Remote Sensing of Environment 210 (2018) 473–489

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A spatial ensemble approach for broad-area mapping of land surface properties

Sam Hooper*, Robert E. Kennedy

College of Earth, Ocean, and Atmospheric Sciences, 104 CEOAS Admin Bldg., Oregon State University, Corvallis, OR 97331, United States

ARTICLE INFO

Keywords:
Broad-area mapping
Time series
Spatial scale
Forest canopy cover
Impervious surface cover
Regression trees

ABSTRACT

Understanding rapid global change requires land cover maps with broad spatial extent, but also fine spatial and temporal resolution. Developing such maps presents a unique challenge, as variability in relationships between spectral characteristics (i.e., predictors) and a response variable is likely to increase with the size of the region across which a model is built and applied. Although most mapping approaches apply the same predictor-response relationships globally across the entire modeling region, learned relationships from one local area may be invalid for another when predicting across broad extents. Here, we adapted a spatial ensemble approach borrowed from species distribution modeling to land cover mapping, and evaluated whether the approach could faithfully represent spatial variation in relationships between land cover and spectral data. The spatiotemporal exploratory model (STEM) uses an ensemble of regression trees defined within spatially overlapping support sets, producing a broad-extent map that reflects variability at the spatial scale of each constituent support set. As test cases for reference maps, we used 30-m-resolution forest canopy and impervious surface cover layers from the 2001 U.S. National Land Cover Database (NLCD) for the states of Washington, Oregon, and California. When testing strategies for support set size and sampling intensity, we found that predictor-response relationships were strongest when individual components of the spatial ensemble were small and when sampling intensity was high. Compared to aspatial bagged decision tree and random forest models, we found that the STEM approach successfully captured variation in our source maps, both globally and at scales smaller than the modeling region. Leveraging the spatial structure of a STEM, we also mapped per-pixel spatial variation in prediction confidence and the importance of different predictor variables. After testing appropriate spatial ensemble and sampling strategies, we extended the predictor-response relationships gleaned from the 2001 source maps into a yearly time series based on temporally-smoothed spectral data from the LandTrendr algorithm. The end products were yearly forest canopy and impervious surface cover time series representing 1990–2012. Formal evaluation showed that our temporally extended maps also closely resembled NLCD maps from 2011. The aim of this research was to cultivate the implicit relationships between spectral data and a given map, not improve them, but as the need for time series maps produced at both broad extents and fine resolutions increases, our results demonstrate that an ensemble of locally defined estimators is potentially more appropriate than conventional ensemble models for land cover mapping across broad extents.

1. Introduction

Regional and continental scale land cover maps produced from satellite remote sensing are a key tool in understanding impacts of global change (Gray and Song, 2013; White et al., 2014; Wang et al., 2015). Because land cover condition and land cover change often operate at fine spatial scales (Sohl et al., 2004), moderate resolution sensors such as Landsat play a central role in mapping efforts (Cohen and Goward, 2004). Although data cost and computational burden long constrained such mapping to local and regional extents, the era of free, consistently-processed Landsat data has led to an increased ability to bring moderate-scale mapping to the broad spatial extents needed to understand global change (Wulder et al., 2012). Previously unattainable global-extent, Landsat-resolution maps of key surface properties are rapidly emerging (Hansen et al., 2013; Kim et al., 2014; Pekel et al., 2016). Moreover, many new time-series algorithms have emerged to tap the temporal richness of the Landsat archive (Zhu, 2017), and computational advances (Gorelick et al., 2017) allow extension of these

* Corresponding author.
E-mail address: hoopersa@oregonstate.edu (S. Hooper).

https://doi.org/10.1016/j.rse.2018.03.032
Received 8 March 2017; Received in revised form 19 March 2018; Accepted 20 March 2018

algorithms to arbitrarily large areas.

These new time-series approaches address a key challenge in mapping land cover change. Making maps of any surface property from remote sensing requires mathematical modeling of the linkage between that property and the spectral space of the remotely sensed data. Random variation in spectral space caused by inexact atmospheric correction and by vagaries of image acquisition date could lead to spurious signals of change (Kennedy et al., 2009). Time-series algorithms rely on models that largely remove the random noise (Zhu and Woodcock, 2014; Kennedy et al., 2010), allowing development of a temporally stabilized spectral data space. In principle, models between that data space and a map developed at one point in time can be applied to any other point in time, allowing large-area maps from a single era to be extended backward and forward in time to better understand change (Bartz et al., 2015; Kennedy et al., 2018; Maclaurin and Leyk, 2016).

Despite these advances in the temporal domain, there remains an important challenge in the spatial domain. Global-scale science benefits from spatial and temporal consistency in mapped surface properties, but the mathematical model linking a response variable and its predictors varies among different ecosystems (Sohl et al., 2004). As the extent of a "modeling region" (the geographic area over which that model is built and applied) increases, spatially monolithic models become less appropriate and prediction accuracy can decrease (Fink et al., 2010). To capture local-scale relationships over broad areas, institutional mapping efforts have developed continent-scale mosaics of maps produced within semi-independent, ecoregion-sized modeling regions where local-scale relationships and expert judgment are invoked to improve overall map quality (Gallant et al., 2004; Homer et al., 2004; Kellndorfer et al., 2013). While such an approach can result in high quality, single-date, large-area maps (Homer et al., 2007), the underlying relationships are spatially variable and potentially unknowable. To extract change information from any given large-area map, we must develop mapping strategies that flexibly handle spatial variability in predictor-response relationships without the need for expert intervention.

A candidate algorithm comes from the literature of species distribution modeling. Originally developed to model bird species presence from citizen science data, the spatiotemporal exploratory model (STEM) is a framework to generate predictions at continental scales (Fink et al., 2010). A STEM aggregates predictions from an ensemble of locally defined, spatially overlapping estimators (i.e., base models). Each pixel in the final prediction map is the average or mode of all overlapping pixels from each estimator prediction, producing a broad-extent map that reflects variability at the spatial scale of each constituent estimator. Although the main goal of Fink et al. (2010) was to predict species distributions with intra-annual spatiotemporal variability, the spatial ensemble approach is a promising strategy to map any phenomenon where predictor-response relationships may vary over space.

The objective of our study was to adapt and evaluate STEM for land cover mapping to produce a broad-extent time series. To simplify evaluation of the method, we confined our efforts to the goal of replicating existing land cover maps rather than creating entirely new maps. Using spectral data produced from time-series segmentation algorithms as predictors, we first attempted to reproduce broad-extent maps of land cover in a single year. As source maps, we chose continuous-field products from the National Land Cover Database (NLCD) that were mosaics of ecoregion-scale mapping products (Homer et al., 2004). Using a suite of testing tools, we evaluated STEM's ability to replicate the spatial and distributional patterns in the original maps. After testing appropriate model parameters, we then extended the predictor-response relationships gleaned from each map into a time series, and evaluated those spatial patterns against separately produced maps from the same database. If successful, this combination of algorithm and predictor dataset could provide a cost-effective approach to producing continental-scale time series from a single map.

2. Materials and methods

2.1. Modeling region description

We conducted this study using data from the west coast of the conterminous United States, covering all of California, Oregon, and Washington. Elevations across the modeling region range from sea level to 4400 m. Both Oregon and Washington are broadly characterized by two vegetation zones, forest and steppe (Franklin and Dyrness, 1988), generally corresponding to maritime and continental climatic zones, respectively. Vegetation zones are similarly divided by the southern Cascades and Sierra mountains in California, although more xeric vegetation communities are common in southern California with coastal chaparral in the west (Ashbaugh and Alwin, 1994, 110) and desert scrublands throughout much of southeastern California (Miller and Hyslop, 1983). Overall, the entire region exhibits a wide variety of forest and vegetation types. Agricultural land dominates much of the low-elevation central valleys throughout the modeling region, with a wide variety of crop types, including cotton, wheat, vegetables, feed crops, and tropical evergreens (Miller and Hyslop, 1983). Most urban development throughout Oregon and Washington is relegated to major urban centers and highway travel corridors (Ashbaugh and Alwin, 1994, 382). The same general pattern is found in California, although development intensity and extent are generally greater across the state. The ecological and population density diversity makes the modeling region ideal for testing this ensemble model, designed to handle globally variable relationships between predictor and response variables.

2.2. Training and reference data

To assess the STEM algorithm, we chose as training maps continuous-field products from NLCD: forest canopy cover (FCC) and impervious surface cover (ISC) for the nominal year 2001. As continuous metrics of land cover essential for understanding drivers of and responses to important landscape change processes (Hansen et al., 2013; Theobald et al., 2009), these two products show markedly different spatial patterns and spectral relationships, providing bookend-type tests of the STEM approach. Although methods used to create these source maps are well documented elsewhere, several points relevant to our testing are notable. Most importantly, the maps were constructed within predefined mapping zones, generally corresponding to ecoregion boundaries (Homer et al., 2004). Estimates of FCC and ISC percent cover (ranging from 0 to 100%) were interpreted on high-resolution orthophotos and related to Landsat composites of imagery dating from 1999 to 2002. Reported accuracy from ten-fold cross-validation for each modeling region within our study area ranged from 79 to 91% for FCC and 83 to 93% for ISC (Homer et al., 2004). While independent assessments have revealed errors in NLCD FCC and ISC data (Greenfield et al., 2009; Nowak and Greenfield, 2010), NLCD maps are considered to be high quality products and are widely used in ecological modeling (e.g., Fink et al., 2010; Pidgeon et al., 2007; Theobald et al., 2009;