Agricultural Land Monitoring on Amazon Web Service

Soravis Supavetch[1] and Tranid Prasertsri[2]
[1]Faculty of Engineering, Kasetsart University, Lat Yao, Chatuchank, Bangkok, Thailand,
Email: fengsvsu@ku.ac.th
[2]AI and Robotics Ventures Co., Ltd., Bhiraj Tower at Sathorn Building C 33,31 31/1 S, Sathorn Road, Yannawa, Sathorn, Bangkok, Thailand,
Email: tranidp@arv.co.th

KEY WORDS: Amazon Web Service, Sentinel-2, Agricultural Land Monitoring Platform

ABSTRACT: The paper demonstrates developing a remote sensing cloud processing platform from the Sentinel-2 dataset (S2) and SMAP supporting the ASEAN agriculture management system. Every month, S2 from the S3 bucket is download to a Varuna Land Monitoring (VLM) platform (on Amazon EC2). Then the pre-processing such as vegetation index (NDVI: Normalized Vegetation Index and NDMI: Normalized Different Moisture Index), cloud masking (using Scene Classification (SC) image), fill-gap (using the mathematical model) are executed for each grid tile over the 98 tiles of the Thailand region (for this initial stage of VLM to the entire ASEAN region further). Each acquisition image is collected and composited into a monthly dataset using the maximum value composite (MVC) method. MVC is a suitable method for plant growth monitoring that is the major purpose of the VLM platform. Then UTM reference image is transformed into a geographic coordinate system. The pixels of each tile are rearranged to fill the gap of each twisted image frame due to the projection transformation process. This process is necessary for the data dissemination platform to cover a large area. The VLM dataset publishes in the OGC web services using GeoServer by ImageMosaic datastore type on the tile caching technology using the GWC S3 BlobStore plugin (expanding the storage throughout S3 bucket technology). Finally, the vegetation indices will publish to geospatial end-user (agricultural researcher community) through WMTS (Web Map Tile Service) protocol. The computing algorithm of the steps is explained and discussed in this paper's sections for the recommendations of the next improvement. Finally, the recommendation is explained in the conclusion of this paper.

1. INTRODUCTION

Varuna Land Monitoring (VLM) platform is built from the cooperation between the private and academic sector; AI Robotics Venture. Co., Ltd. under PTT Exploration and Production Public Company Limited and Kasetsart University, respectively. The platform aims to support high-resolution agricultural relevant data and is proofed in quality from the academic sector then distributed to the ASEAN user communities such as research, industry, and government. The low cost of preprocessed data is the key to driving the satellite data technology from the VLM platform to local implementation. Agriculture in the ASEAN region is unique and differs from others because it is a tropical crop with small farm size and most mixed plants. The moderate resolution satellite data such as MODIS or VIIRS sensors are not sufficient to local need quality. In contrast, high-resolution data have a higher cost from resources for processing and storage. For this reason, VLM is the initiated solution to move the ASEAN agriculture sector from those limits of support to the use of satellite data technology for sustainable agro-industry further.

The ASEAN region is at high risk of climate change affecting agro-industry because most countries have a high GDP percentage of the agriculture sector. Every year ASEAN spends a lot of budget subsidies in agriculture for helping their farmers. Satellite data is the best solution for facing natural disasters to prepare, discover, reduce and mitigate the impact if it does not over increase their investment. Thus, the purpose of the VLM platform is capacity building to take advantage of satellite data to that small farms through private, government, research sectors that fit local demand and characteristics.

2. PLATFORM DESCRIPTION

For supporting various communities, the platform is divided into three subgroups of applications. In Thailand case, satellite data used in three kinds of format, i.e., use bands composition for visualizing and analysis by an expert user such as land use classification (Land Development Department), index-based data supporting analysis direct (i.e., normalized difference vegetation index, normalized difference water index, soil moisture, and temperature), and the analysis-ready data such as yield, health, harvest detection, age, evapotranspiration, carbon credit, and so on. Thus, VLM platforms are designed into three types: (1) Sensor Data Records (SDR) provides the false-color composite image to a visualized use such as expert user used for monitoring the change of land use and land cover type. (2) Environmental Data Records (EDR) for analysis application. For example, yield prediction, vegetation health monitoring, farm fertilizer management, crop stress monitoring, and so on. This EDR is a solution from an industry and government requirement because they have an analysis system support. (3) Application Data Records (ADR), the business-based functional data format. This supporting data type design for no expertise exists in the organization, which requires a full-function platform implementation (Figure 1).

VLM was developed using opensource based libraries and applications in the Amazon cloud platform. In data preparation, open platform tools such as gdal, ogr, osgeo, and numpy base on python language are used to automate downloads and data preparation sections. GeoServer, opensource software used to provide OGC Web Map Service (WMS), Web Map Tile Service (WMTS), and Tile Map Service (TMS) to service users with GeoWebCache (GWC) S3 (simple cloud storage) BlobStore technology (GeoServer, 2021). Python-based data preparation platform connects to JAVA-based software (using Apache Tomcat as a web HTTP server) such as GeoServer through REST service API for layers (satellite dataset) management.

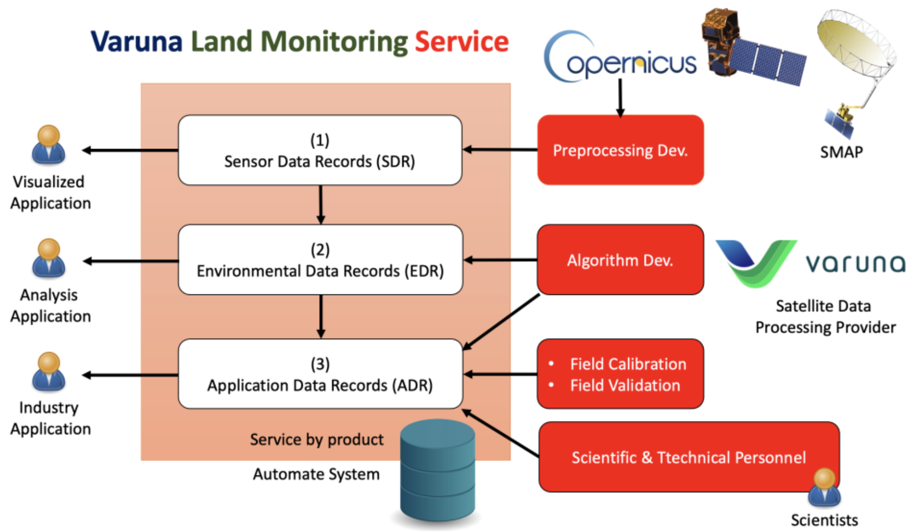


Figure 1 Varuna Land Monitoring Service Platform

Sentinel-2 and SMAP (Soil Moisture Active Passive) are based satellites for the VLM platform and distributed through three data services categories mentioned above. Sentinel-2 multispectral sensors use for vegetation health monitoring. SMAP data for crop growth modeling such as monthly average surface temperature and surface soil moisture. Soil moisture and temperature data are used to interpret the climate change effect on vegetation due to climate factors such as rainfall, humidity, or pressure in the climate cycle that have changed over time.

Sentinel-2 (TOA: Top Of Atmosphere) L1C and (BOA: Bottom Of Atmosphere) L2A download from open data in AWS S3 bucket using a command-line interface (CLI) every ten days. SMAP-Sentinel Level 2 Soil Moisture Active-Passive (L2SMSP) data is also downloaded every ten days through NSIDAC (National Snow & Ice Data Center Distributed) DAAC (Distributed Active Archive Center) using python script.

3. SYSTEM DESCRIPTION

Amazon Elastic Compute Cloud (Amazon EC2) is used for the platform computing for VLM. AWS EC2 provides secure, and resizable compute capacity in the cloud that suitable platform for VLM further supports the processing covering the ASEAN region in the future. VLM processing develops in two processing chains, i.e., Sentinel-2 and SMAP. Sentinel-2 for Thailand covers 98 grid tiles (each tile cover 100x100 square kilometers) is processing in every ten days. SMAP 1x1 square kilometers resolution covers 213 overlapped grid tiles (each tile cover 290x220 square kilometers) for the monthly dataset.

3.1 Sentinel-2 Data Processing

Sentinel-2 L1C product download from AWS S3 bucket every ten days to VLM then atmospheric correction is executed using Sen2Cor tools to produce L2A product in 20x20 meters resolution and store in Amazon EBS (Elastic Block Store). Scene classification (SCL) layer masks cloud and shadow coverage area from vegetation index image. Due to the equatorial climate, cloud density in ASEAN countries is high in the rainy season between May and September. Then monthly composition for cloud-free is recommended to use instead of a daily dataset for crop growth monitoring. Due to Sentinel-2 coordinates reference system is UTM when used to compose the whole region and publish through OGC WMS/WMTS/TMS and use to industry application the preferable coordinate system is geographic coordinates system. UTM reference datasets are converted to the geographic coordinate reference system. Available pixel data are twisted due to geographic grid projection that will be a problem when published in the overlap area. GeoServer does not support the compensation mechanism over the overlap area when it composes a raster dataset to respond GetMap request. For solving this limit, VLM re-arranges pixels (in the overlap area) using a competitive selection algorithm over the candidate overlap pixels. Finally, the 98 raster tiles publish to a layer in GeoServer through REST API into the ImageMosaic datastore type. GeoServer manages grid tile requested using GeoWebCache with S3 BlobStore plugin for scalable tile caching storage in AWS S3 bucket technology (Figure 2) as a suggestion in Testbed 10 Performance of OGC Services in the Cloud engineering report (OGC 14-028r1).

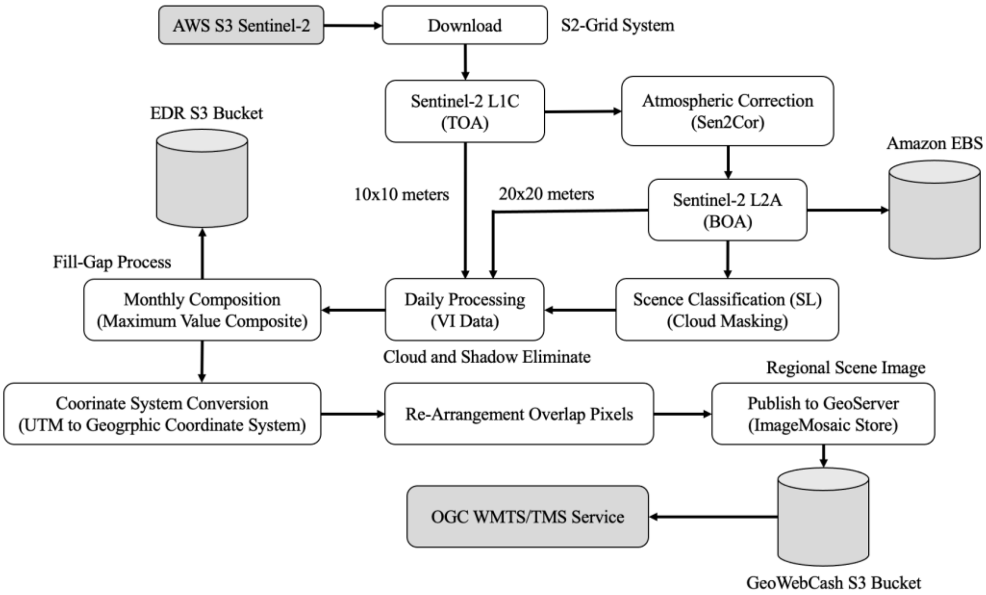


Figure 2 Sentinel-2 data processing chain

3.2 SMAP Data Processing

SMAP-Sentinel Level 2 Soil Moisture Active-Passive (L2SMSP) product is the product using Sentinel-1A/1B data in the SMAP active-passive algorithm for the potential of obtaining the disaggregated brightness temperature and soil moisture at much more satisfactory spatial resolutions (less than 3 km) (Das 2020). VLM downloads SMAP L2SMSP product from NSIDAC DAAC and temporal store in Amazon EBS. It extracts soil moisture and temperature 1 km resolution sub dataset from Equal-Area Scalable Earth-2 (EASE2) grid every ten days to the geographic coordinates reference system. Then composite daily data using an average, maximum, minimum method on overlap pixels into a monthly dataset for an equivalent period with Sentinel-2 vegetation indices. The three types of statistics support such analysis as evapotranspiration using satellite-based data like the Hargreave-Samani equation (George H. Hargreaves 2003) or another crop model.

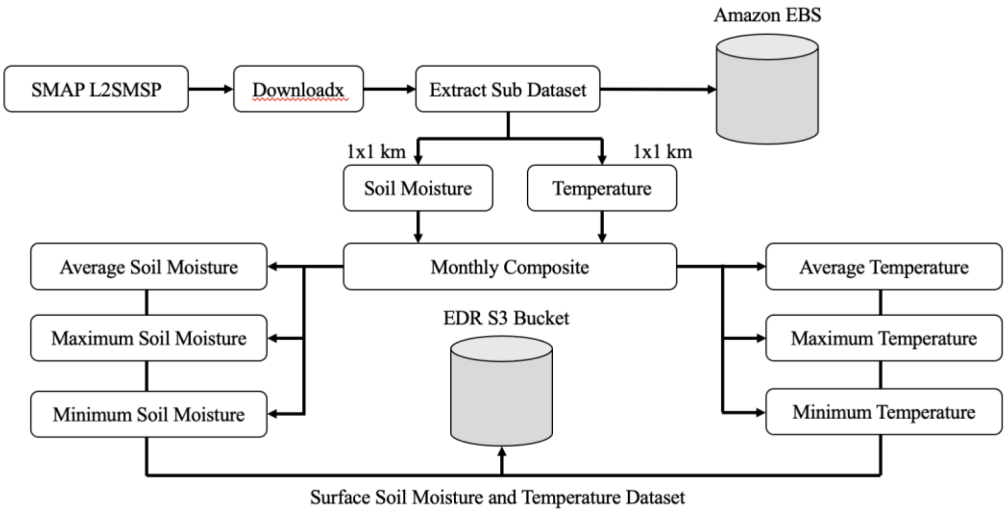


Figure 3 SMAP L2SMSP data processing chain

Figure 4 displays volumetric soil moisture between January - August 2021 from EDR. This product is under the field calibration and validation for downscaling and improve precision to use with 20x20 square meters NDVI in crop monitoring model. However this 1x1 square kilometers product is be able to use in weather forcasting model or drought monitoring.

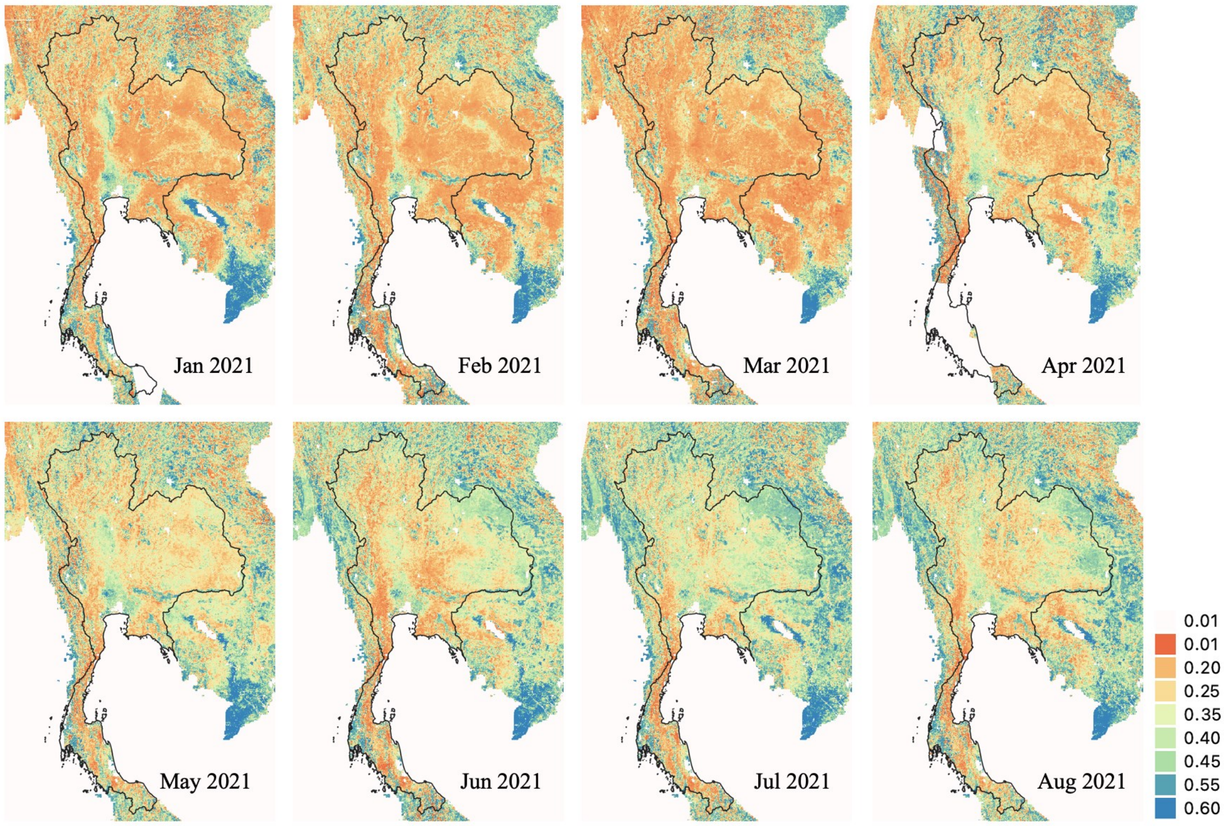


Figure 4 SMAP L2SMSP soil moisture (monthly composite)

4. THAILAND IMPLEMENTATION CASES

These are the example cases that demonstrate how VLM supports industry and government by using high-resolution satellite data. In the sugarcane industry, mills investigate and monitor cane farms by sending their person to visit the contract farmer and their farm for more than a thousand per month. A lot of work process is repeated every month before a harvest period. Satellite data (NDVI time series) is used to monitor a farm and estimate the yield since a growing period. The mill can monitor cane health by reading the growth of NDVI instead of a field visit at all farms. Many ten thousand farms are watching by the satellite's eyes.

Government disaster solving the problem, air pollution case study caused by agriculture fire, satellite data reveal the farm was burning for new crop or burning for harvesting at date and time. The study showing NDVI time series can identify the harvest and burn period or can be said that satellite data could investigate those farm activities.

4.1 Sugarcane Industry Application

The monthly NDVI for each farmland from the EDR S3 bucket displays sugarcane growth during the cultivation period. In the implementation to the factory, NDVI display to users on the web platform, which is calculated using zonal statistics with farm polygon. NDVI series demonstrate the health of growth and infer to yield at harvest time. Yield, which is predicted from a mathematical model of NDVI series, indicates each farm growth quality. Factory users utilize NDVI to plan support to farm and field visits later on. Geospatial application technologies such as GeoTools, GDAL/OGR, Leaflet JS, and GeoServer are included to develop the factory's actual implementation, as demonstrated in Figure 5.

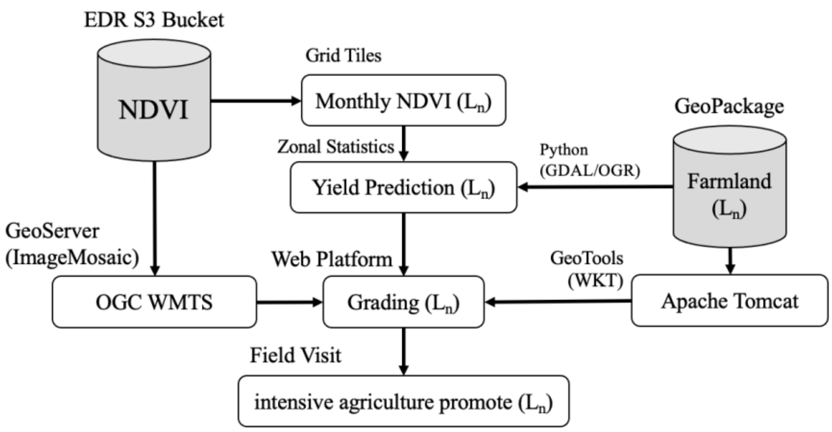


Figure 5 Workflow of the EDR supporting sugarcane industry

The growth of NDVI in stages, i.e., germination, tillering, grand growth, and maturity, displays the health of cane. In our study, the NDVI time series during this period can be used with a proper mathematical model to predict the yield at the end of the season. Thus, the prediction using NDVI time series are collected and display to the cane mill for grading farms to promote a fertilizer to the low productivity farmland. Figure 6 demonstrates the front-end application (web application) in which the NDVI time series of each farmland is listed and used to calculate a yield prediction. Each yield indicates low or high productivity for which needs a fertilizer promotion. Cane health also displays the effect of a natural disaster such as drought and flood in which sugarcane responds to that influenced by display lower NDVI than the average value of each growth period.

ลำดับ	รหัส	การเจริญเติบโต											คุณภาพ	พื้นที่ (ไร่)	ตัน/ไร่	ผลผลิต (ตัน)	การเจริญเติบโต
		ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.					
1	2	0.62	0.48	0.23	0.38	0.51	0.53	0.58	0.74	-	-	-	Good	12.5	11.4	142	
2	3	0.61	0.46	0.22	0.39	0.63	0.61	0.66	0.77	-	-	-	Good	21.7	12.6	272	
3	4	0.64	0.28	0.35	0.62	0.72	0.72	0.67	0.74	-	-	-	Very Good	5.0	14.8	74	
4	5	0.69	0.24	0.33	0.62	0.73	0.75	0.68	0.75	-	-	-	Excellent	11.8	15.0	177	
5	6	0.71	0.27	0.34	0.61	0.72	0.74	0.67	0.73	-	-	-	Excellent	11.0	15.0	164	
6	7	0.73	0.39	0.41	0.65	0.73	0.73	0.69	0.72	-	-	-	Excellent	5.0	16.3	81	
7	8	0.72	0.52	0.51	0.69	0.76	0.77	0.76	0.79	-	-	-	Excellent	0.5	18.3	8	
8	9	0.54	0.29	0.28	0.37	0.45	0.52	0.56	0.66	-	-	-	Average	8.1	9.8	79	
9	10	0.46	0.29	0.27	0.32	0.38	0.50	0.62	0.59	-	-	-	Average	3.8	8.9	34	
10	11	0.55	0.42	0.42	0.59	0.59	0.53	0.50	0.63	-	-	-	Very Good	5.0	13.0	65	

Figure 6 Sugarcane health and yield prediction during plantation period

NDVI time series visualized as a map as Figure 7 demonstrates homogeneity of a cane and where that need fertilizer or re-cultivate. Each month, the mill investigates this farmland before planning to visit a farmer and farmland.

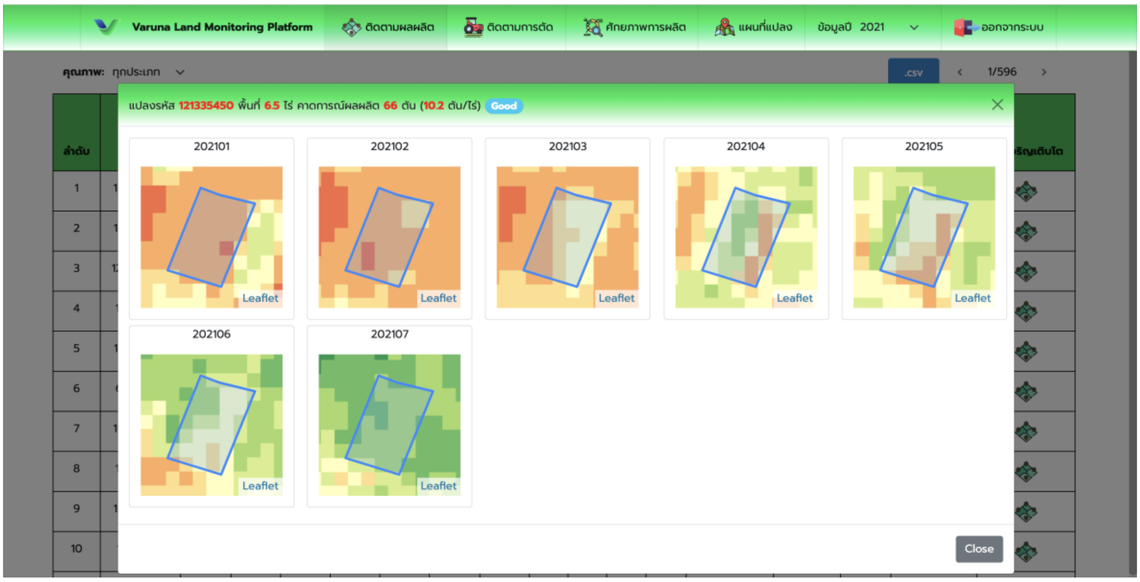
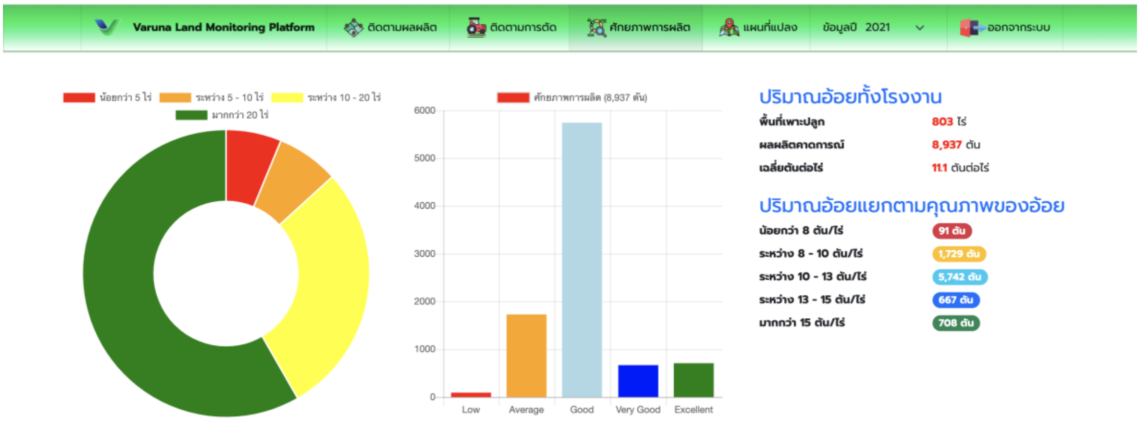
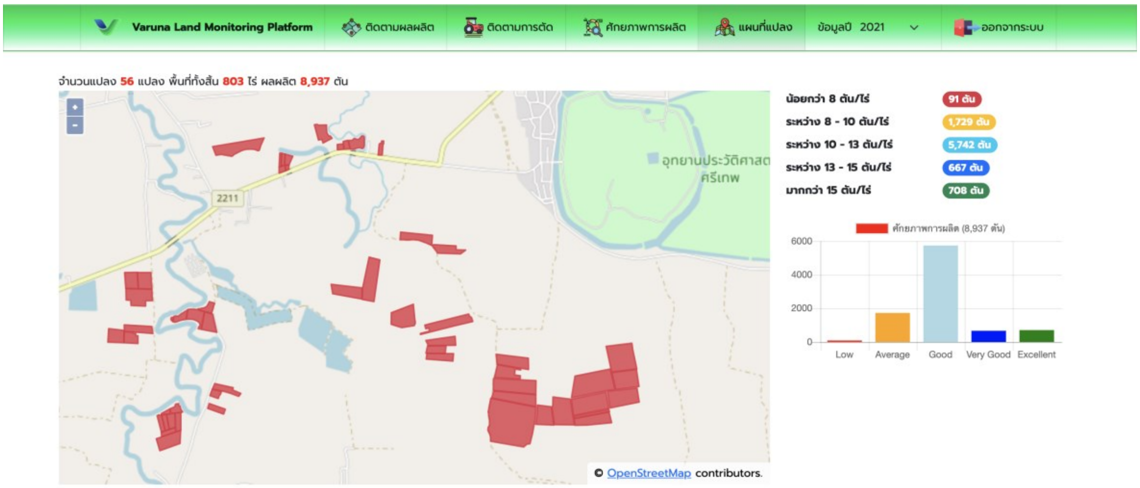


Figure 7 Visualized monitoring from OGC WMTS through a geospatial web application

The VLM implementation platform's great advantage is that it makes a lower cost and time dependency of practical works such as staffing, field visit, and the investment (the crop-related observation instrument) over the large area. At the same time, it improves precision and accuracy from regional-scale remote sensing to farm-scale remote sensing. In Figure 8, factory users read a farm assessment from the platform every month and report specific low yields for farmers contacting and helping them improve productivity by offering fertilizer and another type of support. The factory productivity summary gives the factory a plan to increase the yield reaching the target yield. This approach precisely shows that high-resolution data ready from the VLM platform yet collapses the budget barriers of the general precision farming on a factory scale.



(a)



(b)

Figure 8 Visualized monitoring from OGC WMTS through a geospatial web application. (a) (b)

4.2 Agriculture Fire Study

Fire in agriculture case study differs from previous industry scenario because it needs the solution as evidence which can prove to compare with the field investigation. Fire in the harvesting season, including wildfire, is a significant cause of air pollution in Thailand between November to April (pollution peak season). The government tries to develop an approach to reduce the fire in both types using policy, advertisement, promotion, etc. The human induces fire, which is used in agricultural activity, burns for two primary purposes, (1) burns for harvesting, and (2) burns for agricultural waste clearance such as leaf sheaths. The burns in both activities are related to a cost compared to other methods such as plow up and landfill. Then farmers decided to burn a fire due to its make plantation cost lower.

The workflow corresponding to the government requirement shows in Figure 9. NDVI time series identifies a crop activity in a harvesting case by their phenotyping construction. When it is harvesting, the NDVI time series will display the value's reducing rate that decreases to 0.3 (bare soil or no vegetation). The FIRMS (Fire Information and Resource Management System) identifies hotspot detected location from VIIRS sensor (Suomi-NPP satellite). It is then used as a starting point for phenotyping by collecting the NDVI time series to display the crop cycle at the vicinity spot. Finally, the hotspot's date and time in that crop cycle will able to identify their fire activity types.

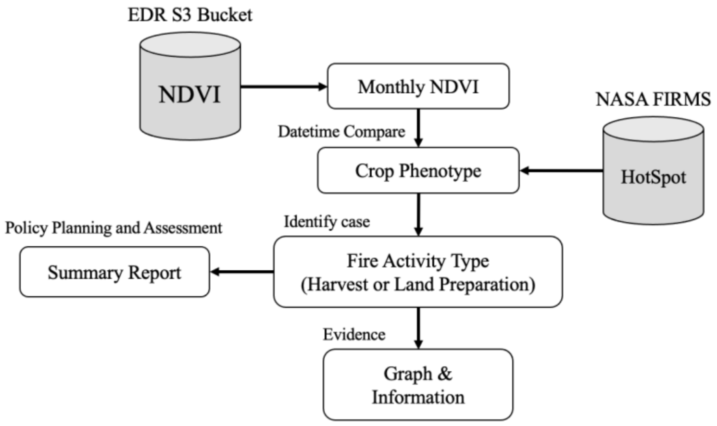


Figure 9 Workflow of fire activity detection

Figure 10 demonstrates the detection evidence of burn for waste clearance type, which fire activity detected after the harvest period. Figure 11 presents the burn for harvest, such as sugarcane, due to its reduced cost from labor. These two cases explain that NDVIs from EDR can be used as farming activity detection and crop growth description ability. In addition, time marks in the NDVI series are crucial for any method that supports the government's crop-related policy response assessment as show in this workflow scenario.

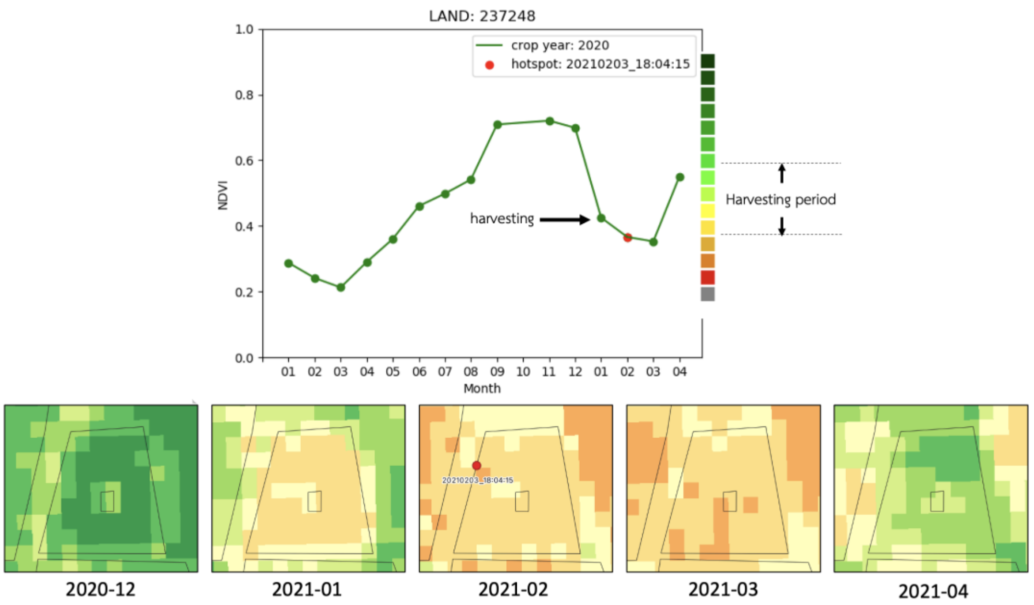


Figure 10 Burn for agricultural waste clearance

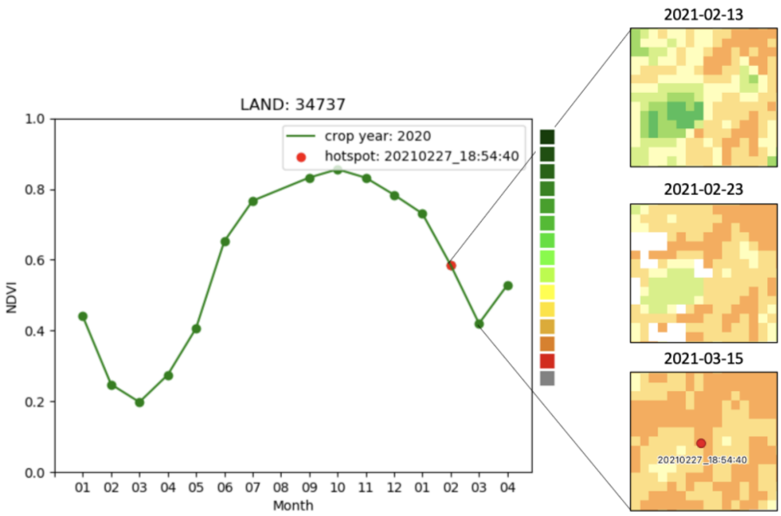


Figure 11 Burn for harvest (sugarcane)

5. CONCLUSION AND FUTURE WORK

Varuna Land Monitoring (VLM) produces a successful high-resolution vegetation index that enables use in large-scale coverage areas but at a precision farming level, as shown in both scenarios. The 20x20 square meters resolution data is a proper resolution for the majority ASEAN farm size. But necessary factors for crop modeling cannot produce in this resolution, for example, soil moisture and temperature from SMAP. The improvement model is required to fulfill the objective of the system. Open-source libraries and software in the VLM platform are proof to be used as the central processing core of an enterprise system. Amazon ES2 and S3 are also good platforms for scalable support of Big data storage and processing resource.

6. REFERENCES

Amazon, 2021. Amazon Elastic Block Store (EBS) Easy to use, high performance block storage at any scale, Retrieved August 12, 2021, from https://aws.amazon.com/ebs

Amazon, 2021. Amazon EC2 Secure and resizable compute capacity to support virtually any workload, Retrieved August 12, 2021, from https://aws.amazon.com/ec2

Amazon, 2021. Amazon S3Object storage built to store and retrieve any amount of data from anywher, Retrieved August 12, 2021, from https://aws.amazon.com/ec2

Das, Narendra N., D. Entekhabi, S. Dunbar, S. Kim, S. Yueh, A. Colliander, T. J. Jackson, P. E. O'Neill, M. Cosh, T. Caldwell, J. Walker, A. Berg, T. Rowlandson, J. Martínez-Fernández, Á. González-Zamora, P. Starks, C. Holifield-Collins, J. Prueger, and E. Lopez-Baeza, (2020). Assessment Report for the L2_SM_SP Version-3 Release Data Products, SMAP Project, JPL D-56549, Jet Propulsion Laboratory, Pasadena, CA, pp. 4-5.

Edric Keighan, Benjamin Pross, Hervé Caumont, (2014). Testbed 10 Performance of OGC Services in the Cloud: The WMS, WMTS, and WPS cases, pp. 15-20.

George H. Hargreaves, F. R. (2003). History and Evaluation of Hargreaves Evapotranspiration Equation. Journal of Irrigation and Drainage Engineering, pp. 53-63.

World Bank, 2021. GDP share of agriculture in ASEAN, Retrieved August 12, 2021, from https://www.theglobaleconomy.com/rankings/share_of_agriculture/ASEAN/