APPLICATION OF MODIS 250M IMAGES WITHOUT IN SITU OBSERVATIONS FOR MAPPING MEKONG RIVER BASIN LAND COVER

Kithsiri Perera ¹*, Srikantha Herath¹, Armando Apan¹, Lal Samarakoon³

¹Engineering and Surveying and Australian Centre for Sustainable Catchments University of Southern Queensland, West Street, Toowoomba 4350 QLD Australia; E-Mail: perera@usq.edu.au, apana@usq.edu.au
²Institute for Sustainability and Peace United, Nations University, 5-53-70, Jingumae, Shibuya-ku, Tokyo 150-0001, Japan; E-Mail: hearath@iu.edu
³Geoinformatics Center, Asian Institute of Technology, Bangkok, Thailand; E-Mail: lal@ait.ac.th

ISPRS Commission VIII WG 4 - Water

KEY WORDS: MODIS, Lower Mekong River Basin, Land cover, Maximum likelihood classification, Mekong Delta

ABSTRACT:

Mekong River runs from Hengduan Mountains in central-west China to Vietnam covering 805,604 sq km of land by its basin. The Lower Mekong Basin (LMB), the region mapped in this study, covers nearly 3/4 of the entire basin. About 90% of the population and agricultural activities of the Mekong River basin is located in this fertile LMB which faces disastrous floods almost annually. Mapping LMB at moderate resolution gives number of advantages for studies of flood mitigation and land utilization. However, compiling a cloud free mosaic and collecting ground truth data for training samples and map validation make map production a challenging task. This study utilized MODIS 250m image data of the region obtained in 2005 February. Dry weather in Jan-Apr makes the sky relatively free of clouds and 2005 February also had fewer disturbances coming from smoke of biomass burning. The methodology of the study substantially relied on high resolution images in Google Earth for collection of training sample for supervised classification and accuracy assessment. Arc- GIS generated KMZ file of unclassified and classified maps used to overlay image and map on Google Earth for identifying training site and field information extraction for accuracy assessment. Also ground information collected by a different research projects in 2008 were combined with information gathers from Google Earth images. The classified map showed 29.2% of the LMB under forest, 36.5% under Scrubland, when combined its Highland and Lowland subcategories. Three subcategories of paddy cultivated area covered 27.9% of LMB. Accuracy assessment conducted with randomly selected 200 points against high resolution images gave an overall accuracy of 80.7% in major land cover classes. According to the 250m resolution, urban features have not clearly separated though large urban areas like Phnom Penh and Can Tao have accurately classified. The methodology of this study produced a noteworthy success in classifying land cover of large areas like LMB, without expensive data sources and difficult and costly field investigations.

1. THE LOWER MEKONG BASIN (LMB)

The Mekong River basin is one of the few river basins of the world that has an immense social and economic impact on the regional and the global economy. Starting at 4,500m high Tibet Plateau, Mekong runs southward passing 4,800km, before it enters into South China Sea. The massive river system which covers parts of China, Myanmar, Lao PDR, Thailand, Cambodia and Viet Nam, is one of the 10 greatest rivers of the world, by the volume of water discharge to the sea. The total area under Mekong River, 795,000 square kilometers was divided into two major parts as upper and lower basin (see Figure 1). The upper Mekong consists of steep mountain slopes and only contributes about 15-20% of total volume of water (Mekong River Commission, 2005). The Lower Mekong basin covers nearly 75% of the land area of the total river basin.

About 90% of the population and agricultural activities of the total river basin are located in this lower part. The total contribution from agriculture (farming, fishing and forestry) provides living for 85% of people in the Lower Mekong Basin (LMB), making the river an integral part of domestic life. About 60 million people in the region annually produce enough rice to feed 300 million people (Mekong River Commission, 2005). In production of rice, Vietnam farmers alone produce over 3.5 million tones, annually, while Laos became self sufficient with rice in 1999. The LMB also enjoys a healthy consumption of fish and fish products annually. The per capita consumption of fish within the region in 2002 was 36 kg/person/year. This is a significantly larger volume compared to average consumption in EU countries, which was 26 kg/person/year in 2003 (Kent and Simon, 2003). With the growing population which is estimated to increase up to between 75 to 90 million people by 2025 (Mekong River Commission, 2005), LMB is facing many challenges to maintain its biodiversity while feeding the huge population.

The construction of new dams in upper Mekong Basin is affecting the biodiversity of LMB (WWF, 2009), while annual occurrences of floods are threatening the life and physical environment. Although, there is no significant change in annual flood levels in recent decades (Flood situation report, 2008), ever increasing human interaction with the river has caused extensive damage from flood in recent years.

*Corresponding Author; Kithsiri Perera, FOES, University of Southern Queensland E-Mail: perera@usq.edu.au; Tel.: +61-7-4631-2543
These complicate and important environmental and socio-economic conditions are encouraging to conduct extensive studies on the basin to maintain its sustainable use. However, the lack of detailed land cover products (at least at 250m resolution) covering entire LMB in public domain is a major inadequacy for researches based on the region. Even before the map produced in present study is published in any journal, it comes as the 1st image when search in “Google search - image” using the search phrase “Land cover map of Mekong River Basin”. Providing low cost mapping products to cover LMB is an essential requirement in this regard.

2. MAPPING LOWER MEKONG BASIN

The 56.3 million people live in LMB are belong to six countries (Table 1). Despite its high overall rural population density (71 people per km²) and widespread agricultural activities, some of the areas of LMB are very remote and difficult to access through surface transportation.

Table 1. Area under each country within LMB

<table>
<thead>
<tr>
<th>Country</th>
<th>China</th>
<th>Myanmar</th>
<th>Laos PDR</th>
<th>Thailand</th>
<th>Cambodia</th>
<th>Vietnam</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area of</td>
<td>165</td>
<td>24</td>
<td>202</td>
<td>184</td>
<td>155</td>
<td>65</td>
<td>795</td>
</tr>
<tr>
<td>(1000 km²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of LMB</td>
<td>21</td>
<td>3</td>
<td>25</td>
<td>23</td>
<td>20</td>
<td>8</td>
<td>100</td>
</tr>
</tbody>
</table>

Source: Mekong River Commission, 2005

Nearly all big cities like, Can Tho, Phnom Penh, Pakse, and Vientiane, in LMB are located along the river, where major transportation source is river itself. Apart from well accessible Thailand portion of LMB, regions within Laos and Cambodia, or about 45% of the LMB, have very poor road access (Hook at el., 2003). This remoteness and involvement of six countries within LMB made mapping the region a time and money consuming task and limits broad-level mapping projects to cover entire LMB. The Mekong River Commission (MRC) takes the leading role among these works.

Application of satellite data for land cover mapping tasks is becoming standard among researchers in recent years (Hook at el., 2003, Saito at el., 2007, Shigeo at el., 2006). Most of the mapping and related studies in Mekong Basin have focused on delta area where extensive paddy cultivation and flood activities are going on, or forest cover changes are taking place (Stibig at el., 2004). Among existing land cover maps at public domain, the forest and land cover map of 1997, published by MRC as a part of the LMB atlas can be highlighted (Hook at el., 2003, Hook at el., 2003). Also, NOAA 8km resolution land cover map of Mekong River Basin (Saito at el., 2007), 1km resolution the forest cover map of South East Asia (Shigeo at el., 2006), and International Geosphere Biosphere Programme (IGBP) 500m land cover products are in public domain, but have differences in resolution and volume of information of LMB, compared to the map we produced in this study. The land cover map in MRC atlas is based on Landsat TM satellite image hardcopies and contains 13 land cover classes at 1:250,000 scale. However, according to the MRC, the map has focused on vegetation types while land cover types have been mapped with a lesser degree of detail, e.g. there is only 1 class of permanent agriculture, which contains rice fields as well as agricultural plantations. This background encourages working on low cost and medium resolution approach to map the entire LMB without the use of administratively difficult and costly in situ data.

3. THE METHODOLOGY

Since MODIS (MODerate-resolution Imaging Spectrometer) data became freely available (Barnes at el., 2003), the scientific community has successfully made numerous land cover maps and/or clarified land cover status using these data (Friedl at el., 2002, Hall at el., 2002, Takeuchi and Yasuoka, 2006, Zhan et al., 2002). The MODIS data products such as MODIS NDVI (Ray, 1994, Tsuchiya, 2000, USGS, 2007) and true color image data (Gumley at el., 2003, NASA, 2006) became freely available through NASA for the global community. Also, free data availability and near-daily repetition of MODIS coverage (NASA, 2006) helps to update MODIS based works such as, flood monitoring, land cover changes, and forest fire monitoring efficiently. When dealing with land cover characteristics of large areas, MODIS 250m resolution data or widely known as moderate resolution data became a vital research tool (Perera and Armando, 2009), specially for large areas with less accessibility like LMB.

3.1 Research objects

The present study focused on producing a land cover map of entire LMB to fill the gaps in available land cover data sets while satisfying the following main objectives;

- Map with a medium level (or moderate) spatial resolution
- Map completion without in situ data
- Low production cost
- Ease of map updating

Here, the term “medium level” was decided by spatial resolution of existing land cover maps, existing free satellite data, and land area of the LMB. Even though the spatial resolution of land cover map produced in this study is 250m, it offers greater details which can be beneficial to MRC data sets. An important aspect of the new land cover map is completion of the map without using in situ observations. This was achieved through acquiring freely available field information in public domain and application of knowledge based ground truth selecting method to evade costly field investigations. A simplified research methodology is presented by the flow chart in figure 2.

3.2 Data acquisition

As common to most tropical regions with heavy rainfall, most parts of the LMB sky covered with a near-perennial cloud cover throughout the year. A systematically selected sample image set of lower section of the LMB, which apart each image by 30 to 33 days, is presented in figure 3 (from 002 to 294 Julian day of 2005 or January to October). All these images are showing heavy cloud cover over different locations with different volumes. Furthermore, it is evident that surface visibility of less clouded months (Jan – Mar) is frequently affected by haze caused from biomass burning. Red dots on images represent those fire spots. In dry early months (January – April) of the year, forest fires set by farmers to clear the land are common in remote areas.

Precipitation data over the basin were also examined to justify the driest season for most parts of the LMB. Figure 4 shows monthly rainfall distribution pattern of two rainfall stations in LMB, where both locations have lowest precipitation from Dec to Mar.

After observing cloud patterns and rainfall trends, February was selected to acquire MODIS images, with the consideration of low cloud cover, low disturbances form biomass fire smoke, and low water levels in the river basin.

Through an intensive search over daily MODIS data of February 2005, 15 relatively cloud-free images were selected from 2005031(Julian day) to 2005065 (WIST, 2009). Low-level MODIS data are corrected for atmospheric gases and aerosols,
yielding a level-2 basis for several higher-order gridded level-2 (L2G) and level-3 products. Although there may be other improved versions, these data freely provided by NASA are ready for use in scientific research works (USGS-LPDCCA, 2008). Availability of high-speed internet connection will be an advantage for speedy data download from MODIS web server (LP DAAC, 2008).

Figure 2. The research methodology explains here with regard to research objects, main data processing steps, and results.

3.3 Ground truth data and image processing

Initial image processing conducted using ER-Mapper 5.5, which produced a mosaic of the LMB at 250m resolution using true color FAS (USDS Foreign Agriculture Service) project and AERONET (Aerosol Robotic Network) project subset data. A base map produced by digitally interpolating 1 km DEM (Digital Elevation Model) data and 1 km land cover data into 250m resolution to conduct geometric corrections for image potions to replaced scattered cloud spots in MODIS mosaic. The 250m MODIS image mosaic of LMB falls within image rectangle of 5075 x 6870 pixels which is an easily manageable size in a standard PC.

The selection of ground truth data for classification and map validation plays a significant role to complete the map successfully. But to maintain map accuracy at satisfying level while minimizing the cost, the methodology was completely based on “free of in situ data”, and used number of information sources to collect ground truth information.

Figure 3. The perennial cloud cover and forest fire spots in Mekong Basin.

Figure 4. Monthly rainfall distribution pattern of 2 rainfall stations; Pakse (Laos) and Chau Doc (Vietnam), in LMB.
based on “free of in situ data”, and used number of information sources to collect ground truth information.

1. Available high resolution satellite images of the LMB. In this paper, the term “high resolution” is used to explain all finer resolution images available in Google Earth data set, ranging from 1.84m (DigitalGlobe) to 10m (SPOT) or 30m (Landsat TM). The coverage of these free images is highly irregular, but has a high practical use when access through KML (Keyhole Markup Language) file environment, which facilitate overlay other image/map layers on Google Earth. MODIS images are available in KML format, and classified image can be converted into KML format through ArcGIS to display in Google Earth. Figure 5 shows the process of extracting ground truth information from available super resolution images in Google Earth.

2. Previously collected photographic information of LMB area. A set of field photos with GPS information from Mekong delta region were also utilized to identify ground truth information. The figure 6 shows field photos of paddy cultivation, which is the dominant land cover type in the delta and mix vegetation within paddy.

3. Information derived from other existing land cover maps of LMB (LMB atlas, Asian 30 minute land cover data set).

4. Knowledge based visual identification to support collection of ground truth information for different land cover types. Visual comparison of high resolution images with 250m MODIS true colour images was supported by geographical knowledge of the region to establish knowledge-based criteria (Crowther and Hartnett, 2003) for identifying training sites. This approach can be itemized as follows:
   • Fundamental knowledge of the area.
   • Relational knowledge on spatial relationships among land surface components (e.g. forest is common in rural mountains).
   • Pattern and texture of features. This is basically used in visual interpretation of air photos and here, it was useful to identify training sites in paddy and forest.
   • Image classification knowledge, or understanding about digital classification.

The 2nd stage of image processing to produce the classification was conducted using ENVI 4.4 software package, after installing MODIS conversion toolkit into the PC to convert MODIS HDF file format. Also preprocessed KMZ (Zipped version of KML) MODIS files were collected to use in Google Earth. Table 2 presents comparison between MODIS true color bands and Landsat TM natural color image bands to show similarities in spectral characteristics and major applications to justify the use of MODIS true color images for selecting training samples. The same image band combination of true color mosaic (Table 2) was used for classification at 250m spatial resolution.

![Figure 5. Searching ground truth information using KML files of MODIS and classified images over Google Earth. This approach helps to locate the exact spot on Google Earth to visually verify the land cover type on high resolution images.](image)

![Figure 6. Field photos paddy associated and mix vegetation character and GPS locations used for ground truth information. These photos were collected from different sources and not through a specific field investigation for the map.](image)

<table>
<thead>
<tr>
<th>BAND</th>
<th>RANGE nm</th>
<th>KEY USE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODIS TRUE Color</td>
<td>1(R)</td>
<td>620–670</td>
</tr>
<tr>
<td></td>
<td>4(G)</td>
<td>545–565</td>
</tr>
<tr>
<td></td>
<td>3(B)</td>
<td>439–479</td>
</tr>
<tr>
<td>Landsat TM 5 TRUE Color</td>
<td>3(R)</td>
<td>630–690</td>
</tr>
<tr>
<td></td>
<td>2(G)</td>
<td>520–600</td>
</tr>
<tr>
<td></td>
<td>1(B)</td>
<td>450–520</td>
</tr>
</tbody>
</table>

Along with the true color mosaic, a combination of band 7-2-1 was used to produce the water bodies and the ocean layer. In band 7-2-1 combination, water is visible in dark blue and no sub colors are prominently appearing within water. When selecting training samples, forest cover in mountain was assumed as having fewer fluctuations in green biomass, due to tropical climatic conditions. Pixel values within each training site were used as the initial criteria to discriminate each land cover type, but knowledge about the region and existing map information were also used to build training sites as previously explained. Each training sample contained well over 1000 pixels to give enough representation of the category. Image classification was conducted by the supervised maximum likelihood classifier and final land cover map produced through combining 26 initial classes into 11 Land cover classes.

4. THE LAND COVER MAP

The establishment of a land cover classification scheme is a challenging task for any land cover mapping task. MODIS images are greatly helpful for vegetation monitoring (Perera and Tsuchiya, 2009) but its 250m spatial resolution naturally hides or segments some scattered smaller features like, urban characters and smaller open soil patches, and smaller water bodies, into one pixel (see figure 05). The present classification scheme has paid attention to these issues when selecting ground truth data. The Mekong River Commission’s vegetation and land cover map has 13 categories and gives greater information on vegetation types with lesser attention to paddy cultivation. In present study, a special attention was given to map differences in paddy lands where frequent flood damage occur. The final land cover map is presented in figure 7. A short description of each land cover category is included in Table 3. Though there are 11 categories in the classification legend, within the LMB boundary, the total area under clouds and sea is less than 0.01%. The forest covers 29.2% of the LMB, while Scrubland classified as the largest category (36.5%), if combined its Highland and Lowland subcategories.
Three subcategories of paddy cultivated area accounted for 27.9% of LMB.

### Table 3. The classification scheme.

<table>
<thead>
<tr>
<th>Class</th>
<th>Class Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Forest</td>
<td>Native forest in mountains and hilly terrain</td>
</tr>
<tr>
<td>2</td>
<td>Scrubland</td>
<td>- Mainly in SE Cambodia, perhaps due to long-term deforestation</td>
</tr>
<tr>
<td>3</td>
<td>Scrubland Lowland</td>
<td>- Flat and low lands close to rivers and paddy.</td>
</tr>
<tr>
<td>4</td>
<td>Paddy in low wet lands</td>
<td>- Vast paddy lands in floodplains</td>
</tr>
<tr>
<td>5</td>
<td>Paddy dominant in hinterland</td>
<td>- Paddy dominant low lands in hinterland and next to floodplains.</td>
</tr>
<tr>
<td>6</td>
<td>Paddy dominant mix vegetation</td>
<td>- Available in areas with high rural population. This fact was checked through higher number of human settlements visible in high resolution images.</td>
</tr>
<tr>
<td>7</td>
<td>Mix vegetation</td>
<td>- A mixed environment of perennial corps and woodlands (mostly villages), but also may be a mixture of forest and scrublands.</td>
</tr>
<tr>
<td>8</td>
<td>Non-Vegetated</td>
<td>- Urban, other buildup and open areas.</td>
</tr>
<tr>
<td>9</td>
<td>Water Bodies</td>
<td>- Water bodies including river pixels.</td>
</tr>
<tr>
<td>10</td>
<td>Sea</td>
<td>- Sea including few pixels mixed in Mekong Delta.</td>
</tr>
<tr>
<td>11</td>
<td>Cloud</td>
<td>- Clouds (only included in the area outside the LMB).</td>
</tr>
</tbody>
</table>

### Table 4. Confusion matrix of error assessment.

<table>
<thead>
<tr>
<th>Category (% in map)</th>
<th>Forest (29.2)</th>
<th>Scrubland (36.3)</th>
<th>Paddy (27.9)</th>
<th>Mix Veg. (4.7)</th>
<th>Total</th>
<th>User’s accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed from high resolution images</td>
<td>50</td>
<td>57</td>
<td>10</td>
<td>5</td>
<td>72</td>
<td>68</td>
</tr>
<tr>
<td>Forest</td>
<td>4</td>
<td>2</td>
<td>11</td>
<td>38</td>
<td>56</td>
<td>90.9</td>
</tr>
<tr>
<td>Scrubland</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Paddy</td>
<td>17</td>
<td>4</td>
<td>11</td>
<td>38</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Mix Veg.</td>
<td>5</td>
<td>1</td>
<td>38</td>
<td>53</td>
<td>71.7</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>72</td>
<td>68</td>
<td>15</td>
<td>45</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>Producer’s accuracy</td>
<td>83.8</td>
<td>53</td>
<td>71.3</td>
<td>71.7</td>
<td>Overall accuracy = 80.8%</td>
<td></td>
</tr>
</tbody>
</table>

The same approach explained in section 3.3 was use for map accuracy assessment, which did not use the in situ observations. The accuracy test was only administrated on major land cover types identified due to the high representation of those classes. Initially, 200 random sample points were generated in ARCGIS environment, which distributed as 56(forest), 80(scrubland), 11(paddy), and 53(mixed vegetation) among categories. These points were checked in Google Image, by converting the Arc MXD file into an Arc KML file. This KML file contained three data layers, i.e., land cover map, LMB boundary, and 200 sample pints, which registered the location of each point on the ground or Google Earth. The accuracy assessment data were tabulated into a confusion matrix which gave the overall accuracy of 80.8% (Table 4). User’s and Producer’s accuracy values have some discrepancies, which may have been caused by the use of highly difference resolutions in mapping process of accuracy assessment. Even though the high overall accuracy value can be considered as an indication of better map quality, increase of sample points through a stratified random sampling method and inclusion of every category in the error assessment is vital.

### 5. DISCUSSION

This research has produced a land cover data of Lower Mekong River Basin without cost for data, ground truth data collection, and field verification. Though the spatial resolution is medium (250m), the land cover map showed the detailed information of the basin. Apart from the map’s overall accuracy of 80.8%, a visual comparison between a high resolution image section (south-central LMB) clipped from Google Earth (A) and map (B) in figure 8 shows some significant overall agreement between the image (or the ground conditions) and classified map.

Another comparison between new map and MRC land cover map was presented in Figure 9 and it shows the level of details achieved with MODIS data. In MRC land cover map, a huge area in Mekong Delta region is falling under Agriculture crop while the new map found many sub classes within the same area. Also, Non Vegetated areas (in red) are significantly matching with actual location of urban features like Can Tho city of Vietnam and Phnom Penh city of Cambodia. Future research interests of this study will be focused on updating the land cover map with more ground observations to justify the validity of application of freely available high resolution images as single source to replace in-situ data. Also, it’s worth to produce land cover maps of selected sample sites with fine resolution image data (SPOT, ALOS) to build a relationship between medium and high resolution data, which will help to attest the use of MODIS data as an economical alternation for land cover mapping. Furthermore, it’s important to improve present classification scheme with the hierarchical approach based FAO land cover classification system (Gregorio, 2005), to conduct comparison studies with land cover maps of other parts of the world.

### 6. CONCLUSIONS

The study produced a land cover data set to cover entire Lower Mekong River Basin (LMB), using 250m MODIS images. The accuracy assessment was based on random point samples tested against high resolution images in Google Earth which yield an accuracy of 80.8% for four major land cover classes of the area, which share 98.3% of the LMB. The methodology of the study did not use any direct field investigations and tested the application possibility of very high to high resolution (0.8m to 20m) images available in Google Earth to extract ground truth information. The classification scheme with 11 classes designed to match with the limitations of 250m MODIS image resolution and available ground truth information. Three paddy cultivating areas were identified in the map according to the spectral differences associated with local geography, such as delta and central-north hinterland of the basin. This helped to delineate the vast paddy area around Mekong Delta from rest of the map. Non vegetated areas (open soil/urban/sand) are also classified though 250m pixel size may causes errors in this category. It’s notable that water bodies also may contain lower accuracy due to the spatial resolution of the image and the dryness of the season which original MODIS images were acquired.

Apart from land cover information, the new map is useful in flood related studies on the LMB. Also, it will be a supporting data source for studies in rich biodiversity of the basin. The map is already in use under the Environmental and sustainable Development Program of United Nations University, Tokyo, Japan. The second edition of the map is already planned and images processing will commence in due course with more ground observations to validate the map while using Google Earth images to represent as source of in-situ data.
Acknowledgements: Authors are thankful to United Nations University, Tokyo, Japan for funding this work in 2005/2006. A special gratitude is due to K. MacDougal, Southern Queensland University, Australia, for providing institutional facilities and Professor Ryutarou Tateishi, CEReS, Chiba University, Japan, for valuable comments on classification scheme and accuracy assessment.

References:


6. Gregorio A. D, Fao Land Cover Classification System, Classification concepts and user manual, software version 02. 2005


26. WIST, Warehouse Inventory Search Tool (WIST), 2009, https://wist.echo.nasa.gov/wist/apr/mwswelcome/
