Remote Sensing of Environment 205 (2018) 131–140

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A LandTrendr multispectral ensemble for forest disturbance detection

Warren B. Cohen[a,*], Zhiqiang Yang[b], Sean P. Healey[c], Robert E. Kennedy[d], Noel Gorelick[e]

[a] USDA Forest Service, PNW Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331, United States
[b] Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, United States
[c] USDA Forest Service, RMRS Research Station, 507 25th Street, Ogden, UT 84401, United States
[d] College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331, United States
[e] Google Switzerland GmbH, Zurich, CH 8002, Switzerland.

ARTICLE INFO

Keywords:
Forest disturbance
Landsat time series
Multispectral change detection
Disturbance signal-to-noise ratio
Stacking ensemble

ABSTRACT

Monitoring and classifying forest disturbance using Landsat time series has improved greatly over the past decade, with many new algorithms taking advantage of the high-quality, cost free data in the archive. Much of the innovation has been focused on use of sophisticated workflows that consist of a logical sequence of processes and rules, multiple statistical functions, and parameter sets that must be calibrated to accurately classify disturbance. For many algorithms, calibration has been local to areas of interest and the algorithm's classification performance has been good under those circumstances. When applied elsewhere, however, algorithm performance has suffered. An alternative strategy for calibration may be to use the locally tested parameter values in conjunction with a statistical approach (e.g., Random Forests; RF) to align algorithm classification with a reference disturbance dataset, a process we call secondary classification. We tested that strategy here using RF with LandTrendr, an algorithm that runs on one spectral band or index. Disturbance detection using secondary classification was spectral band- or index-dependent, with each spectral dimension providing some unique detections and different error rates. Using secondary classification, we tested whether an integrated multispectral LandTrendr ensemble, with various combinations of the six basic Landsat reflectance bands and seven common spectral indices, improves algorithm performance. Results indicated a substantial reduction in errors relative to secondary classification based on single bands/indices, revealing the importance of a multispectral approach to forest disturbance detection. To explain the importance of specific bands and spectral indices in the multispectral ensemble, we developed a disturbance signal-to-noise metric that clearly highlighted the value of shortwave-infrared reflectance, especially when paired with near-infrared reflectance.

1. Introduction

Employing Landsat time series for the characterization and mapping of forest disturbance has received considerable attention over the past decade (Hansen et al., 2013; Kim et al., 2014), since the opening of the image archive when data became freely available in a highly-calibrated format (Roy et al., 2014). Many newer algorithms have used all cloud-free observations, either directly or after data reduction to derive annual composites, before subjecting the time series of data to various sophisticated algorithm functions designed to detect disturbances (Hermosilla et al., 2015). For example, Brooks et al. (2014) identified abrupt disturbances with all available data using residuals from harmonic regression and statistical quality control charts, DeVries et al. (2015) used harmonic regression with a breakpoint seeking method called "moving sums", Kennedy et al. (2010) subjected annual composite time series to temporal segmentation with the goal of mapping both abrupt and gradual change, and Huang et al. (2010) highlighted spectral anomalies in moving multi-year windows to characterize disturbances. Prior to this new era of freely available, well-calibrated data, most applications of Landsat time series to map forest disturbance were limited to less dense time series (Cohen et al., 2002; Masek et al., 2008). These applications commonly relied on traditional statistical methods, such as post-classification map comparison, bi-temporal differencing, principal components analysis, and supervised classification (Coppin et al., 2004; Healey et al., 2005).

Denser time series data and more sophisticated approaches facilitate detection of subtler disturbance signals, which has led to a move away from an almost exclusive characterization of stand replacement disturbances (Healey et al., 2008; Wulder et al., 2004) towards the exploration of partial (i.e., non-stand replacement) disturbances

* Corresponding author.
E-mail addresses: wcohen@fs.fed.us (W.B. Cohen), zhiqiang.yang@oregonstate.edu (Z. Yang), seanhealey@fs.fed.us (S.P. Healey), rkennedy@coas.oregonstate.edu (R.E. Kennedy), gorelick@google.com (N. Gorelick).

https://doi.org/10.1016/j.rse.2017.11.015
Received 6 July 2017; Received in revised form 6 November 2017; Accepted 16 November 2017
0034-4257/ Published by Elsevier Inc.

associated with forest thinning, degradation, and insect and disease activity that unfolds over multiple years (Meigs et al., 2011; Meddens and Hicke, 2014; Cohen et al., 2016; Hughes et al., 2017). Exploring subtler signals within time series data has an attendant risk of false detection of change associated with noise, as indicated by time series studies from other disciplines (Trenberth, 1984; Pohmann et al., 2016). However, as forest management policy has shifted away from stand replacement harvests towards maintenance of healthy forest systems (Moeur et al., 2011), international agreements on forest monitoring have begun to include forest degradation along with deforestation (Kissinger et al., 2012), and recognition that climate change is making forests more vulnerable to mortality associated with increasing physiological stress (Allen et al., 2015; Mildrexler et al., 2016), there are now greater demands on remote sensing to provide a full range of detection capabilities from subtle to dramatic forest disturbance (McDowell et al., 2015).

When attempting to detect low magnitude disturbances with Landsat time series, the signal associated with spectral change due to disturbance may be masked by noise associated with normal temporal variation from imperfect atmospheric and geometric corrections, vegetation phenology, sun angle variations, and sensor degradation. In this regard, Kennedy et al. (2010) found that different spectral indices had varying abilities for accurate detection of subtler disturbance signals in western Oregon, with the normalized difference vegetation index (NDVI) performing less effectively than the normalized burn ratio (NBR) or Tasseled Cap Wetness (TCW). Because many of the newer algorithms employ a limited set of spectral bands or indices to detect disturbance, such as the NBR (Kennedy et al., 2012), Forestness Index (Huang et al., 2010), NDVI and SWIR-NIR (shortwave-infrared, near-infrared) ratio (Vogelmann et al., 2012), or Tasseled Cap Angle (TCA, Brooks et al., 2014), careful consideration of the comparative signal-to-noise (SNR) strengths among spectral indices is important. To address this need, we formalized the derivation of a disturbance SNR (DSNR) metric and used that to test the effectiveness of the six primary Landsat reflectance bands (i.e., TM/ETM + bands 1–5, 7) and a host of common spectral indices for detecting forest disturbance across a wide variety of forest types in the US.

Application of any given algorithm or approach for detecting disturbance requires one or more thresholds and calibration steps to separate disturbance signal from temporal noise. These are usually derived using statistical procedures, but also involve a great degree of heuristics. For example, Huang et al. (2010), Kennedy et al. (2010), Brooks et al. (2014), and Hughes et al. (2017) all describe the complexity of their unique Landsat-based forest disturbance detection algorithms, the in-depth rigorous steps involved in calibration for local conditions, and the hands-on assessments and related cautions regarding potential limits of the calibrated parameters in new forested systems. Given the effort involved to recalibrate complicated, but effective algorithms for new forest systems and conditions, a reasonable question to ask is: Could these algorithms be applied in new forest types or locations using well-tuned parameter sets from a limited set of localized applications, with an additional, bulk statistical calibration from a reference dataset and a secondary statistical classification approach such as Random Forest (RF, Breiman, 2001)? We test this idea of secondary classification here using the LandTrendr (Landsat-based detection of Trends in Disturbance and Recovery) algorithm (Kennedy et al., 2010).

LandTrendr runs on a single band or spectral index (Kennedy et al., 2012), which may unnecessarily limit its value as a forest disturbance detection algorithm. In a recent study (Cohen et al., 2017), multiple algorithms were run on a common Landsat dataset across six diverse forested areas in the US, with each algorithm using different spectral bands and indices. When the maps from those algorithms were compared against each other they were found to be quite different, suggesting that, at least in part, spectral bands/indices used was a factor in the differences among maps. If calibration of LandTrendr through a secondary classification model is effective, there would also be the opportunity to run the algorithm multiple times, each time using a different band or spectral index, before integrating the results from all runs as a multispectral ensemble using the secondary classification model.

The ensemble integration of maps from a variety of algorithms using RF was recently tested by Healey et al. (in press). In that study, empirical weights among an ensemble of map products were generated through a process called stacking (stacked generalization), in conjunction with reference data acquired through visual interpretation of Landsat time series data using a tool called TimeSync (Cohen et al., 2010). Disturbance mapping errors from the ensemble, relative to the individual maps from each algorithm, were greatly reduced when compared to the reference data. Healey et al. (in press) showed that adding informative, non-overlapping predictor information from different algorithms improved ensemble change detection performance. In this study, we tested the idea that valuable, non-overlapping information can be generated from a single algorithm operating on different parts of the electromagnetic spectrum. Specifically, we tested the stacking ensemble approach using LandTrendr and a combination of the six primary Landsat reflectance bands plus seven commonly used vegetation indices. This is similar to an approach used by Schultz et al. (2016), where maps from different indices derived from the BFAST Monitor algorithm (DeVries et al., 2015) were fused to create a single, improved map of deforestation in the tropics.

Three main objectives were addressed in this study:

- Quantify distributions of forest DSNR values for the original Landsat spectral bands and selected spectral indices, and the relationship between DSNR and disturbance detection error rates;
- Test secondary classification of LandTrendr when run on a single band/index using RF, and determine if there is a relationship between classification error rates and DSNR values; and
- Combine single band/index outputs of LandTrendr in a RF stacking ensemble to understand the power of secondary classification in a multispectral context, and evaluate the complementarity among bands/indices for forest disturbance detection.

2. Methods

2.1. The disturbance signal-to-noise ratio (DSNR) metric

To calculate the DSNR we used a TimeSync (Cohen et al., 2010) reference dataset that was collected from 1800 single pixel-sized plots (300 randomly selected per scene) over six, largely forested Landsat scenes widely dispersed across the conterminous US (Cohen et al., 2017). The forests consisted largely of a variety of needleleaf evergreen and broadleaf deciduous tree species commonly found across the different forested regions of the US (see Table 2, Cohen et al., 2017 for details). Of the 1800 plots, 1303 were forested, as determined by visual interpretation of high spatial resolution images in Google Earth. The Landsat time series data for each plot was temporally segmented by human interpretation using the TimeSync tool, which integrates simultaneous viewing of an annual series of full resolution Landsat image chips, temporal trajectories for each plot in a variety of spectral bands and indices, and the high-resolution images within Google Earth centered on the plot. To temporally segment the time series from 1984 to 2013 for a given plot, multiple spectral bands and indices (along with other tools) and breaks in the trends of spectral values were evaluated and identified (see Fig. 3 in both Cohen et al., 2010 and Kennedy et al., 2010). Using TimeSync, each segment was assigned a label represented by three types of observed forest processes – disturbance, growth, and stable – based on expert opinion. By definition, each segment was at least one year in length and bounded by two break points (start vertex and end vertex). For single segment plots the start vertex was 1984 and end vertex was 2013, whereas for multiple segment plots there were