Comparison of Google Earth Engine (GEE)-based Machine Learning Classifiers for Mangrove Mapping

Muhammad Kamal (1), Ilham Jamaluddin (1), Artha Parela (1), Nur Mohammad Farda (1)

[1]Department of Geographic Information Science, Faculty of Geography, Universitas Gadjah Mada, Yogyakarta, 55281 Indonesia
Email: m.kamal@ugm.ac.id; ilham.jamaluddin@mail.ugm.ac.id; artha.p@mail.ugm.ac.id; farda@ugm.ac.id

KEY WORDS: Google Earth Engine, Mangroves, SVM, Random Forest, CART

ABSTRACT: Providing fast, up-to-date and accurate mangrove extent map is essential to support mangrove monitoring and management actions. Conventional scene by scene remote sensing image classification approach is inefficient and time consuming. Google Earth Engine (GEE) provides a cloud platform to access and seamlessly process large amount of freely available satellite images. The GEE also provides a set of the state-of-the-art classifiers for pixel-based classification that can be used for mangrove mapping. This study aims to compare the performance of three machine learning classifiers available in GEE, namely support vector machine (SVM), random forest (RF), and classification and regression trees (CART), for discriminating mangrove and non-mangrove objects. We used Landsat 8 OLI scene over Agats Papua area, Indonesia, as the main image for this study (path 102 row 64) acquired at October 19, 2014. A total of 838 points samples were collected representing mangroves (244), non-mangroves (161), water bodies (311), and cloud (122) class. These training areas were used by all classifiers in GEE to classify the image into targeted land cover classes. The classification results show that mangrove objects could be detected by all classifiers. To assess the accuracy of the map produced, the map results were compared to the visually interpreted mangrove map of the study area and previous mangrove map by Giri et al. (2011) as the references. From this comparison we found that SVM outperformed the classification results of CART and RF significantly. Most of the miss-classification were occurred at the shadow objects beneath the clouds on the image, where shadows were miss-classified as mangroves. This study shows the potentials of GEE for mangrove extent mapping. Future works need to focus on the sample refinement and testing the applicability of the classifiers at other mangrove sites.

1. INTRODUCTION

Mangroves are typical coastal vegetation that can be found in the intertidal area of coastal ecosystem. It grows and develop at the low wave energy where anaerobic substrate deposited and provides a link between land and sea ecosystem in the world's subtropics and tropics (Alongi 2002). Mangrove ecosystems that live in the intertidal areas have a variety of important roles that are beneficial to human and nature. Mangroves have many ecological and economic functions, including protecting coastal areas from ocean currents and winds, providing spawning grounds for marine organisms, maintaining the food cycle of marine organisms, maintaining the quality of coastal waters, shelter of various wild fauna and as a tourist attraction (Giri et al., 2011; Green et al 1998). At the same time, these ecosystem functions are threatened by anthropogenic disturbances such as mangrove forest clearing, land conversions, infrastructure development, aquaculture, etc. World Resources Institute (WRI, 2015) mapped and monitored the world mangrove change using Global Forest Watch (GFW), an online forest monitoring platform, and found that the world lost 192,000 hectares of mangroves from 2001 to 2012. It is a total loss of 1.38 percent since 2000 (or 0.13 percent annually). Of this figure, mangroves in Asia reported to have annual loss rates nearly double the global average due to the fast growing of aqua culture industry, agricultural expansion, oil-spills and other chemical pollution, and coastal and urban development. Therefore, it is required to have a rapid and efficient method for mapping mangrove cover status throughout the time to support management and conservation actions in this ecosystem.

The biggest challenge in mapping mangrove area is the condition of the mangrove ecosystem which is located in the intertidal area with unconsolidated sediments, making it difficult to access. In this case, remote sensing data can be applied to map the mangrove area because it can represent the mangrove object simultaneously in one image scene. However, the use of conventional remote sensing methods for large area mangrove mapping is inefficient. It is because the remote sensing processing for conventional mangrove mapping requires large volume of image data and large data processing resources. Recent technological developments help in the advancement of remote sensing processing approach, one of which is the advent of the Google Earth Engine (GEE) cloud platform. GEE is a platform that enable mapping and geospatial analysis to be performed using cloud computation, this includes information derivation from remote sensing images available in GEE (Butler, 2006). The GEE platform supports the processing of cloud-based image classifications that make it possible to implement machine learning algorithms in the classification process. The machine learning based classification system at GEE can be used as a basis for mapping various land cover including mangrove coverage.

Machine learning is computer-based machine learning that is widely used to build a system that able to learn and solve problems automatically. In the supervised classification, machine learning method can be an alternative algorithm used for classification with large dimension of input data input (big data). To use machine learning methods, training and testing data are needed; the training data is used to train the algorithm to recognize object, while testing data is used to determine the performance of the algorithms. The GEE platform has a variety of machine learning algorithms that can be used for supervised classification of image data. In this study, three machine learning algorithms were used to map mangrove extent, namely support vector machine (SVM), random forest (RF), and classification and regression tree (CART). This study aims to compare the performance of these three machine learning classifiers available in GEE for discriminating mangrove and non-mangrove objects.

2. METHODS

2.1 Study Site and Image Data

This study uses Landsat 8 OLI images of Asmat Regency, Papua Province, Indonesia (Figure 1). Asmat Regency was chosen because it is one of the areas that has large mangrove cover in Indonesia, thus making it easier for testing the image and algorithm performance. The Landsat image used are located at path 102 and row 64 acquired on October 24, 2014. This image is used as input data for mangrove mapping with three different machine learning methods. It has been calibrated up to top of atmosphere (ToA) reflectance correction level by GEE algorithm based on the procedure from Chander et al. (2009).

2.2 Machine Learning Algorithms

There are three machine learning algorithms used in this study; support vector machine (SVM), random forest (RF), and classification and regression trees (CART). All three algorithms will be tested for suitability and accuracy for mapping mangrove vegetation. The limitation of the study using these three methods is that the parameters used are the default parameters available in GEE. The brief explanation of each algorithms is as follow:

- SVM is an alternative algorithm that can be used to overcome various weaknesses in classification using remote sensing data (Huang, 2002). SVM is a classification algorithm that formulates a classification function based on an optimal hyperplane to separate classes from training data based on the process that is performed. The best hyperplane to separate two classes can be known by measuring the hyperplane's margin and finding the maximum value. Margin is the distance between hyperplane and the closest pattern of each class. This closest pattern is also called a support vector.

- RF is an algorithm developed from a decision tree based on multiclassification or regression based on many decision trees, where the random forest itself consists of several decision tree models (Jin et al., 2018). This algorithm builds statistical data (ensemble) from the decision tree to overcome the weaknesses of the decision tree algorithm related to sensitivity to training data (Breiman, 2001).
- CART is a classification algorithm based on the concept of entropy structure (decision tree) where the classification of objects is based on the characteristics that exist in a class (training data). It can also be said as an IF-THEN function in object classification (Li et al., 2011). This algorithm has a weakness that is high sensitivity to training data, changes in training dataset will cause different classification results (Bishop, 2006).

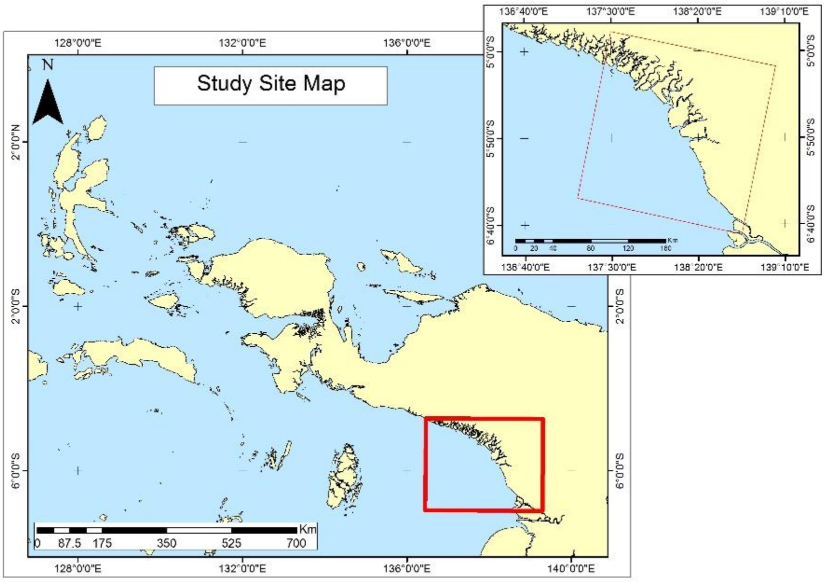


Figure 1. Study site: Asmat Regency, Papua Province, Indonesia.

2.3 Training Sample and Result Validation

The use of machine learning methods for mangrove mapping using supervised classification approach requires training samples for classification input. The sample model used in this study consisted of four landcover classes, namely mangroves (244 points), non-mangroves (161 points), water bodies (311 points), and clouds (122 points) (Figure 2). The sample model is used as input to guide the formation of decision tree in the classification process using machine learning methods. The selection of model samples in this study is based on visual interpretation of true color (RGB 432) and pseudo-color (RGB 567) image composition of Landsat. The use of both image color composition in the process of determining sample points is intended to make it easier to determine the decision making of each class.

The accuracy assessment of the three algorithms was carried out with (1) visual observations related to the detection of mangroves throughout the image scene and (2) quantitative accuracy assessment was carried out to determine the accuracy of mangrove identification of the three machine learning algorithms. The reference data in which the accuracy assessment will be based on was produced from visual interpretation of the Landsat data. Accuracy assessment is conducted by visually comparing the reference map with the mangrove mapping resulted from the classification of the three machine learning algorithms. Quantitative accuracy test is done by selecting validation samples on the reference map and then comparing it to the validation sample from the machine learning classification results. Based on the results of the accuracy assessment, the highest accuracy machine learning algorithms at GEE for mangrove vegetation mapping will be obtained.

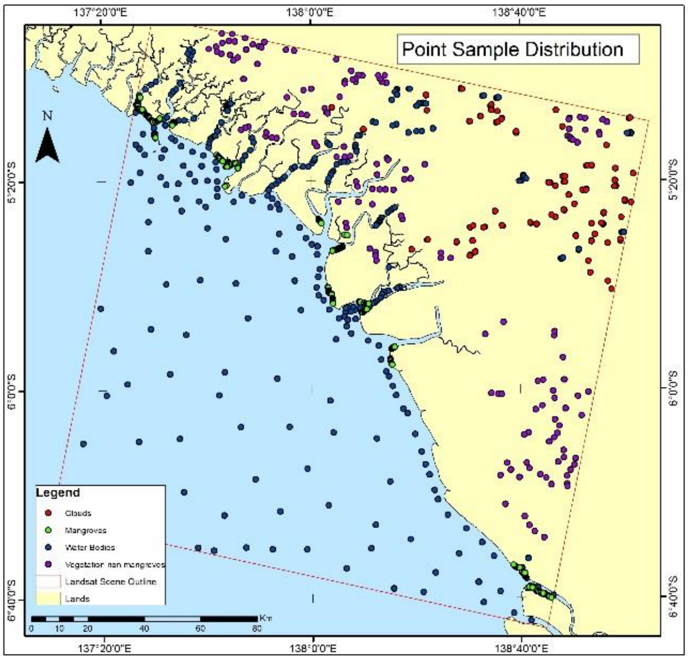


Figure 2. Point sample distribution.

3. RESULTS AND DISCUSSION

3.1 Machine Learning Algorithms

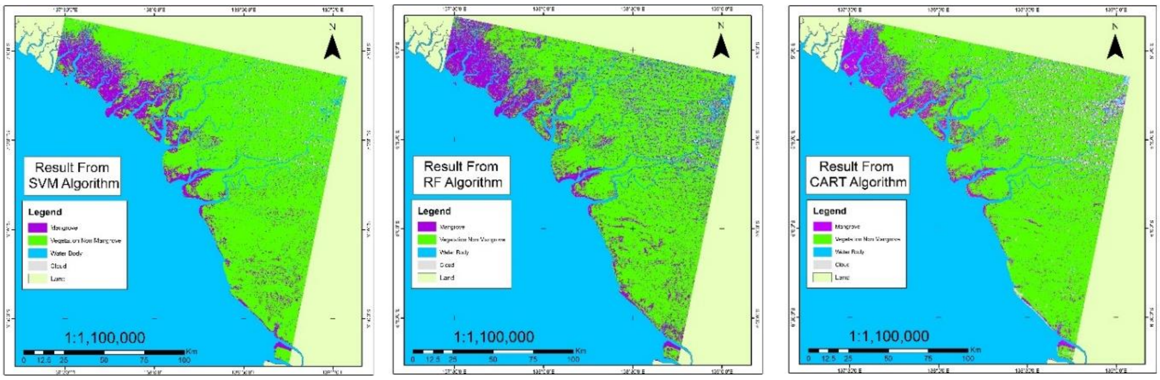
The script in GEE for mangrove mapping using machine learning algorithms generally starts from combining sample points in feature collection, followed making decision tree in the three machine learning algorithms which are then applied to the Landsat image. The bands used as input for the three machine learning algorithms were blue (450-510 nm), green (530-590 nm), red (640-670 nm), near infrared (850-880 nm), shortwave infrared1 (1570- 1650 nm), and shortwave infrared2 (2110-2290 nm) bands. Generally, the image classification script sequence procedure in SVM, RF, and CART algorithms is similar. The only different is the machine learning algorithm itself. A summary of the classification procedure of the three algorithms in mangrove mapping using GEE is explained in Table 1.

Table 1. Classification procedure summary.

Steps	Task Description
1	To define Landsat 8 OLI variable
2	To choose Landsat 8 OLI scene with parameter cloud cover percentage, region of interest, date acquired, and bands used
3	To change pixel value to be median value
4	To combine sample point in feature collection
5	To get pixel value every sample form Landsat 8 OLI
6	To define SVM algorithm or RF algorithm or CART algorithm
7	To run SVM algorithm or RF algorithm or CART algorithm to build tree scheme
8	To classify Landsat 8 OLI path 102 row 64 based on tree scheme
9	Set zoom level and map center
10	Export classification result

3.2 Mangrove Delineation Results from Machine Learning Algorithms

The machine learning algorithms of SVM, RF, and CART on GEE successfully identified the mangrove vegetation from Landsat image. The SVM classification result shows mangrove objects are located in the intertidal areas of Agats Regency. However, there are some miss-classifications on pixels that have a mangrove-like reflectance value, such as inland vegetation that are close to water and inland vegetation that are located close to cloud shadows. The result of mangrove mapping using the RF algorithm shows delineation of mangrove vegetation that is almost the same as the results of the SVM method. The difference is that there are more miss-classifications on land vegetation near the aquatic object. While the result of mangrove mapping with CART shows a lot of miss-classification caused by vegetation objects that are in the shadow of clouds, which was detected by the algorithm as mangroves. The results of mangrove mapping using the three machine learning algorithms is shown in Figure 3.

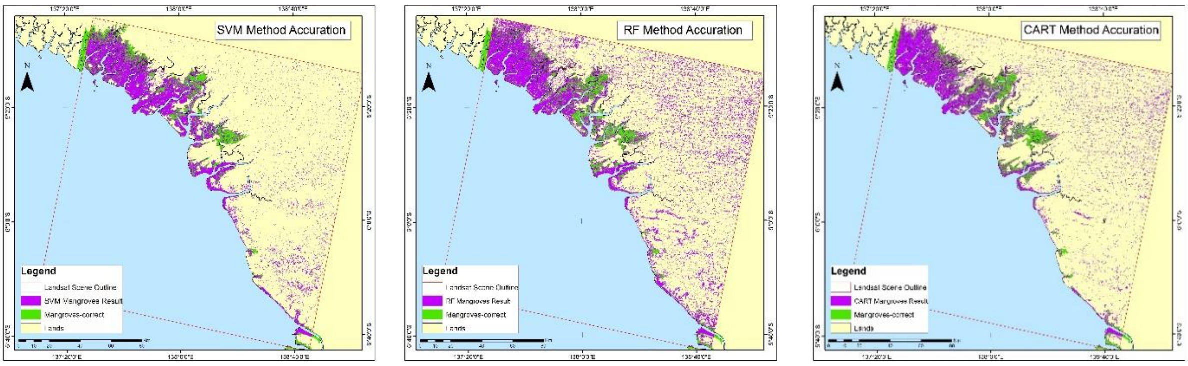


(a) (b) (c)

Figure 3. Mangrove delineation result from (a) SVM, (b) RF, and (c) CART algorithms.

3.3 Accuracy Assessment and Validation

A visual accuracy assessment was performed by comparing the results of mangrove delineation using machine learning algorithms with a reference map resulted from visual image interpretation. The visual comparison shows that SVM was more optimum in identifying mangrove vegetation compared to the RF and CART. This was indicated by the minimum miss-classification found in the SVM result, and the degree of classification agreement between SVM and the reference map (Figure 4).



(a) (b) (c)

Figure 4. Mapping accuracy assessment between reference map and classification results from (a) SVM, (b) RF, and (c) CART.

When we focus on a smaller area, it is obvious that the result SVM algorithm provided better delineation of mangrove objects than the RF and CART (Figure 5). In the southeastern part of this area mangroves could be identified by the SVM algorithm, but not so much identified by

RF or CART. In addition, mangroves located around the mainland (eastern part) could be identified better from SVM. In the western part, there are lots of miss-classification resulted by RF algorithm where the terrestrial object is classified as mangrove. Likewise, there were also miss-classification produced by RF in the land area which were contained by mangroves. This miss-classification can be one of the causes of the classification results with the RF algorithm has the lowest accuracy value among the three algorithms.

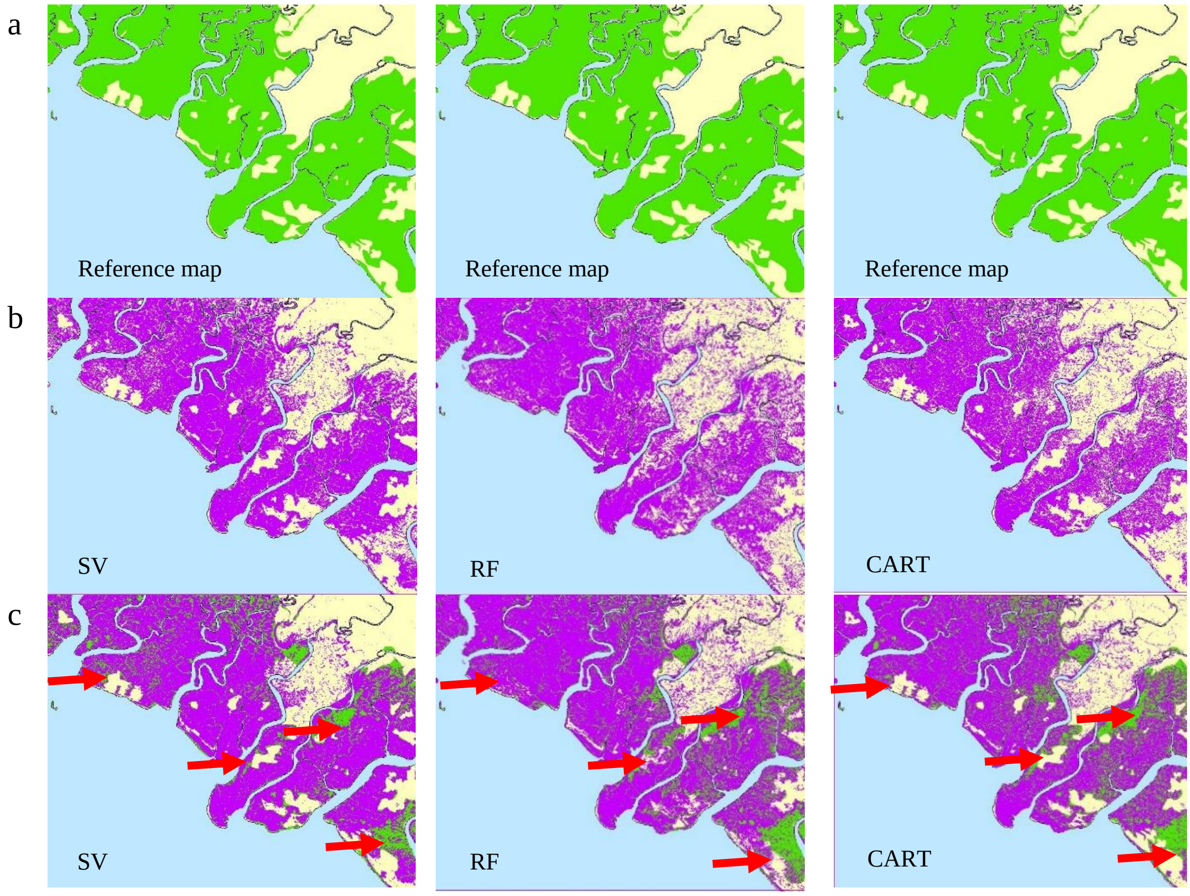


Figure 5. Comparison of classification results among SV, RF, and CART: (a) reference map, (b) classification results, and (c) reference-classification comparison.

Quantitative accuracy assessment was conducted to the results of mangrove mapping using SVM, RF, and CART. This accuracy assessment was intended to determine the absolute accuracy value of mangrove delineation based on the results of the three algorithms. This accuracy assessment process started with the determination of mangrove vegetation validation samples on the reference map. The validation sample used consisted of 57 samples representing all mangrove vegetation in the study area (Figure 6). The validation sample was then compared with the results of mangrove mapping using SVM, RF, and CART. Based on the results of the calculation of accuracy, the results of mangrove mapping using the SVM method showed an accuracy value of 66.67%. It means that there are 33.33% of the total study area that mangroves should be classified as non-mangroves, or vice versa. RF accuracy calculation results show the accuracy of 57.89% and 42.11% are miss-classified pixels. The results of accuracy calculation of CART show the accuracy of 61.40%, while 38.60% were miss-classified pixels. The results of the quantitative accuracy assessment measurements are shown in Table 2.

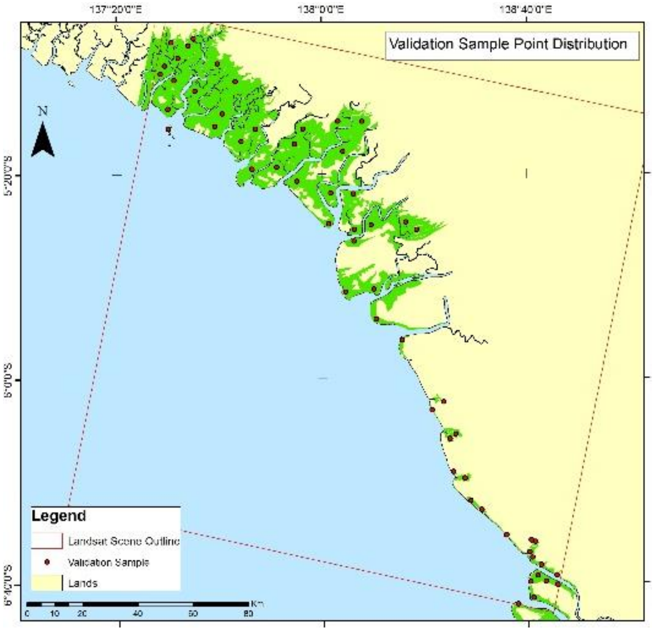


Figure 6. Validation sample distribution.

Table 2. Mangrove delineation accuracy result.

Accuracy	SVM	RF	CART
Mangrove Accuracy	66.67%	57.89%	61.40%
Miss-classification	33.33%	42.11%	38.60%

4. CONCLUSION

Based on the results of this research, the three machine learning algorithms that are applied to GEE produced mangrove vegetation map with the accuracy of above 50%. According to the results of the visual accuracy assessment, the SVM algorithm shows high consistency and accuracy in mapping mangrove vegetation. While the quantitative accuracy assessment results obtained the value of mangrove vegetation accuracy for SVM, RF, and CART algorithm were 66.67%, 57.89%, and 61.40%, respectively. The main finding of this research is that SVM algorithm is the best machine learning algorithm in mapping mangrove vegetation using GEE. Future works will be focus on applying SVM algorithm for mangrove mapping at larger area and testing the consistency of SVM algorithm result for other image datasets.

5. ACKNOWLEDGEMENT

The authors would like to thank Department of Geographic Information Science, Faculty of Geography, Universitas Gadjah Mada for providing support for this research, Google Earth Engine for providing image cloud computing platform, and USGS for providing free access of Landsat 8 OLI data for this research.

6. REFERENCES

Alongi, DM 2002, 'Present state and future of the world's mangrove forests', *Environmental Conservation*, vol. 29, pp.331-349.

Bishop, C. M. 2006. Pattern Recognition and Machine Learning. New York: NY: Springer.

Breiman, L. 2001. ‘Random Forests’, *Machine Learning*, 45,5-32.

Butler, D. 2006. ‘The Web-Wide World: Virtual Globe’, *Nature*, 439, 776-778.

Chander, G, Markham, BL & Helder, DL 2009, 'Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors', *Remote Sensing of Environment*, vol. 113, pp.893-903.

Giri, C, Ochieng, E, Tieszen, LL, Zhu, Z, Singh, A, Loveland, T, Masek, J & Duke, N 2011, 'Status and distribution of mangrove forests of the world using earth observation satellite data'. *Global Ecology and Biogeography*, vol. 20, pp. 154-159.

Green, EP, Clark, CD, Mumby, PJ, Edwards, AJ & Ellis, AC 1998, 'Remote sensing techniques for mangrove mapping'. *International Journal of Remote Sensing*, vol. 19, pp. 935-956.

Huang, C. 2002, 'An Assessment of Support Vector Machines for Land Cover Classification'. *International Journal of Remote Sensing*, vol.23, no.4, 725-749.

Jin, Y. Lu, X. Chen, Y. and Liang, X. 2018. 'Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong', *International Journal of Remote Sensing*, VOL. 39, NO. 23 8703–8723.

Li, D. Liu, Y. and Chen, Y. 2011. 'Classification Methods of Remote Sensing Image Based on Decision Tree Technologies.' *Computer and Computing Technologies in Agriculture IV* 353–358.

World Resource Institute 2015, 'Satellite Data Reveals State of the World's Mangrove Forests', available online at https://www.wri.org/blog/2015/02/satellite-data-reveals-state-world-s-mangrove-forests, accessed at 7 September 2019.