

Article

# Developing Land Use Land Cover Maps for the Lower Mekong Basin to Aid Hydrologic Modeling and Basin Planning

Joseph Spruce <sup>1,\*</sup>, John Bolten <sup>2</sup>, Raghavan Srinivasan <sup>3</sup> and Venkat Lakshmi <sup>4</sup>

<sup>1</sup> Science System and Applications, Inc. Consultant, 88384 Diamondhead Drive East, Diamondhead, MS 39525, USA

<sup>2</sup> Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Mail Code 617, Greenbelt, MD 20771, USA; john.bolten@nasa.gov

<sup>3</sup> Spatial Sciences Laboratory, Department of Ecosystem Science and Management, Texas A&M University, College Station, TX 77843, USA; r-srinivasan@tamu.edu

<sup>4</sup> School of Earth Ocean and Environment, University of South Carolina, Columbia, SC 29208, USA; vlakshmi@geol.sc.edu

\* Correspondence: j\_spruce@outlook.com; Tel.: +1-228-342-5505

Received: 23 October 2018; Accepted: 20 November 2018; Published: 29 November 2018



**Abstract:** This paper discusses research methodology to develop Land Use Land Cover (LULC) maps for the Lower Mekong Basin (LMB) for basin planning, using both MODIS and Landsat satellite data. The 2010 MODIS MOD09 and MYD09 8-day reflectance data was processed into monthly NDVI maps with the Time Series Product Tool software package and then used to classify regionally common forest and agricultural LULC types. Dry season circa 2010 Landsat top of atmosphere reflectance mosaics were classified to map locally common LULC types. Unsupervised ISODATA clustering was used to derive most LULC classifications. MODIS and Landsat classifications were combined with GIS methods to derive final 250-m LULC maps for Sub-basins (SBs) 1–8 of the LMB. The SB 7 LULC map with 14 classes was assessed for accuracy. This assessment compared random locations for sampled types on the SB 7 LULC map to geospatial reference data such as Landsat RGBs, MODIS NDVI phenologic profiles, high resolution satellite data, and Mekong River Commission data (e.g., crop calendars). The SB 7 LULC map showed an overall agreement to reference data of ~81%. By grouping three deciduous forest classes into one, the overall agreement improved to ~87%. The project enabled updated regional LULC maps that included more detailed agriculture LULC types. LULC maps were supplied to project partners to improve use of Soil and Water Assessment Tool for modeling hydrology and water use, plus enhance LMB water and disaster management in a region vulnerable to flooding, droughts, and anthropogenic change as part of basin planning and assessment.

**Keywords:** land use land cover mapping; SWAT hydrologic modeling disaster management; water resource management; agricultural monitoring

## 1. Introduction

This paper discusses results of a study that was conducted to update and validate Land Use Land Cover (LULC) maps for the Lower Mekong Basin (LMB) as part of a larger, collaborative project between National Aeronautics and Space Administration (NASA), United States Geological Survey (USGS), Texas A&M University, University of South Carolina, and the Mekong River Commission (MRC). The project was supported by the NASA Earth Sciences Disasters Program (<https://disasters.nasa.gov/>) and aimed to provide updated and improved hydrological monitoring and modeling

tools for the Lower Mekong River Basin. This effort used NASA remote sensing systems and data to help improve Mekong River Commission (MRC) Soil Water Assessment Tool (SWAT) models for sub-basins (SBs) 1–8. The project included computation and updates of LULC maps that were in turn used as a primary input to MRC SWAT hydrologic models developed for multiple sub-basins of the LMB. Updated LULC were required to optimize MRC SWAT Hydrologic models that are used in turn to aid water management and basin planning via better land surface representation and soil moisture characterization inputs to SWAT and improved hydrologic modeling outputs (e.g., in terms of surface runoff streamflow). The LULC component of the project is discussed in this paper in terms of introductory background information, methods, results, and conclusions.

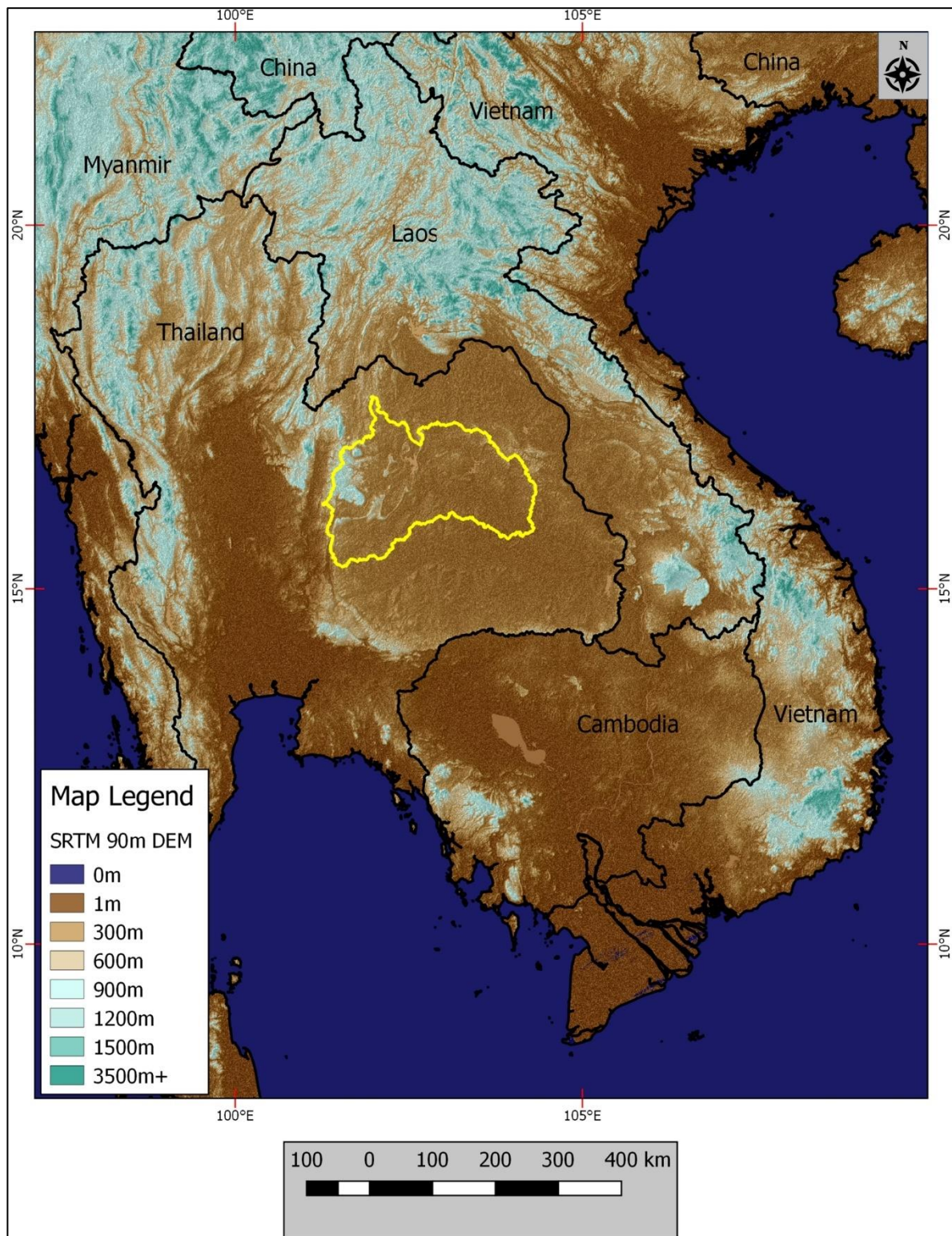
### 1.1. Description of Problem

The Mekong River region contains the longest river in Southeast Asia [1]. It is currently home to more than 70 million people inhabiting six countries [2]. The river is heavily used for enabling transportation, supplying water to urban areas, irrigating agricultural crops, managing fisheries, conserving wildlife, generating electric power, and enabling ecotourism [3]. The region is being increasingly utilized for hydroelectric power generation, which has led to a growing network of dams that threaten certain land uses such as agriculture and fishing. Several dams have been built in the mid to upper reaches of the river and more are being constructed or planned. The region also is highly vulnerable to climate change related disasters, such as seasonal flooding and droughts [4]. The LMB (Figure 1) includes over 60 million people with mostly agricultural areas in the lower elevations and forests in the higher elevations [3]. Agricultural crop security is an issue in the LMB due to increased hydropower development (including damming) in the upper reaches of the primary river and its main tributaries [4,5]. Climate change is another factor impacting crop security in the LMB [3,6].

### 1.2. Literature Review

Selective LULC types within portions of the LMB has been mapped previously using multispectral satellite data from either Landsat or Moderate Resolution Imaging Spectroradiometer (MODIS) sensors supported by NASA. Prior to this project, the MRC's SWAT model utilized a circa 1997 LULC map of the LMB [7]. The latter map was compiled via image analyst interpretation and delineation of 1993 and 1997 Landsat false color composite image hard copies output at a 1:250,000 scale [8]. In producing this map, image analysts interpreted the Landsat image hard copies to delineate LULC patches according to a 0.5 square kilometers (km<sup>2</sup>) minimum mapping unit (MMU), as part of the Mekong Forest Cover Monitoring Project funded by the country of Germany [9]. When input into SWAT, this map was generalized and then used with lookup tables and other Geographic Information System (GIS) data to predict various hydrologic parameters. SWAT uses LULC data in part because hydrologic flow can be affected in part by the amount and location of LULC type patches within a given watershed [1]. LULC can also affect riverine water quality in terms of instream sedimentation. In terms of aiding water management, the MRC considers forests to be a preferred LULC type for catchments being used for water supplies [3].

LULC mapping in the study area and LMB at large is difficult to do at both a detailed LULC type specificity and a moderately high classification accuracy due to multiple factors, including cloud contamination, availability of cloud free data, LULC patch characteristics, plus the available spatial, temporal, and spectral resolutions of satellite data. Some LULC types, such as orchards, can occur as fine scaled patch sizes that occupy a fraction of a pixel. LULC mapping in the tropical LMB region can be subjected to high cloud frequency, especially during the wet season [10]. Even for the multi-month dry season, cloud-free or nearly cloud-free Landsat scenes can be scarce and/or spotty in occurrence, usually limited to one or two quality scenes per location annually (typically during the dry season).



**Figure 1.** Location of sub-basin 9 study area (in yellow) with respect to countries (in black) in the Lower Mekong Basin region of Southeast Asia. The underlying 90 m elevation map is from NASA Shuttle Radar Topography Mission (SRTM) data.

Unfortunately, effective mapping of certain LULC types in the LMB can require more than one date of remote sensed data annually, especially for LULC class with specific foliar greenness phenology, such as certain agricultural and forest types. Native forest types are often described and categorized



according to an apparent level of deciduousness. Mapping such types can require high temporal resolution satellite data to observe greenness levels across multiple times of the growing season. Attempts to map forest deciduousness from only one date of the dry season may only be partially successful due to variable phenology of leaf senescence that can be evident for specific forest types and locations. Some agricultural types also require multiple dates of remote sensing data to map the number of crops per location annually [11,12]. LMB LULC mapping conducted with Landsat data typically limits classification scheme specificity for certain agricultural LULC types. For example, for another project, the MRC generated a 2010 LULC map based on Landsat data that contained only one rice class [13].

MODIS data collected twice daily offers a potential means to map LULC types with LULC specific vegetation greenness phenology. For example, MODIS data collected was combined with other geospatial data to derive a regional land cover map for the Mekong region [10]. The latter observed that several LULC types can be identified by image classification of 16-day Enhanced Vegetation Index (EVI) data sets from combined MODIS Terra and Aqua data in conjunction with other data such as SRTM DEM data [10]. They found that mapping LULC in this cloud prone tropical region required a custom temporal data preprocessing method to derive cloud-mitigated, useable EVI time series [10]. These researchers also noted that the high cloud frequency during the rainy season can exceed or approach 90% based on MODIS Terra data [10]. Even in seasonally cloudy temperate broadleaved forests, the estimated cloud frequency using MODIS Terra daily data during late spring to mid-summer was approximately 78% [14].

Normalized Difference Vegetation Index (NDVI) data is frequently employed to monitor vegetation greenness phenology. NDVI was first used for monitoring vegetation canopy greenness in the 1970s [15]. NDVI has had a long history of application since then and can be generated with any multispectral system that contains red and near infrared spectral bands. NDVIs theoretically varies between  $-1$  and  $+1$ . Based on personal experience, MODIS quarter kilometer (qkm) NDVI data is practically always less than the maximum native value of 1 even for mature, high biomass tropical rain forests. NDVI has been used in conjunction with lookup tables to compute Leaf Area Index (LAI) image products that can saturate in high biomass vegetative conditions [16]. EVI has been used for vegetation monitoring instead of NDVI because it is less vulnerable to saturation from high foliar biomass and is also more atmospherically resistant than NDVI [16]. However, NDVI is better suited for monitoring vegetation in rugged terrain [17].

Cloud cover on multispectral satellite data (e.g., MODIS) tends to reduce NDVI values for most LULC types compared to NDVIs of the same types on cloud free data. Such reductions are undesirable for monitoring change in vegetation greenness and classifying LULC class based on multitemporal NDVI data. One way to mitigate cloud impacts on NDVI data products is through use of maximum value NDVI temporal compositing which favors noncloudy NDVIs which are usually higher than cloud contaminated NDVIs [18]. In addition, constrained view maximum NDVI temporal compositing has been used to improve NDVI compositing quality by excluding data with extreme off nadir views [16]. Other techniques have also been applied to further reduce noise in NDVI composites, such as the use of methods for temporal smoothing of NDVI times series through use of Savitzky–Golay filtering and curve fitting techniques [19].

Landsat data at 30m spatial resolution has been used to map a variety of LULC classes in the LMB [9,13], though full use of annually collected imagery can be limited by a lack of cloud-free data, especially during the rainy season. Single dates of cloud free LMB Landsat data during the dry season are more likely to be available than for the rainy season. However, multiple dates of cloud-free data across the dry season are rarely if ever regionally available. On the other hand, MODIS data collected twice daily is more temporally resolute than Landsat data collected every 16 days, enabling the possibility of mapping LULC in the LMB based on vegetation greenness phenology characteristics [10]. However, MODIS data even at its finest spatial resolution (qkm) can be too spatially coarse for optimally mapping certain LULC types that can be locally common yet regionally sparse, such as

urban or industrial forest plantation areas. For the LMB, the combined use of MODIS and Landsat data could enable an improved means to map a broader range of LULC types than what is feasible singularly using either Landsat or MODIS data. However, previous efforts to apply a combination of Landsat and MODIS data for mapping multiple LULC types in the LMB or other regions was not indicated in the review of published literature. Most of the LULC mapping studies focus on use of a singular data source, such as from Landsat or MODIS or other satellite sensors. However, sometimes the requirements of a given LULC mapping project may not be met using one source of remotely sensed data to generate the needed map. Given that LULC mapping applications can have different requirements (e.g., in terms of classification scheme, mapping accuracy, and minimum mapping unit), sometimes more than one data source may ultimately be needed to produce required LULC maps.

There have been a variety of Landsat and/or MODIS data processing techniques used to generate LULC maps for the LMB, including semi-automated unsupervised and/or supervised image classification methods [13,20], as well as LULC maps based on analyst-based manual interpretation and digitization of satellite and/or aerial imagery either in hardcopy or digital form [9]. Each of these map generation methods has advantages and disadvantages as described in numerous remote sensing text books [21]. The supervised classification method can be used to produce superior maps, though the onus is on the map developer to obtain and properly use enough training data needed to produce viable signatures for classifying each targeted LULC class in the study area. The unsupervised method can also be used to produce impressive LULC maps, though successful use usually depends on how the method is applied, the skill sets of the map developer, and the availability of reference data. Sometimes, a combination of supervised and unsupervised methods is used to compute needed LULC maps [21]. In addition, GIS-based spatial models with data integration decision rules can be used to merge multiple input classifications into holistic “wall to wall” LULC maps [22].

The analyst-based digitization of LULC mapping method was used to generate to map LULC in the LMB for 1997, though this method was expensive to apply and required experienced image analysts, field surveys, and subjective delineation of LULC polygons [9]. Such a manual method is more subjective compared to semi-automated image classification methods and is not amenable to automated map production.

### 1.3. Project Objectives

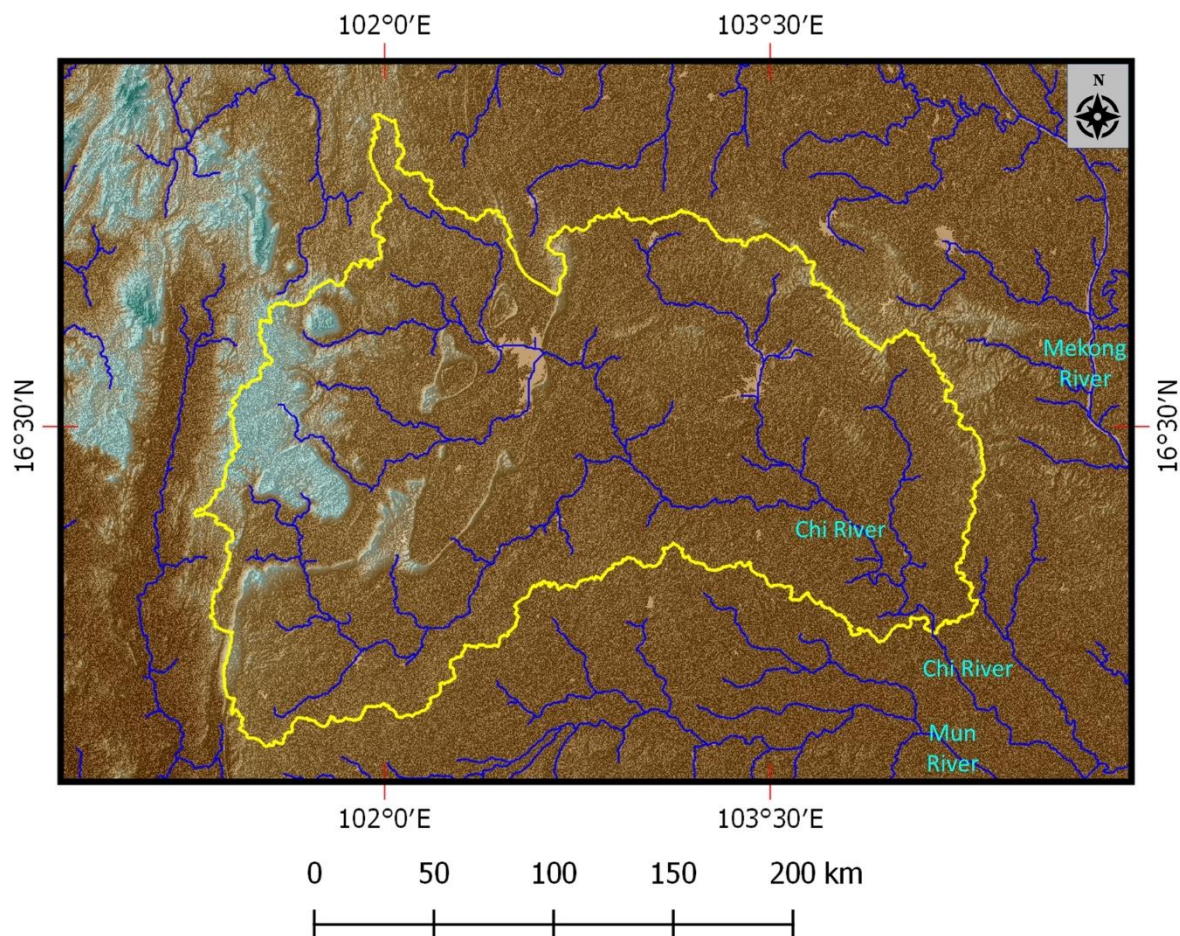
Prior to the project described herein, the LULC map used in the MRC SWAT model was based on data from 1997 [7]. Consequently, there was a need to update the LULC maps used in MRC SWAT models. In response, we conducted a project that included the goal of updating the LULC maps to 2010 so that LMB SWAT models for SBs 1–8 could be improved. The 2010 date was selected given the project was initiated in 2012 and MRC’s interest in updating LULC to 2010. The LULC mapping task for the project included multiple objectives. (1) Compute updated LULC maps of LMB SB’s 1–8 using circa 2010 MODIS and Landsat data. (2) Validate representative LULC areas. (3) Assess the utility of the LULC maps and the methods for producing such maps. (4) Contribute the final LULC maps to the MRC for updating SB-specific SWAT models. This paper presents and discusses LULC mapping and map validation methods and results for SB 7 of the LMB as an example LMB SB that contains a mixture of forest, agriculture (including crop rotation information), water, and barren LULC types common to the LMB at large.

## 2. Materials and Methods

### 2.1. Study Area

The study area is in the sub-basin (SB) 7 of the Lower Mekong Basin (LMB) (Figure 1). The LMB includes areas in Cambodia, Laos, Myanmar, Thailand, and Vietnam. SB 7 is one out of 8 SBs that the MRC uses with the SWAT model to aid water and disaster management in the region [1]. The study area includes in the central portion of the Khorat Plateau of Thailand and varies in elevation from 90 to

1326 m, based on NASA SRTM 30-m Digital Elevation Model (DEM) data [23]. As the main waterway in the SB, the Chu River runs into the Mun River which then flows into the Mekong River (Figure 2).

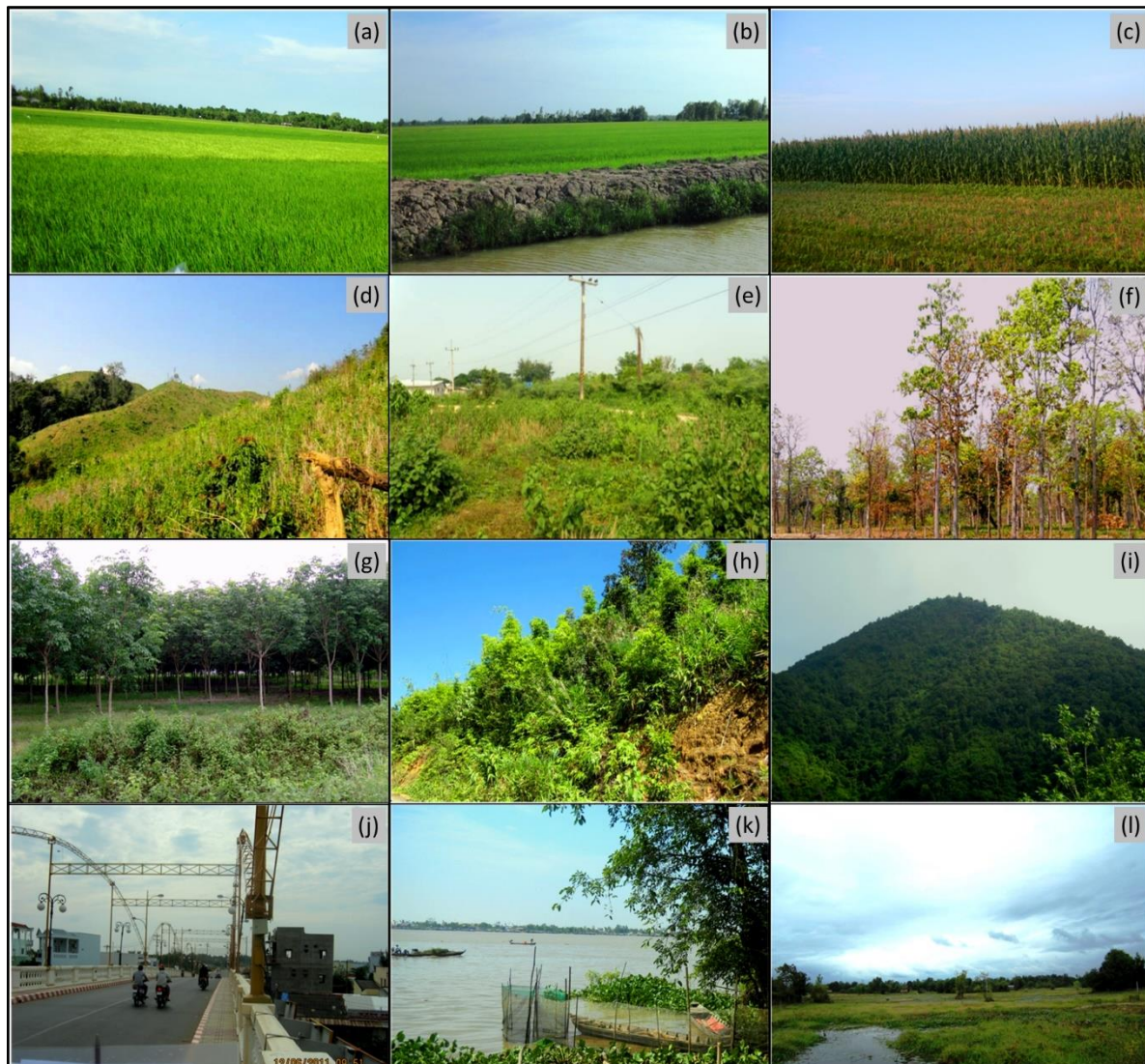


**Figure 2.** Terrain and principal waterways of the study area (sub-basin 7). The study area boundary is shown in yellow and the waterways are shown in blue. The underlying elevation map is from 30 m SRTM data in conjunction with the same legend shown in Figure 1.

The study area contains several LULC classes (Figure 3) including multiple natural forest and agricultural types of interest to MRC SWAT modelers and relevant to water resource management. For the most part, the study area is an agriculturally dominated landscape in which single season rainfed rice and other mixed annual crops are grown. Along more prominent waterways, irrigated double-cropped rice can commonly occur. Shifting cultivation “slash and burn” agriculture is rare, compared to more forest dominated SBs in the LMB, such as in Lao PDR. Within the study area, some industrial forest plantations (e.g., rubber) occur amongst the main agricultural areas. Deciduous and evergreen broadleaved forests occupy portions in the study area and are primarily limited to higher elevation areas that are mainly managed as national parks. Bamboo scrub/forest occurs in some forest dominated areas. Wetlands with nonagricultural vegetation in the study area are comparatively rare, due to most wetland sites being converted to and managed for rice agriculture. A detailed description of forest LULC types of the study area is provided by [24].

The study area includes a dry season that typically occurs from October through April. A wet (i.e., monsoon) season takes place from May to October [24]. LMB cloud frequency is typically much higher during the wet season [10].





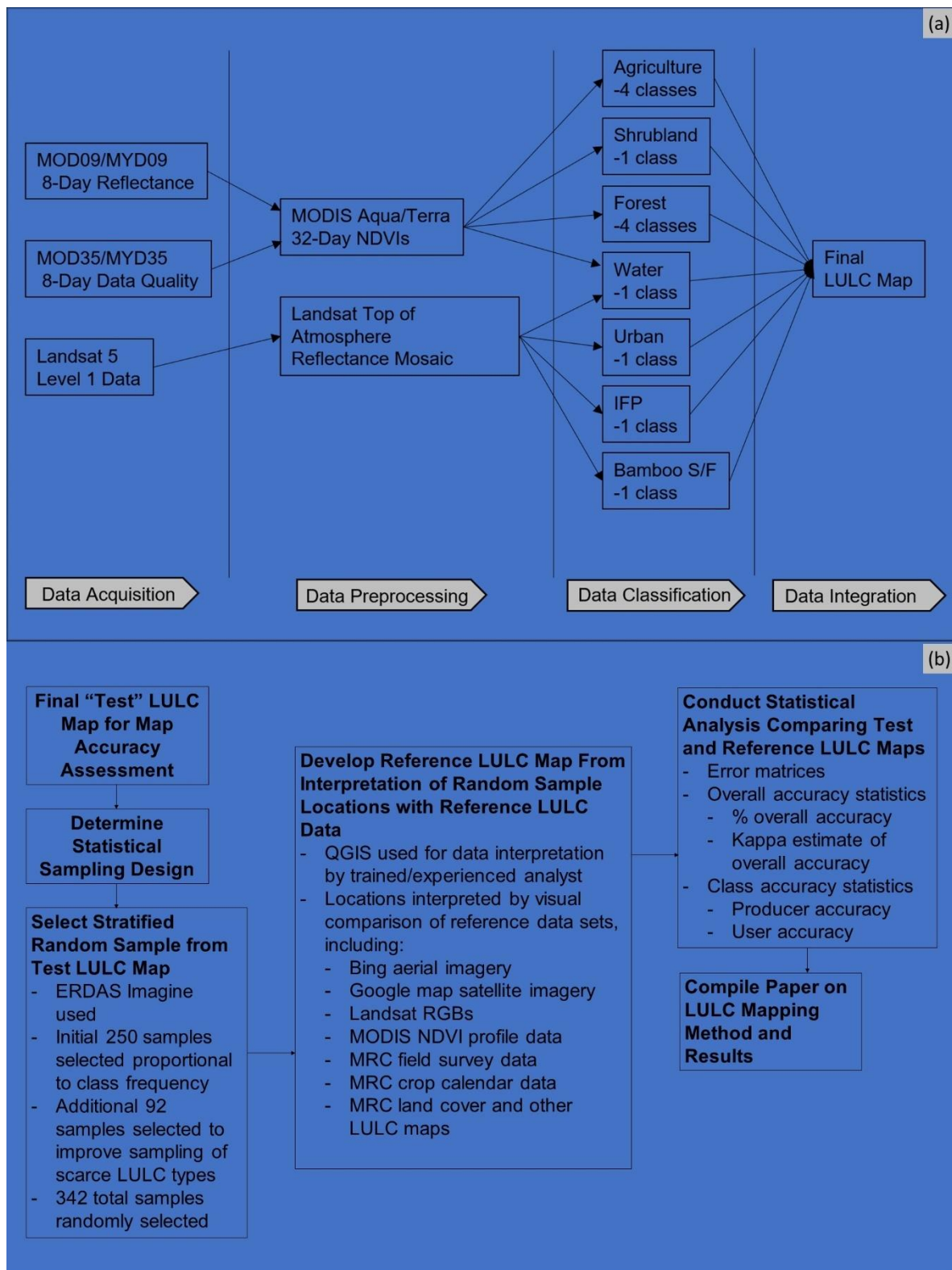
**Figure 3.** Example Land Use Land Cover (LULC) classes occurring in SBs 1–8 of the lower Mekong Basin (LMB): (a) rainfed rice; (b) irrigated double-cropped rice; (c) mixed annual crops other than rice; (d) shifting cultivation areas; (e) shrubland; (f) deciduous forest; (g) industrial forest plantation (e.g., rubber); (h) bamboo scrub/forest; (i) evergreen broadleaved forest; (j) urban areas; (k) water bodies; and (l) wetlands. The above imagery was provided to project by the Mekong River Commission (MRC).

## 2.2. Data Acquisition

The basic work flow used in generating the LULC map for SB 7 is shown in Figure 4. To initiate the mapping process, MODIS Collection 5 MOD09 and MYD09 8-day composited reflectance data at a 1 km spatial resolution [25,26] were acquired for 2003 to 2011 from the LP DAAC, including data for tiles h27v06, h27v07, h28v07, and h28v08. Corresponding MODIS MOD35 and MYD35 cloud masking products [27] were also acquired to assess and clean MODIS reflectance data for known data quality issues.

Level 1 Landsat Thematic Mapper data collected during the dry seasons (February) of 2009 to 2011 was downloaded from the USGS GloVis web site [28], including data sets from 6 scenes with paths 127 (rows 48 and 49), 128 (rows 48 and 49), and 129 (rows 48 and 49). In addition, SRTM DEM data was acquired for the study area at 1 arc second (i.e., 30-m) resolution [23]. Additional higher

spatial resolution true color satellite and aerial imagery was accessed from Google and Bing web sites via the QGIS OpenLayers plugin [29].



**Figure 4.** Generalized work flow for LULC mapping (a) and LULC map accuracy assessment (b) components of project. For acronyms in above graphic, IFP = Industrial Forest Plantation and S/F = Scrub/Forest.



### 2.3. Data Preprocessing

Although MODIS Aqua and Terra sensors together collect two images per day for most of the human inhabited world, the data from these systems typically needs additional temporal processing to reduce cloud and cloud shadow impacts, especially for the cloud prone tropics. The LMB area is frequently cloudy, especially during the rainy season [10,24]. Even in temperate environments, effective production and use of MODIS-based time series of vegetation index products typically must include temporal processing to remove bad data and mitigate pixels for dates of data affected by atmospheric contamination. Also known as temporal processing, this type of data preprocessing is necessary as a precursor to LULC classification based on MODIS NDVI time series data sets.

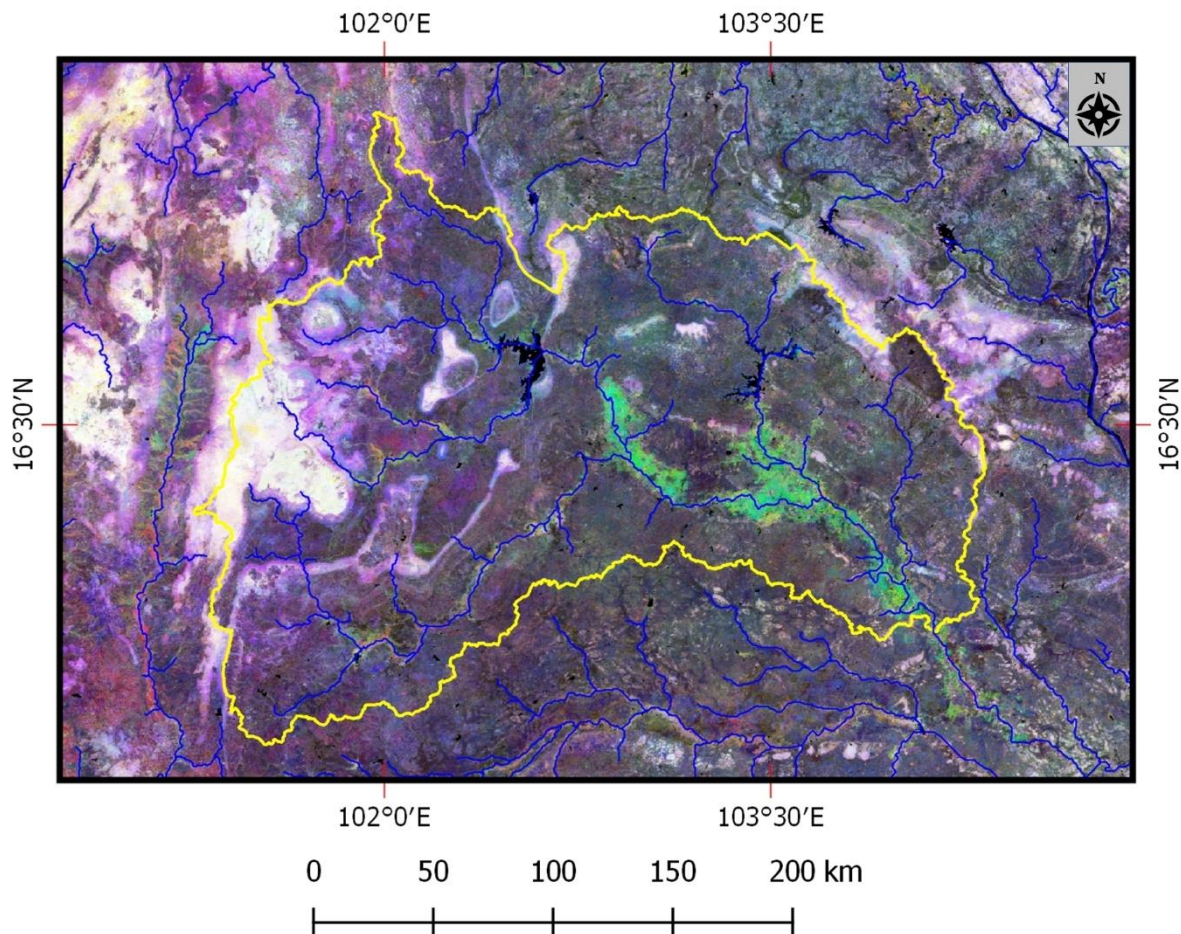
The Time Series Product Tool (TSPT) was used in this project to perform temporal processing of the MODIS MOD09 and MYD09 8-day reflectance data. TSPT is a custom program developed with MATLAB computer programming software [30–32]. TSPT has been used on numerous occasions to derive temporally process time series of vegetation index products. It enables such time series data sets to be processed in terms of bad data removal, data fusion, noise reduction, and data void interpolation [14,30–32].

TSPT data processing began with the ingest of MODIS Terra (MOD09) and Aqua (MYD09) 8-day data on an individual tile basis for a given year. Such 8-day products are less frequently cloud covered than daily reflectance data but still can include extensive areas (e.g., pixels) affected by atmospheric contamination, poor viewing geometry, and inherent data quality artifacts (e.g., from pixel and line drop outs). In this project, the goal of using TSPT was to temporally process both the 8-day MODIS Aqua and Terra reflectance data to derive annual 32-day merged MODIS Aqua/Terra NDVI time series that were much less visibly cloud-impacted than the original 8-day data sets and therefore more amenable to LULC classification. The processing of the annual time series for 2010 included 96 total MODIS 8-day reflectance data sets, including 48 each from MODIS Terra and Aqua sensors.

TSPT was used to initially ingest MODIS reflectance data in HDF format and subsequently process NIR and red reflectance data into a NDVI imagery, processing the MODIS Terra and Aqua data separately. Then, depending on the sensor, MODIS MOD35 or MYD35 “Quality Assurance” (QA) data was used to eliminate bad data due to atmospheric contamination. Afterward, each date of QA filtered MODIS NDVI data was subject to a noise detection and removal routine to delete unrealistic NDVI values with abnormally high or low data values compared to NDVIs of adjacent dates in the time series. In doing so, a threshold of  $\pm 0.25$  to detects abnormal “spikes” (aberrant increases or decreases) in 8-day NDVI from one date to the next within an annual sequential time series. If a pixel was determined to have a bad NDVI value, then it was assigned a no data value. Acceptable NDVIs were then retained for subsequent processing. After the spike removal process, the residual NDVI data were further processed to eliminate NDVIs with extreme off nadir viewing geometries (pixels with sensor zenith angles exceeding  $\pm 50$  degrees). Once all the MODIS Terra and Aqua NDVI time series data was cleaned with TSPT, resulting data from each sensor was then merged into 8-day maximum value composited NDVIs on a date by date basis. The merged NDVI images included some data voids that were then mitigated using a bilinear void interpolation routine that estimates the NDVI of a given void by considering adjacent dates in the annual time series that have quality NDVI values.

After void interpolation, the NDVI time series data was smoothed using temporal curve fitting routine based on a weighted Savitzky–Golay (SG) filtering algorithm [33]. The parameters for the latter routine was custom set through trial and error analysis for the study area. In doing so, the Full Width Half Maximum (FWHM) was set to 2 (i.e.,  $\pm 16$  days within a given date), the Frame Size was set to 3 (i.e.,  $\pm 24$  days within a given date), and the Polynomial Order was set to 1. This method is used to reduce residual noise in the 8-day NDVI time series due to factors such as atmospheric contamination [19]. After SG filtering, the processed 8-day products were then aggregated for each tile into 32-day NDVIs on an annual basis that included 12 dates of nonoverlapping 32-day NDVIs for each calendar year. The final NDVI output from TSPT was output as generic binary image files with ENVI header files. The latter NDVIs for each tile was then imported into ERDAS Imagine which was then

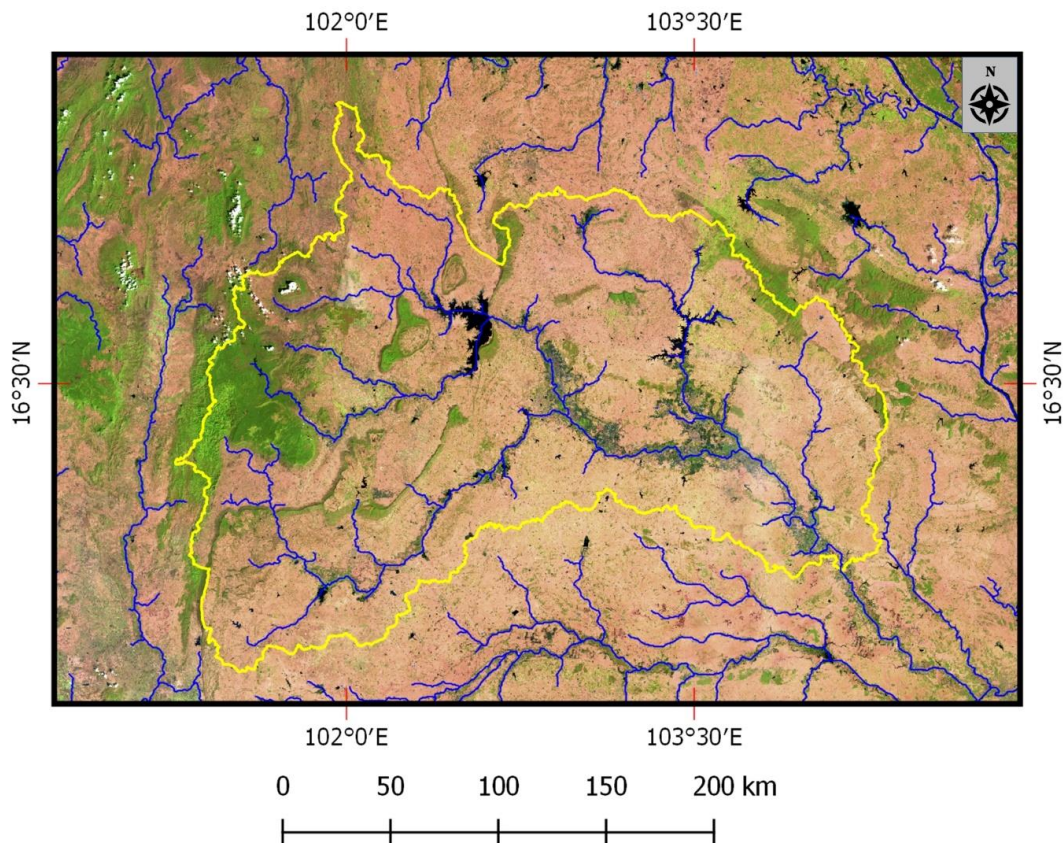
applied to stack each annual set of 32-day NDVIs for each of the individual tiles. Once stacked, the individual tiles for each year were stitched into an annual region wide mosaic that were subsequently re-projected from the native sinusoidal to the geographic coordinate system. The final NDVIs were scaled in terms of an unsigned 8-bit dynamic range in which Digital Numbers (DNs) ranged from 1 to 201. This scaling was accomplished by multiplying the native NDVI by 100 and then applying an additive offset of 101. In doing so, the resulting NDVIs for most water areas were greater than 0 and less than or equal to 101. The NDVIs for land areas were typically between 101 and 201. An example 3-date monthly MODIS NDVI RGB for 2010 is shown in Figure 5.



**Figure 5.** Example false color composite image of study area (outlined in yellow) based on 2010 monthly MODIS NDVI data in which dates 1, 3, and 5 are assigned to the RGB color guns. The RGB shows irrigated, double-cropped rice along waterways in a cyan tone. The waterways (in blue) are included to aid interpretation.

Level 1 Landsat data was acquired for the project from the USGS GloVis web site [28] and was subsequently preprocessed with the QGIS Semi-Automatic Classification Plugin (SCP) plugin into top of atmosphere reflectance data. Afterwards, the Dark Object Subtraction (DOS) method, developed previously [34], was applied to reduce haze effects of each utilized Landsat data set. In the tropics, it is common for even during the dry season that atmospheric contamination effects can occur, and such effects tend to be more evident in the shorter visible wavebands. Such contamination can hinder the production of viable LULC classifications. Consequently, each Landsat data set was subset to include only the red, near infrared, and two short wave infrared bands as a precursor to LULC image classification. Each Landsat scene was then stitched together into a mosaic covering the study area. The preprocessed Landsat mosaic was then assessed for quality assurance (Figure 6).





**Figure 6.** Circa 2010 Landsat TM mosaic of study area (yellow), based SWIR-1 (1.65 micron), NIR, and red reflectance bands loaded into the RGB color guns. The waterways (in blue) are included to aid interpretation.

#### 2.4. LULC Mapping

Unsupervised and supervised classification methods were both used to process MODIS NDVI and Landsat multispectral data into the needed LULC map of the study area. The unsupervised method was primary technique used for LULC classification. MODIS 32-day NDVI data for 2010 was used to classify regionally evident natural forest and agricultural cover types that exhibited unique vegetation greenness phenology either for the dry season or across a given calendar year. The Landsat mosaic was used to classify locally common but regionally-scarce LULC types that were too fine scaled for detection with MODIS qkm NDVI data.

The LULC classification process began with open water areas being mapped using the MODIS 32-day NDVI data stack for 2010 by (1) extracting the minimum value NDVI considering all 12 dates and (2) using a spatial model to apply a threshold range of NDVI values 1 through 101 to remove all apparent water dominated areas on the minimum value NDVI. The resulting water mask was used to derive a land NDVI data stack. The latter was subjected to a series of unsupervised classifications to map regionally evident agriculture and forest LULC types.

ERDAS Imagine software was used to derive all unsupervised classifications, using an ISODATA clustering routine. The classification parameters were setup as follows. (1) The convergence level to 99.95 percent, (2) the total number of iterations to 100, (3) means initialization along the principal axis, and (4) scaling range using the auto option. In addition, the number of clusters per classification run was set according to the purpose of the classification (e.g., whether the classification was for the whole or a subset of the map area or for a subset of targeted LULC classes).

For unsupervised classifications of study area LULC based on MODIS NDVI data, each cluster class was assessed by comparison to available reference data and then assigned to the apparent predominant LULC class. Reference data for cluster class assessment and LULC class assignment



included Landsat RGB multispectral image displays, MODIS NDVI temporal data stacks, Google Map and Bing aerial/satellite high resolution imagery, and assorted reference data from the MRC. The latter included a land cover map for 2010 [13]; 1997 LULC map with class descriptions [9], 2010 survey data points, and field photography [35]; and crop calendars [36]. The Bing and Google Map imagery were viewed using QGIS software and its OpenLayers plugin [29]. Some LULC classes (e.g., classes with specific vegetation greenness phenology) required consideration of specific reference data (e.g., 2010 MODIS monthly NDVI data stacks). In some cases, geo-tagged field photography resident to Google Map was also considered.

Although used as a reference, the 2010 land cover map from the MRC [13] was not produced for aiding the SWAT model and does not include certain LULC classes (e.g., double-cropped rice) that are needed for optimizing SWAT modeling results for applicable LMB SBs 1–8. Other available LULC maps of the LMB region were also not used for this project because they did not include agricultural types needed for optimizing LMB SWAT. The mentioned MRC “2010” land cover map does not distinguish between single and double cropped rice agriculture types, which is regarded as suboptimal for the LMB SWAT models.

The “land” NDVI stack was next applied to produce an unsupervised classification containing 30 cluster classes. The latter was evaluated to identify clusters that were either agriculture or forest dominated. This classification was then recoded to derive two preliminary masks, including one each for apparent agriculture and forest areas. The masks were then used to subset the land NDVI stack into a NDVI stack for agriculture and another for forest. Each of these were then subject to a reclassification process known as cluster busting [21]. In doing so, Isodata unsupervised clustering was performed independently on agriculture and forest specific NDVI data stacks. The cluster busting process was used to reduce classification confusion for the agriculture and forests occurring in the study area.

The NDVI data for the agricultural areas was used to produce an ISODATA unsupervised classification with 24 total cluster classes. Each of the cluster classes was then assessed and assigned to the apparent predominant LULC class. To aid identification of single versus double-cropped rice areas, an analyst assessed the signatures of each agricultural cluster class that included generation of cluster class specific mean NDVI profiles. In the reclassification of agricultural areas, most of the cluster classes pertained to specific regionally evident agricultural types, including single cropped rice, double-cropped rice, and mixed annual crops that were not rice. A few of the cluster classes (4 total) pertained to forest dominated areas. A lookup table was then developed to recode cluster classes to LULC categories.

MODIS NDVI-based reclassification of the forested areas required a different approach because MODIS NDVIs for mountainous forests had some occasional atmospheric contamination during the rainy season months. Unfortunately, these artifacts persisted even after the data was temporally processed with TSPT. Given the natural forests were to be classified according to deciduousness that typically occurs during the dry season [37], only dry season NDVI data was used to refine the forest classifications, which included MODIS 32-day NDVI for 32-day dates 1–4 and 10–12. This subset of dates enabled an unsupervised cluster classification of forests that was more visibly effective compared to what could be derived using all 12 dates.

Consequently, the forest 2010 dry season NDVI data stack was used with ISODATA clustering to derive a classification of 16 cluster classes. These were then analyzed using reference data and recoded into LULC types that included four types of forest according to deciduousness level and another related shrubland class that includes mixtures of very young forest regeneration and abandoned, old fallow former agricultural areas.

After preprocessing described earlier, the Landsat mosaic (Figure 6) was employed to classify additional LULC types that were less regionally common and/or finer scaled compared to MODIS qkm NDVI data. LULC classes mapped with Landsat data included water, barren, urban, bamboo scrub/forest, and industrial forest plantation patches. Unsupervised classification via ISODATA clustering was used to derive these products, though Area of Interest (AOI) vector polygons were also

used in some cases to either isolate areas with targeted LULC types or else to reduce classification errors. Initially, the Landsat mosaic was subjected to unsupervised ISODATA clustering with the output containing 40 cluster classes. While some of the cluster classes were associated with the targeted LULC types, additional classification refinement was performed for each Landsat-based LULC class prior to integrating into the final LULC map. More information on mapping of specific LULC classes from Landsat data is given below.

Each LULC class mapped with Landsat 30 m data was recoded into a binary map in which pixels representing the targeted class were set to 100. Afterwards, each 30 m binary map was spatially averaged 8 by 8 pixels with the ERDAS Imagine Degrade function, resulting in a 240-m map indicating the relative fractional abundance of the targeted LULC type. A LULC class specific threshold was then applied to derive a final map of each targeted LULC type. Water, urban, and industrial forest plantation areas were mapped when the per pixel fractional abundance of the targeted LULC type equaled or exceeded 25 percent. Barren and bamboo forest areas were included on the final LUC map when their per pixel fractional abundance equaled 50 percent or more. The water, urban, and IFP classes were assigned a lower fractional abundance threshold because they tended to occur in finer scaled patches that would have been eliminated otherwise. The fractional abundance thresholds were set to a higher value (50%+) for the bamboo forest and barren areas because these types in the study area appeared to primarily occur as larger patches relative to the more frequently finer scaled water, urban, and IFP patches.

The cluster classes representing water were recoded into a binary map that was then used as a mask to isolate Landsat reflectance data for water areas. The latter data was then used to derive an unsupervised classification of 20 cluster classes of which 16 pertained to predominantly water areas. The latter was recoded into a 1 class binary map of water set to a DN of 100 that was then spatially averaged to a 240 m spatial resolution.

The 40-class wall to wall cluster map also contained two clusters that included areas with industrial forest plantations (IFP). Most of the IFP areas were in a highly green state on the Landsat mosaic and thus easy to visualize. The IFP clusters were recoded into a binary map to a DN of 100 that was then spatially averaged to 240 m and edited with AOIs to reduce classification errors.

Urban areas were mapped with Landsat data by initially using AOIs to isolate and subset Landsat data with commonly occurring urban development. The subset imagery was then subjected to an unsupervised classification that contained 30 cluster classes of which four were predominantly urban. The signatures from the unsupervised classification were then input into a maximum likelihood supervised classification routine, along with the entire Landsat mosaic, to produce a wall to wall classification. The urban classes for the latter were recoded into binary urban map to a DN of 100 that was then subjected to an error reduction routine. Errors due to bright bare soil in agricultural areas were reduced by using a spatial model that regarded urban areas to be those mapped by Landsat that also had an annual 2010 MODIS 32-day peak NDVI of less than 160 but greater than or equal to 106.

Bamboo scrub/forest areas were then classified from the Landsat data, using the initial 40 class unsupervised classification to identify four cluster classes containing bamboo areas. Such scrub/forests have a comparatively unique spectral tone and generally occur in higher elevations amongst broadleaved forest types. These bamboo cluster classes also contained some commission error, which was reduced using screen digitized AOIs to more clearly define locations where such areas typically occur. The bamboo scrub/forest clusters were recoded into a binary map to a DN of 100 that was then spatially averaged to 240m and edited with additional AOIs to reduce classification errors.

Once the individual LULC classes were mapped, a series of raster GIS-based spatial models within ERDAS Imagine were then constructed and applied sequentially to compile a master final LULC map. Six individual models were used to compile the final LULC map in the following sequential steps that added: (1) MODIS-based agriculture and forest LULC classes; (2) MODIS-based shifting cultivation areas; (3) Landsat-based urban zones; (4) Landsat-based industrial forest plantations; (5) MODIS and

Landsat-based open water surfaces; and (6) Landsat-based bamboo scrub/forest patches. The final LULC map was coded according to a standard LULC classification scheme and color table.

The 2010 LULC map from this project was then compared visually to the MRC 1997 LULC map. To facilitate this comparison, the 1997 LULC map were recoded to approximate the 2010 LULC classification scheme, given that the 1997 contained some differently defined LULC classes and contained fewer agricultural LULC classes than the 2010 LULC map.

### 2.5. Method for LULC Accuracy Assessment

The accuracy assessment for the SB 7 LULC map was performed as described previously [38,39]. The work flow for this LULC map accuracy assessment is depicted in Figure 4. The main goal of the LULC map accuracy assessment was to assess the overall accuracy of the final map. A secondary goal was to provide some basic information on individual LULC class mapping performance, knowing that the total samples per scarce individual LULC class was too small for an in-depth assessment of individual class accuracy. The total sampling samples drawn overall and per class considered in part the resources and time available to do the accuracy assessment.

The random selection of sample locations was also designed to assess performance of LULC classifications within the interior of mapped LULC patches as opposed to edges of patches. In doing so, LULC class sample locations were drawn randomly using a 3 by 3-pixel window in which candidate center pixel was compared to the surrounding 8 pixels. Random sampled pixels were selected for accuracy assessment when (1) the center pixel was the targeted class and (2) most of the pixels (at least 5) in the 3 × 3 window were of the same LULC type as the center pixel in the 3 × 3-pixel window.

Initially the sample design was for a stratified random sample of 250 total pixels that were drawn from the LULC classification using ERDAS Imagine software. The number of samples drawn per LULC class was initially based on the observed class frequency on the final LULC map. In doing so, the sample size allocated to each stratum was proportional to the percent area of the stratum relative to the total mapped area. The number of random samples drawn per class depended in part on the LULC class frequency and map extent, as well as the need when feasible to assess at least 10 sample locations per scarce LULC class (Table 1).

**Table 1.** 2010 LULC class frequency versus samples drawn per class for the SB 7 final map.

LULC Class Description	# Pixels	Frequency	Proportional Allocation per 250 Total Samples	# Drawn Samples
Water	22,648	2.5%	6.3	20
Urban	8005	0.9%	2.2	15
Agriculture—rice—one crop per year	428,365	47.7%	119.3	123
Agriculture—rice—two crops per year	43,461	4.8%	12.1	20
Agriculture—mixed annual crops—other than rice	201,494	22.4%	56.1	58
Agriculture—shifting cultivation—cleared in 2010	1012	0.1%	0.3	2
Agriculture—shifting cultivation—partially cleared in 2010	421	0.0%	0.1	0
Deciduous shrubland—mixed Scrub/herbaceous/low broadleaved forest	55,918	6.2%	15.6	20
Forest/scrub—deciduous broadleaved—low height	56,569	6.3%	15.8	20
Forest—deciduous/evergreen broadleaved—low/medium height	19,325	2.2%	5.4	20
Forest—evergreen broadleaved—medium/tall height	33,597	3.7%	9.4	20
Forest—evergreen/deciduous broadleaved—low/medium height	25,620	2.9%	7.1	20
Bamboo scrub/forest—low height—mostly evergreen	621	0.1%	0.2	3
Industrial forest plantation—low/medium height	734	0.1%	0.2	1
TOTAL	897,790	100%	250	342

In drawing the sample, the observed LULC classes on the final map were categorized according to four frequency classes: (1) rare = 0 to 0.5 percent of total mapped area; (2) scarce = 0.5 to 10 percent; (3) moderately common = 10 to 20 percent; and (4) highly common = 20 percent or more. Most of the observed LULC classes were relatively scarce compared to the apparent two most frequent types in the



final LULC map for SB 7 (Table 1). These two most common LULC types together represented about 70 percent of the classified pixels. A few of the LULC classes were rare in occurrence. To help improve information on the agreement of individual scarce classes, additional sample locations were drawn for the scarce classes. Fifteen to 20 sample locations were drawn for each scarce class. Doing so enabled the number of sample locations drawn per class to be either equal to or exceeding what the sample size would have been if the sampling allocation was strictly proportional to class frequency. The randomly drawn number of sample locations per common LULC classes were roughly proportional to class frequency for a total sample size of 250. The rare LULC classes were sampled proportional to the class frequency. A total number of 342 random sample locations were drawn considering all LULC classes.

The sample locations for a given SB once selected were converted from an ERDAS Imagine tabular format to spreadsheet and GIS formats using MS Excel, ArcGIS, and QGIS software. QGIS was used to do assess the apparent LULC class for each drawn sample location. Individual points were displayed in QGIS at a 1:10,000 scale according to sample identification # using a zoom to point function and then compared to reference data discussed in Section 2.4.

Reference imagery used in the map accuracy assessment included Landsat false color RGB mosaic imagery from the dry season of 2009 to 2011. The date of the Landsat data was considered when interpreting sample locations. Also, the apparent vintage of the Google Map satellite imagery, geo-tagged Google Map field photos, and Bing aerial imagery were assessed to make sure that high spatial resolution imagery showed comparable LULC patterns relative to the vintage of the MODIS and Landsat imagery used to produce the final LULC map. Other reference documentation, such as field survey-based notes and example field photos from the MRC was also used.

The Landsat RGBs were used as a general reference to help distinguish between general cover types (e.g., forest versus agriculture) and certain specific cover types that were visually apparent (e.g., urban, water, and bamboo forests). For some of the LULC types (e.g., agricultural types), other reference data had to be used, such as the higher spatial resolution aerial and satellite imagery resident to either Bing or Google Maps. In most cases, multiple sources of reference data were cross compared to assess the apparent predominant LULC type for each randomly sampled pixel.

To assess LULC class within a given sample pixel, the QGIS grid overlay display function was employed to view a given sample location boundary as a vector outline in relation to reference imagery. This display function was used to assess the predominant LULC class for each sample location. In doing so, the grid was proportioned so that it has the same dimensions as the MODIS pixel (~231.65 by 231.65 m). To center the grid around each sample point, one had to apply offsets in the X and Y direction. A trained analyst interpreted and recorded the apparent LULC for each sample location by comparing the location to reference data and determining the predominant LULC type. In using the grid function, the display map projection was set to the Pseudo Mercator map projection so that the dimensions of the grid could be aligned with the boundary of a given sample MODIS qkm pixel.

The LULC map accuracy assessment for certain LULC classes required use of specific reference data. For example, the assessment of sample locations with apparent agricultural and forest LULC types necessitated consideration of the 32-day MODIS NDVI data to assess phenological greenness characteristics. MODIS data was used to assess vegetation greenness phenology, given it was the only synoptic Earth Observation data available at a high temporal resolution for assessing such conditions. However, Landsat, Bing, and Google Maps image data was also used to help assess such sample locations in a supplemental capacity, along with LULC data (e.g., agricultural crop calendars and LULC class descriptions) from the MRC.

Sample locations of rice agricultural areas were assessed in part using pixel specific 32-day MODIS NDVI profiles for 2010 to determine if such areas were cropped one or two times per year. In doing so, a spectral/temporal profiling plugin for QGIS was used to view NDVI profiles of specific pixels. Landsat data was also employed for assessing agricultural type locations, especially with respect to rice types. Although rare in occurrence, sample locations for apparent shifting cultivation agriculture

class were assessed using MODIS monthly NDVI data for dates 12 versus 1 of 2010, along with dry season Landsat data of comparable vintage (to the 2010 MODIS data).

Randomly sampled locations of apparent forest were mainly assessed for deciduousness by considering MODIS monthly NDVI values from the dry season of 2010, including 32-day NDVI dates 1–4 and dates 10–12. Such NDVI data were used to quantify two deciduousness indicator metrics for sampled forested pixels: (1) percent NDVI change comparing the median versus maximum NDVI and (2) percent NDVI change comparing the sum of the three lowest versus sum of the three highest NDVIs. Dry season NDVIs were used because of observed higher noisiness of the NDVIs in mountainous forests during the rainy season. Given the same time of year and location, residual unmasked cloud contaminated NDVIs tend to be lower than cloud free NDVIs and can contribute to suboptimal LULC classifications. Consequently, only dry season NDVI data was used to map the forest LULC classes.

For each applicable sample location, the calculated values for each of the two deciduousness indices were compared to decision rules to determine forest class according to an observed level of deciduousness (Table 2). The thresholds for the different deciduous forest classes were determined previously by analysis of signatures for each predominantly forest cluster class that were derived from the MODIS NDVI unsupervised classification (described in earlier section on LULC map development).

**Table 2.** Decision rules used in classification of forest deciduousness, based on comparison of MODIS monthly (32-day) NDVI data for peak senescence versus peak greenness observed during the dry season months.

LULC Forest Deciduousness Class	NDVI Change Level for Senesced versus Peak Green State
Deciduous Forest/Scrub	$\leq -15\%$ change in NDVI
Deciduous/Evergreen Forest	$\leq -10\%$ change in NDVI and $> -15\%$ change in NDVI
Evergreen/Deciduous Forest	$\leq -5\%$ change in NDVI and $> -10\%$ change in NDVI
Evergreen Forest	$\leq 0\%$ change in NDVI and $> -5\%$ change in NDVI

For each apparent forest sample location, a trained analyst assessed the deciduousness level using two deciduousness indicators. The latter were calculated based on MODIS 32-day NDVI data values observed for each sample location considered to be forest. One of these deciduousness indicators (percent change based on median versus maximum NDVI from MODIS data) was used primarily to estimate deciduousness of each apparent forest sample location. For non-obvious borderline cases, estimates from the second index (change based on sum of the three lowest versus sum of three highest dry season NDVIs) were also considered. In addition, Landsat, Google Maps, and Bing imagery were viewed for each sample location. The use of Landsat, Google, and Bing imagery included consideration of the possibility that there may be temporal decorrelation between such data and the MODIS data. A given form of reference data was considered only if it appeared to be of similar vintage to the data used to construct the 2010 LULC map.

Given the cloud frequencies for the region and availability of cloud-free data, Landsat imagery was used as a secondary data source for assessing forest deciduousness. The Landsat data used for a given year was from a single date acquired during the dry season. As such, the timing of peak deciduousness is not necessarily recorded or else highly apparent on the observed Landsat image data. The deciduousness may be only subtly indicated on a given date of dry season Landsat data, depending on the selected date and whether the dry season was wetter, dryer or normal in nature. Consequently, the employed “single snap shot in time” Landsat data did not always show the full degree of forest deciduousness—only what was evident for the acquired date. In addition, even for the selected “best available” Landsat data sets, clouds and shadows sometimes occluded areas of interest.

Results of a given SB LULC accuracy assessment (e.g., for SB 7) were then summarized into error matrices that were subsequently used to calculate overall agreement as well as individual class agreement in terms of user and producer agreement. For a given classification scheme specificity, the overall agreement was then adjusted for random chance effects via use of the Kappa statistic.

### 3. Results

The final 2010 LULC map for SB 7 is shown in Figure 7. This map contains 14 LULC classes that varied in frequency from rare to scarce to common.

The results of LULC map accuracy assessment are given as error matrices in Tables 3 and 4. The error matrix for LULC classification at full classification scheme specificity is shown in Table 3, whereas Table 4 reports the error matrix for a more simplified LULC classification in which deciduous forest classes were grouped into one singular class.

**Table 3.** SB7 LULC map error matrix at full scheme specificity for 342 total sample locations (PA = Producer Agreement and UA = User Agreement).

SB 7 LULC Map	LULC Class on Reference Data													Total	% UA	
	C10	C16	C21	C22	C23	C25	C31	C32	C33	C34	C35	C36	C42			
C10	18														18	100
C16		13	2												15	87
C21	2	1	109	1	3										116	94
C22		1		19		2									22	86
C23			9		50		5	2							66	76
C25						0									0	N/A
C31			3		2		15								20	75
C32					2			17	10						29	59
C33									9		10				19	47
C34									1	15	2				18	83
C35										5	8				13	62
C36												3			3	100
C42					1			1					1		3	33
<b>Total</b>	20	15	123	20	58	2	20	20	20	20	20	3	1		342	
<b>% PA</b>	90	87	89	95	86	0	75	85	45	75	40	100	100			

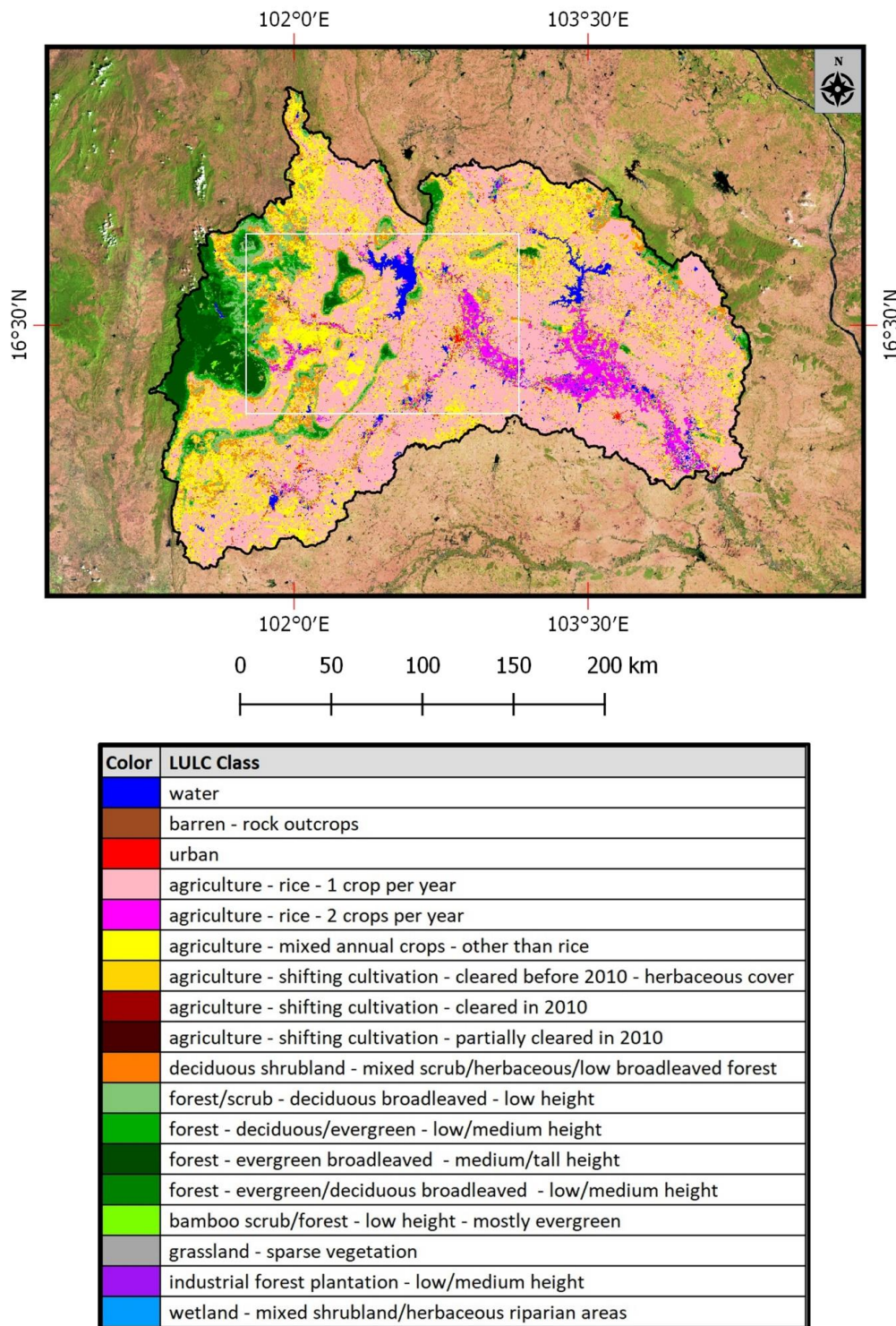
C10 = Water (class # 10 in LULC classification scheme). C16 = Urban. C21 = Agriculture—rice—one crop per year. C22 = Agriculture—rice—two crops per year. C23 = Agriculture—mixed annual crops—other than rice. C25 = Agriculture—shifting cultivation—cleared in 2010. C31 = Deciduous shrubland—mixed scrub/herbaceous/low broadleaved forest. C32 = Forest/scrub—deciduous broadleaved—low height. C33 = Forest—deciduous/evergreen—low/medium height. C34 = Forest—evergreen broadleaved—medium/tall height. C35 = Forest—evergreen/deciduous broadleaved—low/medium height. C36 = Bamboo scrub/forest—low height—mostly evergreen. C42 = Industrial forest plantation—low/medium height.

**Table 4.** SB7 LULC map error matrix for scheme with only one deciduous forest class instead of three (PA = Producer Agreement and UA = User Agreement). This matrix is identical to the one shown in Table 3 except for the deciduous forests being grouped as a singular class. See Table 3 for descriptions of the non-merged class numbers in the table below.

SB 7 LULC Map	LULC Class on Reference Data											Total	% UA		
	C10	C16	C21	C22	C23	C25	C31	X01	C34	C36	C42				
C10	18													18	100
C16		13	2											15	87
C21	2	1	109	1	3									116	94
C22		1		19		2								22	86
C23			9		50		5	2						66	76
C25						0								0	N/A
C31			3		2		15							20	75
X01					2			54	5					61	89
C34								3	15					18	83
C36										3				3	100
C42					1			1				1		3	33
<b>Total</b>	20	15	123	20	58	2	20	60	20	3	1			342	
<b>% PA</b>	90	87	89	95	86	0	75	90	75	100	100				

X01 = all three classes of deciduous forest grouped into one (based on C32, C33, and C35 as individually described in Table 3).





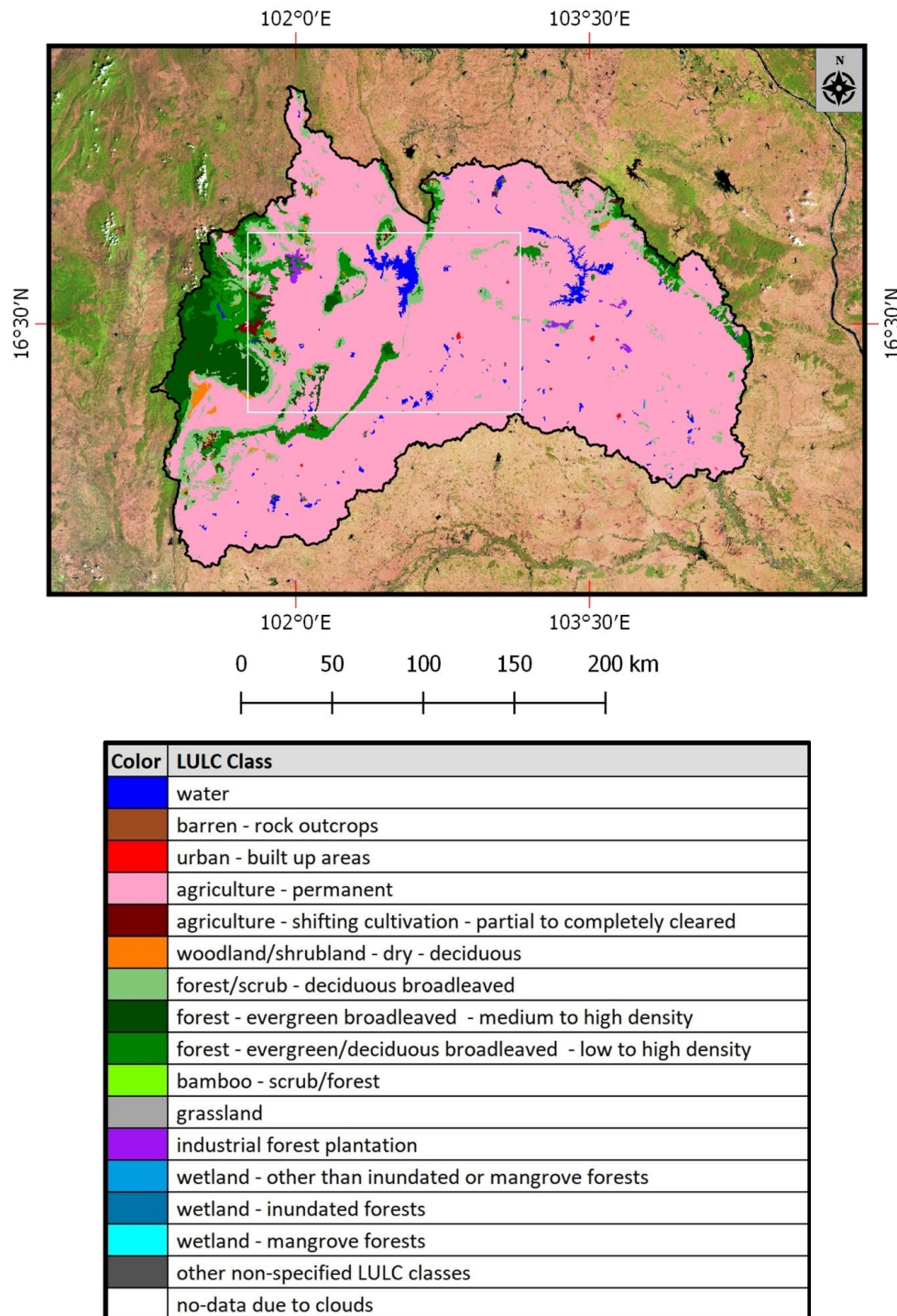
**Figure 7.** 2010 LULC map of study area produced by project. The study area is outlined in black. The white rectangle shows location of Figure 9.

The overall agreement for the SB7 2010 LULC map at full scheme specificity was ~81% (with Kappa coefficient of 0.77) compared to available reference data (Table 3). This level of overall agreement is moderately high and acceptable in that it exceeds 80%. The overall agreement improved to ~87% (with Kappa value of 0.84) when deciduous forest classes were all grouped into one class (Table 4).

At full scheme specificity, the LULC map shows a moderately high level of overall agreement (81%) with reference data. Scarce to common agricultural classes yielded moderate to high agreement

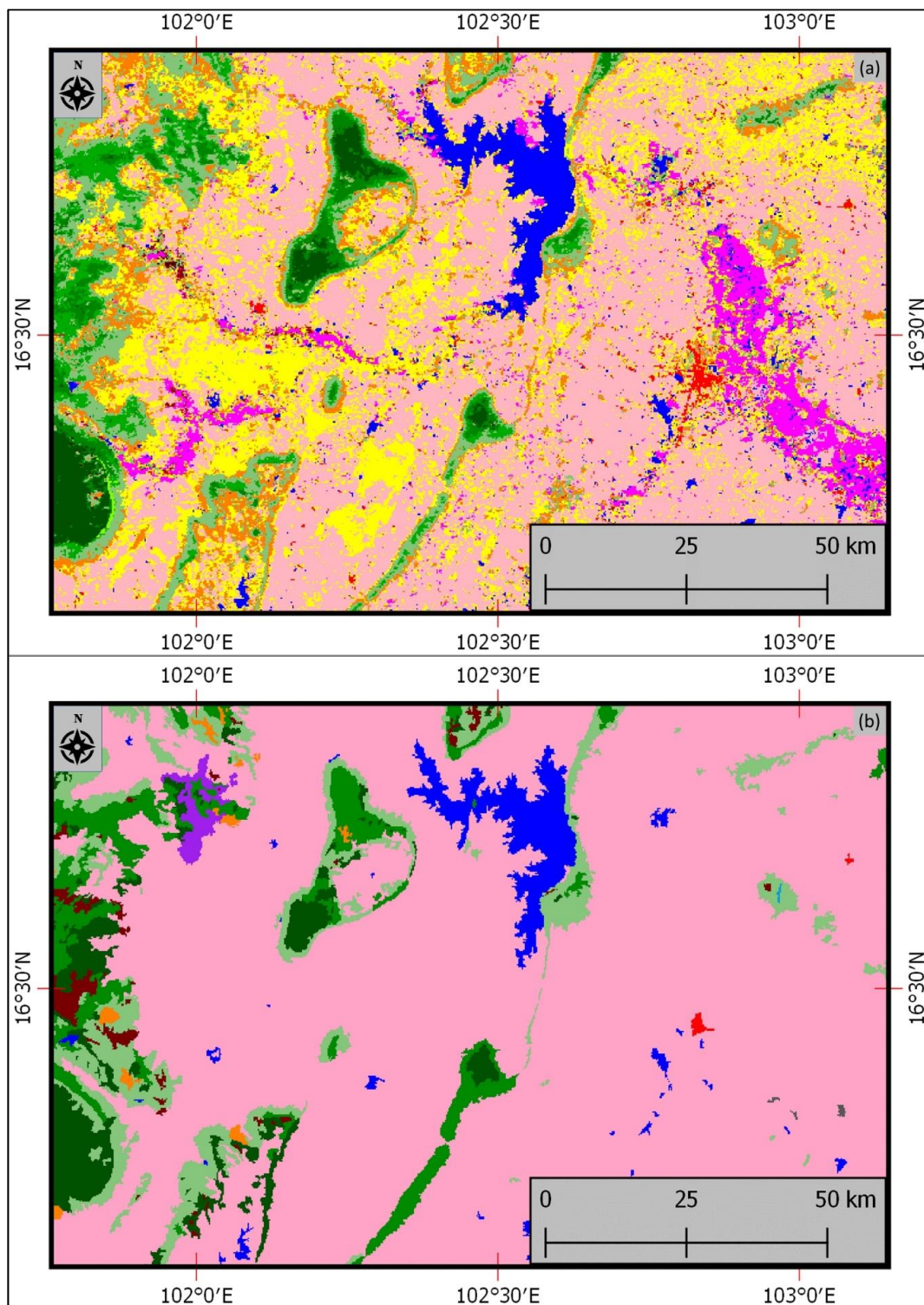
with reference data. Single and double-cropped rice classes had user and producer agreements that exceeded 85%. The mixed annual crop other than rice LULC type had a high producer agreement (86%) and a moderately high user agreement (76%).

Figure 8 displays a version of the MRC 1997 LULC map that was recoded to approximate the classification scheme used for the 2010 LULC map generated with MODIS and Landsat data (Figure 7). Figure 9 shows an enlarged view of the 2010 versus the 1997 LULC maps.



**Figure 8.** Recoded version of the 1997 LULC map of study area (latter outlined in black). The shifting cultivation and forest classes were recoded to aid comparison to the 2010 LULC map. The white outline shows location of Figure 9.





**Figure 9.** Enlarged subset of study area, comparing the (a) 2010 LULC map from the project to (b) recoded 1997 LULC map from MRC. The legend for the 2010 LULC map is given in Figure 7 and the legend for the 1997 LULC map is shown in Figure 8.

#### 4. Discussion

Most of LULC classes were scarce relative to the overall geographic extent of SB 7 and the two most common LULC classes (single annual rice crops and mixed annual crops other than rice). However, the regionally-scarce classes still represented sizeable areas in terms of mapped hectares and were also



sometimes locally common. For example, double-cropped rice and certain native forest types were sometimes common in certain locations but regionally-scarce on the final LULC map.

Based on the accuracy assessment results, the monthly MODIS NDVI data in conjunction with unsupervised classification appeared to be useful for mapping the more common permanent agricultural cover types in SB 7, particularly the permanent agriculture classes (i.e., irrigated rice, rainfed rice, and annual crops other than rice such as row crops). This is an improvement in terms of increased agricultural classification scheme specificity compared to the preceding LULC map being used for SWAT modeling—the 1997 map maintained by the MRC. The latter shows all forms of permanent agriculture as one singular class. The SB 7 area does not contain extensive shifting cultivation areas, which are more evident in certain other SBs, such as those in Laos PDR. More research is needed to assess the accuracy of using MODIS monthly NDVI data to classify the shifting cultivation LULC types.

In this study, regionally-scarce in occurrence, shrubland, and forest LULC classes showed variable agreement with reference data. The scrub/shrub and evergreen forest types had user and producer agreements that were either moderately high or high. However, the deciduous forest classes (all deciduous, mixed deciduous/evergreen, and mixed evergreen/deciduous forest types) were more frequently confused with each other than hoped. The MODIS data used in LULC classification showed an improved ability to map forest classes according to its deciduousness level once the deciduous related forest classes were reduced from three to one class.

Deciduous forest LULC classification confusion may be due to multiple factors, including map production and evaluation methods. At the outset of the project, it appeared that deciduous forest could be separated via unsupervised image classification of monthly MODIS NDVI data for a calendar year into deciduous forest, mixed deciduous/evergreen forest, and mixed evergreen/deciduous forest types. Such types are known to occur in Thailand based on descriptions by the authors of a previous paper [24]. On the full classification scheme LULC map, the three-mapped forest deciduous classes tended to be confused with each other. The method for mapping deciduous forest classes in part considered forest deciduousness in relation to cluster class mean 32-day NDVIs for the dry season months. The cluster class signatures also included variability with respect to mean NDVIs per cluster class that may be a factor contributing to the observed confusion between deciduous forest LULC classes. As a result, additional accuracy assessment of forest LULC classes in other SBs with more frequently occurring forests is desirable.

It is possible that the classification of such forest classes may be improved by additional classification and cluster class signature analysis. Also, the forests in SB 7 were much less frequent in occurrence than the main agricultural types. The agricultural classes in some portions of the study area visibly fragmented some of the forested areas, given also that some of the Khorat plateau area was formerly forest prior to agricultural conversion. It is conceivable that the classification of deciduous forest types could improve in other SBs of the LMB in which there is greater representation and perhaps less fragmentation of such forest types.

The overall agreement of the SB 7 LULC map with reference data improved from 81 to 87% when deciduous forest classes were regrouped from three into one class. Once merged, the single deciduous forest class yielded high user and producer class agreement.

Even with only one deciduous forest class, the use of MODIS dry season 32-day NDVI data was able to map two native forest types (deciduous and evergreen) with moderately high to high levels of agreement to available reference data (with user or producer accuracies ranging from 75 to 90%). The ability to monitor forest deciduousness across multiple dates of the dry season was not feasible with Landsat 5 data, due to the low temporal resolution of and the inconsistent availability of cloud-free data (even during the dry season).

The project's LULC map at both levels of classification scheme specificity yielded an overall agreement with reference data that exceeded 80%. This amount may be conservatively estimated, given that the scarce LULC types were oversampled relative to their percent occurrence. A good land cover map for a given purpose can be deemed as good or acceptable if it has an overall accuracy greater

than or equal to 80 to 85% overall accuracy. The latter is comparable to some land cover mapping products that have specifications for accuracy, such as NOAA Coastal Change Analysis Program (C-CAP) land cover maps with multiple wetland classes for the temperate United States. However, the latter C-CAP maps are from Landsat data and do not include as many agricultural classes, compared to the LULC maps we produced for the LMB region in Southeast Asia.

The use of monthly NDVI data derived from twice daily MODIS data collections helped to classify deciduous versus evergreen forest based on seven dates of monthly NDVI for the dry season months of 2010. From a temporal resolution perspective, it appears that the seven dates of MODIS 32-day data enabled an improved means to classify forest deciduousness over using one or two dates of Landsat data, especially given that the duration of the leaf-off state of deciduous forests in Thailand can vary from 2 to 21 weeks [37].

Even though the project used dry season Landsat 5 data from 2009 to 2011 for LULC classifications (Figure 6), the cloud frequency in the study area was a factor in assembling the Landsat mosaic in that there was insufficient cloud free data to compile a “clear view” mosaic just from 2010 data. As a result, dry season data from 2009 to 2011 was used instead. Based on this study, the use of a single date of Landsat data per pixel for classifying forest deciduousness can only produce deciduous and evergreen forest maps relative to the foliar state of the forest at the time of data acquisition. The study area mosaic included portions of six Landsat scenes from three swaths. The mosaic based on scenes from three different years and therefore was less suitable for classifying deciduousness forest types.

Given that forest deciduousness patterns can vary from one year to the next, 2010 dry season MODIS 32-day NDVIs (Figure 5) was used instead of Landsat data (Figure 6) for mapping native forest types. Landsat false color imagery was useful as an ancillary data source for aiding the interpretation of relative forest deciduousness in the study area however.

In addition, Landsat data showed utility for mapping some of finer scaled LULC types in the study area that are difficult to map as accurately with MODIS data. For example, the Landsat data was used to map regionally infrequent but locally common LULC types. In this study, regionally-scarce water and urban classes both showed high agreement with the reference data. The water class was derived using both MODIS and Landsat data, while the urban class was mapped entirely from Landsat data. Landsat helped to improve effective mapping of finer scaled water bodies. The rare LULC classes recorded on the final LULC map showed variable levels of agreement with reference data, though the very small sample size and class rarity were factors that impeded the ability to fully assess the accuracy of such LULC classes. Some of the rare LULC classes for SB 7 (e.g., industrial forest plantations and shifting cultivation classes) are more common in other SBs of the LMB. Additional accuracy assessment of the rare LULC classes on the SB 7 map is therefore recommended using LULC maps produced from this project for other SBs in the LMB.

The ability to map the specific land cover types at the MODIS qkm spatial resolution is useful for aiding SWAT LMB hydrologic modeling applications given that (1) it is still more spatially resolute than some other SWAT inputs, such as soils and (2) stream flow and runoff characteristics depend in part on the LULC type, according to US Department of Agriculture (USDA) published runoff curves [40]. For example, the runoff from forests is much less than that from urban areas dominated by impervious surfaces. In addition, the qkm resolution 2010 LULC map is also at a finer scaled MMU (0.0625 km<sup>2</sup>) compared to the MRC LULC map for 1997 (latter with a 0.5 km<sup>2</sup> MMU according to [9]). Consequently, the finer resolution MMU for the 2010 LULC map tended to enable more smaller sized patches to be more discretely mapped compared to the 1997 LULC map.

The 1997 LULC map only classifies permanent agriculture as single class (Figure 8), while the 2010 LULC map from our project contains multiple permanent agricultural LULC classes (Figure 7). Both dates of LULC maps show forests generally occurring in the same locations, though the 1997 LULC map has a few locations with shifting cultivation that have apparently reverted to some form of woody vegetation on the 2010 map (Figure 9). The latter map shows some urban expansion since 1997 and apparent finer scale definition of water bodies.

Given that the 1997 LULC map includes a single permanent agricultural class, a 2010 versus 1997 LULC change map at the 2010 classification scheme specificity was not possible. However, in terms of future work, it is currently feasible to update the 2010 LULC map computed in this project, given that MODIS and Landsat are both still collecting viable data. Also, alternatives to both MODIS (e.g., VIIRS) and Landsat (e.g., Sentinel-2) currently exist.

The project's LULC mapping method is mostly based on unsupervised classification techniques. It makes use of some supervised techniques, though mainly in terms of how the unsupervised method is applied. Singular use of the supervised approach can be advantageous if there is enough ground reference data to characterize the spectral variability of the targeted LULC types if one uses multispectral imagery like Landsat as a classification input - or the vegetation greenness variability if one uses a year's worth of monthly MODIS NDVI as an input to classification. For this study, the supervised method was not used due to insufficient training data to account for the apparent spectral variability for some of the targeted types. Also, the cloud frequency on the Landsat data was problematic for singularly using the supervised method in that there was only cloud free data during part of the dry season. That time of year does not necessarily clearly show where all the targeted LULC types are located. In addition, the MODIS data's spatial coarseness is a factor impeding the singular use of the supervised method.

## 5. Conclusions

The project combined use of 2010 MODIS 32-day NDVI and Landsat data from 2009 to 2011 to produce an updated LULC map for SB 7 of the LMB. The final LULC map for this SB yielded high overall agreement with available reference data. The MODIS NDVI data was used to map regionally evident LULC types, including certain agricultural and forest types that had specific vegetation greenness phenology signatures. Such LULC classes could not be mapped using monthly composited Landsat data, given that insufficient cloud free data exist for multiple portions of a given calendar year. The MODIS data used in this study enabled regionally common agricultural LULC classes to be mapped at moderately high to high levels of agreement with reference data. The results indicated that deciduous and evergreen broadleaved forest can be classified with MODIS NDVI data at a moderately high to high level of agreement with reference data. The results suggest that MODIS NDVI data in conjunction with image classification techniques was not able to consistently map multiple deciduous forest classes in terms of pure deciduous, mixed deciduous/evergreen, and mixed evergreen/deciduous forest types. However, by combining the three deciduous forest classes into one, the technique was able to map a single class of deciduous forest/scrub with a high level of agreement to reference data.

Landsat data appeared to be effective for mapping certain fine scaled LULC classes that were less detectable on the 1 km MODIS NDVI data, including urban and fine scaled waterways. The Landsat data showed some potential for mapping other finer scaled LULC types (e.g., IFP areas) output at the MODIS 1 km scale, though additional quantitative accuracy assessment of other SBs is needed to confirm this belief.

The project yielded a new technique for mapping tropical deciduous forest classes that included utilization of forest deciduousness indicators computed from 2010 MODIS 32-day NDVI data. The latter data was useful for viewing per pixel fluctuations in vegetation canopy greenness that occurs during the LMB dry season when forest leaf drop primarily occurs.

The project enabled updated LULC maps to be produced for all eight SBs of the LMB that are currently being analyzed with MRC SWAT hydrologic models. The LULC maps from the project are being used to help SWAT modelers improve predictions of water flow dynamics that in turn is being used for aiding water management and basin planning in the LMB. Updated LULC maps along with soils, terrain, and precipitation maps from the project are being used to generate needed SWAT modeling products [41–43]. The LULC maps from the project also provided more specific maps of rice paddies (single- versus double-cropped or rainfed versus irrigated) that are potentially helpful for



aiding other basin planning efforts in the LMB, including flood damage monitoring and assessment applications [44,45].

**Author Contributions:** J.S. processed and assessed land cover mapping products in consultation with coauthors J.B., S.R., and V.L. The manuscript was developed primarily by J.S. and later revised with advice from coauthors J.B., S.R., and V.L.

**Funding:** This work was funded primarily from a 2011 NASA ROSES A.33 Earth Science Applications: Disasters grant (NNH11ZDA001N-DISASTER). Additional funding was received through a NASA Goddard Space Flight Center (GSFC) Applied Sciences grant (NNG15HQ01C) to Science Systems and Applications, Inc. (SSAI) that supports the Hydrospheric and Biospheric Sciences (HBS) code 610 at NASA GSFC in various research and engineering activities.

**Acknowledgments:** The refinement of the land use land cover mapping products in this project was aided by helpful, timely comments from Ibrahim Mohammed. Much of the MODIS monthly NDVI data used in this project was generated previously through a project conducted at NASA Stennis Space Center in support of the USGS Forecast Mekong project. Opinions, findings, and conclusions or recommendations given in this work are those of the author(s) and do not necessarily reflect the views of NASA, SSAI, and or other assisting organizations.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Rossi, C.G.; Srinivasan, R.; Jirayoot, K.; Le Duc, T.; Souvannabouth, P.; Binh, N.; Gassman, P.W. Hydrologic evaluation of the Lower Mekong River Basin with the soil and water assessment tool model. *Int. Agric. Eng. J.* **2009**, *18*, 1–13.
- Food and Agriculture Organization (FAO). *Irrigation in Southern and Eastern Asia in Figures: AQUASTAT Survey—2011*; FAO Water Report 37; FAO: Rome, Italy, 2012.
- Mekong River Commission (MRC). *State of the Basin Report 2010*; Mekong River Commission: Vientiane, Laos, 2010.
- Rodriguez, F. *Annual Report 2014*; Mekong River Commission: Vientiane, Laos, 2014.
- Pearse-Smith, S.W.D. The impact of continued Mekong basin hydropower development on local livelihoods. *Consilience* **2012**, *7*, 73–88.
- Mekong River Commission (MRC). *Crop Production for Food Security and Rural Poverty—Baseline and Pilot Modeling*; Mekong River Commission: Vientiane, Laos, 2014.
- Dat, N.D. *Model Setup in Mekong Basin*; MRC Modeling Team from the Information and Knowledge Management Programme; NASA: Washington, DC, USA, 2013.
- Mekong River Commission (MRC). *State of the Basin Report 2003*; Mekong River Commission: Vientiane, Laos, 2013.
- Malyvanh, M.; Feldkotter, C. Application of Remote Sensing and GIS for Forest Cover Monitoring in LAO P.D.R. In Proceedings of the Application of Resource Information Technologies in Forest Land & Resources Management, Hanoi, Vietnam, 10–20 October 1999; pp. 48–61.
- Leinenkugel, P.; Kuenzer, C.; Oppelt, N.; Dech, S. Characterisation of land surface phenology and land cover based on moderate resolution satellite data in cloud prone areas—A novel product for the Mekong Basin. *Remote Sens. Environ.* **2013**, *136*, 180–198. [[CrossRef](#)]
- Sakamoto, T.; Yokozawa, M.; Toritani, H.; Shibayama, M.; Ishitsuka, M.; Ohno, H. A crop phenology detection method using time-series MODIS data. *Remote Sens. Environ.* **2005**, *96*, 366–374. [[CrossRef](#)]
- Son, N.T.; Chen, C.F.; Chen, C.R.; Duc, H.N.; Chang, L.Y. A Phenology-Based Classification of Time-Series MODIS Data for Rice Crop Monitoring in Mekong Delta, Vietnam. *Remote Sens.* **2014**, *6*, 135–156. [[CrossRef](#)]
- Kityuttachai, K.; Heng, S.; Sou, V. *Land Cover Map of the Lower Mekong Basin*; MRC Technical Paper No. 59; Information and Knowledge Management Programme; Mekong River Commission: Phnom Penh, Cambodia, 2016.
- Spruce, J.P.; Sader, S.; Ryan, R.E.; Smoot, J.; Kuper, P.; Ross, K.; Prados, D.; Russell, J.; Gasser, G.; McKellip, R.; et al. Assessment of MODIS NDVI time series data products for detecting forest defoliation by gypsy moth outbreaks. *Remote Sens. Environ.* **2011**, *115*, 427–437. [[CrossRef](#)]
- Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]

16. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
17. Matsushita, B.; Yang, W.; Chen, J.; Onda, Y.; Qiu, G. Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest. *Sensors* **2007**, *7*, 2636–2651. [CrossRef] [PubMed]
18. Holben, B.N. Characterization of maximum value composites from temporal AVHRR data. *Int. J. Remote Sens.* **1986**, *7*, 1417–1434. [CrossRef]
19. Chen, J.; Jönsson, P.; Tamura, M.; Gu, Z.; Matsushita, B.; Eklundh, L. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sens. Environ.* **2004**, *91*, 332–344. [CrossRef]
20. Leinenkugel, P.; Oppelt, N.; Kuenzer, C. A new land cover map for the Mekong: Southeast Asia’s largest transboundary river basin. *Pac. Geogr.* **2014**, *41*, 10–14.
21. Jensen, J.R. *Introductory Image Processing: A Remote Sensing Perspective*; Prentice Hall: Saddle River, NJ, USA, 1996; 318p.
22. Spruce, J.P.; Smoot, J.C.; Ellis, J.T.; Hilbert, H.; Swann, R. Geospatial method for computing supplemental multi-decadal US coastal land use and land cover classification products, using Landsat data and C-CAP products. *Geocarto Int.* **2014**, *29*, 470–485. [CrossRef]
23. NASA JPL. NASA Shuttle Radar Topography Mission Global 1 Arc Second V003. NASA LP DAAC. Available online: <https://lpdaac.usgs.gov/node/527> (accessed on 22 September 2018).
24. Williams, L. *Vegetation of Southeast Asia Studies of Forest Types 1963–1965*; USDA ARS Prepared Report to the DOD Advanced Research Projects Agency, Order No. 121; U.S. Department of Agriculture: Washington, DC, USA, 1965; 302p.
25. Vermote, E. *MOD09Q1 MODIS/Terra Surface Reflectance 8-Day L3 Global 250m SIN Grid V006*; NASA EOSDIS Land Processes DAAC: Sioux Falls, SD, USA, 2015.
26. Vermote, E. *MYD09Q1 MODIS/Aqua Surface Reflectance 8-Day L3 Global 250m SIN Grid V006*; NASA EOSDIS Land Processes DAAC: Sioux Falls, SD, USA, 2015.
27. Ackerman, S.A.; Frey, R. *MODIS Atmosphere L2 Cloud Mask Product (35\_L2)*; NASA MODIS Adaptive Processing System, NASA Goddard Space Flight Center: Greenbelt, MD, USA, 2015.
28. USGS. GloVis. Available online: <https://glovis.usgs.gov/> (accessed on 29 August 2018).
29. QGIS. *QGIS Training Manual*; Release 2.14. 2017. Available online: <https://docs.qgis.org/2.14/pdf/en/QGIS-2.14-UserGuide-en.pdf> (accessed on 24 November 2018).
30. McKellip, R.; Ryan, R.E.; Prados, D.; Blonski, S. Crop surveillance demonstration using a near-daily MODIS vegetation index time series. In Proceedings of the 2005 International Workshop on the Analysis of Multi-Temporal Remote Sensing Images, Biloxi, MS, USA, 16–18 May 2005.
31. Prados, D.; Ryan, R.E.; Ross, K.W. Remote Sensing Time Series Product Tool. In Proceedings of the 2006 Fall Meeting, San Francisco, CA, USA, 11–15 December 2006.
32. McKellip, R.; Prados, D.; Ryan, R.; Ross, K.; Spruce, J.; Gasser, G.; Greer, R. Remote sensing time series analysis, a vegetation monitoring tool. *NASA Tech. Briefs* **2008**, *32*, 63–64.
33. Savitzky, A.; Golay, M. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
34. Chavez, P.S. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sens. Environ.* **1988**, *24*, 459–479. [CrossRef]
35. Nam, V.T.B.; Haase, M.; Kityuttachai, K.; Virak, S. *Land Cover Information Catalogue of the Lower Mekong Basin*; MRC Technical Paper, Information and Knowledge Management Programme; Mekong River Commission: Phnom Penh, Cambodia, 2015; 18p.
36. Halcrow Group Ltd. *SWAT and IQQM Models*; MRC Technical Reference Report DSF 620; Water Utilisation Project Component A: Development of Basin Modelling Package and Knowledge Base (WUP-A), Mekong River Commission: Phnom Penh, Cambodia, 2004.
37. Williams, L.J.; Bunyavejchewin, S.; Baker, P.J. Deciduousness in a seasonal tropical forest in western Thailand: Interannual and intraspecific variation in timing, duration and environmental cues. *Oecologia* **2008**, *155*, 571–582. [CrossRef] [PubMed]
38. Congalton, R.G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [CrossRef]

39. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2009; 183p, ISBN 978-1-4200-5512-2.
40. United States Department of Agriculture (USDA). *Urban Hydrology for Small Watersheds*; Technical Release 55, Version 2.31; USDA Natural Resource Conservation Service, Engineering Division: Washington, DC, USA, 1986; 164p.
41. Mohammed, I.N.; Bolten, J.D.; Srinivasan, R.; Lakshmi, V. Improved Hydrological Decision Support System for the Lower Mekong River Basin Using Satellite-Based Earth Observations. *Remote Sens.* **2018**, *10*, 885. [[CrossRef](#)] [[PubMed](#)]
42. Mohammed, I.N.; Bolten, J.D.; Srinivasan, R.; Lakshmi, V. Satellite observations and modeling to understand the Lower Mekong River Basin streamflow variability. *J. Hydrol.* **2018**, *564*, 559–573. [[CrossRef](#)] [[PubMed](#)]
43. Mohammed, I.N.; Bolten, J.D.; Srinivasan, R.; Meechaiya, C.; Spruce, J.P.; Lakshmi, V. Ground and satellite based observation datasets for the Lower Mekong River Basin. *Data Brief* **2018**, in press. [[CrossRef](#)]
44. Ahamed, A.; Bolten, J.D. A MODIS-based automated flood monitoring system for southeast Asia. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *61*, 104–117. [[CrossRef](#)]
45. Oddo, P.C.; Ahamed, A.; Bolten, J.D. Socioeconomic Impact Evaluation for Near Real-Time Flood Detection in the Lower Mekong River Basin. *Hydrology* **2018**, *5*, 23. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).