

## DEVELOPMENT OF VALIDATION DATA SETS FOR GLOBAL LAND COVER CLASSIFICATION USING ALOS/AVNIR-2 DATA

N.Soyama <sup>a</sup>, K.Muramatsu <sup>b</sup>, S.Furumi <sup>c</sup>, M. Daigo <sup>d</sup>

<sup>a</sup> Center for Research and Development of Liberal arts Education, Tenri University  
1050 Somanouchi, Tenri, Japan – soyama@sta.tenri-u.ac.jp

<sup>b</sup> Laboratory of Nature Information Science, Department of Information and Computer Sciences, Nara Women's  
University, Nara, Japan – muramatu@ics.nara-wu.ac.jp

<sup>c</sup> Preschool education, Nara Saho College, Nara, Japan – furumi@saho.ac.jp

<sup>d</sup> Faculty of Economics, Doshisha University, Kyoto, Japan – daigo@doshisha.ac.jp

### Commission VIII, WG VIII/8

**KEY WORDS:** Land cover classification, Validation data sets, Vegetation coverage degree, Land cover class definitions, ALOS AVNIR-2, UPDM

### ABSTRACT:

The GCOM Climate first generation satellite (GCOM-C1) project started in October 2009 and each research has proceeded to the aim of project. Our final aim in GCOM-C1 project is producing global land cover data sets that provide much useful information for the study of global environmental changes such as global biogeochemical cycles and climate change, and for the estimation of net primary production. In this study, we proposal a land cover class structure of global land cover, and apply to produce validation data sets for the land cover classification system with low resolution data using high resolution satellite data sets, ALOS/ AVNIR-2 data sets. To determine the classification conditions in our classification system, we used the universal pattern decomposition method (UPDM) coefficients and the modified vegetation index based on the UPDM (MVIUPD). As validation data sets, we produced vegetation coverage degree rank data sets and dominant class data sets. Comparisons between the ground truth data sets and the results of our land cover classification system with ALOS/AVNIR-2 showed many areas where classes agreed. Using the classification results, validation data sets were produced.

## 1. INTRODUCTION

Global Change Observation Mission (GCOM) was planned by JAXA as the comprehensive observation system of the Earth's environmental changes. The GCOM Climate first generation satellite (GCOM-C1) project started in October 2009, and each research has proceeded to the aim of project. The GCOM-C1 satellite will be equipped with the Second-generation Global Imager (SGLI) which will have special features of wide spectral coverage from 380nm to 12um, a high spatial resolution of 250m, a field of view exceeding 1000km, two-direction simultaneous observation, and polarization observation. GCOM-C1 is planned to be launched in fiscal 2013.

Our final aim in the GCOM-C1 project is producing global land cover data sets that provide much useful information for the study of global environmental changes such as global biogeochemical cycles and climate change, and for the estimation of net primary production. We have to produce validation data sets for global land cover classification by that time.

Though it is difficult to identify the land cover type of a pixel that contains a mixture with low resolution such as SGLI data sets, higher quality of the global land cover classification is needed. For extracting the land characteristic of a pixel of low resolution, it is effective to produce validation data sets with high resolution data set, and to get to know the land cover situation and vegetation conditions.

In this study, we produce validation data sets with high resolution data sets, ALOS/AVNIR-2 data sets, for the global land cover classification system which will be developed using SGLI data sets. ALOS satellite was launched in 2006. It has

AVNIR-2 sensor with 10-m spatial resolution and four spectral bands 460, 560, 650 and 830 nm. Because of the spatial resolution and multi-spectral information, AVNIR-2 sensor data is expected for local area study and environmental managements.

To determine the classification conditions in our classification system, we used the universal pattern decomposition method (UPDM) coefficients and the modified vegetation index based on the UPDM (MVIUPD), which is calculated using the UPDM. The UPDM is suitable for multi-dimensional reflectance values. UPDM coefficients and the MVIUPD values are useful for determining land cover conditions.

Validation data sets are consists of dominant class information and vegetation coverage degree rank information. First we produce the land cover products consisted of seven classes, which are needle leaf forest class, broadleaf forest class, grassland class, cropland class, urban class, barren class and water class. After verifying that the land cover classification results are agreed with ground data sets, we produce vegetation coverage degree rank data sets and dominant class data sets.

## 2. CLASSIFICATION SCHEME

### 2.1 Land cover class definitions for global land cover classification

In our global land cover classification system, land cover classes are designed with hierarchical structure as shown in the Table 1. Level-1 classes are guaranteed higher accuracy than Level-2 classes or Level-3 classes. To estimate Gross Primary Production (GPP) or Biomass, not only land cover class but

also vegetation coverage degree information is required. We propose to add the information, which is consisted of three ranks: high, middle and low, into our global land cover classification system.

Class Rank	Level 1	Level 2	Level 3 (IGBP Land Cover Units)	Ancillary information
Class Category	Forest	Evergreen forest	Evergreen Needleleaf Forests	
			Evergreen Broadleaf Forests	
		Deciduous forest	Deciduous Needleleaf Forests	
			Deciduous Broadleaf Forests	
			Mixed Forests	
	Grasslands	Shrublands	Closed Shrublands	Vegetation growth season, Vegetation density level (High , Medium, Low)
			Open Shrublands	
		Savannas	Woody Savannas	
			Savannas	
	Grasslands			
	Wetlands		Permanent Wetlands	
	Croplands		Croplands	Harvest season, Irrigated land or Dry land, Vegetation density level (High , Medium, Low)
			Cropland/Natural Vegetation Mosaics	
	Urban		Urban and Built-Up Lands	
	Other		Snow and Ice	
		Barren		
Water		Water Bodies		

Table1. Land cover classification definitions for global land cover classification.

In this study, we produce vegetation coverage degree rank data sets and dominant class data sets as validation data sets. These validation data sets are produced after classifying each pixel of ALOS/AVNIR-2 into seven classes; needle leaf forest class, broadleaf forest class, grassland class, cropland class, urban class, barren class and water class.

The vegetation coverage degree rank is determined by the ratio of vegetation classes to total pixels of AVNIR-2 corresponded to a pixel of SGLI-250m as follows; for high rank, the ratio is more than 70%, for low rank, the ratio is less than 30%; for middle rank, the remaining range. The dominant class is the most class in total pixels of AVNIR-2 corresponded to a pixel of SGLI-250m. When the percentage of the most class is less than 60%, the next largest class is also saved as validation data.

## 2.2 Data used in this study

### Satellite data sets

We used ALOS/AVNIR-2 data sets for producing validation data sets of high resolution. ALOS satellite was launched in 2006. It has AVNIR-2 sensor with 10-m spatial resolution and four spectral bands 460, 560, 650 and 830 nm. Because of the spatial resolution and multi-spectral information, AVNIR-2 sensor data is expected for local area study and environmental managements.

Clear data sets are selected for study area around Kii-Peninsula in Japan. ALOS/AVNIR-2 data observed on 9 October in 2006 was used. The spectral radiance  $L(\lambda)$  at the centre wavelength  $\lambda$  is converted from the DN recorded by the AVNIR-2. The calibration constants from each satellite sensor project are available to the public. The total in-band solar radiance ( $L_{\Delta\lambda}^{solar}$ ) outside the atmosphere is

$$(L_{\Delta\lambda}^{solar}) = \frac{E_{\Delta\lambda}^{mean} \cos\theta_0}{\pi d^2}. \quad (1)$$

Here  $E_{\Delta\lambda}^{mean}$  ( $Wm^{-2}$ ) is the solar spectral irradiance over the spectral bandwidth ( $\Delta\lambda$ ) of the band at the mean earth-sun distance, and the value from the each project is available to the public.  $\theta_0$  is sun's angle from the zenith, and the value is in the data header.  $d$  is the earth-sun distance on the observation day in astronomical units and was calculated manually.

The path radiance ( $L_{\Delta\lambda}^{path}$ ) depends on changes in solar illumination and atmospheric conditions. The main process leading to the path radiance is Rayleigh scattering. The cross section (Hill and Sturm 1991) of Rayleigh scattering varies as  $\lambda^{-4}$  and is calculated for each band using the solar zenith angle and the earth-sun distance for each image datum.

### Validation data sets

To validate land cover classification results using ALOS/AVNIR-2, we use ground truth data sets which are collected around KINKI area. For vegetation coverage degree results, we used the digital map 5000 (land use) are produced by geospatial information authority of Japan, hereafter abbreviated to DLU. This is used for survey of land use trend on 2001. For using these DLU data sets as validation data sets, the following problems may be pointed out. The DLS data sets are old to compare with ALOS/AVNIR-2 data sets may be pointed out. For that, these are trivial problems since we select relatively less secular changed area in this study. As another problem, some class of DLU disagree with land cover class of our system, since class definitions of DLU are decided for land use. In this study, we counted agricultural land, park area and a forest area as vegetation coverage type, however some of them may different from the real land states; for example, all of park areas are not covered with vegetation subjects.

## 2.3 AVNIR-2 data analysis method

### 2.3.1 UPDM and MVIUPD

The UPDM decomposes a set of  $n$  reflectance channel values for each pixel using standard spectral patterns for water, vegetation, and soil, and the supplementary pattern coefficients. Thus, the multi-dimensional spectral reflectance data are transformed into three coefficients that almost directly correspond to actual objects on the ground, i.e., the water coefficient, vegetation coefficient, soil coefficient (Zhang, 2006). Since we define each standard spectral pattern as a continuous spectral function from 350–2500 nm, the UPDM coefficients are independent of the sensor. The MVIUPD function is defined by the four UPDM coefficients as follows:

$$MVIUPD = \frac{C_v - 0.2 \times C_s - C_w}{C_w + C_v + C_s}. \quad (2)$$

where  $C_w + C_v + C_s$  represents the sum of the total reflectance of all bands, and  $C_w$  renders MVIUPD negative for surfaces such as snow-covered areas that lack vegetation. The 0.2 coefficient for represents a small correction for the relationship between vegetation photosynthesis and the cover ratio (Xiong, 2005).

### 2.3.2 UPDM and MVIUPD characteristics

Figure 1 (a) shows typical reflectance patterns for the spectral reflectance of ALOS/AVNIR-2 data sets for ground, reservoir, and forest samples. These samples were taken from Nara prefecture in Japan. Figure 1 (b) shows the UPDM coefficients ( $C_w, C_v, C_s$ ) for the same samples shown in Fig. 1 (a). For the reservoir samples in Fig. 1 (b),  $C_w$  is the highest of the three coefficients, and every coefficient is much lower than that of the other samples. For the forest,  $C_v$  is the highest of the three coefficients. The  $C_s$  values are highest for ground. Thus, we can determine the characteristics of each area by comparing three coefficient values.

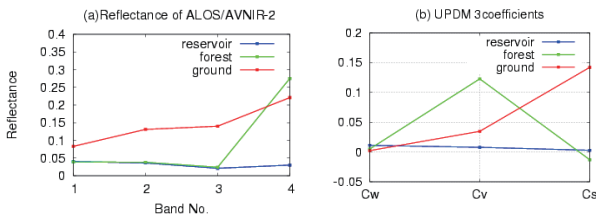


Figure 1. (a) Spectral reflectance of ALOS/AVNIR-2 data sets. (b) UPDM coefficients calculated from AVNIR-2 spectral reflectance data for reservoir, forest and ground samples.

### 2.4 Classes and criteria for determination

The condition values used in this land cover classification system were determined as follows. The UPDM coefficients and the MVIUPD of sample data sets were calculated for each class. Sample data sets were taken from Nara Prefecture. For build-up, grassland and cropland samples were taken from Nara city. Broad leaf forest samples were taken from Ikoma city and Nara city. Water and ground samples were taken from Tenri city. Needle leaf forest samples were taken from Yoshino country.

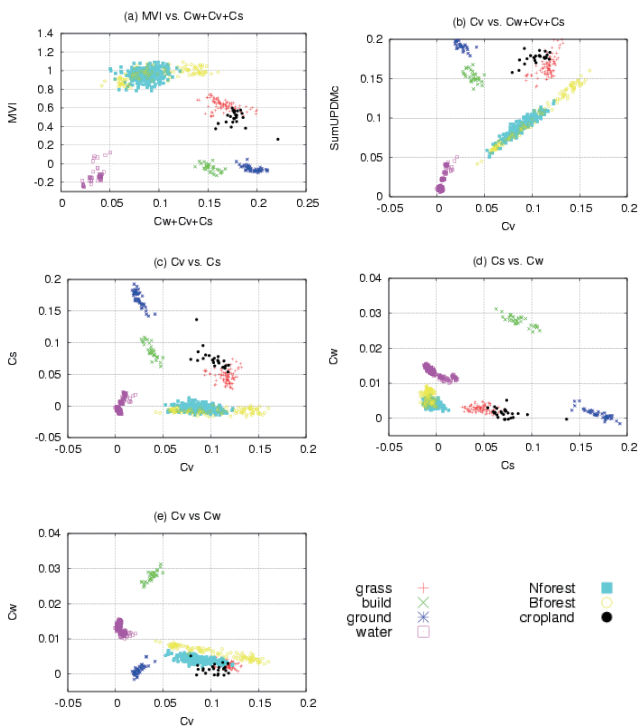


Figure 2. (a) The relationship between the MVIUPD and  $C_w+C_v+C_s$  of sample data for build-up land, ground, needle leaf forest, broad leaf forest, grassland, cropland and water, (b) The relationship between the  $C_v$  and  $C_w+C_v+C_s$ , (c) The relationship between the  $C_s$  and the  $C_v$ , (d) The relationship between the  $C_s$  and the  $C_w$ , (e) The relationship between the  $C_w$  and  $C_v$ .

Figure 2 (a) shows the relationship between the MVIUPD and  $C_w+C_v+C_s$  of sample data for build-up land, ground, needle leaf forest, broad leaf forest, grassland, cropland and water. We can see five clusters in Figure 2.(a), the left upper cluster is consisted of needle leaf forest and broad leaf forest, the second upper cluster is consisted of grass and cropland, the left lower cluster is water, the centre lower cluster is build-up land, the right lower cluster is ground. We can roughly distinguish between cropland and grassland in Figure 2.(d). In Figure 2.(e), we can distinguish between needle leaf forest and broad leaf forest. The threshold values of classification conditions were determined using these features.

Classification processing was conducted according to the flowchart in Figure 3. The first step of the classification process was to distinguish the water class using sum of three UPDM coefficients. In second step, urban and barren classes are extracted using MVIUPD,  $C_w$  and  $C_v$ , and then two forest classes are detected, in the final part, grassland and cropland classes are classified.

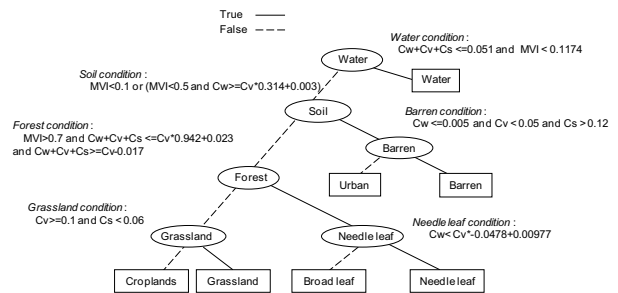


Figure 3. Land cover classification procedure flowchart.

## 3. RESULTS AND DISCUSSION

### 3.1 Classification results

Figure 4 shows the land cover classification map of KINKI area produced by the land cover classification system in this study. The land cover classification map of Figure 4 is roughly similar the actual vegetation map which is the 5th national survey on the natural environment (vegetation) 1994-1997, produced by Ministry of the Environment.

Table 2 is the confusion matrix derived from the ground truth data sets and the classification results in this study. We used ground truth data sets of sites chosen from Kinki area but Nara prefecture as the ground truth data sets.

The overall accuracy of land cover classification was about 65%, or 0.58 in terms of the kappa coefficient of agreement. The correct ratios for cropland and grassland classes are not high. Possible reason for this disagreement may be only one season's data sets used for our classification. If we use seasonal data sets, the precision of the classification is expected to be improved.

With the results of this land cover classification, we produced dominant class data sets of which each pixel is corresponded to a pixel of SGLI-250m.

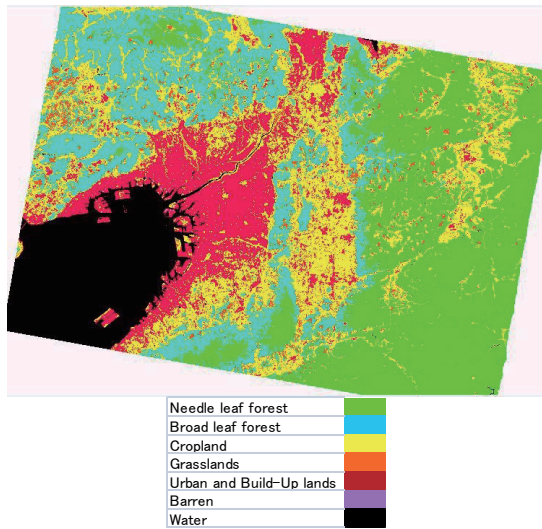


Figure 4. Land cover classification results in Kinki area in Japan.

		Ground truth data							
		Needle leaf forest	Broad leaf forest	Cropland	Grassland	Urban	Barren	Water	Row total
Land cover classification results	Needle leaf forest	21	16			5			44
	Broad leaf forest	7	32		27	10			76
	Cropland			24	16	5			49
	Grassland			8		23			31
	Urban						18		18
	Barren							10	10
	Water							17	17
Column Total		28	48	32	31	28	18	17	222
Correct Ratio		75.0%	66.7%	47.4%	51.6%	82.1%	100.0%	58.8%	

Table 2. Confusion matrix showing agreement between the ground truth data (columns) and land cover data determined by this study land cover classification system.

### 3.2 Vegetation coverage degree rank data sets.

The vegetation coverage degree rank is determined by the ratio of vegetation classes; needle leaf forest, broad leaf forest, cropland and grassland, to total pixels of AVNIR-2 corresponded to a pixel of SGLI-250m. Figure 5(a) shows the vegetation covered map produced using the land cover classification system and Figure 5(b) is the map drawn with DLU data sets for 3km × 4km area in Nara city and Yamatokoriyama city. For the DLU, forest, cropland, public good facilities and park are painted in green as vegetation classes, residential area, road, commercial area, vacancy area and industry area are painted in red as soil, river, lake, pond are painted in black as water.

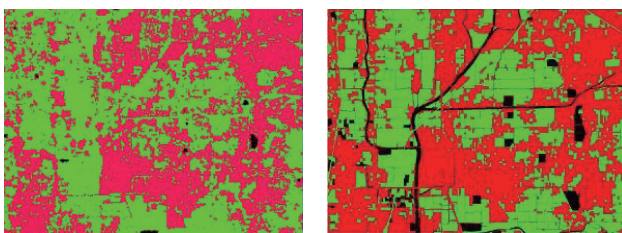


Figure 5. (a) Left figure is the map made with land cover products in this study, (b) Right figure is the map drawn with DLU data sets.

Figure 5(a) is roughly similar in position to Figure 5(b) except water. Since the grass grows thick in bank of river and pond in this area, these areas were classified as vegetation classes. In this study, the vegetation coverage degree rank data sets were produced by classifying into three ranks after calculating the ratio of vegetation classes to total pixels of AVNIR-2 corresponded to a pixel of SGLI-250m.

## 4. SUMMARY

This paper describes the algorithm used to produce validation data sets for global land cover classification system with GCOM-C/SGLI data sets. We created qualitatively realistic products on a local scale by applying our land cover classification to ALOS/AVNIR-2 data sets, and then we produced validation data sets for the global land cover classification system with low resolution data sets. For the study of estimating GPP, vegetation covered density information is very useful. In this study, we produced vegetation coverage degree rank data sets and dominant class data sets for part of KINKI area using the results of land cover classification with high resolution data sets. We will produce validation data sets of another country in the future.

## References :

Hill, J., Sturm, B., Radiometric correction of multitemporal Thematic Mapper data for use in agricultural land-cover classification and vegetation monitoring, *International Journal of Remote Sensing*, Vol.7, 1471-1491.

Muramatsu, K., Tahara, A., Soyama, N., Daigo, M., Fujiwara, N., Vegetation species classification using ALOS/AVNIR-2 data, *Proceeding ACRS 2009*.

Xiong, Yan, A study on algorithm for estimation of global terrestrial net primary production using satellite sensor data, Doctor thesis, Nara Women's Univ., 2005.

Zhang, L.F., Furumi, S., Muramatsu, K., Fujiwara, N., Daigo, M., Zhang, L.P., Sensor-independent analysis method for hyper-multispectral data based on the pattern decomposition method, *Int. J. of Remote Sensing*, Vol. 27. Nos. 21-22, Nov. 2006, 4899-4910.

Zhang, L.F., Furumi, S., Muramatsu, K., Fujiwara, N., Daigo, M., Zhang, L.P., A new vegetation index based on the universal pattern decomposition method, *Int. J. of Remote Sensing*, Vol. 28. Nos. 1-2, Jan. 2007, 107-124.

## Acknowledgements :

This study was supported by the National Space Development Agency (JAXA) of Japan as part of GCOM-C1 and ALOS projects.