Remote Sensing of Environment 195 (2017) 13–29

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A multi-resolution approach to national-scale cultivated area estimation of soybean

LeeAnn King [a,*], Bernard Adusei [a], Stephen V. Stehman [b], Peter V. Potapov [a], Xiao-Peng Song [a], Alexander Krylov [a], Carlos Di Bella [c], Thomas R. Loveland [d], David M. Johnson [e], Matthew C. Hansen [a]

[a] Department of Geographical Sciences, University of Maryland, College Park, MD 20742, United States
[b] Department of Forest and Natural Resources Management, SUNY, Syracuse, NY 13210, United States
[c] Institute of Climate and Water- INTA Castelar, Buenos Aires, Argentina
[d] United States Geological Survey, Sioux Falls, SD, United States
[e] National Agricultural Statistics Service, USDA, Washington, DC 20250, United States

ARTICLE INFO

Article history:
Received 24 May 2016
Received in revised form 28 February 2017
Accepted 30 March 2017
Available online xxxx

Keywords:
Agriculture
Landsat
MODIS
Area estimate
Sample
Soybean
United States
Argentina

ABSTRACT

Satellite remote sensing data can provide timely, accurate, and objective information on cultivated area by crop type and, in turn, facilitate accurate estimates of crop production. Here, we present a generic multi-resolution approach to sample-based crop type area estimation at the national level using soybean as an example crop type. Historical MODIS (MODerate resolution Imaging Spectroradiometer) data were used to stratify growing regions into subsets of low, medium and high soybean cover. A stratified random sample of 20 km × 20 km sample blocks was selected and Landsat data for these sample blocks classified into soybean cover. The Landsat-derived soybean area was used to produce national estimates of soybean area. Current year MODIS-indicated soybean cover served as an auxiliary variable in a stratified regression estimator procedure. To evaluate the approach, we prototyped the method in the USA, where the 2013 USDA Cropland Data Layer (CDL) was used as a reference training data set for mapping soybean cover within each sample block. Three individual Landsat images were sufficient to accurately map soybean cover for all blocks, revealing that a rather sparse sample of phenological variation is needed to separate soybean from other cover types. In addition to stacks of images, we also evaluated standard radiometrically normalized Landsat inputs for mapping blocks individually (local-scale) and all at once (national-scale). All tested inputs resulted in area estimates comparable to the official USDA estimate of 30.86 Mha, with lower accuracy and higher standard error for national-scale mapping implementations. The stratified regression estimator incorporating current year MODIS-indicated soy reduced the standard error of the estimated soybean area by over 25% relative to the standard error of the stratified estimator. Finally, the method was ported to Argentina. A stratified random sample of blocks was characterized for soybean cultivated area using stacks of individual Landsat images for the 2013–2014 southern hemisphere growing season. A subsample of these blocks was visited on the ground to assess the accuracy of the Landsat-derived soy classification. The stratified regression estimator procedure performed similarly to the US application as it resulted in a reduction in standard error of about 25% relative to the stratified estimator not incorporating current year MODIS-indicated soybean. Our final estimated soybean area was 28% lower than that reported by the USDA, corresponding to a 20% field-based omission error related to underdeveloped fields. Lessons learned from this study can be ported to other regions of comparable field size and management intensity to assess soybean cultivated area. Results for the USA and Argentina may be viewed and downloaded at http://glad.geog.umd.edu/us-analysis and http://glad.geog.umd.edu/argentina-analysis, respectively.

1. Introduction

National-scale crop area estimates are typically obtained using ground data (Craig and Atkinson, 2013). However, many countries do not have resources to produce accurate and precise in situ estimates. Additionally, ground-based methods may differ between countries, limiting the ability to generate consistent and intercomparable global data for a given commodity crop. Earth observation data offer an alternative data source for improving the consistency of cropland assessments (Ozdogan and Woodcock, 2006). Given standard and repeatable methods and well-calibrated and freely available data, directly comparable area estimates may be achieved for key global commodity crops.

* Corresponding author.
E-mail address: mkinglee@terpmail.umd.edu (L. King).

http://dx.doi.org/10.1016/j.rse.2017.03.047
0034-4257/

In this study, we employ Moderate Resolution Imaging Spectroradiometer (MODIS) data to construct strata for selecting a sample from which Landsat imagery is used to map soybean cover in estimating national-scale soybean area. While the generic procedure has been prototyped for gross forest cover loss (Hansen et al., 2008), crop type is a more challenging land use theme, made difficult due to a diversity of crop phenologies and a corresponding need for high cadence image inputs, variable field sizes and shapes, and management practices. We employ a high-quality reference, United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) (Johnson and Mueller, 2009; Han et al., 2012), to demonstrate and evaluate the approach for the United States, using CDL labels in both calibration and validation. We vary the Landsat inputs and spatial scales of algorithmic implementation to evaluate the performance of different soybean classifications. We then test the method in Argentina in assessing its viability for soybean cultivated area estimation in the absence of CDL-like data.

MODIS data are collected daily globally, providing detailed temporal information on phenological variation that can facilitate crop-type identification. A number of pathfinding studies employed MODIS data to characterize crop type over large areas (Lobell and Asner, 2004; Chang et al., 2007; Wardlow and Egbert, 2008; Ozdogan, 2010; Shao et al., 2010). MODIS has recently been employed for mapping soybean area using vegetation indices (Wardlow and Egbert, 2008) and spectral profiles (Souza et al., 2015), temporal unmixing with Independent Component Analysis (Ozdogan, 2010), two-step filter approach (Sakamoto et al., 2011) and multiresolution approaches in conjunction with Landsat and aerial photography (Rudorff et al., 2011). However, the coarse spatial resolution of MODIS (250 m - 1000 m) data limit their use in area estimation given typical field size distributions (Fritz et al., 2015; Yan and Roy, 2014). Instead, using MODIS as a crop indicator can facilitate crop type area estimation using sample-based methods. Specifically, MODIS-derived crop information can assist in sampling design (stratification) and estimation (regression estimator) procedures (Hansen et al., 2008).

Another source of globally acquired and freely available data is the Landsat series of medium spatial resolution (30 m) sensors. The limitation of Landsat for crop monitoring has been the lack of temporal detail, with a nominal repeat coverage of 16 days per Landsat satellite. However, since the launch of Landsat 8, both the Operational Land Imager (OLI) onboard the Landsat 8 spacecraft and the Enhanced Thematic Mapper Plus (ETM+) onboard the Landsat 7 spacecraft have steadily increased growing season acquisitions (Wulder et al., 2015). The improved temporal cadence of Landsat's 30 m spatial resolution data make crop type identification and mapping a realistic possibility (Zhong et al., 2014), with cloud cover the primary limiting factor for optical systems such as Landsat (Waldner et al., 2015). For sample-based approaches, the use of Landsat in crop area estimation (measurement) is a viable option.

An invaluable resource for testing the mapping capabilities of satellite data is USDA's CDL, a medium spatial resolution crop-type map for the continental US generated annually since 2008. The CDL is a supervised classification derived from Landsat and Landsat-like imagery through a decision-tree methodology that is primarily trained and validated using field level USDA administrative planted crop type information (Boryan et al., 2011). Classification accuracy within the CDL for the dominant commodity crops like corn, soybeans, and wheat in intensively cultivated regions is typically very high (95%) and thus can act as a quality surrogate as validation for other crop type mapping applications. However, most countries do not have the information inputs for training and calibrating national-scale crop type mapping. As a result, extending the CDL methodology outside of the US is impractical. For global commodity crop monitoring, a more generic approach is needed.

The goal of this research is to evaluate the potential of freely available time-series satellite data for national-scale crop type mapping and area estimation. Gallego (2004) presents an overview of remote sensing data for land cover area estimation, describing approaches that use remote sensing data as the primary input. Our approach is one of many described by Gallego (2004), where "information from high resolution images plays the role of ground survey and is assumed to be accurate and unbiased." Specifically, we investigate the use of multi-temporal MODIS and Landsat inputs for accurate classification and area estimation through a sample-based approach, using soybean as an example commodity crop type. Although the method can be applied to any large-scale commodity crop, such as corn, soybean, wheat or rice, soybean is an appropriate thematic target to evaluate the methodology. Soybean is a crop that has been rapidly expanding in area over the past 30 years. Escalating demands for soybean over the next twenty years are anticipated to be met by an increase of 1.5 times the current global production, resulting in an expansion of soybean cultivated land area by nearly the same amount (Masuda and Goldsmith, 2009). Given such dynamics, consistent, intercomparable and repeatable assessments of soybean cultivated area are warranted.

2. Data and methods

2.1. Study areas and overall methodology

Our study areas consist of the soybean growing regions of the conterminous United States (CONUS) and Argentina, each divided into 20 km by 20 km blocks. For the CONUS, the study area consists of 4052 blocks that cover over 99% of soybean production between 2009 and 2012 based on a set of annual MODIS-derived soybean indicator maps (See Section 2.3.1 for details). Similarly for Argentina, the study area consists of 4268 blocks that cover over 99% of soybean production between 2007/2008 and 2012/2013 also based on a set of annual MODIS-derived soybean indicator maps (See Section 2.4.1 for details). Cultivated soybean area is the theme of interest and defined as planted soybean that reaches maturity in terms of plant growth and development. Our definition includes both single and double-cropped soybean.

Fig. 1 illustrates a flowchart of our methodology for estimating national-scale soybean area. Our multi-resolution approach consists of three major sequential modules: (1) spatially explicit stratification using coarse-resolution wall-to-wall MODIS data; (2) pixel-based soybean classification of moderate spatial resolution Landsat sample data; (3) statistical estimation of cultivated soybean area. The US was used to test various data inputs, classification models and area estimators, due to the availability of CDL data to evaluate and compare results of different experiments. Data inputs were applied at the local scale where each sample block was trained and classified independently (hereafter referred to as the local-scale model) and the national scale where all sample blocks were trained and classified at once (hereafter referred to as the national-scale model). Argentina was used as an example implementation of the method for a country lacking CDL-like reference data.

2.2. MODIS data

MODIS Terra MOD44C 16-day surface reflectance composite inputs were used as inputs for the respective US and Argentina soybean growing seasons. We used three of the seven MODIS land bands, 250 m band 1 (620–670 nm, red), 250 m band 2 (841–876 nm, near-infrared) and 500 m band 7 (2105–2155 nm, shortwave infrared) and brightness temperature data from the 1 km band 31 (10.780–11.280 μm, thermal). The shortwave infrared and thermal bands were resampled to 250 m resolution for this task. In addition to the spectral bands, Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) were calculated as inputs (Pittman et al., 2010, Jia et al., 2014, Dempewolf et al., 2014). The MODIS 16-day composites were converted to growing season phenological metrics (Chang et al., 2007) and used as independent variables in a regression tree model. Metrics are statistical transformations of image time-series that capture the