Mekong River Delta Crop Mapping Using a Machine Learning Approach

December 2022

Surajit Ghosh[1], Michael Wellington[1,2] and Bunyod Holmatov[1]

[1]International Water Management Institute, Colombo, Sri Lanka

[2]Australian National University, Canberra, Australia

Abstract

Agricultural land use and practices have important implications for climate change mitigation and adaptation. It is, therefore, important to develop methods of monitoring and quantifying the extent of crop types and cropping practices. A machine learning approach using random forest classification was applied to Sentinel-1 and 2 satellite imagery and satellite-derived phenological statistics to map crop types in the Mekong River Delta, enabling levels of rice intensification to be identified. This initial classification differentiated between broad and prevalent crop types, including perennial tree crops, rice, other vegetation, oil palm and other crops. A two-step classification was used to classify rice seasonality, whereby the areas identified as rice in the initial classification were further classified into single, double, or triple-cropped rice in a subsequent classification with the same input data but different training polygons. Both classifications had an overall accuracy of approximately 96% when cross-validated on test data. Radar bands from Sentinel-1 and Sentinel-2 reflectance bands were important predictors of crop type, perhaps due to their capacity to differentiate between periodically flooded rice fields and perennial tree cover, which were the predominant classes in the Delta. On the other hand, the Start of Season (SoS) and End of Season (EoS) dates were the most important predictors of single, double, or triple-cropped rice, demonstrating the efficacy of the phenological predictors. The accuracy and detail are limited by the availability of reliable training data, especially for tree crops in small-scale orchards. A preliminary result is presented here, and, in the future, efficient collection of ground images may enable cost-effective training data collection for similar mapping exercises.

1. Introduction

The Mekong River Delta is a densely populated part of southwest Vietnam. It supports a large population engaged primarily in irrigated agriculture and fisheries **(Hoanh et al. 2014; World Bank 2021**). Much of the Delta is flooded at least annually and irrigated rice activities have relied on landscape alteration, such as irrigation channels, dykes, embankments, or levies, to manage water inundation of fields and to manage the mixing of fresh and brackish water in the lower Delta **(Hoanh et al. 2014**). Further, water and water infrastructure are used to support the extensive aquaculture activities in the Delta. Intensification of the rice system to allow triple-cropping is necessary to improve productivity and resource use efficiency. However, this would require investment in new irrigation infrastructure to allow for dry-season crop production (**Hoanh et al. 2014**). Rice production, especially in the wet season, is especially sensitive to temperature and rainfall variability in the Mekong Delta (**Nhan et al. 2011; Schneider & Asch 2020**). Conversely, shrimp production and aquaculture are less sensitive to weather anomalies than rice production, which characterizes those industries' ability to support farmers' resilience to climate change **(Nhan et al. 2011; Hoanh et al. 2014; World Bank 2021**). Sea level rise, saltwater intrusion, and soil salinization present other challenges related to climate and weather for farmers of the Mekong River Delta (**Nhan et al. 2012**). These are exacerbated by land use-related environmental impacts such as erosion (**Anthony et al. 2015**). The agroecosystems of the Mekong River Delta are, therefore, likely to change over the coming decades and will necessitate adaptation and, in some cases, transformative changes such as novel land use (The World Bank 2021). This means that land use and management can greatly impact the mitigation of and adaptation to climate change (**Sebesvari et al. 2011).** Emissions are likely to vary across land uses, so they must be monitored for accurate landscape-scale emissions calculations.

Temporal mapping of crop type is one of the major focus areas to understand the trend of land use and land cover changes. Google Earth Engine (GEE) is a platform enabling research to be conducted on a range of global datasets to map land cover and land use (**Gorelick et al. 2017; Wu 2020**). Integration of GEE products and machine learning algorithms with accessible and reproducible code documentation, such as in Jupyter notebooks, offers large-scale crop-type mapping opportunities. Decision tree, or classification and regression tree (CART) approaches, are useful non-parametric approaches that have the advantage of being interpretable and explainable with regard to prediction pathways (**James et al. 2014**). However, they are more prone to bias and overfitting training data. Random forest, a forest of decision trees, has become a popular algorithm for crop type mapping (**Pareeth et al. 2019**). As an ensemble method, it often produces accurate results and can be less prone to overfitting (**James et al. 2014**). It is also relatively easy to program in a workflow, so it is a useful staple algorithm for crop type mapping. The random forest algorithm was used to map the rice and other land cover type.

2. Data and Methods

2.1 Training data collection

The Mekong River Delta is comprised of 13 provinces. Training data were collected by drawing and labelling polygons based on visual inspection of public satellite imagery in QGIS software. Some crops, such as oil palm, were able to be identified via supplementary and secondary data like Street View and user images on Google Maps, which may become a more dominant means of ground data collection in the future. The proportion of data points (**Table 1**) broadly reflected the areas evident from visual sampling. While effective, this approach is not as accurate as comprehensive on-ground survey efforts and does not allow for as many crop types to be classified due to uncertainty with identification. The workflow was designed with this method as a preliminary step. This training data could be easily placed with ground-collected data in the future with minimal change to the overall workflow. The classes and visual inspection rules that were followed when drawing and labelling polygons are shown in **Table 1**, along with the number of polygons generated in that class. The distribution of training polygons broadly reflected each class's area observable from the satellite imagery inspection.

Table 1 *Classes of training data prepared for the classification.*

Class	**Identification Rules**	***Number of locations***
Perennial Tree	Individual trees are evident in cropland, not identified as oil palms (e.g., coffee).	102
Rice	Uniform light/dark green/brown = harvested / fallow field.	159
Other Veg.	Heterogeneous veg in cropland, likely to be abandoned fields	5
Oil Palm	Oil Palm trees are identifiable, confirmed with nearby Google Maps/street view images.	122
Other Crops	Evident cropland management, unable to be identified as trees or rice, likely to be maize, cane, or annual horticultural crops.	35

The workflow covered in this report comprises two separate classifications: the initial classification of crop type to classes in **Table 1** and the subsequent classification of rice seasonality within areas identified as rice from the initial classification. The rice seasonality classification followed the same procedure as the crop type classification but used different training data. It is difficult to collect accurate training data on rice seasonality without detailed time series, so results obtained by Son et al. (2014) were used to inform polygon drawing and labelling.

2.2 Satellite imagery

2.2.1 Masking

Recent land use and land cover products, such as the European Space Agency (ESA) WorldCover product available on GEE (**Zanaga et al. 2022**). **Figure 1** shows the land cover types for the region of interest as defined by ESA WorldCover. This product has the advantage of being Sentinel-2 derived, so it is available globally in 10-m resolution enabling higher resolution crop type mapping.

It is evident that some perennial tree crop areas in the Delta are classified as 'Tree cover' rather than 'Cropland', so the two classes need to be combined to produce a useful map for crop type mapping. Another challenge in the Delta is that some frequently flooded rice cultivation areas may become classified as 'Permanent water'. The 'Permanent water' class assists with removing streams and rivers from the crop type map.

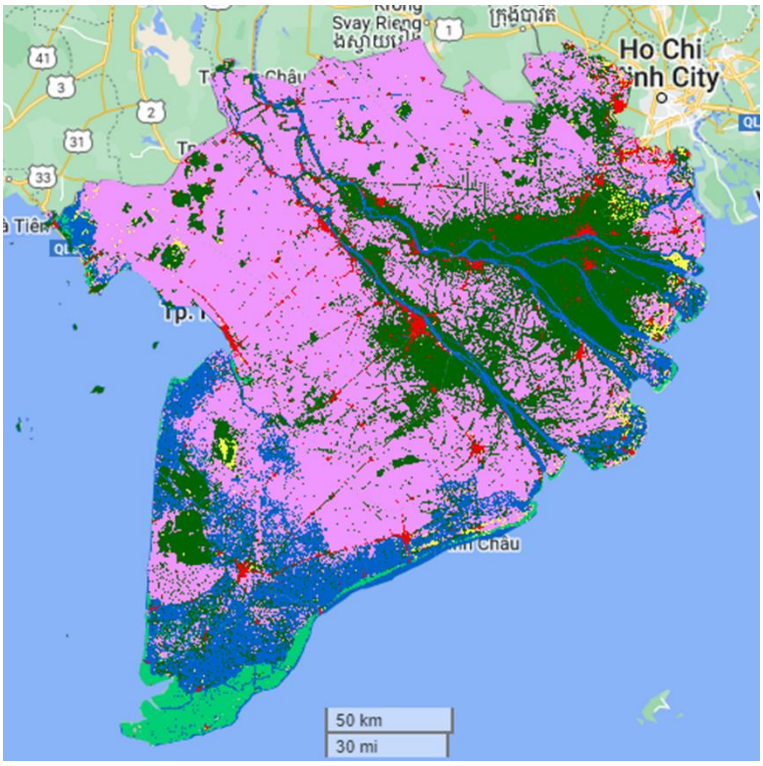

***Figure 1** Land cover types in the Mekong River Delta as defined by the European Space Agency (ESA) WorldCover product.*

The final resolution and demonstrated the scalability of the workflow were key considerations in this mapping exercise, so the ESA WorldCover product was chosen after merging the 'Cropland' and 'Tree cover' classes to create a mask.

2.2.2 Sentinel-1 and Sentinel-2

The Sentinel mission satellites offer high spatial and temporal resolution imagery for numerous applications. Since their launch, they have been widely applied to cropland mapping and monitoring activities. Sentinel-1 collects synthetic aperture radar data which offers the benefit of being unaffected by cloud and haze interference. Sentinel-2 collects optical band reflectance, which clouds, and other interferences can inhibit. The optical bands in Sentinel-2 allow observation of vegetation dynamics and vegetation indices such as Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), which enable the calculation of phenological statistics.

Sentinel-1 and 2 images were collected from GEE for 2020, with a cloudy pixel percentage set to below 20% for Sentinel-2. Once images were loaded, cloud masks were used to eliminate cloudy pixels from Sentinel-2. Sentinel-1 is consistent due to no cloud interference, while clean Sentinel-2 images are scarce during the wet and rainy seasons from June to November. The Sentinel-1 and 2 bands used to train and test the machine learning algorithm are shown in **Table 2**.

Table 2 Band designations and descriptions of Sentinel instruments used in the crop type mapping workflow.

Satellite	Band ID	Band Description	Approx. Central Wavelength (nm)
Sentinel-2	B2	Blue	495
	B3	Green	560
	B4	Red	665
	B5	Red Edge 1	704
	B6	Red Edge 2	740
	B7	Red Edge 3	780
	B8	Near Infra-Red	834
	B8A	Red Edge 4 / narrow Near Infra-Red	864
Sentinel-1	VV	Vertical transmit/ vertical receive	-
	VH	Vertical transmit/horizontal receive	-

The annual median images were taken to produce training data, and the time series were used to generate the phenological statistics. Using temporal aggregations can reduce overfitting and reduce the 'curse of dimensionality' in machine learning approaches.

2.3 Methodology

2.3.1 Calculation of phenological statistics

Temporal patterns in vegetation indices can be used to calculate phenological statistics for cropland, such as the Start of Season (SoS) and End of Season (EoS). This workflow adapted some code written in GEE JavaScript due to its applicability to GEE objects. Other methods exist to calculate phenological statistics in Python packages such as xarray, though converting GEE objects to the required data formats can be memory intensive and may not be possible, depending on the Jupyter Notebook's environment. The method used in this workflow utilizes the cloud computing capacity of GEE.

Calculation of phenological statistics requires a vegetation index. In this case, the EVI was calculated on Sentinel-2 imagery as follows (**Justice et al. 1998**)

$$EVI = 2.5\,x\,\frac{nir - red}{(nir + 6\,x\,red - 7.5\,x\,blue) + 1}$$

EVI is similar to the Normalized Difference Vegetation Index (NDVI), though it corrects for some atmospheric factors and performs better in high biomass conditions, such as in the orchards and plantations in the Mekong River Delta (**Justice et al. 1998**).

2.3.2 Random forest classification

Random forest classification was used to train and predict crop types. GEE allows using the *smileRandomForest* function, which performs rapid training and prediction computation for large geospatial datasets. Before training, the labelled geometries were randomly divided into a training set of 80% polygons and a test set of 20%. Training data comprising values from the satellite bands in **Table 2** and SoS and EoS statistics were generated.

2.3.3 Accuracy assessment

Cross-validation was used to assess model accuracy using the test data, which was withheld from the initial training step. The accuracy of the machine learning classifications was ultimately quantified by calculating overall accuracy on the test dataset, which is the proportion of test sites classified correctly. Producing and inspecting a confusion matrix depicting observed and predicted classes for the test data enabled observation of potentially misclassified classes.

3 Results

3.1 Accuracy of classification

The overall accuracy of the random forest classification with 500 trees was 96%, the same as the overall accuracy of the random forest classifier with 100 trees. The confusion matrix (**Figure 2**) for the random forest classification with 500 trees did not reveal any major misclassifications. The 'Other_Veg' category was not classified as accurately as other classes, likely because of its low overall occurrence.

3.2 Variable importance

The variable importance in the random forest classifier is available using the GEE explainer command. The crop type classification relied heavily on the Sentinel-1 and 2 bands (**Figure 3**). The Sentinel-1 bands could likely differentiate frequently flooded rice fields from perennial tree crops and drier agricultural land types.

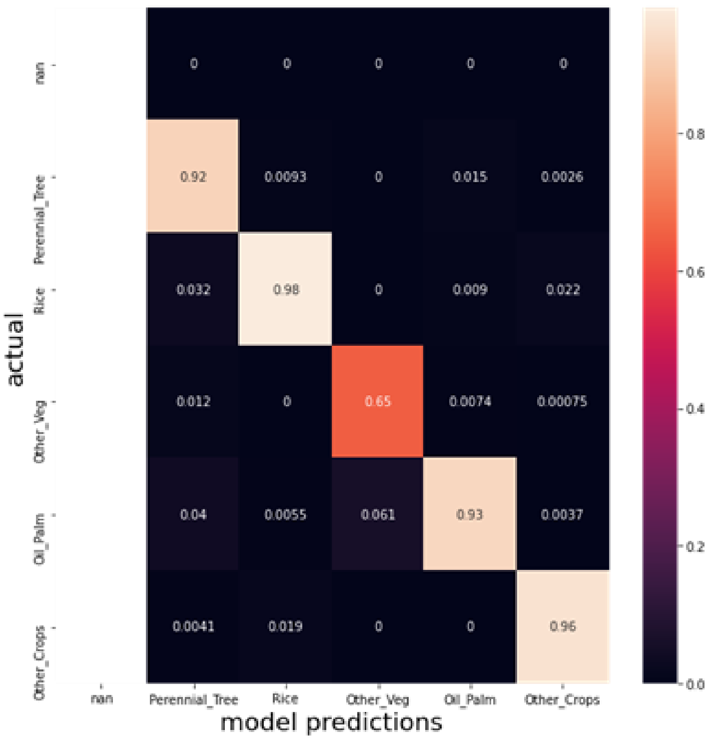

Figure 2 *Confusion matrix computed on test data for crop types with a model prediction from a random forest classifier.*

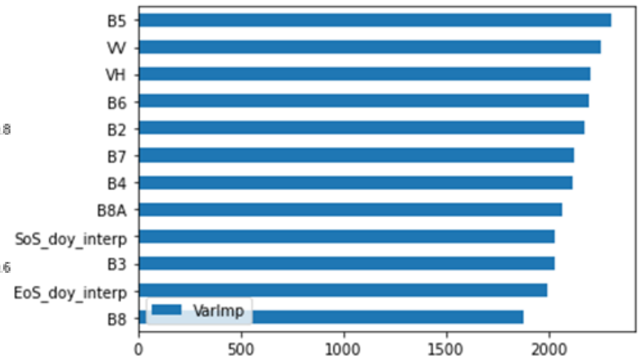

Figure 3 *Variable importance for crop type classification computed as Gini impurity index over all trees in the forest.*

3.3 Area and distribution

The classification indicates that rice and perennial tree crops dominate the Mekong River cropland, accounting for over 90% of the cropland when combined (**Table 3**). Further division of perennial tree crops in future classifications will inform their distribution. Oil palms and other crops accounted for nearly 9% of the area.

Table 3 *Area of different land classes in the Mekong River Delta based on prediction from a random forest classification.*

Province	Crop Type	Area (sq. km)	Area (%)
An Giang	Perennial Tree	576.3	25.2
	Rice	1,610.5	70.5
	Other Veg.	-	-
	Oil Palm	65.8	2.9
	Other Crops	31.1	1.4
Ben Tre	Perennial Tree	593.9	35.7
	Rice	738.2	44.4
	Other Veg.	-	-
	Oil Palm	305.0	18.3
	Other Crops	26.3	1.6
Dong Thap	Perennial Tree	480.7	28.8
	Rice	1,090.2	65.3
	Other Veg.	-	-
	Oil Palm	47.2	2.8
	Other Crops	52.7	3.2
Kien Giang	Perennial Tree	856.3	23.6
	Rice	2,564.2	70.6
	Other Veg.	3.9	0.1
	Oil Palm	51.9	1.4
	Other Crops	155.0	4.3
Long An	Perennial Tree	830.6	31.8
	Rice	1,457.0	55.8
	Other Veg.	-	-
	Oil Palm	177.5	6.8
	Other Crops	145.3	5.6
Tien Giang	Perennial Tree	951.3	66.7
	Rice	389.0	27.3
	Other Veg.	-	-
	Oil Palm	79.3	5.6
	Other Crops	5.6	0.4
Soc Trang	Perennial Tree	517.8	22.2
	Rice	1,625.0	69.6
	Other Veg.	5.9	0.3
	Oil Palm	145.1	6.2
	Other Crops	41.2	1.8
Tra Vinh	Perennial Tree	581.5	31.6
	Rice	1,026.9	55.9
	Other Veg.	-	-
	Oil Palm	151.7	8.3
	Other Crops	78.1	4.2
Vinh Long	Perennial Tree	648.3	66.3
	Rice	259.6	26.6
	Other Veg.	-	-
	Oil Palm	66.7	6.8
	Other Crops	2.9	0.3
Bac Lieu	Perennial Tree	72.8	4.1
	Rice	1,666.3	93.7
	Other Veg.	-	-
	Oil Palm	4.9	0.3
	Other Crops	33.9	1.9
Ca Mau	Perennial Tree	520.8	14.3
	Rice	2,862.2	78.3
	Other Veg.	1.9	0.1
	Oil Palm	30.4	0.8
	Other Crops	238.0	6.5
Can Tho City	Perennial Tree	245.7	45.7
	Rice	247.4	46.1
	Other Veg.	-	-
	Oil Palm	39.0	7.3
	Other Crops	5.0	0.9
Hau Giang	Perennial Tree	664.8	58.0
	Rice	401.7	35.1
	Other Veg.	1.0	0.1
	Oil Palm	72.1	6.3
	Other Crops	5.9	0.5
Total	**Perennial Tree**	**7,794.0**	**30.5**
	Rice	**15,853.0**	**62.0**
	Other Veg.	**11.8**	**0.0**
	Oil Palm	**1,101.7**	**4.3**
	Other Crops	**790.8**	**3.1**

Figure 4 and **Figure 5** show the distribution of crop types in the Mekong River Delta. Perennial tree crops and Oil Palm are concentrated in the central area of the Delta, whereas rice is more prevalent in the upper and lower parts of the Delta.

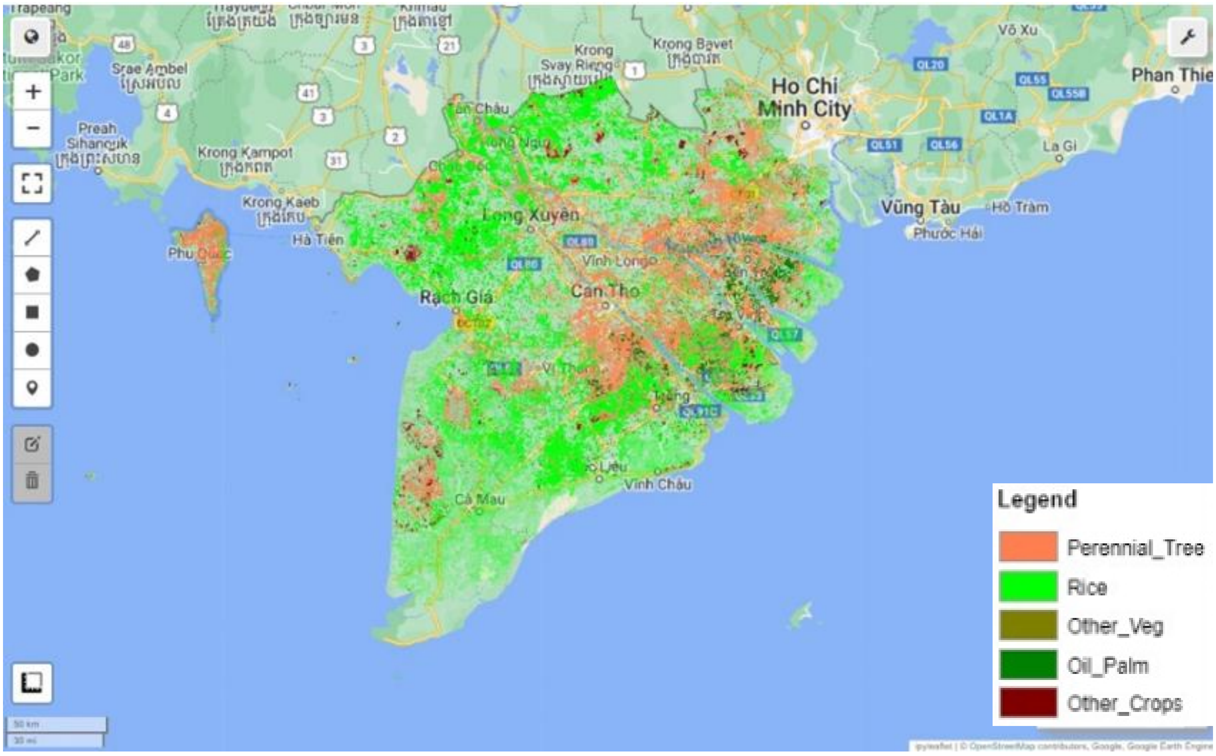

Figure 4 *Crop type map of the Mekong River Delta produced from predictions of a random forest classification algorithm.*

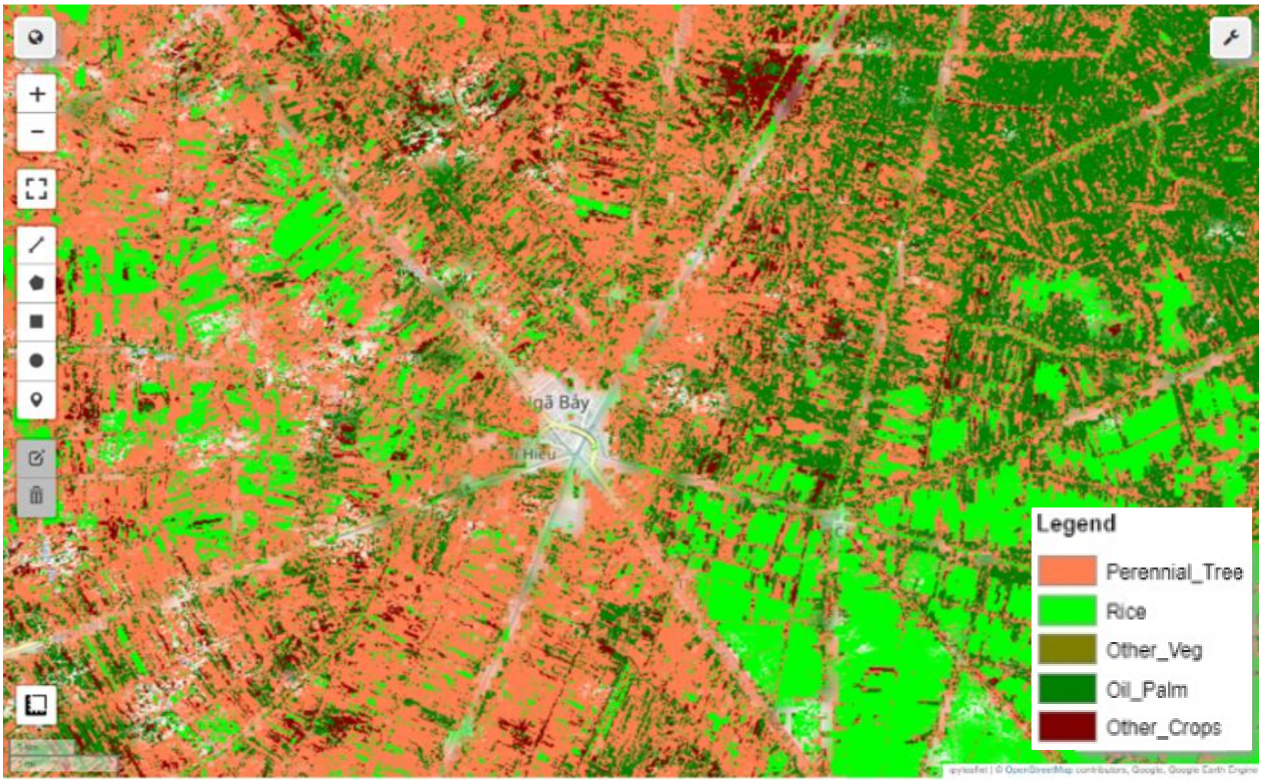

Figure 5 *Detailed crop type mapping in a central area of the Mekong River Delta showing the distribution of perennial tree crops, other crops, rice, and Oil Palms in a small area.*

The rice seasonality map (**Figure 6**) shows the concentration of triple rice cropping in the central Delta and single and double rice cropping in the southwest and coastal areas. **Figure 6** shows that triple rice cropping is the dominant strategy, followed by single cropping. Double cropping is less prevalent.

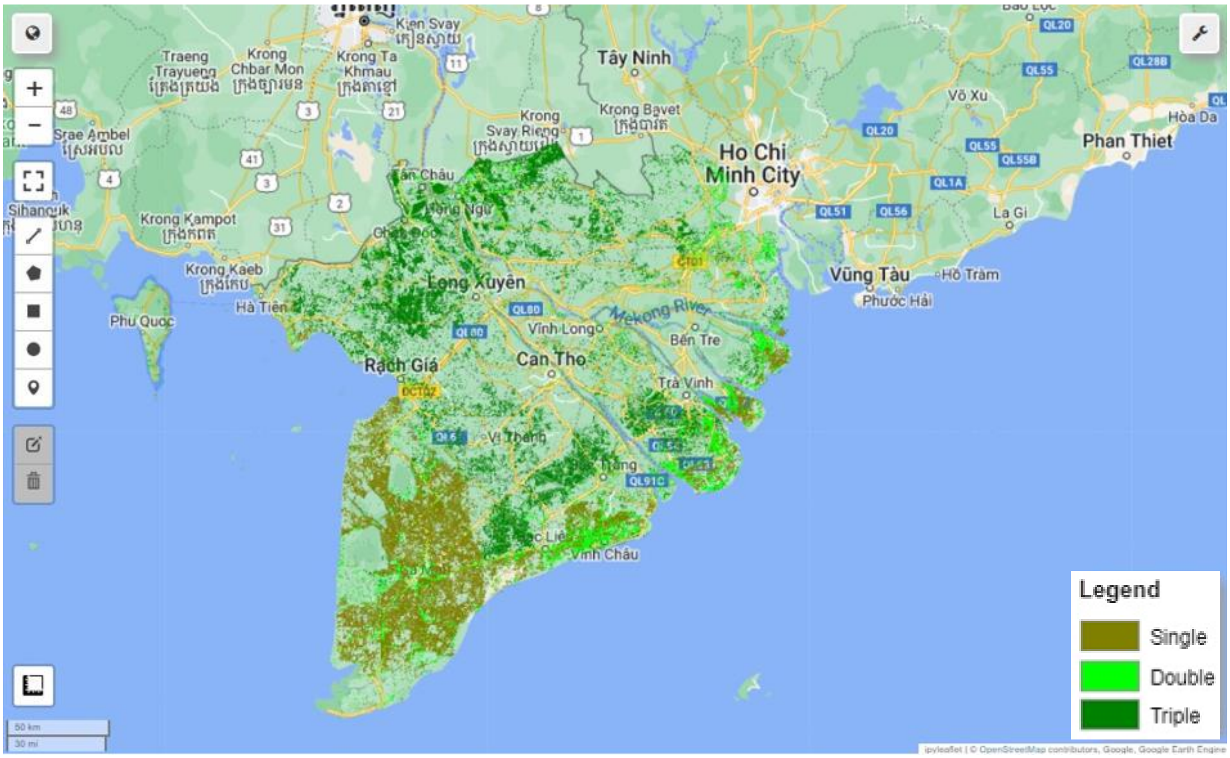

Figure 6 Distribution of single, double, and triple rice in the Mekong River Delta.

4 Discussion and Conclusion

The workflow developed can be rapidly adapted to crop-type mapping in numerous South-East Asian regions, depending on the availability of reliable training data. Jupyter Notebooks and cloud computing on GEE allow users to alter the region of interest, input data, and other parameters and run large-scale classifications on any machine with a reliable internet connection. Researchers can adapt and update crop-type classifications as necessary. Some aspects of the workflow are regionally specific. For example, the application of Sentinel-1 is likely to be especially useful in Delta agroecosystems where irrigated rice and aquaculture dominate. The high accuracy of random forest classification on test data in this workflow demonstrates its ability to differentiate between major crop types. Increasing the number of classes may decrease classification accuracy, though initial results are very promising.

In any case, the workflow demonstrates that the phenological approach can accurately distinguish rice crop sequences. This workflow adapted the approach taken by **Son et al.** (2014) with moderate resolution MODIS imagery to higher resolution Sentinel-2 derived data. The number of crop type classes presented in this workflow was limited by the availability of reliable and accurate training data, especially for perennial tree crop types. The visual inspection of satellite images and public Street View imagery approach to training data collection is also associated with some uncertainty.

Training data collection is often challenging for accurate crop type mapping and can be difficult and expensive to collect, especially when seasonality is considered **(Waldner et al. 2019)**. Automated geotagged ground imagery is increasingly being explored for the cost-effective collection of training data for crop-type mapping (**Tseng et al. 2021**). Ongoing additions to public imagery, such as Google Maps Street View, may also allow for cost-effective ground data collection.

The workflow presented in this report and its associated Jupyter Notebook presents a scalable method for crop type classification in the Mekong River Delta, which can be applied to other South-East Asian countries. It utilizes public satellite data repositories and cloud computing facilities which means that data can be streamed, and large-scale classification computation can be run on any machine, with results produced relatively rapidly. The workflow, therefore, aligns with the principles of Findability, Accessibility, Interoperability, and Reuse (FAIR) due to its reliance on open-access datasets, code, environments and documentation for reproducibility. From an agronomic perspective, incorporating phenological predictors allows crop sequences to be differentiated within a crop type, such as single, double, and triple rice.

References

Anthony, E, Brunier, G, Besset, M, Goichot, M, Dussouillez, P & Nguyen, V 2015, 'Linking rapid erosion of the Mekong River delta to human activities', *SCIENTIFIC REPORTS*, vol. 5, p. 14745.

Gorelick, N, Hancher, M, Dixon, M, Ilyushchenko, S, Thau, D & Moore, R 2017, 'Google Earth Engine: Planetary-scale geospatial analysis for everyone', *Remote Sensing of Environment*, vol. 202, pp. 18-27.

Hoanh, CT, Suhardiman, D & Anh, LT 2014, 'Irrigation development in the Vietnamese Mekong Delta', *International Journal of Water Governance*, vol. 2, no. 2.

James, G, Witten, D, Hastie, T & Tibshirani, R 2014, *An Introduction to Statistical Learning: with Applications in R*, Springer Publishing Company Incorporated.

Justice, CO, Vermote, E, Townshend, JR, Defries, R, Roy, DP, Hall, DK, Salomonson, VV, Privette, JL, Riggs, G & Strahler, A 1998, 'The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 4, pp. 1228-49.

Nhan, DK, Phap, VA, Phuc, TH & Trung, NH 2012, 'Rice production response and technological measures to adapt to salinity intrusion in the coastal Mekong delta', *Can Tho: Can Tho University*.

Nhan, DK, Trung, NH & Sanh, NV 2011, 'The impact of weather variability on rice and aquaculture production in the Mekong Delta', in *Environmental change and agricultural sustainability in the Mekong Delta*, Springer, pp. 437-51.

Schneider, P & Asch, F 2020, 'Rice production and food security in Asian Mega deltas—A review on characteristics, vulnerabilities and agricultural adaptation options to cope with climate change', *Journal of Agronomy and Crop Science*, vol. 206, no. 4, pp. 491-503.

Sebesvari, Z, Le, TTH & Renaud, FG 2011, 'Climate change adaptation and agrichemicals in the Mekong Delta, Vietnam', in *Environmental change and agricultural sustainability in the Mekong Delta*, Springer, pp. 219-39.

Son, N-T, Chen, C-F, Chen, C-R, Duc, H-N & Chang, L-Y 2014, 'A Phenology-Based Classification of Time-Series MODIS Data for Rice Crop Monitoring in Mekong Delta, Vietnam', Remote Sensing, vol. 6, no. 1, pp. 135-56.

World Bank 2021, *For Mekong Delta farmers diversification is the key to climate resilience*, The World Bank, <https://www.worldbank.org/en/news/feature/2021/10/21/for-mekong-delta-farmers-diversification-is-the-key-to-climate-resilience>.

Tseng, G, Zvonkov, I, Nakalembe, CL & Kerner, H 2021, 'Cropharvest: A global dataset for crop-type classification', in *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Waldner, F, Chen, Y, Lawes, R & Hochman, Z 2019, 'Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods', *Remote Sensing of Environment*, vol. 233, p. 111375.

Wu, Q 2020, 'geemap: A Python package for interactive mapping with Google Earth Engine', *Journal of open source software*, vol. 5, no. 51, p. 2305.

Zanaga, D, Van De Kerchove, R, Daems, D, De Keersmaecker, W, Brockmann, C, Kirches, G, Wevers, J, Cartus, O, Santoro, M & Fritz, S 2022, 'ESA WorldCover 10 m 2021 v200'.

SUGGESTED CITATION

Ghosh, S.; Wellington, M.; Holmatov, B. 2022. *Mekong River Delta crop mapping using a machine learning approach*. Colombo, Sri Lanka: International Water Management Institute (IWMI). CGIAR Initiative on Low-Emission Food Systems (Mitigate+). 11 p.

FUNDING ACKNOWLEDGEMENT

We would like to thank all funders who supported this research through their contributions to the CGIAR Trust Fund: https://www.cgiar.org/funders/

To learn more, please contact: s.ghosh@cgiar.org

ABOUT MITIGATE+

Mitigate+ is a One CGIAR initiative that aims to reduce annual global food systems emissions by working closely with key actors in the target countries to ensure they are equipped to make evidence-based decisions and address challenges in food systems discourse, policy development, and implementation to reduce greenhouse gas emissions. For more details about the initiative see https://www.cgiar.org/initiative/low-emission-food-systems/

ABOUT CGIAR

CGIAR is a global research partnership for a food secure future. Visit https://www.cgiar.org/research/cgiar-portfolio to learn more about the initiatives in the CGIAR research portfolio

DISCLAIMER