Automated Landsat 8 data preprocessing for national forest monitoring system

Hudjimartsu, Sahid, Prasetyo, Lilik, Setiawan, Yudi, Suyamto, Desi, Ikbal, Wim, et al.

Sahid Hudjimartsu, Lilik Prasetyo, Yudi Setiawan, Desi Suyamto, Wim Ikbal, Kustiyo, "Automated Landsat 8 data preprocessing for national forest monitoring system," Proc. SPIE 10773, Sixth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2018), 107730R (6 August 2018); doi: 10.1117/12.2326100

Event: Sixth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2018), 2018, Paphos, Cyprus

Automated Landsat 8 Data Preprocessing for National Forest Monitoring System

Sahid Hudjimartsu *[a], Lilik Prasetyo [a], Yudi Setiawan [a], Desi Suyamto [a], Wim Ikbal [a], Kustiyo [b]
[a] Forests2020 Programme, Environmental Analysis and Spatial Modelling Laboratory Department of Forest Resource Conservation, Faculty of Forestry, Bogor Agricultural University, Bogor, Indonesia;
[b] Indonesian National Institute of Aeronautics and Space (LAPAN), Deputy for Remote Sensing Affairs, Indonesia

ABSTRACT

Precise digital classification for Landsat 8 data of remote sensing images require pre-processing steps. The pre-processing consist of conversion from digital numbers (DN) to top of atmosphere (TOA) reflectance, cloud and cloud shadow masking, topographic correction and image normalization. In general, pre-processing steps were implemented to National scale (Indonesia) excluding topographic correction. The topographic correction algorithm is required to avoid reflectance bias from terrain effects due to shading. The highest mountains in Indonesia were selected as window areas, considering the reflectance bias is produced due to terrain effects. The results showed that algorithm is able to solve overcorrection problems and will be implemented into LAPAN's system of image pre-processing for National scale. This research is a collaboration between Bogor Agricultural University (IPB) with National Institute of Aeronautics and Space (LAPAN) under Forests2020 Programme, in order to produce Landsat 8 data with the minimal cloud over Indonesia annually and then to automatically digital classification for forest monitoring. The automated system of pre-processing was developed with Perl and Python programming languages.

Keywords: Digital Classification, Landsat 8, Preprocessing, Topographic Correction, National Scale

1. INTRODUCTION

Current and past Landsat missions collect an earth observation data. Landsat data is suitable for detecting changes on earth's surface including changes in forest cover [1-4]. The changes can be analyzed by using three classification methods in visual or manual interpretation and automated classification [5]. In automated method, precise digital classification requires pre-processing steps to correct radiometric anomalies of satellite[6]. The pre-processing consist of conversion from digital numbers (DN) to top of atmosphere (TOA) reflectance, cloud and cloud shadow masking, topographic correction and image normalization. Several pre-processing steps were implemented to National scale in LAPAN's system of image pre-processing excluding topographic correction.

Terrain elevation variations also cause shading that can introduce errors in image understanding in general or differences illumination conditions due to solar position with respect to slope and aspect that can produce reflectance bias of pixel in the same category [7, 8]. To avoid reflectance bias from terrain effects due to shading, there are several algorithms for topographic correction remotely sense data from satellites with optical sensors, i.e. (1) cosine correction, (2) Minnaert correction, (3) statistical–empirical correction, and (4) C-correction [9]. The cosine and Minnaert correction models are based on the radiance from idealized surface, known as a Lambertian surface, the Lambertian surface assumption causes overcorrection of topographic effect[10]. The Statistical-empirical and C models are wave-length dependent, representing a specific wave-length, developed to overcome overcorrection problems for wave-length independent correction models (cosine and Minnaert). According to [11], introduced empirical rotation model for topographic correction and does not assume a lambertian surface. The model compared with consine and C models and the result showed the rotation model can solve overcorrecting problem which occurred with models based on Lambertian surface assumptions[10].

The main objective of this study was to implement the topographic correction model to National scale especially in Indonesia. The algorithm is required for spectral correction from terrain effects due to shading for automated digital classification. Automated digital classification is needed to improve the classification method of visual interpretation to

Sixth International Conf. on Remote Sensing and Geoinformation of the Environment (RSCy2018),
K. Themistocleous, G. Papadavid, S. Michaelides, V. Ambrosia, D. G. Hadjimitsis, Eds., Proc. of SPIE Vol. 10773,
107730R · © 2018 SPIE · CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2326100

automate classification. LAPAN currently relies on manual or visual interpretation to classify land cover in Indonesia in order to produce Landsat 8 data with the minimal cloud over Indonesia annually. The automated LAPAN's system of pre-processing was developed with Perl and Python programming languages.

2. STUDY SITE AND DATA

2.1 Study site

The highest mountains in Indonesia were selected as the study sites to evaluate the topographic correction algorithm. Selection the mountain based on spreading from west to east Indonesia, there are Mount Kerinci (± 3,805 m) on Sumatra Island, Mount Pangrango (± 3,019 m) on Java Island, Mount BuyuBalease (± 3,016 m) on Sulawesi Island, Mount Rinjani (3,726 m) on Lombok Island and Mount Jayawijaya (4,884 m) on Papua Island (see Figure. 1).

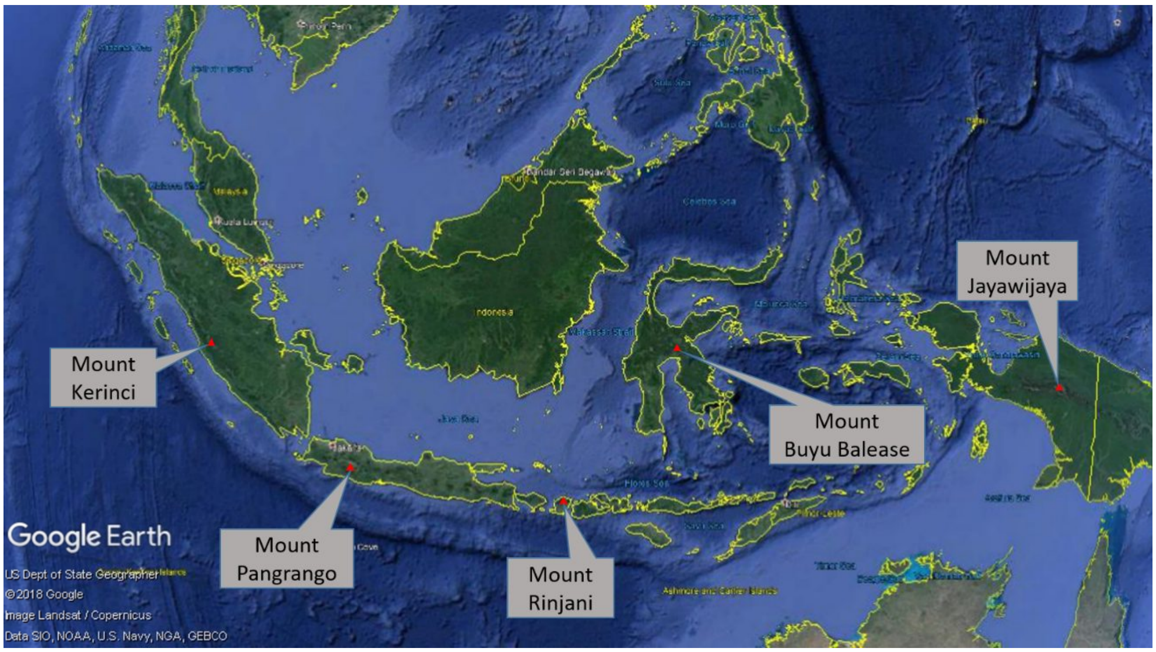

Figure 1. The highest mountains site in Indonesia (Image Landsat from Google Earth on the background)

2.2 Data

This study used three data to automate the pre-processing Landsat 8 data in particular topographic correction for National scale. Five Landsat 8 data are used as input for the algorithm and moreover SWIR, NIR, and red bands of Landsat 8 (i.e. band 6, band 5, band 4) were selected to evaluate the algorithm. The second is digital elevation model (DEM) as the key to input for illumination correction algorithm. The DEM data used in this study is the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 2 (GDEM V2) with spatial resolution of this data set is 30 m. The third is land cover data to determine areas classified as forest and non-forest. The land cover data applied is sourced from forest planning agency (BAPLAN) part of Ministry of Environment and forestry (MoEF) data 2016.

3. METHODS

Several stages in order to automate pre-processing for topographic correction, basically combines illumination modelling and simple statistical approaches to correct the reflectance of satellite imageries from terrain effects, mostly adapted from [10]. The topographic correction procedures can be seen in Figure 2. The model was developed using Python and available at https://github.com/hudji/SCRIPT/blob/master/Auto/SCRIPT/Topographic.py.

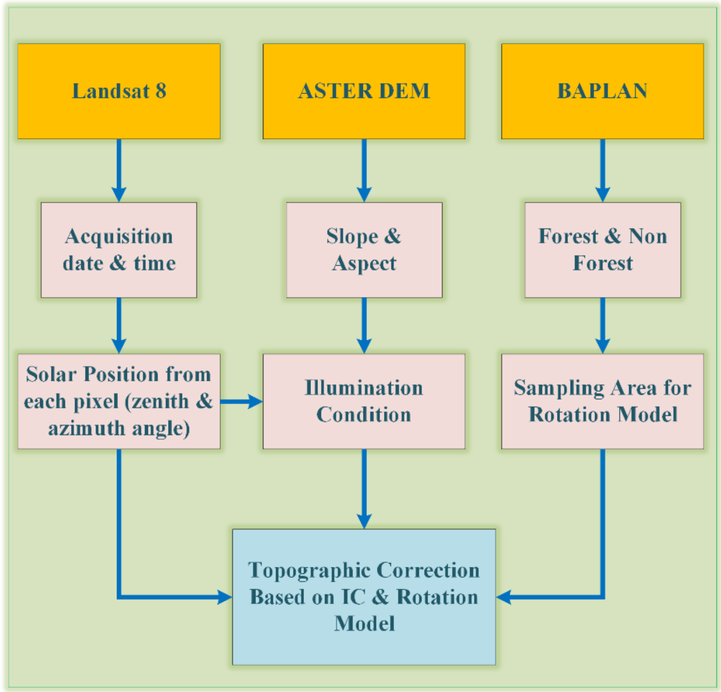

Figure 2. The topographic correction procedures

3.1 Mathematical model of solar position

Uncertainties in knowledge of the Sun position can increase to significant errors in shadow area estimation in rough terrain[7]. By calculating the solar position to estimate a zenith and azimuth angle each pixels of the Sun from a given date and time of Landsat 8 metadata, are estimated using the equations was adopted from[12] and suggested by [13].

First is to calculate the fractional year (γ):

$$\gamma=\frac{2\pi}{\sum d}*\left(d-1+\frac{hour-12}{24}\right) \tag{1}$$

where $\sum d$ is total number of days in year (366 for leap years instead 365 in the denominator). The fractional year (γ) is used to estimate the equation of time and the sun declination angle (radians). The equation of time and the sun declination angle are shown by Equation 2 and 3.

$$eqtime=229.18*(0.000075+0.001868\cos(\gamma)-0.032077\sin(\gamma)-0.014615\cos(\gamma)-0.040849\sin(2\gamma)) \tag{2}$$

and

$$\varphi=(0.006918-0.399912\cos(\gamma)+0.070257\sin(\gamma)-0.006758\cos(2\gamma)+0.000907\sin(2\gamma)+0.000907\sin(2\gamma)-0.002697\cos(3\gamma)+0.00148\sin(3\gamma) \tag{3}$$

According to equation 2, we can calculate time offset (minutes) as follows:

$$t_{offset} = eqtime + 4*latitude - 60*timezone \tag{4}$$

Time zone is according to hours from UTC, in Indonesia is divided into three zones: Western Indonesian Time = +7 hours, Central Indonesian Time = +8 and Eastern Indonesian Time = +9. Moreover, in order to calculate the sun hour angle (degrees) can be seen Equation 5:

$$\alpha=\frac{60hh+mn+sc/60+t_{offset}}{4}-180 \tag{5}$$

From equation 5, where hh is hours, mn is minutes and sc is seconds. Final step is to calculate the solar zenith and azimuth angle, which are derived from sun hour angle (α), latitude (degrees) and sun declination angle (φ). The equation is shown as follows:

$$\phi=\cos^{-1}(\sin lat \sin \varphi+ \cos lat \cos \varphi \cos \alpha) \tag{6}$$

and

$$\theta=180\pm\cos^{-1}\left(-\frac{\sin lat \cos \phi - \sin \varphi}{\cos lat \sin \phi}\right) \tag{7}$$

From Equation 6 and 7, where ϕ and θ are zenith and azimuth angle (^{o}), lat is latitude from each pixel (□).

3.2 Mathematical model of illumination condition

Illumination condition model was calculated by equation adopted from [10] as follows:

$$ic= \cos \phi \cos slope+ \sin \phi \sin slope \cos(\theta\text{-aspect}) \tag{8}$$

where slope and aspect of each pixel are derived from DEM respectively, the ϕ and θ are zenith and azimuth angles of the sun from each pixel at given date and time, respectively. The illumination condition is denoted as ic of each pixel at a given sun position and terrain attributes.

3.3 Sampling method

Sampling method is required to obtain area that can be used as sample areas for rotation model. Sample area are applied to calculate slope of linear regression between illumination condition and wavelength variables (i.e digital number or reflectance). To determine a sample area in Landsat 8 imagery can be calculated Normalized Difference Vegetation Index (NDVI) value as initial in term of forest. In general, dryland forest in Indonesia is dominated by montane forest. Montane forest commonly has high altitude and moderate slope, by combining NDVI value and montane forest can determine sample area. This study, to classify dryland forest is using land use classification map from BAPLAN 2016.

3.4 Empirical rotation model

According to illumination model, each wave-length of band Landsat 8 is corrected empirically, based on the sample area that used to calculate the slope of linear regression between illumination condition and wave-length variables, as follows [10]:

$$\lambda_{corrected}=\lambda_{uncorrected}-\beta(ic-\phi) \tag{9}$$

where $\lambda_{corrected}$ and $\lambda_{uncorrected}$ are wave-length variables after and before topographic correction and their also free from cloud and cloud shadow, β is the slope of linear regression between illumination condition of each pixel and wave-length variables and ϕ is cosine of zenith angle of the sun at given date and time.

4. RESULTS AND DISCUSSION

4.1 Illumination condition

Each of site area is calculated with different acquisition date and time based on information from Landsat 8 data to calculate illumination condition, respectively. Illumination condition from five the highest mountains in Indonesia based on illumination model are shown in Figure 3-5. The result is showing IC value ranges from −1 (minimum illumination) to +1 (maximum illumination).

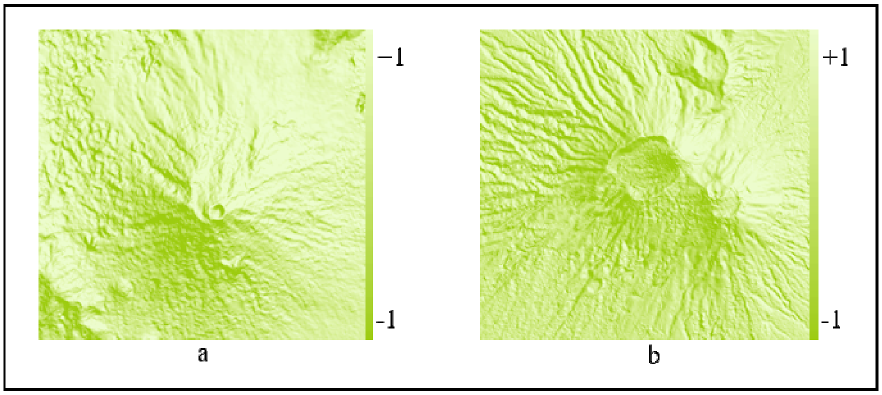

Figure 3: Illumination model of Mount Kerinci (a) and Mount Pangrango (b)

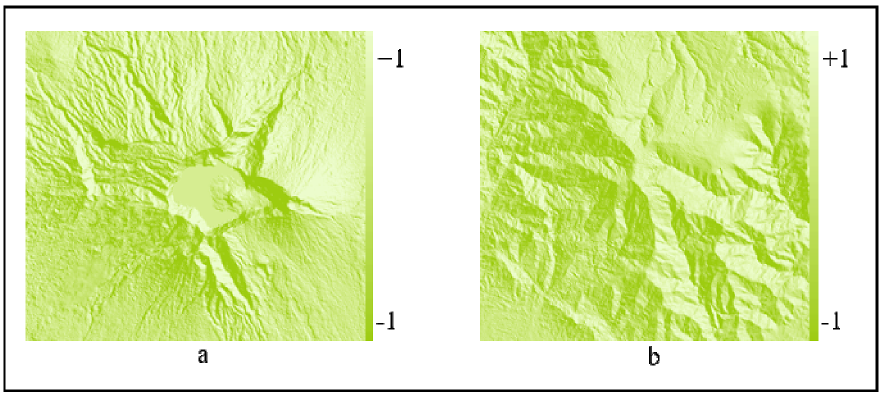

Figure 4: Illumination model of Mount Rinjani (a) and Mount Buyu Balease (b)

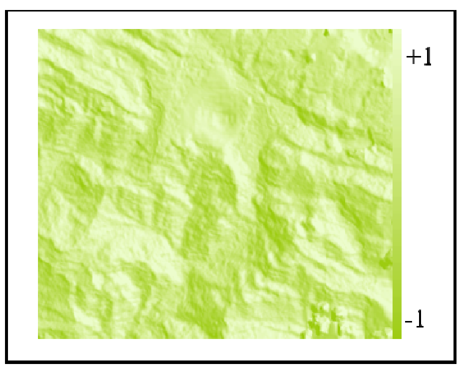

Figure 5: Illumination model of Mount Jayawijaya

4.2 Sampling Method

Sample area(s) are derived from combining NDVI value and land cover classification map (BAPLAN 2016). In this research many experimental to retrieve value of NVDI that can be used as base value to distinguish in term of forest. The result showed with more than 0.5 NDVI value can estimate a forest and also cloud free area as a constraint for topographic correction algorithm. By combining NDVI value 0.5 and land cover as forest (BAPLAN 2016), both of parameter are capable of determining sample area for rotation model in topographic correction algorithm. The parameter was applied to five highest mountains in Indonesia and the result showed sample(s) area are effectively for rotation model. The result of automated sampling area can be seen in Figure 6.

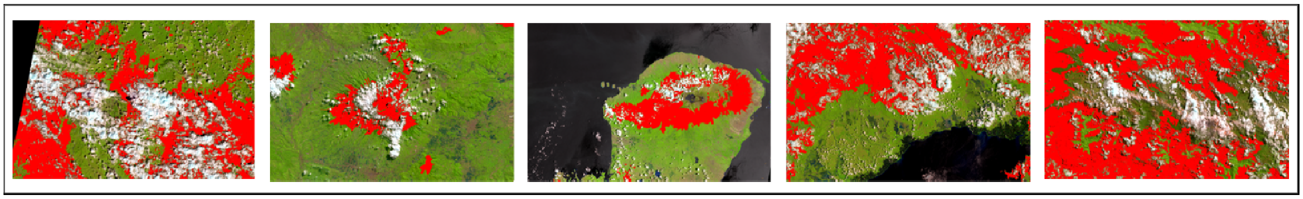

Figure 6. The highest mountains sample area from west to east Indonesia: Mount Kerinci, Mount Pangrango, Mount Rinjani, Mount Buyu Balease and Mount Jayawijaya.

4.3 Topographic correction effects on Landsat 8 imagery

According sample area(s) are used for rotation model, Figure 7 - 9 show the effects of topographic correction using the algorithm on the composites of SWIR (Band 6), NIR (Band 5) and red (Band 4) bands of Landsat 8. Composite of SWIR, NIR and red bands of Landsat 8 can used as vegetation analysis, which is can determine where area is classified as forest or not.

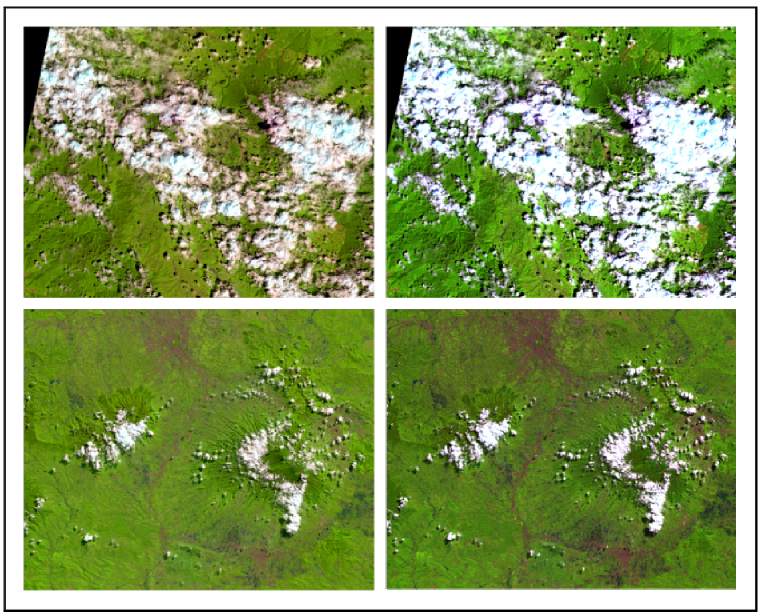

Figure 7. Composite of bands 6, 5, 4 before (left) and after (right) topographic correction of Landsat 8 imageries using empirical rotation model for window areas Mount Kerinci (top) and Mount Pangrango (bottom).

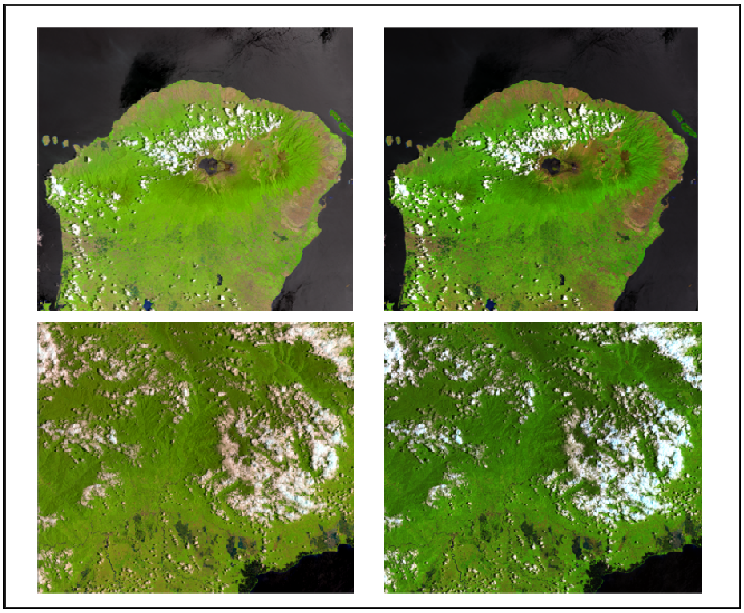

Figure 8. Composite of bands 6, 5, 4 before (left) and after (right) topographic correction of Landsat 8 imageries using empirical rotation model for window areas Mount Rinjani (top) and Mount Buyu Balease (bottom).

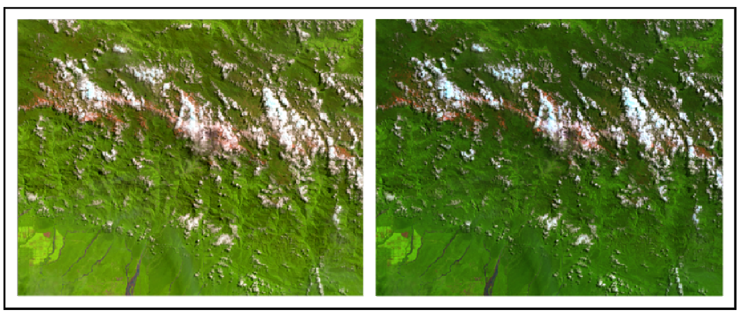

Figure 9. Composite of bands 6, 5, 4 before (left) and after (right) topographic correction of Landsat 8 imageries using empirical rotation model for window areas Mount Jayawijaya.

4.4 Correlation and Sensitivities of SWIR, NIR and Red bands

From Figure 10-14, in generally uncorrected reflectance images of Landsat 8 have correlation with illumination condition in range value 0.01 to 0.24. Furthermore, corrected reflectance images of Landsat 8 almost have not correlation with illumination condition with value less than 0.0000. Consistent correlation between uncorrected reflectance images have correlation with illumination condition in SWIR, NIR and red bands. In corrected reflectance images also are consistent have not correlation with illumination condition in SWIR, NIR and red bands.

Sensitivities of SWIR, NIR, red bands of Landsat 8 to topographic correction using the algorithm for study sites are relatively consistent. The strength sensitivities in SWIR and NIR bands otherwise red band has weak sensitivities (see Figure 10-14).

Findings from this study that red band has relatively weak correlation with illumination condition, thus not sensitive to the topographic correction. Further exploration regarding the insensitivity of red band should be carried out.

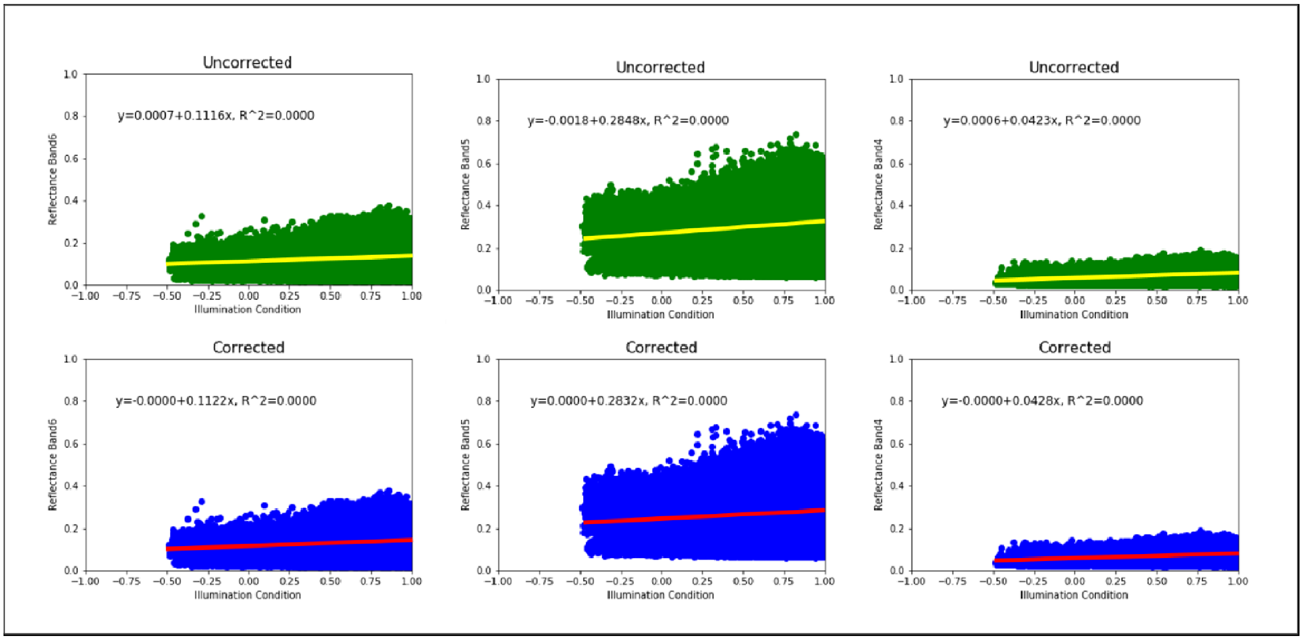

Figure 10. Composite of bands 6, 5, 4 before (Top) and after (bottom) topographic correction of Landsat 8 imageries using empirical rotation model for window areas Mount Kerinci.

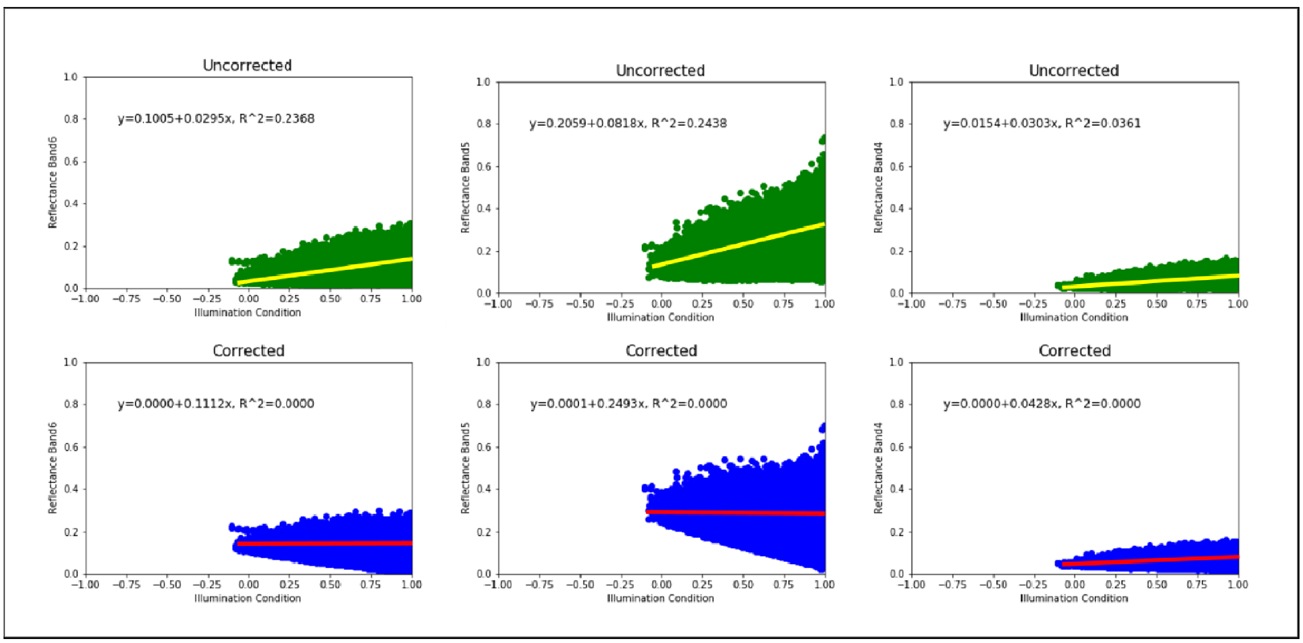

Figure 11. Composite of bands 6, 5, 4 before (Top) and after (bottom) topographic correction of Landsat 8 imageries using empirical rotation model for window areas Mount Pangrango.

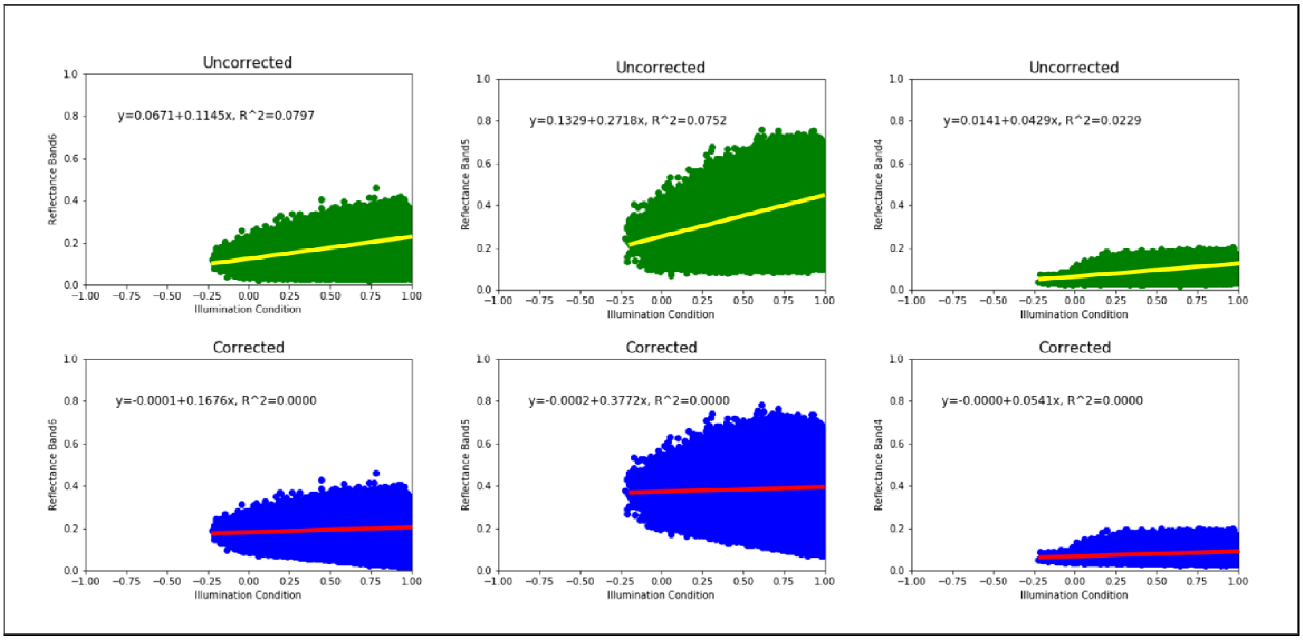

Figure 12. Composite of bands 6, 5, 4 before (Top) and after (bottom) topographic correction of Landsat 8 imageries using empirical rotation model for window areas Mount Rinjani.

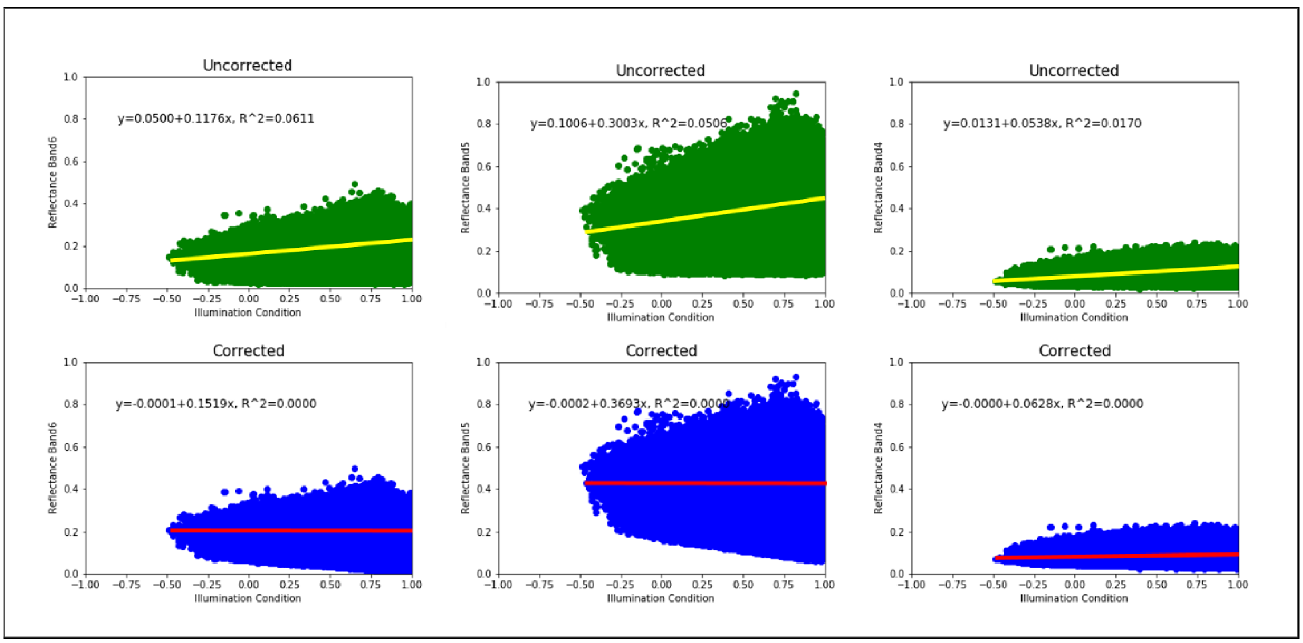

Figure 13. Composite of bands 6, 5, 4 before (Top) and after (bottom) topographic correction of Landsat 8 imageries using empirical rotation model for window areas Mount Buyu Balease.

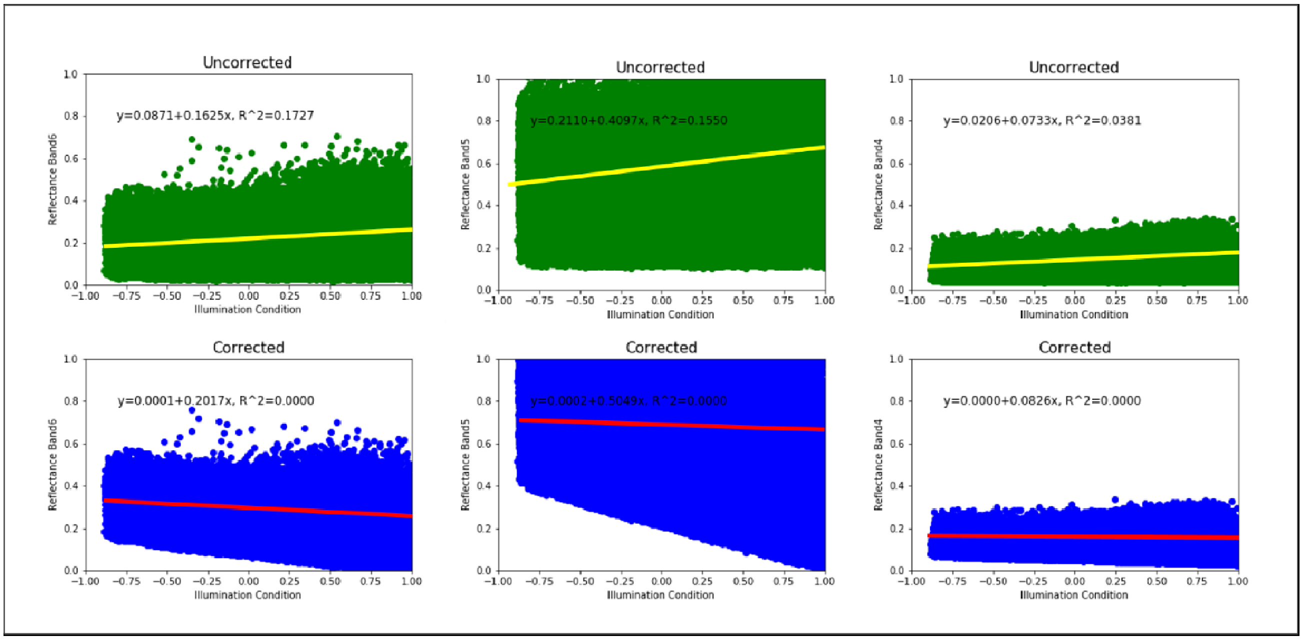

Figure 14. Composite of bands 6, 5, 4 before (Top) and after (bottom) topographic correction of Landsat 8 imageries using empirical rotation model for window areas Mount Jayawijaya.

5. CONCLUSION

1. The parameters of sample area were proven to be able to determine area for rotation model in topographic correction algorithm.
2. The automated topographic correction algorithm developed in this study combines illumination modelling, sampling method and empirical rotation model approaches, through explicit consideration on diversity of solar position and land use area.
3. The result relatively consistent correlation and sensitivities in SWIR, NIR and red bands from different acquisition dates and times.

ACKNOWLEDGMENT

We are grateful for the financial assistance from the International Partnership Programme (IPP) of the UK Space Agency through Ecometrica under Forests2020 Project (https://ecometrica.com/forests2020).

REFERENCES

[1] P. R. Coppin and M. E. Bauer, "Digital change detection in forest ecosystems with remote sensing imagery," *Remote Sensing Reviews,* vol. 13. 207-234, (1996).
[2] N. G. Steve, C. C. Nicholas, A. W. Michael, E. G. Sarah, and N. Trisalyn, "Multitemporal remote sensing of landscape dynamics and pattern change: describing natural and anthropogenic trends," *Progress in Physical Geography: Earth and Environment,* vol. 32. 503-528, (2008).
[3] M. A. Wulder, J. C. White, S. N. Goward, J. G. Masek, J. R. Irons, M. Herold*, et al.*, "Landsat continuity: Issues and opportunities for land cover monitoring," *Remote Sensing of Environment,* vol. 112. 955-969, (2008).

[4] A. Banskota, N. Kayastha, M. J. Falkowski, M. A. Wulder, R. E. Froese, and J. C. White, "Forest Monitoring Using Landsat Time Series Data: A Review," *Canadian Journal of Remote Sensing,* vol. 40. 362-384, (2014).
[5] M. Özdoğan, "Image Classifcation Methods in Land Cover and Land Use," in *REMOTELY SENSED DATA CHARACTERIZATION, CLASSIFICATION, AND ACCURACIES*. vol. 1, P. Prasad S. Thenkabail, (Ed.), ed Boca Raton, FL, USA: CRC Press, 231-241, (2016).
[6] L. R. Pettinger, "Digital classification of Landsat data for vegetation and land-cover mapping in the Blackfoot River watershed, southeastern Idaho," Report 1219, 1982.
[7] P. M. Teillet, "Overview of Satellite Image Radiometry in the Solar-Reflective Optical Domain," in *REMOTELY SENSED DATA CHARACTERIZATION, CLASSIFICATION, AND ACCURACIES*. vol. 1, P. Prasad S. Thenkabail, (Ed.), ed Boca Raton, FL, USA: CRC Press, 87-99, (2016).
[8] R. Gens and J. C. Rosselló, "Remote Sensing Data Normalization," in *REMOTELY SENSED DATA CHARACTERIZATION, CLASSIFICATION, AND ACCURACIES*. vol. 1, P. Prasad S. Thenkabail, (Ed.), ed Boca Raton, FL, USA: CRC Press, 133-140, (2016).
[9] S. Narumalani and P. Merani, "Digital Image Processing: A Review of the Fundamental Methods and Techniques," in *REMOTELY SENSED DATA CHARACTERIZATION, CLASSIFICATION, AND ACCURACIES*. vol. 1, P. Prasad S. Thenkabail, (Ed.), ed Boca Raton, FL, USA: CRC Press, 133-140, (2016).
[10] B. Tan, J. G. Masek, R. Wolfe, F. Gao, C. Huang, E. F. Vermote*, et al.*, "Improved forest change detection with terrain illumination corrected Landsat images," *Remote Sensing of Environment,* vol. 136. 469-483, (2013).
[11] B. Tan, R. Wolfe, J. Masek, F. Gao, and E. F. Vermote, "An illumination correction algorithm on Landsat-TM data," in *2010 IEEE International Geoscience and Remote Sensing Symposium*, 2010, pp. 1964-1967.
[12] NOAA, "*General Solar Position Calculations* "< http://www.esrl.noaa.gov/gmd/grad/solcalc/solareqns.PDF>, (1 November 2014).
[13] G. Prinsloo and R. Dablo, *Solar Tracking: High precision solar position algorithms, programs, software and source-code for computing the solar vector, solar coordinates & sun angles in Microprocessor, PLC, Arduino, PIC and PC-based sun tracking devices or dynamic sun following hardware*. (2nd Edition). South Africa: Stellenbosch: SolarBooks, 2015.