

TECHNICAL ADVANCE

Shape selection in Landsat time series: a tool for monitoring forest dynamics

GRETCHEN G. MOISEN¹, MARY C. MEYER², TODD A. SCHROEDER^{1,3}, XIYUE LIAO², KAREN G. SCHLEEWEIS¹, ELIZABETH A. FREEMAN¹ and CHRIS TONEY¹

¹Rocky Mountain Research Station, US Forest Service, 507 25th Street, Ogden, UT 84401, USA, ²Department of Statistics, Colorado State University, 212 Statistics Building, Fort Collins, CO 80523-1877, USA, ³ASRC Federal InuTeq, U.S. Geological Survey, Earth Resources Observation and Science Center, Sioux Falls, SD 57198, USA

Abstract

We present a new methodology for fitting nonparametric shape-restricted regression splines to time series of Landsat imagery for the purpose of modeling, mapping, and monitoring annual forest disturbance dynamics over nearly three decades. For each pixel and spectral band or index of choice in temporal Landsat data, our method delivers a smoothed rendition of the trajectory constrained to behave in an ecologically sensible manner, reflecting one of seven possible ‘shapes’. It also provides parameters summarizing the patterns of each change including year of onset, duration, magnitude, and pre- and postchange rates of growth or recovery. Through a case study featuring fire, harvest, and bark beetle outbreak, we illustrate how resultant fitted values and parameters can be fed into empirical models to map disturbance causal agent and tree canopy cover changes coincident with disturbance events through time. We provide our code in the R package ShapeSelectForest on the Comprehensive R Archival Network and describe our computational approaches for running the method over large geographic areas. We also discuss how this methodology is currently being used for forest disturbance and attribute mapping across the conterminous United States.

Keywords: attribution, canopy change activities, change agents, forest disturbance, landcover change, R package, regression splines, tree canopy cover

Received 16 January 2016; revised version received 15 April 2016 and accepted 28 April 2016

Introduction

Understanding trends in forest disturbance and their effects on forest parameters such as tree canopy cover and biomass is important for carbon assessments, as well as for forest management decisions and scientific investigations across the globe. Data from the Landsat suite of remote sensing satellites offer a historically robust collection of earth observations which can be used to understand forest dynamics at a variety of spatial and temporal scales. Collected across many parts of the globe at least every 16 days, Landsat offers both the spatial (30 m pixels) and spectral capabilities (seven reflectance bands covering the visible, near-infrared, and shortwave-infrared wavelengths) necessary to effectively study vegetation disturbance and recovery dynamics (Goward *et al.*, 2008; Cohen *et al.*, 2016). Starting in 2008, the longest continuous record of medium resolution (<50 m) satellite images became freely available to the scientific community (Woodcock *et al.*, 2008). This free, temporally dense data from Landsat

(and other new Landsat-like sensors, e.g., Sentinel 2A, Drusch *et al.*, 2012) opens the door for dramatic new analyses of land use land cover change that are affecting this planet (Wulder *et al.*, 2012; Kennedy *et al.*, 2014).

Tremendous progress has been made to process and analyze forest dynamics using time series of Landsat imagery. Recent work uses annual observations of Landsat bands or indices to identify anomalies or disturbances in normal growth patterns of forests, either through fitting the trajectories directly, or through the application of thresholding approaches. Some widely used algorithms include the following: the Vegetation Change Tracker (VCT) (Huang *et al.*, 2010a), Vegetation Continuous Fields (VCF) (Potapov *et al.*, 2012), Land-Trendr (Kennedy *et al.*, 2010), Continuous Change Detection and Classification (CCDC) (Zhu *et al.*, 2012), Multi-Index Integrated Change Algorithm (MIICA) (Fry *et al.*, 2011), a Fourier regression algorithm (Brooks *et al.*, 2014), and a gradual ecosystem change algorithm (Vogelmann *et al.*, 2012).

Here we present a new method based on the non-parametric statistical literature (Meyer, 2008, 2013b) to objectively fit Landsat trajectories developed with four

Correspondence: Gretchen Moisen, tel. +1 801 625 5384, fax: +1 801 625 5723, e-mail: gmoisen@fs.fed.us

different spectral indices. The method was developed for the conterminous United States as part of the North American Forest Dynamic project (Goward *et al.*, 2008; Masek *et al.*, 2013). Through exploration of data collected in diverse pilot scenes, seven possible ‘shapes’ were identified to represent common spectral responses to different disturbances occurring in various forest ecosystems throughout the country. The algorithm picks the optimal shape based on goodness of fit and a penalty for model complexity. Output from this algorithm is not intended as a final change detection map. Rather, the ‘shapes’ and their associated metrics can be used as predictor variables for modeling and mapping a large suite of both continuous and discrete forest attributes, as well as ancillary data to improve precision in estimates of forest attributes through time.

In this manuscript, we outline the preprocessing steps for developing the Landsat trajectories, give details of the shape selection algorithm, then provide a simple case study to illustrate how the shape selection output can be used to map forest disturbance agents (discrete) and tree canopy cover (continuous) through time in a Landsat scene located in the central Rocky Mountains of Colorado, USA. We also discuss how this methodology is currently being applied for forest disturbance and attribute mapping across the conterminous United States.

Materials and methods

Landsat data

Preparing Landsat time series for input into the shapes algorithm requires four steps. First, a near cloud free image is selected during peak growing season for each year in the time series (1984–2010 in this case study). Second, the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek *et al.*, 2006) is run to atmospherically correct the images and convert them from digital numbers to earth’s surface reflectance. Third, a clear view compositing algorithm is run to replace any pixels with bad data, cloud, or shadow with clear pixels from peak green growing season imagery within the same year (Huang *et al.*, 2010b). Pixel values in image years where a clear view is not obtainable are linearly interpolated from their nearest temporal neighbors in the Landsat time series. Fourth, Band 5 (B5; shortwave infrared) is used directly, along with three vegetation indices including: the forestness index (FI) (Huang *et al.*, 2008, 2009) derived as an integrated z-score from the visible and short-wave IR bands, the normalized burn ratio (NBR) (Key & Benson, 2005) derived from the near-infrared and short-wave bands, and the normalized difference vegetation index (NDVI) (Rouse *et al.*, 1973) derived from the visible and near-infrared bands.

We then run the shape selection algorithm on each pixel’s spectral trajectory over the 26-year time series for each of the four chosen indices. Using multiple indices helps improve

detection of the wide range of disturbance agents which occur in forested ecosystems (Schroeder *et al.*, 2011). For example, short-wave infrared data (e.g., B5, FI, and NBR) are sensitive to changes in leaf moisture content and shadowing, and near-infrared vegetation indices (e.g., NDVI and NBR) are sensitive to changes in plant vigor and canopy density. In order for these indices to respond to disturbance in a similar numerical direction, NBR and NDVI were multiplied by -1 and increased by a constant large enough to ensure all values were positive. The shapes algorithm can be applied to any other metrics whose temporal patterns are described by the seven possible shapes, such as tasseled cap indices (Crist & Cicone, 1984).

Shape selection

We have a scatterplot of Landsat band or index measurements y_i against times t_i , for $i = 1, \dots, n$, where it is assumed that the measurements are signal plus noise, that is,

$$y_i = f(t_i) + \varepsilon_i,$$

where f is the trend over time and ε_i is random error. We assume that the trend functions represent various possible phenomena captured by Landsat trajectories (based on B5, FI, -NDVI, and -NBR measures) on a single pixel through time.

Seven trend patterns (‘shapes’) were identified to reflect the behavior of these forested Landsat pixel trajectories under a range of disturbance scenarios (Fig. 1). A *flat* shape indicates a forest in a relatively stable condition. A decreasing (*decr*) shape indicates a forest accumulating biomass, potentially in a recovery or young growth stage. A sudden *jump* is indicative of a change in forest canopy or structure that happens over a relatively short period of time, typical of a harvest or fire pattern. A double jump (*2jump*) is indicative of two distinct disturbance events, such as two harvests over short rotation periods, or a fire followed by a salvage harvest. A *vee* indicates a forest that is, at first, growing or stable, but then encounters a slow disturbance mechanism that results in a gradual increase in reflectance or values of the vegetation indices common in forest cover loss. Conversely, an inverted vee (*inv*) illustrates a spectral signature of a forest in gradual decline, followed by a gradual recovery. Finally, an increase (*incr*) is a linear pattern that reflects instances where a disturbance may have occurred very early in (or before) the time series of imagery began, then shows signs of increasing spectral bands/indices related to slowly declining canopy. Because of the random errors, the measurements will not exactly follow any of these trend functions, and the goal of our shapes algorithm is to determine which of the shapes best describes the underlying phenomenon in each trajectory. Below we describe details on how we fit all shapes to each scatterplot (bands/index values through time on a single pixel) and apply an information criterion to choose which shape fits best.

For all shapes, we assume the trend is ‘smooth’, that is, the first derivative is continuous (except for jumps within the *jump* and *2jump* shapes). Quadratic regression spline functions are flexible smooth curves appropriate for estimating these trends. These are linear combinations of the smooth basis functions in Fig. 2a, with knots ξ_1, \dots, ξ_k marked as x

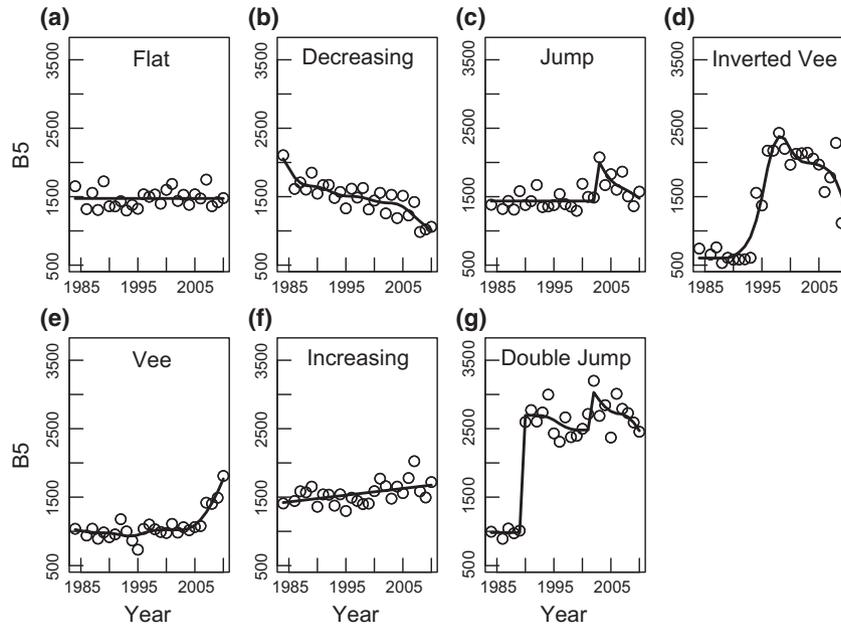


Fig. 1 Seven possible ‘shapes’ describing temporal patterns in Landsat bands and indices. Flat and decreasing shapes are often associated with stable or growing forest conditions. Jumps and inverted vee’s often reflect a rapid reduction in forest canopy due to events like harvest and fire. Vee’s and increasing shapes often capture slow onset disturbances such as insect and disease or drought. A double jump enables capture of two disturbance events in one trajectory.

(the times t_i are scaled to be in $[0,1]$). Between each pair of knots, each basis function is a piece of a parabola, and these parabolas are pieced together smoothly so that the first derivative is continuous and piecewise linear (Fig. 2a).

Let $\delta_j(t), j = 1, \dots, m$ be the basis functions (m is the number k of knots, plus one). Then, a ‘spline function’ with these knots is

$$f(t) = \sum_{j=1}^m b_j \delta_j(t),$$

and this is guaranteed to be smooth because its components are. Now given our scatterplot, we can find $b = (b_1, \dots, b_m)$ to minimize

$$\text{SSE} = \sum_{i=1}^n \left[y_i - \sum_{j=1}^m b_j \delta_j(t_i) \right]^2$$

This is the unconstrained spline fit and is shown as the blue curve on the scatterplot of observations in Fig. 2b. The

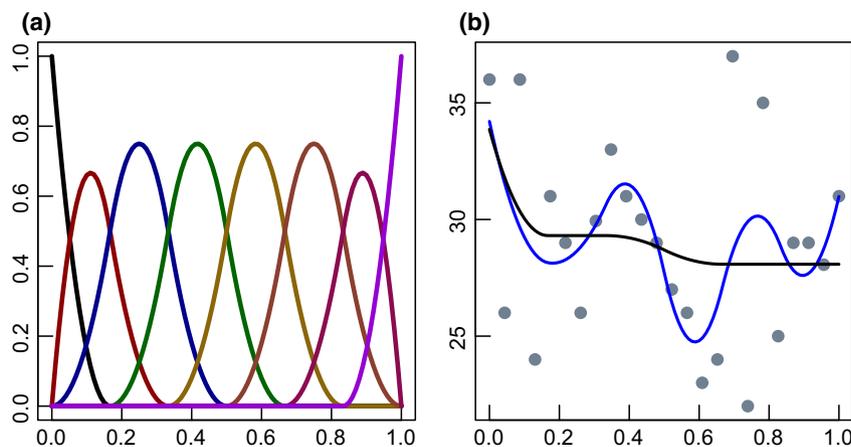


Fig. 2 The building blocks for many nonparametric fits to scatterplots are spline basis functions, shown on the left in (a). An unconstrained regression spline will piece together basis functions to form a very erratic fit [blue line on the right in (b)]. Newly developed constrained regression splines (Meyer 2013) force the fits to behave in a certain way. The black line on the right is an example of a regression spline constrained to always be decreasing.

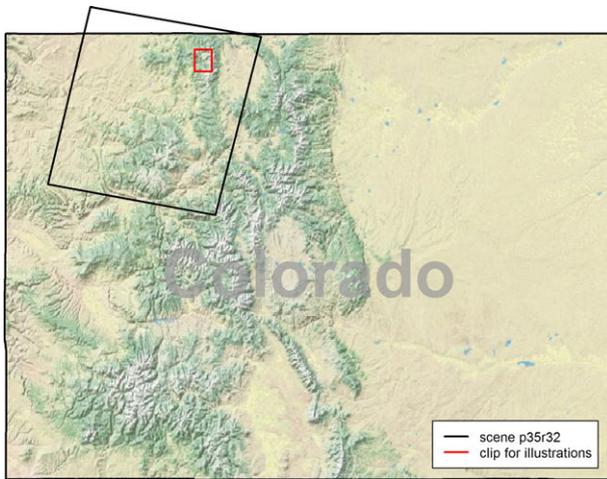


Fig. 3 Landsat path 32 row 35, shown in black, is located in the mountains of northern Colorado near Steamboat Springs. The red box indicates a small clip of this scene used in several of the mapping illustrations.

unconstrained fit is obtained by the usual least squares procedures, that is,

$$\hat{b} = B(B'B)^{-1}B'y,$$

where the matrix B contains the spline basis functions evaluated at the time points, that is $B_{ij} = \delta_j(t_i)$. If the underlying trend is actually decreasing, the fit can also be constrained to be decreasing; this is shown as the black curve in Fig. 2b.

Because the derivative of a piecewise quadratic function is piecewise linear, if we constrain the slope of the spline function to be negative at the knots, then the spline function will be decreasing over the entire range. Therefore, we define a $k \times m$ matrix S of slopes at the knots, that is, $S_{ij} = \delta_j(\xi_i)$ where ξ_1, \dots, ξ_k are the knots. The slope of the spline function at the i th knot is $\sum_{j=1}^m b_j \delta_j'(\xi_i)$, so a vector of slopes at the knots is Sb .

To obtain the constrained least squares fit, we find $b \in \mathbb{R}^m$ to minimize the SSE in (1) subject to the linear inequality constraints $Sb \leq 0$. This is a quadratic programming problem that is easily solved with the R package `CONEPROJ` (Liao & Meyer, 2014). To ensure computational efficiency, the main routines in `coneproj` were written and compiled in C++ and called using the package `RCPP`, which allows seamless integration of R and C++.

We can constrain the fit to be decreasing on $[0, c)$ for some $c \in (0, 1)$, and subsequently increasing on $(c, 1]$ (the *vee* shape) by swapping the sign of the last rows of S , corresponding to the knots to the right of c . The *inv* shape is increasing and then decreasing. The *jump* shape is decreasing, then has a discontinuity or jump, then is decreasing again. We can again define a matrix to constrain the spline coefficients b to produce such a jump. The *2jump* shape follows the decreasing patterns from the end of the jump shape with a second discontinuity, again ending with a decreasing trend. The best ‘change point’, or point of discontinuity, has to be estimated for the *vee*, *inv*, and *jump*. To do this, we fit the scatterplot $k - 1$ times with c in between the observed knot points, and we choose the change point that minimizes the SSE over all fits. Two change points have to be estimated for the *2jump* shape which involves fitting the scatter plot $n-3$ choose two times, requiring change points to be at least 2 years apart.

For each trajectory, the algorithm fits all seven shapes, finding the best change point(s) for the *vee*, *inv*, *jump*, and *2jump* shapes. The information criterion used to choose between the fits is the SSE penalized by adding a measure of model complexity that is a function of the degrees of freedom of the model. For the unconstrained spline, the degrees of freedom are the number of spline basis functions; the eight degrees of freedom in the above example make the fit quite flexible; in fact, it ‘over-fits’ by following the wiggles in the scatterplot caused by the error rather than the underlying trend. The decreasing fit sets some of the slopes at the knots to zero, thus ‘using’ fewer degrees of freedom. Unlike in the unconstrained case, the degrees of freedom for the constrained models are a random variable (Meyer, 2008), taking integer values that can be as small as zero or as large as the number of basis functions. We use the null expected degrees of freedom as a measure of model complexity. For a given shape, this is computed by generating many data sets from a trend, and taking the average of the used degrees of freedom of the fits. The null expected degrees of freedom is determined by the shape, basis function specification, and sample size. For the change-point models, we start with the degrees of freedom for the best SSE change point and add one degree of freedom for each change point in the shape. The *flat* shape only uses one degree of freedom because it is essentially a linear model with a slope of 0. The *incr* shape uses only 1.5 degrees of freedom because it is a linear model where the slope is estimated, but is constrained to be >0 . For more information about computing constrained fits and degrees of freedom, see Meyer (2013a). For each trajectory, all the shapes are fit and an information criterion (IC) is computed for each fit. The smallest information criterion is the

Table 1 Number of training plots by disturbance causal agent used in the forest disturbance model. Plots in the row labeled ‘Probabilistic’ were collected under a stratified random design. Plots in the row labeled ‘Purposive’ were selected to ensure each agent class had a sufficient number of plots for training the empirical model

	Conversion	Fire	Harvest	Stress	Stable	Total
Probabilistic	0	25	14	76	114	229
Purposive	14	58	18	14	8	112
Total	14	83	32	90	122	341

winner. Two options for the information criteria are considered. One is the standard Bayesian Information Criteria,

$$IC_{BIC} = -n \log(SSE) + \log(n) * df0$$

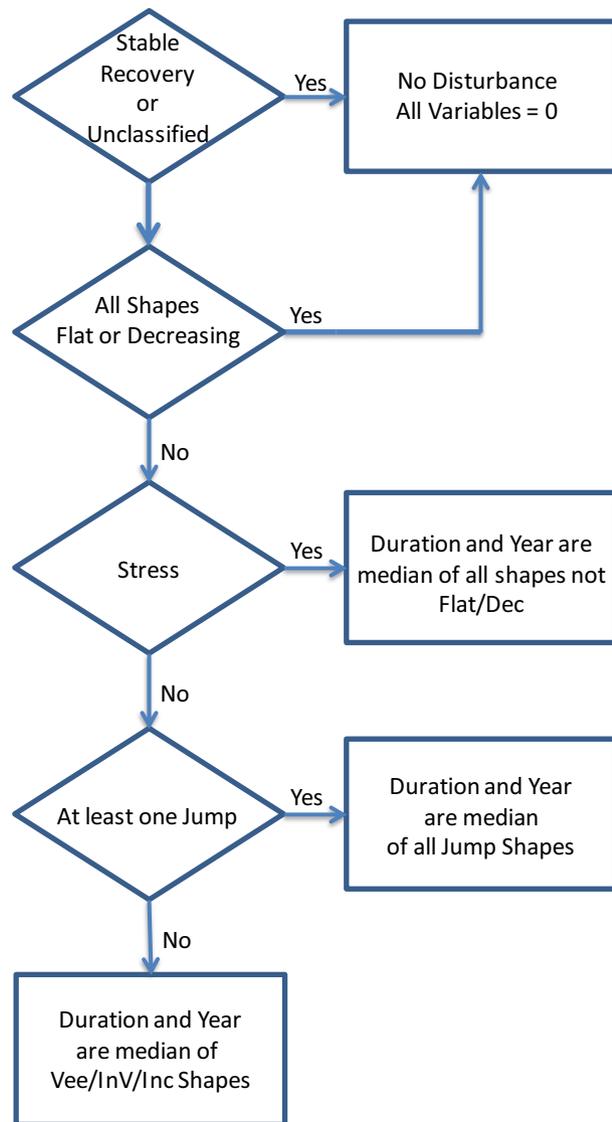


Fig. 4 Flow chart for combining the shape parameters and predicted causal agent to determine year of disturbance and duration. If the predicted causal agent is 'None', the assigned year and duration are both zero. If the predicted causal agent is something other than 'None', but shapes from all four spectral bands are either flat or decreasing, year and duration are coded zero and the predicted causal agent converted to 'None'. If the predicted causal agent is 'Stress', the duration and year are the median of associated band parameters not assigned a flat or decreasing shape. For causal agents other than 'None' or 'Stress' where at least one band or index produces a jump, the assigned year and duration are the median of associated band parameters not assigned a jump shape.

The other is the Cone Information Criterion (Meyer, 2008) which tends to provide greater sensitivity to choosing shapes with change points,

$$IC_{CIC} = -\log(SSE) + \log\left(2 * \frac{(df0 + 1)}{(n - 1 - 1.5 * df0)} + 1\right)$$

Once the best shape is chosen for each pixel, a number of parameters are derived from the model fit that summarize information about the pattern of that trajectory. These parameters include the winning shape, year(s) of change point(s), two measures of magnitude for each change point (i.e., absolute and relative magnitude), duration (i.e., the number of years or image intervals that the spectral change occurs), annual rate of growth prior to change point(s), and annual recovery rate after the change point(s) (Appendix S1). Another useful set of outputs from the shape selection process is the predicted values from the fitted trajectories which provide a smoothed representation of the spectral pattern found at each pixel.

Case study

As an example, we apply the shape-fitting routine to one Landsat scene (path32 row35), located in the mountains of northern Colorado, USA, near Steamboat Springs (Fig. 3). Forests dominate the landscape with approximately 80% of the area occupied by spruce/fir, aspen, and lodgepole pine types. Pinyon juniper and oak woodlands occupy more xeric sites on the fringe of the rangelands. The region is interesting for its complex mix of forest disturbance regimes and presence of large wildfires, harvesting activities, as well as beetle outbreak (labeled here as 'stress'). Although this method is currently being applied nationally, this single scene is used to illustrate the use of outputs from the shapes algorithm in two different applications.

Disturbance attribution maps

Our first application illustrates how to map forest disturbance attribution classes at 30 m resolution annually (1984–2010) in the Colorado scene. Attribution classes included disturbance from fire, harvest, stress (i.e., subtle change brought on by insects, disease, and drought), and stable forest (or no disturbance). The target population is defined by a temporal forest mask layer produced by the VCT disturbance mapping algorithm (Huang *et al.*, 2010a). This mask was also used to form two prestratification classes which included persistent forest and disturbed forest. Training data consisted of a stratified random sample of 229 plots (50% in the disturbed forest stratum, and 50% in the persistent forest stratum) collected using human interpretations made on sample plots using annual Landsat imagery, visualizations of spectral trajectories, and aerial photography through an interpretation tool, TimeSync (Cohen *et al.*, 2010). An additional 112 plots were purposively sampled to augment training data for modeling. Excluding the samples for recovering forests, the number of plots by disturbance causal agent and sampling scheme is given in Table 1.

Modeling disturbance causal agent through time was a two-step process we call ‘flat-to-annual’. First, we built a temporally indifferent (flattened through time) disturbance attribution model where the relationship between the first disturbance on the plots was modeled empirically as a function of a set of spatially explicit predictor variables through Random Forests (Breiman, 2001) using the R package MODELMAP (Freeman *et al.*, 2009). Temporally dynamic predictor variables from the shape selection process included shape, magnitude, duration, pre-, and postdisturbance growth/recovery rates from each of the four spectral bands/indices. Temporally static predictor variables included elevation, forest type derived from Ruefenacht *et al.* (2008), and sine and cosign of aspect. Second, the year of a predicted disturbance was modeled using rules applied to shape parameters involving shape, disturbance year, and duration (Fig. 4) through the flat2parameters function in the ShapeSelectForest package (Meyer *et al.*, 2015).

Validation of the attribution model included examining the out-of-bag contingency table, and estimates of the errors of omission and commission using only the probabilistic training plots weighted appropriately to account for the prestratified design. For the year assignment, the percentage of plots whose disturbance was within specified time intervals of the truth (± 0 years, 1 year, 2–5 years, 5–10 years, and >10 years) were computed.

Mapping tree canopy cover

In our second application, we illustrate how to use shapes output to map tree canopy cover through time. Tree canopy cover is a continuous variable ranging from 0% to 100% within the same forested target population as above. Training data consisted of a systematic sample of 326 plots collected by the US Forest Service, Forest Inventory and Analysis (FIA) program (Reams *et al.*, 2005) ranging in date from 2000 to 2010.

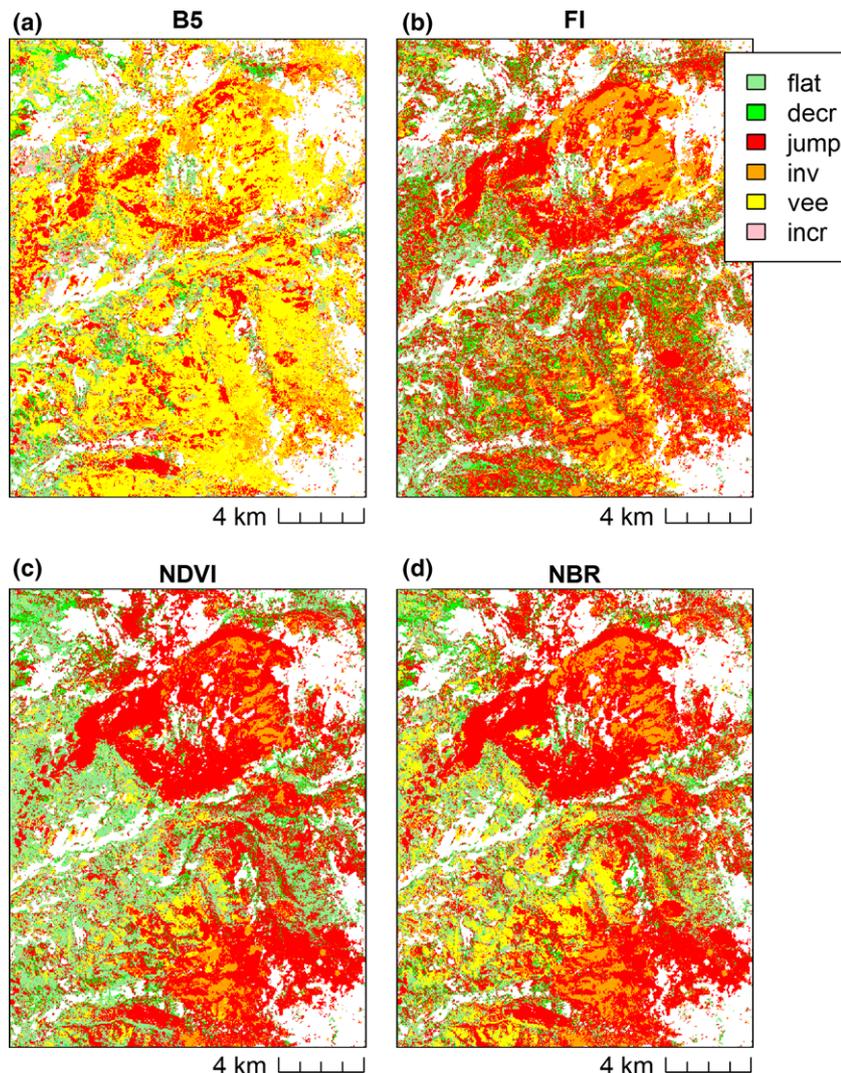


Fig. 5 The shapes algorithm produces dramatically different spatial patterns using different spectral bands/indices for the same small geographic area in the Colorado scene.

Predictor variables included both the 'raw' and 'smoothed' spectral values from each of the four bands/indices at the date corresponding to the year a particular plot was visited in the field. 'Raw' refers to values prior to smoothing from the shapes algorithm, 'smoothed' the converse. Additional static predictor layers included those used in the disturbance mapping above and in Freeman *et al.* (2015). We developed a single random forest model using the training data from 2000 to 2010, then applied that model to predict tree canopy cover for all 26 images (1984–2010) in the Colorado time series.

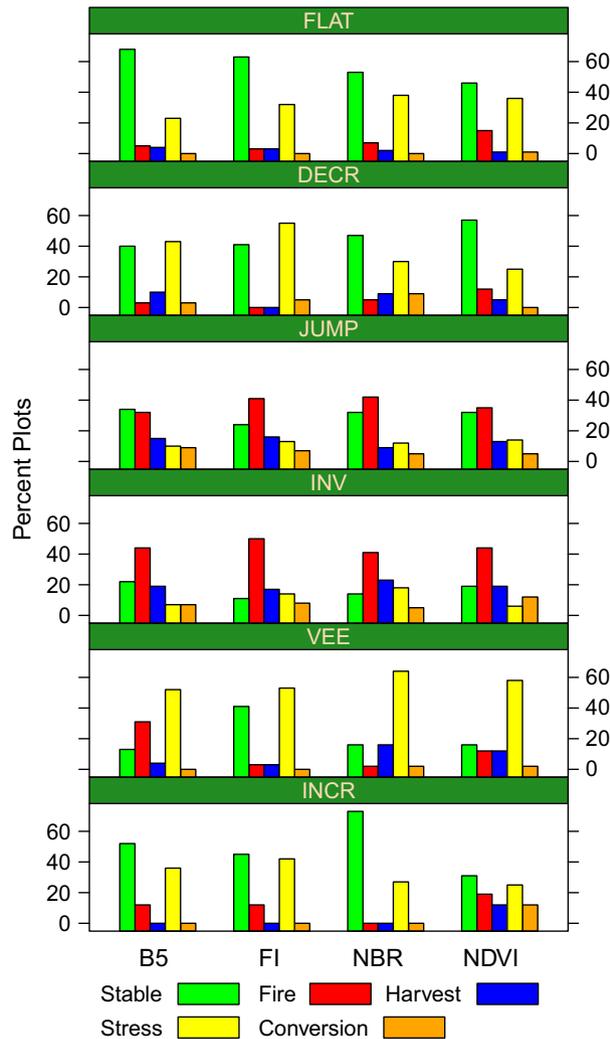


Fig. 6 Distribution of disturbance agent within the selected shapes expressed as percent of training plots. This figure illustrates that the shape parameters for each band contributes different information to the empirical models of forest disturbance. For all four bands, the FLAT, DECR, and INCR shapes are dominated first by stable conditions then by the difficult to decipher stress. VEE's are predominantly stress. JUMP's and INV's are dominated by fire in this particular landscape, but still carry a large proportion of stable as well as the other disturbance agents.

Out-of-bag RMSE and Spearman and Pearson correlation coefficients were used to validate the models and compare performance using both the smoothed and raw spectral values, with, and without static variables.

Results and discussion

Disturbance attribution maps

The shapes output for each of the four bands/indices produced dramatically different spatial patterns (Fig. 5). Figure 6 illustrates the distribution of disturbance agents within the winning shapes by band. Together, these figures illustrate how the shapes algorithm provides unique spatial and temporal information from the bands/indices to help the empirical Random Forests model distinguish between disturbance agents.

The first model in this application, the flat-to-annual disturbance attribution model, predicted the disturbance agent at each Landsat pixel in the scene regardless of year. Table 2 illustrates out-of-bag accuracies obtained by disturbance agent through the entire time series. Here, we have partitioned the contribution of the shapes parameters by running the model with static variables alone, dynamic (shapes) variables, and the two groups combined, illustrating accuracies obtained under each scenario.

The second model in this application annualized the temporally flattened prediction, assigning a date. The majority of the disturbed plots (71.4%) were given the exact year of disturbance. Another 19.5% of the plot predictions fell within 1–3 years of the correct date; 2.6% were within 4–6 years; another 2.6% off by 7–9 years; and finally 3.9% fell 10 years or greater beyond the true date from the reference plots. Figure 7 illustrates patterns of disturbance for a sample of years in the time series, showing the progression of insect disturbance throughout the scene. It should be noted that in this scene, 97% of the training plots had only one disturbance. In other parts of the country dominated by frequent harvests, or where substantially more training data are collected, multiple disturbances could be handled using this same methodology, but adding new classes like 'Harvest–Harvest', 'Fire–Harvest', and 'Stress–Fire'. The behavior of the shapes algorithm for a few instances where we had training data with multiple disturbances is illustrated in Appendix S2.

Modeling disturbance in two steps allows for the capture of disturbance events whose detection in the training data does not coincide exactly with the timing of detection using the shapes algorithm on the spectral data. It is the convergence of evidence in space that

Table 2 Errors of omission and commission using only the data from the probabilistic sample, accounting for the stratum weights and including an estimate of standard error, are shown for three model runs using: static predictors only, dynamic predictors only, and all the predictor variables together. The decrease in errors of omission and commission illustrate the relative contribution of these predictor sets to classifying agents of change.

Agent	Commission error (SE)						Omission error (SE)					
	Static variables		Dynamic variables		All variables		Static variables		Dynamic variables		All variables	
Fire	0.751	(0.064)	0.522	(0.091)	0.387	(0.088)	0.440	(0.100)	0.280	(0.090)	0.240	(0.086)
Harvest	0.773	(0.153)	0.364	(0.146)	0.182	(0.117)	0.857	(0.094)	0.500	(0.134)	0.357	(0.129)
Stable	0.431	(0.049)	0.346	(0.047)	0.328	(0.046)	0.389	(0.050)	0.269	(0.045)	0.227	(0.041)
Stress	0.459	(0.071)	0.354	(0.066)	0.344	(0.063)	0.598	(0.061)	0.494	(0.062)	0.488	(0.062)

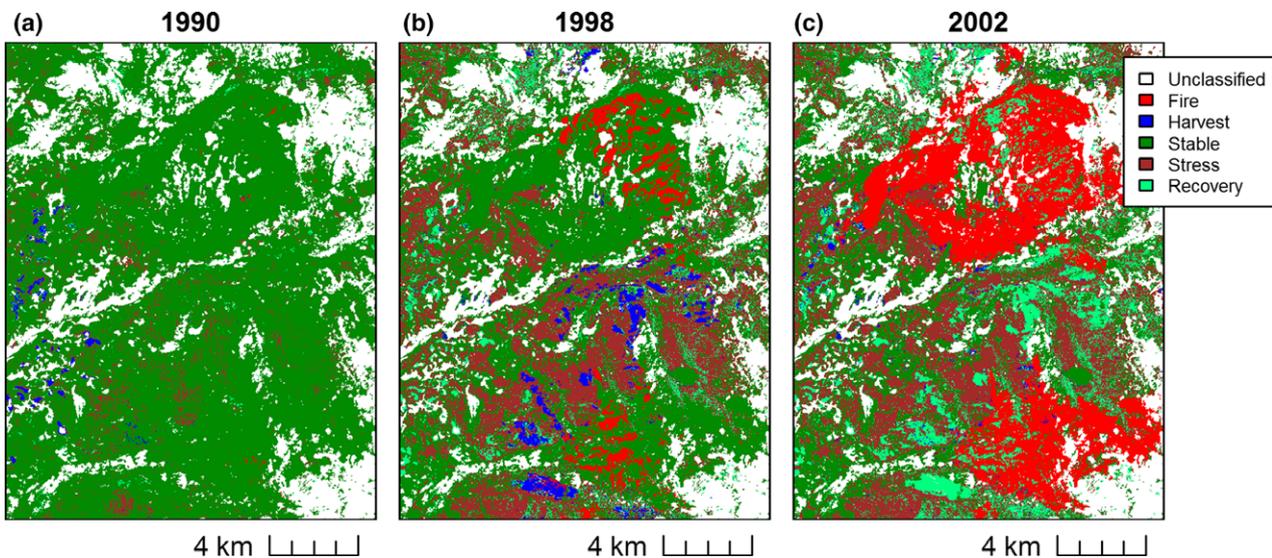


Fig. 7 Predicted disturbance agent over the small clip for a sample of 3 years in the time series, illustrating the progression of beetle outbreak, harvest and fire events. Multiple disturbances such as “Stress – Fire” or “Stress – Harvest” were not modeled here because of insufficient training data in these classes. Appendix S2 illustrates the inflections detected by the shapes algorithm in several of these instances in the Colorado scene.

Table 3 Out-of-bag RMSE, Spearman, and Pearson correlation coefficients obtained using the raw spectral values, raw combined with static predictors, smoothed spectral values, and smoothed combined with static predictors. Little was gained in terms of these accuracy metrics using the smoothed over the raw spectral predictors, but adding static predictors to the spectral data did improve the model outcomes

Model	RMSE	Spearman	Pearson
Raw	21.68	0.48	0.52
Raw + static	19.15	0.61	0.65
Smoothed	20.88	0.54	0.56
Smoothed + static	19.17	0.63	0.65

gives power to the disturbance agent model, while assigning the timing of disturbance through a rule-based approach allows the inclusion of human logic.

Tree canopy cover map

Fitted values can also be used to empirically model and map tree canopy cover through the duration of the Landsat time series. This allows the tracking of forest attributes and their changes coincident with disturbance events. We compare the out-of-bag accuracy metrics from models built using combinations of raw, smoothed, and static predictor variables in Table 3. While using smoothed values from the shapes instead of raw values does not markedly improve the accuracy metrics, it does stabilize the temporal pattern considerably, preventing tree canopy cover from jumping erratically through time (Fig. 8). Further, raw predictions themselves could be run through the shapes algorithm, similar in spirit to approaches invoked by (Powell *et al.*, 2010; Pflugmacher *et al.*, 2012, 2014).

R package and computational issues

The shapes algorithm described here is available as the ShapeSelectForest package (Meyer *et al.*, 2015) on the

CRAN. The package currently consists of three major sets of functions. First, the shape function fits all shapes and chooses the winning shape for each trajectory. Second,

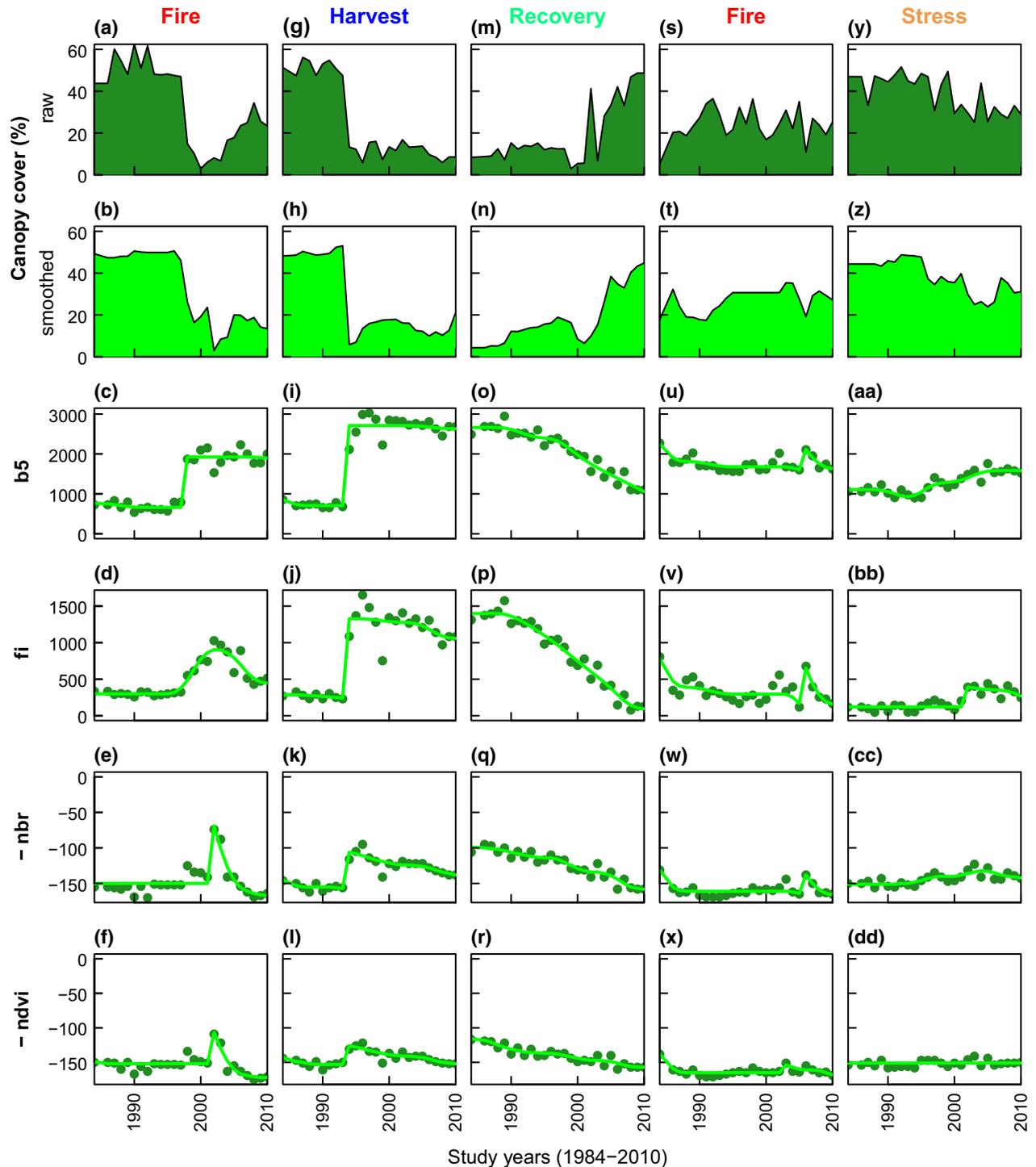


Fig. 8 Predicted tree canopy cover through time for five different disturbance agents (columns). Models using raw spectral values (dark green pattern top row) and shape-smoothed spectral values (light green pattern second row) differ in terms of predictive smoothness through time. Smoothing improves ecological feasibility of tree canopy cover behavior. The last four rows depict corresponding predictor variables (both raw and smoothed) from each of the four bands/indices.

Published 2016.

This article is a U.S. Government work and is in the public domain in the USA, *Global Change Biology*, doi: 10.1111/gcb.13358

the `shapeparams` function derives parameters for empirical models from the shape selection output. Third, a suite of ‘flat-to-annual’ functions annualize the spatial predictions of forest disturbance agent by deriving a year and duration of disturbance.

Using the NASA Earth Exchange (NEX) computing environment (Nemani *et al.*, 2011), we have implemented shapes algorithm on 434 Landsat scenes covering the conterminous United States with 28 years of imagery at 30 m resolution. For national runs, each scene was processed in parallel, using R’s `SNOW` package, across 434 nodes with 16 cores for a total national run time of approximately 21 h, excluding the double jump. Scene-level trials with the double jump option suggest a fivefold increase in computing time over forests if all seven shapes are fit at once. However, computational time for double jumps can be reduced by first filtering out areas with flat, decreasing, or increasing patterns.

US applications

This paper documents methods currently being applied across the conterminous United States using the Landsat data record and forest history metrics from the shapes parameters coupled with information from VCT to map the location and timing of harvest, fire, stress, wind, and conversion events on forest land cover between 1985 and 2010 (Moisen *et al.*, 2012). Work is underway to classify and date two sequential disturbance agents through the double jump and subtle inflections in other shapes. Parameters from the shapes algorithm are being tested for their contributions to mapping changes in tree canopy cover in the next round of the National Land Cover Database (Coulston *et al.*, 2012). The Landscape Change Monitoring System (LCMS) (Masek & Healey, 2012) is also leveraging the diversity of forest dynamics information derived from multiple algorithms including shapes, `LandTrendr`, `VCT`, `CCDC`, `MIICA`, and others through an empirical modeling approach to disturbance mapping in the United States, ultimately over the full Landsat data record (1972—Current). As part of this effort, the shapes algorithm will be implemented in Google Earth Engine in 2016, providing shapes parameters and fitted trajectories for use in other disturbance, tree canopy cover, and biomass mapping projects.

Acknowledgements

The authors would like to thank NASA’s Carbon Cycle and Applied Science programs (grant numbers NNX11AJ78G and NNH14AY63I) as well as the very talented scientists and staff at NASA’s Earth Exchange, University of Maryland, and the US Forest Service, FIA Program. Thanks also go out to the members

of the LCMS team for their efforts pushing change mapping forward in the US. In addition, we are grateful to the anonymous reviewers whose thoughtful comments helped in our revisions.

References

- Breiman L (2001) Random forests. *Machine Learning*, **45**, 5–32.
- Brooks EB, Wynne RH, Thomas V, Blinn CE, Coulston JW (2014) On-the-fly massively multitemporal change detection using statistical quality control charts and Landsat data. *Geoscience and Remote Sensing, IEEE Transactions*, **52**, 3316–3332.
- Cohen WB, Yang ZQ, Kennedy R (2010) Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync – tools for calibration and validation. *Remote Sensing of Environment*, **114**, 2911–2924.
- Cohen WB, Yang Z, Stehman SV *et al.* (2016) Forest disturbance across the conterminous United States from 1985–2012: The emerging dominance of forest decline. *Forest Ecology and Management*, **360**, 242–252.
- Coulston JW, Moisen GG, Wilson BT, Finco MV, Cohen WB, Brewer CK (2012) Modeling percent tree canopy cover: a pilot study. *Photogrammetric Engineering and Remote Sensing*, **78**, 715–727.
- Crist EP, Cicone RC (1984) A physically-based transformation of Thematic Mapper data—The TM Tasseled Cap. *Geoscience and Remote Sensing, IEEE Transactions, GE22*, 256–263.
- Drusch M, Del Bello U, Carlier S *et al.* 2012 Sentinel-2: ESA’s optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, **120**, 25–36.
- Freeman EA, Frescino TS, Moisen GG (2009) *ModelMap: an R Package for model creation and map production using Random Forest and Stochastic Gradient Boosting*. CRAN package.
- Freeman EA, Moisen GG, Coulston JW, Wilson BT (2015) Using random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. *Canadian Journal of Forest Research*, **46**, 323–339.
- Fry JA, Xian G, Jin S *et al.* (2011) Completion of the 2006 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, **77**, 858–864.
- Goward SN, Masek JG, Cohen W *et al.* (2008) Forest disturbance and North American carbon flux. *Eos Transactions*, **89**, 105–116.
- Huang C, Song K, Kim S, Townshend JRG, Davis P, Masek JG, Goward SN (2008) Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, **112**, 970–985.
- Huang C, Goward SN, Masek JG *et al.* (2009) Development of time series stacks of Landsat images for reconstructing forest disturbance history. *International Journal of Digital Earth*, **2**, 195–218.
- Huang C, Goward SN, Masek JG, Thomas N, Zhu Z, Vogelmann JE (2010a) An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment*, **114**, 183–198.
- Huang C, Thomas N, Goward SN, Masek JG, Zhu Z, Townshend JR, Vogelmann JE (2010b) Automated masking of cloud and cloud shadow for forest change analysis using Landsat images. *International Journal of Remote Sensing*, **31**, 5449–5464.
- Kennedy RE, Yang Z, Cohen WB (2010) Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—temporal segmentation algorithms. *Remote Sensing of Environment*, **114**, 2897–2910.
- Kennedy RE, Andréfouët S, Cohen WB *et al.* (2014) Bringing an ecological view of change to Landsat-based remote sensing. *Frontiers in Ecology and the Environment*, **12**, 339–346.
- Key C, Benson N (2005) *Landscape assessment: remote sensing of severity, the normalized burn ratio and ground measure of severity, the composite burn index*. FIREMON: Fire effects monitoring and inventory system Ogden, Utah: USDA Forest Service, Rocky Mountain Res. Station.
- Liao X, Meyer MC (2014) `coneproj`: an R package for the primal or dual cone projections with routines for constrained regression. *Journal of Statistical Software*, **61**, 1–22.
- Masek JG, Healey SP (2012) Monitoring U.S. forest dynamics with Landsat. In: *Global Forest Monitoring from Earth Observation* (eds Frederic A, Hansen MC), pp. 225–242. CRC Press, Boca Raton, FL, USA.
- Masek JG, Vermote EF, Saleous NE *et al.* (2006) A Landsat surface reflectance dataset for North America, 1990–2000. *Geoscience and Remote Sensing Letters, IEEE*, **3**, 68–72.
- Masek JG, Goward SN, Kennedy RE, Cohen WB, Moisen GG, Schleeweis K, Huang C (2013) United States forest disturbance trends observed using landsat time series. *Ecosystems*, **16**, 1087–1104.
- Meyer MC (2008) Inference using shape-restricted regression splines. *The Annals of Applied Statistics*, **2**, 1013–1033.

- Meyer MC (2013a) Semi-parametric additive constrained regression. *Journal of non-parametric statistics*, **25**, 715–730.
- Meyer MC (2013b) A simple new algorithm for quadratic programming with applications in statistics. *Communications in Statistics-Simulation and Computation*, **42**, 1126–1139.
- Meyer MC, Liao X, Freeman EA, Moisen GG (2015) ShapeSelectForest: shape selection for Landsat time series of forest dynamics. CRAN package version 1.1.
- Moisen GG, Schroeder TA, Schleeweis KG, Toney JC, Cohen WB, Goward SN (2012) Attributing causal agents to nationwide maps of forest disturbance. In: *Moving from status to trends: Forest Inventory and Analysis (FIA) symposium 2012*, Morin RS, Liknes GC, comps. 2012 December 4–6; Baltimore, MD. Gen. Tech. Rep. NRS-P-105, pp. 61–64. Department of Agriculture, Forest Service, Northern Research Station. [CD-ROM], Newtown Square, PA, USA.
- Nemani R, Votava P, Michaelis A, Melton F, Milesi C (2011) Collaborative supercomputing for global change science. *Eos Transactions*, **92**, 109.
- Pflugmacher D, Cohen WB, Kennedy ER (2012) Using Landsat-derived disturbance history (1972–2010) to predict current forest structure. *Remote Sensing of Environment*, **122**, 146–165.
- Pflugmacher D, Cohen WB, Kennedy RE, Yang Z (2014) Using Landsat-derived disturbance and recovery history and lidar to map forest biomass dynamics. *Remote Sensing of Environment*, **151**, 124–137.
- Potapov PV, Turubanova SA, Hansen MC *et al.* (2012) Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM data. *Remote Sensing of Environment*, **122**, 106–116.
- Powell SL, Cohen WB, Healey SP, Kennedy RE, Moisen GG, Pierce KB, Ohmann JL (2010) Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: a comparison of empirical modeling approaches. *Remote Sensing of Environment*, **114**, 1053–1068.
- Reams GA, Smith WD, Hansen MH, Bechtold WA, Roesch FA, Moisen GG (2005) The Enhanced Forest Inventory and Analysis Program — national sampling design and estimation procedures. Southern Research Station General Technical Report SRS-80, Asheville, NC.
- Rouse JW, Hass RH, Schell JA, Deering DW (1973) *Monitoring vegetation systems in the Great Plains with ERTS*. In: 3rd ERTS Symposium. NASA SP-351.
- Ruefenacht B, Finco MV, Nelson MD *et al.* (2008) Conterminous US and Alaska forest type mapping using Forest Inventory and Analysis data. *Photogrammetric Engineering and Remote Sensing*, **74**, 1379–1388.
- Schroeder TA, Wulder MA, Healey SP, Moisen GG (2011) Mapping wildfire and clearcut harvest disturbances in boreal forests with Landsat time series data. *Remote Sensing of Environment*, **115**, 1421–1433.
- Vogelmann JE, Xian G, Homer C, Tolck B (2012) Monitoring gradual ecosystem change using Landsat time series analyses: case studies in selected forest and rangeland ecosystems. *Remote Sensing of Environment*, **122**, 92–105.
- Woodcock CE, Allen R, Anderson M *et al.* (2008) Free access to landsat imagery. *Science*, **320**, 1011.
- Wulder MA, Masek JG, Cohen WB, Loveland TR, Woodcock CE (2012) Opening the archive: how free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, **122**, 2–10.
- Zhu Z, Woodcock CE, Olofsson P (2012) Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment*, **122**, 75–91.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Details on parameters derived from the shapes algorithm.

Appendix S2. Examples of the behavior of the shapes algorithm in the presence of two disturbances through time.