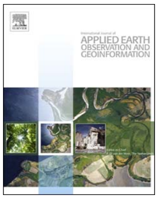
Int J Appl Earth Obs Geoinformation 69 (2018) 110–118

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag



Towards machine ecoregionalization of Earth's landmass using pattern segmentation method

Jakub Nowosad, Tomasz F. Stepinski*

Space Informatics Lab, Department of Geography and GIS, University of Cincinnati, Cincinnati, USA

ARTICLE INFO

Keywords:
Global ecoregions
Environmental variables
Regionalization
Segmentation
Pattern

ABSTRACT

We present and evaluate a quantitative method for delineation of ecophysiographic regions throughout the entire terrestrial landmass. The method uses the new pattern-based segmentation technique which attempts to emulate the qualitative, weight-of-evidence approach to a delineation of ecoregions in a computer code. An ecophysiographic region is characterized by homogeneous physiography defined by the cohesiveness of patterns of four variables: land cover, soils, landforms, and climatic patterns. Homogeneous physiography is a necessary but not sufficient condition for a region to be an ecoregion, thus machine delineation of ecophysiographic regions is the first, important step toward global ecoregionalization. In this paper, we focus on the first-order approximation of the proposed method – delineation on the basis of the patterns of the land cover alone. We justify this approximation by the existence of significant spatial associations between various physiographic variables. Resulting ecophysiographic regionalization (ECOR) is shown to be more physiographically homogeneous than existing global ecoregionalizations (Terrestrial Ecoregions of the World (TEW) and Bailey's Ecoregions of the Continents (BEC)). The presented quantitative method has an advantage of being transparent and objective. It can be verified, easily updated, modified and customized for specific applications. Each region in ECOR contains detailed, SQL-searchable information about physiographic patterns within it. It also has a computer-generated label. To give a sense of how ECOR compares to TEW and, in the U.S., to EPA Level III ecoregions, we contrast these different delineations using two specific sites as examples. We conclude that ECOR yields regionalization somewhat similar to EPA level III ecoregions, but for the entire world, and by automatic means.

1. Introduction

Terrestrial ecoregions (hereafter referred to as ecoregions) are the result of regionalization of land into areal units of homogeneous ecosystem which contrast from surroundings. Because the means of such regionalization are not the part of their definition, ecoregions are an umbrella term with a clear general intent, but with specifics depending on how ecosystems are described and compared (Gonzales, 1966; Jax, 2006; Haber, 2011), on the spatial scale considered, and on the approach to the regionalization procedure.

The need for ecoregions was initially driven by conservation planning (Larsen et al., 1994), but their usage has since expanded to tabulating environmental information in general. Ecoregions are mapped at different scales from global to local. At the broadest scale regionalization of ecoregions relies on climatic, geologic, and geomorphologic divisions (Bailey, 2014). At the finer spatial scale more attention is given to landscape patterns, vegetation types and biodiversity, and, eventually, at the local scale, attention shifts to specific species of flora and fauna (see, for example, Blasi et al., 2014).

Several different approaches have been applied to a delineation of ecoregions on the broad scale. Bailey (1989, 2014) developed a deductive approach wherein delineation of ecoregions follows from identifying environmental variables responsible for differentiating between ecosystems and drawing boundaries where these variables change significantly. Resulting regionalization is known as Bailey's Ecoregions of the Continents (BEC). Olson et al. (2001) applied a synthetic approach wherein ecoregions are delineated based on a large body of previous biogeographical studies. Existing information was refined and synthesized using expert judgment. Resulting regionalization is referred to as Terrestrial Ecoregions of the World (TEW). The similar synthetic methodology was applied on a regional scale to develop the Digital Map of European Ecological Regions (DMEER) (Painho et al., 1996) and the Interim Biogeographic Regionalisation for Australia (IBRA) (EA, 2000). Omernik (1987) used a weight-of-evidence approach to delineate ecoregions in the conterminous U.S. In this approach maps of environmental variables are overlaid and ecoregions

---

* Corresponding author.
E-mail address: stepintz@uc.edu (T.F. Stepinski).

https://doi.org/10.1016/j.jag.2018.03.004
Received 8 September 2017; Received in revised form 7 March 2018; Accepted 12 March 2018

are delineated by expert judgment through reconciling differences between variability of individual variables. The difference between Bailey's deductive approach and the weight-of-evidence approach is that whereas in the first the reconciliation follows an a priori determined scheme while in the second it is done on the case-by-case basis.

The issue with the synthetic approach to ecoregionalization (TEW, DMEER, IBRA) lies in the lack of quantitative framework. TEW is a compilation of local regions taken from pre-existing, independently conducted studies. On one hand, this may be viewed as a positive because TEW combines expert knowledge of the broad community. On the other hand, there are no straightforward means to inspect materials and protocols that contributed to the creation of TEW. As there is no underlying quantitative framework, there are no quantitative criteria to assess the quality of TEW. Therefore, no systematic checks, modifications or objective updates to TEW are possible. Moreover, although many individual regions in TEW may be well-delineated, as a whole, TEW lacks overall consistency. A user has no means of knowing which regions are well-delineated and which are not. TEW legend conveys a short description of a region which usually pertains to a combination of region's geography, climate, and flora. Because regions in TEW lack quantitative description, the inter-regions comparison is limited to contrasting their short descriptions in the legend.

The weight-of-evidence approach (Omernik, 1987; Omernik and Griffith, 2014) also lacks quantitative framework, but, it is rooted in a clear conceptual framework – "Ecoregions should depict areas of similarity in the collective patterns of all biotic, abiotic, terrestrial, and aquatic ecosystem components with humans being part of the biota" (Omernik and Griffith, 2014). Regions are delineated manually by experts on the basis of visually perceived breaks in aforementioned patterns. In this approach the resulting ecoregionalization may be consistently delineated (to a degree that humans perception can be consistent), but, like in the case of TEW, a user has no means of determining the quality of the regionalization. Omernik's legend has the character similar to that in TEW, the inter-regions comparison is limited to contrasting their descriptions in the legend.

In BEC a delineation of regions follows the Köppen–Trewartha climate classification modified by land cover information (Bailey, 2014). BEC legend conveys regions' climatic and floristic character. Because of its reliance on the climate, BEC offers only the broadest scale regionalization.

An attempt to automate the ecoregionalization process using a multivariate *k*-means clustering algorithm was made by Hargrove and Hoffman (2005) and followed up by Kumar et al. (2011). In such framework vectors of environmental variables are associated with each pixel (a tract of land corresponding to the resolution of the data) and pixels agglomerated into larger zones (ecoregions) on the basis of the Euclidean distance between these vectors. Such automated approach addresses issues related to objectivity, consistency, and inter-region comparability (see our discussion above), however, its ability to yield a useful ecoregionalization is limited by the choice of clustering as a technique enabling the automation. Clustering leads to a delineation of non-contiguous, highly fragmented zones, with the fragments spread over wide areas. Clustering may be well-suited for classification but it is ill-suited for mapping. Mapping needs to be based on characteristics which are macroscopically recognizable (Klijn et al., 1995), which environmental variables measured on the scale of an individual pixel are not.

In this paper, we propose and describe an approach to data-driven machine regionalization of the entire terrestrial landmass capable of producing a useful global map of ecophysiographic regions. We call the resultant regions "ecophysiographic" because they are mapped based on physiography but aim at delineating ecosystems as well. This is consistent with the notion that ecoregionalization on larger scales should be based on physiography (Klijn et al., 1995; Sayre et al., 2014). Following Omernik and Griffith (2014), our mapping is based on macroscopically recognizable *patterns* of physiographic categorical variables, but a decision on where to put boundaries between the regions is made by a segmentation algorithm instead of a committee of experts. Segmentation is a natural choice for machine delineation of regions because it is an algorithmic implementation of regionalization. Quantitative assessment of segmentation quality corresponds directly to the qualitative notion (McMahon et al., 2001; Loveland and Merchant, 2004; Omernik and Griffith, 2014) that regions should be internally as homogeneous as possible with respect to the environment, and they should stand out from adjacent regions.

Pattern-based segmentation is the enabling technology behind our proposed method but it also presents a big challenge. This recently developed technology (Jasiewicz et al., 2015, 2017) works at present only with patterns of a single variable, not with patterns of multiple variables as our proposed framework calls for. However, we find a high level of spatial association between categories of various physiographic variables, thus we can achieve a viable regionalization by segmenting the landmass on the basis of patterns of the land cover alone. The quality of such approximation is checked a posteriori.

The goals of this paper are as follows. (1) To describe how pattern-based segmentation technique can be used for automatic creation of a global map and the legend of ecophysiographic regions. (2) To demonstrate that a segmentation based only on patterns of land cover yields a viable ecoregionalization. (3) To compare such ecoregionalization with TEW. (4) To provide a spatial database of delineated regions with a detailed quantitative description of patterns in each region.

2. Data and methods

Table 1 lists four global physiographic datasets we used to calculate associations between categories of land cover, climate, topography, and soils, and to calculate homogeneity of delineated regions. Our choice of environmental variables is very similar to that made by Sayre et al. (2014) except we use newly available (Hengl et al., 2017) soil types data (reclassified to 12 orders) instead of lithology used by Sayre et al. (2014) as a proxy for soils. We also use the newest global land cover dataset – the European Space Agency (ESA) Climate Change Initiative (CCI) global land cover map (thereafter referred to as CCI-LC). Note that all variables are categorical. Land cover is arguably the most ecologically important of the four variables because it was demonstrated to provide the first-order information about geographical distribution of biodiversity and ecological processes (Siriwardena et al., 2000; Maes et al., 2003; Eyre et al., 2004; Heikkinen et al., 2004; Fuller et al., 2005; Luoto et al., 2006). Details about the CCI-LC land cover dataset including its accuracy can be found in the Land Cover CCI Product User Guide V.2 (ESA, 2017).

2.1. Pattern-based segmentation of Earth's landmass

Segmentation was performed using the Geospatial Pattern Analysis Toolbox (GeoPAT) (Jasiewicz et al., 2015, 2017) – a collection of

Table 1
Global environmental datasets.

Variable	Dataset	Data type	Res.	Source
Land cover	CCI-LC 2010	Categorical grid (22 classes)	300 m	http://maps.elie.ucl.ac.be/CCI
Climate	Bioclimatic classification	Categorical grid (37 classes)	250 m	Sayre et al. (2014) modified from Metzger et al. (2013)
Topography	Landforms classification	Categorical grid (17 classes)	250 m	Karagulle et al. (2017)
Soil	SoilGrids soil classification	Categorical grid (12 classes)	250 m	Hengl et al. (2017)