



Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr – Temporal segmentation algorithms

Robert E. Kennedy^{a,*}, Zhiqiang Yang^a, Warren B. Cohen^b

^a Department of Forest Ecosystems and Society, Oregon State University, 321 Richardson Hall, Corvallis, OR 97331, United States

^b Pacific Northwest Research Station, USDA Forest Service, Corvallis, OR 97331, United States

ARTICLE INFO

Article history:

Received 17 January 2010

Received in revised form 22 July 2010

Accepted 24 July 2010

Keywords:

Change detection

Land cover dynamics

Landsat

Time series

Forest disturbance

Forest growth

Temporal segmentation

ABSTRACT

We introduce and test LandTrendr (Landsat-based detection of Trends in Disturbance and Recovery), a new approach to extract spectral trajectories of land surface change from yearly Landsat time-series stacks (LTS). The method brings together two themes in time-series analysis of LTS: capture of short-duration events and smoothing of long-term trends. Our strategy is founded on the recognition that change is not simply a contrast between conditions at two points in time, but rather a continual process operating at both fast and slow rates on landscapes. This concept requires both new algorithms to extract change and new interpretation tools to validate those algorithms. The challenge is to resolve salient features of the time series while eliminating noise introduced by ephemeral changes in illumination, phenology, atmospheric condition, and geometric registration. In the LandTrendr approach, we use relative radiometric normalization and simple cloud screening rules to create on-the-fly mosaics of multiple images per year, and extract temporal trajectories of spectral data on a pixel-by-pixel basis. We then apply temporal segmentation strategies with both regression-based and point-to-point fitting of spectral indices as a function of time, allowing capture of both slowly-evolving processes, such as regrowth, and abrupt events, such as forest harvest. Because any temporal trajectory pattern is allowable, we use control parameters and threshold-based filtering to reduce the role of false positive detections. No suitable reference data are available to assess the role of these control parameters or to test overall algorithm performance. Therefore, we also developed a companion interpretation approach founded on the same conceptual framework of capturing both long and short-duration processes, and developed a software tool to apply this concept to expert interpretation and segmentation of spectral trajectories (TimeSync, described in a companion paper by Cohen et al., 2010). These data were used as a truth set against which to evaluate the behavior of the LandTrendr algorithms applied to three spectral indices. We applied the LandTrendr algorithms to several hundred points across western Oregon and Washington (U.S.A.). Because of the diversity of potential outputs from the LTS data, we evaluated algorithm performance against summary metrics for disturbance, recovery, and stability, both for capture of events and longer-duration processes. Despite the apparent complexity of parameters, our results suggest a simple grouping of parameters along a single axis that balances the detection of abrupt events with capture of long-duration trends. Overall algorithm performance was good, capturing a wide range of disturbance and recovery phenomena, even when evaluated against a truth set that contained new targets (recovery and stability) with much subtler thresholds of change than available from prior validation datasets. Temporal segmentation of the archive appears to be a feasible and robust means of increasing information extraction from the Landsat archive.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Landsat instruments have witnessed decades of unprecedented change on the Earth's surface (Wulder et al., 2008), and have the spatial and temporal properties needed to capture the processes driving that change (Cohen and Goward, 2004). Traditionally, change detection studies have focused on comparisons between two images:

one before and one after a change (Coppin et al., 2004; Lu et al., 2004). To more fully tap the long archive of Landsat data, some studies use multiple two-date comparisons in sequence to summarize multi-temporal trends over time (Cohen et al., 2002; Jin and Sader, 2005; Olsson, 2009). Although these latter approaches are powerful, change analyses based on two-date change detection methods do not fully tap the interrelationships among many multitemporal images, and may not be able to separate from background noise the subtle or long-duration changes in cover condition and vigor that are expected under climate change (Hicke et al., 2006; Logan et al., 2003). Recognizing this limitation, methods that simultaneously consider the signal from

* Corresponding author.

E-mail address: Robert.kennedy@oregonstate.edu (R.E. Kennedy).

multiple images have been developed over many years with the goal of improving the signal-to-noise ratio in change detection, and these methods have begun to flourish with the recent opening of the Landsat archive by the U.S. Geological Survey (USGS) (Garcia-Haro et al., 2001; Goodwin et al., 2008; Healey et al., 2006; Hostert et al., 2003; Huang et al., 2010; Lawrence and Ripple, 1999; Röder et al., 2008; Viedma et al., 1997; Vogelmann et al., 2009).

These true multitemporal change detection approaches using Landsat data can be grouped largely by whether they seek deviations or trends. Those focusing on deviation use multiple instances of a presumed stable condition to better define when a persistent change has occurred that moves the spectral signal away from that stable condition, allowing better separation of subtle but true change from background noise (Garcia-Haro et al., 2001; Goodwin et al., 2008; Healey et al., 2006; Huang et al., 2010). Those that seek trends, in contrast, use time-series fitting algorithms to separate longer-duration signals from year-to-year noise caused by phenology, sun angle, sensor drift, atmospheric effects, and geometric misregistration (Hostert et al., 2003; Lawrence and Ripple, 1999; Röder et al., 2008; Viedma et al., 1997; Vogelmann et al., 2009). In essence, the two groups of algorithms capture either abrupt events or slower processes, respectively, and both assume the absence of the other type of effect. To describe and understand our increasingly human-dominated Earth system (Vitousek et al., 1997), algorithms must be developed that can flexibly capture both abrupt disturbance events and longer-term stress-induced degradation and ecological change induced by human and natural processes.

In this paper, we describe and test LandTrendr (Landsat-based Detection of Trends in Disturbance and Recovery), a set of processing and analysis algorithms that move toward the goal of capturing both trends and events. We previously described one approach toward meeting this goal by comparing spectral trajectories against idealized temporal models of land cover change (Kennedy et al., 2007b), but with LandTrendr we take an entirely different strategy. Rather than prescribe models of change, the new strategy is one of arbitrary temporal segmentation: using straight line segments to model the important features of the trajectory and to eliminate the noise. No predetermined models of change are required – the data themselves determine the shape of the change trajectory. The result of temporal segmentation is a simplified representation of the spectral trajectory, where the starting and ending points of segments are vertices whose time position and spectral value provide the essential information needed to produce maps of change.

Three broad strategies underlie the LandTrendr approach. First, we seek to track change in cover condition that is durable across years, not to capture intra-year trends. Because of cloud cover, data collection gaps, and the 16-day satellite repeat cycle, the annual time step is the temporal scale for which we believe the majority of the Landsat archive is best suited. Year-to-year variability is thus considered noise. Second, we move toward a purely pixel-based structure for analysis. Multiple images per year can be fed to the algorithm, but each pixel's temporal trajectory is constructed from the series of single best values for each pixel for each year. Clouds, cloud shadows, and gaps caused by the Landsat 7 scan-line corrector failure can thus be avoided on a pixel-by-pixel basis. Finally, the algorithms allow both for the temporal smoothing of spectral noise in long-duration signals similar to trend-seeking approaches (e.g. Röder et al., 2008; Viedma et al., 1997) and for the unsmoothed capture of spectral change for abrupt events similar to the deviation-seeking strategies (e.g. Huang et al., 2010). Allowing both in a single pass through the data requires the development of hybrid fitting approaches that form the core of our analysis. The primary objective of our paper is to describe the processing and analysis algorithms needed to achieve temporal segmentation.

By allowing any arbitrary segmentation, we increase the risk of over- and under-fitting. Therefore, the algorithm requires a series of control parameters and filtering steps designed to reduce overfitting

while still capturing the desired features of the trajectories. The secondary objective of our paper is to provide an overview of the effects of those parameters on the final segment-based representation of the trajectory, and ultimately on the ability to capture change.

Evaluating parameter effects poses a challenge, however. First, the algorithm results must be compared to a reference dataset that captures all of the phenomena of interest (abrupt events as well as longer-duration processes) over large areas at a spatial grain and temporal density commensurate with Landsat data. No such datasets exist. Second, methods to assess change detection algorithms often focus on verifying capture of abrupt events (Cohen et al., 1998; Hayes and Sader, 2001), which is only part of our goal. Therefore, in a companion paper (Cohen et al., in review), we describe the fundamentally new conceptual approach needed for expert interpretation of the image stacks themselves, allowing delineation of both abrupt and long-duration processes, and we document the stand-alone software package developed to facilitate that approach (known as TimeSync). In this paper, we use TimeSync interpretations to aid in characterizing LandTrendr performance and sensitivity to parameter value settings.

2. Methods

2.1. Study area

The project was carried out on a set of randomly-located testing plots distributed across four Landsat scenes in the Pacific Northwest of the U.S.A. The work is part of a USDA Forest Service-funded project to map disturbance within the geographic bounds of the Northwest Forest Plan (NWFP), a federal management plan designed to protect the northern spotted owl and other endangered and threatened forest-dependent species (Haynes et al., 2006; USDI and USDA, 1994). Forest types are diverse, including wet temperate rainforests near the Pacific coast; high productivity temperate coniferous forests across much of mid-elevation areas; high elevation, cold-tolerant forests in mountainous areas; broadleaf-dominated forests along riparian areas and in some successional classes throughout; and lower productivity conifer forests and shrublands in the dry interior regions (Franklin and Dyrness, 1988; Ruefenacht et al., 2008). The dominant processes occurring in the forests of the region are anthropogenic harvest (both clear- and partial-cutting), fire, insect-related mortality, and post-disturbance regrowth. For the purposes of the NWFP project, desired detection processes were disturbance timing, magnitude, and duration, as well as recovery magnitude and onset.

We identified all Landsat scenes (where “scene” refers to the geographic footprint of a single Landsat World Reference System II [WRS-II] address) that intersect the NWFP area of Oregon and Washington, and for each scene delineated the non-overlapping Thiessen (or Voronoi, see Okabe et al., 2000) polygon that describes the area closer to the scene center than to adjacent scenes, which we refer to as the Thiessen scene area (TSA; Fig. 1). From this set, we selected four TSAs that captured a wide range of environmental and management conditions found in the larger NWFP. All subsequent steps were then carried out for the TSAs of these four pilot scenes.

2.2. Overall processing flow

The processing flow of LandTrendr contains familiar steps of image selection, preprocessing, and analysis, but the requirements of the trajectory-based approach affect decisions throughout.

2.3. Preprocessing

2.3.1. Image selection

We constructed image stacks for the four pilot TSAs, composed primarily of images acquired from mid-July to late August, with nearly

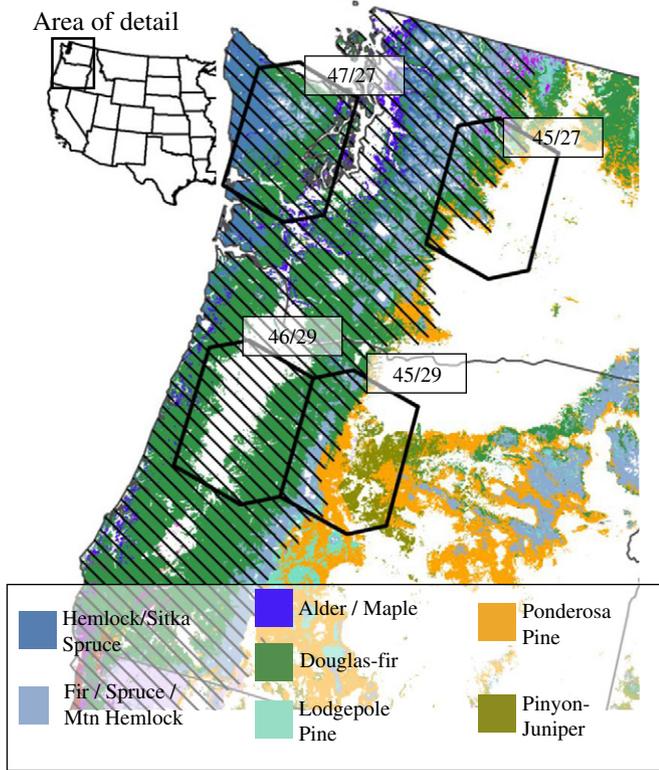


Fig. 1. The study area includes four Landsat Thematic Mapper scenes (black polygons) that sample a diversity of forest types from within the area of the Northwest Forest Plan (NWFP; hatched area) in the Pacific Northwest, U.S.A. Forest type labels are taken from the USDA Forest Service Forest Inventory and Analysis Program’s national forest type map (Ruefenacht et al., 2008).

every year between 1985 and 2007 represented (Fig. 2). For years where clouds were more prevalent or where Landsat 7 SLC-off images with data gaps were acquired, we used multiple images per year to improve spatial coverage of usable pixels. In image selection for LandTrendr, consistency of seasonality trumps absence of clouds as the highest priority. Clouds and their shadows are typically easier to mask (see Section 2.3.5) than phenological and sun-angle effects are

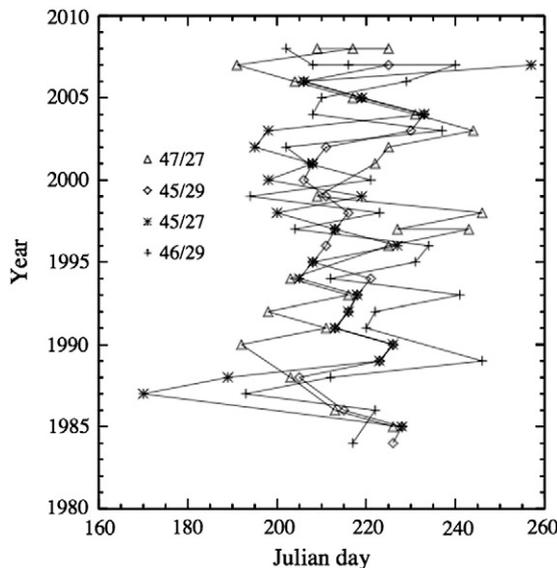


Fig. 2. Scene acquisition dates by year for the four pilot scenes. Julian days are calculated relative to January 1 without reference to leap year; for reference, Julian day 220 is August 8th.

to model, and the processing algorithms can ingest multiple partially-clouded images per year to create yearly “on-the-fly” mosaics with minimal clouds (see Section 2.4). Even for those remaining pixels that happen to be cloudy in all images of one year, there are frequently preceding or subsequent images in the annual time series where these pixels are not clouded, and thus spectral trajectories can still be fit even with occasional cloud-obscured views.

For efficient management of stacks of imagery, we tracked all relevant file and processing information in a meta-data structure for each stack that was passed from module to module in the processing flow.

2.3.2. Geometric correction

Images acquired after the USGS opened the Landsat archive typically had sub-pixel geolocation accuracies and required no further geometric processing, but because the project began before that time, we constructed LTS from a range of sources and, where necessary, used our automated tie point selection algorithm (Kennedy and Cohen, 2003) to locate two to five hundred tie points per image for orthorectification.

2.3.3. Radiometric normalization and transformation

Following Schroeder et al. (2006), we used the cosine-Theta (COST) correction of Chavez (1996) to remove most atmospheric effects for a single reference image in a given LTS. Dark object values for use in the COST calculations were based on visual assessment by a trained interpreter following the rules outlined in protocols found in Kennedy et al. (2007a). All remaining images in the LTS of a single scene were then normalized to the COST image using the MADCAL (multivariate alteration detection and calibration) algorithms of Canty et al. (2004), which identify stable pixels within a small subset of the larger image. We expanded the basic approach to allow use of multiple subsets across each scene and across the stack, greatly speeding the normalization process relative to a manual scene-by-scene approach. The exceptions were occasional images in a stack where subsets would violate the assumptions of the MADCAL algorithm and introduce clouded pixels into the no-change population. To aid the interpreter in identifying these scenes, we produced band-wise scatterplots as well as stack-wise summary statistics to identify anomalies. We emphasize that the method is not limited to cloud-free scenes, but only that some of the subsets in a given image be cloud free.

2.3.4. Tasseled-cap transformation

Because the normalization process makes the spectral space relatively consistent across sensors, it is imperative that a single spectral transformation be used for all images in the LTS. We chose the tasseled-cap transformation, and used the coefficients defined for reflectance data (Crist, 1985), regardless of Landsat sensor.

2.3.5. Cloud, smoke, snow, and shadow screening

Our strategy to identify clouds, smoke, snow and shadows was to contrast each subject image in the time series with a single reference image using scores that accentuate the target effect, and then use manual interpretation to define cloud and shadow thresholds based on those scores. At their most basic level, the cloud/snow/smoke score and shadow scores increased with increasing and decreasing tasseled-cap Brightness (TCB), respectively, in the subject image relative to the reference image. These scores were further modified to reduce confusion caused by actual land surface change. To reduce effects of vegetation clearing, the cloud score also increased with increasing tasseled-cap Wetness (TCW; indicative of clouds, but not cleared areas) and with decreasing tasseled-cap Greenness (TCG) relative to TCB and TCW. To improve sensitivity in low-brightness areas, the shadow score also increased with increasing TCW; to reduce confusion with vegetation regrowth, shadow scores were masked out where TCG increased relative to TCB.

Once score images were calculated, the interpreter then manually determined thresholds of cloud score and shadow score for each

image by visual comparison of score images against the actual images. Interpreter derived thresholds were applied to the score images to create [0,1] masks for cloud/snow/smoke and for shadow, then combined and spatially filtered with a 3 by 3 median filter to expand the mask near edges of clouds and shadows, where cloud effects can bleed into adjacent pixels.

Although we report the details of our cloud masking strategy here, we note that the mechanics of cloud masking are not our overall focus in this research. Rather, we focus the algorithm development on segmentation once clouded pixels have been identified, and have developed our workflow to allow easy integration of cloud masks from other sources as necessary (such as those developed for the Vegetation Change Tracker [VCT] of Huang et al. (2010)). Thus, the particulars of our masking approach (manual thresholds, post-mask filtering) are less integral to the method.

2.4. Image stack mosaicking and trajectory extraction

Once the images in a TSA were normalized and cloud mask images were created, the image stack was arranged for trajectory segmen-

tation. For years where multiple images were available, an on-the-fly mosaic was created, where each pixel's spectral value was taken from the image closest to the median Julian day (1 to 365) of the images in the whole stack, or iteratively from the next closest in date if the pixel in the target date was clouded.

Spectral trajectories were then extracted at the pixel scale and passed to the trajectory segmentation algorithms. Spectral values for each year can be taken from any arbitrary window kernel centered on the pixel of interest; for this paper, we chose to use the mean value in a 3 by 3 window as a compromise between spatial detail and robustness to pixel-wise misregistration across images in the stack. Any single spectral index could be used.

2.5. Trajectory segmentation

Temporal segmentation is the core approach we use to capture both abrupt and slow phenomena in the Landsat time series. By “temporal segmentation,” we refer to modeling a pixel's spectral time series as a sequence of straightline segments that capture the broad features of the trajectory, eliminating noise but without sacrificing necessary detail.

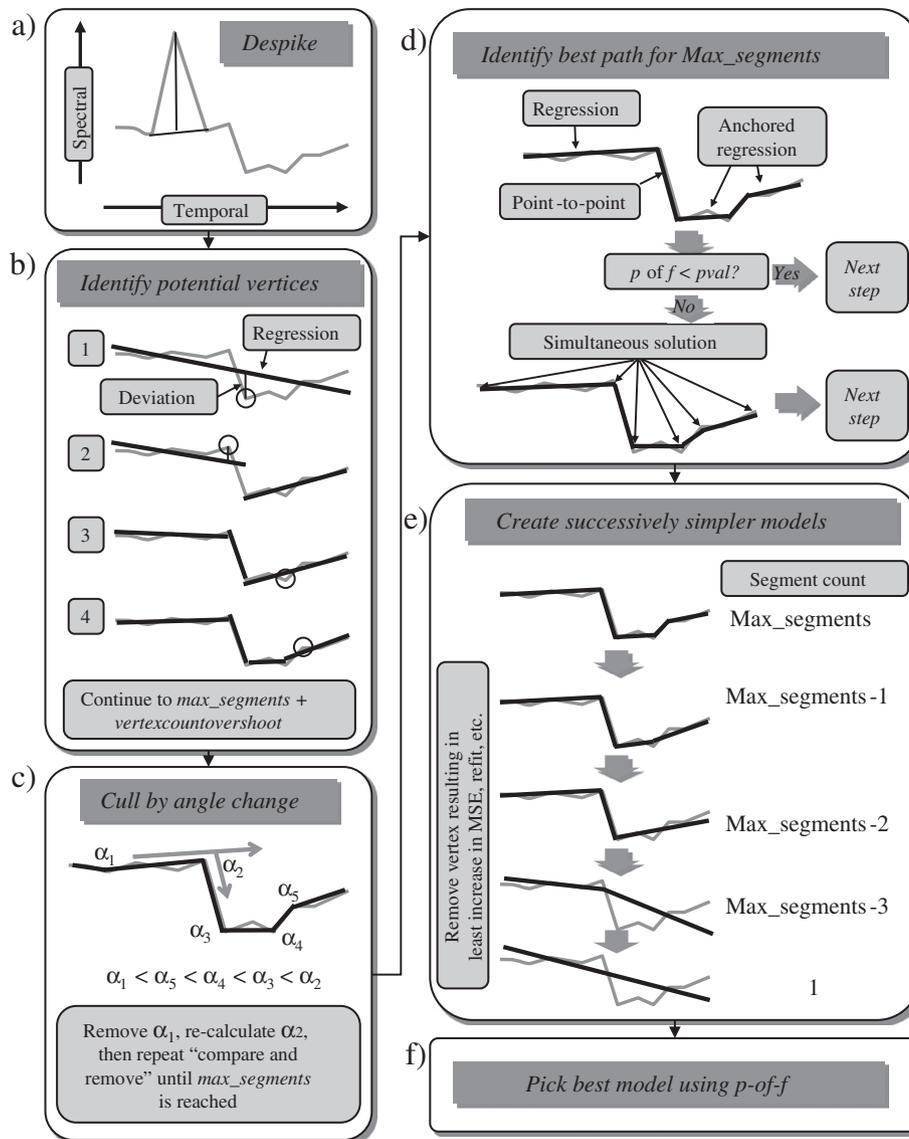


Fig. 3. The LandTrendr segmentation process. After removal of ephemeral spikes (a), potential vertices are identified using deviation from simple regression lines (b), with more vertices identified than needed. Excess vertices are removed based on low angle change (c). A single path through the vertices is chosen using flexible fitting rules (d), and then segments are removed to create successively simplified models of the trajectory (e). The model with the best fit (accounting for increased model complexity) is chosen (f).

The overall strategy (Fig. 3) is to identify a maximally-complex model of the trajectory (Sections 2.5.1, 2.5.2, and 2.5.3), then iteratively simplify the model (Section 2.5.4), choose the best model based on simple fitting statistics (Section 2.5.5), and finally remove change that is considered noise based on a vegetative cover criterion (Section 2.5.6).

2.5.1. Remove spikes

We assumed that noise-induced spikes (residual clouds, snow, smoke or shadows) would be ephemeral, with the spectral index returning to pre-spike value in the post-spike year, while real land cover changes would more persistent, with the post-spike spectral value not returning to the pre-spike value quickly (Fig. 3a). Therefore, the despiking algorithm would only remove spikes where the spectral value before and after the spike was similar, with the necessary degree of similarity set by a control parameter (“despike”; Table 1). The process was iterative, correcting only the worst spike at each iteration. If no spikes exceeded the despike parameter, no changes were made.

2.5.2. Identifying potential vertices

Conceptually, a vertex in the time series is the [year, spectral value] combination that separates a distinct period before it from a distinct period after it. We employed two complementary strategies to identify and then cull vertices.

Regression-based vertex identification used a residual-error criterion to identify vertices (Fig. 3b). The first and last years of the entire time series were defined as initial vertices, and a least-squares first-order regression of spectral index versus year was calculated for all of the points in the time series. The point with the largest absolute deviation from the fitted regression line was designated as the next vertex, defining the break between a segment that runs from the first year to that vertex and a segment that runs from that vertex to the final year. For each of those two segments, new regressions were calculated, and then used to determine each segment’s mean square error (MSE). The segment with the larger MSE (greatest residual error) was identified for the next split, repeating until the segment count reached the number set by adding the “max_segments” and “vertexcountovershoot” control parameters (Table 1).

Excess segments were then culled using the angle criterion (Fig. 3c). Segments were calculated connecting actual values (not regression fitted) of vertices in [year, spectral value] space. To allow a single angle

criterion to have meaning across spectral indices with different absolute values, however, spectral values were scaled to make the year and spectral_value ranges equivalent. The vertex with the shallowest angle was removed, resulting in the elimination of one segment. A new segment was drawn between the two vertices adjacent to the one removed, and then all vertex angles recalculated. This culling process was repeated until the segment count reached max_segments.

2.5.3. Fitting trajectories

Once a final set of candidate vertices was identified, a second set of fitting algorithms was used to determine the spectral values for each vertex year that would result in the best continuous trajectory through the time series (Fig. 3d). Conceptually, the vertex algorithms in the prior section focused on the x-axis of the time series (identifying the years of breaks in the time series) while the trajectory-fitting algorithms honed the y-axis (the spectral values) given the x-axis vertex points.

To capture the full range of land cover dynamics, we propose that trajectory-fitting algorithms must smooth noise over trends but faithfully capture abrupt change without smoothing—goals that are often at odds. We thus employed a flexible fitting approach that allowed both point-to-point and regression-based fitting between vertices, working from early to late through the vertices. For the first segment, either a regression or vertex-to-vertex line (using actual vertex y-axis values) was used depending on which approach resulted in lower MSE for the segment. For segments other than the first, the position of the start year of the segment was constrained to connect with the prior segment; the latter was connected either to the actual point of the next vertex (point-to-point) or smoothed with a simple regression anchored at the starting point (regression-based fitting). The result of the process was a series of connected segments (with the number of segments controlled by the max_segments parameter) that best connected the vertices chosen in the prior step.

As an indicator of the goodness of fit, we calculated the *p*-value of a standard *F*-statistic for the fit of this model across the entire trajectory. We adjusted degrees of freedom for vertex-to-vertex segments, which have zero deviation, to avoid erroneous reduction in *p*-value. Note that this *p*-value is not a true measure of goodness of fit, because there are no *a priori* models being tested and because temporal autocorrelation is not accounted for. Rather, it provides a comparison among segmentation models with different numbers of segments (predictor variables) that,

Table 1
Value ranges for control parameters used in test runs of the LandTrendr algorithms.

Parameter	Description	Values tested
<i>despike</i>	Before fitting, spikes are dampened if the spectral value difference between spectral values on either side of the spike is less than 1-despike desawtooth proportion of the spike itself. Lower values filter spikes more aggressive. Nc setting to 1.0 turns off.	1.0, 0.9, 0.75
<i>pval</i>	If best fitted trajectory’s <i>p</i> -of- <i>F</i> value exceeds this threshold, the entire trajectory is considered no-change.	0.05, 0.1, 0.2
<i>max_segments</i>	The maximum number of segments allowed in fitting	4, 5, 6
<i>recovery_threshold</i>	During fitting, if a candidate segment has a recovery rate faster than 1/recovery_threshold (in years), that segment is disallowed and a threshold different segmentation must be used. Setting to 1.0 turns off this filter.	1, 0.5, 0.25
<i>vertexcount overshoot</i>	The initial regression-based detection of potential vertices can overshoot (<i>max_segments</i> + 1) vertices by this value; angle-based culling is used to return to the desired number of vertices if overshoot occurs. Allows a mix of criteria for vertex identification.	0, 3
<i>pct_veg_lossl*</i>	Disturbance segments of 1-year duration are considered no-change if their change in spectral value, when converted to percent vegetative cover, is less than this threshold.	10, 15
<i>pct_veg_loss20*</i>	The same as prior, but for segments with duration 20 years or greater.	3, 5, 10
<i>pre_dist_cover*</i>	Disturbance segments that start in percent-cover conditions lower than this value will be considered no-change. Filters out change in.	10, 20, 40
<i>Pct_veg_gain*</i>	Recovery segments of any duration are considered no-change if their change in spectral value, when converted to percent vegetative cover is less than this threshold.	3, 5, 10

* All parameters were varied for Phase 1, but only those marked with asterisks were used in Phase 2.

through experimentation, we found provided the same relative information as Akaike's information criterion and as a separate autocorrelation-resistant measure, but at much greater speed.

If the fitting based on the early-to-late regression approach resulted in a p -value greater than a user-defined threshold (parameter "pval"), then the entire fitting process was repeated with an approach that allows all vertex y -values to vary simultaneously rather than in an early-to-late sequence. This approach was based on an implementation of the Levenburg–Marquardt fitting algorithms (MPFIT, Craig Markwardt, <http://cow.physics.wisc.edu/~craigm/idl/idl.html>), and was much more demanding computationally. The outcome of this second fitting process was retained, regardless of p -value.

2.5.4. Simplifying models

The result of the prior section was the maximally-complex segmentation model of the trajectory. The next step was to iteratively simplify and re-fit the trajectory using successively fewer segments, resulting in models of the trajectory for segment counts ranging from $max_segments$ downward to 1 (Fig. 3e). Each simplification of the trajectory was accomplished through removal of the weakest vertex, defined by either a recovery rate or an MSE criterion. The recovery rate criterion (control parameter $recovery_threshold$) removes a segment if its slope of spectral recovery would span the entire spectral range of the trajectory (i.e. minimum to maximum spectral value for the individual trajectory) in $1/recovery_threshold$ years. It reflects the fact that extremely fast recovery in many ecosystems is more likely caused by inadequate shadow or cloud masking than by unrealistically robust conditions on the ground.

For each reduced model, the fitting process described in Section 2.5.3 was re-applied, and fitting statistics recalculated. Because models with more segments have greater degrees of freedom, simpler models can have higher fitting scores than the more complex models.

2.5.5. Determining the best model

The final step was to identify the best model using the p -value for the F -statistic (p -of- f). From the best model, each segment's type (disturbance or recovery), vertex years and vertex values formed the core outputs of the LandTrendr segmentation algorithms that were evaluated against an independent dataset (see Section 2.6). We refer to these outputs as Phase 1 to distinguish them from those after filtering described in the next section.

2.5.6. Filtering by change in vegetative cover

To remove small spectral changes captured by the raw segmentation algorithms, we examined the impact of filtering using a magnitude-of-change threshold.

Because the spectral indices themselves are of little inherent biophysical meaning, we filtered based on estimated percent vegetative cover change. Vegetation cover estimates were determined from aerial photos using methods described in Cohen et al. (2010). We used two approaches to link spectral data with these interpreted values to estimate percent-cover change. In the static model approach, we developed a regression between spectral values and percent-cover estimates, then used that static model to calculate percent cover both before and after each change, and finally subtracted the resultant percent-cover estimates to calculate vegetation cover change. In the delta model approach, we modeled change in cover directly. We used a

jackknifing procedure to randomly pair different plots, and with regression we modeled the difference in percent cover between the plots with the difference in spectral value between the plots. For both types, we developed models using three different spectral indices: the tasseled-cap wetness ("wetness"; Crist, 1985), the normalized burn ratio (NBR; van Wagtenonk et al., 2004), or the normalized difference vegetation index (NDVI; Tucker, 1979). Summaries of the percent-cover models are shown in Table 2.

Percent-cover change for each segment was then compared against control parameters (Table 1) to determine if the segment's change merited retention. Filtering was done on a sliding scale as a function of the duration of the segment: changes detected as consistent trends over 20 years of imagery are less likely to be caused by noise, and therefore could be labeled as change at a lower threshold (control parameter pct_veg_loss20) than 1-year changes, which are more susceptible to ephemeral spectral noise and labeled change at a potentially higher threshold (control parameter pct_veg_loss1). Thresholds for intermediate-duration trends were calculated with simple linear interpolation of the 1 and 20 year thresholds. If a segment's change in spectral index values was less than a user-defined parameter, that segment was considered a "stable" segment. We refer to segmentation results after percent-cover filtering as Phase 2 results.

2.6. Evaluating segmentation results

As a potentially complex system with many control parameters, the algorithms described in Section 2.5 must be evaluated to determine how changing those parameters affects final outcomes. Our evaluation approach was similar to a sensitivity or optimization analysis: the complex model (LandTrendr) was run using many combinations of parameter values and compared to a reference dataset (determined using the TimeSync tool introduced in our companion paper (Cohen et al., 2010)) using a variety of scoring metrics.

2.6.1. Parameter variations

TimeSync interpretation was conducted at 543 plots distributed randomly across the four Landsat TSAs described in Section 2.1. These plots were then evaluated using the TimeSync tools for interpretation described in Cohen et al. (2010). Of those, 388 were labeled as being "forest" at some point in the time period and were used as the reference dataset for evaluation. For each of the forested plots, we repeatedly applied the LandTrendr segmentation algorithms described in Section 2.5, varying the control parameters across all ranges of values shown in Table 1 for the wetness, NBR, and NDVI spectral indices. Each of these "runs" resulted in a set of segments and vertices for each of the plots that was compared against the segments and vertices defined by the human interpreter.

As noted above, an important distinction was between Phase 1 runs with and Phase 2 runs without filtering by vegetative cover. Variation in parameter values for Phase 1 resulted in 2916 unique combinations of parameter values. Addition of the parameter values for Phase 2 increased the count to 157,464.

2.6.2. Scoring metrics

There is no simple, unified metric that can be used to compare how well the automated algorithm captured all of the features of the trajectory identified by the interpreter. For some purposes, accurate

Table 2
Regression models (and r^2 values) linking interpreted vegetative cover* to spectral values.

Spectral index	Static model	Delta model
NBR	$16.12 + 104.65^* nbr$ (0.75)	$108.46^* \Delta nbr - 0.22$ (0.75)
NDVI	$1.12 + 84.23^* ndvi$ (0.56)	$84.17^* \Delta ndvi - 0.03$ (0.56)
Wetness	$100 - 100^* (1 - \exp(21^* wetness))^8$ (0.82)	$412.6^* \Delta wetness + 1.48$ (0.74)

* Vegetative cover interpreted at 313 plots (see Cohen et al., in review).

timing of turning points (vertices) is critical, but for others overall agreement in trajectory shape or temporal overlap of long-duration processes is more important.

Consequently, the effectiveness of the algorithm can be evaluated solely by comparing vertices or by evaluating overall match in shape. The vertex-focused approach evaluates how well the algorithm matches the timing and direction of turning points in the spectral trajectory (vertices) defined by the interpreter. For either the interpreter- or algorithm-derived segmentation, each year in a spectral trajectory is labeled as a disturbance, recovery, or stable vertex depending on the label for the subsequent segment, or as no-vertex for years where none is assigned. In contrast, the trajectory match approach evaluates how well the algorithms match the overall shape of the spectral trajectory determined by the interpreter. For each year in the trajectory, the disturbance, recovery, or stable label defined by the algorithms is compared against that made by the interpreter, and agreement is scored when the labels agree in the same year, regardless whether at a vertex point or at a midpoint of a segment. For either the vertex or overall trajectory match case, standard contingency tables can be constructed (Fig. 4). From the vertex-based or trajectory match contingency tables, eight summary metrics were then calculated for each run: five focusing on timing of disturbance vertices only (*Disturbance false negative*, *disturbance false positive*, *disturbance matched*, *disturbance accuracy* and *disturbance kappa*), two on timing of transitions to any type (*accuracy* and *kappa*), and one on overall segment agreement across all types (*trajectory match score*).

3. Results and discussion

3.1. Example segmentation results

As intended, the algorithms captured abrupt disturbance events, including fire and harvest, as well as longer-duration processes such as post-disturbance growth as well as slow loss of vigor caused by insects (Fig. 5). By flexibly recording both events and processes, the algorithm also captures situations where one precedes or follow the other, the sequence of which is often useful for interpretation: for example, the algorithm can distinguish among abrupt disturbance that follows decline, growth, or relative stability, allowing distinction between fire that burns through insect-damaged stands from fire that burns through regrowing forest. We are currently unaware of other automated approaches that can capture and characterize arbitrary shapes and durations of land cover change using a single fitting algorithm.

		TimeSync			
		Disturbance(d)	Recovery (r)	Stable(s)	No-Vertex(n)
Land-Trendr	Disturbance (d)	dd	dr	ds	dn
	Recovery (r)	rd	rr	rs	rn
	Stable (s)	sd	sr	ss	sn
	No-Vertex(n)	nd	nr	ns	nn

Fig. 4. Contingency matrix used for calculation of scoring metrics. Cells within the full 4 × 4 table were used for vertex-based scores: *Matched disturbance*, *false negative* and *positive disturbance*, *overall accuracy*, and *kappa*. For calculation of *disturbance accuracy* and *disturbance kappa*, cells with same shading were aggregated to a 2 × 2 table. For segment matching, a similar table is constructed by evaluating match for every year, not just vertices, making the no-vertex not applicable. Proportional agreement on each trajectory was calculated as $dd + rr + ss / (n_years)$, where n_years is number of years in the time series. Mean proportional agreement across all trajectories in a run yielded the *trajectory match score* for that run.

3.2. TimeSync interpretation

The TimeSync interpretation of the 388 forested plots resulted in 703 unique segments, of which one quarter were disturbance, one third recovery, and the remainder stable. Of the disturbances, 126 were harvest, 31 fire, 14 insect or pathogens, and 5 of other type. Interpreted disturbance intensity across all disturbances was approximately even distributed among magnitude categories. Definitions used for reference data interpretation and comparisons with other reference datasets are provided in our companion paper (Cohen et al., 2010).

3.3. Parameter evaluation

Although our parameter evaluation scheme involved many repetitions of the segmentation model under different parameter value combinations, we did not conduct a true sensitivity analysis or parameter optimization for two reasons. First, landscapes contain many more combinations of phenomena and timing than can be observed in a reasonably-sized reference dataset; optimization or parameter quantification would be an exercise in overfitting the particular random draw of the landscape captured in our reference points. Second, our purpose is to broadly describe the impact of the parameters on the algorithm in a manner that could be applied more generally in other ecosystems; this is better achieved through a holistic evaluation of parameter effects and groupings. We evaluated parameter effects in two ways: by parameter across all runs, and across parameters within the best runs.

For the former, we arranged all of the runs with a given parameter value as cumulative distributions ordered according to each of the eight summary metrics (Phase 1 NBR overall accuracy score shown in Fig. 6a). The left–right shift of the cumulative distributions is determined by the variation in the parameter of interest: the greater the effect of a particular parameter value, the greater the separation among runs. Although combinations of other parameter values at a given percentile along the cumulative distribution curve are not necessarily the same, the rank comparison at a given percentile rank of the curve nevertheless provides a first approximation of the effect of varying the parameter, especially if one value always outperforms the others. This rank effect can then be more concisely captured by calculating a simple ratio of accuracy scores of the two lower-ranking parameter values versus the best-ranked parameter value, within percentile bins of the distributions (Fig. 6b), which can then be expressed with a standard box-plot approach (Fig. 6c). These box plots can then be evaluated to determine which value of each parameter was best when evaluated from the perspective of any of the eight summary metrics, or whether there was no separation among runs of different parameter values (Fig. 7).

Because the cumulative distributions show the impact of only a single parameter value at a time, they cannot be used to determine which parameters determine the overall robustness of the runs when combined with the other parameters. This leads to our second parameter evaluation approach. We tallied the proportion of runs associated with each parameter value for the top 5% of all scores of overall kappa (vertex-based), matched disturbance (vertex-based), and trajectory match score (overall-agreement based). The more balanced the counts among parameter values in these best runs, the less that parameter mattered. If a single value of a parameter dominated the best runs, we interpret that parameter's value to be important in controlling overall segmentation success. We used simple stacked bar plots (Fig. 8) to evaluate parameter value dominance. Detailed evaluation of parameter effects across all scoring metrics are provided as Tables A1–A4 in Appendix 1.

Evaluating both within and cross-parameter results (summarized in Table 3 from Tables A1–A4), several patterns emerge. First, thresholding change by magnitude of change (here expressed as

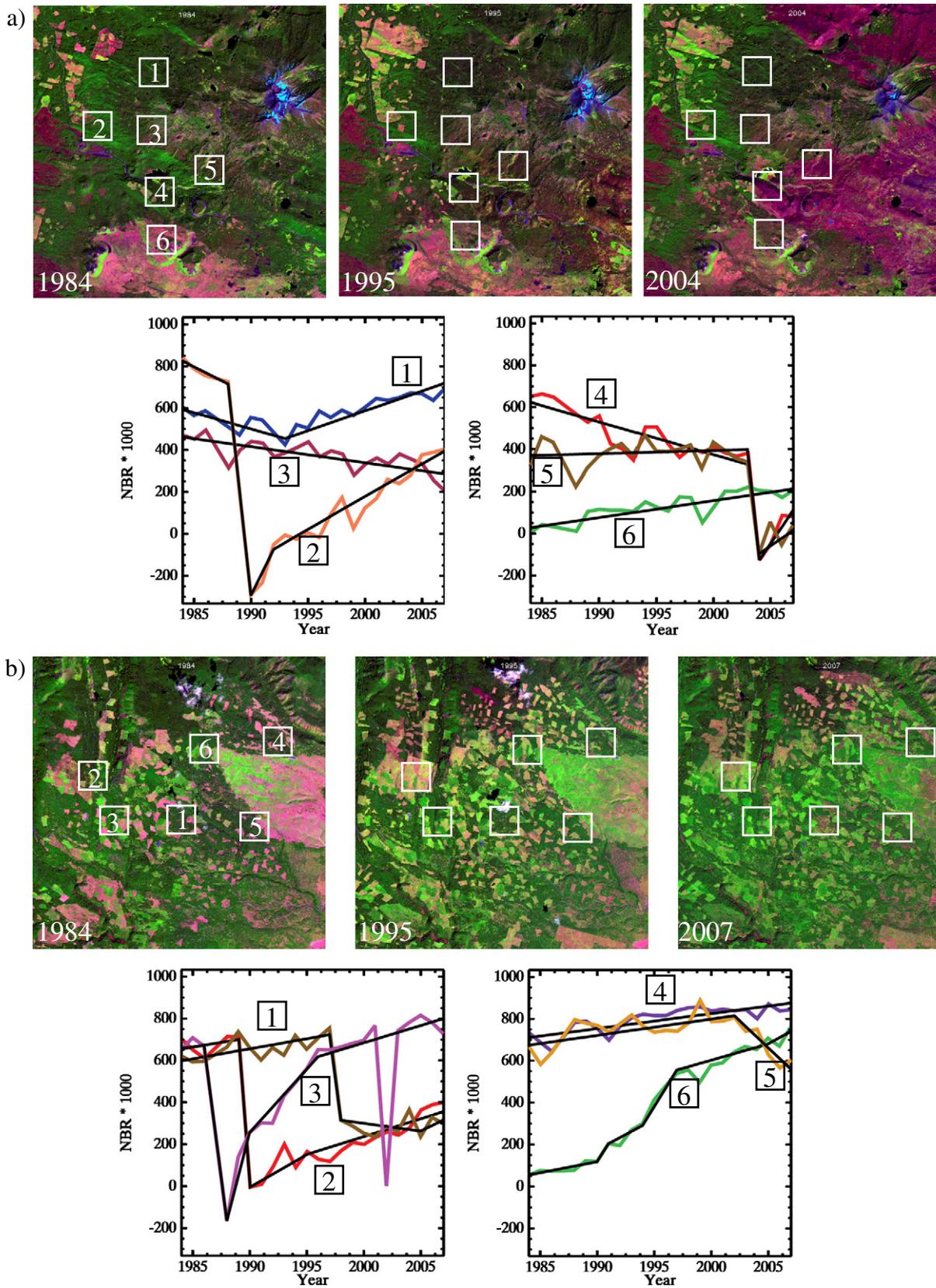


Fig. 5. Segmentation results (black lines), source spectral data (colored lines), and three years of Landsat imagery displayed using RGB = bands 5, 4, and 3. a) An area in Oregon's central Cascades shows insect-related mortality (trajectories 1 and 3), insects followed by fire (4), as well as clearcut harvest (2), stability followed by fire (5), and recovery from prior fire (6). b) Further north in the Cascades, segmentation captures harvests of different intensity and timing (1, 2, and 3) as well as regrowth (4 and 6), with low-magnitude disturbance emerging after regrowth (5).

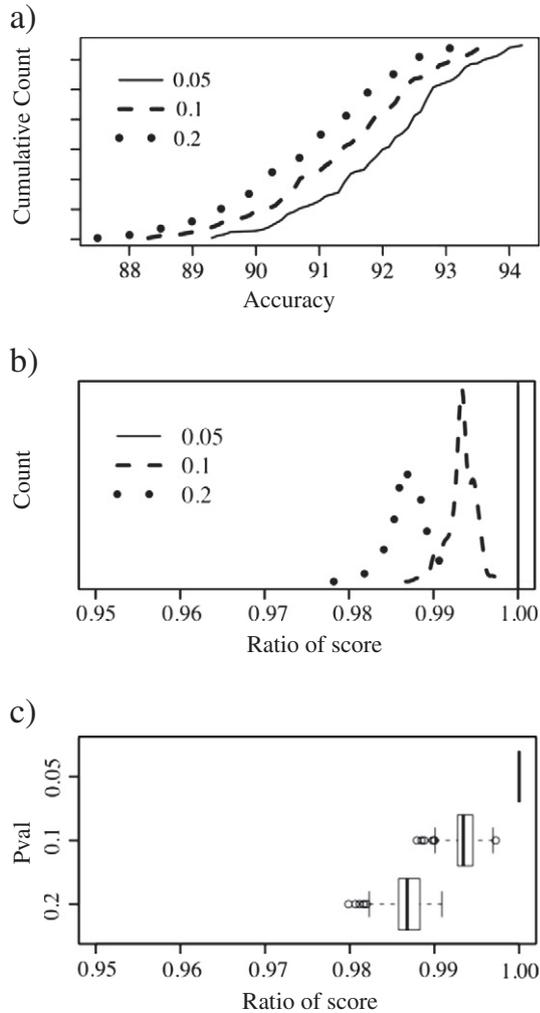


Fig. 6. An example of evaluating the effects of a single parameter across all combinations of the other control parameters. a) Distribution of overall accuracy scores for three different values of the control parameter *pval* Phase 1 NBR runs. At a given cumulative run count (y-axis), the runs with the *pval* of 0.05 have scores that always exceed the scores when *pval* is set to 0.1 or 0.2. To capture the relative rank of the three parameter values in this case, the parameter value with the highest mean rank within cumulative distribution of scores is identified as the best value; within each cumulative distribution bin, the overall accuracy score of the lesser parameter values can be divided by the score of the best parameter value, resulting in a distribution of score ratios (b) that can then be expressed concisely as a box plot (c) that distills the relative impact of the parameter values.

percent cover) always improves accuracy by diminishing false positives. This is expected, but highlights the penalty paid by this flexible fitting method against one that pre-defines the type of allowable change (such as that we reported earlier; Kennedy et al., 2007b). Second, selection of potential vertices is a key step, and must consider both regression and angle-based criteria (as indicated by a universal need to set the *vertexcountovershoot* parameter to 3). This emphasizes the importance of a strategy that identifies vertices in a forward pass through the trajectory as a separate step from the actual fitting once vertices are identified. Third, parameter settings that favor capture of disturbance also result in higher false positive counts, making them less favorable for capturing the timing of recovery or stable period or for matching overall trajectory shape.

Indeed, this latter point emphasizes a key characteristic of the algorithm corroborated by our parameter tests: the central tension in the algorithm is between high sensitivity to abrupt events and robust capture of long-term trends. Moreover, the apparent complexity of both the fitting algorithm and hence parameter choices distills to essentially two fairly simple groupings of parameter based on that

tension. Success at capturing the most disturbance is achieved primarily by increasing the maximum number of segments in play (*max_segments*) and by lowering the percent-cover thresholds (*pct_veg_loss1* and *pct_veg_loss20*). To better capture all types of change, fewer segments should be allowed, robustness of fit should be set stringently (*pval* set lower), percent-cover thresholds should be increased (particularly *pct_veg_loss20*), and the prohibition on over-fast recovery should be enforced (setting *recovery_threshold* to more stringent values).

3.4. Comparison among spectral indices

Although the purpose of our study is to evaluate effects of parameters on fitting, comparison of the spectral indices used for fitting is also instructive. We calculated the mean summary metric scores for the top 5% runs for all three spectral indices (Table 4). For our sample, NDVI performed more poorly than the two other indices, supporting the use of the short-wave infrared region for tracking change in forests (Fraser and Latifovic, 2005; Hais et al., 2009; Jin and Sader, 2005; Wilson and Sader, 2002). NBR was most sensitive to capture of disturbance events (high “matched disturbance” scores), but was also more susceptible to noise and therefore had slightly lower overall accuracy and trajectory match scores than did wetness for Phase 2 results.

3.5. Overall algorithm performance

To maintain reasonable scope, we have focused in this paper on describing our algorithms and on judging their relative performance under thousands of runs using different combinations of control parameters and using different spectral indices. The *absolute* performance is difficult to judge and compare against other change detection tools, even from the summary metrics in Table 4, because 1) the reasons for omission and commission cannot be identified without detailed evaluation of a single run, and 2) the truth set against which we compare our results (the TimeSync interpretation set) includes far more subtlety than has been available to other change algorithms for comparison to date. Therefore, we have split the evaluation of the interpretation tool and the sources of omission and commission into a companion paper (Cohen et al., 2010). Based on detailed evaluation of results from single runs, we found that the algorithm captures traditional disturbances as well or better than two-date change methods have in the past (Cohen et al., 1998), detects with reasonable robustness a wide range of other dynamics such as insect-related disturbance and growth (Fig. 5), and that errors of omission and commission are generally confined to very subtle phenomena.

3.6. Algorithm assumptions and weaknesses

Several assumptions and weaknesses not immediately evident in the summary data deserve note. First, the core segmentation approach judges change in a completely relative sense: the definition of change in any time period of a trajectory depends on spectral properties of the rest of time series, even if the absolute magnitude of change is small. This allows capture of subtle but consistent trends, but also leads to many small-magnitude false positive changes. Thus, the absolute magnitude-filtering of Phase 2 is useful. This leads to a second issue: if each year's spectral deviation is judged entirely with regard to the years before and after it, the spectral deviations of the first and last years of the trajectory are by definition more difficult to judge than deviations in all other years. This makes it imperative that cloud and shadow screening be particularly aggressive in those years. Finally, the segmentation method considers each plot as an isolated entity; the behavior of spatially-adjacent areas is not used to improve robustness of detection.

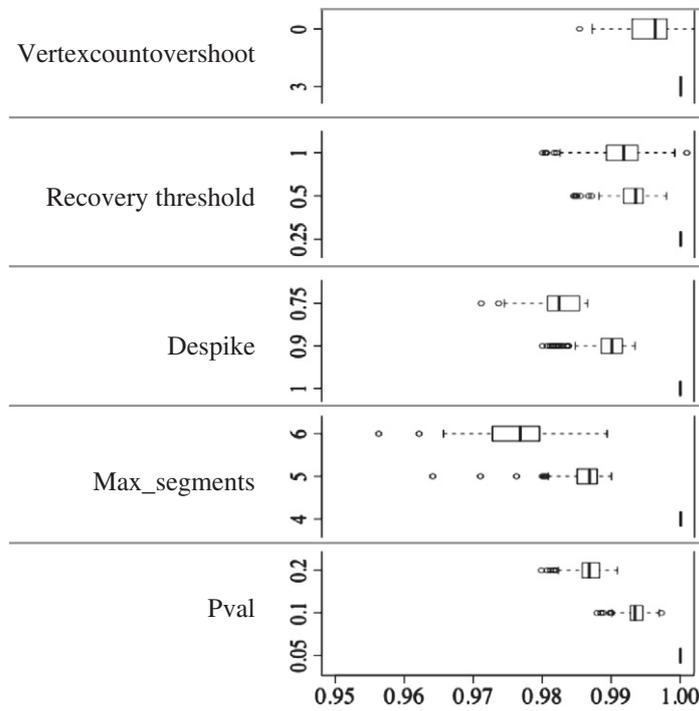


Fig. 7. Using box plots of cumulative score ratios (see Fig. 6c) to interpret how variation in parameter values affects summary metrics. The greater the separation among score distributions for different values of a parameter, the more distinct the effect of that parameter value on the summary metric. For the overall accuracy scores of Phase 1 NBR runs shown here, for example, the *pval*, *max_segments*, and *despike* parameters had clear effects, while the *vertexcountovershoot* and *recovery_threshold* parameters had lesser effects. These interpretations are captured in Table 3 for this and all other summary metrics for NBR runs.

3.7. Applicability to other ecoregions and sensors

Although the tests reported here focus on forested systems in the Pacific Northwest, U.S.A., the general conclusions appear to be applicable for LandTrendr runs we have begun applying in the northern

temperate forests of Minnesota and Wisconsin, mixed and dry forests in the Sierra Nevadas of California, xeric shrublands, woodlands, and mixed forests in the American Southwest, high elevation and boreal systems in southwest Alaska, and temperate mixed forest landscapes in the Carpathians in southeastern Europe. In all of these situations, cloud

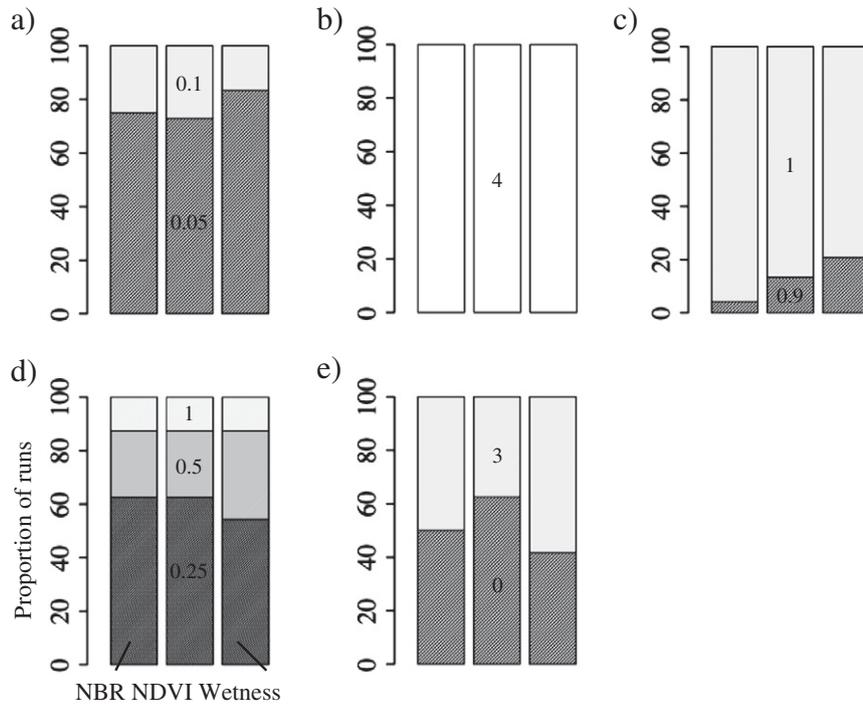


Fig. 8. Dominance of parameter values for runs scoring in the top 5% of kappa scores for all Phase 1 runs. Shown are proportions for (a) *pval*, (b) *max_segments*, (c) *despike*, (d) *recovery_threshold*, and (e) *vertexcountovershoot*. Values noted within each bar indicate the parameter value associated with that bar's proportion. When a single parameter value dominates the top 5% of runs (as in *max_segments* = 4), it is considered exclusive; otherwise parameter effects are judged by relative proportions of runs.

Table 3
Parameter effects summarized by fitting goal.

Parameter	Fitting goal		
	Timing of disturbance	Timing of all types	Overall shape
<i>despike</i>		Only minor effect; should be set to moderate value	
<i>pval</i>	Equivocal	Somewhat improved by setting to lower value	
<i>max_segments</i>	Critical to set high	Critical to set low	
<i>recovery_threshold</i>	Equivocal	Important to set low	Somewhat improved by setting low for some indices
<i>vertexcount overshoot</i>		Must be set high for all cases	
<i>pct_veg_loss1</i>	Somewhat improved by low values	Equivocal	
<i>pct_veg_loss20</i>	Improved by low values	Improved by higher values	
<i>pre_dist_cover</i>		Generally equivocal	
<i>pct_veg_gain</i>	Equivocal	Low impact, but avoid high or low values	

screening and consistency of phenological state in image dates are critical. When sufficient density of imagery across years is available, a relatively stable set of parameters can be used across all of these systems, and disturbance and growth patterns are detected robustly. Where recovery rates after disturbance are very fast, setting the recovery threshold to 1.0 (turned off) is often required. Where robust percent-cover models are not available, a first approximation based on a linear fit to spectral values at the minimum and maximum of vegetation cover appears to function well. Where persistent cloudiness or gaps in the archive significantly reduce the density of available pixels in the time series, we must reduce the number of segments sought by the algorithm, which changes the focus of change detection from yearly mapping to simple detection of change across years and detection of longer-term processes. Thus, although the ability to separate ephemeral change from real change in these systems is still an advantage, other algorithms targeted to these areas may be more appropriate if year-to-year change of high-magnitude phenomena is critical.

The temporal segmentation approach need not be limited to the Thematic Mapper sensors. Conceptually, application to time series of other sensors, including the multi-spectral scanner (MSS) or the moderate resolution imaging spectrometer (MODIS) can easily be envisioned. In these cases, however, we anticipate that the differing signal-to-noise ratios and spectral properties of the different instruments may require alteration of the despiking, the percent cover, and the *p*-value thresholds.

4. Conclusions

Segmentation of the time-domain in Landsat imagery appears to be both a feasible and a powerful approach to capture diverse land cover dynamics in the forested ecosystems tested here. The use of regression- and vertex-to-vertex-based trajectory fitting allows detection of abrupt events such as disturbance as well as longer-duration processes such as regrowth, and control parameters can be used to reduce problems associated with overfitting. Because of the variety of potential phenomena that can be detected, different control parameters are likely to be useful for different ultimate mapping

goals. The range of possible mapping is likely greater than hitherto described, and reflects the power and wisdom of investment in a consistent, long-term land monitoring satellite system, as well as the added information content accessible with free access to the entire Landsat archive.

Acknowledgments

The development and testing of the LandTrendr algorithms reported in this paper were made possible with support of the USDA Forest Service Northwest Forest Plan Effectiveness Monitoring Program, the North American Carbon Program through grants from NASA’s Terrestrial Ecology, Carbon Cycle Science, and Applied Sciences Programs, the NASA New Investigator Program, the Office of Science (BER) of the U.S. Department of Energy, and the following Inventory and Monitoring networks of the National Park Service: Southwest Alaska, Sierra Nevada, Northern Colorado Plateau, and Southern Colorado Plateau. We wish to particularly thank Dr. Melinda Moeur (Region 6, USDA Forest Service) for her vision in supporting this work from the proof of concept to the implementation phase. We also wish to thank Peder Nelson and Eric Pfaff for their work on the project, and Ray Davis for helpful evaluations of the early results from the algorithms. This work was generously supported by the USGS EROS with free, high quality Landsat data before implementation of the new data policy making such data available free of charge to everyone.

Appendix 1. Detailed parameter evaluation results

For the runs of Phase 1, two primary patterns emerge (Table A1). First, the control parameter value of 3 was always the best for the *vertexcountovershoot*. Second, the effect of varying parameter values depended on whether capturing the most disturbance was the goal (as represented by the *disturbance_match* summary metric) or whether overall accuracy was the goal (as represented by all the other summary metrics). The behavior of these metrics under parameter change was generally as expected: more stringent values

Table 4
Dominant parameter values across top 5% Phase I NBR runs evaluated using three summary metrics.

Parameter	Values	Summary metrics		
		Kappa	Matched disturbance	Trajectory match score
<i>pval</i>	0.05, 0.1, 0.2	0.05+*	0.1, 0.2 even, no 0.05	0.05++, no 0.2
<i>max_segments</i>	4, 5, 6	4+++	6+++	4++ (no 5)
<i>recovery_threshold</i>	0.25, 0.5, 1.0	0.25+	1, 0.5, no 0.25	0.25
<i>despike</i>	0.75, 0.9, 1.0	1++	0.9, no 1.0	1, no 0.75
<i>vertexcount overshoot</i>	0, 3	–	0	3++

* Degree of imbalance among parameter values: None (i), slight (number only), large (+), very large (++), exclusive (+++). Absence of a value is indicated with “no XXX,” where XXX is the parameter value.

of *pval* and *recovery_threshold* led to fewer false positives, which improved overall accuracies, but increased false negatives and reduced the absolute number of disturbances captured. Variation in parameter values caused similar responses in both the vertex-based and trajectory match scores, as well as in the disturbance and overall accuracy and kappa metrics. The same patterns of relative effect emerged for the wetness and the NDVI indices (not shown).

The best Phase 1 runs exhibited patterns (Table A2) generally consistent with those observed when considering each parameter separately. If all types of change are desired, then parameter values that avoid false positives are favored: *pval* of 0.05, *max_segments* of 4, and *recovery_threshold* of 0.25. If maximum capture of disturbance is favored, then the converse is true: *pval* of 0.1 or 0.2, *max_segments* of 6, and *recovery_threshold* of 1.0 (turned off). In most cases, the *vertexcountovershoot* parameter should be set to 3. These conclusions also hold for wetness and NDVI (not shown), with the exception of the *despike* parameter. For NBR, the best *despike* value is 1.0 (turned off), but for wetness it is either 0.9 or 0.75, and for NDVI it is dominantly 0.75.

When filtering by percent cover was introduced (Phase 2), the patterns of parameter effect were similar in some ways to Phase 1, but became more variable across summary metrics and across spectral indices (static model results; Table A3). As before, the *vertexcountovershoot* parameter value of 3 was consistently favored. Also as before, the capture of the most disturbance events was favored when *max_segments* was 6, *pval* was 0.1 or 0.2, and the *recovery_threshold* was 1 (turned off) or 0.5, while the best capture of all types of process was favored when those parameters were set at the complementary values. Unlike the Phase 1 results, however, the two disturbance summary metrics (disturbance accuracy and disturbance kappa) differed from the other metrics on the effect of *max_segments* and *despike*, with high segment counts (5 and 6) and greater spike-removal (*despike* values of 0.75 and 0.9) becoming favored in Phase 2. Variability among NBR, wetness, and NDVI also increased, particularly in the *max_segments* and *despike* parameters.

The parameters associated specifically with the percent-cover modeling (Phase 2 parameters) affected the summary metrics in generally predictable ways (also Table A3). More disturbance events were captured when percent-cover thresholds were reduced, but false positives increased and overall accuracies went down. Overall accuracies increased when *pct_veg_loss20* and *pct_veg_gain* increased, but were generally unaffected by changes in *pct_veg_loss1* or *pre_dist_cover*. For disturbance accuracy and kappa metrics, *pct_veg_gain* had little effect, but were improved when *pre_dist_cover* and *pct_veg_loss1* were increased.

For the most part, the patterns observed when the static change model was used were exactly matched when the delta change model was used (not shown). For wetness, however, there were several differences: *max_segments* had a reduced effect on both disturbance accuracy and false positives, *despike* had an increased effect on the same (with *despike* of 0.75 improving scores), and *pct_veg_gain* was strongly favored at a value of 10 (versus 3 or no effect for the static model).

The dominance patterns of parameter values in the top 5% of Phase 2 runs were generally consistent with those seen when parameters were considered separately (Table A4). The parameter *vertexcountovershoot* set to 3 was always favored; *pval* set to 0.05 was generally favored but set to any value was sufficient for capture of disturbance; *recovery_threshold*, and *max_segments* set to 0.25 and 4, respectively, favored generality and set to 1.0 and 6, respectively, favored capture of disturbance. Among parameters used for cover filtering, *pct_veg_loss1* and *pre_dist_cover* were equivocal for overall accuracy, but set to 10 and 10, respectively, favored capture of disturbance; *pct_veg_loss20* set to 10 favored generality but set lower captured disturbance. The parameter *pct_veg_gain* was not relevant for the disturbance matches, but set to 10, it favored vertex-based accuracy; set to 3 or 5 it favored overall trajectory match accuracy. As with other Phase 2 results, the *despike* showed the most variability among spectral indices and across summary metrics.

Table A1

Best parameter values across all Phase 1 NBR runs evaluated using eight summary metrics.

Parameter	Values tested	Summary metrics			
		Disturbance: false negative	Disturbance: false positive	Disturbance: matched	Trajectory match score
<i>pval</i>	0.05, 0.1, 0.2	0.1, 0.2*	0.05++	0.1, 0.2–	0.05++
<i>max_segments</i>	4, 5, 6	6++	4+	6++	4+
<i>recovery_threshold</i>	0.25, 0.5, 1.0	0.5, 1.0	0.25	0.5, 1.0+	0.25+
<i>despike</i>	0.75, 0.9, 1.0	0.75–	1	0.75, 0.9	1
<i>vertexcount overshoot</i>	0, 3	3	3+	3	3
Parameter	Values tested	Summary metrics			
		Overall: accuracy	Overall: kappa	Disturbance: accuracy	Disturbance: kappa
<i>pval</i>	0.05, 0.1, 0.2	0.05+	0.05++	0.05+	0.05+
<i>max_segments</i>	4, 5, 6	4++	4+	4+	4
<i>recovery_threshold</i>	0.25, 0.5, 1.0	0.25+	0.25+	0.25+	0.25
<i>despike</i>	0.75, 0.9, 1.0	1+	1	1	1.0, 0.9+
<i>vertexcount overshoot</i>	0, 3	3	3	3	3++

* Separation among parameter values: None (i), moderate (number only), large (+), very large (++). When two parameter values are nearly identical, both are listed, and the separation refers to the separation between that group and the third parameter.

Table A2

Dominant parameter values across top 5% Phase 1 HER runs evaluated using three summary metrics.

Parameter	Values	Summary metrics		
		Kappa	Matched disturbance	Trajectory match score
<i>pval</i>	0.05, 0.1, 0.2	0.05+*	0.1, 0.2 even, no 0.05	0.05++, no 0.2
<i>max_segments</i>	4, 5, 6	4+++	6+++	4++ (no 5)
<i>recovery_threshold</i>	0.25, 0.5, 1.0	0.25+	1, 0.5, no 0.25	0.25
<i>despike</i>	0.75, 0.9, 1.0	1++	0.9, no 1.0	1, no 0.75
<i>vertexcount overshoot</i>	0, 3	–	0	3++

* Degree of imbalance among parameter values: None (i), slight (number only), large (+), very large (++), exclusive (+++). Absence of a value is indicated with “no XXX,” where XXX is the parameter value.

Table A3
Best parameter values across all Phase 2 NBR runs evaluated using eight summary metrics.

Parameter	Values	Summary metrics								
		Disturbance: false negative			Disturbance: false positive			Disturbance: matched		
		NBR	Wetness	NDVI	NBR	Wetness	NDVI	NBR	Wetness	NDVI
<i>pval</i>	0.05, 0.1, 0.2	0.2, 0.1–	.	0.2, 0.1–	.	0.05	.	0.1, 0.2–	.	0.1, 0.2–
<i>max_segments</i>	4, 5, 6	.	6+	.	4, 5+	.	.	6++	.	.
<i>recovery_threshold</i>	0.25, 0.5, 1.0	0.5, 10	0.5, 1.0	1	0.25–	.	0.5, 1.0	0.5, 1.0	.	1
<i>despike</i>	0.75, 0.9, 1.0	0.9	0.9, 0.75	0.9, 0.75	0.9	.	0.75	0.9	0.9, 0.75	0.9
<i>vertexcount_overshoot</i>	0, 3	.	3	.	3	.	3	.	.	.
<i>pct_veg_loss1</i>	10, 15	.	10	.	15	.	.	10	.	.
<i>pct_veg_loss20</i>	3, 5, 10	.	.	.	10++
<i>pre_dist_cover</i>	10, 20, 40	.	20, 40–	.	40–	.	.	10, 20–	.	.
<i>pct_veg_gain</i>	3, 5, 10

Parameter	Trajectory match score			Overall: accuracy			Overall: kappa		
	NBR	Wetness	NDVI	NBR	Wetness	NDVI	NBR	Wetness	NDVI
<i>pval</i>	0.05+	0.05+	0.05+	.	0.05	.	.	0.05	.
<i>max_segments</i>	4, 5	4	4	.	4	.	.	4	.
<i>recovery_threshold</i>	0.25, 0.5–	0.25+	.	0.25	.	.	0.25	.	.
<i>despike</i>	.	1	.	0.9, 1.0	1	.	0.9, 0.1	1	0.9, 1.0–
<i>vertexcount_overshoot</i>	.	3+	.	.	3	.	.	3	.
<i>pct_veg_loss1</i>	.	.	15–
<i>pct_veg_loss20</i>	10	.	10	10–	.	10–	10–	.	10–
<i>pct_dist_cover</i>	.	40–
<i>pct_veg_gain</i>	.	5	.	10	.	10	10	.	10

Parameter	Disturbance: accuracy			Disturbance: kappa		
	NBR	Wetness	NDVI	NBR	Wetness	NDVI
<i>pval</i>	.	0.05	.	.	0.05	.
<i>max_segments</i>	5.6–	4.5	4–	6	5	6
<i>recovery_threshold</i>	.	0.25–
<i>despike</i>	0.9	0.75–	0.75	0.9	0.75	0.75
<i>vertexcount_overshoot</i>	.	3	.	.	3	.
<i>pct_veg_loss1</i>	.	15	.	.	15	.
<i>pct_veg_loss20</i>	.	10+	.	.	10+	.
<i>pre_dist_cover</i>	40–	40+	.	.	40+	.
<i>pct_veg_gain</i>

*Separation among parameter values. None (i), slight (–), moderate (number only), large (+), very large (++). When two parameter values are nearly identical, both are listed, and the separation refers to the separation between that group and the third.

Table A4
Dominant parameter values in top 5% Phase 2 runs evaluated using three summary metrics.

Parameter	Values	Kappa						Matched disturbance						Trajectory match score						
		Static cover model			Delta cover model			Static cover model			Delta cover model			Static cover model			Delta cover model			
		NBR	Wetness	NDVI	NBR	Wetness	NDVI	NBR	Wetness	NDVI	NBR	Wetness	NDVI	NBR	Wetness	NDVI	NBR	Wetness	NDVI	
<i>pval</i>	0.05, 0.1, 0.2	0.05+	.	0.05+	0.05+	.	.	0.05+	.	.	0.05+	
<i>max_segments</i>	4, 5, 6	4++	.	4++	.	6+++	.	6+++	.	4	.	4	.	4	.	4	.	4	.	
<i>recovery_threshold</i>	0.25, 0.5, 1.0	0.25	0.25	0.25	0.25	0.25+	1, 0.5	1, 0.5	.	0.25	.	0.25	.	0.25	.	0.25	.	0.25	.	
<i>despike</i>	0.75, 0.9, 1.0	1	1+	.	1	0.9	0.9	0.75	0.9	0.9	0.75	0.9	0.9	0.75	1, 0.9	1, 0.9	.	1, 0.9	.	
<i>vertexcount_overshoot</i>	0, 3	3++	.	3++	.	.	.	3	.	3++	.	3++	.	3++	.	3++	.	3++	.	
<i>pct_veg_loss1</i>	10, 15	10++	.	10++
<i>pct_veg_loss20</i>	3, 5, 10	10	.	10+	.	10+	3, 5	3, 5	.	10	.	10+	10	.	10+	10+	.	10+	.	
<i>pre_dist_cover</i>	10, 20, 40	10, 20	.	10, 20
<i>pct_veg_gain</i>	3, 5, 10	10	5	10	10	5+	5	3+	5+	5	5	5	5	5	10+	

*Dominance of indicated parameter(s): No dominance (i), slight (number only), large (+), very large (++), and exclusive (+++).

References

- Canty, M. J., Nielsen, A. A., & Schmidt, M. (2004). Automatic radiometric normalization of multitemporal satellite imagery. *Remote Sensing of Environment*, 91, 441–451.
- Chavez, P. S., Jr. (1996). Image-based atmospheric corrections – Revisited and improved. *Photogrammetric Engineering and Remote Sensing*, 62, 1025–1036.
- Cohen, W. B., Fiorella, M., Gray, J., Helmer, E. H., & Anderson, K. (1998). An efficient and accurate method for mapping forest clearcuts in the Pacific Northwest using Landsat imagery. *Photogrammetric Engineering and Remote Sensing*, 64, 293–300.
- Cohen, W. B., & Goward, S. N. (2004). Landsat's role in ecological applications of remote sensing. *Bioscience*, 54, 535–545.
- Cohen, W. B., Spies, T., Alig, R. J., Oetter, D. R., Maiersperger, T. K., & Fiorella, M. (2002). Characterizing 23 years (1972–95) of stand replacement disturbance in western Oregon forests with Landsat imagery. *Ecosystems*, 5, 122–137.
- Cohen, W. B., Yang, Z., & Kennedy, R. E. (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.
- Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: A review. *International Journal of Remote Sensing*, 25, 1565–1596.
- Crist, E. P. (1985). A TM tasseled cap equivalent transformation for reflectance factor data. *Remote Sensing of Environment*, 17, 301–306.
- Franklin, J. F., & Dyrness, C. T. (1988). Natural vegetation of Oregon and Washington. Corvallis, OR: OSU Press.
- Fraser, R. H., & Latifovic, R. (2005). Mapping insect-induced tree defoliation and mortality using coarse spatial resolution satellite imagery. *International Journal of Remote Sensing*, 26, 193–200.
- García-Haro, F. J., Gilabert, M. A., & Meliá, J. (2001). Monitoring fire-affected areas using Thematic Mapper data. *International Journal of Remote Sensing*, 22, 533–549.
- Goodwin, N. R., Coops, N. C., Wulder, M. A., & Gillanders, S. (2008). Estimation of insect infestation dynamics using a temporal sequence of Landsat data. *Remote Sensing of Environment*, 112, 3680–3689.
- Hais, M., Jonášová, M., Langhammer, J., & Kučera, T. (2009). Comparison of two types of forest disturbance using multitemporal Landsat TM/ETM+ imagery and field vegetation data. *Remote Sensing of Environment*, 113, 835–845.
- Hayes, D. J., & Sader, S. A. (2001). Comparison of change-detection techniques for monitoring tropical forest clearing and vegetation regrowth in a time series. *Photogrammetric Engineering and Remote Sensing*, 67, 1067–1075.
- Haynes, R. W., Bormann, B. T., Lee, D. C., & Martin, J. R. (2006). *Northwest Forest Plan – The first 10 years (1994–2003): Synthesis of monitoring and research results, PNW-GTR-651*. Portland: US Department of Agriculture, Forest Service OR 292.
- Healey, S. P., Yang, Z., Cohen, W. B., & Pierce, D. J. (2006). Application of two regression-based methods to estimate the effects of partial harvest on forest structure using Landsat data. *Remote Sensing of Environment*, 101, 115–126.
- Hicke, J. A., Logan, J. A., Powell, J., & Ojima, D. S. (2006). Changing temperatures influence suitability for modeled mountain pine beetle (*Dendroctonus ponderosae*) outbreaks in the western United States. *Journal of Geophysical Research, Biogeosciences*, 111, G02019. doi:10.1029/2005JG000101.
- Hostert, P., Roder, A., & Hill, J. (2003). Coupling spectral unmixing and trend analysis for monitoring of long-term vegetation dynamics in Mediterranean rangelands. *Remote Sensing of Environment*, 87, 183–197.
- Huang, C., Goward, S. N., Masek, J. G., Thomas, N., Zhu, Z., & Vogelmann, J. E. (2010). An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment*, 114, 183–198.
- Jin, S., & Sader, S. A. (2005). Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote Sensing of Environment*, 94, 364–372.
- Kennedy, R. E., & Cohen, W. B. (2003). Automated designation of tie-points for image-to-image coregistration. *International Journal of Remote Sensing*, 24, 3467–3490.
- Kennedy, R. E., Cohen, W. B., Kirschbaum, A. A., & Haunreiter, E. (2007a). Protocol for Landsat-based monitoring of landscape dynamics at North Coast and Cascades Network parks. *U.S. Geological Survey Techniques and Methods* <http://pubs.usgs.gov/tm/2007/tm2g1/>: USGS Biological Resources Division.
- Kennedy, R. E., Cohen, W. B., & Schroeder, T. A. (2007b). Trajectory-based change detection for automated characterization of forest disturbance dynamics. *Remote Sensing of Environment*, 110, 370–386.
- Lawrence, R., & Ripple, W. J. (1999). Calculating change curves for multitemporal satellite imagery: Mount St. Helens 1980–1995. *Remote Sensing of Environment*, 67, 309–319.
- Logan, J. A., Regniere, J., & Powell, J. A. (2003). Assessing the impacts of global warming on forest pest dynamics. *Frontiers in Ecology and the Environment*, 1, 130–137.
- Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing*, 25, 2365–2407.
- Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (2000). *Spatial tessellations: Concepts and applications of Voronoi diagrams*. Chichester; New York: Wiley.
- Olsson, H. (2009). A method for using Landsat time series for monitoring young plantations in boreal forests. *International Journal of Remote Sensing*, 30, 5117–5131.
- Röder, A., Udelhoven, T., Hill, J., del Barrio, G., & Tsiourlis, G. (2008). Trend analysis of Landsat-TM and -ETM+ imagery to monitor grazing impact in a rangeland ecosystem in Northern Greece. *Remote Sensing of Environment*, 112, 2863–2875.
- Ruefenacht, B., Finco, M. V., Neslon, M. D., Czaplewski, R., Helmer, E. H., Blackard, J. A., et al. (2008). Conterminous U.S. and Alaska forest type mapping using forest inventory and analysis data. *Photogrammetric Engineering and Remote Sensing*, 74, 1379–1388.
- Schroeder, T. A., Cohen, W. B., Song, C., Canty, M. J., & Yang, Z. (2006). Radiometric calibration of Landsat data for characterization of early successional forest patterns in western Oregon. *Remote Sensing of Environment*, 103, 16–26.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127–150.
- USDI, & USDA (1994). *Record of decision on management of habitat for late-successional and old-growth forest related species within the range of the Northern Spotted Owl (Northwest Forest Plan)*. : US Department of Agriculture, Forest Service & US Department of Interior, Bureau of Land Management, Portland, OR.
- van Wageningen, J. W., Root, R. R., & Key, C. H. (2004). Comparison of AVIRIS and Landsat ETM+ detection capabilities for burn severity. *Remote Sensing of Environment*, 92, 397–408.
- Viedma, O., Meliá, J., Segarra, D., & García-Haro, J. (1997). Modeling rates of ecosystem recovery after fires by using Landsat TM data. *Remote Sensing of Environment*, 61, 383–398.
- Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science*, 277, 494–499.
- Vogelmann, J. E., Tolck, B., & Zhu, Z. (2009). Monitoring forest changes in the southwestern United States using multitemporal Landsat data. *Remote Sensing of Environment*, 113, 1739–1748.
- Wilson, E. H., & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, 80, 385–396.
- Wulder, M. A., White, J. C., Goward, S. N., Masek, J. G., Irons, J. R., Herold, M., et al. (2008). Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sensing of Environment*, 112, 955–969.