DEVELOPMENT OF ANNUAL COMPOSITE ALGORITHM USING LANDSAT-8 TO MINIMIZES CLOUD (CASE STUDY: SOUTHERN PART OF CENTRAL KALIMANTAN)

Kustiyo
Remote Sensing Technology and Data Centre, LAPAN

kustiyo@lapan.go.id

Abstract. Since January 2013, Landsat-8 data can be freely accessed from LAPAN, making it possible to use the all available Landsat-8 data to produce the cloud-free Landsat-8 composite images. This study used Landsat-8 archive images in 2015, Operational Land Imager (OLI) sensor in 30 meters resolution, geometric correction level of L1T. The eight data in L1T of 118-062,southern part of Central Kalimantanwere used to produce a cloud-free composite image. Radiometric correction using Top of Atmospher (TOA) and Bidirectional Reflectance Distribution Function(BRDF) algorithm to produce reflectance images have been applied, and then the most cloud-free pixels were selected in composite result. Six composite methods base on greeness, open area and haze indexs were compared, and the best one was selected using visual analysis. The analysis shows that the composite algorithm using Max (Max (NIR, SWIR1) / Green) produces the best image composite.

1. Introduction

Since January 2013, National Aeronautics and Space Institute (LAPAN) have been providing free terrain corrected (Level 1T) or systematic terrain corrected (L1GT) of Landsat-8 images, for whole acquisitions of Indonesia. With this dataset, temporally composited, mosaics of Indonesia could be generated periodically. Landsat acquisitions with any cloud cover are processed and users may request any other scene of Indonesia area. The Landsat 8 Operational Land Imager (OLI) sensor is the most recent in a series of Landsat-8 sensors that acquire medium resolution multi-spectral data over an approximately 183 km×170 km extent, with a 16 day revisit capability. Every Landsat overpass of Indonesia is acquired by LAPAN, providing 22 or 23 acquisitions per year per path/row (Ju and Roy 2008).

Regional mosaics of Landsat imagery are increasingly being developed to meet national monitoring and reporting needs across land-use and resource sectors, for example, in Canada (Wulder et al. 2002), the Congo basin (Hansen et al. 2008), and Indonesia (Kustiyo et al. 2015). Large volume Landsat processing was developed by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) that processed over 2100 Landsat Thematic Mapper and ETM+ acquisitions to provide wall-to-wall surface reflectance coverage for North America for the 1990s and 2000s (Masek et al. 2006).

Compositing procedures are applied to reduce cloud and aerosol contamination, fill missing values, and reduce the data volume (Cihlar and Manak 1994). The mosaic processing steps included conversion of digital numbers to calibrated radiance to top of atmosphere reflectance and brightness temperature, per-band radiometric saturation identification, cloud screening, re-projection, and compositing (Roy et al. 2010).

The compositing images could be used in any application, especially in producing land cover maps. Global land cover maps have been produced with multiple-year data from Landsat (Gong et al. 2013) and single- year Landsat-like imagery (Chen et al. 2014), with reported overall accuracies ranging from 65–80%. Land cover maps can be updated by identification and mapping of changed areas,

leaving unchanged areas in the original map intact. Some large-area land cover programs currently apply such a change-updating approach, for example European CORINE Land Cover (Büttner et al. 2004) or the US NLCD (Xian et al. 2009).

The compositing approach in this research is designed to preferentially select valid land surface observations with minimal cloud, and atmospheric haze contamination. The processing approach is intentionally designed to facilitate automated processing with minimal human intervention, including no chronological order of the Landsat acquisition and processing dates, and to provide processing in near-real time i.e., updating composited mosaics shortly after the Landsat data are acquired.

2. Data and Method

2.1. Data and Research Sites

The Landsat-8 data are nominally processed as Level 1 terrain corrected (L1T) data. The L1T data are available in GeoTIFF format in the Universal Transverse Mercator (UTM) map projection with World Geodetic System 84 (WGS84) datum which is compatible with heritage GLS and Landsat MSS data sets. The Level 1T processing includes radiometric correction, systematic geometric correction, precision correction using ground control chips, and the use of a digital elevation model to correct parallax error due to local topographic relief. While most Landsat data are processed as L1T (i.e., precision and terrain-corrected), certain acquisitions do not have sufficient ground control for precision, respectively. In these cases, the best level of correction is applied and, the data are processed to Level 1GT systematic (L1GT).

This researchused the level L1T of Landsat-8 archive data that cover the southern part of Central Kalimantan. Figure 1 shows the images quick look of eight scene Landsat-8 that selected from 23 scene available images using cloud cover less than 50%, and Table 1 shows the detail information about cloud cover and geometric accuracy in X and Y directions.

230115	280315	310515	020715

| 030815 | 190815 | 040915 | 231115 |
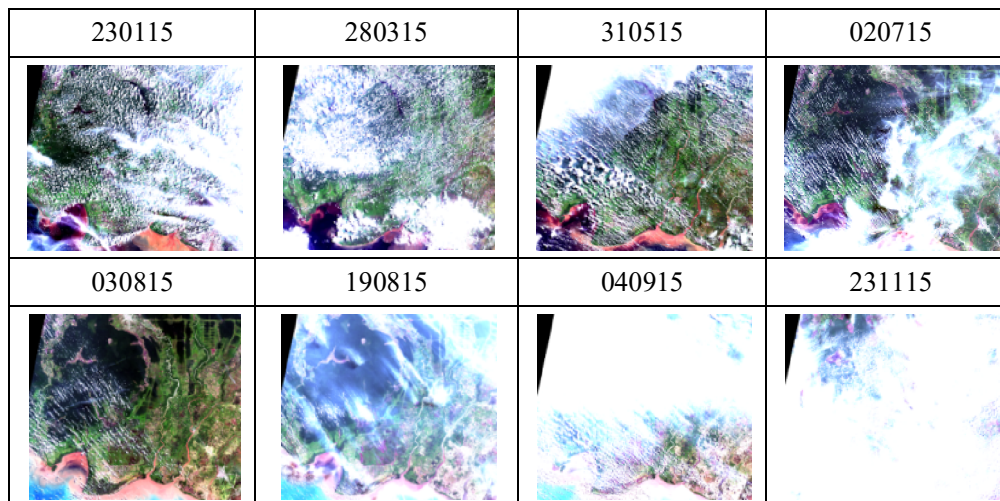
Figure 1. Landsat-8 natural colour combination of scene 118-062, level L1T in 2015 with cloud cover less than 50%

Table 1. Cloud cover and geometric accuracy of Landsat-8 scene 118-062, level L1T in 2015 with cloud cover less than 50%

No	Acquisition Date	Level Data	CC	RMSE-X	RMSE-Y
1	23-01-2015	L1T	30.68	7.369	6.354
2	28-03-2015	L1T	26.54	7.688	5.403
3	31-05-2015	L1T	23.74	7.437	5.110
4	02-07-2015	L1T	8.13	5.140	5.757
5	03-08-2015	L1T	2.04	6.275	5.700
6	19-08-2015	L1T	1.81	6.189	5.877
7	04-09-2015	L1T	38.09	6.249	6.297
8	23-11-2015	L1T	45.82	7.004	6.318

2.2. Methodology

The five composite algorithm was compared to match the best composite result. The six algorithm are (1) Maximum Normalize Difference vegetation Index (NDVI), (2) Maximum from maximum NIR and SWIR divided by Green reflectance, (3) Maximum NIR divided by Green reflectance, (4) Maximum SWIR divided by Green reflectance, (5) Minimum Red reflectance and (6) Minimum Haze Index,. Radiometric correction using Top Of Atmosphere (TOA) and Bidirectional Reflectance Distribution Function (BRDF) correction was applied in the first step of processing. Finally, the best result was selected sing visual investigation. Figure 2 shows the general algorithm in producing the best composite image.
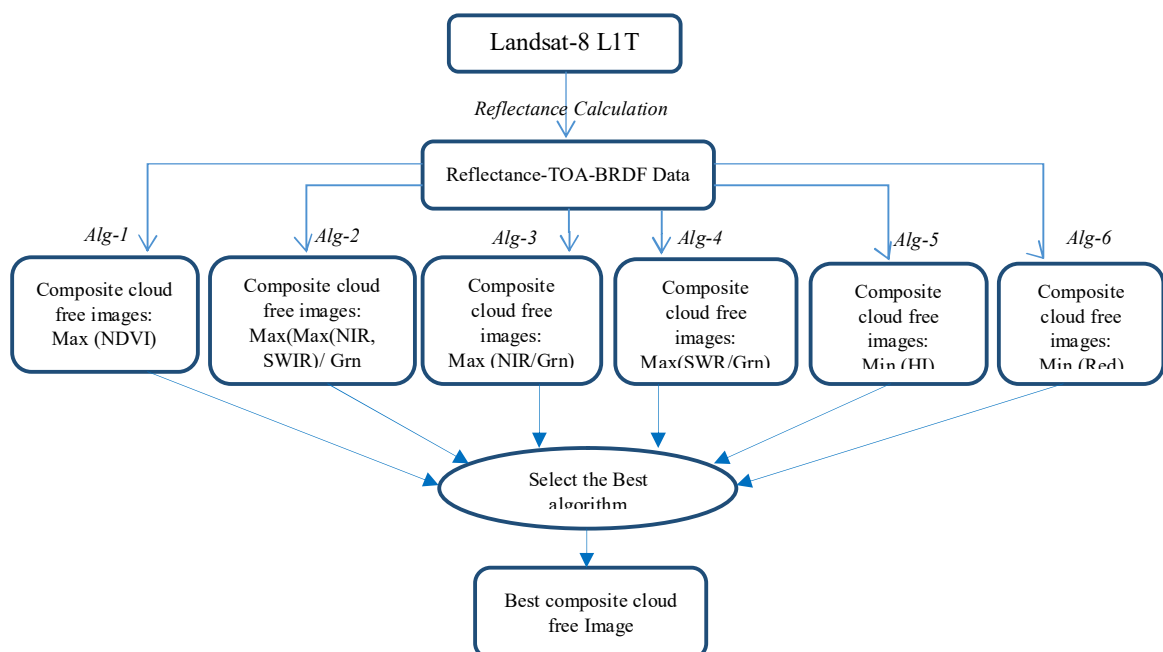
Figure 2. General algorithm in producing composite images

2.3. Pre Processing

The spectral radiance sensed by each OLI detector is stored as an 10-bit digital number. The digital numbers should be converted to radiance (units: W m^{-2} μm^{-1}), to minimize changes in the instrument radiometric calibration, and then converted to top of atmosphere reflectance to minimize remote

sensing variations introduced by variations in the sun–earth distance, the solar geometry, and exoatmospheric solar irradiance arising from spectral band differences (Chander et al. 2009).

The radiance sensed in the Landsat reflective wavelength bands, i.e., the blue, green, red, near-infrared, and the two mid-infrared bands, were converted to top of atmosphere reflectance using the standard formula as:

$$\rho_\lambda = \frac{\pi.L_\lambda + d^2}{ESUN_\lambda .\cos\theta_s} \quad\text{...} \quad (1)$$

where ρ_λ is the top of atmosphere (TOA) reflectance (unit less), L_λ is the TOA spectral radiance (W m^{-2} sr^{-1} μm^{-1}), d is the Earth–Sun distance (astronomical units), $ESUN_\lambda$ is the mean TOA solar spectral irradiance (W m^{-2} μm^{-1}), and θ_s is solar zenith angle at the center of the Landsat acquisition (radians).

Results of reflectance TOA correction is a real value between 0 and 1, and then multiplied by 60000 to be stored in a 16-bit integer. After that, the Bidirectional Reflectance Distribution Function (BRDF) correction was applied in radiometric correction.It gives the reflectance of a target as a function of illumination geometry and viewing geometry. The BRDF depends on wavelength and is determined by the structural and optical properties of the surface, such as shadow-casting, multiple scattering, mutual shadowing, transmission, reflection, absorption, and emission by surface elements, facet orientation distribution, and facet density.

2.4. Composite Algorithm

The composite algorithm selected the best pixel in same location from more than 2 images. The clearest pixel must be selected using certain algorithm. In general, the reflectance from visible to Short Wave Infra-Red (SWIR) that was measured by satellite sensor become higher in the haze and cloud condition, the clearest pixel was the minimum reflectance. More higher the wave length the more effected by haze and cloud condition, the visible band are more effected by haze and cloud condition compare with the Near Infra-Red (NIR) and SWIR bands. Using this idea, this research propose 6 algorithm for selecting the most clear images, and compare them to decide the best one.

The composite algorithm was selecting image n from the m input images, where n<m. In the Maximum NDVI algorithm, the formula is:

$IM_{bx}(i,j) = I_{bx,n}(i,j)$; x: 1,2,.….m
n is image number, so that:
$NDVI_n (i,j) = Max (NDVI_1 (i,j)……NDVI_i (i,j)……NDVI_m (i,j))$ ………………..(2)

Where $IM_{bx}(i,j)$: reflectance band bx,in row colom (i,j) from image mosaic; $I_{bx,n}(i,j)$: reflectance band bx, in row colom (i,j) from image number n; $NDVI_n (i,j)$: NDVI value in row colom (i,j) from image number n; m: number of data used in mosaic. The formula of 6 algorithms that were used are:

$NDVI (i,j) = (I_{NIR}(i,j) -I_{NIR}(i,j)) / (I_{NIR}(i,j) +I_{NIR}(i,j))$ …………………....……(3)
$MaxNirSwir\_Grn (i,j) = Maximum (I_{NIR}(i,j) ,I_{SWIR}(i,j)) / I_{GRN}(i,j)$ ………..(4)
$Nir\_Grn(i,j) = I_{NIR}(i,j) / I_{GRN}(i,j)$ …………………………………….…(5)
$Swir\_Grn(i,j) = I_{SWIR}(i,j) / I_{GRN}(i,j)$ ……………………………………(6)
$Red (i,j) = I_{RED}(i,j)$ …………………………………………………..(7)
$HI(i,j) = (3.2 * I_{BLU}(i,j)) - I_{RED}(i,j)$ …………………………………… .(8)

Where HI is haze index, $I_{bx}(i,j)$: reflectance band bx, in row colom (i,j). For formula 3,4,5, and 6 the selection criteria was maximum, but for formula 7 and 8, the selection criteria was minimum.
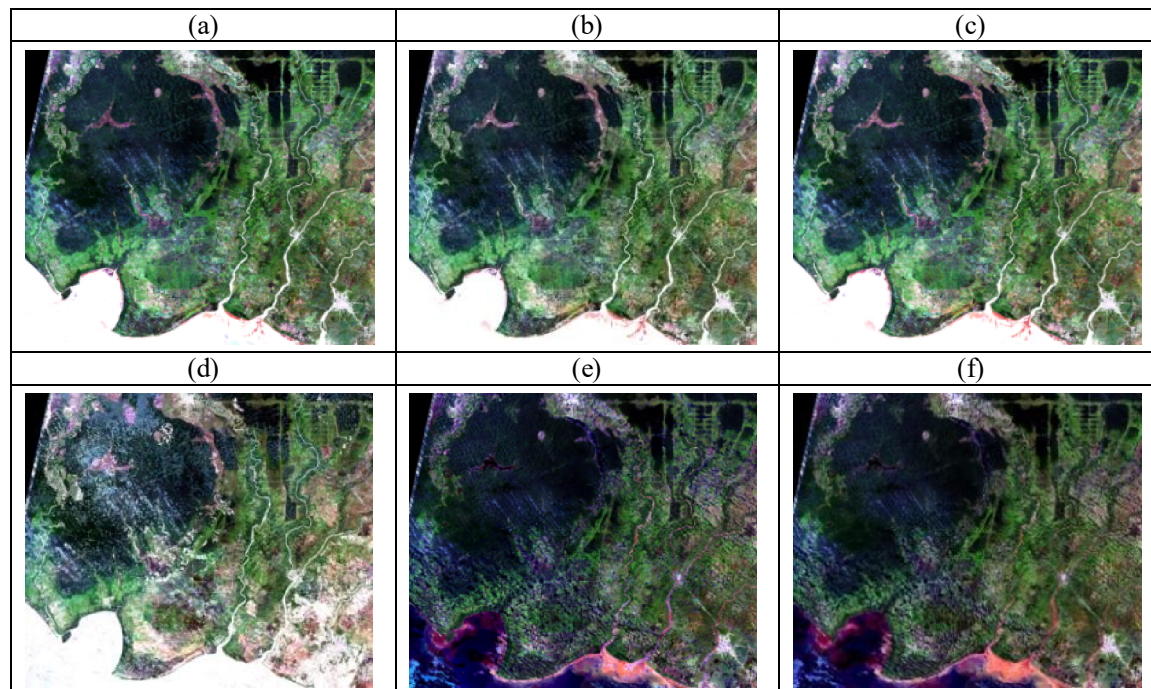
3. Results and Discussion

3.1. Results

The results was provided in 2 visual combinations, true color combination using swir-1, nir, and red in RGB domain, and natural color combination using red, green and blue in RGB domain. The composite results with true color combinations from the six models in this research can be seen in Figure 3. Visually, the composite results can be grouped into three groups, group-1 is (a), (b) and (c), group-2 is (d) and groups-3 is (e) and (f). Group 1, the general appearance is dominated by the green color in the land area, and blue and white color in the sea. In the land area, the vegetation object was selected as composite result, but in the sea area the composite result still cloudy. Group 2, the composite result was dominated by red color, the open land object was selected as composite result. In the sea area, some area still cloudy, this is same as with group 1. Group 3, no cloud in the land area, but many shadow with black color. In the sea area, no cloud is there.

Group 1 and group 2 can be used as cloud free image composite algorithm in land area but cannot be used in water area, especially in the sea. Otherwise, group 3 can be used as cloud free composite algorithm in water area. The results shows that in one side there is no cloud in land, and in other side there are no clouds in water area, combining them the result become better in land, and better in the water.
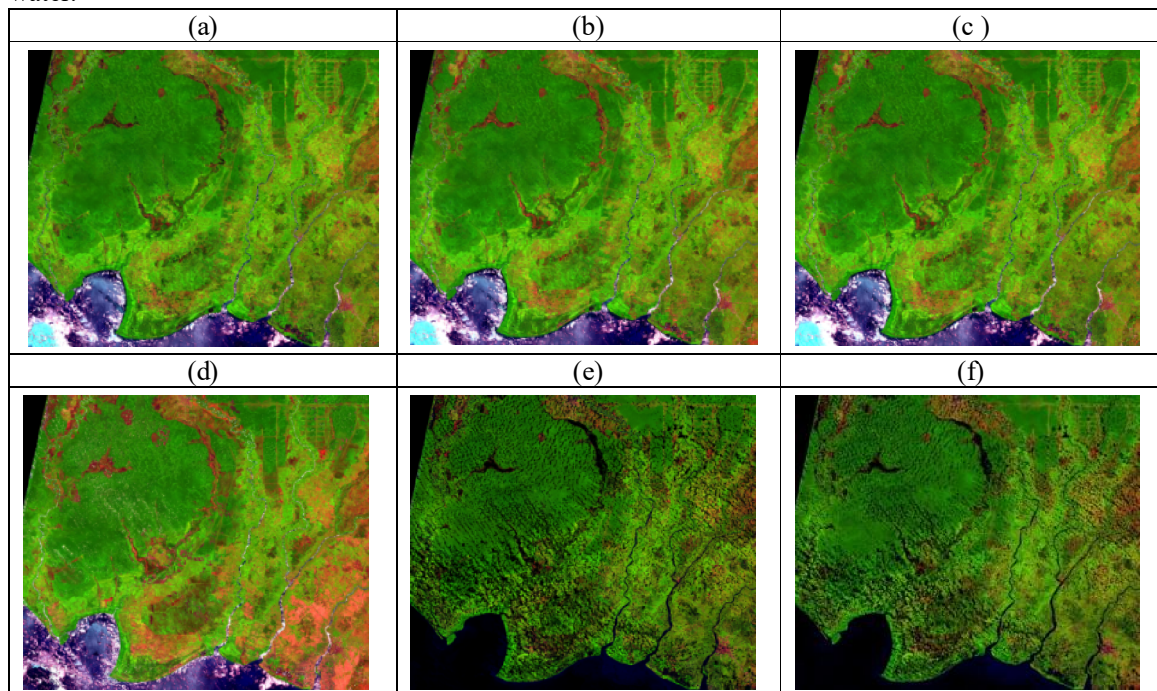
Figure 3. RGB (Swir, NIR, Red) composite result sing difference algorithm (a) maximum NDVI, (b) maximum_maximum_NIR_Swir_div_Green, (c) maximum_NIR_div_Green, (d) Maximum_Swir_div_Green, (e) minimum_red, and (d) minimum_haze_index
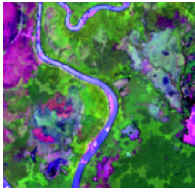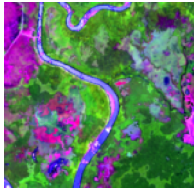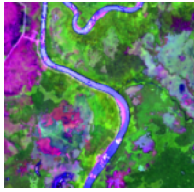
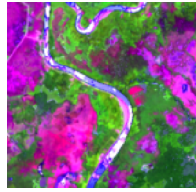Figure 4 shows the composite results in natural colour combination, this natural colour combination is well detect the haze area, because it used the visible wave length, this combination well detect the water condition. All results in Figure 4 shows the some hazy area in the land area, and white colour in water area in group 1 and group 2, especially in group 2 the result is more haze compared with group 1 results. In group 3, the water area is more colourful, and less hazy.

(a)	(b)	(c)
(d)	(e)	(f)

Figure 4. RGB (Red, Green, Blue) composite result sing difference algorithm (a) maximum NDVI, (b) maximum_maximum_NIR_Swir_div_Green, (c) maximum_NIR_div_Green, (d) Maximum_Swir_div_Green, (e) minimum_red, and (d) minimum_haze_index
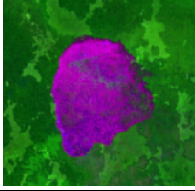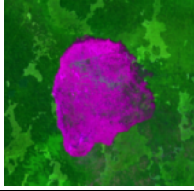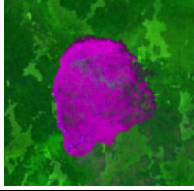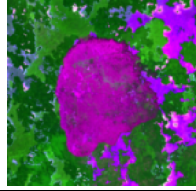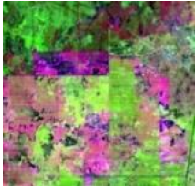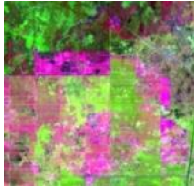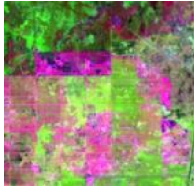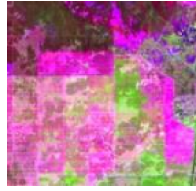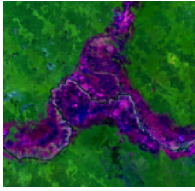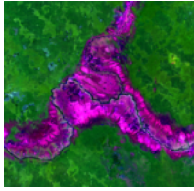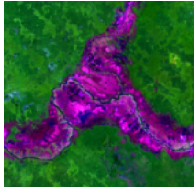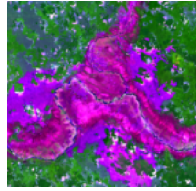
3.2. Discussions

In the results shows that group 1 and group 2 results is better in land area, but group 3 results is better in water area. The detail analysis and discussion will be focu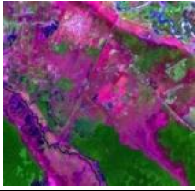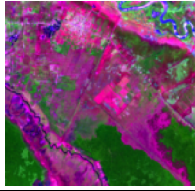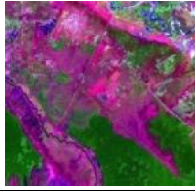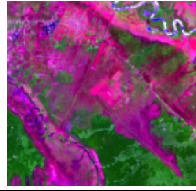sed on land area, so that, only the group 1 and group 2 will be compared more detail. Detail analysis have been done in five difference location, it was shown in Figure 5.

The discussion focused order of the selected pixel in the results. The discussion just for group 1 and group 2. Group 3 was not included because too many mistakes in land area, there are still many shadow area. In the group 3, pixel with shadow was selected in the composite image, because shadow have minimum value of red reflectance and minimum of haze index. Table 2 discuss about the order and some mistake (e.g. wrong choice in selecting the clear pixel) in group 1 and group 2 to match the best algorithm.

Table 2. Order in selecting pixel of the composite results and some mistakes

algorithm	Order in selecting pixel	mistakes
maximum NDVI	(1)vegetation; (2) water ; (3) open area	some shadow border
maximum_maximu m_NIR_Swir_div_G reen	(1)vegetation; (2) water ; (3) open area	-
maximum_NIR_div _Green	(1)vegetation; (2) water ; (3) open area	little shadow border
Maximum_Swir_div _Green	(1)vegetation; (2) water ; (3) open area	• some white cloud • hazy pixel in open land

No	(a)	(b)	(c)	(d)	note
1					in (a)(b)(c) as vegetation, but in (d) as open area
2					hazy in (d)
3					some shadow border in (a) little shadow border in (c)
4					hazy in (d) in (a) (c) as water, but in (b) as open area
5					(b) (e) are clear, but (a) (c) blur

Figure 5. RGB (SWIR, NIR, Red) composite result sing difference algorithm (a) maximum NDVI,(b) maximum_maximum_NIR_Swir_div_Green, (c) maximum_NIR_div_Green, (d) Maximum_Swir_div_Green

4. Conclutions

The research described in this paper represents preliminary results of a project with the goal of providing consistent mosaic of Indonesia using Landsat-8 in 30 meter pixel resolution. Such data are needed to monitor land-cover change, especially in forest monitoring. The result shows that the maximum (maximum(NIR,Swir)Green) algorithm was the best algorithm in compositing the images. The result of true colour combinations is enough in visual interpretation, but using the natural colour combination in some hazy area still appears. Increasing the compositing period reduced the percentage of cloudy and hazy area. Difference algorithm produce difference results in land and water area, future research is needed to combine any algorithm that match in land also in water area.

Acknowledgments

standard imagery data. We are thankful to our colleagues who provided the strong motivation in finishing this paper.

References

Büttner G, Feranec J, Jaffrain G, Mari L, Maucha G and Soukup T 2004 The Corine Land Cover 2000 Project *EARSeL eProceedings* **3** 331–46

Chander G, Huang C, Yang L, Homer C and Larson C 2009 Developing consistent landsat data sets for large area applications: The MRLC 2001 protocol *IEEE Geosci. Remote Sens. Lett.* **6** 777–81

Chen C, Chen Z, Li M, Liu Y, Cheng L and Ren Y 2014 Parallel relative radiometric normalisation for remote sensing image mosaics *Comput. Geosci.* **73** 28–36

Cihlar J and Manak D 1994 Evaluation of Compositing Algorithms for AVHRR Data Over Land *IEEE Trans. Geosci. Remote Sens.* **32** 427–37

Gong P, Wang J, Yu L, Zhao Y, Zhao Y, Liang L, Niu Z, Huang X, Fu H, Liu S, Li C, Li X, Fu W, Liu C, Xu Y, Wang X, Cheng Q, Hu L, Yao W, Zhang H, Zhu P, Zhao Z, Zhang H, Zheng Y, Ji L, Zhang Y, Chen H, Yan A, Guo J, Yu L, Wang L, Liu X, Shi T, Zhu M, Chen Y, Yang G, Tang P, Xu B, Giri C, Clinton N, Zhu Z, Chen J and Chen J 2013 Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data *Int. J. Remote Sens.* **34** 2607–54

Hansen M C, Roy D P, Lindquist E, Adusei B, Justice C O and Altstatt A 2008 A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change and preliminary results for Central Africa *Remote Sens. Environ.* **112** 2495–513

Ju J and Roy D P 2008 The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally *Remote Sens. Environ.* **112** 1196–211

Kustiyo, Roswintiarti O, Tjahjaningsih A, Dewanti R, Furby S and Wallace J 2015 Annual forest monitoring as part of the Indonesia's National Carbon Accounting System *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* **40** 441–8

Masek J G, Vermote E F, Saleous N E, Wolfe R, Hall F G, Huemmrich K F, Gao F, Kutler J and Lim T K 2006 A landsat surface reflectance dataset for North America, 1990-2000 *IEEE Geosci. Remote Sens. Lett.* **3** 68–72

Roy D P, Ju J, Kline K, Scaramuzza P L, Kovalskyy V, Hansen M, Loveland T R, Vermote E and Zhang C 2010 Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States *Remote Sens. Environ.* **114** 35–49

Wulder M, Loubier E and Richardson D 2002 Landsat-7 ETM+ orthoimage coverage of Canada *Can. J. Remote Sens.* **28** 667–71

Xian G, Homer C and Fry J 2009 Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods *Remote Sens. Environ.* **113** 1133–47