

## An Automatic Feature Extraction Method of Satellite Multispectral Images for Interpreting Deforestation Effects in Soil Degradation

Irene Erlyn Wina Rachmawan<sup>#1</sup>, Yasushi Kiyoki<sup>#2</sup>, Shiori Sasaki<sup>#3</sup>

<sup>#</sup>Graduate School of Media and Governance, Keio University, Shonan Fujisawa Campus, Endo 5322  
Fujisawa, 252-0882, Japan

<sup>1</sup> irenerl@sfc.keio.ac.jp, <sup>2</sup> kiyoki@sfc.keio.ac.jp, <sup>3</sup> sashiori@sfc.keio.ac.jp,

---

**Abstract**— Deforestation is still a major nature phenomenon in our society. For assessing deforestation effect, satellites remote sensing provides a fundamental data for observation. While new remote-sensing technologies are able to represent high-resolution forest mapping, the application is still limited only for detecting and mapping the deforestation area. In this paper, we proposed a new method for automatically extract features of Satellite Multispectral images for interpreting deforestation effect in the context of soil degradation. We proposed an idea to interpret reflected “substances (material)” of bare soil in deforested area in spectrum domain into human language. The objectives of this paper are to (1) recognize the deforestation activity automatically. (2) Identify deforestation causes and examines the deforestation effect based on deforestation causes. (3) Scrutinize deforestation effects on soil degradation. (4) Representing nature knowledge of deforestation effect in human language using semantic computing, to bring the clear, comprehensible knowledge even for people who are not familiar with forestry. As for the experimental study, Riau Tropical Forest has been selected as the study area, where the multispectral data was acquired by using Landsat 8 Satellite between 2013 and 2014; Where forest fire and logging activities are reported and detected.

**Keywords**— deforestation; soil; degradation; feature extraction; remote-sensing; semantic analysis

---

### I. INTRODUCTION

Deforestation is a process of clearing and changing the main function of the forest. During the massive activities on deforestation, it has been one of the major natural damage on earth. Forest has never been stay in one condition. Not only it is living, but the changing in short and long term has been detected both caused by nature and human activity [1]. Deforestation activities come in many ways, including forest fire, logging, and mining. These causes of deforestation are relatively bringing different effect on nature. Often, in Tropical Countries, the biggest deforestation activity is the conversion of forest for raising livestock by firing, as the fire was the first great force of man development and it is crucial in the story of deforestation [2], in farmer perspective, fire clears the forest in fast and economically way. Despite that, a fire significantly affects soil properties because organic matter located near the soil surface is rapidly destroyed. The changes of Organic Matters, in turn, could affect several chemicals, physical, and microbiological properties of the underlying soil [3]. Way different with the deforestation by logging or cutting trees activity. Tropical countries government has been stated that they fight deforestation activity through environmental legislation or by place concern to automate deforestation detection. Several

researchers have been trying to produce routine estimation about deforestation in the tropical forest. Many countries use satellite imaginary as their basis evaluation. However, their project particularly aims to bring high resolution and accuracy of deforestation area mapping. In order to present deep knowledge about deforestation, it is more important to analyze the effect of deforestation in the environmental change. And to give the brief analysis of deforestation effect, it is necessary to identify the activity that causes deforestation happens to indicate the effect of deforestation afterward.

Common methods of deforestation are burning trees and clear cutting [11]. These activities leave land completely barren and causing many negative effects. Clear cutting—the process of cutting most of the trees all at once—is a type of logging activities that has been considered as controversial methods because of its effect on the landscape. Burning trees—Process to clear the forest by using fire as a tool—is an activity that commonly used by the farmer because it can run in short time. The ash from burned tree has fertilized the soil in short period of time. These two activities are commonly found as the driver of deforestation in many tropical forest nations. The long-term effect of deforestation on the soil condition can be critical. Clearing the vegetation cover for logging and burning exposes the soil to the strong

intensity of tropical sunlight. Forest soils are usually poor of nutrient, because of the nutrients are stored in the tree stem. Making the removal of tree leave the soil lost its ability to store carbon, the house of plant, and supporting leaving a place for plant. The most research on detected effect regarding soil properties caused by burning shown the increase of pH value and nutrient availability in short-term after the burning occurs, but some parameters such as pH will be changing drastically after several months [14]. In contrast, in the case of clear cutting, pH value seems to be decreased, and nutrients capacities are lower than the natural range [15].

Indonesia is one of the tropical countries where owns a large-scale area of tropical forest, but nowadays, Indonesia also contributes to the massive effect of climate change because Indonesia reported being the second most deforested country in the world. Not only the total loss is reaching to 51 percent in lowland forest, where the forest type for the most prone to deforestation locate but also in wetlands, forest loss is exponentially increasing at 43 percent of total loss Not only the total loss is reaching to 51 percent in lowland forest, where the forest type for the most prone to deforestation locate but also in wetlands, forest loss is exponentially increased at 43 percent of total loss overall [16].

Indonesian government using both manual and remote sensing methods for detecting deforestation and hotspot. But other than the mapping area of current forest area and hotspot data, there is no complete public information about deforestation effect in Indonesia. The common media used are maps which presenting the area by indicating some specific features equally to only labeling deforestation and forestation area and not further information about the damaged level of deforestation. Indonesia current land cover data and land use maps of Indonesia are mad via interpretation methods(MoF 2011). Forest are categorised into two broad categories: primary and secondary forest. Primary forest presents the undisturbed intact forest, while the secondary forest is representing an area of forest affected by commercial logging.

Deforestation rate increases each year, the present works have been dealt with mapping and measuring the rate over time and place. But the effect of deforestation was not well reported. To present deep knowledge about deforestation, after gaining knowledge about deforestation rate, it is necessary to understanding the deforestation impact on actual condition. Therefore in our study, we propose a new system for world reporting system of tropical deforestation effect in order to track the nature impact of deforestation activity by using remote sensing and multispectral images analysis.

In order to test the effect of clear-cutting and forest fire effect described above we compared two different forests area and put the investigation in 20 plots in Riau as representative of different effect clear cutting and forest burning: unlogged old-growth, selectively logged and regenerated following clear cutting. Satellite data were collected in between of 2013 and 2014, two years range of detected deforestation activity. From these data, we derived 3 distinct indices encompassing with soil quality, which we proceed for mining new knowledge of the impact of deforestation activity. We focused on deforestation impact

caused by fire and logging on soil condition by using multispectral reflectance to measure the degree of the damage on deforested area and create the analysis to detect the cause of deforestation automatically. This study aims to: (1) Recognize the deforestation activity automatically. (2) Identity deforestation causes and examines the deforestation effect based on deforestation causes. (3) Scrutinize deforestation effects on soil degradation, carbon stock, and biodiversity. (4) Representing nature knowledge of deforestation effect in human language using semantic computing, to bring the clear, comprehensible knowledge for everybody even people who are not familiar with forestry.

Deforestation is one of the main issues in environmental changes field. Various modelling methods to detect deforestation from satellite images had been proposed by many researchers. C.J Tucker[4] proposed wall-to-wall covered to estimate deforestation rate by using sampling image data from Landsat Image. Marcio Puppim M[5] publish new methods called STARS to detect deforestation by using Modis Satellite, where the study area was located in Brazil. Their proposed idea was successfully implemented to detect deforestation area by distinguishing the spectra value. They use MCI as input attribute and calculate the spectra value to classify the different area. Another researcher has done more specific remote sensing analysis, Badamasi [6] use NDVI threshold classification to calculate the satellite images digital number value. They use NDVI to detect the vegetation cover in Nigeria then analyze the classification result in temporal time. Curtis A. Colling [7] presents a multi-temporal analysis of Landsat data to determine Forest Age Classes by using vegetation indices. As for interpreting the knowledge Yasushi Kiyoki *et al.* [10] proposed a new semantic computing method with multi-spectral images for analyzing and interpreting environmental phenomena and changes occurring in the physical world.

Analyzing and mining information in the temporal dataset and in spatial dataset presented in two separate streams of research, where the current works of spatiotemporal data analysis work by gathering information over places and time. Spatial data represent a unique attribute in that “everything is related to everything else, but nearby things are more related than distant things.” This is also known as Tobler’s first law of geography. Thereupon, the main objective of spatial data analysis is to discover the area by using proximity relationship. Many approaches have been proposed for mining the temporal association patterns, such as cyclic association rules, periodic association rules, and calendric association rules. Chen and Petunias present a framework for mining temporal association patterns. As for spatiotemporal data analysis, it is firstly done by Stolorz, Nakamura, Muntz and provides queries for geographical scientific for a pattern such as cyclones, hurricanes, and fronts [17]. The next researcher works on mining interesting spatiotemporal pattern in earth science data, where they apply existing data mining techniques to find clusters, to analyze the difference of spatial data. Setia Darmawan presenting an identification of deforestation area using MODIS EVI 250 m in 2000 and 2012 to identify deforestation rate in Java island. MODIS EVI is one of a kind MODIS images which is able to detect vegetation based on photosynthesis rate and vegetation density using Fuzzy

C-Means with 13 numbers of clusters [18]. Xuejun Wang also does research to identifying deforestation using Landsat data using two methods; one is the simple linear regression model (SLR).

Some researchers are interested in mining Spatio-temporal patterns in earth science data. Many approaches have been proposed for mining the temporal association patterns, such as cyclic association rules, periodic association rules, and calendric association rules. C.Immaculate Mary and dr. S.V. Kasmir Raja [19] presented ant colony as optimization for K-Means. Stolorz et. al. [20] presented spatiotemporal data analysis and provided queries for geographical patterns such as cyclones, hurricanes, and fronts, did research to identify deforestation using Landsat data using two methods which are simple linear regression model and curve model. Pickup and Foran [21] developed a method to monitor landscapes for pastoralism based on the spatial variability of the vegetation. The spatial autocorrelation function and mean-variance plots of a spectral indicator were found to be successful in discriminating between the cover responses for common annual rainfall rate. In dry condition, the rapid changing of increasing spatial lag will be followed by decreasing value of spatial correlation since the ground surface is bare and most of the vegetation signals come from scattered areas of trees and shrubs. A low decay rate of the autocorrelation function indicated a greater spatial uniformity of the landscape, e.g. during wet periods, when more ground cover, reduces the contrast between the bare soil signal and others produced by trees and shrubs. Similar observations were

made by Lambin over the seasonal and inter-annual cycle of three West African landscapes. Vogt [22] also analyzed the seasonal changes in spatial structure of a West African vegetational landscape, showing that there is a marked seasonal cycle in the spatial structure of a vegetation index (NDVI), and that area of transition during the seasonal dynamic had been; however, the observation of this variability of spatial, only provided for a qualitative description of the cover state.

## II. MATERIALS AND METHODS

Our proposed idea consists of four processes: (A) Recognize the deforestation activity automatically. (B) Identify deforestation causes and examines the deforestation effect based on deforestation causes. (C) Scrutinize deforestation effects on soil degradation (D) Representing nature knowledge of deforestation effect in human language using semantic computing.

### A. Recognizing the Deforestation Activity Automatically

Following steps are applied in order to detect deforestation area automatically: (1) Data Acquisition and preprocessing of Landsat 8, (2) Multispectral image conversion from Digital Number (DN) to Reflectance Conversion, (3) Automatic deforestation area detection by using Automatic Clustering. Steps of our proposed idea are represented in Fig. 2 where the



Fig. 1 Forest Land cover zone in Riau, Forest area are indicated by green color divided into six types of forest: HK, HL, HPT, HP, HPK, AL. (Data source: Ministry of Forestry 2012)

TABLE I  
INDONESIAN FOREST LAND USE ZONES: FOREST CLASSES, FUNCTION, POSSIBLE MANAGEMENT PRACTICES AND CONSEQUENCES OF EACH CLASS

Code	Forest Land Use	Function	Possible management Practices	Consequences
<b>HK</b>	Conservative Forest	Preserving the biodiversity of flora fauna and their ecosystem		To understand the architecture of Cluster Computer.
<b>HL</b>	Protected Forest	Protecting the water system to prevent flooding, control erosion, protect sea water intrusion and maintain soil fertility	Forest protection Stable forests without any	Forest protection Stable forests without any
<b>HPT</b>	Production Forest	Providing the forest products mainly from timber extraction	Forest production	Dynamic deforestation and
<b>HP</b>	Limited Production Forest	Low-intensity logging (due to topographical condition)	Limited logging Very selective logging Very limited clear-cutting Post-logging silvicultural treatments	Forest degradation
<b>HPK</b>	Regular Productive Forest	Logging Forest plantations	Selective logging Post-logging silvicultural treatments Clear cutting	Temporary deforestation Forest degradation
<b>APL</b>	Convertible Productive Forest	Logging	Logging Agriculture Other uses	Permanent and temporary Deforestation Forest degradation

1) *Step 1. Data Acquisition and Preprocessing*

Landsat 8 OLI of Riau is downloaded for data sources in this study. After choosing Landsat OLI (resolution: 30 m, see Table 2 for the spectral characteristic of Landsat 8), the specific area then selected as the observation area. In the next step, the data were reregistered with USGS National Map accuracy standard. The image acquisition dates were selected between 2013 and 2014. Where in 2013, the satellite images capture the earth surface before forest fire occur, and in June 2014 where the forest fire happened at the beginning of March and ended at the end of March 2014.

Along with satellite images, Land use data from local government is collected to merge and give precise deforestation mapping. Land use data for this case study are acquired from Ministry of Forestry for Riau Land use mapping in 2012 (Fig. 1). Forest are divided into six categories, Conservation forest, Protected Forest, Production Forest, Limited Production Forest, Regular Productive

Forest. Deforestation activities those being observed in this study are located inside of the limited production Forest. Satellite data were logged in 8 months interval from April to December. Logged data were downloaded from Landsat data, we used seven bands of Landsat images, to classify and analyze deforestation effect. We excluded the data from the stations whose data were missing for more than 21 days. As a result, data of two places, which are situated on the side of the river and in middle of hills, were chosen.

Short-term missing data were replaced by interpolating with a moving average method. We assumed that deforestation activity caused by fire brings the strong effect on the natural object, such as soil moisture, while illegal logging activity may affect differently. The study area is situated on the south of river Siak, where forest fire occurs in March 2014 and causes a big degradation of forest area. The numeric data of fire node occurs in Riau are used as the training dataset in this experiment.

TABLE II  
SUMMARY OF SPECTRAL CHARACTERISTIC OF LANDSAT 8

Waveband Name	Band Number	Wavelengths
Coastal	1	0.433–0.453
Blue	2	0.450–0.515
Green	3	0.525–0.600
Red	4	0.630–0.680
NIR	5	0.845–0.885
SWIR	6	1.560–1.660
SWIR	7	2.100–2.300
Panchromatic	8	0.500–0.680
Cirrus	9	1.360–1.390
Long Wavelength IR	10	10.6–11.2
Long Wavelength IR	11	11.5–12.5

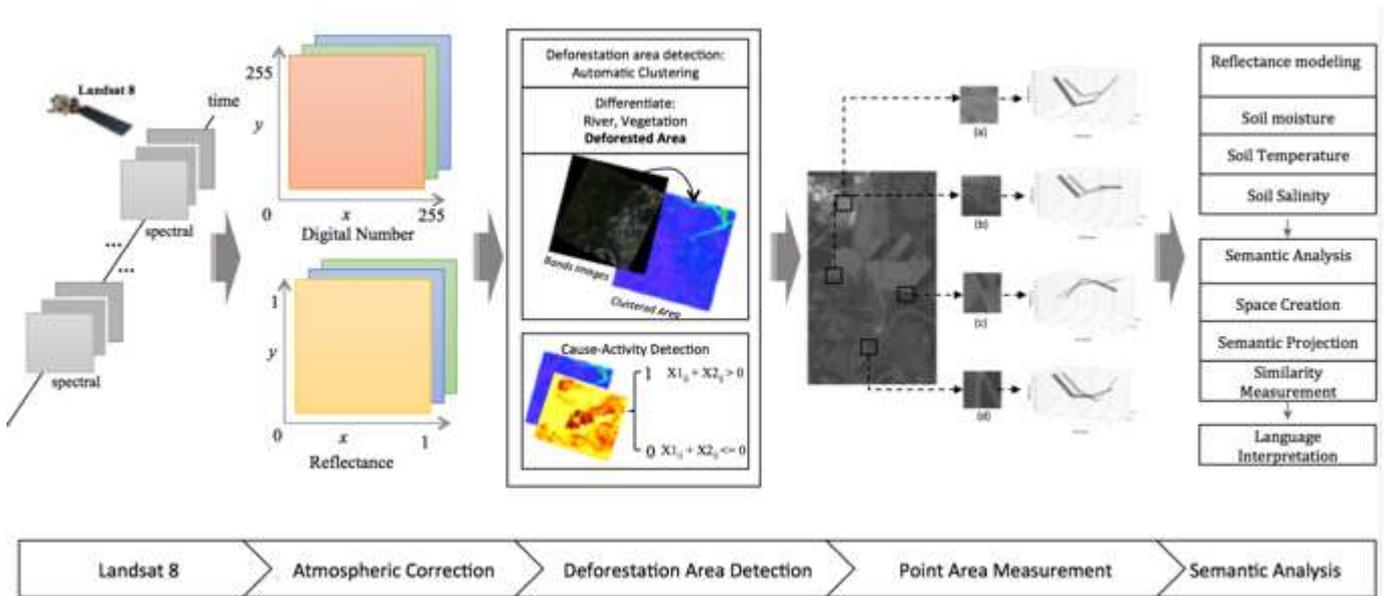


Fig. 2 System design and proposal, processing system are including acquire data, DN to Reflectance Conversion, Deforestation Area detection and automatic cause-activity detection

### 2) Step 2. Digital Number (DN) to Reflectance Conversion

OLI band data is served in Digital Number Value, which can be converted to Reflectance value to describe nature object property. Reflectance is measurement number to represent the object which been lighted by the sun. Its value describes the number, which further use to differ plant, water, soil, and any other object in earth surface. To convert DN data to reflectance conversion, the standard equation for rescaling coefficients listed in the metadata file used and inserted in equation (1) called Top of Atmosphere reflectance (TOA) is applied to this study.

$$\rho\lambda = M_p * Q_{cal} * A_p / \sin \theta_{SE} \quad (1)$$

Where:

$\rho\lambda$  = TOA reflectance, ranges from 0 to 1.

$M_p$  = Multiplicative rescaling factor from the metadata.

$A_p$  = Band-specific additive rescaling factor from the metadata.

$Q_{cal}$  = Quantized and calibrated standard product pixel values (DN).

$\theta_{SE}$  = Local sun elevation angle.

While Digital Numeric range from 0 to 255 and indicate the color difference in each nature object, the Reflectance value range from 0 to 1, which indicate the reflected and absorbed the wavelength of nature object.

### 3) Step 3. Deforestation Area Detection

Automatic Clustering and K-Means are applied in order to detect the deforestation area automatically. Clustering process is a step to classify the area of deforestation to get the information from its clusters. Due to the amount of the area that will be clustered is unknown; it is difficult to cluster the area with usual clustering techniques, which requires the exact number of the cluster should be known.

Therefore, the automatic clustering process will be used for this process where the variance will be calculated and used as the measurement to get the optimum number of cluster. This cluster variance of each moving variance is designed by the variance within the cluster, and it judges having reached the global optimal solution based on this tendency. From the cluster variance, finding the ideal cluster is very difficult because the minimum variance cannot directly be applied to find the global optimum as an ideal cluster. For finding the global optimum of cluster construction and avoiding the local optima, we apply our Valley Tracing method to find the global optimum. First of all, patterns of the moving variance should be described and analyzed the possibility of the global optimum that resides in the valley of patterns. Automatic Clustering is classified as hierarchal clustering methods, where the algorithm is choosing a pair of the closest cluster to merge by considering the optimum value of fitness value or variance (equation 2). It detects the optimum cluster number by analyzing the moving variance.

$$V = (V_w/V_b) * 100\% \quad (2)$$

Where:

V = Global Variance

$V_w$  = Variance within cluster

$V_b$  = Variance Between cluster

The shape of observed moving variance pattern indicated the optimum cluster number. Fig. 2 represents how optimum cluster number is decided.

Pattern	Possible?	Pattern	Possible?
	✓		✗
	✓		✗
	✓		✗
	✗		✓

Fig. 3 Optimum pattern of observed moving variance

### B. Detect the Type Activity that Causes Deforestation

One of the important proposal in our paper is to identify type activity that causes deforestation. Forest fire and Logging are the common activity of deforestation. These two causes, resulting in different behaviour of the after effect from deforested land. Our main idea is to distinguish these differences by following rules: (1) Identify burn scars by using Normalized Burn Ratio (NBR) [12], (2) Eliminate area with burn scar outside of desired time interval by using differential equation of NBR, (3) Identification of the deforestation activity.

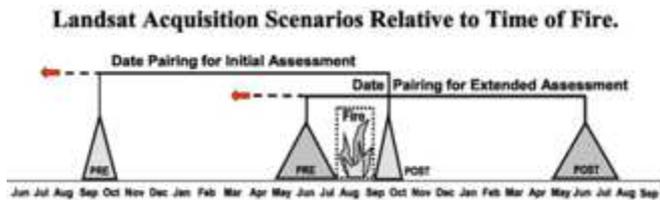


Fig. 4 Pre-fire and Post-fire imagery data analyzing the process for assessment. The different timing will affect significantly the NBR Measurement [12]

#### 1) Step 1. Identify burn scars by using Normalized Burn Ratio

The original step of NBR has calculated the two Landsat bands that respond to heat and vegetation the most, but in opposite ways to burning.

$$NBR = (NIR - SWIR) / (NIR + SWIR) \quad (3)$$

Where:

- $NBR$  = Normalized Burn Ratio
- $NIR$  = Near Infrared Reflectance
- $SWIR$  = Shortwave Infrared Reflectance

Those were determined to be TM/ETM+ Landsat 8 in Band 4 and Band 7. The NBR is indices calculated as shown in equation 3. Where  $NIR$  and  $SWIR$  are the reflectance value calculated from reflectance values per band, which have been converted from digital number to reflectance. The combination of those two bands appears to provide the significant distinction between burned and unburned areas. The (*Reflectance of Band 4 - Reflectance of Band 7*) difference is scaled by the sum of the two bands to normalize for overall brightness that is consistent across the bands. It helps remove within-scene topographic effects and between-scene solar illumination effects. This effectively isolates the real reflective differences between the bands, which enables spatial and multi-temporal comparison of the derived NBR values.

#### 2) Step 2 Identify Deforestation Activity Area with Burn Scar Outside of Desire Time Interval

To distinguish burned and unburned area eliminates area out of time scope by using the differential equation of NBR as shown in Equation 4.

$$dNBR = NBR_{prefire} - NBR_{postfire} \quad (4)$$

Where:

- $dNBR$  = differential of Normalized Burn Ratio
- $NBR_{prefire}$  = NBR value before fire
- $NBR_{postfire}$  = NBR value after fire

This process, measure the change of  $NBR$  value and hypothesize the deforestation area to be caused by fire. Assuming unburned forest is relatively similar in moisture between the two sample dates, and the two datasets are adequately reregistered, background areas take on values near zero in  $dNBR$ . Likewise, burned areas assume strongly positive or negative values, depending on whether the fire distresses or actually enhances productivity on the site. Standard to determine the severity of burn area is shown in Fig. 2. This burn area measures by the absolute change in NBR and scaled by  $10^3$ . To accurately identify the effects, soon after forest fire occurs and before the next growing

TABLE III  
ORDINAL SEVERITY LEVELS OF FOREST FIRE

season is the best time to apply the NBR.

Severity level	dNBR range
Enhanced regrowth, high	-500 to -251
Enhanced regrowth, low	-250 to -101
Unburned	-100 to +99
Low severity	+100 to +269
Moderate-low severity	+270 to +439
Moderate-high severity	+440 to +659
High severity	+660 to +1300

#### 3) Step 3. Identification of the Deforestation Activity

After applying the burn index, the next step is to gain knowledge and labelling the deforested point caused by clear-cutting or burning trees; the proposed mathematical model is expressed in equation 5.

$$Dx = \begin{cases} 1, & x1_{ij} - x2_{ij} \geq 100 \\ 0, & x1_{ij} - x2_{ij} < -100 \end{cases} \quad (5)$$

Where:

- $Dx$  = Category of deforestation activity
- $x1$  = Deforestation Area
- $x2$  = Burned Area

According to equation 5, by measuring the NBR value, we could put the hypothesis on the cause of deforestation. Assuming unburned deforested place is relatively similar in physically, and moisture between the two sample dates, background soil take on values near zero in  $dNBR$ . Otherwise, burned deforested areas assume strongly positive or negative values. Therefore, we could simply put threshold

$\epsilon$  to determine the cause of deforestation. As such, an initial assessment can yield figures on burn size, composition, and complexity within a month or two of the fire. Another action may consider as a factor for many needs concerning public information, planning, and rehabilitation.

C. Parameters for Measuring the Degree of Deforestation Effect by considering Soil Property

Three values of soil properties are importance for measuring the degree of deforestation effect for soil [5].

1) Soil Moisture and Texture

The first important dimension to measure the soil condition is the water held by soil, or the moisture of the soil. Soil Moisture analysis could be computed by using orthogonal indices, the distance value of any soil point in NIR-Red Reflectance to Line L determines the soil moisture condition. The closer the distance, the more moisture soil condition is and *vice versa*. The Reflectance value could be determined as seen Fig. 5.

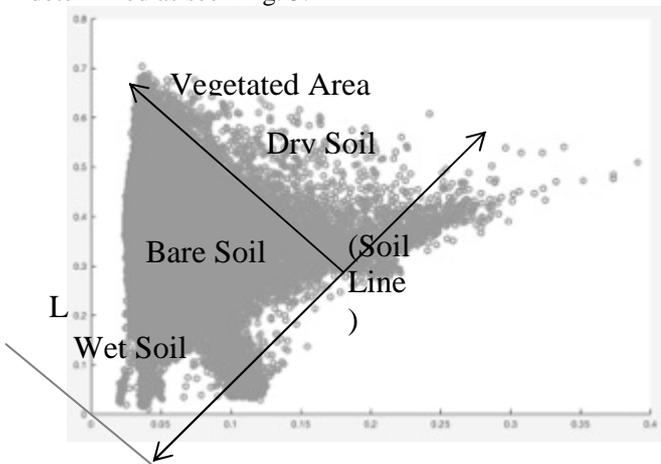


Fig. 5 NIR and red space in Study Area on 2014 dataset

In common condition, the point of nature object placed close to L site is always wet region and soil moisture indices defined number close to, whereas, in the furthest region, soil moisture is close to 0 and determined as an extremely dry region.

$$\text{Soil Line} = (1 / (M^2 + 1)) * (R_{\text{nir}} + MR_{\text{red}}) \quad (6)$$

In the case of soil texture analysis, the most common way to determine the soil texture is by distinguishing the textural classes based on soil texture triangle. The classification is depending on the percent amount of clay, sandy, or silt type.

2) Soil Temperature

In order to represent the soil condition the after deforestation effect, surface temperature of deforested area also needs to be measured. Standard Landsat 8 product acquired not only by the Operational Land Imager but also by Thermal Infrared Sensor. Equation (7) shows how to convert the Reflectance number of TIRS band (Band 10 in Landsat 8) into Kelvin scale [11].

$$T = K_2 / \ln( (K_1 / L_\lambda ) + 1) - 273.15 \quad (7)$$

Where:

$$T = \text{At-satellite brightness temperature (C)}$$

- $L_\lambda$  = TOA spectral radiance (Watts/( m2 \* srad \*  $\mu\text{m}$ ))
- $K_1$  = Band-specific thermal conversion constant from the metadata
- $K_2$  = Band-specific thermal conversion constant from the metadata

3) Soil Salinity Analysis

Soil Salinity analysis is needed to measure the saline level of soil; it is often linked with sustained wetness, therefore increases the vulnerability of soils to erosion. Soil salinity analysis could be calculated by using Equation 8

$$SI = (NIR * R) / G \quad (8)$$

Where:

- SI = Salinity Index
- NIR = Near Infrared Reflectance
- R = Red Reflectance
- G = Green Reflectance

D. Semantic Analysis of Soils that Affect Post Deforestation Management

In order to represent knowledge using semantic analysis, Following steps are determined to analyze deforestation activity.

1) Semantic Space Creation

First, we construct the correlation matrix with respect to the feature. Define the space  $I$  with the  $m$  words are given and the each word is characterized by  $n$  features ( $f_1, f_2, f_3$ ).  $m$  words are words determined context in deforestation assessment, assessment in plant growth and drought are selected in this research. Meanwhile,  $f$  is the feature value of semantic space as describe in in part C (Soil Moisture, soil temperature, and soil salinity). Fig. 6. Shows the representation of metadata item by matrix  $M$ :

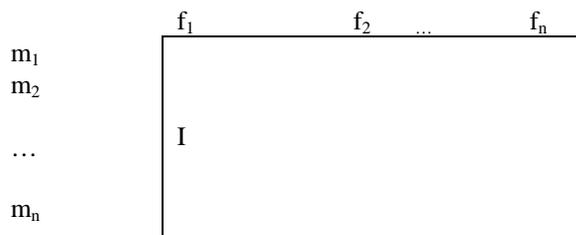


Fig. 6 Representation of metadata items by matrix  $m$

2) Semantic Projection

Semantic Projection is the node to measure the correlation between the given words with the semantic space. All the projections are considered from the space  $I$  and map to the invariant subspaces. That is, this model can express  $2^v$  different phases of meaning, where  $v$  is the number of  $i$  dimensional invariant subspace.

3) Semantic Similarity Measurement

In order to represent deforestation area semantically, the inputted data item with the nearest meaning of value calculation with the semantic projection. (1) Map the point of interest of deforestation area in space  $I$ . (2) sum up values

for each semantic element. Since  $v$  semantic elements existed, we can construct a  $v$  dimensional vector. (3) If the sum obtained in (2) is less than a given threshold  $\epsilon$ , the point of interest will be employed to thus semantic element.

### III. RESULTS AND DISCUSSION

Our proposed idea is to interpret the effect of soil condition in deforestation activity in language interpretation by using semantic analysis by using soil moisture and texture, soil temperature, and soil salinity. After converting multispectral images from DN to Reflectance, and process the language interpretation and semantic analysis, the semantic matrices for deforestation is produced.

#### A. Study Area and Acquired Data

The study area was located at Riau, Indonesia, between 101.79 ~ 102.89 E and 0.23 ~ 0.24 S. One site (1000 m above the sea level) represents a hardwood forest dominated by *Quercus mongolica* and *Acer pseudo-sieboldianum* (Lee et al., 1999). The other (800 m above the sea level) lies within a bare soil area where vegetation and surface litter were cleared. Two cloud free multispectral satellite images between 2013 and 2014 have been selected as a study area, considering the major differential changes caused by a forest fire in March 2014. Another activity of deforestation such as illegal logging also occurs in this years. The experiment will examine the nature effect after two different type of deforestation occurs.

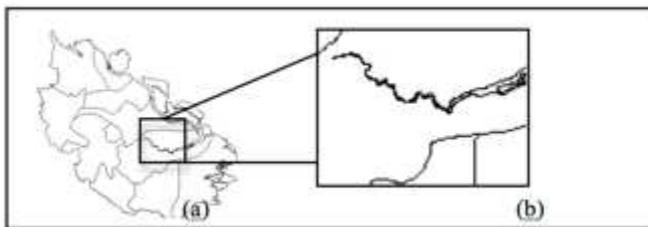


Fig. 7 Study area and acquired data

To bring the limitation of observation object on deforestation area, we assumed that deforestation activity caused by fire brings the strong effect on the natural object, such as soil moisture. The deforestation area itself is supposed to be distinctly different between soil and another substance (known as bare soil), as deforestation is likely to change the forest cover into another usage, while illegal logging activity may affect differently. The study area is situated on the south of river Siak, where forest fire occurs in March 2014 and causes a big degradation of forest area. The numeric data of fire node occurs in Riau are used as the training dataset in this experiment.

#### B. Experimental Results and Discussion

##### 1) Automatic Detection of Deforestation

Automatic detection of deforestation is extremely important for maintaining the information about deforestation. In this experimental study, we apply automatic clustering to generate the automatic number of the cluster could be implemented, after implementing the automatic clustering landscape change detection are applied and remap the clustered data.

In the first step clustering process, Automatic Clustering runs right before K-Means clustering method. The main goal of using Automatic Clustering is to assign the accurate number of cluster. If Automatic clustering is not used, the cluster number will be assigned a static value and leads to misshaping of the cluster member. For examples, if there are 4 distinct areas (in 2004), then the initial number of cluster ( $k$ ) for K-Means is 4. But, if the area is less or more than 4 (in another year), the number of initial  $k$  is changed. So it will be complicated to handle if the number of the cluster in every year changes differently. Therefore, this is the main reason why Automatic Clustering is needed. The Automatic Clustering is to identify the most optimum number of cluster every year. The steps for Automatic Clustering is as the following:

- Resize the multiband images to maximum 30 x 30. This step is crucial because Automatic Clustering is a type of hierarchal clustering methods of which they can not handle big data. The maximum arrays that can be handled by hierarchal clustering methods are 1000 arrays. So the images are resized from the satellite images into compressed the dataset.
- Collect grayscale values of each pixel in every multiband image at one interval time and store them in the database. Because multiband satellite images serve in grayscale format, the step is only to get the value from one element RGB. This step produces a new dataset containing the grayscale value of each multiband images.
- Mention every node as a single cluster. So, there will be 900 single clusters. Then, iterate to move to the nearest two clusters to be one cluster. This step is repeated until the optimum cluster number is found.
- Measure the Variance values from each cluster. Then, do Valley Tracing in Eq. 2 to produce the optimum number of clusters.

The actual optimum number of cluster is to determine by Eq. 2 of which the value of  $\theta$  is the global optimum of cluster number. After the implementation of the steps, the result is used as the K parameter in K-Means for clustering multispectral images. As the result of our experimental study to detecting deforestation activity, Fig. 8 shows the changed area in Riau between 2013 and 2014.

After Automatic clustering has been successfully implemented to detect the deforestation area, Automatic image clustering is the automatic process of dividing an image into appropriate regions that identify specific areas within the scene. The total number of cluster produced by Automatic Clustering is added as the cluster number parameters in K-Means process. The pixel size of multispectral images is compressed before using Automatic Clustering to achieve a good computational time. The pixel size of this images resized to 30 x 30 pixel because automatic clustering can produce the same number of a cluster even using original data and resized data. It is because Automatic Clustering has an analysis to recognize the pattern of an optimum number of a cluster using valley-tracing methods. The original data is resized from 705 x 705 pixel (497025 data) to 30 x 30 (900 data). This size is determined by experiment of optimum small pixel number of

satellite images that provide the same cluster number with the original dataset. There are three different states are mapped into a newly produced map in this area: yellow color identify deforestation activity, blue color identify the new growth of plant has been detected, and green area showing there is no change happen between 2013 to 2014. After detecting deforestation area, we aim to differentiate between burning-trees activity or clear-cutting activities. Therefore we are sampling on two more specified area (a) and (b).

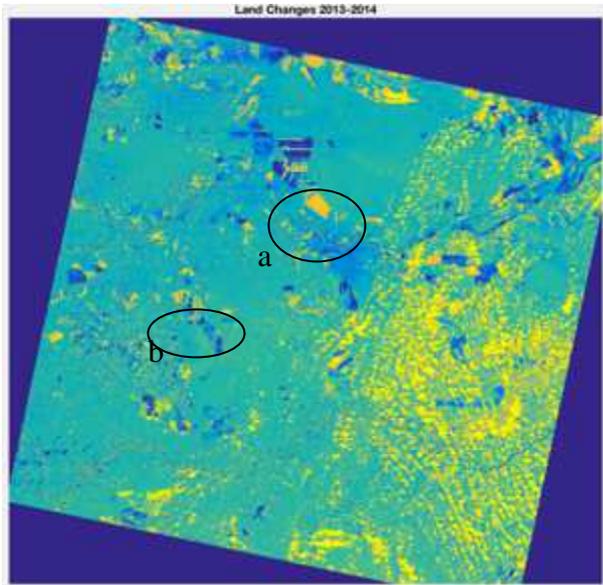


Fig. 8. Automatic detection of deforestation of Riau 2014: Deforested Area indicated by yellow colour

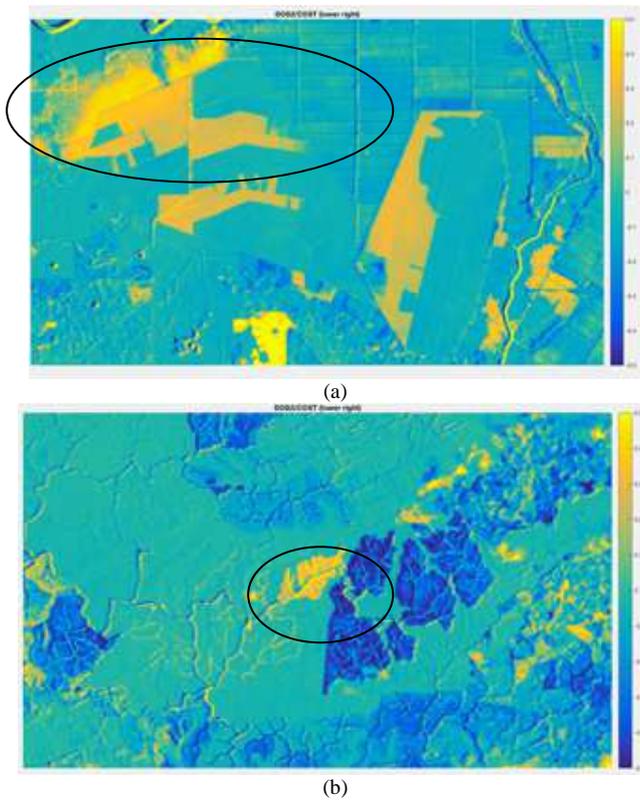


Fig. 9 Two specific deforestation area was chosen, area (a) is a deforested area by burning-trees and area (b) is a deforested are by clear-cutting

Two specific deforestation places were manually chosen in this study by checking the ground truth data of forest fire and detected hotspot by Lapan (Indonesia aerospace agency). In Fig. 9, yellow color indicating changing of the landscape occurs because of deforestation or clearing of the forest. The green color in showing unchanged or a very low landscape change from 2013 to 2014, while the blue color indicates the growing plant or the existence of a plant in 2013 where was not exist in 2013. As for the future comparison, from here, two spots of deforestation was selected. Where (a) is the deforestation place of forest fire and (b) are deforestation of logging or with a different type of cause. The value of burning-trees and clear-cutting are not shown clearly in this phase.

## 2) Detect the Activity Type that Causing Deforestation

Deforestation occurs in two different ways, forest fire or logging, not many information about the cause of deforestation are served. our main proposed idea is to detect the cause of deforestation automatically, In this experimental result, we apply NBR to measure the burning scars of the forest. As mention in eq. (3). This size is determined by experiment of optimum small pixel number of satellite images that provide the same cluster number with the original dataset. There two different states are mapped into a newly produced map in this area: yellow color identifies unburned area, dark red color indicating the possibility of burned area happens between 2013 to 2014. we aim to differentiate between burning-trees activity or clear-cutting activities. Therefore we are sampling on two more specified area (a) and (b).

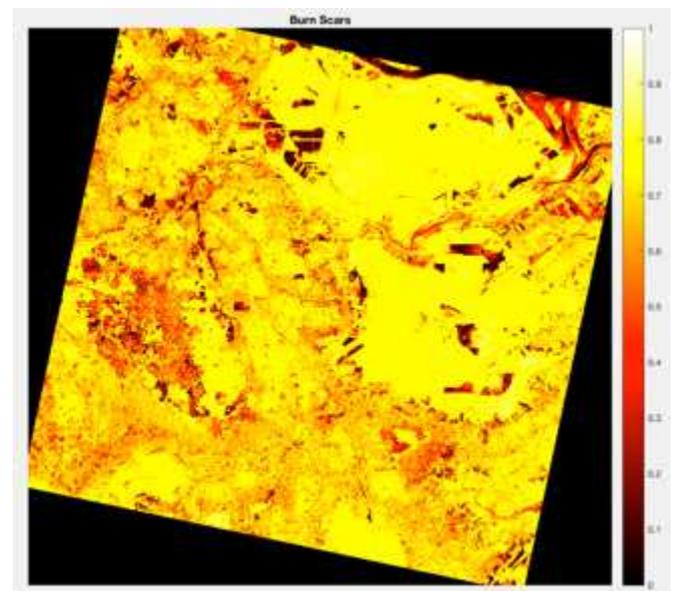


Fig. 10 Normalized burning ratio value of Riau in 2014. Lower value indicates the burned area while higher value indicates unburned area

The experimental result of NBR index for determines the deforestation activity in Riau is shown in Fig. 10. It shows the total NBR value analysis in Riau due to a forest fire in 2014. The darker area indicates the high burn of soil, the NBR showing darken colour in place of bare soil or deforested area. But the value of deforestation activity

caused by burning tree and clear cutting has a slightly different value of NBR.

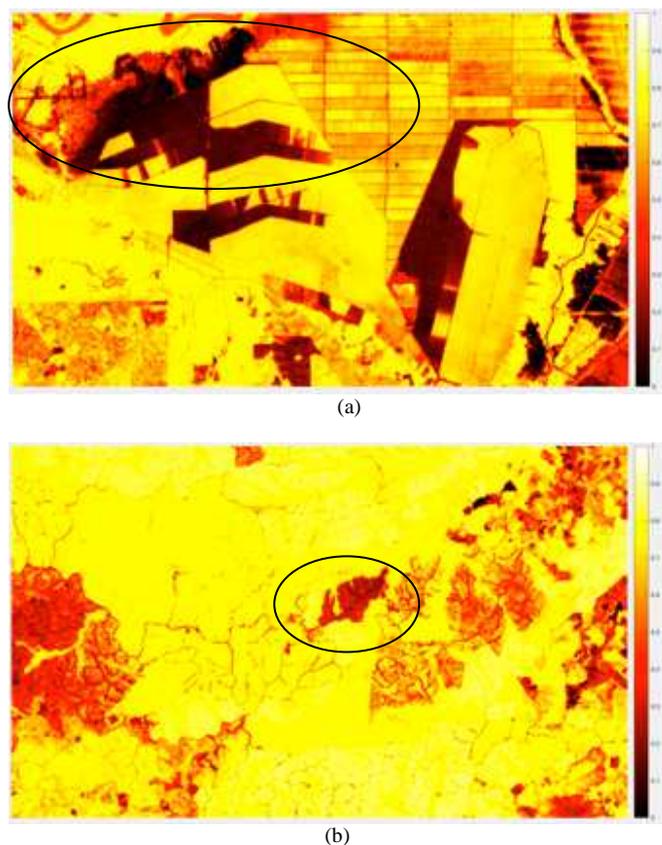


Fig. 11 Normalized Burning Ratio value in two specific deforestation area was chosen, area (a) is a deforested area by burning-trees and area (b) is a deforested are by clear-cutting

Fig. 11 indicated the activities of two different ways of deforestation by the different value of NBR index, for burning trees and clear-cutting. In section (a) represented by Fig. 11 showing the lower value of NBR which indicating the dense burned area, values of burned and unburned area are explained in eq. (3). A lower value (map as darker color in Fig. 11) identified the burn scar of the deforestation place. And by the section (b) in Fig. 11, the value of NBR shows higher value which indicating that the deforestation activity happens to this place is by clear-cutting. The validation method has done by comparing the burned area with open hotspot data provided by Indonesian Aerospace Agency (Lapan).

After assessing the area with NBR calculation, the next step of the experiment is using burn severity to present the knowledge about the degree of severity of the fire, eq. (4) is implemented as shown in two places, where place A shown in Fig. 12(a) and place B is shown in Fig. 12 (b).

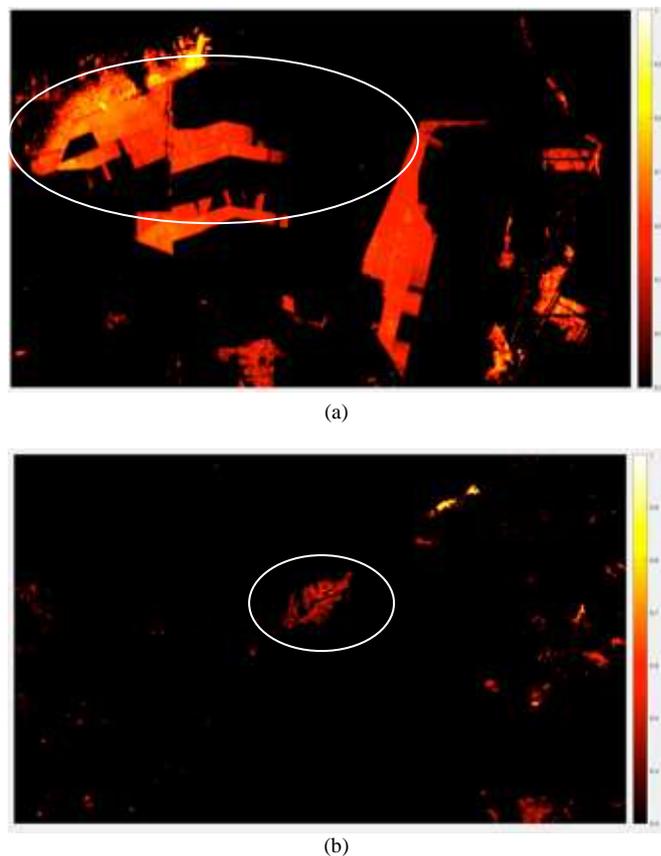


Fig. 12. Normalized Burning Ratio value in two specific deforestation area was chosen, area (a) is a deforested area by burning-trees and area (b) is a deforested are by clear-cutting

It is shown that the value of indices of land with a forest fire, is higher (a) compare to the land with deforestation caused by illegal logging. The comparable value between burning tree and clear-cutting deforestation are identified as shown in Fig. 13.

long	lat	nbr fire	nbr logging
0.83	4136.00	0.83	0.636
0.82	4142.00	0.82	0.658
0.83	4127.00	0.83	0.711
0.82	4151.00	0.82	0.545
0.82	4159.00	0.82	0.408

Fig. 13 NBR value of deforestation because of burning tree(fire) and clear-cutting(logging)

### 3) Characteristic of Parameters of Soil Measurement

Soil moisture and texture are analysed by using Red and NIR Reflectance from the study site in Riau at 2013 and 2014 (Fig. 6). The closer data with the H soil line, the more moisture the soil is, this method particularly useful and sensitive for monitoring changes in soil water content over time.

#### • Soil Moisture

Soil moisture and texture are analysed by using Red and NIR Reflectance from the study site in Riau at 2013 and 2014 (Fig. 6). The closer data with the H soil line, the more moisture the soil is, this method particularly useful and

sensitive for monitoring changes in soil water content over time.

long	lat	moisture fire	long	lat	moisture logging
2804.00	1838.00	0.64	3442.00	4136.00	0.83
2778.00	1863.00	0.66	3447.00	4142.00	0.82
2811.00	1883.00	0.71	2462.00	4127.00	0.83
2740.00	1905.00	0.55	3422.00	4151.00	0.82
2764.00	1857.00	0.41	3443.00	4159.00	0.82

Fig. 14 . Deforestation cause-based effect on temperature of soil

The smaller value of the red reflectance with a higher value of NIR, the dried soil condition will be. Table 1 indicating the correlation between red and NIR to determine the soil moisture. To mining the interesting knowledge in deforestation area, the system select 5 unique point from the study area, where every point should representing soil moisture after deforestation activity. The droughts and moisture of the soil could be identified by the value of NIR and Red Reflectance, with high NIR Reflectance value and a small value of Red.

• *Soil Temperature*

Soil temperature was analysed by using equation (3) and represented in Celsius Degree. The surface or soil temperature at study area in 2013 and 2014 was between 21° C ~ 38° C. While normal soil temperature for the tree to grow is in between 20° C ~ 26° C. Soil temperature was highly dependent on the cause of deforestation. Deforestation because of forest fire has higher temperature compare to logging activity as shown in Fig. 15.

long	lat	nbr fire	temp fire	long	lat	nbr logging	temp logging
2804.00	1838.00	0.83	25.77	3442.00	4136.00	0.64	24.387
2778.00	1863.00	0.82	25.56	3447.00	4142.00	0.66	24.348
2811.00	1883.00	0.83	25.35	2462.00	4127.00	0.71	24.966
2740.00	1905.00	0.82	26.69	3422.00	4151.00	0.55	24.132
2764.00	1857.00	0.82	26.47	3443.00	4159.00	0.41	24.362

Fig. 15 Deforestation cause-based effect on temperature of soil

• *Soil Salinity*

A major constraint to identify salinity is a correlation between spatial and temporal variability in the soil profile. Spectral data acquisition in nowadays research does not allow high precision information to be extracted from the entire soil profile since only the Earth surface is observed. But, it is still reflecting the change of salinity of soil after deforestation effect. Soil salinity index is present in a range between 0 and 1. The higher value of salinity index showing the higher soil salinity in that area, and as seen in Fig.16, the salinity of soil in Study Area have a brief correlation with deforestation activity occur in March 2014, where the salinity in Fig. 8 shows a high number of the saline index which reaches around 0.5.

long	lat	soil salinity fire	long	lat	soil salinity logging
3442.00	4136.00	0.36	3442.00	4136.00	0.47
3447.00	4142.00	0.36	3447.00	4142.00	0.41
2462.00	4127.00	0.47	2462.00	4127.00	0.58
3422.00	4151.00	0.29	3422.00	4151.00	0.55
3443.00	4159.00	0.34	3443.00	4159.00	0.54

Fig. 16 Salinity value and temp value of cause-based deforestation

4) *Semantic Analysis*

Context-dependent in interpreting nature condition is applied to determine the effect of deforestation in this research. The context of forest fire and context of logging

were selected as the semantic words analysis. The combination of three semantic spaces as proposed to determine the meaning of deforestation area stated in Table 1. In order to put an analysis in the context of Forest Fire, soil moisture, temperature and salinity were used to give the meaning. Semantic meanings are interpreted by measuring the distance between the semantic interpretations in Table 1 with the close observance point.

TABLE IV  
SEMANTIC INTERPRETATION IN CONTEXT OF FOREST FIRE AND CONTEXT OF LOGGING

Dryness	Temperature	Salinity	Meaning
<i>Context: Plantation</i>			
0.45 ~ 0.50	27° C ~ 30° C	0.47 ~ 0.55	New Deforested Area, Critical Soil Condition,
0.30 ~ 0.45	25° C - 27° C	0.45 ~ 0.46	Severe Soil Condition,
0.25 ~ 0.30	25° C - 27° C	0.35 ~ 0.40	New Deforested Area, below average Soil Condition.
0.15 ~ 0.25	20° C - 25° C	0.25 ~ 0.35	Good Soil Condition for plant growth.
<i>Context: Soil Degradation</i>			
0.45 ~ 0.50	24° C ~ 25° C	0.37 ~ 0.45	New Deforested Area, Critical Soil Condition,
0.30 ~ 0.45	22° C - 24° C	0.35 ~ 0.45	Severe Soil Condition,
0.30 ~ 0.45	20° C - 22° C	0.25 ~ 0.35	Good Soil Condition for plant growth.

As the experiment result of semantic analysis, we initially put 10 examples point of deforestation area and input to the system to know the effect in language. By applying semantic analysis to determine the effect of deforestation in soil condition, we could interpret reflected “substances (material)” of deforestation area in spectrum domain into the human language in different context. The results of effect classification parameter as effect classification of deforestation effect found 4 classes. Fig. 16. Shows the different point of deforestation activity and which word has the closest meaning to interpret each point.

long	lat	Word1	Word2	Word3	Word4
3047.00	2110.00	0.057	0.197	0.247	0.377
2237.00	3047.00	0.061	0.153	0.203	0.333
2807.00	3045.00	0.098	0.042	0.092	0.222
2802.00	3030.00	0.035	0.105	0.155	0.285
2845.00	2998.00	0.121	0.019	0.069	0.199
1789.00	3272.00	0.174	0.034	0.056	0.146
1925.00	2956.00	0.131	0.045	0.095	0.189
1929.00	2989.00	0.121	0.053	0.103	0.199
2806.00	2173.00	0.071	0.101	0.151	0.249
1892.00	2946.00	0.156	0.026	0.076	0.164
2101.00	2952.00	0.118	0.258	0.308	0.438
2125.00	3033.00	0.143	0.006	0.056	0.177
2244.00	3118.00	0.193	0.253	0.303	0.412
2232.00	3304.00	0.135	0.029	0.079	0.185
1787.00	3184.00	0.236	0.096	0.046	0.084
2862.00	2898.00	0.141	0.005	0.055	0.179
1265.00	6679.00	0.428	0.288	0.238	0.114

Fig. 16 Semantic Interpretation in context of Plantation of Table 4

The system will detect the input value and use range effect for classification, which design and provide in table 4 from knowledgebase discovery in deforestation knowledge area as a knowledge database for analysis process of semantic Analysis.

#### IV. CONCLUSION

Deforestation is driven by two different activities: illegal logging and forest fire. Though the effects of deforestation are varying, these effect is highly dependable to its causes. In order to interpreting the knowledge from deforestation effect from its cause, we proposed a system function to automatically extract the feature of multispectral images for interpreting reflected value of bare soil substances in deforested area in spectrum domain into human language using semantic computing, to bring the clear comprehensible knowledge even for people who are not familiar with forestry. The experimental study has successfully produced the semantic matrices to interpret the nature condition by using three features and convert to axes, which consist of soil moisture, temperature, and soil salinity.

#### ACKNOWLEDGEMENT

This work is supported in part by MEXT Grant-in-Aid for the Program for Leading Graduate School, “Global Environmental System Larders (GESL)” and Multimedia Database Laboratory (MDBL), Graduate School of Media and Governance, Keio University. We thank the anonymous reviewers for their valuable comments and suggestions

#### REFERENCES

- [1] Tropical Deforestation, Gerald Urquhart, Walter Chomentowski, David Skole, And Chris Barber,1999
- [2] Archana K, Impact Of Deforestation On Climate Change, IOSR Journal Of Environmental Science, Toxicology And Food Technology (IOSR-JESTFT) E-ISSN: 2319-2402,P- ISSN: 2319-2399. Volume 4, Issue 2 (May. - Jun. 2013), PP 24-28
- [3] C. J. TUCKER And J. R. G. TOWNSHEND , Strategies For Monitoring Tropical Deforestation Using Satellite Data, Int. J. Remote Sensing, 2000, Vol. 21, No. 6 & 7, 1461–1471
- [4] Marcio Pupin Mello\*, Flora S. R. V. Martins, Luciane Y. Sato, Roberta Z. Cantinho, Daniel A. Aguiar, Bernardo F. T. Rudorff, Rafael D. C. Santos Spectral-Temporal Analysis By Response Surface Applied To Detect Deforestation In The Brazilian Amazon, 978-1-4577-1203-6/11/\$26.00 ©2011 Ieee
- [5] Badamasi, M. M.,I Yelwa, S. A.,2 Abdulrahim, M. A.2 And Noma, S. S NDVI Threshold Classification And Change Detection Of Vegetation Cover At The Falgore Game Reserve In Kano State, Nigeria
- [6] F. Baret, S. Jacquemoud And J. F. Hanocq, About The Soil Line Concept In Remote Sensing, 1993 Cospar Real Time Drought Monitoring Using Remote Sensing Approaches Case Study Western Desert: Kharga And Dakhla Regions Hala M. Ebaid International Journal Of Geomatics And Geosciences Volume 6 Issue 3, 2015
- [7] Yasushi Kiyoki, Xing Chen, Shiori Sasaki And Chawan Koopipat\*\*\*, Globally-Integrated Environmental Analysis And Visualization System With Multi-Spectral & Semantic Computing In “Multi-Dimensional World Map”. Ejc 2016.
- [8] Landsat 8 Handbook, 2013, NASA
- [9] Landscape Assessment (LA), Sampling And Analysis Methods,Carl H. Key Nathan C. Benson
- [10] Semantic Analysis Of Deforestation Soil Effect Using Multispectral Reflectance, Irene Eryln W.R, Yasushi Kiyoki, Internation Electronics Symposium 2016.
- [11] Megan Clark, Deforestation in Madagascar: Consequences of Population Growth and Unsustainable Agricultural Processes. Global Majority E-Journal, Vol. 3, No. 1 (June 2012), pp. 61-71
- [12] b. X. Úbeda, M. Lorca, L. R. Outeiro, S. Bernia, and M. Castellnou, “Effects of prescribed fire on soil quality in Mediterranean grassland (Prades Mountains, north-east Spain),” International Journal of Wildland Fire, vol. 14, no. 4, pp. 379–384, 2005
- [13] c. P. Kutiel and Z. Naveh, “The effect of fire on nutrients in a pine forest soil,” Plant and Soil, vol. 104, no. 2, pp. 269–274, 1987
- [14] d. J. A. Arévalo, S. Fernández-Lugo, A. Naranjo-Cigala et al., “Post-fire recovery of an endemic Canarian pine forest,” International Journal of Wildland Fire, vol. 23, no. 3, pp. 403–409, 2014
- [15] e. D. J. Pennock and C. van Kessel, “Clear-cut forest harvest impacts on soil quality indicators in the mixedwood forest of Saskatchewan, Canada,” Geoderma, vol. 75, no. 1-2, pp. 13–32, 1997
- [16] f. Stolorz, P., Naamura, H., Muntz. Fast Spatio-temporal data mining of large geophysical datasets. Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95).(pp. 300-305).
- [17] g. Setia Darmawan Afandi, Yeni Herdiyeni, Lilik B Prasetyo, Fuzzy C-means for Deforestation Identification Based on Remote Sensing Image, ICACSSIS 2014
- [18] C. Immaculate Mary, S. V. Kashmir Raja, Refinement of Clusters from K-Means with Ant Colony Optimization, Journal of Theoretical and Applied Information Technology, 2005–2009
- [19] Sheng Zheng, Chunxiang Cao, Yongfeng Dang, Haibing Xiang, Jian Zhao, Yuxing Zhang, Xuejun Wang, Hongwen Guo, Retrieval of forest growing stock volume by two different methods using Landsat TM images, International Journal of Remote Sensing, 2014.
- [20] G. Pickup, B.D. Foran, The Use of Spectral and Spatial Variability to Monitor Cover Change on Inert Landscapes, Remote Sensing of Environment 23:351-363, 1987.
- [21] E.F. Lambin, Change Detection at Multiple Temporal Scales: Seasonal and Annual Variations in Landscape Variables, Photogram. Eng. Remote Sensing. 62, 931–938, 1996.
- [22] J. Vogt, Characterizing The Spatio-Temporal Variability of Surface Parameters from NOAA-AVHRR Data, Report EUR 14637 EN, Agriculture Series, pp. 266, Joint Research Centre, Institute for Remote Sensing Applications, Italy, 1992.