Using machine learning to map Indonesia's restoration potential
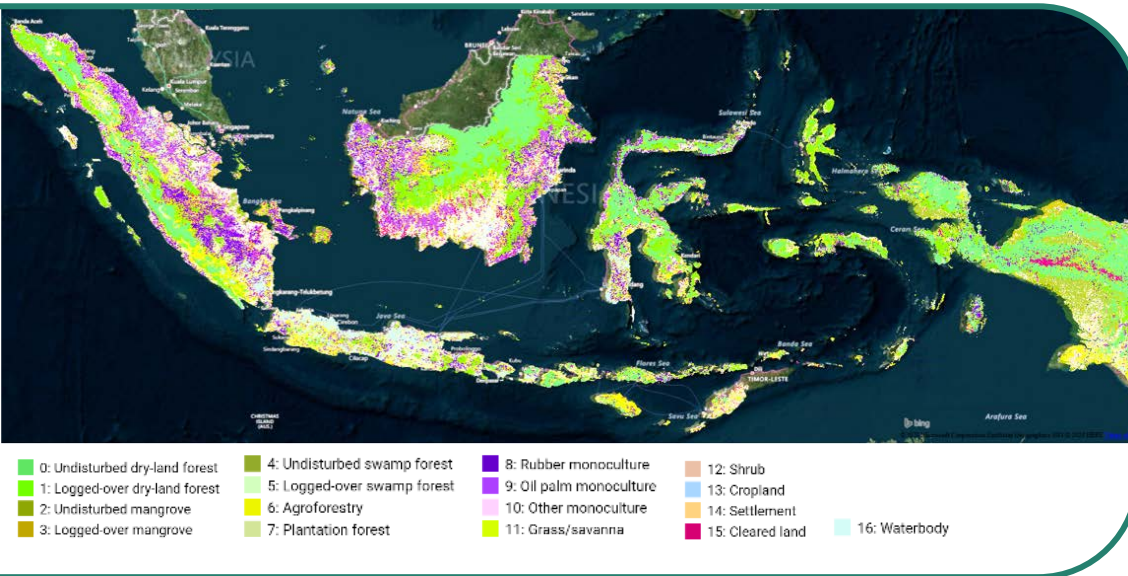
Figure: Restoration potential assessment will be based on the project's experimental countrywide land cover map at 100 m resolution generated by automated supervised classification in Google Earth Engine cloud-based platform.

0: Undisturbed dry-land forest
1: Logged-over dry-land forest
2: Undisturbed mangrove
3: Logged-over mangrove
4: Undisturbed swamp forest
5: Logged-over swamp forest
6: Agroforestry
7: Plantation forest
8: Rubber monoculture
9: Oil palm monoculture
10: Other monoculture
11: Grass/savanna
12: Shrub
13: Cropland
14: Settlement
15: Cleared land
16: Waterbody

› As an initial step for assessing restoration potential, an automated method to analyze countrywide land cover changes based on machine learning and cloud computing technology is being developed

› The development includes series of expert workshops to co-design mapping algorithm and to collect reference data for accuracy assessment

› National-scale crowdsourcing campaign was also conducted to collect a large, geographically diverse training dataset

Background and objective

Towards the national-scale assessment of restoration potential in Indonesia, a countrywide, 100-m resolution land cover map for the year 2018 is being produced (see Figure a). The 2018 map, together with the available reference land cover map for 2010, will allow the assessment of restoration potential, based on changes in land cover between the years 2010 and 2018.

Approach

An automated land cover mapping methodology based on machine learning and cloud computing is developed in the Google Earth Engine (GEE) Javascript API. The method benefits from the open-source platform, transparency, and reproducibility, thus allowing for generation of mapping products that are temporally consistent in a timely manner. As input, a combination of medium resolution optical and radar satellite data was used namely Landsat, PALSAR-2, and Sentinel-1.

In addition, auxiliary maps were used, including elevation (SRTM) and distance to roads/settlement/rivers. For the supervised classification of land cover, the mature Random Forest algorithm (a form of decision tree classifier which operates in an ensemble) was employed. As an initial reference dataset to train the classification model, stratified random samples obtained from an available reference map for 2010 were used. The training was conducted for seven broad regions of Indonesia.

To ensure national relevance, a participatory mapping approach was adopted. Stakeholders from national and sub-national government offices, civil society organizations and universities were involved throughout the whole processes of co-designing the mapping methodology. A series of expert workshops were conducted to build a common definition of land cover typology, collect reference data for map accuracy assessment using the expert tools, and document expert decision rules in interpreting land cover classes.

In addition, training data that are representative of the vast environmental and land management differences in Indonesia were required to train the Random Forest model. In order to collect those data as well as engaging the general public, a nationwide crowdsourcing campaign was conducted via the Urundata platform (https://urundata.id/) and mobile application (Urundata in google playstore: https://bit.ly/UnduhUrundata) (see Figure). More than 80.000 very high-resolution images were interpreted into seven land cover classes namely undisturbed forest, logged-over forest, oil palm monoculture, other tree-based systems not oil palm, cropland, shrub, and grass/savanna.

Interim results and next steps

Several iterations of the national-scale land cover maps for 2018 have been created. The reference data collection for detailed classes using the expert tools Geo-Wiki and a custom GEE application for accuracy assessment of the map will be continued in the next expert workshops.

The collection of these data is expected to be completed in 2020. As next steps, the mapping algorithm will be refined with additional expert rules, ancillary data, training data from the national crowdsourcing campaign, additional covariates (e.g. tree canopy cover and height), and object-based spatial filtering.