Sub-class Clustering of Land Cover over Asia considering 9-year NDVI and Climate Data

Ga-Lam Lee*, Kyung-Soo Han**, †, and Do-Yong Kim***

*Geoinformatic Engineering Research Institute, Pukyong National University, Busan, Korea
**Department of Spatial Information Engineering, Pukyong National University, Busan, Korea
***BK21 Graduate School of Earth Environmental System, Pukyong National University, Busan, Korea

Abstract: In this paper an attempt has been made to classify Asia land cover considering climatic and vegetative characteristics. The sub-class clustering based on the 13 MODIS land cover classes (except water) over Asia was performed with the climate map and the NDVI derived from SPOT 5 VGT D10 data. The unsupervised classification for the sub-class clustering was performed in each land cover class, and total 74 clusters were determined over the study area. Via these clusters, the annual variations (from 1999 to 2007) of precipitation rate and temperature were analyzed as an example by a simple linear regression model. The various annual variations (negative or positive pattern) were represented for each cluster because of the various climate zones and NDVI annual cycles. Therefore, the detailed land cover map as the classification result by the sub-class clustering in this study can be useful information in modelling works for requiring the detailed climatic and vegetative information as a boundary condition.

Key Words: land cover, sub-class clustering, climate map, NDVI, SPOT VGT

1. Introduction

Land cover plays an important in many Earth system process due to the interaction between the Earth’s surface and the atmosphere. Land cover information is a critical parameter in hydrological models (Sandholt et al., 1999; Droogers and Kite, 2002; Yang and Musiake, 2003), biophysical models (Kimball et al., 1999; Liu et al., 2002;Lotsch et al., 2003; Inoue and Olioso, 2004), and climatological models (Champeaux et al., 2000; Chase et al., 2000). Recent works in classifying regional, continental and global land cover have seen the application of multi-temporal remotely sensed data sets, which describe vegetation dynamics by viewing their phenological variation throughout the course of a year (Verhoef et al., 1996). Tucker et al. (1985), Townshend et al. (1987), and Stone et al. (1994) have produced continental-scale classifications of land cover using this approach. For global land cover products, DeFries and Townshend (1994b) derived a one-by-one degree map and more recently an 8 km map.

Received May 27, 2011; Revised June 14, 2011; Accepted June 15, 2011.
† Corresponding Author: Kyung-Soo Han (kyung-soo.han@pknu.ac.kr)
(DeFries et al., 1998) using the Advanced Very High Resolution Radiometer (AVHRR) data. The current global land cover products are much finer in resolution than traditional climate modelers require, although there are some who have begun to take advantage of the additional information in the depiction of landscape heterogeneity provided by finer resolutions (Dickinson, 1995; Hansen et al., 2000).

Use of remotely sensed data from satellite makes susceptible to various interpretations and extraction of thematic information, such as land cover and use, about Earth’s surface at multiple spatial scales (Park et al., 2006). Maps produced using satellite data have advantages over traditional ground based maps due to the continuous coverage and internal consistency of remotely sensed data sets (Hansen et al., 2000). A primary reason for attempting to create maps from these data sets is the potential for creating more accurate products, where the areas of disagreement between products are less than past efforts compiled from ground-based maps (DeFries and Townshend, 1994a; Hansen et al., 2000). Satellite image estimates of vegetation cover provide important and timely information for modeling biochemical cycle and climate, carbon accounting, and monitoring ecosystem condition (Hansen et al., 2002; Jang, 2006; Kim et al., 2007; Lee et al., 2009). Most land cover mapping applications at broad spatial scales have been based on multi-temporal Normalized Difference Vegetation Index (NDVI) data (Tucker et al., 1985; Loveland et al., 1991). Continental and global land cover classification from satellite data have largely been derived from annual time series of the NDVI as a measure of phenology variety throughout the year (Tucker et al., 1985; Townshend et al., 1987; Loveland et al., 1991; Stone et al., 1994; DeFries and Townshend, 1995; DeFries et al., 1998). During the past 15 years, substantial progress has been made in using SPOT VEGETATION (VGT) data for land cover characterization. For example, unsupervised classification (Bartalev et al., 2003; Han et al., 2004; Huang and Siegert, 2006; Vancutsem et al., 2009), supervised classification (Pasqualini et al., 2005), a mixed approach (Latifovic and Olthof, 2004; Stibig et al., 2007), tree approach (Stroppiana et al., 2003; DeFries et al., 1998), logical classification systems (Stibig et al., 2004).

Changes in land cover due to natural and anthropogenic processes can affect the biophysical characteristics of the surface with implications for biogeochemical cycle and the energy budget (Clark et al., 2001; Bounoua et al., 2002; Fuller and Ottke, 2002; Cabral et al., 2003, 2006). Also, there may be various vegetative and climatic conditions in a land cover class. Therefore, it is necessary to consider vegetation and climate changes for the classifying land cover. Champeaux et al. (2004) described the methods used to derive surface parameters at 1km resolution using land cover maps, climate maps, and NDVI datasets. They introduced a new concept of reclassification on a land cover type to identify heterogeneity intra land cover class over Africa and Europe. The study attempted this concept (hereafter, so-called sub-class clustering) over Asia continent. In this study, Sub-class clustering of Asia land cover was performed with SPOT VGT data and climate map, based on the MODerate resolution Imaging Spectroradiometer (MODIS) global land cover map. Via the detailed land cover map as the classification result by sub-class clustering, the annual variations of precipitation rate and temperature were analyzed.

2. Study area and data

The study area encompasses 24 countries in Asia, ranging from 68°E to 180°E and from 55°N to -12°N (Fig. 1). Study area contains a variety of climate
zone, including tropical and subtropical in south area, littoral in southern islands, polar and sub polar in northern continental region. Northern Asia has much the same sort of climate as central Asia, except that is has more rainfall. Winters are extremely cold-the coldest inhabited place in the world is a village in Siberia. Southern Asia is hot all year around and there is a rainy season and a dry season. Eastern Asian is influenced monsoon.

Many kinds of global land cover are currently available, UMD(University of Maryland) 1 km global land cover maps, International Geosphere-Biosphere Programme, Data and Information Systems(IGBP-DIS) land cover, Global Land Cover(GLC) 2000, MODIS global land cover. These global land cover data sets were created for the same fundamental purpose of providing improved global land cover information for environmental modelers. The MODIS land cover map produced by Boston University at approximately 1km spatial resolution (Friedl et al., 2002) was used in this study, and shown in Fig. 2 for the study area. The study area contains 14 land cover classes (water, evergreen needleleaf forest, evergreen broadleaf forest, deciduous needleleaf forest, deciduous broadleaf forest, mixed forests, closed shrublands, open shrublands, woody savannas, savannas, grasslands, croplands, urban and built-up, barren or sparsely vegetated).

The climate map developed by Koepp and De Long (1958) was used in this study to decide the number of sub-class cluster in each MODIS land cover class. The study area contains 16 climate zones (tropical desert, wet and dry tropical, trade wind littoral, semiarid continental, cool marine, humid continental, cold with dry winter, wet equatorial, semiarid tropical, warm littoral, humid subtropical, international desert, polar, extreme subpolar, cold littoral, warm with dry winter).

NDVI derived from SPOT 5 VGT D10 data (Duchemin and Maisongrande, 2002; Hagolle et al., 2004) during 2008 was used in this study. The Earth observation satellite SPOT 5 with VGT 2 instrument onboard was launched on MAY 4th, 2002. The VGT 2 instrument has two types product. One is P-product and the other is S-product. The instrument provides spectral reflectance measurements, which are delivered to users in the form of four standard products as follows:
(a) P product: physical values of spectral reflectance on the top-of-atmosphere;
(b) S1 product: daily maximum of NDVI composite of spectral reflectance at the top-of-canopy;
(c) S10 product: ten-day maximum of NDVI composite of spectral reflectance at the top-of-canopy;
(d) D10 product: ten-day composite of directionally normalized spectral reflectance at the top-of-canopy.

The NCEP/NCAR reanalysis (Kalnay et al., 1996; Kistler et al., 2001) provides long-term model analysis ideal for the study and initialization of global- and regional-climate-scale models. Global reanalysis estimates were generated using a single, ‘frozen’ version of an operational numerical model, which was then run for the period from 1948 to present. All reanalysis values used in this study were analyzed on a 192 × 94 point Gaussian grid (approximate 1.905°N × 1.875°E resolution). The
monthly temperature and daily precipitation rate of NCEP/NCAR reanalysis data during 1999–2007 is used in this study.

The sub-class clustering of Asia land cover was performed by considering climate and vegetation. For determining the initial cluster, the number of climate zone in each MODIS land cover class was detected over the study area, using the climate map. The 64 initial clusters in 13 MODIS land cover classes (except water) over the study area were determined in this study. And then, the sub-class clustering by using the Iterative Self-organizing Data Analysis (ISODATA) unsupervised method (Tou and Gonzalez, 1974) was performed with the VGT D10 NDVI data in each MODIS land cover class. The ISODATA procedure is one of the widely unsupervised clustering algorithms, and commonly used for satellite image classification (Irvin et al., 1997). Spectral reflectances from multiple wavebands (equivalent to attributes) are used to determine cluster in multidimensional attribute space. ISODATA technique method of unsupervised classification uses a maximum-likelihood decision rule to calculate class means that are evenly distributed in the data space and then iteratively clusters the remaining pixels, using minimum distance techniques (Melesse and Jordan, 2002). An unsupervised classification procedure was used for image classification (ERDARS version 8.6), as it allows for the identification of all the important spectral groupings without initially knowing which are the thematically significant (Cihlar et al., 1998; Boles et al., 2004).

3. Results and discussion

The final number of clusters by the sub-class clustering is 74, and shown in Fig. 3. The forest
Fig. 3. The results of sub-class clustering for each MODIS land cover class, (a) evergreen needleleaf forest, (b) evergreen broadleaf forest, (c) deciduous needleleaf forest, (d) deciduous broadleaf forest, (e) mixed forests, (f) closed shrublands, (g) open shrublands, (h) woody savannas, (i) savannas, (j) grasslands, (k) croplands, (l) urban and built-up, (m) barren or sparsely vegetated.
classes for evergreen needleleaf, evergreen broadleaf, deciduous needleleaf, and deciduous broadleaf were classified into 7, 2, 3, and 6 clusters, respectively (Figs. 3a, 3b, 3c, and 3d). The mixed forests class shown in Fig. 3e has 6 clusters. For the closed shrublands, open shrublands, and savannas classes, the number of clusters is 7, 7, 6, and 7, respectively (Figs. 3f, 3g, 3h, and 3i). The grasslands and croplands classes were classified into 6 and 11 clusters, respectively (Figs. 3j and 3k). The urban and built-up class shown in Fig. 3l has 2 clusters. For the barren or sparsely vegetated class, 4 clusters were classified (Fig. 3m). These results show heterogeneity within each land cover class. Because the climate zone and NDVI annual cycle is various for each cluster although they are in the same land cover class. Fig. 4(a) and (b) shows NDVI cycle of each cluster which is included in open shrublands class and grasslands, respectively. Every cluster has different NDVI cycle and, especially, cluster-g6 and cluster-g7 shows growth cycle different with other clusters which show little change of NDVI value. It presents heterogeneity of NDVI within the same land cover class. Table 1 shows that the percentage of
Table 1. The percentage of climatic type which is the largest portion in cluster-b1 and cluster-b2

<table>
<thead>
<tr>
<th>climatic type</th>
<th>Cluster-b1</th>
<th>cluster-b2</th>
</tr>
</thead>
<tbody>
<tr>
<td>percentage (%)</td>
<td>Wet and dry tropical</td>
<td>Warm littoral</td>
</tr>
<tr>
<td>57</td>
<td>72</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Slope range of each level for annual mean precipitation rate and temperature during 1999-2007 in the study area.

<table>
<thead>
<tr>
<th>Level</th>
<th>Slope range for precipitation rate (mm/h)</th>
<th>Slope range for temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level-2</td>
<td>below -0.006</td>
<td>below -0.188</td>
</tr>
<tr>
<td>Level-1</td>
<td>-0.006 ~ -0.0013</td>
<td>-0.188 ~ -0.056</td>
</tr>
<tr>
<td>Level 0</td>
<td>-0.0013 ~ 0.0013</td>
<td>-0.056 ~ 0.056</td>
</tr>
<tr>
<td>Level+1</td>
<td>0.0013 ~ 0.006</td>
<td>0.056 ~ 0.188</td>
</tr>
<tr>
<td>Level+2</td>
<td>over 0.006</td>
<td>over 0.188</td>
</tr>
</tbody>
</table>

The annual variations (from 1999 to 2007) of precipitation rate and temperature as the meteorological and climatological factors for the clusters classified by the sub-class clustering in this study were analyzed by a simple linear regression model. The slope as the regression coefficient of each pixel was calculated by annual mean values, and divided into 5 levels such as Level-2, Level-1, Level 0, Level+1, and Level+2. The levels were decided by histogram using slope of the simple linear regression expression. The slope range of each level for the precipitation rate and temperature is shown in Table 2. The minus and plus signs in each slope level represent negative and positive annual variations, respectively. For example, the Level+2 for the precipitation rate means the increasing (positive) pattern over 0.006 mm/h every year during the study period. For the precipitation rate, the spatial rates of each slope level for (a) cluster-b1 and (b) cluster-b2 in evergreen broadleaf forest class shown in Fig. 3b were analyzed as an example (Fig. 5). The spatial rates of Level-2 and Level-1 in cluster-b1 are 49.79% and 14.46%, respectively (Fig. 5a). It means that the negative pattern area for the annual variation of precipitation rate is wider than the positive pattern area in cluster-b1. In cluster-b2, the similar spatial rate was represented both the negative and positive variations.

![Graph (a)](image1)

![Graph (b)](image2)

Fig. 4. The variation of NDVI in each cluster in 2008; (a) open shrublands, (b) Grasslands

climatic type which is the largest portion in cluster-b1 and cluster-b2. As a result of unsupervised classification, the clusters presented different remarkable climatic type. It means that the classification was influenced on climatic type.
levels (Fig. 5b). The spatial rates of Level-2 and Level+2 in cluster-b2, are 36.41% and 34.67%, and represented over the large part of Sumatra Island and the southwest of Borneo Island in Indonesia, respectively. For the temperature, the spatial rates of each slope level for (a) cluster-g2 in open shrublands class and (b) cluster-k5 in croplands class shown in Figs. 3g and 3k were analyzed as an example (Fig. 6). The large part of cluster-g2 is located around Taklamakan desert, and the positive pattern area is occupied more than 85% (Level+2, 64.93% and Level+1, 22.79%) (Fig. 6a). It means that the annual variation of temperature in cluster-g2 area increases during the study period, and consistent with the Intergovernmental Panel on Climate Change (IPCC) 4th report (IPCC, 2007). Many clusters represent an increase in the annual variation of temperature, but cluster-k5 in croplands class represents the negative pattern. The spatial rates of Level-2 and Level-1 in cluster-k5 are 44.85% and 37.93%, respectively (Fig.
6b). As a whole, the various annual variations of the precipitation rate and temperature for each cluster were represented in the same land cover class. Moreover, the various spatial rates of negative and positive pattern were represented in a cluster.

4. Summary and conclusion

In this study, the sub-class clustering based on MODIS land cover was performed to develop the detailed land cover map over Asia considering climate zone and vegetation. The ISODATA unsupervised method was used for the sub-class clustering. The 64 initial clusters in 13 MODIS land cover classes (except water) over the study area were determined by the climate map. And then, the sub-class clustering was performed with the NDVI derived from SPOT 5 VGT D10 data in each MODIS land cover class, and 74 clusters were determined as follows:
(a) Evergreen needleleaf forest class : 7 clusters
(b) Evergreen broadleaf forest class : 2 clusters
(c) Deciduous needleleaf forest class : 3 clusters
(d) Deciduous broadleaf forest class : 6 clusters
(e) Mixed forests class : 6 clusters
(f) Closed shrublands : 7 clusters
(g) Open shrublands : 7 clusters
(h) Woody savannas : 6 clusters
(i) Savannas : 7 clusters
(j) Grasslands : 6 clusters
(k) Croplands : 11 clusters
(l) Urban and built-up : 2 clusters
(m) Barren or sparsely vegetated : 4 clusters

The annual variations of precipitation rate and temperature were analyzed by the simple linear regression model, using NCEP/NCAR reanalysis data during 1999~2007, via the detailed land cover map as the classification result by the sub-class clustering in this study. The results were discussed with the regression slopes as the annual variation patterns for some cluster as an example. Some cluster represented the positive pattern (increasing trend) of the annual variation for each factor. There were also the negative pattern (decreasing trend) clusters. Moreover, the spatial rates of negative and positive pattern were various in a cluster. Therefore, the various variation patterns of the meteorological and climatological factors in high-resolution studies can be investigated by using the land cover classification by the sub-class clustering considering climate and vegetation. And, the results of the sub-class clustering can be useful information in modelling works for requiring the detailed climatic and vegetative information as a boundary condition. Moreover, the result of sub-class clustering does not need to analyze biophysically but it may also need to be analyzed.

Acknowledgements

This subject is supported by Ministry of Environment as “The Eco-technopria 21 project”(121-091-069).

References


Tou, J. T., and R. C. Gonzalez, 1974. Isodata algorithm, Pattern Classification by Distance Functions, Pattern recognition principles,
Addison-Wesley, Reading, Massachusetts.