Remote Sensing of Environment 186 (2016) 465–478

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A stratified random sampling design in space and time for regional to global scale burned area product validation

Luigi Boschetti [a,*], Stephen V. Stehman [b], David P. Roy [c]

[a] Department of Natural Resources and Society, University of Idaho, Moscow, ID 83843, USA
[b] Department of Forest and Natural Resources Management, State University of New York, Syracuse, NY 13210, USA
[c] Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD 57007, USA

ARTICLE INFO

Article history:
Received 26 November 2015
Received in revised form 28 August 2016
Accepted 14 September 2016
Available online xxxx

Keywords:
Validation
Fire
Global burned area
Accuracy
Probability sampling
MODIS
Landsat

ABSTRACT

The potential research, policy and management applications of global burned area products place a high priority on rigorous, quantitative assessment of their accuracy. Such an assessment can be achieved by implementing validation methods employing design-based inference in which the independent reference data are selected via a probability sampling design. The majority of global burned area validation exercises use Landsat data to derive the independent reference data. This paper presents a three-dimensional sampling grid that allows for probability sampling of Landsat data in both space and time. To sample the globe in the spatial domain with non-overlapping sampling units, the Thiessen Scene Area (TSA) tessellation of the Landsat path/row geometry is used. The TSA grid is combined in time with the 16-day Landsat acquisition calendar to provide three-dimensional elements (voxels). This allows for implementation of stratified random sampling designs, where not only the location but also the time interval of the independent reference data is explicitly drawn by probability sampling. To illustrate this, we use a stratification methodology based on the Olson global ecoregion map and on the MODIS global active fire product. Using the global MODIS burned area product to establish a hypothetical population of reference data, we show that a sampling scheme based on the proposed stratification with equal sample allocation among strata is effective in reducing the standard errors of accuracy and area estimators compared to simple random sampling. Globally, the standard errors were reduced by 63%, 54%, 22% and 53% for overall accuracy, omission error, commission error and total burned area estimates respectively. By incorporating probability sampling in both the spatial and temporal domains, the present study establishes the foundation for rigorous design-based validation of global burned area products and, more generally, of terrestrial thematic products that have high temporal variability.

1. Introduction

The relevance of satellite derived burned area products for research, policy and management applications translates into the need for rigorous, transparent and repeatable product accuracy assessment (Morisette et al., 2006; Mouillot et al., 2014). Intercomparison of products made with different satellite data and/or algorithms provides an indication of gross differences and possibly insights into the reasons for any differences (Chang and Song, 2009; Boschetti et al., 2004). However, product comparison with an independent reference data set, not used to generate the product, is needed to determine accuracy (Justice et al., 2000). Validation is the term used here, and more generally, to refer to the process of assessing accuracy by comparison with independent reference data (Roy and Boschetti, 2009). Validation is required to provide accuracy information to help users decide if and perhaps how to use a product (Mouillot et al., 2014), and, combined with routine quality assessment (Roy et al., 2002), to identify any needed product improvements (Morisette et al., 2002). Burned area products generated with greater spatial and temporal coverage become increasingly difficult to validate in a statistically meaningful way. A number of global burned area products have been derived over the last two decades from coarse resolution satellite data including the GBA2000 (Global Burned Area 2000; Grégoire et al., 2003; Tansey et al., 2004), Globscar (Simon et al., 2004), GLOBCARBON (Plummer et al., 2006), L3JRC (Leicester, Louvain-la-Neuve, Lisbon & JRC, Tansey et al., 2008), Moderate Resolution Imaging Spectroradiometer (MODIS, Roy et al., 2005a; Roy et al., 2008; Giglio et al., 2009) and Fire CCI (Climate Change Initiative) (Alonso-Canas and Chuvieco, 2015) burned area products.

Research to validate global burned area products in a statistically rigorous manner is still ongoing. Ground based burned area measurement is time consuming and difficult to undertake over large regions (Cardoso et al., 2005). Because it is unrealistic to collect a global sample of burned area independent reference data from ground measurements,

* Corresponding author.
E-mail address: luigi@uidaho.edu (L. Boschetti).

http://dx.doi.org/10.1016/j.rse.2016.09.016
0034-4257/© 2016 Published by Elsevier Inc.

independent reference data derived from remotely sensed data must be used instead. The Global Burned Area Satellite Validation Protocol (Boschetti et al., 2009) was endorsed by the Land Product Validation subgroup of the Committee on Earth Observation Satellites (CEOS, http://lpvs.gsfc.nasa.gov/) and is based on a protocol developed for validation of the MODIS burned area product (Roy et al., 2005b). It defines the requirements for the use of satellite data as independent reference data but does not include any recommendations regarding sampling and accuracy metrics (Boschetti et al., 2009). Globally, fire activity is usually concentrated in a relatively short season (from a few weeks to several months) even in places where a large percentage of the landscape burns annually (Csiszar et al., 2006; Giglio et al., 2006a; Boschetti and Roy, 2008). Given the impermanent nature of many burned areas, with a spectral signal that can disappear in as little as a few weeks in certain savanna systems (Trigg and Flasse, 2000), and conversely the persistence of the burned signal for several years in boreal systems (Sukhinin et al., 2004), the time period covered by the reference data must be the same as the time period of the satellite burned area product being validated. The CEOS validation protocol requires that global coarse resolution burned area products (250 m–1 km spatial resolution) be validated using independent reference data derived from two or more Landsat-class images, allowing for comparison between the reference data and the burned areas detected by the global product in the period between acquisitions. The independent reference data must be derived with minimum error, either by visual interpretation (Roy et al., 2005b; Roy and Boschetti, 2009; Giglio et al., 2009) or by application of a semi-automatic algorithm followed by visual checking and manual refinement (Boschetti et al., 2006; Padilla et al., 2014, 2015). The use of bi-temporal image pairs ensures that burned areas that occurred before the first acquisition date are not mistakenly mapped as having burned between the two acquisition dates. Furthermore, the use of two acquisitions provides several interpretative advantages over single date data for mapping burned areas. These include a reduction in the likelihood of confusion with spectrally similar static land cover types (e.g., water bodies or dark soil) and the option to interpret the data by mapping relative changes rather than using single image interpretation approaches (Chuvieco et al., 2002; Roy et al., 2005b).

The characteristics of the independent reference data influence the reliability and the degree to which validation results are representative of the product. Assuming good quality independent reference data, the CEOS endorsed the descriptive validation hierarchy proposed by Morisette et al. (2006) to provide a guide to the degree of reliability of validation reporting with respect to the independent reference data sampling characteristics:

- Stage 1 Validation: Product accuracy has been estimated using a small number of measurements obtained from selected locations and time periods.
- Stage 2 Validation: Product accuracy has been assessed over a widely distributed set of locations and time periods, representative of the full range of conditions present in the product.
- Stage 3 Validation: Product accuracy has been assessed, and the uncertainties in the product established via independent measurements made in a statistically robust way that represents global conditions, and is characterized by the selection of reference data via a probability sampling i.e., design-based validation.
- Stage 4 Validation: Validation results for Stage 3 are systematically updated when new product versions are released, or when the time coverage of existing products expands.

The majority of regional to global scale burned area product validation exercises have been Stage 1, using Landsat data selected on the basis of availability (e.g., Barbosa et al., 1999; Fraser et al., 2000; Silva et al., 2003). An early attempt at Stage 2 validation, where the reference dataset covered a range of representative conditions, was described by Roy et al. (2005b) and Roy and Boschetti (2009). They selected 11 Landsat bi-temporal image pairs distributed across Southern Africa to cover approximately the range of plant water availability conditions, that they considered to be a regionally controlling factor on the type and amount of vegetation, and so indirectly on human population density and land use, that together influence the regional distribution of burned areas (Archibald et al., 2009). Only a limited number of Stage 3 validation exercises have been undertaken and they have implemented spatially stratified sampling designs (Boschetti et al., 2006; Padilla et al., 2014).

In this paper, a stratified probability sampling design that encompasses both the spatial and temporal domain is proposed to achieve Stage 3 and Stage 4 validation criteria. Importantly, sampling of independent reference data with respect to both the spatial and temporal domains is shown to be necessary for statistically rigorous inferences derived from sample-based validation of global burned area products. The sampling design is tailored to the appropriate collection of bi-temporal Landsat images because of their aforementioned use for global coarse resolution burned area product validation, and because the Landsat satellites provide the longest, freely available, satellite record that is continuing with the availability of Landsat 8 data (Roy et al., 2014). The sampling units are defined with respect to the Landsat World-wide Reference System (WRS-2) path/row coordinates and Landsat 16-day acquisition calendar (Arvidson et al., 2006). Regional strata are constructed using a global ecoregion map (Olson et al., 2001) and the global MODIS active fire product is used to stratify the population based on the level of fire activity in each sampling unit. The sampling grid and the protocol for constructing the strata are fully independent from the burned area product that is being validated, and are therefore applicable to any of the current moderate resolution burned area products. To evaluate the performance of the proposed sampling design the global MODIS burned area product is used as a hypothetical population (census) of reference data. The stratified probability sampling design could be adapted for application using other Landsat class global frequent repeat coverage moderate resolution data and using other ecoregion maps to construct regional strata. The evaluation of standard errors presented in the paper serves as an illustrative case study.

The paper is organized as follows. Section 2 describes the datasets used to demonstrate the proposed independent reference data sampling design. Section 3 describes the design, starting with a rigorous definition of the sampling elements for global burned area product validation, and presents formulae using the proposed sampling design for estimating (i) burned area product accuracy metrics and, (ii) total burned area estimates and associated standard errors. The formulae for optimal allocation of independent reference data that minimize the standard errors of (i) and (ii) using the design are presented. Section 4 demonstrates the proposed design by implementing the stratification using one year of the MODIS fire products, and computing the standard errors of the accuracy and area estimators using different sample allocation strategies. The efficiency of the proposed design is evaluated by computing the reduction of the standard errors compared to simple random sampling. The paper concludes with recommendations for future research and application.

2. Data

2.1. Fire datasets

In this study MODIS products for all the globe from January to December 2008 were used. The year 2008 was selected as it has comparable average total global area burned compared to other years (Giglio et al., 2013) and also was the same year used to validate a number of other global burned area products (Padilla et al., 2014, 2015).