Abstract. We are conducting the joint international research project with INPE (Sao Paulo) and INPA (Manaus). This report focuses on the use of MODIS data in forest biomass mapping in the Amazon. This report focuses on the topic, “ecosystem zoning based on forest stand biomass”. We applied our time-series filtering (KLM-FIT) to the MODIS 1-7 reflectance products (MOD09A1 ver.5) and created cloud-free NDVI and NDII dataset. The day-time and night-time temperature data were also processed. The SRTM, biome type map, rainfall data, soil type map, forest cover map in 2001 to 2008 were also used for environmental classification. The data types are numerical (integer and real) data, categorical data and nominal scale data are included. Then we adopted the classification method of symbolic sequence based on the count kernel by converting the continuous number into a symbol. For classifying the site environments, this method applies the Self-Organizing Map algorithm. And a prototype of site environmental map with 500m resolution was successfully created by this methodology.

Keywords: Remote Sensing, Image Processing, MODIS, Amazon forest.

1. Introduction
There are big concerns on Amazon forests because of their influences to global environment. Houghton et al. (2001) compared seven methodologies for estimating forest biomass over the Brazilian Amazon and found that current biomass estimation of the Amazon basin varies widely. Because the entire Amazon is too large to get detail information on the ground and remote sensing technologies have been introduced to monitor forest conditions. MODIS plays important role in forest monitoring in a large area (Anaya et al., 2009; Hansen et al., 2008, Saatchi et al., 2007, Tokunaga et al., 2011).

For example, Saatchi et al (2007) used MODIS to make forest biomass maps of tropical countries based on the correlation with ground field data. However statistical values of MODIS include various errors because of influences of cloud and system noises even though they introduced 32-day composite data.

We are conducting the joint international research project with INPE (Sao Paulo) and INPA (Manaus). This report focuses on the use of MODIS data in forest biomass mapping in the Amazon. More than one thousand forest inventory data with the plot size of 0.25ha (20m
x 125m) have been collected in the project named “Carbon Dynamics of Amazon Forest (2009-2012): CADAF”. Remote sensing technologies were introduced for up-scale these biomass information to the whole Amazon forest in the project. The remote sensing study in the project consists of three themes as follows;

1) Study on forest structure parameters: The objectives of this study is to identify scale parameters by using inventory data, airborne data and high-middle spatial resolution satellite data for understanding carbon stocks with the comparison of the site environment map. The results enable us to estimate carbon stocks and its accuracy at each category of the site environment map.

2) Ecosystem zoning based on forest stand environment: To find out parameters to upscale the field plot data set to regional-scale forest information to evaluate forest carbon dynamics in central Amazon.

3) Study on forest biomass distribution in Amazon: The objectives of this study is to identify scale parameters by using inventory data, airborne data and high-middle spatial resolution satellite data for understanding carbon stocks with the comparison of the site environment map. The results enable us to estimate carbon stocks and its accuracy at each category of the site environment map.

This report focuses on the second topic, “ecosystem zoning based on forest stand biomass”. Uncertainty of forest biomass estimation from remote sensing data depends on the reference data (plot data) and remote sensing data, as well as processing algorithm. Some papers calculate the forest stand biomass based on one allometric equation as a faction of DBH and/or tree height for the tropical area. They ignore the environmental differences and forest area shows the amount of biomass. However, direct measurements of DBH and tree height on the ground are not easy tasks for a large area. Then satellite observation or other remote sensing data are often used to estimate forest structural parameters, including biomass.

2. Methodology

2.1 Pre-processing of MODIS data

Our study area covers the Amazon River basin (10N-20S, 80W-40W). The 8-day composite data of MODIS are strongly influenced by weather and the 8-day composite MODIS original data are not suitable for statistical processing. Then we applied our time-series filtering (KLM-FIT) to the MODIS 1-7 reflectance products (MOD09A1 ver.5) and created cloud-free NDVI and NDII dataset. The day-time and night-time temperature data were also processed. The total amount of data is larger than 500BG.

The KLM processing has three steps

1) Local maximum filtering

\[ d'_i = \text{Min}[\text{Max}(d_{i-w+1}, d_{i-w+2}, \ldots, d_i), \text{Max}(d_i, d_{i+1}, \ldots, d_{i+w-1})] \]

where \( d_i \) is a data at \( t \), \( w \) is the duration for filtering at the time \( t \), \( d' \) is the filtered result

2) Fitting by cyclic function

\[ f_r = c_0 + c_1t + \sum_{i=1}^{N} \left( c_{2i} \sin \left( \frac{2\pi k_i t}{M} \right) + c_{2i+1} \cos \left( \frac{2\pi k_i t}{M} \right) \right) \]

Where \( c_n \) are the coefficients, \( N \): number of cyclic pairs, \( k_i \): cycle parameter (1, 2, 3, 4, 6, or 12 month), Length of a unit cycle (One year=36 when using 10-day composite data)
3) Selection of the best combination by the Akaike Information criteria

$$AIC = D\left[ \log(2\pi\sigma) + 1 \right] + 2(j + 1)$$

where $D$ is the number of data, $j$ is the number of functions

4) Fitting by Kalman filter (LMF-KM)

$$f_t = c_0(t) + \sum_{i=1}^{N} \left\{ c_{2i}(t) \cdot \sin\left(\frac{2\pi k_i t}{M}\right) + c_{2i+1}(t) \cdot \cos\left(\frac{2\pi k_i t}{M}\right) \right\}$$

The structure is the same with the equation 3. The coefficients of the $C$ is the function of $t$

![Figure 1. Example of a LMF-KL processed image (Left: temperature, middle: NDVI, right: color composite of NDVI and Thermal data), 36 cloud-free images/one year are available](image)

2.2 Dataset for environment classification

The SRTM digital terrain data with 90m mesh is used because the terrain condition influences the forest structure. The biome type map, rainfall data, soil type map, forest cover map in 2001 to 2008 were also considered useful for environmental classification. The table 1 shows the dataset collected for this study.

<table>
<thead>
<tr>
<th>No.</th>
<th>Dataset name</th>
<th>Data Type</th>
<th>Units</th>
<th>Sensor/ Original Data</th>
<th>Time interval</th>
<th>Total Number of Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Elevation (DEM)</td>
<td>numeric raster</td>
<td>m</td>
<td>SRTM</td>
<td></td>
<td>400</td>
</tr>
<tr>
<td>2</td>
<td>Plateau Ratio</td>
<td>numeric raster</td>
<td>%</td>
<td>SRTM</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>Slope Angle</td>
<td>numeric raster</td>
<td>degrees</td>
<td>SRTM</td>
<td></td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>Biome Type</td>
<td>categorical vectors</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Averaged annual rainfall</td>
<td>vectors</td>
<td>mm</td>
<td></td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>6</td>
<td>Soil Type</td>
<td>categorical vectors</td>
<td></td>
<td></td>
<td></td>
<td>63</td>
</tr>
<tr>
<td>7</td>
<td>NDVI &amp; NDII Seasonal Change Profile Clusters</td>
<td>nominal raster</td>
<td>MOD09A1</td>
<td>1year</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Daytime &amp; Nighttime Surface Temperature Seasonal Change Profile Clusters</td>
<td>nominal raster</td>
<td>MOD11A2</td>
<td>1year</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Water Coverage Period Map</td>
<td>nominal raster</td>
<td>MOD09A1</td>
<td>1year</td>
<td>46</td>
<td></td>
</tr>
</tbody>
</table>
2.3 Algorithm for environment classification
The data types of them are not same and numerical (integer and real) data, categorical data and nominal scale data are included. Therefore we have developed special software for classification which can classify the environmental characteristics of each pixel. Then we adopted the classification method of symbolic sequence based on the count kernel (Tsuda et al. 2002) by converting the continuous number into a symbol. Then, we developed a site environment map that reflects the diversity of environment conditions of Amazon River basin.

3. Results and Discussion
We developed technologies to collect site environment parameters (ex. surface temperature, period of inundation, and terrain conditions) and create a site environment map that shows the diversity of site environment conditions of Amazon River basin.

3.1 Pre-processed MODIS data
The filtering method “kalmfit” and a discrete time-series modeling method were applied to the 8-day composite MODIS dataset for removing the influences of clouds and other noises on the data. The datasets consist of NDVI, NDII, day and night brightness temperature, and spectral clusters in 2001 to 2009 with 8 day interval (Fig.3).
3.2 Other Dataset
The detail forest distribution maps which were created based on the interpretation of Landsat and CBERS by INPE were used as a base map. The terrain data derived from the SRTM were also introduced and a site environmental map with 500m resolution was created. Water environments, such as inundation period and distribution of flooded forests, are verified by PALSAR and other high-middle resolution satellite data.

3.3 Classification
A classification methodology of site environment conditions was developed. For classifying the site environments, this method applies the Self-Organizing Map algorithm to the terrain condition characterized by SRTM DEM, temperature condition derived from MODIS brightness temperature, water condition obtained from water coverage map, and other GIS data such as soil maps. A prototype of site environmental map with 500m resolution was created by this methodology (Figure 5).
3. Conclusion
We have rewritten source codes of the discrete time-series modeling method in order to execute the processing on the GPGPU (General Purpose Graphics Processing Units). It runs about 26 times faster than CPU processing. That made us possible to create the cloud-free MODIS data set of 10 years. Because the site environmental map has similar color at similar environmental groups, the map shows several environmental zones in the Amazon. Further study will be needed to identify the important environmental factors which reflect the zones.

Acknowledgement
We would like to thank many people who support the CADAF project which is financed by JST and JICA, Japan. Our special thank are for the two leaders of the projects, Dr Moriyoshi Ishizuka, FFPRI, and Dr. Niro Higuchi, INPA, who work hard for giving us important information about Amazon forest.

References


Tsuda, K.; Kin, T.; Asai, K. Marginalized kernels for biological sequences, Bioinformatics, V.1, Suppl.1, p.S268-S275, 2002