EVALUATION OF TIME-SERIES OF MODIS DATA FOR TRANSITIONAL LAND MAPPING IN SUPPORT OF BIOENERGY POLICY DEVELOPMENT

F. Zhou, A. Zhang, H. Wang, G. Hong

Canada Centre for Remote Sensing, Booth Street, Ottawa, Ontario, K1A 0Y7 Canada - (Fuqun.Zhou, Aining.Zhang, Huili.Wang, Gang.Hong)@nrcan.gc.ca

KEYWORDS: Land cover, Vegetation, Change Detection, Data mining, Research, Decision Support

ABSTRACT:

Demanding for information on spatial distribution of biomass as feedstock supply and on land resources that could potentially be used for renewable bioenergy production is rising as a result of increasing government investment for bioenergy and bioeconomy development, and as a way of adaptation to climate warning. Lands transitioned over the past between the types of forest, grassland, forage land, and cropland are considered as the most promising for the production of dedicated bioenergy crops as a primary source of biomass feedstock for the development of the second generation biofuels, without compromising regular agriculture production.
Aimed at the transitional land mapping at a region scale, Earth Observation data with medium spatial resolution are considered as one of the most effective data sources. Time series of 10 days cloud-free composite MODIS images and its derivation, NDVI and vegetation phenology in the vegetation-growing season, are then used to derive the required information. With these datasets, three groups of data combinations are explored for the identification of the best combinations for land cover identification, then for transitional land mapping, using a data mining tool.
Results showed that longer time series of Earth Observation data could lead to more accurate land cover identification than that of shorter time series of data; Bands (1-7) only and NDVI or phenology with other bands (3-7) could yield almost the same highest accurate information. Results also showed that land cover identification accuracy depends on the degree of homogeneity of the landscape of the region under the study.

1. INTRODUCTION

Demanding for information on spatial distribution of biomass as feedstock supply is rising as a result of increasing government investment for bioenergy and bioeconomy development, and as a way of adaptation to climate warning. The recent global food crisis has raised an important question of how to develop a sustainable bioenergy and bioeconomy without compromising the food production for a growing world population. A response to this challenge involves information on land cover, both for lands currently used in crop production and for transitional lands that could be used for growth of dedicated energy crops. Such information will have to be available over large areas, usually at regional, even at national scale (Zhou et al,, 2009; Gang et al., 2010).

Transitional lands refer to the lands transitioned over the past between types such as forest, grassland, forage land, and cropland due to factors such as climate or market fluctuations or other reasons. This category of lands is considered most promising for the production of dedicated bioenergy crops as a primary source of biomass feedstock for the development of the second generation biofuels, without compromising regular agriculture production. Identification of these lands for biomass growth potential is a gap in the Canadian national biomass inventory, which has been undertaken by Canadian Forest Service and Agriculture and Agri-food Canada, with funding from the federal, interdepartmental Program for Energy Research and Development (PERD) operated by Natural Resources Canada. The purpose of the transitional land mapping is to enable the application of earth observation data to fill this gap and to provide science-based information in support of bioenergy policy development.

Aimed at the task at a regional scale, Earth Observation data with medium spatial resolution are considered as the most effective data sources. Time series of the Moderate Resolution Imaging Spectroradiometer (MODIS) images and its derived information, such as Normalized Difference Vegetation Index (NDVI) and vegetation phenology, are then used to derive the required information. Time series of data have many advantages compared to a single time image as the former captures dynamic spectral information at various vegetation growth stages. However, it also poses challenge of such as how to efficiently use the 'high-dimensional' and a large volume of data and extract needed information for the issues at hand (Zhou et al., 2009). In this regard, a data mining tool is applied to explore and identify the optimal data combinations from all the data available in the vegetation growing season for land cover identification, and then for transitional land mapping. Results showed that, in general, longer time series of data set would yield higher accurate land cover classification than that of a shorter time series of data, and due to the medium spatial resolution of MODIS data, land cover identification accuracy also depends on the degree of homogeneity of the landscape setting of the region under study. The following sections will describe in some details about the data used, the methodology developed, and the results produced, as well as give a brief discussion.

2. DATA AND METHOD

MODIS sensors aboard the Terra and Aqua satellites have a revisiting frequency of 1 to 2 days, acquiring data in 7 spectral bands for land surface applications, with a spatial resolution ranging from 250 – 500 metres. It is one of the most advanced sensors available for large-scale terrestrial applications (Salomonson et al., 1989). To facilitate use of the datasets, Canada Centre fore Remote Sensing has developed a new technology to product of 10 days cloud-free composites of MODIS 7 land bands covering Canada and North America. Additional to other characteristics of this product, bands 3-7 are downscaled from the original spatial resolution of 500 meters to 250 meters (Luo et al., 2008). Therefore all the 7 bands of the cloud-free composites have the same spatial resolution and dimension.

For the study, only the data in the vegetation growing season (April to October) of the time series of MODIS 10 days cloud-free composites are analyzed. For the selected datasets there are 3 composites every month, and in total 21 composites are available from the beginning of the vegetation growing season to the end of the growing season. Considering all the 7 bands for each composite, there is a total of 147 'bands' at various periods of the vegetation growing season.

Two additional datasets, NDVI and phenology in the vegetation-growing season, are derived from the time series of MODIS bands. NDVI, representing 'greenness' of vegetation, has been widely used for vegetation mapping. In total, same as the number of the MODIS bands, there are also a time series of 147 NDVI values in the vegetation growing season as they are derived from MODIS band 1 and band 2.

Another derived dataset from the MODIS data is vegetation phenology parameters. Phenology represents vegetation periodic biological phenomenon, which can be derived from NDVI time series by function fitting. There are total 11 vegetation phenology parameters including starting and ending of growing season, seasonal amplitude, seasonally integrated NDVI, rate of increase at the beginning of the season, rate of decrease at the end of the season, etc. Vegetation phenology is derived from the MODIS data by the aid of TIMESAT (Jonsson and Eklundh, 2003) software.

The rationale of using NDVI and vegetation phenology to aid land cover information extraction is that different vegetation types at different growing stages may have distinctive NDVI and phenology phenomenon.

With all the information available (MODIS 10 days cloud-free composite bands, NDVI, and phenology parameters), all possible data combinations for maximizing the overall land cover identification accuracy for targeted land cover types are sought and assessed by using See5/C5.0 data mining tool which has been used or discussed by various studies (Keane et al., 2004; Pal and Mather 2003).

Year 2000 is the first year when MODIS was operational; also there is a good reference of Circa 2000 land cover map (Figure 1a shows the reference land cover map of Saskatchewan, the case study province of the activity) which was generated from Landsat TM images. Therefore year 2000 is set as the starting point for transitional land assessment.

10 major land cover types of Circa 2000 land cover map (Agriculture and Agri-Food Canada, 2008) were mapped in the study. They are Annual Cropland, Water bodies, Developed land, Native Grassland, Shrubland, Perennial (crop and pasture), Wetland, Deciduous, Coniferous and Mixed Forest.

From bioenergy land cover mapping point of view, it is not critical to separate Deciduous, Coniferous, and Mixed Forest land cover. They are grouped into Forest land cover type in the study. Wetland is conservative land, so it was masked out before the mapping. All the analysis and results described below are based on the redefined types for the land cover mapping.

It is found that point samples like ground reference collected from field survey using GPS are not the best representations for model training and for verification of results derived from MODIS images due to the relative coarse spatial resolution of the data. Instead, area sample method was used in the study. Area sampling means that ground reference is established not by a point data, but by an area which has a homogenous land cover.

A homogeneous pixel (area) means that the ground represented by the pixel on the image is covered by only one land type. Truly, land cover is not always homogeneous, and heterogeneity is universal. However, homogeneity and heterogeneity are relative terms. When they are applied to EO-based applications, they are determined by spatial resolution of images in relation to the size of features on the ground. A piece of land is heterogeneous for a coarse resolution image, but may be homogeneous for all its sub-areas with a finer spatial resolution imagery (also depends on object size). Homogeneity and heterogeneity are also affected by the level of a land cover classification system.

For the study, three types of landscape settings are considered within the dimension of a MODIS pixel: they are homogenous, dominant and heterogeneous. As described above, a homogeneous pixel implies that the ground represented by the pixel on the image is covered by only one land type; a dominant pixel means that over a half of the ground area is occupied by one land cover with other land covers mixed; and the rest are heterogeneous cases.

To identify and evaluate the distribution of these types within the study area, the reference Circa 2000 land cover map is geometrically matched to and superimposed on the MODIS images. A MODIS pixel then corresponds to 25 pixels of circa 2000 land cover map (which was rescaled from 30 m to 50m). A MODIS pixel is classified as homogeneous if and only if all the corresponding 25 sub-pixels on the Circa land cover map have the same land cover, or dominant if a land cover type (the dominant one) has more than 13 or more sub-pixels, or heterogeneous (other land cover combinations).

The total number of the MODIS pixels of the study area is 6,355,364, among which, the number of homogeneous pixels is 3,101,288, and the number of the pixels with a dominant land cover type is 2,954,431. The proportion of the two types of MODIS pixels is 48.80%, and 46.49%, respectively. In total, they occupy more than 95% of the study area. Less than 5% of the area is occupied by multiple land cover types without a dominant land cover defined above. Therefore the analysis for the homogeneous and dominant pixels represents over 95% of the landscape of the study region.

3. RESULT ANALYSIS

Using the method developed based on data mining technology for land cover information extraction, we evaluated three groups of variable combinations, 1) band combinations only, 2) band and NDVI combinations, and 3) band and phenology combinations. In the last two cases, in order to avoid redundancy, band 1 and 2 are excluded as they are used to derive NDVI and phenology parameters. In all the three groups of variable combinations, time parameter is factored in, which means that data at various time periods (of the 10 days composites) are combined for optimal classification results.

Sampling data both for training and verification are selected from the reference map: 4500 (500 for each land cover type) samples for training and 4500 (also 500 for each land cover type) samples for verification. Deciduous, Coniferous, and Mixed Forest are sampled and classified separately, and their results are aggregated as Forestland. Therefore, as a group, Forest has 1500 samples in total. In the selection process, it is ensured that the samples were selected independently and randomly over the study region. These samples are considered as area ones.

For accuracy evaluations, two cases are explored: 1) training and verification samples are all homogeneous pixels, and 2) training samples are homogeneous pixels and verification samples are dominant pixels. The former represents around 49% of the situations of the study region while the later represents more than 95% of the situations, so it is more realistic.

3.1 Results and analysis of homogeneous pixels for training and verification

Table 1 lists the lowest and highest Kappa values of the three groups of variable combinations (over 20,000 combinations in tiotal) for land cover identification using homogeneous pixels for both training and verification processes. The highest Kappa values of the three groups of data combinations are almost the same, but the lowest has remarkable variations. For the low end of Kappa, the single band of a single time stamp yields the lowest Kappa (0.24), while the combination of band and vegetation phenology yields the highest Kappa (0.70). An analysis of the Kappa values of different variable combinations reveals that, in general, the longer the length of the timer series of data used, the higher the Kappa value, then the higher identification accuracy. This observation thus suggests that time series data are better than a single snapshot for pattern extraction of land cover. Usually, the longer the time series is, the better. For example, the lowest Kappa for phenology/band combination has much bigger value than that of other two combinations as phenology parameters are derived from all the season data.

Data combination	Kappa	
	Lowest	Highest
Bands in different time period	0.24	0.84
NDVI + Bands in different time period	0.51	0.85
Vegetation phenology + Bands in different time period	0.70	0.85

Table 1. Kappa for land cover classification

Land cover type	Cropland	Forest	Grassland	Shrubland	Perennial	Develoed	Water	%
Cropland	**463**		8	3	18	8		92.60
Forest Land	1	**1447**	1	41	7		3	96.47
Grassland	10	2	**412**	3	62	11		82.40
Shrubland	4	68	10	**388**	27	1	2	77.60
Perennial	34	4	79	12	**359**	12		71.80
Developed	9	1	1	3	13	**473**		94.60
Water	3					5	**492**	98.40
Average								**87.70**

Table 2. Accuracy matrix of a phenology/band combination for land cover information extraction (one of the best combinations)

Table 2 shows the result of one of the NDVI/band combinations with the highest Kappa value against the reference data. The numbers in bold in the diagonal of the matrix represent the number of pixels that are correctly classified. The numbers suggest that the developed method with the data combination can discriminate most of the land cover types with an averaged accuracy percent about 88%. Among the 7 land covers, 4 of them have accuracy over 90%, and two of them are lower than 80% -- they are Perennial (~72%) and Shrubland (~78%) land cover types. Among the targeted land covers, native Grassland is mostly mixed with Perennial land cover (62 out of 500 samples). This is explainable as Perennial land includes tamed Grassland which is similar to native Grassland. Perennial land is also mostly mixed with native Grassland (79 out of 500 samples). This can also be explained with the same reason that native Grassland is mixed with Perennial land in a certain degree in real situations. If the native Grassland and Perennial land cover types were combined (both land types could be potentially used for bioenergy crop growth), the classification accuracy would become higher. Shrubland and Forest land are mostly mixed as thick Shrubland may have the similar spectral signatures as Forest land.

For transitional land mapping, the transitions among Cropland, Forest land, Perennial and Grassland are of the most interest. From Table 2, it can be seen that, Cropland is not confused with Forest land, and only has 1.6% and 3.6% misclassified as Grassland and Perennial land, respectively. It means that if a piece of Cropland is changed to forest land from one year to another, we have a chance of 92.6% that the change is true, and if Cropland is changed to Grassland or Perennial land, the rate of the mistakes made for detecting the change is also very small.

Similarly, Forest land (Coniferous, Deciduous, and Mixed Forest) is almost not confused with Cropland (only 1 out of 1500 forest land samples was classified as cropland). This fact implies that if Forest land cover is changed to Cropland, we have a confidence that the change is true. Table 2 also indicates that Forest land is not mixed much with Perennial and Native Grassland, with only 1 and 7 out of 1500 were classified as Grassland, and Perennial land, respectively. Hence, if a piece of Forest land cover is detected a change to Grassland or Perennial land, the change is likely true.

Although Grassland and Perennial land are mixed by each other, they are separable with Forest land, and only 2 and 4 out of 500 samples of Grassland and Perennial land are classified as Forest land cover, respectively. This means that if a change is detected from Grassland or Perennial land to Forest land, the change is highly likely. The same conclusion can be made for a change from Grassland or Perennial land to Cropland, although the confidence is slightly lower as there are 10 out of 500, and 34 out of 500 samples of native Grassland and Perennial land are classified as Cropland, respectively.

Figure 1 shows the comparison of the reference map and the classified land cover map using the method discussed above with the time-series of 10 days cloud-free MODIS composite data. Figure 1a is Saskatchewan portion of Circa 2000 land cover map. The reference map was generated by Agriculture and Agri-food Canada using 30m Landsat TM data, while Figure 1b is generated by using 250m MODIS data (NDVI + bands) and the developed method. Although Figure 1b is not as rich in term of spatial details, it is evident that the identified land cover types are generally agreeable to the Circa 2000 land cover map.

3.2 Results of homogeneous pixels for training and dominant pixels for verification

The above assessment and analysis are based on the results generated using homogeneous samples for training and verification, which can be applied to the landscape with homogeneous land cover.

For the evaluation, we used the same homogeneous sample pixels for model training, but randomly and independently selected dominant land cover pixels (which could enclose some homogenous pixels) for verification. Statistically, the verification samples represent more than 95% situations of the MODIS pixels of the study region. The same three groups of variable combinations described above were evaluated for the process. Table 3 lists the highest accuracy from the three variable combinations above, respectively.

In comparison with the results of using homogeneous pixels for training and verification, the three variable combinations using homogeneous pixels for training and dominant pixels for verification yield similar but lower accuracy. However, the overall accuracy of the land cover identification reaches about ~80%. Considering the spatial resolution of MODIS, the results are encouraging.

Land Cover Type	Accuracy (%)		
	Bands	NDVI + bands	Phenology + bands
Cropland	89.20	90.40	88.80
Forest land	87.13	87.73	88.40
Grassland	70.40	70.80	70.80
Shrubland	68.60	70.40	70.00
Perennial	70.00	71.80	73.00
Developed	78.00	77.80	76.40
Water	87.40	89.00	88.60
Average	78.68	79.70	79.43

Table 3. Accurate percent of land cover identification using dominant pixels for verification

It can be seen from Table 3 and Table 2, Cropland cover type has similar accuracies from the two methods. This is because Cropland in Saskatchewan has large parcels. Once Cropland cover becomes dominant, it is likely that all the 25 sub-pixels of a MODIS pixel are Cropland cover. Other land covers yield a lower accuracy (~2%-~8% lower) except Perennial land cover. The results are explainable because the land covers other than the dominant one within a MODIS pixel would contribute to the spectral information, and then somewhat confuse the tool for accurately identifying the dominant land cover. The degree of the confusion may depend on the number and the types of land covers within the pixel of the dominant land cover.

An exception is that Perennial land cover which has an equal or a slightly higher accuracy for NDVI and phenology with bands combinations. The reason for the higher accuracy needs further investigation.

4. CONCLUSION AND DISCUSSION

Land cover mapping and its subsequent transitional land assessment at a regional and national level require large coverage and adequate spatial and temporal resolutions of EO data. MODIS data is a reasonable choice. Although promising, however, based on our evaluations, its usefulness depends on two critical variables: 1) landscape and 2) spectral characteristics of targeted objects. Once the targets are determined, the major factors of affecting land cover identification are the spatial distribution patterns of land covers. The degree of landscape heterogeneity under the study area determines the degree of mixed information within a pixel, and then plays a major role in affecting the accuracy of land cover identification.

Our study shows that, in the study region of Saskatchewan, about 49% of landscape is homogeneous with only one land cover, and 46% of the land is dominated by one land cover based on the size of a MODIS pixel and Circa 2000 land cover map. The accuracies of the land cover identification for the homogeneous and dominant landscape (includes homogeneous landscape) are about 88% and 80%, respectively. These suggest that MODIS may provide valuable information for the transitional land mapping for Saskatchewan region although further evaluation is needed such as improving the quality of the time-series of MODIS data and the method developed. For other regions of Canada, the similar analysis needs to be conducted in order to determine if medium spatial resolution of EO data is efficient for transitional land mapping.

The current step is to identify land cover (biomass) classes and to map its spatial distribution as a potential feedstock for bioenergy production. The next step is, through analyzing multi-year time-series of MODIS, to identify the transitional lands that are potentially most sensitive to climate and market conditions, and therefore most likely to become available for bioenergy cropping when these environments change which is underway now.

REFERENCES

Agriculture and Agri-Food Canada, 2008. Land Cover. http://www4.agr.gc.ca/AAFC-AAC/display-afficher.do?id=1227635802316&lang=eng

Hong G., A, Zhang, F. Zhou, L. Townley-Smith, B. Brisco, I. Olthof. 2010. Biomass type identification potential of Radarsat-2 and MODIS images for the Canadian prairies. Submitted to Canadian Journal of Remote Sensing. 2010.

Jönsson, P. and Eklundh, L. 2003. Seasonality extraction from satellite sensor data. In *Frontiers of Remote Sensing Information Processing*, edited by Chen, C.H. World Scientific Publishing. pp 487-500.

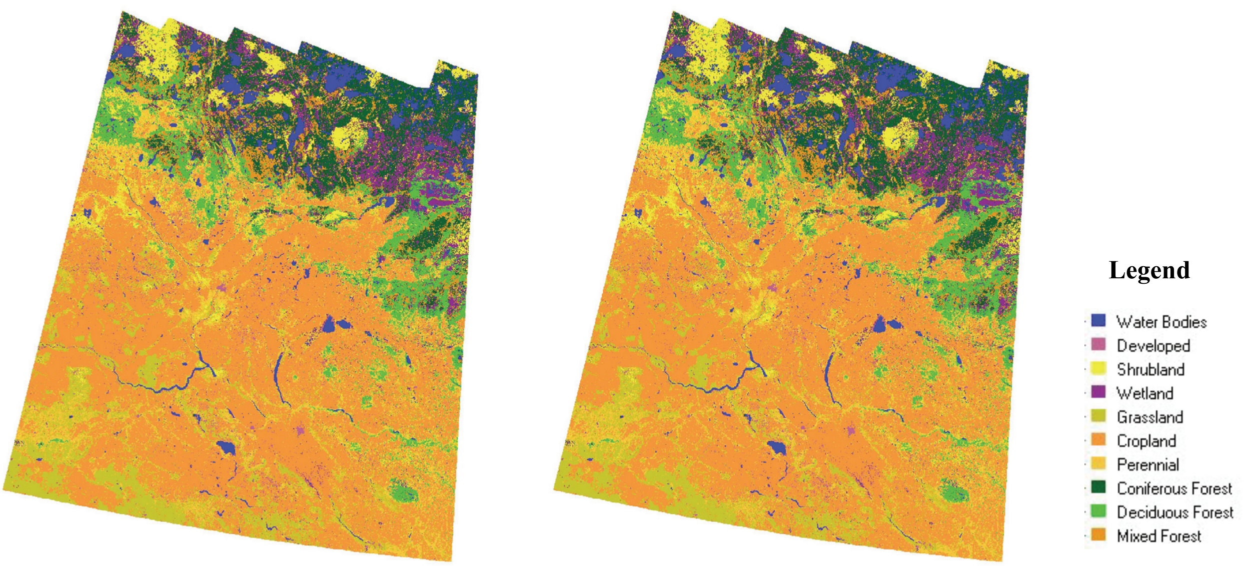
Luo, Y, A. P. Trishchenko, K. V. Khlopenkov, 2008. Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America. *Remote Sensing of Environment*, *Volume 112, Issue 12*, *15 December 2008*, pp 4167-4185.

Keane R. E., G. J. Cary, I. D. Davies, M. D. Flannigan, R. H. Gardner, S. Lavorel, J. M. Lenihan, C. Li, T. S. Rupp. 2004. A classification of landscape fire succession models: spatial simulations of fire and vegetation dynamics. *Ecological Modelling*, 179 (2004) pp.3-27.

Pal, M. and Mather, P.M. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens. of Env*., 86:554-565.

Salomonson V.V., W.L. Barnes, P.W. Maymon, H.E. Montgomery and H. Ostrow, 1989. MODIS: Advanced facility instrument for studies of the earth as a system, *IEEE Transactions on Geoscience Remote Sensing* **27** (2) (1989), pp. 145–153.

Zhou, F., A. Zhang, G. Hong, L. Sun, 2009. NDVI derived vegetation phenology for land use change investigation. *Oral presentation at The Second Conference on Earth Observation for Global Changes* (EOGC2009) in Chengdu, China, 25-29 May 2009.



1a 1b

Figure 1. Land cover maps of Saskatchewan in 2000
1a) Land cover circa 2000 (30m spatial resolution)
1b) Land cover map generated by using MODIS data (250m spatial resolution)