



Random Forest Classification for Plantation of Rubber Area Using Digital Number versus Reflectance Values

Tengku Noradilah Tengku Jalal, N. Rajkumar A/L V. Nagarethinam,
Nur Hurriyatul Huda Abdullah Sani, Nurul Aishah Rahman

Core Team Big Data Analytics,
Department of Statistics, Malaysia



9TH MALAYSIA STATISTICS CONFERENCE

Department of Statistics Malaysia

4TH OCT. 2022
(VIRTUAL)
&
5TH OCT. 2022
(ILSM, SUNGKAI, PERAK)



Dealing with Uncertainties: Unearthing Measures for Recovery

Organised by:



PRIME MINISTER'S DEPARTMENT
DEPARTMENT OF STATISTICS MALAYSIA

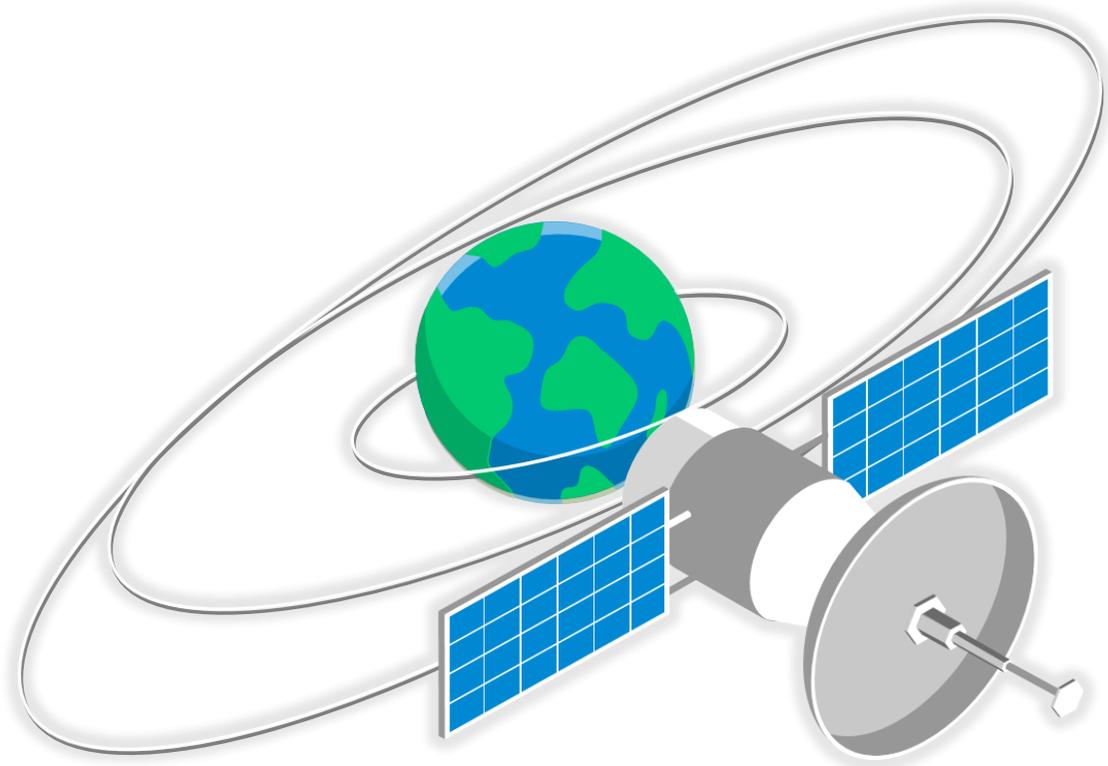


BANK NEGARA MALAYSIA
CENTRAL BANK OF MALAYSIA



MALAYSIA INSTITUTE
OF STATISTICS

OUTLINE



INTRODUCTION

METHODOLOGY

RESULT

DISCUSSION AND CONCLUSION



9TH MALAYSIA STATISTICS CONFERENCE



Organised by:



PRIME MINISTER'S DEPARTMENT
DEPARTMENT OF STATISTICS MALAYSIA



BANK NEGARA MALAYSIA
MALAYSIA INSTITUTE OF STATISTICS



MALAYSIA INSTITUTE OF STATISTICS



INTRODUCTION



9TH MALAYSIA STATISTICS CONFERENCE



Organised by:



RUBBER



Rubber statistics were first published in 1965

Total area planted with rubber in 1965 was 788.5 thousand hectares, with a tapped area of 542.3 thousand hectares producing 507.9 thousand tonnes of natural rubber.



In 2019, the planted area was only 95.4 thousand hectares with a tapped area of 43.6 thousand hectares, which produced 61.2 thousand tonnes of natural rubber production.



9TH MALAYSIA STATISTICS CONFERENCE



Organised by:



PRIME MINISTER'S DEPARTMENT
DEPARTMENT OF STATISTICS MALAYSIA



BANK NEGARA MALAYSIA
MALAYSIA INSTITUTE OF STATISTICS



MALAYSIA INSTITUTE OF STATISTICS



RUBBER MONITORING



Traditionally, rubber tree monitoring is **time-consuming and labour-intensive.**

The collection of ground data relies heavily on **conventional monitoring** methods.



Restricted control movement during pandemic Covid-19 has made it even more challenging for field data collection.



9TH MALAYSIA STATISTICS CONFERENCE



Organised by:



REMOTE SENSING AND SATELLITE-BASED TECHNOLOGIES



Remote sensing is about acquiring images and information about the Earth's land and water surfaces from an overhead perspective by employing electromagnetic radiation where a distinction is made between reflected or emitted electromagnetic radiation recorded by the sensors

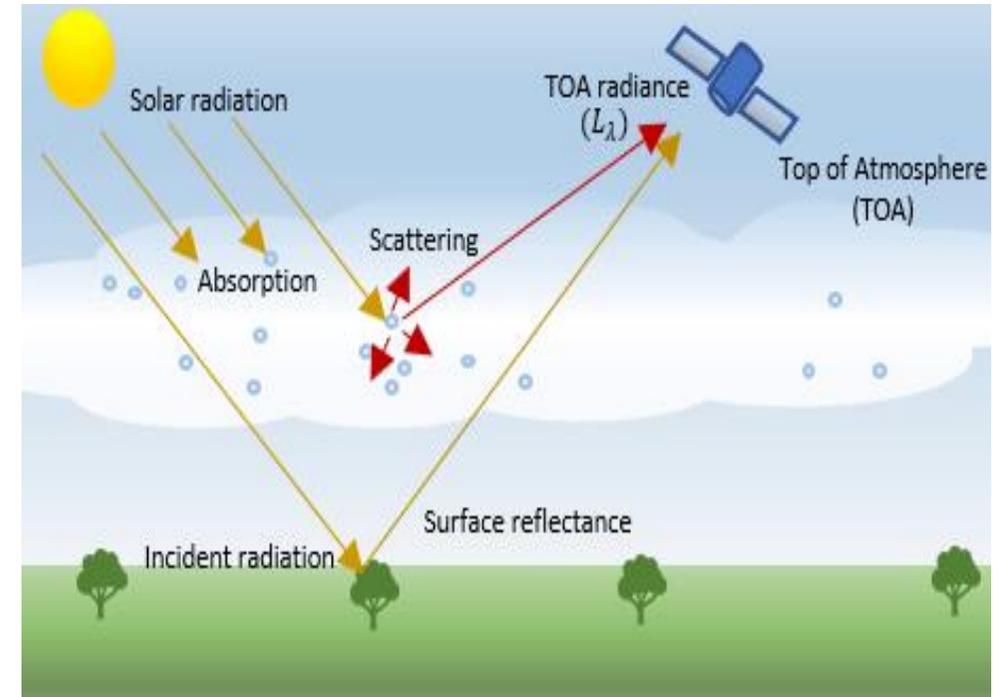
(Campbell and Wynne, 2011)



REMOTE SENSING AND SATELLITE-BASED TECHNOLOGIES



- ❖ Sensor turns the energy into a voltage, which an analog to digital converter turns into a single integer value called the **Digital Number**.
- ❖ **Radiance** is the amount of radiation coming from an area. Radiance includes radiation reflected from the surface, bounced in from neighbouring pixels, and reflected from clouds above the area of the pixel.
- ❖ **Reflectance** is the proportion of the radiation striking a surface to the radiation reflected off of it. Some materials can be identified by their reflectance spectra, so it is common to correct an image to Reflectance Value as a first step toward locating or identifying features in an image (Shippert, 2017).

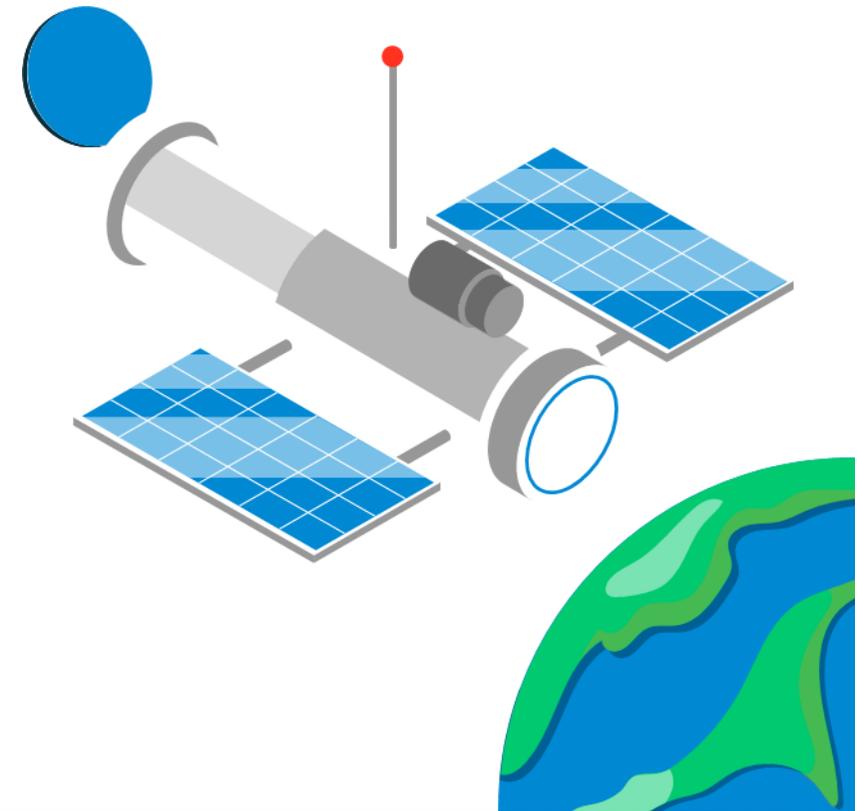


MACHINE LEARNING AND SATELLITE IMAGE



Satellite Imagery (SI) and Machine Learning (ML) offers great opportunities and potential in advancing and modernising rubber data collection methods.

This study aims to **determine plantation of rubber area using two different inputs data namely digital number (DN) and reflectance values (RV).**



9TH MALAYSIA STATISTICS CONFERENCE



Organised by:





METHODOLOGY



9TH MALAYSIA STATISTICS CONFERENCE



Organised by:



PRIME MINISTER'S DEPARTMENT
DEPARTMENT OF STATISTICS MALAYSIA



BANK NEGARA MALAYSIA
MALAYSIA INSTITUTE OF STATISTICS



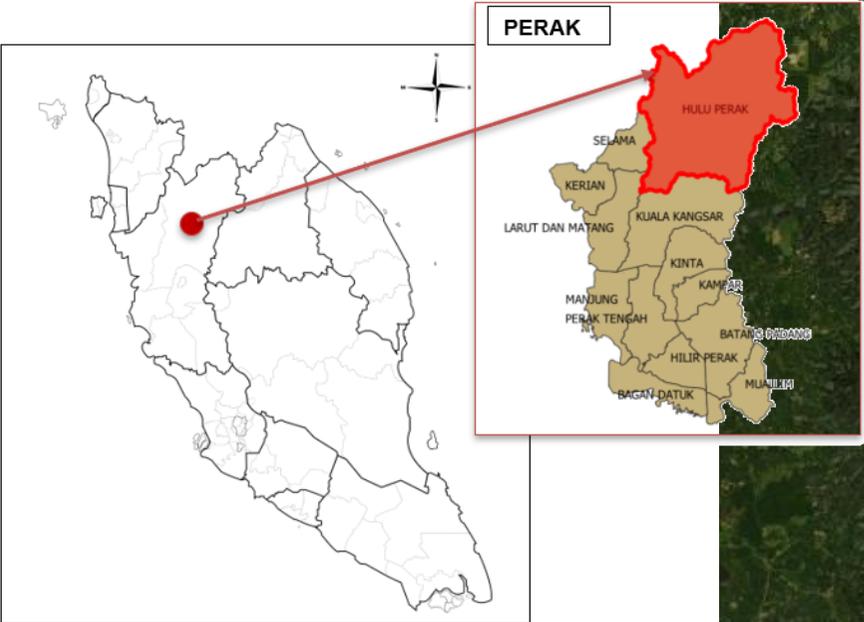
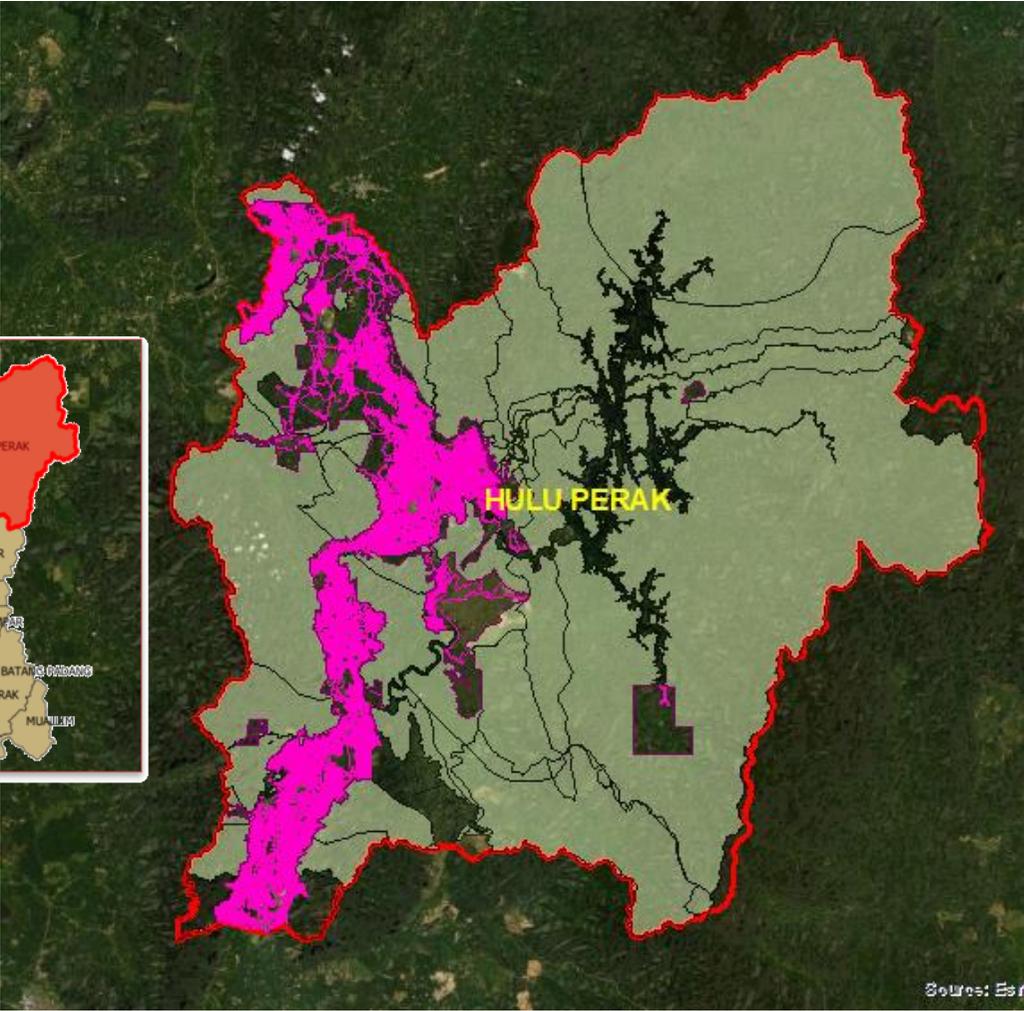
MALAYSIA INSTITUTE
OF STATISTICS



AREA OF INTEREST



Hulu Perak, the biggest district (6,613 km²) in Perak, Peninsular Malaysia.



Geographically diverse, with the economic activity based on agriculture such as the cultivation of rubber, palm oil and other agricultural activities, as well as eco-tourism.



9TH MALAYSIA STATISTICS CONFERENCE

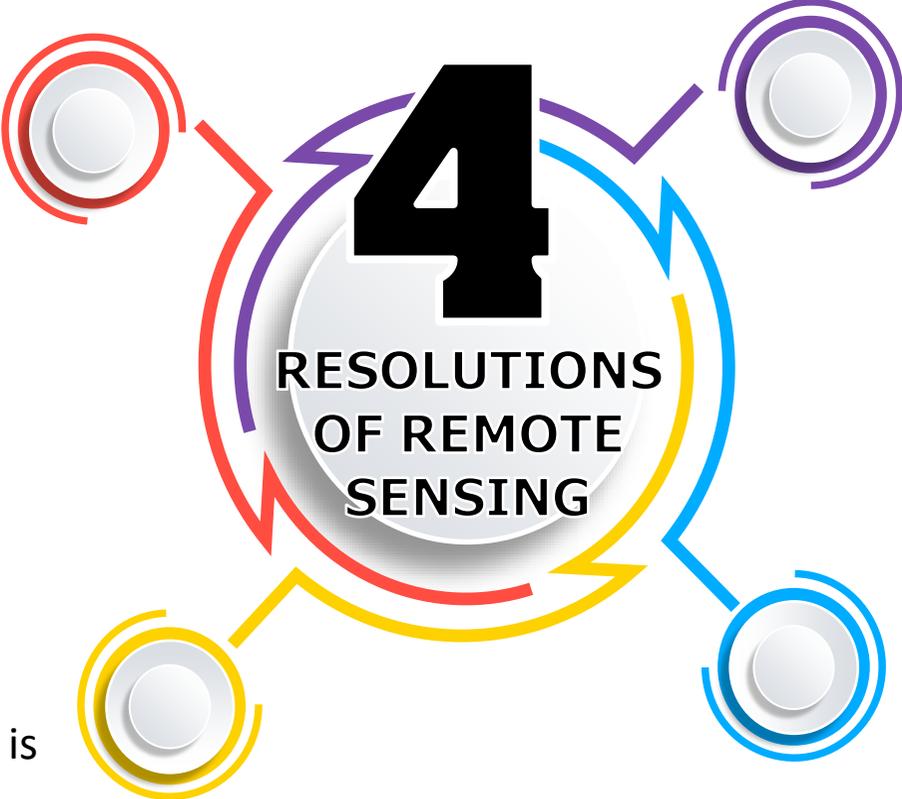


RESOLUTIONS OF REMOTE SENSING



SPATIAL RESOLUTION
measures the smallest angular separation between two objects

TEMPORAL RESOLUTION
the frequency at which imagery is recorded for a particular area



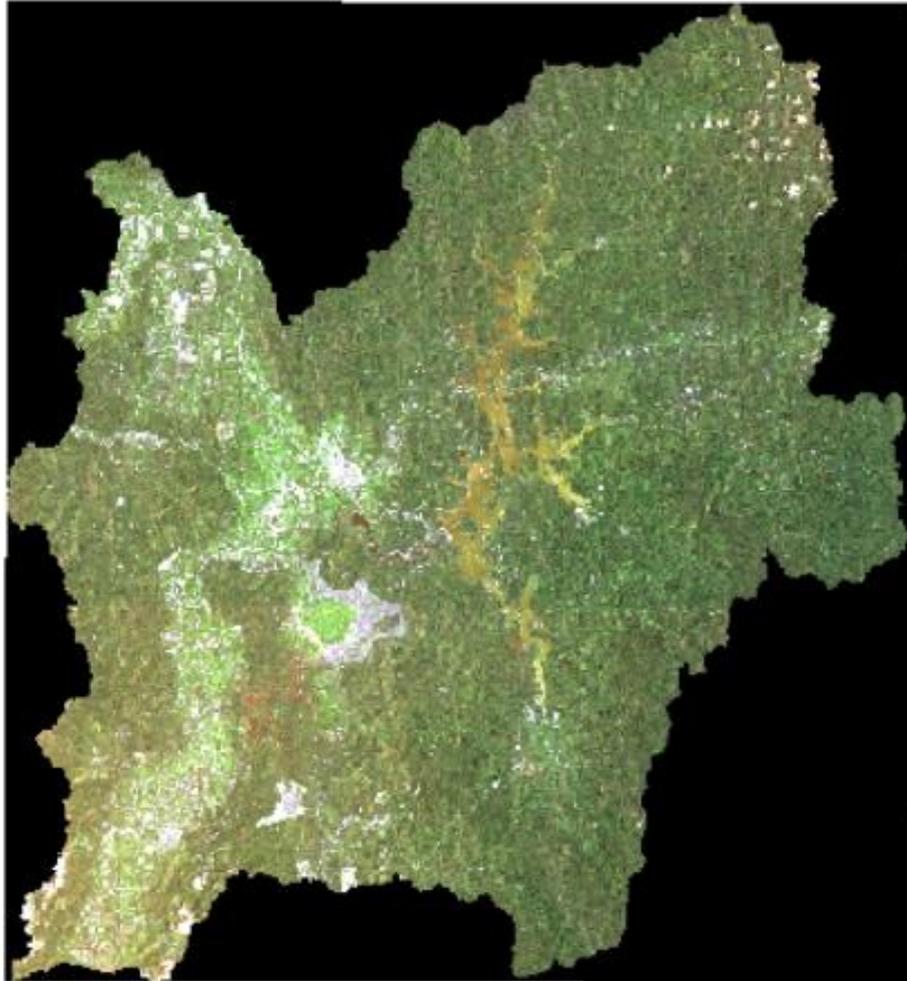
SPECTRAL RESOLUTION
Specific wavelength intervals that a sensor can record

RADIOMETRIC RESOLUTION
the sensitivity of a remote sensing platform to detect slight differences in energy



9TH MALAYSIA STATISTICS CONFERENCE





For the purpose of this study, spectral resolutions from **Sentinel-2A satellite** image data of **15th March 2019** were acquired and used for **image classification of rubber and other classes**.

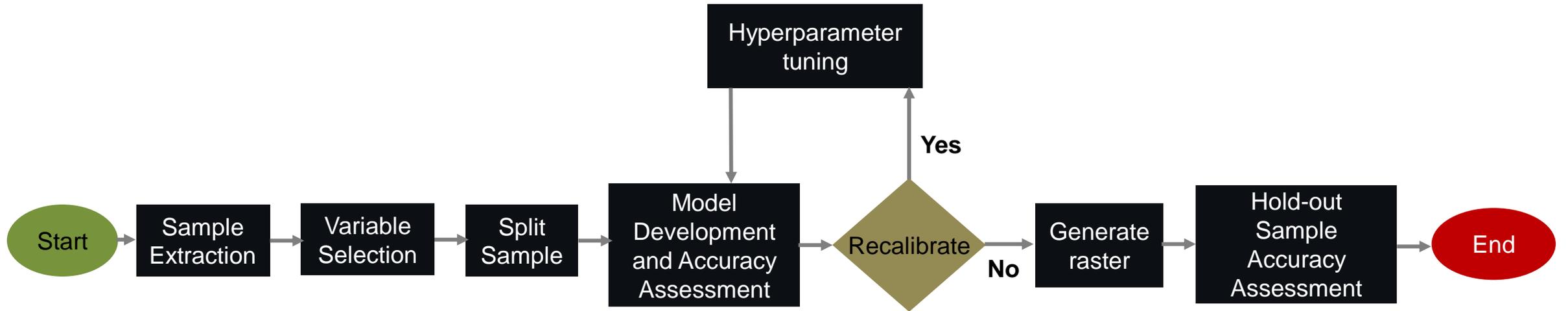
These images were downloaded using **Semi-Automatic Classification Plugin (SCP)**, a plugin available in **Quantum Geography Information System (QGIS)**.



SAMPLE SELECTION & VALIDATION:

Geospatial data (shape file) was obtained from the Population and Demographic Statistics Division (DOSM) and Department of Town and Country Planning Malaysia (PLANMalaysia)

FLOWCHART



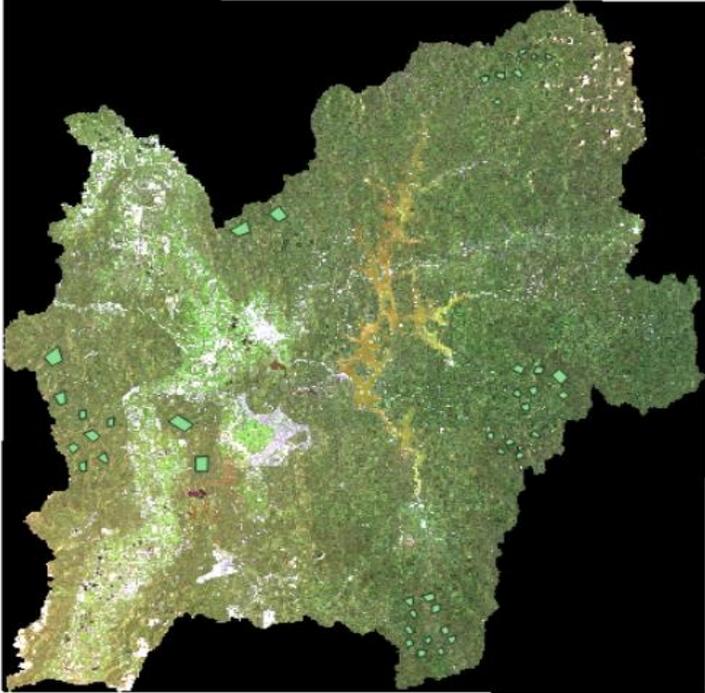
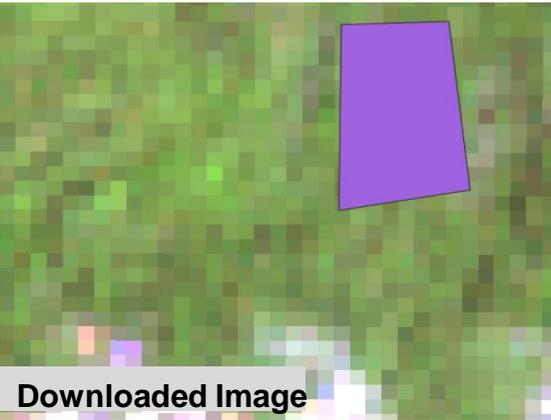
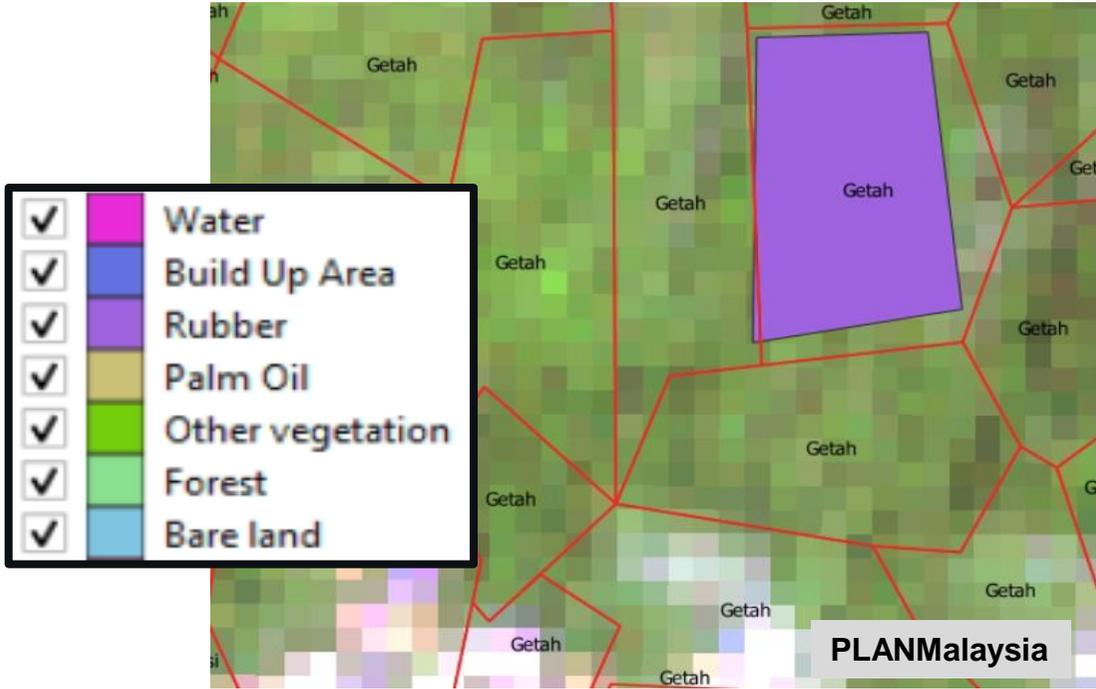
9TH MALAYSIA STATISTICS CONFERENCE



Organised by:



SAMPLE EXTRACTION



TRAIN
(17,285)

VALIDATION
(7,409)

TEST
(6,175)



9TH MALAYSIA STATISTICS CONFERENCE



VARIABLES



Sentinel-2 Bands	Resolution (m)
Band 2 - Blue	10
Band 3 - Green	10
Band 4 - Red	10
Band 5 - Vegetation Red Edge (VRE1)	20
Band 6 - Vegetation Red Edge (VRE2)	20
Band 7 - Vegetation Red Edge (VRE3)	20
Band 8 - Near Infrared (NIR)	10
Band 8A - Vegetation Red Edge (VRE4)	20
Band 11 - Short-wave infrared (SWIR1)	20
Band 11 - Short-wave infrared (SWIR2)	20
Normalised Difference Vegetation Index (NDVI)	-
Normalise Difference Building Index (NBDI)	-
Enhanced Vegetation Index (EVI)	-

Initially, there are 4 bands with 10 metre resolution, 6 bands with 20 metre resolution plus with additional 3 derived indices NDVI, NBDI, and EVI.

Bands with 20 metre resolutions were resampled to 10 metre resolution to standardise the resolution.

Pearson product-moment correlation coefficient and Tukey's Honest Significant Difference test conducted to check for association of the variables.

Variables with weak correlation and equal mean difference will be removed from the study.



9TH MALAYSIA STATISTICS CONFERENCE



MODEL DEVELOPMENT



Model used: Random Forest Classifier

- ❖ simple to implement, fast in operation and have proven to be extremely successful in various domains.
- ❖ robust to outliers and indifferent to non linear features.

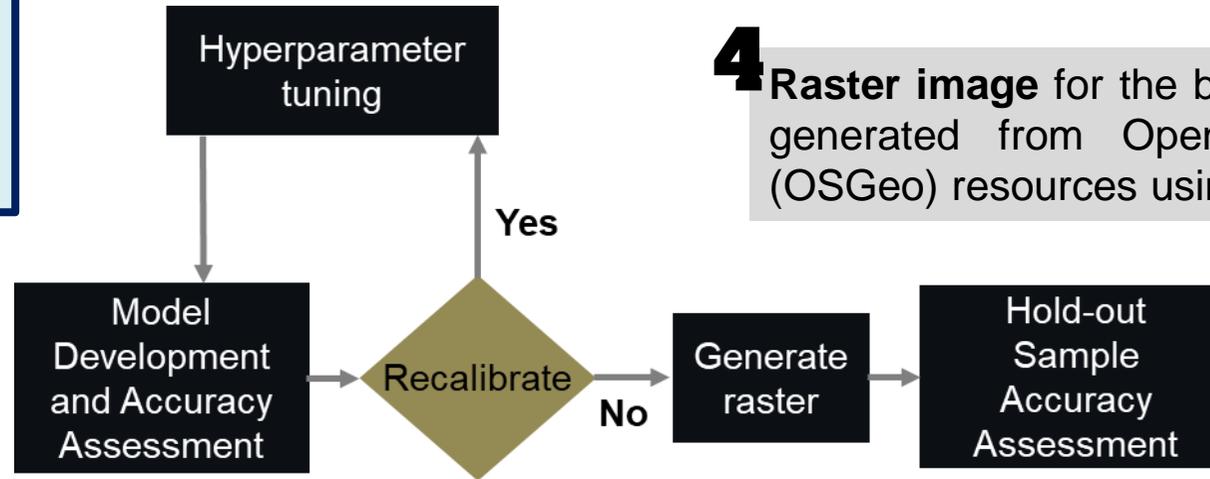
3 Hyperparameter tuning is an essential part of controlling the behaviour of a machine learning model.

4 Raster image for the best RF Model is being generated from Open Source Geospatial (OSGeo) resources using Python Scripts.

1 Default parameters were used to build the **initial model** and later assessed using Confusion Matrix.

2 Model is further **recalibrated** and tweaked using hyperparameter tuning to improve the model accuracy

5 Hold-out or **validation accuracy** is performed to ensure that the best model performs optimally and unbiased in the wild.





RESULT



9TH MALAYSIA STATISTICS CONFERENCE



Organised by:



PRIME MINISTER'S DEPARTMENT
DEPARTMENT OF STATISTICS MALAYSIA



BANK NEGARA MALAYSIA
BERSEKUTUAN BERKUALITI

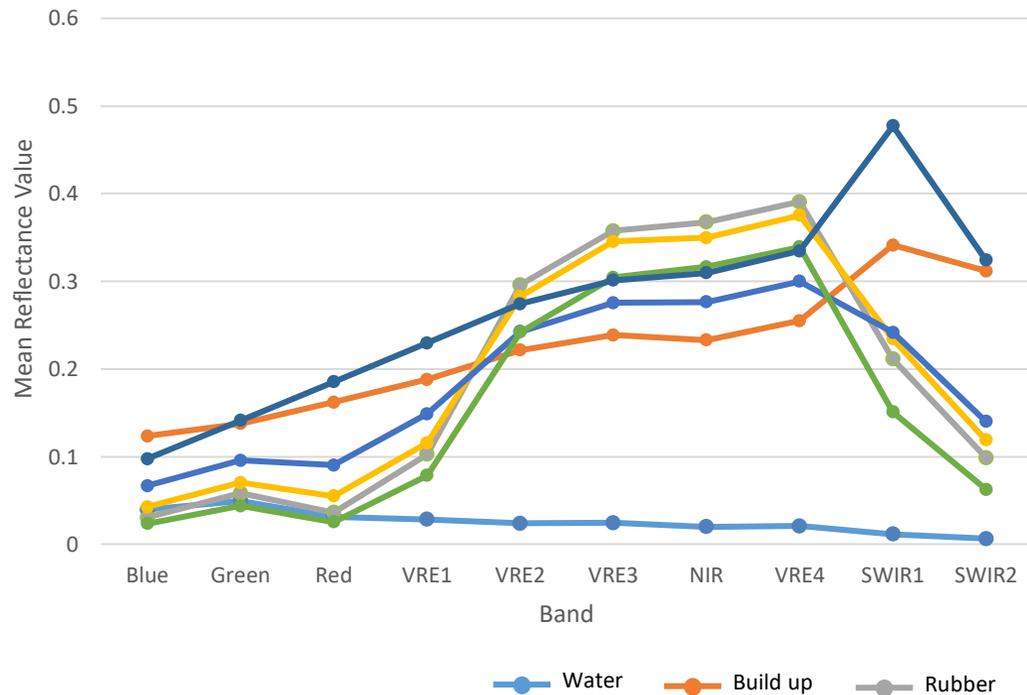


MALAYSIA INSTITUTE
OF STATISTICS

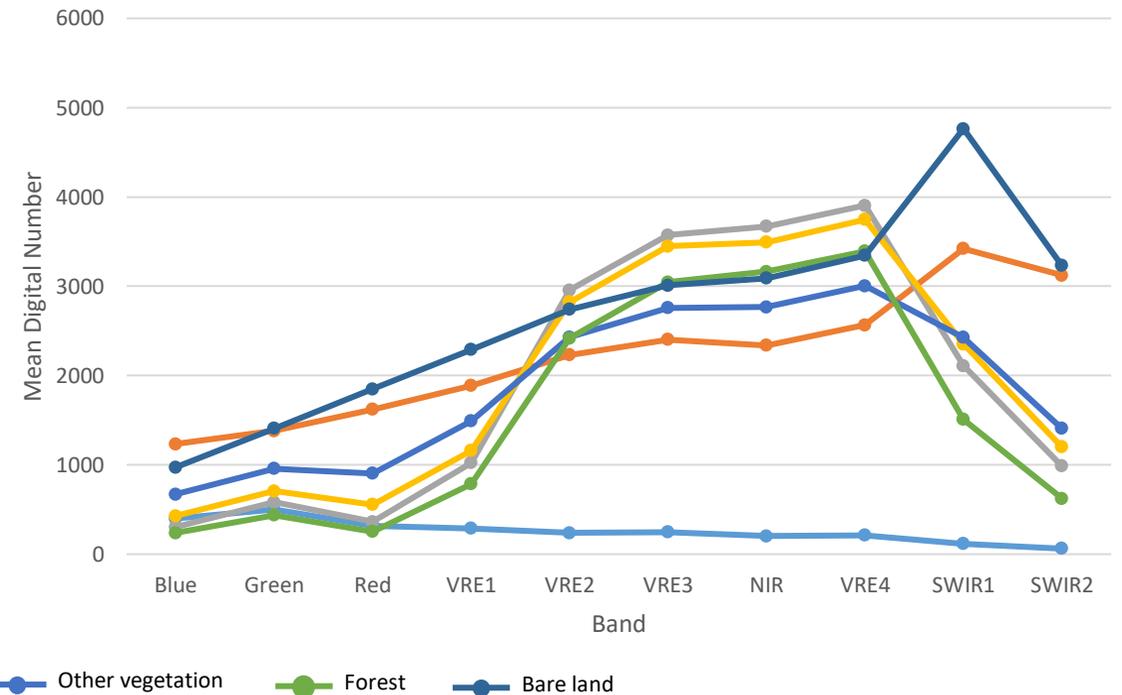
OVERALL SPECTRAL CHARACTERISTICS



Comparison of the mean of different classes for Reflectance Values



Comparison of the mean of different classes for Digital number values



Rubber and Palm oil have the closest mean and the same spectral characteristics. Forest also shows the same spectral characteristics but with lower mean.



9TH MALAYSIA STATISTICS CONFERENCE



VARIABLE SELECTION



EVI indicates a very weak association with the dependent variable (-0.002 for DN and -0.004 for RV) and there was no statistically significant difference between the mean of EVI and other independent variables.

Pearson product-moment correlation coefficient for EVI

		Variables											
		Blue	Green	Red	VRE1	VRE2	VRE3	NIR	VRE4	SWIR1	SWIR2	NDVI	NDBI
Input	DN	-0.015	-0.014	-0.013	-0.013	-0.010	-0.009	-0.008	-0.009	-0.011	-0.012	0.021	-0.009
	RV	-0.011	-0.011	-0.011	-0.011	-0.009	-0.008	-0.009	-0.007	-0.009	-0.010	0.004	-0.003



Stepwise selection suggested that the number of variables be reduced to 7 variables.

Final models were run with only **band blue, VRE1, VRE2, VRE3, NIR, VRE4 and SWIR2** as the independent variables

ACCURACY ASSESMENT



Comparison of model and rubber area accuracy assessment for digital number and reflectance values

			Digital Number Values	Reflectance Values
INITIAL MODEL	Model	Overall Accuracy	0.962	0.961
	Rubber	F1-Score	0.93	0.92
FINAL MODEL	Model	Overall Accuracy	0.967	0.968
		Cohen Kappa	0.956	0.954
	Rubber	F1-Score	0.94	0.94

Overall accuracy of **initial model** with **digital number input** is **0.1% more accurate** compared to reflectance values. F1-score for rubber classification also shows that **digital number score 0.01 higher** than reflectance values.

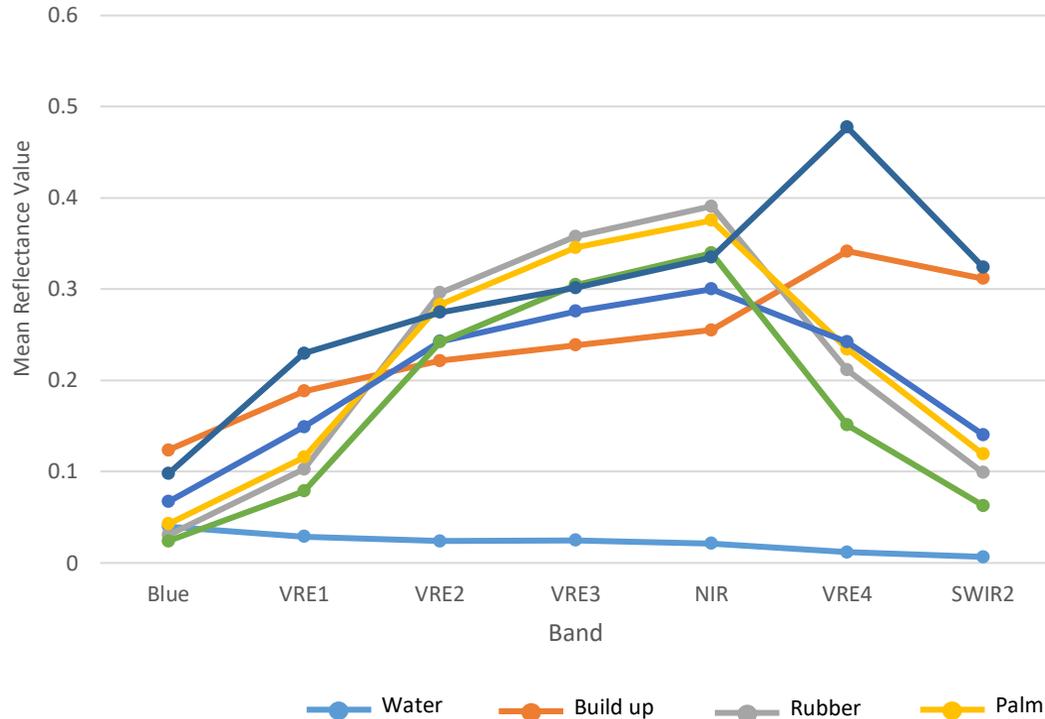
Final model shows a mere difference of overall accuracy with the result of **96.7% accuracy for digital number values** and **96.8% accuracy for reflectance values**. The result of Cohen's Kappa value which is 0.956 for digital number values and 0.954 for reflectance values shows that both inputs **strongly help** the model to successfully map all the classes. F1-score of rubber classification from the model stated that both input data gives the same accuracy score.



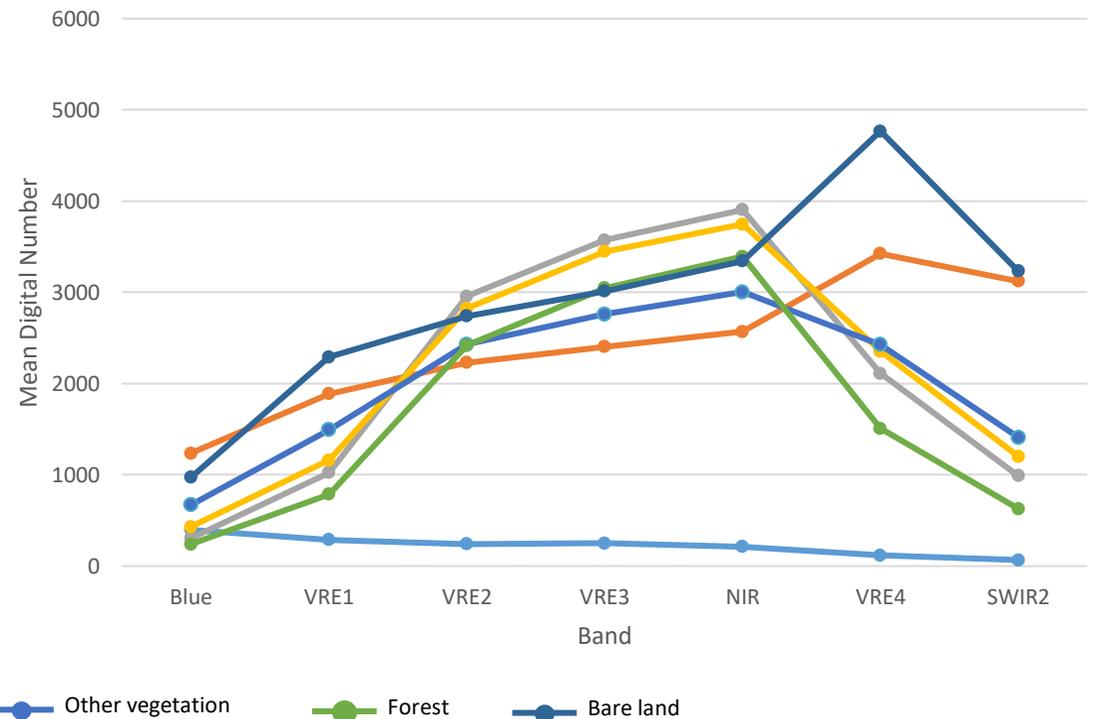
SPECTRAL CHARACTERISTICS



Comparison of the mean of different classes for Reflectance Values



Comparison of the mean of different classes for Digital number values



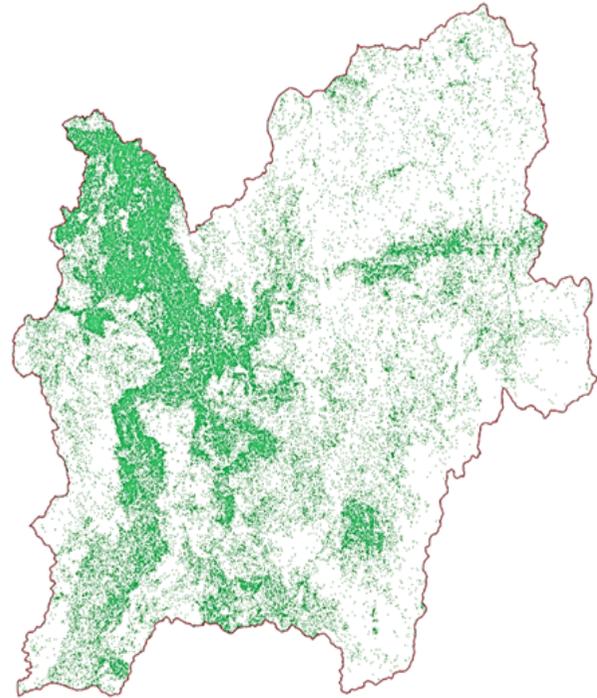
Rubber has **similar spectral characteristics** with palm oil and forest. Mean of digital number and reflectance value for rubber are slightly different to palm oil while forest has a lower mean.



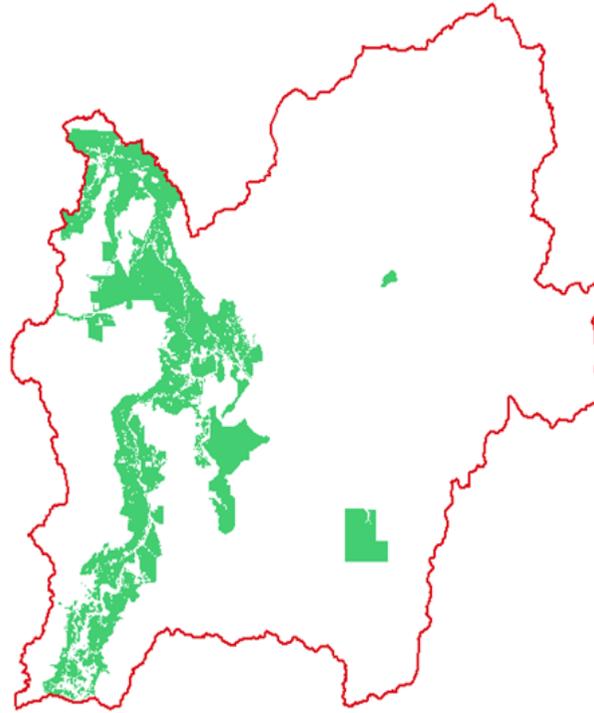
COMPARISON OF THE RUBBER DISTRIBUTION



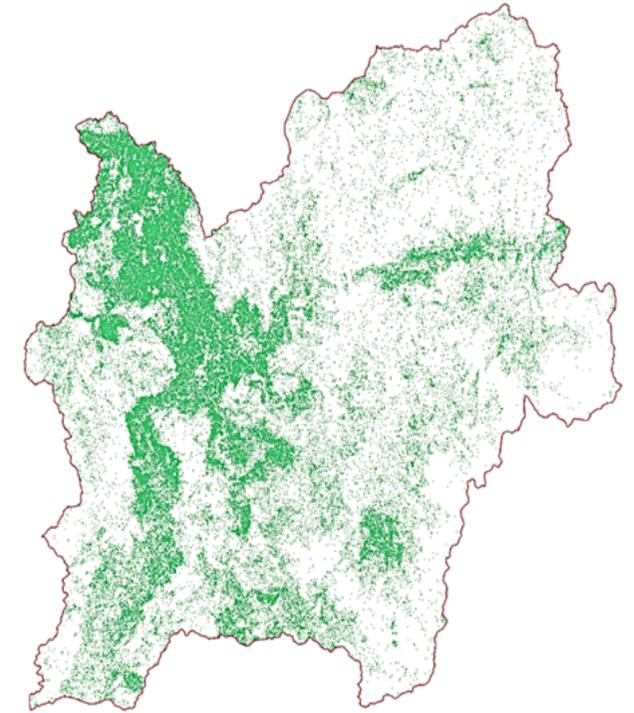
RF classification based on digital number values



Rubber distribution stated by PLANMalaysia



RF classification based on reflectance values



Result of DN and RV shows a **similar pattern** with the data from PLANMalaysia. **Mixed pixels** effect may give some effect to the classification result since the spectral characteristics of rubber were similar to forest and palm oil. In addition, the classification results may also be affected by the **rubber trees that grow in the forest**.





CONCLUSION



9TH MALAYSIA STATISTICS CONFERENCE



Organised by:



PRIME MINISTER'S DEPARTMENT
DEPARTMENT OF STATISTICS MALAYSIA



BANK NEGARA MALAYSIA
BANK OF MALAYSIA



MALAYSIA INSTITUTE
OF STATISTICS

CONCLUSION



Study aim to develop an RF based classification method with the difference input of sentinel-2 data (digital number and reflectance values).

Study demonstrates that there is no difference in accuracy of the models when the inputs were changed.

The spectral characteristics show that rubber has similar characteristics to forest and palm oil. Thus, mixed pixels effect may affect the result.



THANK YOU



StatsMalaysia



www.dosm.gov.my



9TH MALAYSIA STATISTICS CONFERENCE

Department of Statistics Malaysia

4TH OCT. 2022
(VIRTUAL)
&
5TH OCT. 2022
(ILSM, SUNGKAI, PERAK)

Dealing with Uncertainties: Unearthing Measures for Recovery

Organised by:



PRIME MINISTER'S DEPARTMENT
DEPARTMENT OF STATISTICS MALAYSIA



BANK NEGARA MALAYSIA
CENTRAL BANK OF MALAYSIA



MALAYSIA INSTITUTE
OF STATISTICS