ISPRS Journal of Photogrammetry and Remote Sensing 137 (2018) 29–46

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Harmonic regression of Landsat time series for modeling attributes from national forest inventory data

Barry T. Wilson[a,*], Joseph F. Knight[b], Ronald E. McRoberts[a]

[a] Northern Research Station, USDA Forest Service, 1992 Folwell Avenue, St. Paul, MN 55108 USA
[b] Department of Forest Resources, University of Minnesota, 1530 Cleveland Avenue N., St. Paul, MN 55108 USA

ARTICLE INFO

Article history:
Received 7 September 2017
Received in revised form 2 January 2018
Accepted 7 January 2018

Keywords:
Landsat time series
Image compositing
Harmonic regression
National forest inventory
Regression models
Classification models

ABSTRACT

Imagery from the Landsat Program has been used frequently as a source of auxiliary data for modeling land cover, as well as a variety of attributes associated with tree cover. With ready access to all scenes in the archive since 2008 due to the USGS Landsat Data Policy, new approaches to deriving such auxiliary data from dense Landsat time series are required. Several methods have previously been developed for use with finer temporal resolution imagery (e.g. AVHRR and MODIS), including image compositing and harmonic regression using Fourier series. The manuscript presents a study, using Minnesota, USA during the years 2009–2013 as the study area and timeframe. The study examined the relative predictive power of land cover models, in particular those related to tree cover, using predictor variables based solely on composite imagery versus those using estimated harmonic regression coefficients. The study used two common non-parametric modeling approaches (i.e. *k*-nearest neighbors and random forests) for fitting classification and regression models of multiple attributes measured on USFS Forest Inventory and Analysis plots using all available Landsat imagery for the study area and timeframe. The estimated Fourier coefficients developed by harmonic regression of tasseled cap transformation time series data were shown to be correlated with land cover, including tree cover. Regression models using estimated Fourier coefficients as predictor variables showed a two- to threefold increase in explained variance for a small set of continuous response variables, relative to comparable models using monthly image composites. Similarly, the overall accuracies of classification models using the estimated Fourier coefficients were approximately 10–20 percentage points higher than the models using the image composites, with corresponding individual class accuracies between six and 45 percentage points higher.

1. Introduction

The use of remotely sensed data for inventory and mapping of agricultural and natural resources has a long history. Such data have typically been collected on airborne or spaceborne platforms by either passive sensors that use solar radiation as a source of electromagnetic energy, or active sensors that provide their own radiation source. Both passive and active sensors have demonstrated utility as sources of predictor variables that can be used to model a variety of natural resource phenomena.

In the context of using these data in combination with national forest inventory (NFI) data for the purposes of monitoring forest resources over long periods of time and large geographic areas, the Landsat platform has garnered particular interest, as noted by the review of McRoberts et al. (2010a). This is due to several factors, which will be discussed in greater detail below. First, the Landsat Program provides the longest-running continuous collection of Earth imagery of any satellite program, with sensors designed to maintain consistency in resolution (i.e. spatial, spectral, radiometric, and temporal) across missions. Second, the multispectral characteristics of the Landsat sensors enable the derivation of metrics with biophysical meaning. Third, the spatial resolution of the latter generation of sensors provides a better match to the size of the typical NFI plot than that of satellite sensors with coarser spatial resolution. Finally, with the adoption of the open access data policy for the Landsat Program in 2008, the entire archive of Landsat imagery is freely available for use.

1.1. Landsat as a source of auxiliary data

1.1.1. Length of the Landsat record

From the Multispectral Scanner (MSS) sensor used onboard the first five Landsat satellites to the Thematic Mapper (TM) sensor of

* Corresponding author.
E-mail address: barrywilson@fs.fed.us (B.T. Wilson).

https://doi.org/10.1016/j.isprsjprs.2018.01.006
0924-2716/Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS).

Landsat-4 and Landsat-5, the Landsat program has striven for consistency in the data record while simultaneously incorporating improved technological capabilities in terms of spatial and spectral resolution. This consistency extends to the sensors used onboard the Landsat satellites in orbit today, as well as those planned for future missions.

One of the two Landsat sensors currently in operation is the Landsat-7 Enhanced Thematic Mapper Plus (ETM+). In May of 2003, the Landsat-7 scan line corrector (SLC), a small rotating mirror that compensates for the forward motion of the spacecraft, failed. This resulted in the loss of data in wedge-shaped areas on either side of the image, with more missing data further away from nadir. These SLC-off gaps amount to a loss of approximately 22% of the data for any given scene (Ju and Roy, 2008). These gaps have limited the utility of ETM+ imagery, although there have been several approaches proposed to dealing with them, including compositing of several images (Roy et al., 2010), interpolation using SLC-on imagery (Maxwell et al., 2007; Chen et al., 2011), data fusion with MODIS imagery (Roy et al., 2008), and geostatistical methods (Zhang et al., 2007; Pringle et al., 2009).

1.1.2. Development of Landsat spectral indices

Multiple spectral indices have been developed to establish the relationship between the spectral and radiometric response measured by remote sensors and the presence of various land covers, especially vegetation. Bannari et al. (1995) reviewed over forty vegetation indices that have been developed for sensors ranging from ground-based to spaceborne systems. Most such indices are based on the fact that vegetation has wavelength-dependent absorption, transmission, and reflection properties, in particular the differential response in the red and near-infrared (NIR) wavelengths, and have been shown to provide a better indication of the amount of vegetative land cover than any single band alone (Curran, 1980).

Kauth and Thomas (1976) developed a linear transformation of all four of the original MSS bands, named the tasseled cap transformation (TCT), to produce indices related to not only growing vegetation ('green stuff'), but also soils ('brightness'), senescent vegetation ('yellow stuff'), and shadows ('non-such') to better differentiate various crops, as well as phases of crop development over time. Crist and Cicone (1984a, 1984b) extended TCT for use with the six reflective TM bands, simultaneously dropping some features ('yellow stuff' and 'non-such') while defining new ones (e.g. 'wetness'). Huang et al. (2002) and Baig et al. (2014) developed comparable transformations for ETM+ and the Operational Land Imager (OLI) respectively.

Numerous studies have shown TCT components derived from Landsat imagery to be useful for mapping forest characteristics. These studies, using the components of brightness, greenness, and/or wetness, demonstrate their utility across a range of forest mapping applications, including land cover (Byrne et al., 1980; Yuan et al., 2005), forest types (Dymond et al., 2002), succession (Helmer et al., 2000), stand-replacing disturbance (Cohen et al., 1998; Jin and Sader, 2005; Healey et al., 2005), pest damage (Skakun et al., 2003), growing stock volume (Zheng et al., 2014), canopy cover and biomass (Karlson et al., 2015), and recovery from disturbance (Pickell et al., 2016).

1.1.3. Size and configuration of the NFI plot footprint

The current annual NFI conducted by United States Forest Service Forest Inventory and Analysis (FIA) program exhibits many of the characteristics observed for NFI globally that are pertinent to its use with Landsat imagery. A comprehensive description of the FIA program is provided in Bechtold and Patterson (2005). FIA inventory plots are established according to a well-defined sample design, with a sampling frame that is used to generate a spatially-balanced sample of plots. Each plot is a cluster of sub-plots. Cluster plots are often used for NFI because the determination of the relative locations of the sub-plots is typically more accurate, their layout is generally faster due to the smaller distances measured and traveled (possibly over rugged terrain), and they capture more variability in the population than one larger plot (Kangas and Maltamo, 2006).

Since the nationwide start of the annual NFI program in the US in 1998, FIA plots comprise four circular sub-plots, each with a radius of 7.3152 m. Circular plots are often used in forest inventory because they are easy to establish for small radii and are prone to less error in plot area, for the same reasons given previously for using cluster plots (Kangas and Maltamo, 2006). Circular plots also minimize edge effects, since they have the smallest possible perimeter for a given area by the isoperimetric inequality. The sub-plots are arranged with the center of one sub-plot defining the center of the cluster plot. The centers of each of the other three sub-plots are equally spaced about plot center, oriented so one sub-plot is due north of plot center, and each is 36.576 m distant from plot center.

Although other NFI programs have different sample designs, sampling frames, and plot configurations, the information presented for the FIA program remains instructive for comparisons to Landsat and other satellite imagery. Each FIA sub-plot constitutes an area of about 168 m^2, approximately 19% the area of a TM, ETM+, or OLI pixel for the reflective bands, but only about 0.27% the area of a 250-m MODIS pixel. The smallest circle that circumscribes all four sub-plots has an area of about 6052 m^2, or just under seven 30-m pixels. The four sub-plots cover approximately 11% of this area, and about 1% of the area of a MODIS pixel.

1.1.4. Open data access policy for the Landsat archive

Ease of access to data from the Landsat Program has been variable over time. During the commercial operations period, there were high financial costs and restrictive copyright rules in place that limited sharing of access to imagery. With the assumption of mission operations by the United States Geological Survey (USGS), purchased imagery could be shared more freely. Wulder et al. (2012) suggest that the USGS Landsat Data Policy that took effect in 2008 has allowed the scientific community to finally realize the full value of the Landsat Program, as indicated by the dramatic rise in the use of its data globally. This data policy provides unrestricted access to the entire USGS National Satellite Land Remote Sensing Data Archive (NSLRSDA), with selected products made available for retrieval over the Internet at no financial cost to users (Woodcock et al., 2008).

1.2. Approaches to analyzing dense time series of satellite imagery

Free and open access to the Landsat Archive (i.e. NSLRSDA) permits new approaches to image analysis that were not previously available to users of Landsat imagery. With the high costs of Landsat scenes under earlier data policies, users typically selected only those scenes of interest that were predominantly free of clouds, exhibiting a 'scene-centric' focus. Under the new data policy, as well as due to continuing advancements in data storage and computing power, users have begun developing methods that utilize all scenes over a period of interest to improve the information collected for pixels of interest, shifting to a 'pixel-centric' focus.

There are several approaches to processing and analyzing dense time series of satellite imagery. Most were developed for use with other satellite platforms, particularly those having finer temporal resolution than Landsat. One of the first was published by Goward et al. (1985) for analyzing daily observations of NDVI across North America using the Advanced Very High Resolution Radiometer (AVHRR) onboard the NOAA-7 satellite. The daily data