Worldwide inventory of landscapes through segmentation of global land cover dataset

J. Nowosad*[1] and T. F. Stepinski[1]

[1]University of Cincinnati, Cincinnati, OH, 45221, USA

*Email: nowosad.jakub@gmail.com

Abstract

Complete global inventory of landscapes (patterns of land cover) is obtained by segmentation/classification of the CCI-LC – a high resolution global land cover dataset. The CCI-LC is first segmented into a large number of small land units. The pattern of CCI-LC categories within each unit is tightly controlled by segmentation algorithm's merging parameter. Second, land units are classified into more manageable number (400-600) of landscape classes (LANDCs) based on composition and character of their patterns. Pattern properties are reflected in automatically generated class labels. The set of LANDCs provides a global inventory of landscapes. The final result of this work is a suite of vector files containing maps of segmentations and LANDCs at three different levels of spatial scale of pattern (9, 15, and 30 km). These maps differ from the GLC maps by having coarser spatial resolution but higher thematic resolution; they are also SQL-searchable. They have applications in macroecology to serve as proxies of vegetation structure, biotic composition, and can provide first-order information about geographical distribution of biodiversity. The method can be extended to multilayer segmentation for delineation of ecoregions.

Keywords: Global land cover, CCI-LC, Segmentation, Spatial pattern.

1 Introduction

High resolution global land cover (GLC) grids are obtained by classifying pixels in a global mosaic of Earth observation (EO) images into a several categories of surface's properties. They are most frequently used as a convenient repositories from which to obtain land cover for a particular region of interest. They are less frequently used in their entirety. This is because topics that require global coverage of land cover are low in numbers although they are very important. They include modeling of climate change, global assessment of land cover change, and delineation of ecoregions. The first two topics are well served by the GLC grid but the last one is not.

Recently developed GLC grid – the CCI-LC 2010 (http://maps.elie.ucl.ac.be/CCI) – has been developed especially for the purpose of climate change modeling; it has high spatial resolution (300 m) but comprise of only 22 land cover categories based on the FAO/UNEP Land Cover Classification System. An important part of delineation of ecoregions is to identify distinct landscapes – defined as land units having distinct patterns of land cover. For such purpose the resolution of CCI is too high and its thematic resolution is too low (Riitters, 2011; Omernik and Griffith, 2014).

Here we report on development of a novel global map and inventory of landscapes. This is achieved by a segmentation of the CCI-LC into a set of mutually exclusive and exhaustive segments – land units defined as tracts of land encompassing a cohesive pattern of CCI categories and dissimilar from adjacent segments. Cohesiveness of a pattern within each segment is controlled by segmentation algorithm's merging threshold parameter and is high – making a segment an ideal spatial unit of landscape. In addition, segments are classified into a number of landscape classes (LANDCs) – archetypes of land cover patterns; there are much more LANDCs than there are original CCI categories resulting in a significant increase in thematic resolution (at the cost of decreased spatial resolution). Our final products are SQL-searchable vector maps of segments and LANDCs providing delineation and complete inventorying of different landscapes worldwide. This products are available for download (`http://sil.uc.edu/cms/index.php?id=data-1#geocomp17`).

2 Methods

2.1 Segmentation of CCI-LC 2010

Unique patterns of land cover are delineated using the Geospatial Pattern Analysis Toolbox (GeoPAT) (Jasiewicz et al., 2015) – a collection of GRASS GIS (GRASS Development Team, 2016) modules for carrying out pattern-based analysis of large categorical spatial datasets. First, the global CCI-LC grid is tessellated into small square tiles (of the size $k \times k$ of cells). A mosaic of land cover categories (the category co-occurrence histogram of its constituent cells) for each tile encapsulates a local landscape (land cover pattern) at the scale of a tile. Afterwords, similar adjacent tiles are merged by a segmentation algorithm based on the principle of seeded region growing (SRG) (Adams and Bischof, 1994) as long as their internal cohesion (controlled by a value of merging parameter) is maintained. At the end of the segmentation the CCI-LC grid is partitioned into multiple segments (landscape units) collectively covering the entire Earth's land surface. Fig.1 (left panel) shows a fragment of CCI-LC covering a small region located in eastern Australia. The right panel of Fig.1 shows segmentation (with k =9 km) of of this region; note homogeneity of the landscape in each segment.

2.2 Classification of landscape units

Segmentation results in a very large number of landscape units with tightly controlled level of homogeneity. For global analysis it is useful to classify those units into LANDCs. Because we want LANDCs to be specific we aim at a relatively large number of them (still much smaller than the number of segments). For this reason we designed a custom classification scheme instead of relying on clustering algorithm. The classification is based on two properties of segments – the largest share and the second largest share. The largest share is the percentage of segment's area occupied by the most abundant CCI-LC category, and the second largest share is the percentage of the second most abundant category. Based on these two properties we create two new metrics, (largest share + second largest share) and (largest share - second largest share). The left panel in Fig.2 is a scatterplot of segments as represented by points having coordinates equal to the values of the two metrics. Note that set of all points form of a triangular region on the scatterplot. We divided this region into four zones called "matrix" (MX), "simple mosaic" (SM), "complex mosaic" (CM), and "intermediate mosaic" (IM). Points (segments) located in each zone are classified accordingly. Finally, we complete the classification of segments by adding a postfix to a zone-derived

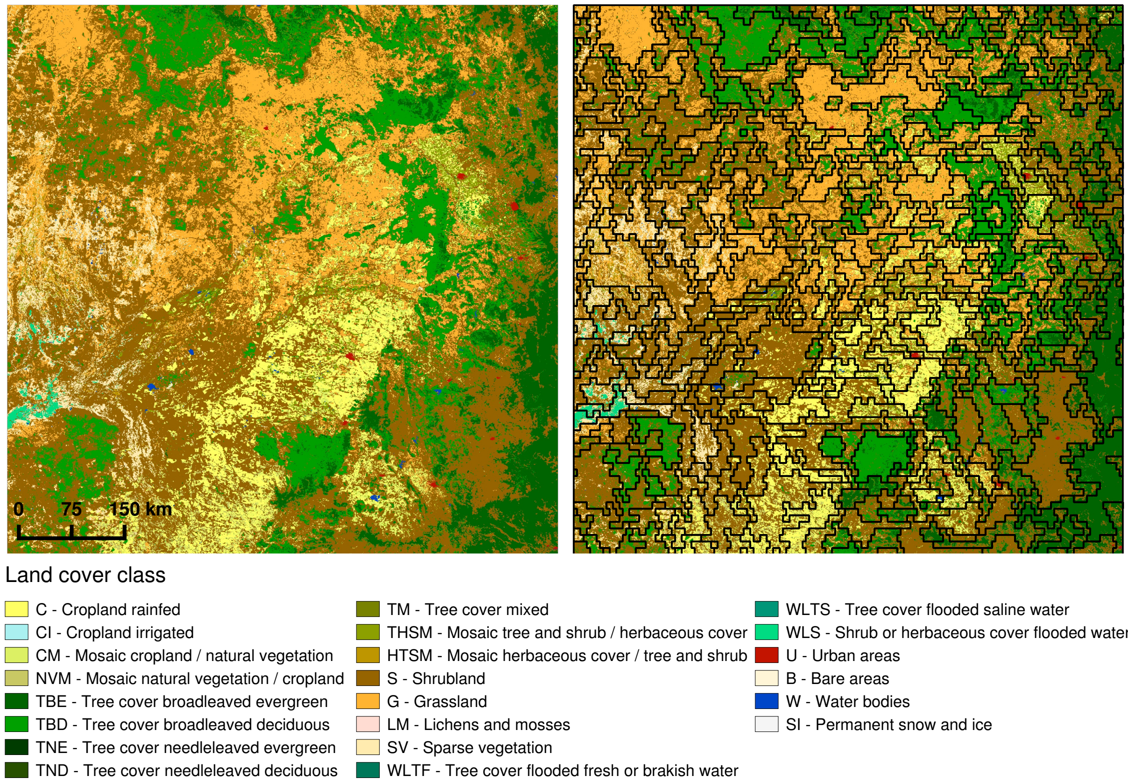

Figure 1: (Left) Fragment of CCI-LC covering a region located in eastern Australia, colors correspond to LC categories as described by the legend. (Right) The same region segmented into homogeneous landscapes.

label. These postfixes are abbreviations of name/names of dominant CCI-LC category/categories as listed in the legend to Fig. 1. Theoretically, this scheme allows for 6975 LANDCs, but much less are actually present and they are further reduced to the 200 most abundant + an additional number of LANDCs which are not abundant but are very distinct. The right panel in Fig.2 shows the map of LANDCs for a regions featured in Fig.1; there are 42 different LANDCs in this region. This map is a generalization of the segmentation map (Fig.1(right)). In comparison to the CCI-LC map (Fig.1(left)) it has a coarser spatial resolution but higher thematic resolution.

3 Results

Using a methodology described above we generated global maps of segments and global maps of LANDCs for three values of a tile's size k, 30 km (9,946 segments; 402 LANDCs), 15 km (36,284; 477), and 9 km (101,274; 594). The smaller the size of the tile the more fine-scale is the LANDCs map; maps in Fig.1 and Fig.2 are based on 9 km tiles. LANDCs map reveal that 63% of Earth's land surface area is covered by MX patterns (dominated by a single land cover category), 18.4% by SM patterns (two dominant land covers categories), 6.1% by CM patterns (several land cover categories but none dominant), and 12.5% by IM (one major land cover category and several secondary categories). Ten LANDCs with largest coverages are all MX, with the largest, MX-B, covering 12.5% of the land surface area. The SM–C.G, ranked 11, is the largest coverage SM pattern

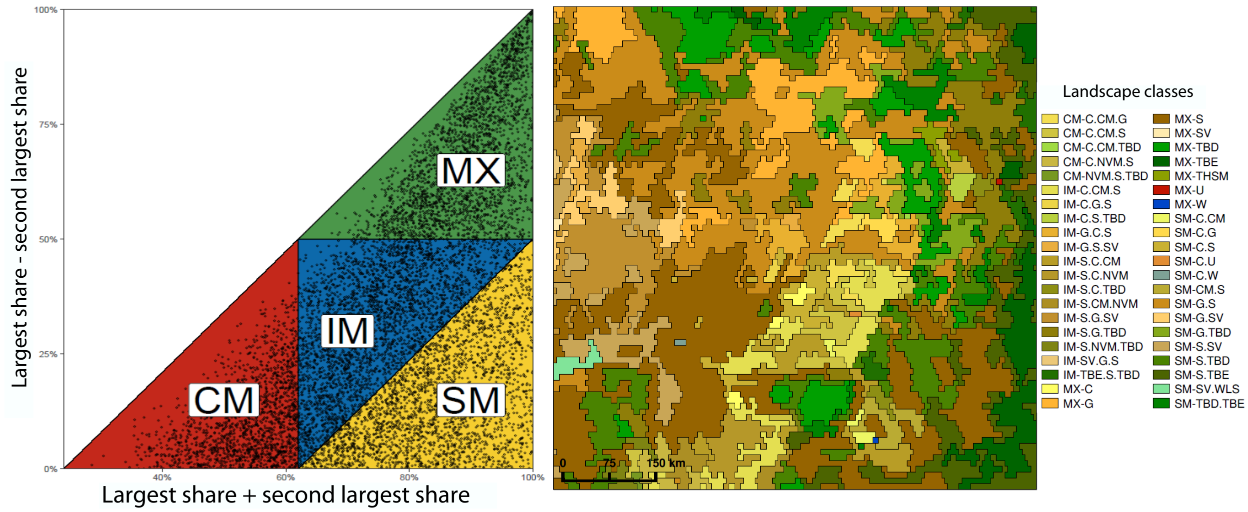

Figure 2: (Right) Classification of segments into four broad types corresponding to different characters of their land cover patterns. (Left) Map of 42 LANDCs in the region shown in Fig.1 and the legend describing each landscape.

with 1.4% of global coverage. The CM–C.CM.NVM, ranked 34, is the largest coverage CM pattern with 0.4% of global coverage, and the IM–C.CI.CM, ranked 36, is the largest coverage IM pattern with 0.36% of global coverage.

All products described in this paper, are freely available for download from University of Cincinnati Space Informatics Lab website at http://sil.uc.edu/cms/index.php?id=data-1. This include vector and PDF maps of segments and LANDCs, and an online version of LANDCs maps.

4 Conclusions and future work

The suite of global land cover pattern maps generated using a process described in this paper is the first set of high resolution vector maps delineating and inventorying different landscapes worldwide. The presented methodology of segmentation–classification of land cover raster provides a robust framework to automatically generate map of landscapes. Previously, such pattern-based maps were drawn manually (Wickham and Norton, 1994) and thus were restricted to very small areas. The key technology is the segmentation algorithm especially designed to work with pattern objects.

Land cover patterns maps (segments and LANDCs) are the first step for automatic delineation of ecoregions. They are created from a single thematic layer – land cover, whereas ecoregions need to be delineated from multiple thematic layers (land cover, terrain, climate). An extension of our method to work with multiple layers is conceptually straightforward although technical details remain to be worked out. In the meantime LANDCs are useful in their own right. In macroecology, they can replace GLC maps as proxies for vegetation structure at the general biome level, and, by conjecture, as proxies for biotic composition (Coops et al., 2009). They can provide first-order information about geographical distribution of biodiversity and ecological processes (Heikkinen et al., 2004; Fuller et al., 2005).

5 Acknowledgements

This work was supported by the University of Cincinnati Space Exploration Institute and by the grant NNX15AJ47G from the National Aeronautics and Space Administration.

6 References

Adams, R. and L. Bischof
1994. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647.

Coops, N. C., M. A. Wulder, and D. Iwanicka
2009. Exploring the relative importance of satellite-derived descriptors of production, topography and land cover for predicting breeding bird species richness over Ontario, Canada. *Remote Sensing of Environment*, 113:668–679.

Fuller, R. M., B. J. Devereux, S. Gillings, G. S. Amable, and R. A. Hill
2005. Indices of bird-habitat preference from field surveys of birds and remote sensing of land cover: a study of south–eastern England with wider implications for conservation and biodiversity assessment. *Global Ecology and Biogeography*, 14:223–239.

GRASS Development Team
2016. *Geographic Resources Analysis Support System (GRASS) Software*. Open Source Geospatial Foundation, USA.

Heikkinen, R. K., M. Luoto, R. Virkkala, and K. Rainio
2004. Effects of habitat cover, landscape structure and spatial variables on the abundance of birds in an agricultural-forest mosaic. *Journal of Applied Ecology*, 41:824–835.

Jasiewicz, J., P. Netzel, and T. Stepinski
2015. GeoPAT: A toolbox for pattern-based information retrieval from large geospatial databases. *Computers & Geosciences*, 80:62–73.

Omernik, J. M. and G. E. Griffith
2014. Ecoregions of the conterminous united states: evolution of a hierarchical spatial framework. *Environmental management*, 54(6):1249–1266.

Riitters, K. H.
2011. Spatial patterns of land cover in the United States: a technical document supporting the Forest Service 2010 RPA Assessment. Technical Report Gen. Tech. Rep. SRS-136, Department of Agriculture Forest Service, Southern Research Station, Asheville, NC.

Wickham, J. D. and D. J. Norton
1994. Mapping and analyzing landscape patterns. *Landscape Ecology*, 9(1):7–23.