

TECHNICAL REPORT ON FOREST-COVER CHANGE DETECTION
IN THE PREY LANG PROTECTED AREA OF CAMBODIA



A study authored by the Ecology Program of Jesuit Service Cambodia (JSC)
in partnership with Cambodian Youth Network (CYN)
Released August 2020

Table of Contents

Acronyms	1
Introduction.....	2
Materials.....	3
Raw Data and Software Used.....	3
Methods.....	3
Pre-processing method	4
Cloud Masking and Image Mosaicking.....	4
Creating Training and Validation Data	5
Supervised Classification	6
Accuracy Assessment	7
Confusion Matrix.....	7
Overall accuracy.....	8
Producer’s accuracy	8
User’s accuracy	8
Kappa coefficient	8
Results.....	19
Year 2000 to 2018	20
Year 2000 to 2019	21
Discussion	26
Conclusion	26
References.....	27

List of Tables

Table 1	Downloaded satellite images used for this project	3
Table 2	Codes used for training data to perform direct-change classification	5
Table 3	Attribute codes used for validation points	6
Table 4	Codes used in setting up <i>R Studio</i>	7
Table 5	Sample confusion matrix.....	8
Table 6	Accuracy assessment for the change detection between years 2000 and 2018.....	21
Table 7	Computed areas for each class for the period ca. 2000 to ca. 2018	21
Table 8	Accuracy assessment for the change detection between years 2000 and 2019.....	22
Table 9	Computed areas for each classification for the period ca. 2000 to ca. 2019	22

List of Figures

Figure 1	Flow chart of processing and creating change detection maps.	10
Figure 2	Settings used in pre-processing <i>Landsat 5 and 8</i> dataset using the <i>Semi-Automatic Classification Plugin</i>	11
Figure 3	Preprocessed <i>Landsat 5</i> (year 2000) image displayed as <i>True Color Composite</i> in QGIS.....	11
Figure 4	Preprocessed <i>Landsat 8</i> (year 2018) image displayed as <i>True Color Composite</i> in QGIS.....	12
Figure 5	Clipped satellite image showing Prey Lang boundary (March 2000)	12
Figure 6	Clipped satellite image showing Prey Lang boundary (March 2018)	13
Figure 7	Clipped satellite image showing Prey Lang boundary (February 2019)	14
Figure 8	Clipped satellite image showing Prey Lang boundary (February 1999)	15
Figure 9	Clipped satellite image showing Prey Lang boundary (February 2018)	16
Figure 10	Interface of <i>cloud masking tool</i> of <i>SCP plugin</i> in QGIS	16
Figure 11	Interface of <i>mosaic band sets tool</i> of <i>SCP plugin</i> in QGIS.....	17
Figure 12	Generated training data (shown as yellow polygons) for year 2000-2018 change detection	17
Figure 13	Generated training data (shown as yellow polygons) for year 2000-2019 change detection	18
Figure 14	Validation points (shown as orange points) for year 2018	18
Figure 15	Validation points (shown as pink points) for year 2019	19
Figure 16	Computation for the overall accuracy	19
Figure 17	Computation for the producer's accuracy.....	19
Figure 18	Computation for the user's accuracy	20
Figure 19	Computation for the kappa coefficient	20
Figure 20	Change detection map for year 2000-2018 generated from <i>R Studio</i>	23
Figure 21	Change detection map of Prey Lang for the period ca. 2000 to ca. 2018.....	24
Figure 22	Change detection map for year 2000-2019 generated from <i>R Studio</i>	25
Figure 23	Change detection map of Prey Lang for the period ca. 2000 to ca. 2019.....	26

Acronyms

CCIPh	– Center for Conservation Innovation Ph
DN	– Digital Numbers
OLI	– Operational Land Imager
QGIS	– Quantum GIS
ROI	– Regions of Interest
SCP	– Semi-Automatic Classification Plugin
NASA	– National Aeronautics and Space Administration
QGIS	– Quantum GIS
SCP	– Semi-Automatic Classification Plug-in for QGIS
R	– Refers to an open-source statistical software package
TIRS	– Thermal Infrared Sensor
TOA	– Top of Atmosphere
USGS	– United States Geological Survey

Acknowledgement

We wish to acknowledge the technical support given by Center for Conservation Innovation Ph and its technical staff, as follows: Dr. Oliver G. Coroza, geospatial specialist, for his project leadership and technical review; Ms. Kristine Andaya, geographer, who spent countless hours on the image processing; Engr. Regine Escala, agricultural and biosystems engineer, who documented the process; and Engr. Angelica Kristina Monzon-Jaojoco, geodetic engineer, who ensured quality of output. We also acknowledge the open source makers of *QGIS*, its *SCP plugin*, *Random Forest*, *R statistics*, *R Studio* and open data distributors such as USGS/ NASA, Google Earth, Bing Maps, ArcGIS Earth and Open Development Cambodia.

Technical Report on Forest-Cover Change Detection in the Prey Lang Protected Area of Cambodia

Introduction

Prey Lang, which literally means “*our forest*”, is situated on the west of the Mekong River in the northern part of Cambodia. According to Prey Lang Community Network (2019), it is the “last major lowland rainforest on the Southeast Asian mainland”, and has a forest cover of approximately 5,000 square kilometers. It has been reported that it is a home to 250,000, mostly indigenous people, and a sanctuary to 530 plant and tree species, as well as, to 393 animal species. There are major threats to the forest, which is why it needs to be monitored and protected.

In line with this, forest-cover change detection maps are essential information in support of monitoring and managing the area. This technical report aims to describe the data and method in creating and analyzing the change in the forests of Prey Lang applied by the Center for Conservation Innovation Ph, technical resource commissioned by JSC Ecology Program. The original timeframe of change detection is between the years 2000 and 2018, but later when better satellite imagery (in terms of lesser cloud cover) was collected a circa 2019 became available. This extended the amount of time to complete the project, but it was worth it due to the recently dated good results.

The output of this project should be able to provide the answers to us to the questions of how much forest was lost and how much forest has recovered or has been gained back between the two time periods mentioned above, i.e., 2000-2018 and 2000-2019. The expected outputs are a change detection map and corresponding statistics on: forest loss and gain; current forest extent within the boundary of the reserve; and data uncertainty of the produced map in terms of its accuracy, as interpreted from satellite imagery.

Generating the outputs has taken awhile due to: processing of two time periods; the large areal size of the protected area about 433,388 has.; confusion with plantation forest against the natural forest; and availability of only secondary data for training and validation purposes, as opposed to getting field data, which was constrained by budget limitations. In the following paragraphs, we describe and explain the materials used and methods applied, the results and corresponding discussions of these, and some conclusions.

Materials

In this section, we describe the raw data we used in terms of satellite sensor, their resolution and date of collection. Because of the desire to have this easily replicated, we resorted to using non-commercial and open-source software packages.

Raw Data and Software Used

The satellite images for years 1999, 2000, 2018, and 2019 were downloaded from <https://earthexplorer.usgs.gov/>. Table 1 shows the details of the downloaded satellite images.

Table 1. Downloaded satellite images used for this project.

Date	Sensor	Spatial Resolution
February 20, 1999	Landsat 5	30-m
March 26, 2000	Landsat 5	30-m
February 8, 2018	Landsat 8	30-m
March 12, 2018	Landsat 8	30-m
February 27, 2019	Landsat 8	30-m

The downloaded images were then pre-processed using the QGIS software and were interpreted or classified accordingly using the *Semi-Automatic Classification Plugin (SCP)* of QGIS, *R* statistical package and *R Studio* to interface with R as its console for editing, executing codes, plotting of results and debugging.

Data for training the software to recognize the classification categories (e.g., forest and non-forest) were derived from secondary data such as published forest cover data of the Democratic Kampuchea and high-resolution images from Google Earth Pro. Similar data sources were also tapped to obtain classification categories for validation purposes to support accuracy assessments of the classification results or elected output.

Methods

Under this section we explain the methods executed in generating the expected outputs of the project through remote sensing techniques. Figure 1 is a graphic visual of the process flow in creating a change detection map. In the following subsections, we describe and explain the methods in the order illustrated in Figure 1: preprocessing method; cloud masking and image mosaicking; training and validating data; supervised classification; and accuracy assessment.

Preprocessing Method

Preprocessing is necessary because remote sensing data comes in different levels of readiness for analysis. Products from USGS come in quantized and calibrated scaled Digital Numbers (DN) representing multispectral image data acquired by both the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS), which is not yet ready for classification. Moreover, these products are delivered in 16-bit unsigned integer format, which can be rescaled to the Top of Atmosphere (TOA) reflectance and/or radiance using radiometric rescaling coefficients provided in the product metadata file (MTL file).

In QGIS, the *Semi-Automatic Classification Plugin* (SCP) was used to preprocess the raw data. Figure 2 shows the *SCP* window, which appears after clicking the *Preprocessing* icon, and shows the settings used in preprocessing the image. The *Landsat* dataset can be loaded by navigating to the directory where the dataset is stored on the “*Directory containing Landsat bands*” input area. In the same folder directory, the *Landsat* MTL file was selected on the input space of “*Select MTL file (if not in Landsat directory)*”. After the dataset was loaded and the proper settings were selected, the preprocessing was started by clicking the “*Run*” button at the lower right corner of the *SCP* window.

Once completed, the output images were added to the *Layers Panel* of QGIS. Figure 3 and Figure 4 show the preprocessed *Landsat 5* (year 2000) and *Landsat 8* (year 2018) images, respectively. The clipped images of the Prey Lang boundary for different time frames are shown in Figure 5, Figure 6, and Figure 7. Upon reviewing the preprocessed images, we observed that cloud cover was present on both March 2000 and March 2018 images. Due to cloud cover, the clouds were masked in the March 2000 and March 2018 images, and then mosaicked to the satellite images from February 1999 (Figure 8) and February 2018 (Figure 9), respectively.

Cloud Masking and Image Mosaicking

To perform cloud masking to remove unwanted cloud cover, the *cloud masking tool* of *SCP plugin* in QGIS was used. This tool creates a masked band output for each band of a band set. Figure 10 shows the *cloud masking tool* of *SCP plugin* interface. After applying the cloud mask to March 2000 and March 2018 images, these were then mosaicked to the February 1999 and February 2018 images, respectively, using the *mosaic band sets tool* of *SCP plugin* in QGIS (see Figure 11).

The resulting mosaicked images of 2000 and 2018 were then stacked together, using the *band set tool* of the *SCP plugin*, to generate a multi-date stack image. The same multi-stacking was done for the 2000 and 2019 timeframe. After stacking the images, the training data were then ready to be created.

Creating Training and Validation Data

To create the training data, stratified sampling of a group of pixels, demarcated as a polygon, on the multi-temporal stack image, which should homogeneously represent a specific class they identify with, e.g., "forest", "non-forest" or "forest to non-forest." The decision which similar pixels can be grouped as a selection is guided by supplementary data such as the secondary data, available as a forest cover map, or from Google Earth or by visual inspection of the stacked image based on familiarity with the area. Resource-wise it was unfeasible to do primary data collection of the ground situation. Hence, to gain familiarity the analyst resorted to consulting secondary forest information about the study area. This was based on contemporary and historical data from forest cover maps generated by Open Development Cambodia (2019)¹ supplementing the secondary data gathered from Google Earth. The latter provided additional information on where plantations are possibly located.

A new vector layer was created for the selected training data to group them as separate polygons for depicting the identified classes. Table 2 shows the codes used for their attributes. These polygons are, alternatively, called regions of interest (ROIs) to use in the direct-change classification method.

Table 2. Codes used for training data to perform direct-change classification.

Code for Change	Classification
11	Forest – Forest
12	Forest – Non-Forest
14	Forest - Plantation
21	Non-Forest – Forest
22	Non-Forest – Non-Forest
33	Water Body

Initially, as a rule of thumb, at least a number of 50 training data were determined for each classification category. However, since in the earlier iterations an over classification occurred, the number was refined and adjusted. A total of 282 training polygons were used for year 2000-2018, as well as, for year 2000-2019. Figure 12 and Figure 13 show a portion of training data used for change detection in 2000-2018 and 2000-2019, respectively.

¹ The term "forest" takes on the definition for forest, as adapted by Open Development Cambodia (2019) in generating their maps. Open Development Cambodia (2019) indicated that they followed "definition of Forestry Administration's Cambodia Forest Cover publication dated June 2008."

The validation data were selected randomly over the study area. Selection is derived from the imagery or forest or land cover, as secondary data, acquired at or near the same time as the classified image and by using the attribute codes and their corresponding classes shown in Table 3. When the cover data are insufficient, assessment is done on the current image to check whether we can directly obtain validation data guided with any adjacent information on the vegetation cover. At least 50 validation points per class were created as a rule of thumb. Figure 14 and Figure 15 show a portion of the validation points used for year 2018 and year 2019, respectively.

Table 3. Attribute codes used for validation points.

FNFCODE	Classes
1	Forest
2	Non-Forest
3	Water bodies

After generating the training data and identifying validation points, supervised classification was then performed using *Random Forest* algorithm (Breiman, 2001), as classifier, through *R* (R Core Team, 2016) and its interface *R Studio* (R Studio Blog, 2011). *Random Forest* has a machine learning algorithm, a subset of artificial intelligence (Copeland, 2016).

Supervised Classification

For supervised classification, the *R-3.4.0-win* and *R Studio-1.0.143* were installed. After opening the *R-Studio*, the working directory of processing was set, the required packages were installed, and the libraries of the installed packages were loaded by writing the codes shown in Table 4.

After setting up the *R-Studio*, the code for *Random Forest* was used to run this classifier to do the direct classification for change. The resulting stack images, mentioned above, and their corresponding training data were used as input into the *Random Forest classifier* to generate the change detection maps. The process to produce these maps is an iterative process and is usually dictated by the level of accuracy of the classification. For instance, further iteration was needed by editing the form of the training data and changing them from bigger polygons into smaller ones, to produce change detection maps with higher accuracy.

Table 4. Codes used in setting up *R Studio*.

Function	Code
Sets up the working environment	<code>getwd()</code>
	<code>setwd("C:/R working directory")</code>
Installs the required packages	<code>install.packages("ggplot2")</code>
	<code>install.packages("RColorBrewer")</code>
	<code>install.packages("sp")</code>
	<code>install.packages("maptools")</code>
	<code>install.packages("randomForest")</code>
	<code>install.packages("raster")</code>
	<code>install.packages("rgdal")</code>
Loads the libraries of the installed packages	<code>library("ggplot2")</code>
	<code>library("RColorBrewer")</code>
	<code>library("sp")</code>
	<code>library("maptools")</code>
	<code>library("randomForest")</code>
	<code>library("raster")</code>
	<code>library("rgdal")</code>

Accuracy Assessment

After generating the change detection maps from the classified images, their accuracy can be assessed. Accuracy assessment is performed by comparing the map, created through a remote sensing method just described above, to a reference map obtained from a different information source. One of the primary purposes of accuracy assessment or error analysis is to permit quantitative comparisons of different interpretations of images into classes. Classifications done from images acquired at different times, processed through different remote sensing procedures, or produced by different individuals can be evaluated using a pixel-by-pixel or point-by-point comparison with a reference image or map. With the limited resource and budget constraint, validation samples were created for 2018 and 2019 images only.

Confusion Matrix

A *confusion matrix* or *error matrix* is usually used as an illustrative and quantitative method of characterizing or assessing image classification accuracy (Congalton, 1991). It is a table that shows correspondence between the classification result and a reference image or map. To create the confusion matrix, the validation points are the ground-truth data needed. For example, Table 5 shows the confusion matrix generated for year 2019.

Table 5. Sample confusion matrix.

Classified Image	Reference Image			Total
	Forest	Non-Forest	Water	
Forest	106	6	0	112
Non-Forest	3	98	1	102
Water	0	0	66	66
Total	109	104	67	280

Four types of accuracy were then computed, namely *overall accuracy*, *producer's accuracy*, *user's accuracy*, and *kappa coefficient*, using the confusion matrix.

Overall accuracy

The *overall accuracy* of the classification map is determined by dividing the total correct pixels (sum of the major diagonal) by the total number of pixels in the error matrix (N). It measures the accuracy of the entire image without reference to the individual class categories. It is sensitive to differences in sample size, and biased towards classes with larger samples. The problem with the overall accuracy is that the summary value is an average. It does not reveal if error was evenly distributed between classes or if some classes were really bad and some really good, which is why other types of accuracy should be included in the accuracy assessment. Figure 16 shows how to compute for the overall accuracy.

Producer's accuracy

The *producer's accuracy* is the total number of correct pixels in a class category that is divided by the total number of pixels of that class category as derived from the reference data. This measure indicates the probability of a reference pixel being correctly classified and is a measure of omission or exclusion error. The producer or the analyst of the classification is interested in how well a certain area, represented in the image as a group of pixels, can be classified. Figure 17 shows how to compute for the producer's accuracy.

User's accuracy

The *user's accuracy* is the total number of correct pixels in a class category that is divided by the total number of pixels that were actually classified in that category. The result is a measure of commission or inclusion error. This measure is the probability that a pixel classified on the map actually represents that category on the ground. Figure 18 shows how to compute for the user's accuracy.

Kappa coefficient

Kappa analysis yields a statistic K_{hat} , which is an estimate of Kappa. It is a measure of agreement between the classification map, derived from the image processing of remote sensing, and the reference map, as indicated by the major diagonal of the matrix, and the chance agreement, which is indicated by the row and column totals, called as *marginals*. The level of agreement expresses the degree of accuracy. A Kappa of 0.8 or above is considered a good classification, while 0.4 or below is considered poor (Penn State Science 2011). Figure 19 shows a sample computation of the kappa coefficient for year 2019.

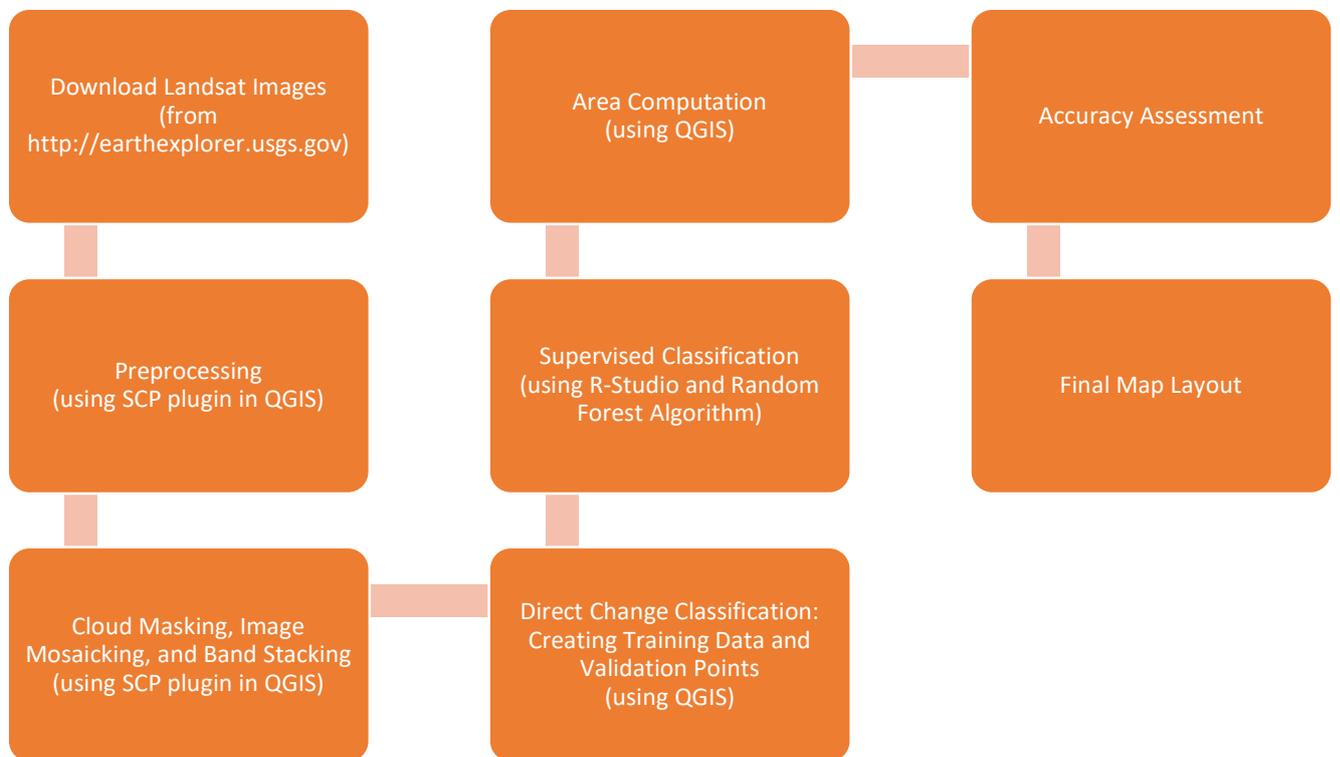


Figure 1. Flow chart of processing and creating change detection maps.

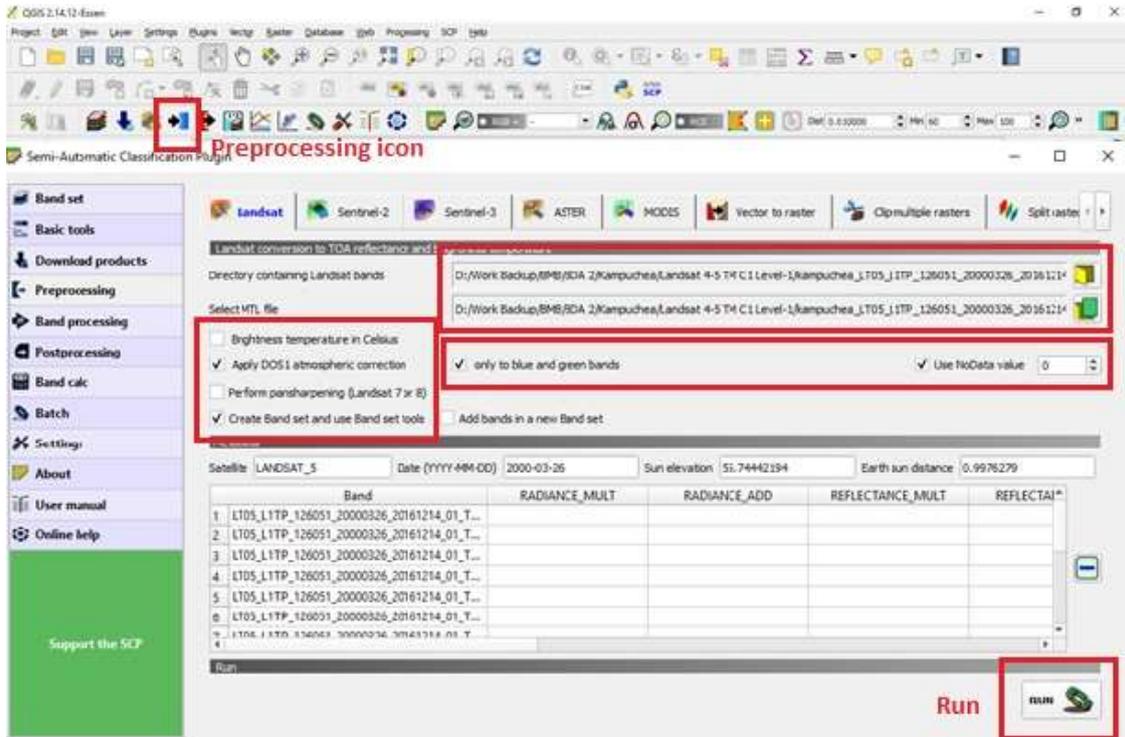


Figure 2. Settings used in pre-processing *Landsat 5 and 8* datasets using the *Semi-Automatic Classification Plugin*.

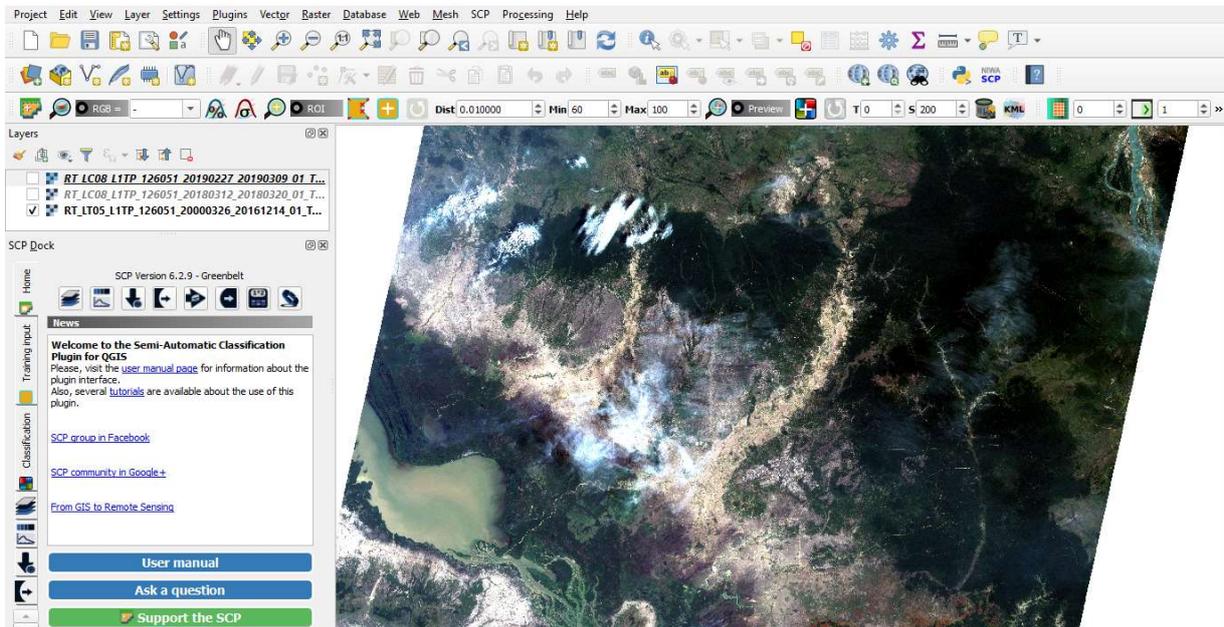


Figure 3. Preprocessed *Landsat 5* (year 2000) image displayed as *True Color Composite* in QGIS.

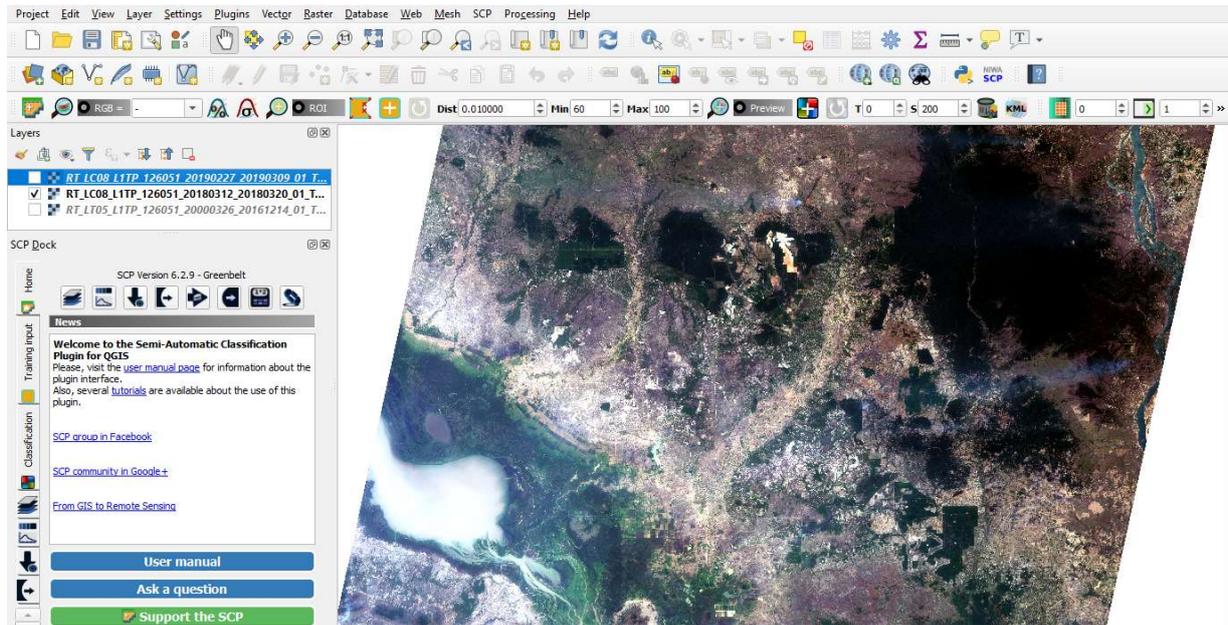


Figure 4. Preprocessed *Landsat 8* (year 2018) image displayed as *True Color Composite* in QGIS.



Figure 5. Clipped satellite image showing Prey Lang boundary (March 2000).



Figure 6. Clipped satellite image showing Prey Lang boundary (March 2018).



Figure 7. Clipped satellite image showing Prey Lang boundary (February 2019).

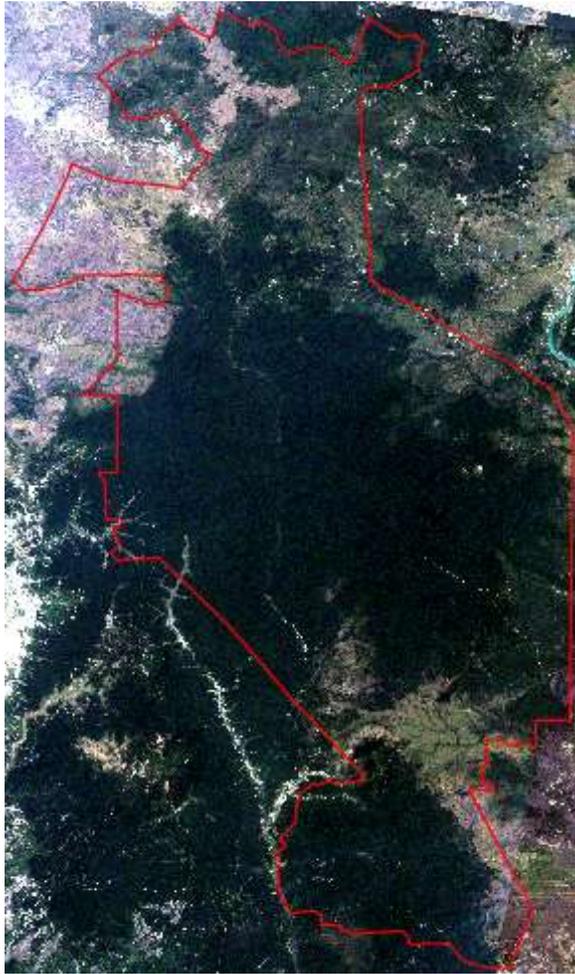


Figure 8. Clipped satellite image showing Prey Lang boundary (February 1999).



Figure 9. Clipped satellite image showing Prey Lang boundary (February 2018).

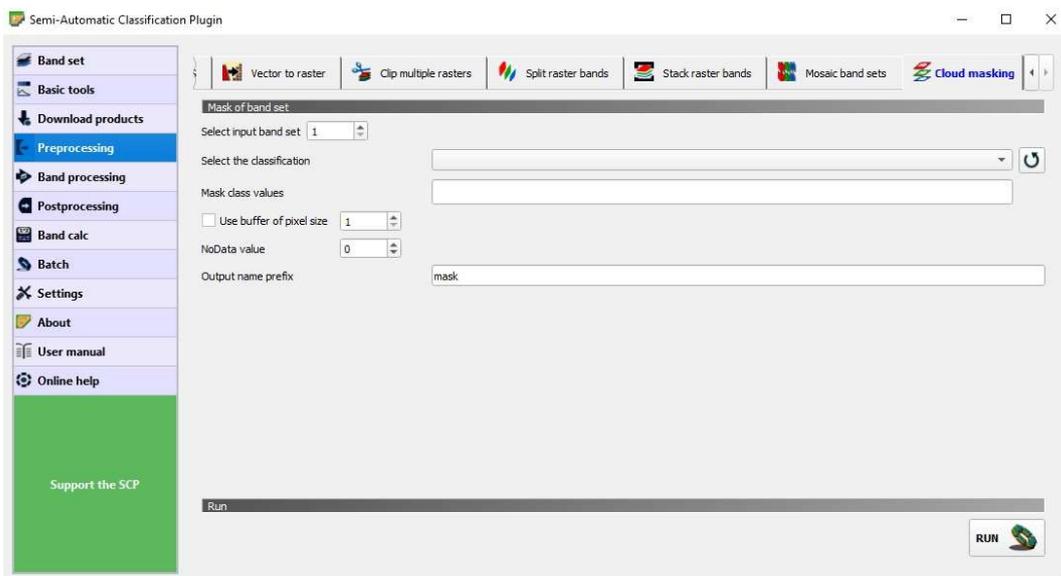


Figure 10. Interface of *cloud masking tool* of *SCP plugin* in QGIS.

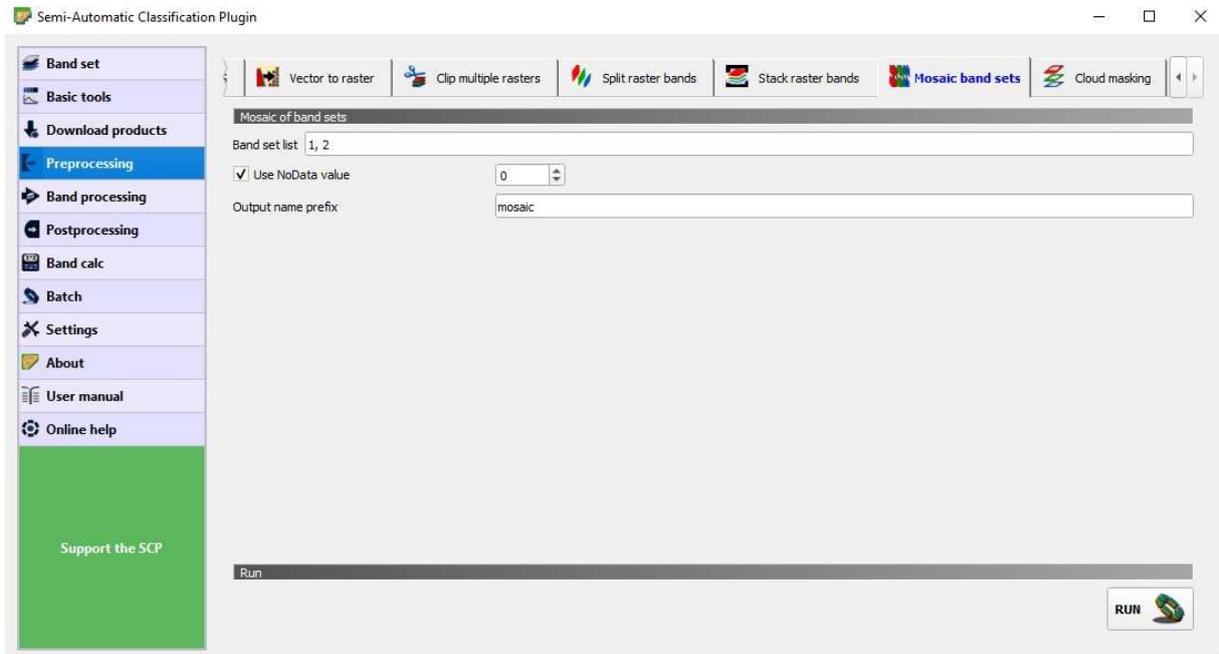


Figure 11. Interface of mosaic band sets tool of SCP plugin in QGIS.



Figure 12. Generated training data (shown as yellow polygons) for year 2000-2018 change detection.



Figure 13. Generated training data (shown as yellow polygons) for year 2000-2019 change detection.



Figure 14. Validation points (shown as orange points) for year 2018.



Figure 15. Validation points (shown as pink points) for year 2019.

Classified Image	Reference Image			Total
	Forest	Non-Forest	Water	
Forest	106	6	0	112
Non-Forest	3	98	1	102
Water	0	0	66	66
Total	109	104	67	280

Overall accuracy = $(106+98+66)/280 = 270/280 = 96.43\%$

Figure 16. Computation for the overall accuracy.

Classified Image	Reference Image			Total
	Forest	Non-Forest	Water	
Forest	106	6	0	112
Non-Forest	3	98	1	102
Water	0	0	66	66
Total	109	104	67	280

Producer's accuracy:

- Forest = $106/109 = 97.25\%$
- Non-Forest = $98/104 = 94.23\%$
- Water = $66/67 = 98.51\%$

Figure 17. Computation for the producer's accuracy.

Classified Image	Reference Image			
	Forest	Non-Forest	Water	Total
Forest	106	6	0	112
Non-Forest	3	98	1	102
Water	0	0	66	66
Total	109	104	67	280

User's accuracy:

Forest = 106/112 = 94.64%

Non-Forest = 98/102 = 96.08%

Water = 66/66 = 100%

Figure 18. Computation for the user's accuracy.

Classified Image	Reference Image			
	Forest	Non-Forest	Water	Total
Forest	106	6	0	112
Non-Forest	3	98	1	102
Water	0	0	66	66
Total	109	104	67	280

$$\hat{K} = \frac{M \sum_{i=j-1}^r n_{ij} - \sum_{i=j-1}^r n_i n_j}{M^2 - \sum_{i=j-1}^r n_i n_j}$$

$$= \frac{(280 \times 270) - [(112 \times 109) + (102 \times 104) + (66 \times 67)]}{280^2 - [(112 \times 109) + (102 \times 104) + (66 \times 67)]} = 0.95$$

Figure 19. Computation for the *kappa coefficient*.

Results

Below are the results from the satellite image processing, through the use of a *random forest* classifier, applied to generate the necessary change-detection classes and assessment of classification accuracy. With a review of the possible number of classes, which can be gleaned from the available secondary map information and focal image, we deemed it necessary to have a separate "forest to plantation" change class from that of "forest to agriculture/non-forest" change class. This particular class of "forest to plantation" should allow us ambiguity in classifying this later as forest loss or forest gain in a situation that which might be considered relevant. In the discussions following each sub-section beneath, the accuracy results and the corresponding change-detection output are described for each of the time periods of ca. 2000-2018 and ca. 2000-2019.

Year 2000 to Year 2018

Figure 20 shows the resulting classification from *R Studio*. Table 6 shows the results of the accuracy assessment, and Table 7 shows the area sizes, inside the protected area boundary, for each class category confined only to our focal classes on forest, non-forest and water due to the limited, secondary reference data. Based on this accuracy assessment, the resulting classification is considered good and highly accurate within the bounds of the protected area boundary. As can be seen in Table 7, a majority of the area is composed of forest (293,184 has.), which remained as forest between ca. 2000 to ca. 2018. However, there was a correspondingly great loss of the forest from year 2000 to year 2018. For better display of the classification, a change detection map layout is shown in Figure 21.

Table 6. Accuracy assessment of the resulting classified image of 2018.

Overall	94.91%
Producer's	
Forest	95.38%
Non-Forest	90.36%
Water	100.00%
User's	
Forest	93.94%
Non-Forest	92.59%
Water	100.00%
Kappa Coefficient	0.92

Table 7. Computed areas for each class for the period ca. 2000 to ca. 2018.

Prey Lang (2000-2018)	Area (Ha)
Forest (No Change)	293,184.36
Forest to Agriculture/Non-Forest (Loss)	69,926.04
Forest to Plantation	30,112.02
Non-Forest to Forest (Gain)	1,494.09
Non-Forest (No Change)	38,757.33
Water	2.97
Total	433,476.81

Year 2000 to Year 2019

Figure 22 shows the image classification generated from *R Studio*. Table 8 shows the results of the accuracy assessment, and Table 9 shows the area size for each class category on forest and non-forest and water. Based on the accuracy assessment, the resulting classification is considered good and highly accurate, even yielding a higher accuracy than that of the period from 2000 to 2018. Similarly, as shown in Table 9, the majority of the area is composed of forest (302,106 has.), which remained there (no change) since ca. 2000. A corresponding forest loss, however, occurred from year 2000 to year 2019. Vast area of land cover has also been changed from forest to plantation about almost twice that interpreted from 2000-2018. For better display of the classification, a change detection map layout is shown in Figure 23.

Table 8. Accuracy assessment of the resulting classified image of 2019.

Overall	96.43%
Producer's	
Forest	97.25%
Non-Forest	94.23%
Water	98.51%
User's	
Forest	94.64%
Non-Forest	96.08%
Water	100.00%
Kappa Coefficient	0.95

Table 9. Computed areas for each classification for the period ca. 2000 to ca. 2019.

Prey Lang (2000-2019)	Area (Ha)
Forest (No Change)	302,105.52
Forest to Agriculture/Non-Forest (Loss)	44,297.64
Forest to Plantation	58,138.29
Non-Forest to Forest (Gain)	1,110.51
Non-Forest (No Change)	27,822.69
Water	2.16
Total	433,476.81

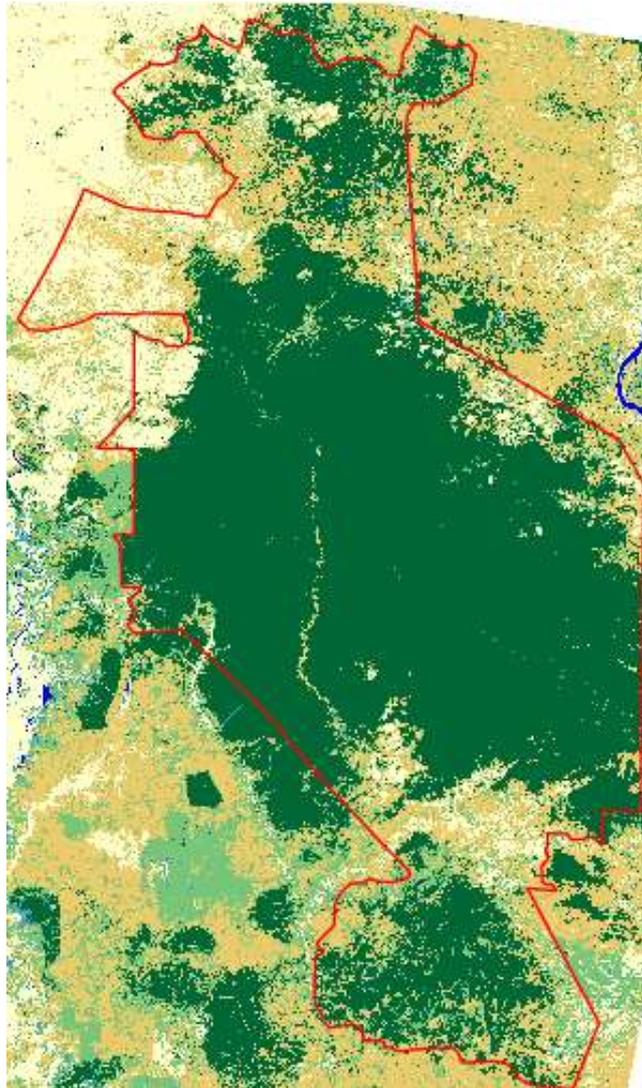


Figure 20. Change detection map for year 2000-2018 generated from *R Studio*.

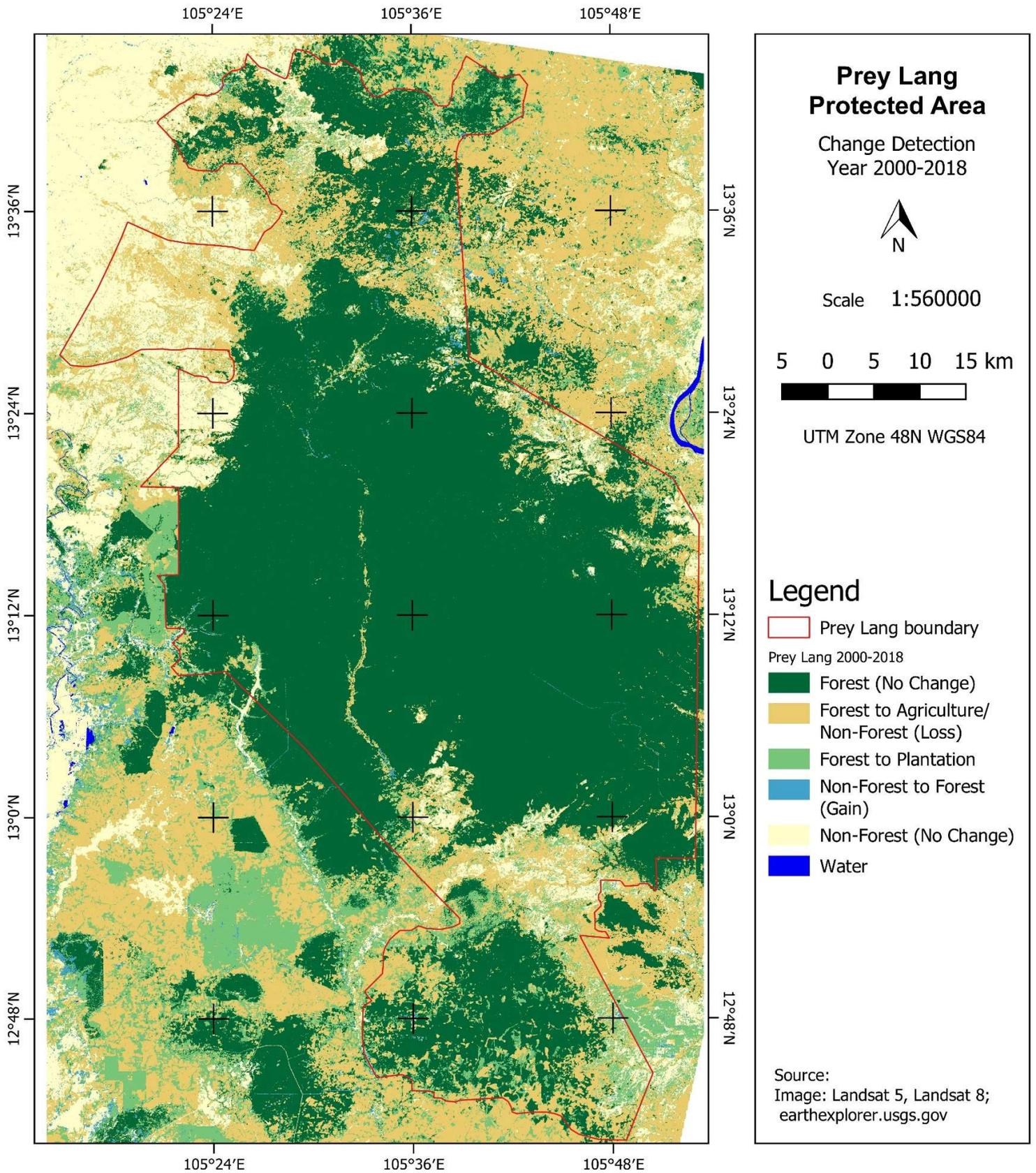


Figure 21. Change detection map of Prey Lang for the period ca. 2000 to ca. 2018.

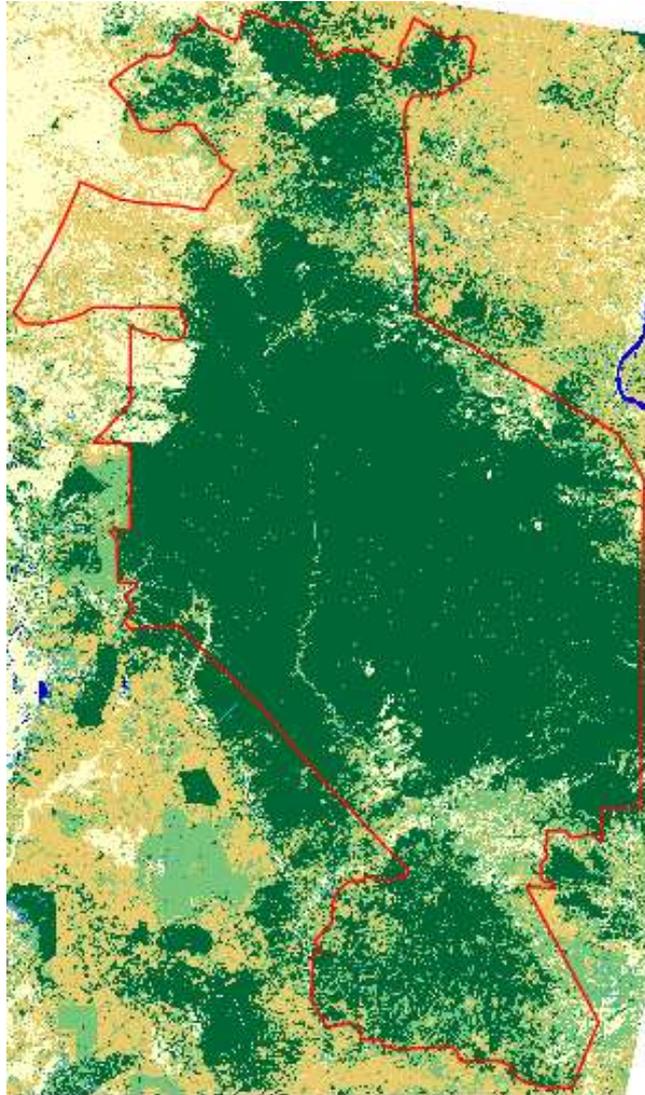


Figure 22. Change detection map for year 2000-2019 generated from *R Studio*.

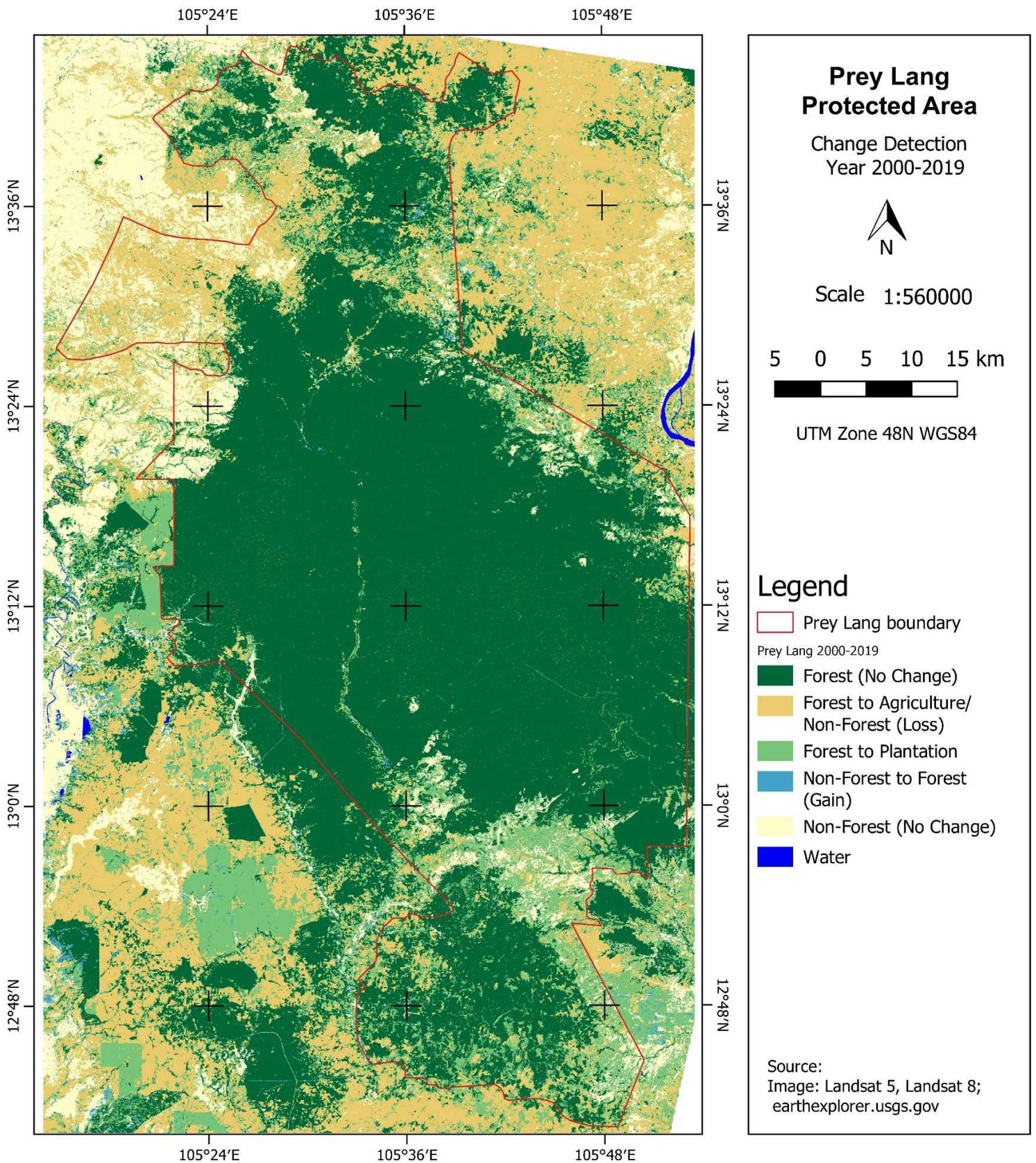


Figure 23. Change detection map of Prey Lang for the period ca. 2000 to ca. 2019.

Discussion

The change detection outputs for the timeframes 2000-2018 and 2000-2019 showed accuracies within a very good range. However, in observing the area sizes, there appear to be some remarkable views about their values when comparing 2000-2018 and 2000-2019. Table 9 indicates a higher figure in "forest" that remained intact since 2000 up to 2019 compared to that of 2000-2018 (Table 7), but a lower figure in change from "forest to non-forest or agriculture" class in 2000-2019 than in 2000-2018. These two trends were also similar in the case of "forest to plantation" change in which a higher figure persisted in 2000-2019 than in 2000-2018, and, succeedingly, the non-forest class, which did not change over time, had a lower figure in 2000-2019 than in 2000-2018. There was somewhat a closer figure in values between the two time periods pertaining to "non-forest to forest gain" and water body.

As of this writing, we can offer no readily available physical explanation on the ground as to why the remaining forests (no change) in 2018 suddenly increased by about 3% (~8,921 has.) in a span of one year in 2019, except, perhaps, on the condition of the preprocessed image data used. The 2018 image was a mosaic of two images from March 12, 2018 and February 2, 2018 in an attempt to remove the clouds. Although there were clouds removed, there was no guarantee that we could also remove haze, due to atmospheric disturbances above the forest canopy, appearing in the March 2018 image. This March 2018 image also appear darker in tone compared to the February 2019 image. Hence, we refer to the quality of the images, which were used in the processing. Although the resulting classified image of vintage 2000-2019 had higher accuracy by about two percent than 2000-2018, visual inspection of the images of each of the two timeframes revealed that the February 2019 image appeared better in quality. There appears to be more pixels revealing forest and also plantations in the north-northwest sector in the 2000-2019 timeframe than in the 2000-2018 timeframe. The two notches up in accuracy of the former than the latter enable us to put more confidence on the 2000-2019.

Conclusion

This technical report has accomplished its aim of describing the data, method and results in developing the change detection maps for Prey Lang. Although it has taken double the original number of days to complete the project, we did due diligence in investigating two timeframes, because of data quality and the level of training and validation data available.

From the derived map outputs, we have been able to answer the questions of how much forest was lost and how much forest was gained between the two-time periods of 2000-2018 and 2000-

2019 and the remaining forest extent in 2018 and 2019. The corresponding accuracy statistics on the classified images for change detection indicated acceptable output, but there was additional uncertainty created by the quality of the 2018 image.

Additional processing time might be necessary to evaluate and rectify the 2018 output. Further, additional facility to obtain training or validation data will be needed, which would require ground-truthing. In view of these constraints and to move forward, we recommend the use of the 2000-2019 change detection map for the purpose this mapping project was conceived by its proponents.

References

- Breiman, L (2001). Random forests. *Machine Learning*, 45, 5-32. 28pp. doi:10.1023/A:1010933404324.
- Congalton, R.G (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37(1): pp. 35-46.
- Copeland, M (2016). What's the Difference Between Artificial Intelligence, Machine Learning and Deep Learning? Accessed on August 28, 2019 from <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>
- Open Development Cambodia (2019). Forest cover in Cambodia (1973-2014). Accessed on May 14, 2019 from <https://data.opendevdevelopmentcambodia.net/en/dataset/forest-cover-in-cambodia-1973-2014>.
- Penn State Science (2011). *Measure of Agreement: Kappa*. Accessed on August 28, 2019 from <https://newonlinecourses.science.psu.edu/stat504/node/99/>
- Prey Lang Community Network (2019). *Prey Lang*. Accessed on July 29, 2019 from <https://preylang.net/about/the-forest/>
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- R Studio Blog (2011, February 28). RStudio, new open-source IDE for R. Accessed on July 29, 2019 from <https://blog.rstudio.com/2011/02/28/rstudio-new-open-source-ide-for-r/>