Int J Appl Earth Obs Geoinformation 68 (2018) 127–138

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

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

Improved regional-scale Brazilian cropping systems' mapping based on a semi-automatic object-based clustering approach

Beatriz Bellón[a,b,*], Agnès Bégué[a,b], Danny Lo Seen[a,b], Valentine Lebourgeois[a,b], Balbino Antônio Evangelista[c], Margareth Simões[d,e], Rodrigo Peçanha Demonte Ferraz[d]

[a] CIRAD, UMR TETIS, F-34398 Montpellier, France
[b] TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, IRSTEA, Montpellier, France
[c] Embrapa Pesca e Aquicultura, Prolongamento da Avenida NS 10, cruzamento com a Avenida LO 18, sentido Norte, loteamento Água Fria, 77008-900 Palmas, TO, Brazil
[d] Embrapa Solos, Rua Jardim Botânico, 1024, 22460 – 000, Rio de Janeiro, RJ, Brazil
[e] Rio de Janeiro State University (UERJ), Rua São Francisco Xavier, 524, Maracanã, 20550-900, Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Keywords:
Cropping systems
Geographic object-based image analysis (GEOBIA)
Stratification
Clustering
Moderate resolution imaging spectroradiometer (MODIS)
Landsat

ABSTRACT

Cropping systems' maps at fine scale over large areas provide key information for further agricultural production and environmental impact assessments, and thus represent a valuable tool for effective land-use planning. There is, therefore, a growing interest in mapping cropping systems in an operational manner over large areas, and remote sensing approaches based on vegetation index time series analysis have proven to be an efficient tool. However, supervised pixel-based approaches are commonly adopted, requiring resource consuming field campaigns to gather training data. In this paper, we present a new object-based unsupervised classification approach tested on an annual MODIS 16-day composite Normalized Difference Vegetation Index time series and a Landsat 8 mosaic of the State of Tocantins, Brazil, for the 2014–2015 growing season. Two variants of the approach are compared: an hyperclustering approach, and a landscape-clustering approach involving a previous stratification of the study area into landscape units on which the clustering is then performed. The main cropping systems of Tocantins, characterized by the crop types and cropping patterns, were efficiently mapped with the landscape-clustering approach. Results show that stratification prior to clustering significantly improves the classification accuracies for underrepresented and sparsely distributed cropping systems. This study illustrates the potential of unsupervised classification for large area cropping systems' mapping and contributes to the development of generic tools for supporting large-scale agricultural monitoring across regions.

1. Introduction

Agricultural systems are constantly evolving to meet the increasing global demand for food production (FAO, 2009). Both the cropland expansion and the intensification of management practices contribute to meet such demand, increasing the concern on the environmental impact associated with these land-use changes (Tilman et al., 2011). In view of these large-scale dynamics, there is a growing need for regularly updated spatial information on the cropping systems at regional scales.

Cropping systems result of a combination of human and environmental factors, which naturally vary through space. A given cropping system thus presents local specificities, leading to a high spatial variability over large areas. In addition, when considering large areas, the cropping systems are embedded in a complex landscape mosaic composed of a high diversity of land use and land cover (LULC) types. Therefore, large-area mapping of cropping systems and their management practices, remains a challenging task, and as a result, their description and location remain rather unclear for most regions (Leenhardt et al., 2010).

Recent studies have however shown the usefulness of earth observation (EO) data to successfully map cropping systems, and some management practices over large areas (Atzberger, 2013; Bégué et al., 2015). In the last decade, various of these studies have successfully used vegetation index (VI) time series, mostly MODIS-derived, to map, at a regional scale, intensification practices such as the adoption of sequential cropping (e.g. Arvor et al., 2011; Brown et al., 2013; Cheema and Bastiaanssen, 2010; Guan et al., 2016; Gumma et al., 2016; Mingwei et al., 2008; Wardlow and Egbert, 2008)

Most of these studies, however, rely entirely on reference data, usually ground observations, to train supervised classifications (e.g.

* Corresponding author.
E-mail addresses: beatriz.bellon@cirad.fr (B. Bellón), agnes.begue@cirad.fr (A. Bégué), danny.lo_seen@cirad.fr (D. Lo Seen), valentine.lebourgeois@cirad.fr (V. Lebourgeois), balbino.evangelista@embrapa.br (B.A. Evangelista), margareth.simoes@embrapa.br (M. Simões), rodrigo.demonte@embrapa.br (R.P. Demonte Ferraz).

https://doi.org/10.1016/j.jag.2018.01.019
Received 6 November 2017; Received in revised form 29 January 2018; Accepted 30 January 2018

Arvor et al., 2011; Brown et al., 2013; Mingwei et al., 2008). Gathering reference data twice or more during a growing season to train classifications of sequential cropping systems represents a time and resource consuming task. This is especially the case for large areas since it is difficult to extensively cover the study area to guarantee representativeness of the intra-class spectral variability linked to the high spatial variability (Franklin and Wulder, 2002).

Contrastingly, unsupervised classifications, being solely based on the VI time series' information, are more operational without affecting the accuracy of the resulting maps, as shown by Cheema and Bastiaanssen (2010), Wardlow and Egbert (2008), Gumma et al. (2016). Hyperclustering techniques are commonly used, producing an initial large number of clusters that are then reduced by successive processing steps. This technique ensures that under-represented classes of interest are correctly isolated, and has the advantage of reducing the impact of the user-defined parameters (e.g. the number of clusters) on the results (Cihlar, 2000). However, the classification refinement techniques used in these studies, such as merging or splitting clusters, are highly expert knowledge-driven and often rely on external data (e.g. cropland masks, reference ground data), which limits their reproducibility.

In order to deal with the high intra-class spectral variability without depending on external data, cropping systemś mapping over large areas would largely benefit from spectral variability reduction steps prior to classification. Geographic object-based image analysis (GEOBIA) techniques are widely used to reduce spectral variability by grouping individual pixels into thematically meaningful objects that summarize their spectral properties (Blaschke et al., 2014; Bunker et al., 2016; Hay and Castilla 2008). Another spectral variability reduction strategy, referred to as stratification, consists in partitioning the study regions into "biophysical-meaningful" sub-regions prior to classification (Cai et al., 2011; Cano et al., 2017; Franklin and Wulder, 2002; Vintrou et al., 2012; Wardlow and Egbert, 2008; Xiong et al., 2017). Traditionally, external maps are used (e.g. ecoregion, agro-ecological zoning maps), but some novel approaches have started to explore the potential of EO data-based landscape stratifications for large area land-cover classifications (e.g. Cano et al., 2017; Vintrou et al., 2012).

Based on this context, we hypothesize that the use of stratification prior to unsupervised classification improves the accuracy of the regional-scale cropping systems' classifications at the field-level by reducing the spectral variability of the EO data. To test our hypothesis, we propose a new semi-automatic GEOBIA clustering approach, and compare the classification accuracies obtained with two variants of the approach: an hyperclustering approach, which includes a field-level segmentation step prior to classification, and a landscape-clustering approach, which combines two spectral variability reduction steps prior to classification, a field-level segmentation and a landscape stratification.

This work contributes to the development of simple and generic tools to improve large-area cropping systems' mapping with new classification methods fully based on EO data. The approach benefits from the high spatial resolution of a Landsat mosaic and the high temporal resolution of a MODIS VI time series, and is tested on the Tocantins state for the 2014–2015 growing season. The classification results are evaluated with ground-truth data and the annual crop estimates of the Brazilian Institute of Geography and Statistics (IBGE) (IBGE, 2014).

2. Materials

2.1. Study region

The Tocantins state, located in the centre-north of Brazil (Fig. 1a), is an agricultural region of strong anthropic influence which has undergone fast land-use changes over the last decades. Since its creation in 1988, Tocantins experiences a growing population and a fast-evolving economy based on a developing agricultural sector (Pedroso da Silva and de Almeida, 2007; Fornaro, 2012). Apart from the extensive development of the livestock systems which has been induced by the spreading of the transport network since the 1970s, this region undergoes, since the last decade, an expansion of the cropland mainly to the detriment of areas of degraded pasture, but also of some native vegetation areas (Barona et al., 2010; Silva, 2007; Fornaro, 2012). This trend is greatly driven by the global demand for Brazilian soybean, which is responsible for the expansion of the soybean cultivated areas in the latest agricultural frontier of Brazil, the MATOPIBA region, named after the states of Maranhão, Tocantins, Piauí and Bahia (Bolfe et al., 2016; Fornaro, 2012).

As a result, three soybean-based cropping systems, characterized by highly mechanized management practices and large field sizes, of around 100 ha, dominate the agricultural landscape of Tocantins. One of the cropping systems is a single cropping system, characterized by a single soybean crop cycle, and the other two cropping systems are characterized by a double-cropping pattern with two distinct crop cycles in the same growing season. This form of intensification is commonly adopted in this region since it allows increasing yields while reducing the need for expansion of the cropland areas (Spangler et al., 2017).

In one of the double cropping systems, henceforth referred to as the 'soybean-cereal double cropping system', a soybean crop is grown in the summer season, followed, in succession, by a cereal crop (mostly maize) during the end of summer to fall season. The other double-cropping system henceforth referred to as the 'rice-soybean double cropping system', is localized in the lowlands of the Formoso River basin and concerns a rainy season rice crop (grown under flood irrigation) succeeded by a late soybean crop.

2.2. Ground-truth data collection

Field surveys were conducted in the study area during October 2015, at the end of the 2014–2015 growing season, just before the rainy summer period and beginning of the next season. The sampling strategy aimed at collecting multiple observations for the three major cropping systems' classes and other LULC types over a wide variety of zones.

In particular, different agricultural zones, scattered throughout the study area were targeted and surveyed thoroughly, in order to capture the regional variability of the cropping systems. The selection of zones to survey was assisted by the NTSA (Thematic Unit on Agricultural Systems) of EMBRAPA (Brazilian Agricultural Research Corporation) Fisheries and Aquaculture, and was mainly focused on identifying zones where the crop fields are spatially "concentrated".

The survey equipment included a Trimble Yuma2 rugged tablet with an integrated GPS and Quantum GIS software, which allowed real-time navigation and direct recording of the waypoints in a spatial database. The tablet was equipped with the multispectral Landsat 8 OLI 2014–2015 mosaic (presented in Section 2.4.2), overlaid by the field-level segmentation (presented in Section 3.1.1). The sampling consisted in associating the observed class to its corresponding object in the segmentation layer. The commonly adopted no-tillage system allowed us to identify the cropping systems by the crop residues that cover the fields after the harvest of the crops.

A total of 900 GPS waypoints belonging to 35 different municipalities were registered (presented in Fig. 2), representing the three major cropping systems belonging to the annual cropland and other LULC types.

2.3. Area-based agricultural statistics

Area-based estimates of the soybean, second crop maize and rice crops were acquired from the Municipal Agricultural Production database (PAM) of the IBGE (2014). The 2015 harvested area estimates of the 139 municipalities of Tocantins were used to evaluate our classification results. The 'second crop maize' area estimates were used to evaluate the classification of the soybean-cereal double cropping