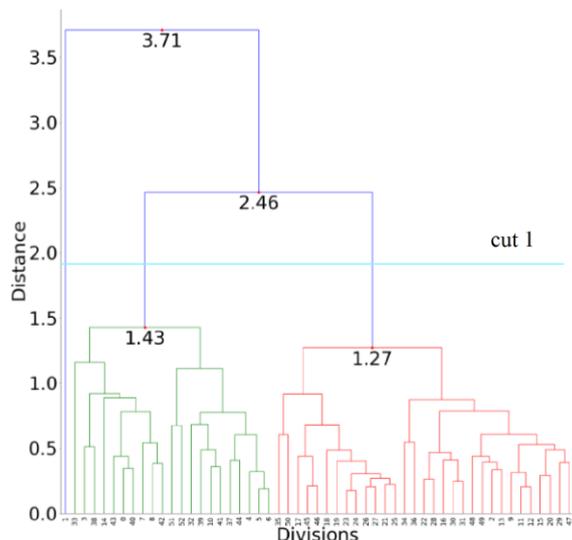


Figure 7. Hierarchical Clustering dendrogram using Complete Linkage algorithm cut into 3 clusters

Figure 7 represents the dendrogram resulting from using Complete Linkage algorithm in the hierarchical clustering function. The number of clusters which is three was decided based on the density of prominent clusters and the value of the distance units between them. The first cluster in red, Cluster 1 with 31 divisions, contains majority of Divisions in the identified active mango orchard as with the Cluster 1 in Figure 6. The next cluster in green, Cluster 2 with 21 clusters, contains 7 out of the 9 Divisions in the previously identified abandoned mango orchard. Division 1, with a relatively higher distance from the rest, has its own cluster. Similar to Figure 6, Division 2 is included in Cluster 1 rather than Cluster 2 where most of the Divisions in the abandoned orchard are clustered.

Two major types of clusters emerged from using the two algorithms, the first manifested by Cluster 1, contains Divisions in mango orchards that are still actively bearing fruits. This is because the time series of each Division has more variance and their NDVI values are higher. The variance could be attributed to the changing phenologies of each tree in each stage of their fruiting cycle and the high NDVI values could be attributed to the new leaves sprouting from the trees after they are pruned. The second cluster, represented by Cluster 2, contains most of the Divisions in the previously identified mango orchard. Other divisions in Cluster 2 are divisions that seem to have juvenile trees based on their size in the orthophotos compared to the other trees surrounding them. Another reason why other divisions are clustered together with the Divisions representing the abandoned trees and the juvenile trees, could be that they have also been dormant during the dates in the study.

## 5 CONCLUSIONS

Two similar major cluster emerged from both clustering methods were observed. One cluster containing majority of the Divisions in the active orchards and the other cluster, containing most trees from the abandoned orchard, Divisions containing juvenile trees, and others. The first cluster has more varied time series attributed to changes in phenology at each stage in their fruiting cycle and higher NDVI values which could have been caused by new leaves sprouting from pruning. The second cluster contains several divisions representing trees that were dormant or still juvenile, so activity is minimal. Based on this, the time series of NDVI values of mango trees provides a way to distinguish trees with active/dormant phenological changes during a period. This study, if further improved, can provide a way to monitor the state of mango trees and orchards in general in aiding the efforts of the Department of Agriculture in rehabilitating the mango trees in the province and to make the mango industry healthy and sustainable.

## 6 RECOMMENDATIONS

More ground data should be collected in the future. These data include the dates of each stage in the fruiting of the mango trees and harvesting, average yield per tree, weather of each date, and dates of pruning. Orchards in the province are commonly small scale compared to other provinces, so coordination with different owners and farmers could be hard. Another area to look into is the elevation of each orchard to see if it affects how data behaves especially Cebu has a mountainous landscape where more mango trees are located at the side of hills or mountains rather than in a flatter terrain. It could also be better if more abandoned orchards can be included in the data so there would be more samples to study their behaviour. Future work should also explore using other Vegetation and Blooming Indices to compare how they fare in capturing the phenological stages of the mango trees. Other clustering methods and distance measures like Distance Time Warping should be explored too in the time series clustering aspect.

### REFERENCES

Aktaş, Ayda & Üstündağ, Burak, 2017: Phenology based NDVI time-series compensation for yield estimation analysis. 1-5. doi.org/10.1109/Agro-Geoinformatics.2017.8047038.

Briones, R. M., 2013. Market structure and distribution of benefits from agricultural exports: the case of the Philippine mango industry. PIDS Discussion Paper Series, 16, 3-21. Retrieved from https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidsdps1316.pdf.

Chávez, R.O., Moreira-Muñoz, A., Galleguillos, M., Olea, M., Aguayo. J., Latín, A., Aguilera-Betti, I., Muñoz A.A., Manríquez, H., GIMMS NDVI time series reveal the extent,

duration, and intensity of "blooming desert" events in the hyper-arid Atacama Desert, Northern Chile. International Journal of Applied Earth Observation and Geoinformation, 76, 193-203. doi.org/10.1016/j.jag.2018.11.013.

Drisya, J., Kumar, S. D., Roshni, T., 2018: Spatiotemporal Variability of Soil Moisture and Drought Estimation Using a Distributed Hydrological Model. Integrating Disaster Science and Management, 451-460. doi.org/10.1016/B978-0-12-812056-9.00027-0.

Europe Space Agency, n.d. Multi Spectral Instrument (MSI) Overview. In Europe Space Agency Earth Online. Retrieved August 30, 2019, from https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument

Fernandez-Stark, K., Couto, V., Gereffi G., 2017: The Philippines in the mango global value chain, 40. http://industry.gov.ph/wp-content/uploads/2017/08/The-Philippines-in-the-Mango-Global-Value-Chain.pdf

Gonçalves, R., Zullo Junior, J,. Ferraz do Amaral, B., Parros Machado Sousa, E., Luciana, R., 2018: Agricultural monitoring in regional scale using clustering on satellite image time series. doi.org/10.5772/intechopen.71148.

Jamali, S., Jönsson, P., Eklundh, L., Ardö, J., Seaquist, J., 2015: Detecting changes in vegetation trends using time series segmentation. Remote Sensing of Environment, 156, 182-195. doi.org/10.1016/j.rse.2014.09.010.

Johansen, K., Raharjo, T., 2017. Multi-temporal assessment of lychee tree crop structure using multi-spectral RPAS imagery. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W6, pp.165-170. doi.org/10.5194/isprs-archives-XLII-2-W6-165-2017.

Kunze, L., Amaral, T., Moraes, L., Oliveira, J., Junior, A., Sousa, E. and Cordeiro, R., 2018: Classification analysis of NDVI time series in metric spaces for sugarcane identification. ICEIS, 1, 162-169 ISBN: 978-989-758-298-1.

Pan. Z., Huang, J., Zhou, Q., Wang, L., Cheng, Y., Zhang, H., Blackburn G.A., Yan, J., Liu, J., 2015: Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data. International Journal of Applied Earth Observation and Geoinformation, 34, 188-197. doi.org/10.1016/j.jag.2014.08.011.

Peña, M., Brenning A., Renfang, L., 2017: Classifying fruit-tree crops by Landsat-8 time series. 1-4. doi.org/10.1109/GRSS-CHILE.2017.7995998.

Philippines Statistics Authority, 2018: Open Statistics Application. In Philippine Statistics Office Online. openstat.psa.gov.ph
Rani, S., Sikka, G., 2012. Recent Techniques of Clustering of Time Series Data: A Survey. International Journal of Computer Applications (0975 – 8887), 52 (15), 1-9. doi.org/10.5120/8282-1278.

Roelofsen, P., 2018. Time series clustering. Vrije Universiteit Amsterdam, Amsterdam, Netherlands. Retrieved from https://www.math.vu.nl/~sbhulai/papers/thesis-roelofsen.pdf.

Sakamoto, T., Yokozawa, M, Toritani, H, Shibayama, M., Ishitsuka, N., Ohno, H., 2005: A crop phenology detection method using time-series MODIS data. Remote Sensing of Environment, 96 (3–4), 366-374. doi.org/10.1016/j.rse.2005.03.008.

Schinasi, L. H., Benmarhnia, T., De Roos, A. J., 2018: Modification of the association between high ambient temperature and health by urban microclimate indicators: A systematic review and meta-analysis. Environmental Research, 161, 168-180.

Seo, B., Lee, J., Lee, K.D., Hong, S., Kang, S, 2019: Improving remotely-sensed crop monitoring by NDVI-based crop phenology estimators for corn and soybeans in Iowa and Illinois, USA. Field Crops Research, 238, 113-128. doi.org/10.1016/j.fcr.2019.03.015.

Sun, C., Fagherazzi, S., Liu, Y., 2018: Classification mapping of salt marsh vegetation by flexible monthly NDVI time-series using Landsat imagery. Estuarine, Coastal and Shelf Science, 213, 61-80. doi.org/10.1016/j.ecss.2018.08.007.