Remote Sensing of Environment 114 (2010) 2911–2924

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync — Tools for calibration and validation

Warren B. Cohen [a,*], Zhiqiang Yang [b], Robert Kennedy [b]

[a] USDA Forest Service Pacific Northwest Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331, USA
[b] Department of Forest Science, Oregon State University, 321 Richardson Hall, Corvallis, OR 97331, USA

ARTICLE INFO

Article history:
Received 20 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
Calibration
Validation

ABSTRACT

Availability of free, high quality Landsat data portends a new era in remote sensing change detection. Using dense (~annual) Landsat time series (LTS), we can now characterize vegetation change over large areas at an annual time step and at the spatial grain of anthropogenic disturbance. Additionally, we expect more accurate detection of subtle disturbances and improved characterization in terms of both timing and intensity. For Landsat change detection in this new era of dense LTS, new detection algorithms are required, and new approaches are needed to calibrate those algorithms and to examine the veracity of their output. This paper addresses that need by presenting a new tool called TimeSync for syncing algorithm and human interpretations of LTS. The tool consists of four components: (1) a chip window within which an area of user-defined size around an area of interest (i.e., plot) is displayed as a time series of image chips which are viewed simultaneously, (2) a trajectory window within which the plot spectral properties are displayed as a trajectory of Landsat band reflectance or index through time in any band or index desired, (3) a Google Earth window where a recent high-resolution image of the plot and its neighborhood can be viewed for context, and (4) an Access database where observations about the LTS for the plot of interest are entered. In this paper, we describe how to use TimeSync to collect data over forested plots in Oregon and Washington, USA, examine the data collected with it, and then compare those data with the output from a new LTS algorithm, LandTrendr, described in a companion paper (Kennedy et al., 2010). For any given plot, both TimeSync and LandTrendr partitioned its spectral trajectory into linear sequential segments. Depending on the direction of spectral change associated with any given segment in a trajectory, the segment was assigned a label of disturbance, recovery, or stable. Each segment was associated with a start and end vertex which describe its duration. We explore a variety of ways to summarize the trajectory data and compare those summaries derived from both TimeSync and LandTrendr. One comparison, involving start vertex date and segment label, provides a direct linkage to existing change detection validation approaches that rely on contingency (error) matrices and kappa statistics. All other comparisons are unique to this study, and provide a rich set of means by which to examine algorithm veracity. One of the strengths of TimeSync is its flexibility with respect to sample design, particularly the ability to sample an area of interest with statistical validity through space and time. This is in comparison to the use of existing reference data (e.g., field or airphoto data), which, at best, exist for only parts of the area of interest, for only specific time periods, or are restricted thematically. The extant data, even though biased in their representation, can be used to ascertain the veracity of TimeSync interpretation of change. We demonstrate that process here, learning that what we cannot see with TimeSync are those changes that are not expressed in the forest canopy (e.g., pre-commercial harvest or understory burning) and that these extant reference datasets have numerous omissions that render them less than desirable for representing truth.

Published by Elsevier Inc.

1. Introduction

The earth's terrestrial biosphere has changed dramatically during the past several centuries, due primarily to land use (Lambin et al., 2001). In addition, climate change associated with both CO_2 and non-CO_2 greenhouse gases (Hansen et al., 2000) is having an increasing effect on biological trends (Parmesan and Yohe, 2003). Understanding causes and consequences of biosphere change requires monitoring and modeling (Running et al., 2004). At a global scale, characterizing changes in vegetation has become routine with the aid of remote sensing (DeFries et al., 2000; Friedl et al., 2002). At more local-to-regional scales, Landsat has long been the workhorse sensor for monitoring vegetation change

* Corresponding author. Tel.: +1 541 750 7322; fax: +1 541 758 7760.
E-mail address: warren.cohen@oregonstate.edu (W.B. Cohen).

0034-4257/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.rse.2010.07.010

(Cohen & Goward, 2004; Healey et al., 2008). This paper concerns new trends in Landsat-based vegetation monitoring that involve analyses of large numbers of annual time series.

Forest change detection with Landsat has a history as long as the Landsat program itself (Heller, 1975). With few exceptions, most research and application have focused on 3–10 years or greater image intervals and one or two Landsat scenes (Singh, 1989; Coppin & Bauer, 1994; Masek, 2001; Lunetta et al., 2004; Masek & Collatz, 2006). Some studies have encompassed large numbers of Landsat scenes, but by necessity were limited to relatively coarse time intervals (Skole & Tucker, 1993; Cohen et al., 2002; Healey et al., 2008; Masek et al., 2008). The few studies that have evaluated annual or near-annual datasets have mostly involved only one Landsat scene (Kaufmann & Seto, 2001; Healey et al., 2006; Schroeder et al., 2006).

Dense Landsat time series (LTS) (i.e., approximately annual interval) for forest change detection over multiple scenes is poised to become the norm. There are several important reasons for this. First, there is currently tremendous need for temporally and spatially detailed forest change information over vast areas for carbon modeling (Turner et al., 2007; Goward et al., 2008) and forest management and policy considerations (Moeur et al., 2005). Second, the entire historic Landsat archive in the USGS holdings is now available online for free in highly preprocessed, standard format (Landsat Science Team, 2008). Third, advanced automated algorithms capable of processing annual LTS are currently being developed, tested, and operationalized (Kennedy et al., 2007 and 2010; Huang et al., 2010).

The development of automated algorithms that use LTS for detection of forest change over large geographic areas creates a new problem for validation. For any remote sensing change detection exercise, validation reference data can be difficult or costly to obtain because historic observations for the period(s) of interest can be rare and/or challenging to retrieve. With LTS, this problem is greatly exacerbated by an annual interval over a period potentially as long as the Landsat archive. Moreover, there are currently no comprehensive methods for corroborating the output from automated LTS algorithms that produce unprecedented types of change information.

1.1. Objectives

This paper introduces TimeSync, a LTS visualization and data collection tool developed to accommodate the need for assessment of the veracity of output from automated LTS algorithms. One such algorithm, LandTrendr, is described in a companion paper (Kennedy et al., 2010). LandTrendr uses temporal segmentation to capture a wide range of forest change processes ranging from abrupt disturbances and chronic mortality to varying rates of vegetation recovery. As LandTrendr and other novel processing algorithms use dense LTS to extend the types of change that can be detected, methods to corroborate or assess algorithm performance must be similarly extended. With TimeSync, this is accomplished by using the same data as the automated algorithm (i.e., the LTS), but by performing a comparable, human-interpreted segmentation that is otherwise totally independent of the automated interpretation.

The objectives of this study are to: (1) introduce and summarize LTS data collected with TimeSync, (2) illustrate how to compare human and automated interpretations of LTS data, and (3) compare TimeSync data with other, extant forest change data. We focus on a sample of four Landsat scenes in the Pacific Northwest, USA, that represent a broad range of forest cover and cover change processes.

2. TimeSync

TimeSync is an image time series visualization and data collection tool that consists of four components: an image chip window, a trajectory window, Google Earth (http://earth.google.com/), and a Microsoft Access database (Fig. 1). When TimeSync is invoked, both the image chip and trajectory windows are opened. From the later window, the user selects an LST database file to work with and selects a plot from the database to examine. For each image date in the LTS, the chip window (Fig. 2) displays a plot within a fixed 3×3-pixel square at the center of each chip. The number of pixels displayed in each chip is selectable via zooming, allowing one to see each pixel in the plot, as well as the greater landscape around the plot for spatial context. The

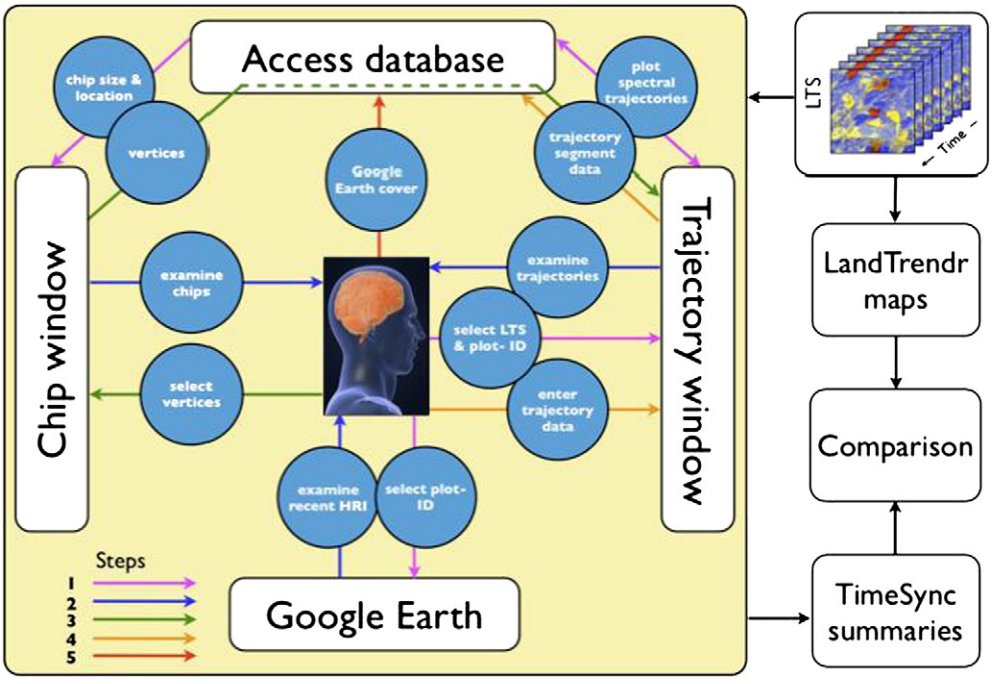


Fig. 1. How validation with TimeSync works. TimeSync is controlled through the chip and trajectory windows, and through a Google Earth window. The analyst first selects a LTS and plot to examine. The plot is examined in all three windows: as a series of image chips, as a spectral time series in bands and indices of interest, and on a high-resolution image (HRI). Vertices are selected by clicking the mouse on the image chips associated with the dates of the vertices selected. The database then draws the segment lines in the trajectory window, where relevant data are entered. When desired, HRI imagery is evaluated for cover percentages and entered into the database. Comparisons between TimeSync and LandTrendr involve several new validation approaches, as described in the text.

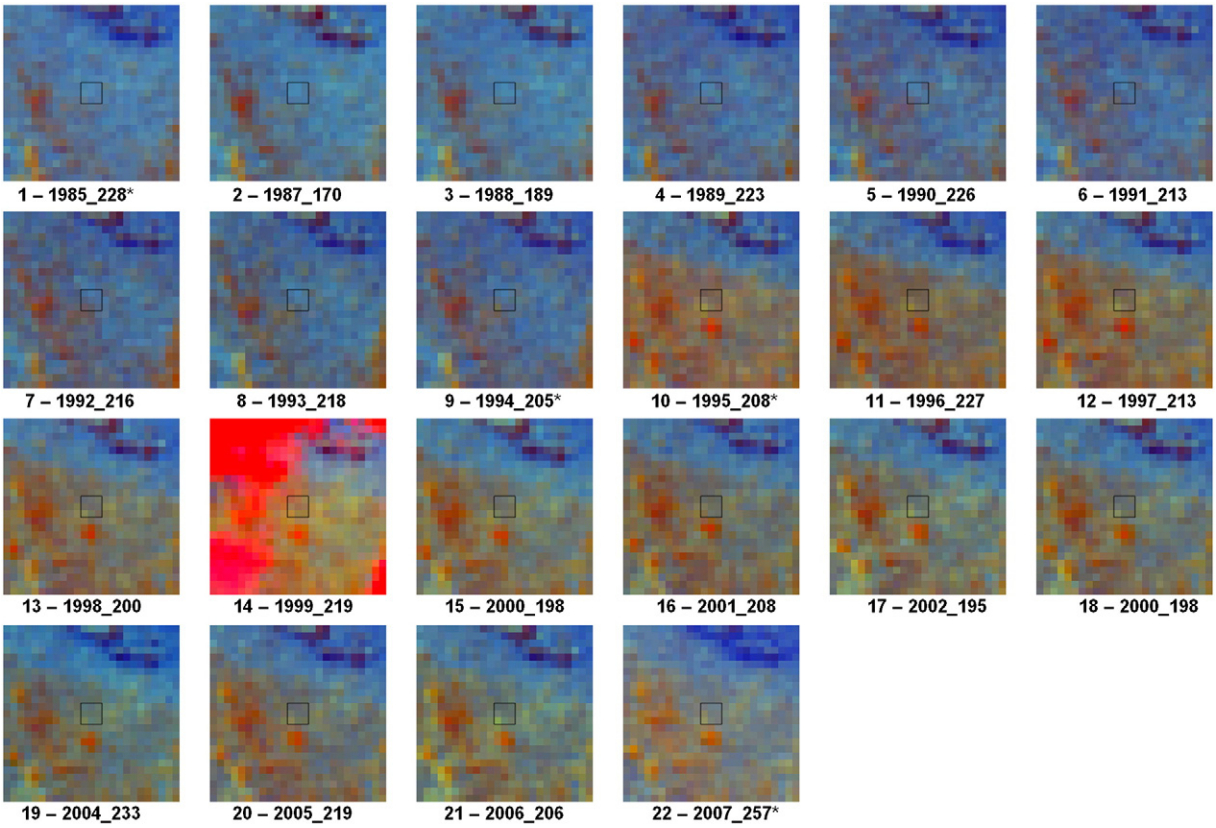
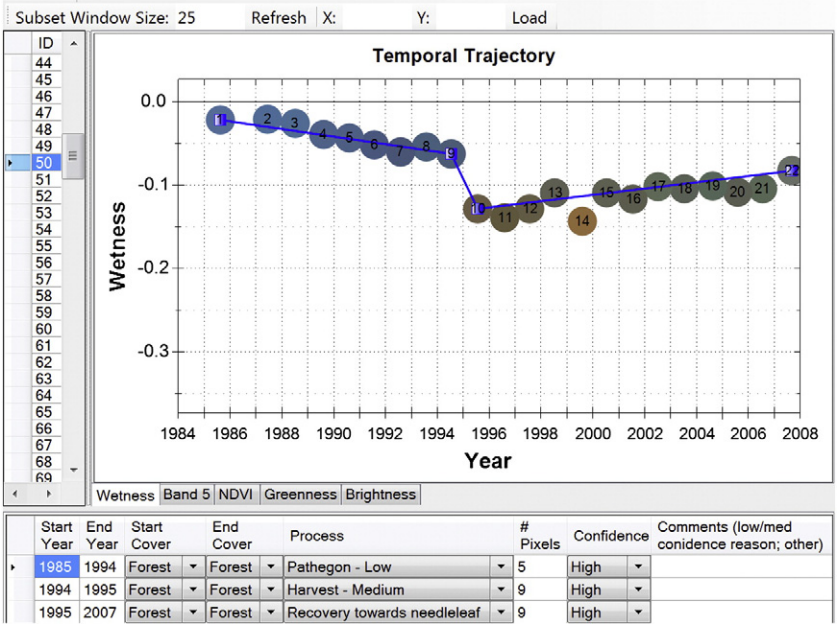


Fig. 2. Image chips (1985–2007) displayed in the chip window for a plot and its neighborhood. The 3 × 3-pixel plot is displayed at the center of each chip, which in this case contains 35 × 35 pixels.

trajectory window displays the 3 × 3-mean plot spectral response over time in any Landsat band or user-defined spectral index (Fig. 3), and the analyst can toggle amongst the various bands and indices to determine the plot's temporal spectral behavior in the context of forest change viewed in the chip window. Regardless of band or index examined, the plot spectral response is shown as a series of colored dots, where the colors are the 3 × 3-mean plot Tasseled Cap brightness (red), greenness (green), and wetness (blue) values for each image date. Any band or index combination could be used, but we rely on the Tasseled Cap because of its general utility (Cohen & Goward, 2004).

To collect data, each plot's trajectory is handled as a series of segments, where a given segment has two vertices (start date and end



Start Year	End Year	Start Cover	End Cover	Process	# Pixels	Confidence	Comments (low/med conidence reason; other)
1985	1994	Forest	Forest	Pathegon - Low	5	High	
1994	1995	Forest	Forest	Harvest - Medium	9	High	
1995	2007	Forest	Forest	Recovery towards needleleaf	9	High	

Fig. 3. The trajectory window for the plot shown in Fig. 2 after the vertices were selected and data entered. Displayed is the wetness index. Tabs are available to display and toggle among various other indices or bands. This trajectory consists of three segments: a low intensity, long duration disturbance associated with pathogens, followed by a medium intensity harvest, and a recovery towards needleleaf forest.

date) and is associated with a linear trend in spectral response that can be attributed to disturbance, recovery, or stability (Fig. 3). From drop-down lists, the analyst selects items that describe each segment, while simultaneously viewing high-resolution images (multi-date if available) in Google Earth for the same locations, with the plot boundaries displayed within Google Earth (Fig. 1). Each vertex of a given segment is labeled according to its land cover/use class, and the process (disturbance, recovery, or stability) describing the segment is selected (Table 1). For disturbance, the agent (e.g., harvest, fire, insect, etc.), relative intensity (low, medium, and high—the magnitude of spectral change from pre-disturbance condition to the spectral condition of bare soil), and number of pixels (from 1 to 9) affected are selected. As vertices are defined, the segments are plotted in the trajectory window (Fig. 3). There are confidence fields (high, medium, low) and a comments field for every segment.

3. Methods

3.1. Study area

This study was conducted over an area represented by the intersection of the Northwest Forest Plan area in the States of Oregon and Washington, USA (Moeur et al., 2005) and four Landsat scenes (see Fig. 1, Kennedy et al., 2010). The Landsat scene areas used (Table 2) were limited to the non-overlapping portions of Landsat scenes, known as Theissen scene areas (TSAs) derived using a Voronoi tessellation (Kennedy et al., 2010).

The total area of the study was nearly 53,000 km^2, most of which is forested (Table 2). The four scenes included in this study represent the strong climatic and topographic gradients of the region, as well as high diversity in vegetation due to a wide array of private and public owners and associated management practices and natural disturbance regimes (Moeur et al., 2005). Coniferous forests dominate most of the area, with temperate rainforest near the coast and drier montane forests in the interior (Franklin & Dyrness, 1988). The dominant species are Douglas-fir, western hemlock, western red cedar, and lodgepole and ponderosa pine. For more comprehensive descriptions refer to Franklin and Dyrness (1988) and Moeur et al. (2005).

3.2. Sample design

Sampling design flexibility is a key attraction of the TimeSync approach. Traditional reference data are typically available from sources that sample only a particular ownership or a subset of years over which change must be validated. With TimeSync, the entire landscape can be sampled, and because it utilizes the entire LTS, all years can be evaluated.

Table 1
Land use/cover classes available for labeling each vertex and the process that can be attributed to each segment in a given trajectory within TimeSync.

Land use/cover class	Process
Forest	Stable
Needleleaf forest	Recovery
Broadleaf forest	Recovery towards needleleaf
Mixed forest	Recovery towards broadleaf
Woodland	Woody encroachment
Shrubland	Harvest
Grassland	Fire
Agriculture	Pathogen
Wetland	Blowdown
Urban	Other
Snow/ice	
Barren	
Water	
Other	

Table 2
Basic characteristics of, and number of plot for, the four LTS used in this study.

WRS path-row	Area (km^2)	Number of plots	Forest area (km^2)
45–27	8826	89	5392
45–29	7939	82	7292
46–29	19,382	200	12,340
47–27	16,690	172	14,490
Totals	52,837	543	39,514

To illustrate this flexibility here, we used a fully random sample, stratified by TSA. The goal was to sample approximately 200 plots per full TSA, with no bias toward any land cover class, ownership, or time period. The number of plots per TSA was scaled to the proportion of the full TSA overlapping the study area. For this study, a total of 543 plots was interpreted over the four scenes (Table 2). This number of plots was a compromise between the time it takes to collect the information with TimeSync and a larger number that might be required to more fully sample every type of change encountered.

Another sample design is possible, such as one stratified by change type, which may make better use of a smaller sample size, but may be less representative of the map as a whole. Ultimately, the sample design should be chosen to meet specific objectives, and TimeSync facilitates the implementation of any desired design because the full population of pixels over the full time period can be sampled.

3.3. Vegetation cover interpretation

For LandTrendr, vegetation cover and cover change models are used for filtering potential segmentations (see Kennedy et al., 2010), requiring that cover be characterized for forested plots. This is accomplished within TimeSync, using the high-resolution imagery available for viewing via Google Earth. Of the total number of interpreted plots, 388 were forested near the end of the time series (~2005), when high spatial resolution true-color images were available in Google Earth. For each forested plot we quantified four components: live tree, other live vegetation, shadow, and open (dead vegetation and barren) cover (summing to 100%). This was done ocularly by an experienced photo-interpreter, without reference to any aids. For a discussion of how these data are used by LandTrendr see Kennedy et al. (2010).

3.4. Comparisons with LandTrendr

3.4.1. LandTrendr data used

LandTrendr has numerous parameters that define how it segments each pixel in an LTS. As described by Kennedy et al. (2010), 150,000 combinations of LandTrendr parameters were tested against the 388 TimeSync-interpreted forested plots to determine which sets of parameters were the most robust to the wide variety of changes observed.

We chose two runs as examples for evaluation in this paper. One run was selected for each of two indices evaluated: the normalized burn ratio (NBR, Key and Benson, 2005) and tasseled cap wetness (Crist & Cicone, 1984). Wetness and NBR both take advantage of the contrast between shortwave- and near-infrared reflectance, which we have found to be the most advantageous multispectral contrast for characterizing vegetated systems (Cohen & Goward, 2004). Wetness is a well-established and studied index in this context, and is a linear combination of all six Landsat reflectance bands. Although NBR is relatively new, it has been demonstrated to be particularly useful for monitoring wildfire in coniferous systems, such as those studied here. As NBR is a ratio-based index, in this study, we were interested in examining NBR's value for change detection more generally, in contrast to wetness. The parameters of the two runs were chosen to roughly balance omission and commission errors while maintaining

high agreement; they were not selected as the best run for any particular change detection goal. As such, the results presented here do not represent a validation of the best output from LandTrendr.

LandTrendr output was produced for each plot interpreted with TimeSync. To match the TimeSync interpretations, which were based on 9-pixel plots, the 3×3-mean index (NBR and wetness) values for each plot were calculated for each year of the LTS. LandTrendr was then run on these plot-level spectral index trajectories. The trajectories were fitted and labeled as described by Kennedy et al. (2010). The approach used was consistent with TimeSync, in that each trajectory was fitted as a series of segments having vertices and labels. TimeSync and LandTrendr rely on direction of spectral change to label segments as disturbance, recovery, or stable. Disturbance and recovery in relation to direction of spectral change is spectral index-dependent. For both NBR and wetness, across our study area, disturbance is almost always associated with negative change and recovery with positive change.

3.4.2. Validation approaches and metrics

LandTrendr provides a wide array of products (Kennedy et al., 2010) based on various combinations of segment label (e.g., disturbance and recovery) and duration, vertex-year, and change magnitude. This unprecedented product richness suite can be evaluated for errors in a variety of ways, heretofore either irrelevant or impossible without use of LTS. In this paper we explore a number of ways to assess error in LandTrendr output as a demonstration of different strategies for error assessment of LTS products using TimeSync. Our assessments are of three basic types: vertex-based, segment-based, and match scores.

3.4.2.1. Vertex-based assessments. Vertices define the beginnings and endings of segments within a given trajectory. Beginning (or start) vertices define the years when spectral trajectories change direction. As such, vertex-based approaches evaluate agreement between the human interpreter and automated algorithm in terms of when changes begin. Because segments have labels, we can partition this agreement into segment category labels (*disturbance*, *recovery*, and *stable*) associated with the segment tied to each start vertex. The vertex-based approach requires a fourth category (*no vertex*) for when no vertex was observed by either the human or the algorithm, or both. Traditional error matrices are essentially of this type; i.e., examine if a change mapped as occurring during a specific interval of time actually occur during this interval, if at all. However, recovery and stability have not heretofore been included in such matrices.

To construct error matrices, it was necessary to tally the segment label for each year of a given segmented spectral trajectory (or plot) and accumulate the results across plots. There were four potential table category labels: (i) *disturbance*, (ii) *recovery*, or (iii) *stable* if a start vertex was present, or (iv) *no vertex* if no vertex was present, for the two independent segmentations (i.e., TimeSync and LandTrendr). Tallies for each cell of this 4×4 contingency table were used to calculate standard accuracy-assessment statistics (Congalton, 1991; Cohen et al., 2002), considering the TimeSync assignments as truth.

To derive contingency tables, four different rule sets were used. (1) All four table category labels were used, and a vertex-year match-rule was strictly enforced (i.e., no relaxation in the exact timing of vertices was allowed). (2) When segments of either recovery or stable follow each other they were collapsed so only the earliest vertex was retained for evaluation. This focused the table on three categories: (i) *disturbance*, (ii) *recovery/stable*, and (iii) *no vertex*, minimizing any mismatches between the human and the algorithm in the distinction between recovery and stable. (3 and 4) Like 1 and 2 above, except that we allowed for a relaxation in agreement on timing of vertices.

Relaxation allowed for agreement when vertices of the same type were recorded by the algorithm and human interpreter as being within one year of each other plus an additional allowable offset of one-quarter of the duration of the segment following the vertex. This timing relaxation allowed for two conditions for which we did not want to penalize LandTrendr. These included (a) situations where the human and the algorithm agree about the occurrence of long duration segments with only subtle spectral change at the start vertex, but disagree on the exact start year and (b) agree about the occurrence of an abrupt disturbance, but because of thin cloud, cloud shadow or haze the algorithm could not identify the disturbance until the following year.

3.4.2.2. Segment-based assessments. Segment-based approaches are independent of timing, focusing instead on numbers and labels of segments. This affords an assessment of the degree of over- or under-fitting of the automated algorithm, both independent of and with attention to the specific segment labels. The most basic comparison ignored timing and label, comparing only the number of segments across plots. This was done both for all three categories and for two categories (disturbance vs. no disturbance). Another comparison summarized the number of segments across plots by category (disturbance, recovery, and stable).

3.4.2.3. Match scores. Match scores focus primarily on segment labels. Trajectory match scores summarize label agreement for every year of each trajectory. To calculate this score, for each year of a given trajectory, a 1 or a 0 was assigned when the category labels matched or did not match, respectively. Proportion was calculated as the sum of the 1s over the total number of years in the trajectory. This was examined both across and by category (for both the three and two segment category cases above). This focused our attention away from matching segment start time to duration of segment overlap. This novel assessment was not relevant in prior, more interval-based assessments that considered disturbances as events without duration and did not consider recovery as a process that was explicitly mapped.

A final type of match score, the summary score, is an integrated measure of the agreement between the algorithm and reference data, in terms of how the area or region under study has changed over the full time period of analysis. At all times, any given plot was in one of three segment categories (or states): disturbance, recovery, or stable. To calculate the summary score we derived the proportion of time each plot was in each category, and calculated the mean proportions across plots. Unlike other comparisons, this assessment completely ignores timing of agreement and numbers of segments. It is a very general assessment of agreement in terms of landscape dynamics without specific regard to timing or location.

3.4.3. Understanding agreement and outliers

In addition to providing a basis for quantifying agreement between an automated algorithm and reference data, TimeSync was used to evaluate disagreement, as a means to understand under what conditions disagreement occurs. We re-examined with TimeSync all occurrences of vertex disagreement about disturbance (both false positive and false negative), noting the most likely cause for disagreement.

We also examined vertex agreement and disagreement with respect to magnitude of spectral change, by producing histograms of spectral change magnitude for segments where TimeSync and LandTrendr agreed that there was a disturbance (agree disturbance) or a recovery (agree recovery) and where there was disagreement about disturbance (false positives and negatives).

Finally, to understand the likely causes of plots having low trajectory match scores, we examined all plots with trajectory match scores below 0.8 and noted the likely cause of disagreement.

3.5. Secondary validation

The primary reason we developed TimeSync was to fill the void in calibration and validation data for forest change detection algorithms and maps. Field visits today are not particularly useful for studying

Table 3
Number of TimeSync segments interpreted, by category, type and intensity, over forested plots.

Segment category by intensity (disturbance only)	Number	Per disturbance type	Per disturbance intensity	Percent
Harvest-high	50	–	–	–
Harvest-medium	38	–	–	–
Harvest-low	38	–	–	–
Harvest	–	126	–	0.72
Fire-high	7	–	–	–
Fire-medium	19	–	–	–
Fire-low	5	–	–	–
Fire		31	–	0.18
Pathogen-high	0	–	–	–
Pathogen-medium	6	–	–	–
Pathogen-low	8	–	–	–
Pathogen		14	–	0.08
Other-high	2	–	–	–
Other-medium	0	–	–	–
Other-low	3	–	–	–
Other	–	5	–	0.03
Disturbance type total	–	176	–	1.00
High	–	–	59	0.34
Medium	–	–	63	0.35
Low	–	–	54	0.31
Disturbance intensity total	–	–	176	1.00
Disturbance total	176	–	–	0.25
Recovery total	250	–	–	0.36
Stable total	277	–	–	0.39
Segment total	703	–	–	1.00

when, at what intensity, or by what agent a plot was disturbed in the past. We can accumulate existing field data from historic surveys and use historic airphotos, but gathering these data is costly and cumbersome, and they may or may not exist for the areas under study. Moreover, they do not facilitate the use of a statistically unbiased sample design. All of these disadvantages associated with using extant data immediately dissolve when using the exact same Landsat time series as used by an automated algorithm. The problem, however, is that we cannot simply assume that a human LTS-interpreter makes perfectly accurate decisions regarding the changes that have occurred on any given plot.

To check our interpretations against other, independent observations, where those observations existed, we accumulated and evaluated several other datasets. These included burn severity maps from the Monitoring Trends in Burn Severity (MTBS) project (Eidenshink et al., 2007), the US Forest Service Forest Health Monitoring pathogen dataset (http://fhm.fs.fed.us), the US Bureau of Land Management Forest Cover/Operation Inventory dataset (http://www.blm.gov/or/gis/data-details.php?data=ds000045) and US Forest Service (USFS) Forest Activities Tracking System datasets available via the USFS intranet. These datasets were spatially intersected with our plots, the relevant forest change information extracted, and cross-tabulations constructed. We declared a match (i.e., agreement) if the plot intersected with a given disturbance polygon from a given ancillary dataset, and the datasets agreed on timing within one year. Otherwise "not disturbed" was declared for that observation.

4. Results

In this section, we first present the results of LandTrendr using the NBR index (often simply referred to as NBR, for brevity), to illustrate the different methods of assessment. We then present the results for LandTrendr using the wetness index, as a means of illustrating how TimeSync can be used to compare the relative value of different indices for characterizing forest change.

4.1. Segment-based assessments

Across the 388 forested plots interpreted with TimeSync, there were a total of 703 individual segments (Table 3). Of these, 25% were labeled disturbance, and 36% and 39% were labeled recovery and spectrally stable, respectively. The 176 segments identified as disturbance were mostly associated with harvest activity (72%), followed by fire (18%), pathogens (8%) and other (3%), which included utility right-of-way clearing through a forested plot, road maintenance clearing, tree death in years subsequent to a fire event, and unknown causes. There was a fairly uniform distribution of disturbance intensity levels, with 34%, 35%, and 31% associated with high, medium, and low.

The number of distinct segments per plot identified with TimeSync was highly variable, ranging from one to seven for the three segment categories (Table 4). Over one-third of all segments were within single segment plots, all of which were in the stable category. More than one-half of the segments were in plots identified as having three or fewer segments.

The number of segments fitted by the automated algorithm relative to the number fitted using TimeSync provides an indication of the degree to which the algorithm under-fit or over-fit the time series of the plots evaluated. The LandTrendr NBR run used here characterized a total 962 segments, or 37% more than TimeSync (Table 4). For the two segment category case, NBR and TimeSync were in very close agreement on total number of segments. The number of segments by category reveals that over-fitting by LandTrendr for these runs, relative to TimeSync, was exclusively associated with recovery and stable segments (Fig. 4).

4.2. Vertex-based assessments

Error (or contingency or agreement) matrices were derived using four different rulesets, distinguished by number of table categories and degree of relaxation in agreement on timing. The most finely resolved comparison used four table categories (disturbance, recovery, and stable vertices and no vertex) and no relaxation of timing (Table 5). In this strict comparison, NBR exhibited a disturbance omission rate that was 8% greater than the commission rate. This is consistent with the segment-based assessment, which indicated slight tendency towards omission of disturbance segments (Fig. 4).

Table 4
Number of segments across plots. Totals are numbers of plots weighted by numbers of segments.

	3 categories			2 categories		
Number of segments	TimeSync	LandTrendr (NBR)	LandTrendr (wetness)	TimeSync	LandTrendr (NBR)	LandTrendr (wetness)
1	247	140	212	254	249	262
2	26	68	53	21	20	26
3	79	79	65	83	95	81
4	19	60	38	14	13	10
5	12	37	16	12	10	8
6	4	4	4	3	1	1
7	1	0	0	1	0	0
Total segments	703	962	769	686	682	643

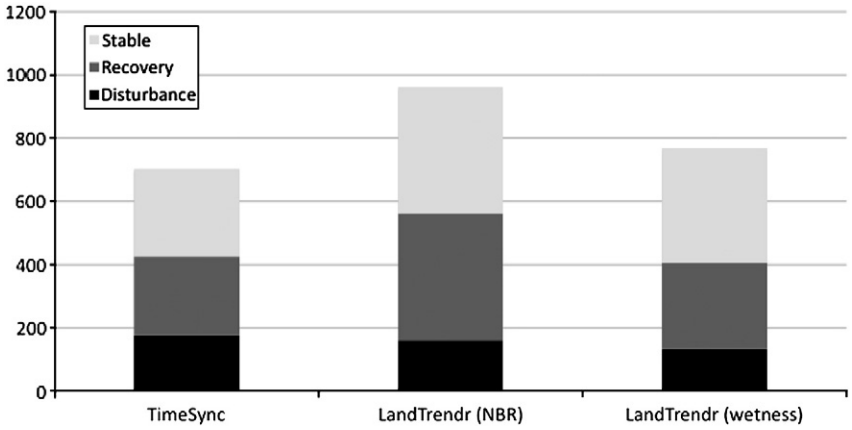


Fig. 4. Number of segments across plots, by category for TimeSync and LandTrendr using both NBR and wetness.

Similarly, we also see here, that NBR overestimated the number of recovery and stable segments relative to TimeSync, with commission rates being 23% and 21% greater than omission rates, respectively. Much of the confusion for the LandTrendr runs, relative to TimeSync, was associated with mislabeling recovery and stable segments, as revealed in the low omission and commission rates for the aggregated recovery and stable case (Table 6).

Allowing for a small degree of timing relaxation greatly improved the results for LandTrendr. The disturbance vertex omission rate for NBR dropped by 13% and the commission rate dropped by 11% (Table 7) relative to the use of a strict timing rule (Table 5). Similarly, omission and commission rates for stable and recovery vertices were also greatly reduced. When timing relaxation and non-disturbance vertex aggregation were considered together (Table 8), error rates were generally quite low.

Table 5
LandTrendr vs. TimeSync vertex-year agreement for disturbance, recovery, and stable as separate categories (restricted year matching).

TimeSync	Disturbance	Recovery	Stable	No vertex	LT omission
LandTrendr (NBR)					
Disturbance	104	2	6	64	0.41
Recovery	0	160	19	70	0.36
Stable	6	39	188	44	0.32
No vertex	46	191	188	7911	0.05
LT Commission	0.33	0.59	0.53	0.02	
LandTrendr (wetness)					
Disturbance	92	1	5	78	0.48
Recovery	2	134	38	75	0.46
Stable	9	21	214	33	0.23
No vertex	32	112	104	8088	0.03
LT commission	0.31	0.50	0.41	0.02	

Table 6
LandTrendr vs. TimeSync vertex-year agreement for disturbance and a combined recovery and stable category (restricted year matching).

TimeSync	Disturbance	Recovery/stable	No vertex	LT omission
LandTrendr (NBR)				
Disturbance	104	0	72	0.41
Recovery/stable	5	443	62	0.13
No vertex	47	54	8251	0.01
LT commission	0.33	0.11	0.02	
LandTrendr (wetness)				
Disturbance	92	1	83	0.48
Recovery/stable	9	427	74	0.16
No vertex	34	54	8264	0.01
LT commission	0.32	0.11	0.02	

Overall accuracy was high (over 90%) in all cases (Table 9), largely because of the large number of agreement observations for the no vertex table category. Kappa statistics are more revealing, in that the

Table 7
LandTrendr vs. TimeSync vertex-year agreement for disturbance, recovery, and stable as separate categories (relaxed year matching).

TimeSync	Disturbance	Recovery	Stable	No Vertex	LT Omission
LandTrendr (NBR)					
Disturbance	126	3	7	40	0.28
Recovery	0	214	20	16	0.14
Stable	6	42	204	25	0.26
No vertex	29	141	170	7994	0.04
LT Commission	0.22	0.47	0.49	0.01	
LandTrendr (wetness)					
Disturbance	116	4	7	49	0.34
Recovery	2	184	40	24	0.26
Stable	9	22	223	23	0.19
No vertex	6	63	93	8172	0.02
LT commission	0.13	0.33	0.39	0.01	

Table 8
LandTrendr vs. TimeSync vertex-year agreement for disturbance and a combined recovery and stable category (relaxed year matching).

TimeSync	Disturbance	Recovery/stable	No vertex	LT omission
LandTrendr (NBR)				
Disturbance	126	0	50	0.28
Recovery/stable	5	477	28	0.06
No vertex	30	44	8277	0.01
LT commission	0.22	0.08	0.01	
LandTrendr (wetness)				
Disturbance	116	1	59	0.34
Recovery/stable	9	466	35	0.09
No vertex	8	43	8300	0.01
LT commission	0.13	0.09	0.01	

Table 9
Overall and kappa agreement statistics for the four cases examined: four table categories, three table categories, and restricted and relaxed year matching.

	4 table categories		3 table categories	
	Overall	Kappa	Overall	Kappa
Unrelaxed				
NBR	92.5	56.5	97.3	80.9
Wetness	94.4	63.2	97.2	79.2
Relaxed				
NBR	94.5	68.1	98.3	87.8
Wetness	96.2	75.4	98.3	87.6

Table 10
LandTrendr (NBR) vs. TimeSync vertex-year omission rates (relaxed) by disturbance type.

LandTrendr (NBR)							
TimeSync	Disturbance	Recovery	Stable	No vertex	LT omission	LT omission by type	LT omission by intensity
Harvest-high	45	0	1	4	0.10		
Harvest-medium	28	0	2	8	0.26		
Harvest-low	14	3	4	17	0.63		
Harvest-all	87	3	7	29		0.31	
Fire-high	6	0	0	1	0.14		
Fire-medium	17	0	0	2	0.11		
Fire-low	5	0	0	0	0.00		
Fire-all	28	0	0	3		0.10	
Pathogen-medium	4	0	0	2	0.33		
Pathogen-low	4	0	0	4	0.50		
Pathogen-all	8	0	0	6		0.43	
Other-high	2	0	0	0	0.00		
Other-low	1	0	0	2	0.67		
Other-all	3	0	0	2		0.40	
All-high	53	0	1	5			0.10
All-medium	49	0	2	12			0.22
All-low	24	3	4	23			0.56

effect of a disparity in the number of observations among table categories is minimized.

Focusing on omission, we begin to understand the limits of an automated algorithm for detecting disturbance in the presence of noise from both the forest system and sensing environment. The most important finding here is how the omission rates change as a function of intensity of disturbance. The NBR run evaluated here omitted only 10% of all high intensity TimeSync-identified disturbances, with increasing rates for medium (22%) and low (56%) intensity disturbances (Table 10).

4.3. Match scores

Trajectory match scores measured the proportion of time where there was agreement between TimeSync and LandTrendr in the segment label for a given plot. These scores are expected to be high if the time series contain long segments (as the generally low number of segments per plot in Table 4 indicates) and there is good vertex agreement (as indicated in the vertex match matrices of Tables 5–8). As such, in general, trajectory match scores indicate a very high degree of concurrence between TimeSync and LandTrendr, especially for the aggregated segment category case, with less than 6% of the plots having scores less than 0.80 for NBR (Fig. 5). For the three segment category case, this percentages was higher, as expected.

By segment category, match scores for NBR ranged between 61%, and 75% (Fig. 6). As these summaries represent the temporal overlap between LandTrendr and TimeSync, they suggest that, for example, 61% of the time that the study areas were in a state of disturbance, as declared by TimeSync, the data from the LandTrendr runs examined here agreed with that assessment.

Knowing specifically when and where changes have happened across a region is critical for several important applications. However, for very broad, general assessments, such as comparing disturbance regimes across regions and using these data in tabular form for carbon dynamics assessments, it may be sufficient simply to know what proportion of time an average forested plot is in a disturbed state. In summary match score terms, LandTrendr and TimeSync agreed quite well for NBR, with less than 4% disagreement for all segment types (Fig. 7).

4.4. Outlier vertices and plots

To better understand the reasons for disagreement about disturbance between TimeSync and the LandTrendr runs used here, we re-examined the false negative and positive disturbance vertices for the relaxed timing case (Table 7). With relaxed timing, there were 50 false negative disturbances for NBR. Not surprisingly, we can see that the proportion of false negatives was a function of numbers of

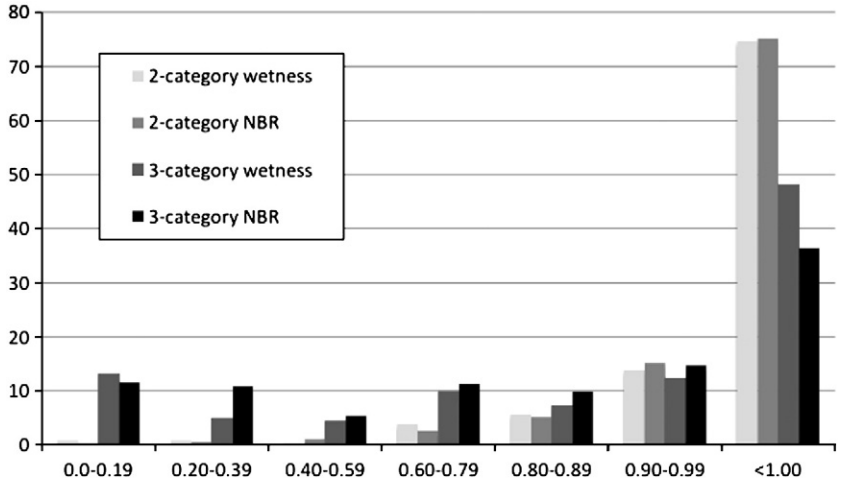


Fig. 5. Cross-category plot-level trajectory match score distributions (% of observations) for both the three category (disturbance, recovery, and stable) and the two-category (disturbance, recovery/stable) cases.

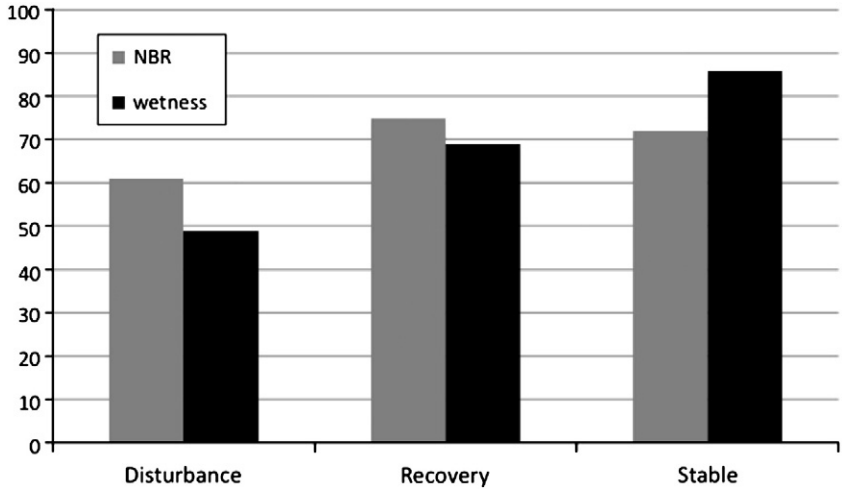


Fig. 6. Trajectory match score distributions by segment category (% of observations).

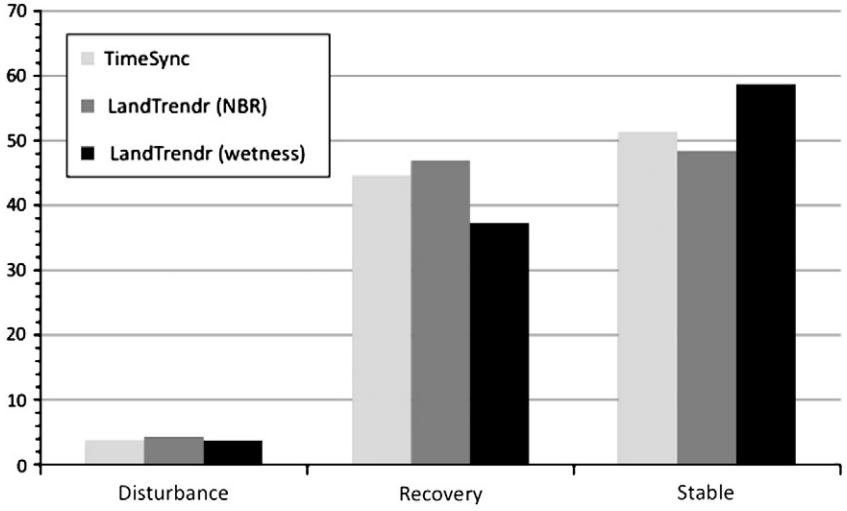


Fig. 7. Summary match scores, as a percent of an average plot trajectory in the three segment categories.

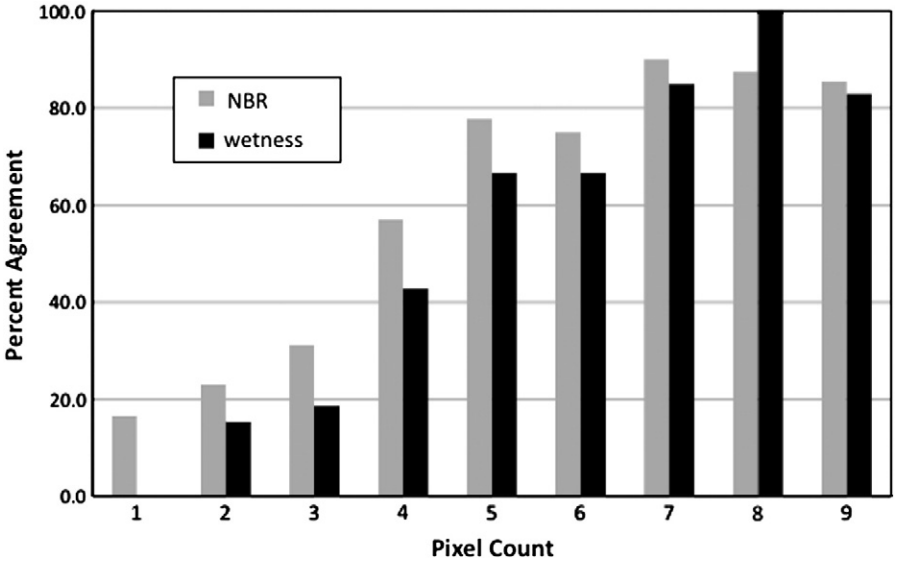


Fig. 8. Percent agreement about disturbance vertices as a function of numbers of pixels affected.

pixels disturbed, as identified by TimeSync (Fig. 8). In particular, when the number of pixels disturbed was in the minority (<5) for a given disturbance vertex, we see that the proportion of false negatives varied from about 42% (4 pixels) to 84% (1 pixel) of the vertices associated with disturbance for NBR. This suggests that the dominant reason LandTrendr missed TimeSync-identified disturbances was that the spectral change magnitude for those disturbances was relatively subtle.

A more direct examination of plot-level spectral change magnitude is to examine histograms for observations of agree disturbance, agree recovery, and false negative and positive disturbances, with respect to TimeSync (Fig. 9). Here we see that the agree disturbance category is strongly dominated by negative spectral change and that agree recovery exhibits positive spectral change. The false negative distribution had a strong modal tendency towards zero spectral change magnitude, which is the most likely reason for LandTrendr not detecting these disturbances.

There were 35 NBR false positive disturbances (Table 7). Our reexamination of these occurrences revealed that these were predominantly from residual clouds and shadows, ephemeral snow cover in the transitional snow zone, phenology, spatial misregistration among image dates, and changes in sun angle and its related topographic effect. All of the false positives exhibited a very subtle spectral change magnitude, that as a distribution tended towards slightly negative, as would be expected for a disturbance (Fig. 9).

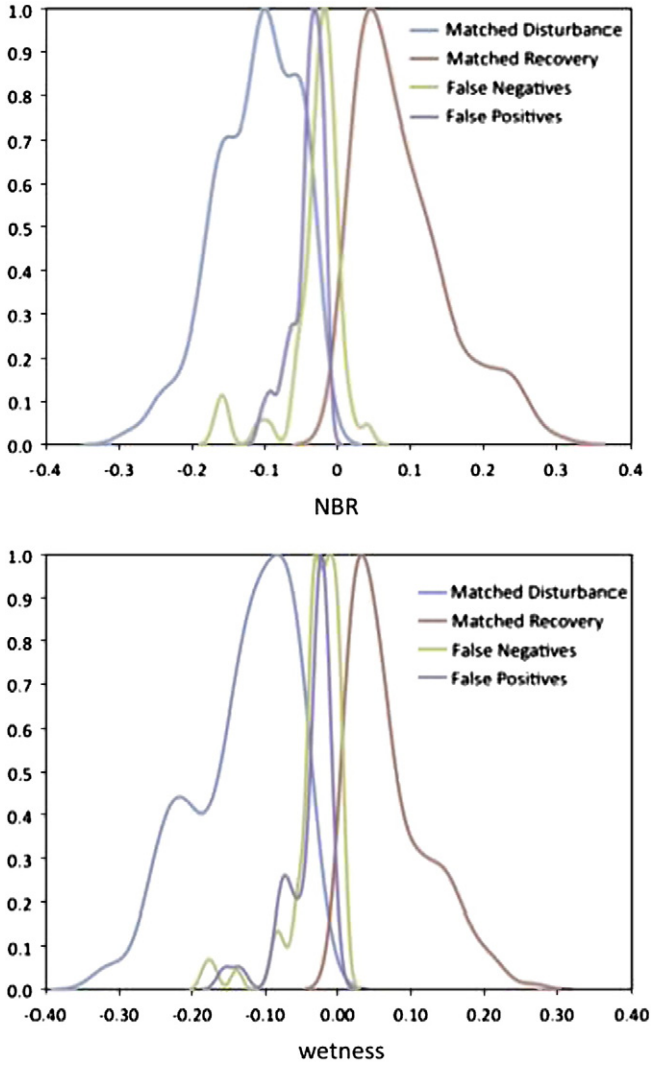


Fig. 9. Normalized frequency distributions for vertex agreement (disturbance and recovery) and disagreement (false positive and false negative disturbance) relative to TimeSync, in terms of NBR and wetness change magnitudes.

Of the trajectory match scores with values less than 0.80, most of these were described by TimeSync as having one or two long segments of recovery and/or stability, where the distinction between spectral stability and spectral change associated with recovery was subtle; i.e., near the saturation of spectral response. It is not surprising that an automated algorithm and a human interpreter would disagree in these cases.

4.5. NBR vs. wetness

Results from the LandTrendr wetness run highlight the different behaviors of the NBR and wetness indices in the context of forest change detection using LTS. With wetness, LandTrendr identified 9% more segments than the TimeSync interpreter, far lower than NBR for the three categories of change (Table 4). This lesser tendency to over-fit trajectories was mostly expressed in the lower number of recovery segments (Fig. 4). However, there was also a greater tendency than NBR to under-fit disturbance relative to the TimeSync interpreter. These observations are confirmed with the vertex-based agreement matrices (Tables 5 and 7) that reveal greater disturbance omission bias and the lesser commission bias for recovery with use of wetness. When the recovery and stable classes are combined (Tables 6 and 8), the most significant difference between the two LandTrendr runs remains the greater disturbance omission error associated with wetness. In terms of overall accuracies and kappas (Table 9), when the distinction between recovery and stability are important, wetness performs better. This is consistent with the observation regarding numbers of segments fit for each spectral index (Table 4). However, when this distinction is unimportant, the two indices performed approximately the same (Table 9). Focusing on omission rates, the greatest difference between NBR (Table 10) and wetness (Table 11), with respect to disturbance intensity, was the higher rate of omission (74%) for wetness associated with low intensity disturbances. This reveals that the main difference between NBR and wetness, as expressed in the vertex agreement matrices for disturbance, was associated with lower intensity disturbances.

Considering whole trajectories, using the trajectory match scores, the difference between the indices was relatively subtle, but most pronounced for the three category case (Fig. 5). The highest scores were associated with wetness, and lower scores were associated with NBR. By segment category (Fig. 6), wetness was more closely matched with TimeSync for the stable category and NBR was more closely matched for the disturbance category. In terms of summary match scores, LandTrendr and TimeSync agreed quite well (Fig. 7). By this measure, for disturbance, wetness actually performed better than NBR, being in near-perfect balance with TimeSync interpretations. The greatest distinction was the relatively greater over-characterization by wetness of stability and under-characterization of recovery. The greater sensitivity of NBR to disturbance was again evident when examining vertex agreement as a function of number of pixels disturbed (Fig. 8).

4.6. Secondary validation

TimeSync and MTBS agreed 98% about fire occurrence (Table 12). One low intensity MTBS fire was labeled as harvest using TimeSync, and was later confirmed to be fire by reexamining the plot with TimeSync. Five fires not within the MTBS database, but identified using TimeSync, were confirmed to be fires using alternative USFS data that included small area fires, below the MTBS threshold size of 1000 acres (Ray Davis, pers. com.).

Comparing the other databases (referred to here as Integrated FACTS–FOI–FHM) with TimeSync, we learned what the limits are of human interpretations of Landsat time series data for detecting forest change (Table 13). The one commercial harvest we could not detect with TimeSync was a salvage harvest after a high intensity wildfire.

Table 11
LandTrendr (wetness) vs. TimeSync vertex-year omission rates (relaxed) by disturbance type.

LandTrendr (wetness)							
TimeSync	Disturbance	Recovery	Stable	No vertex	LT omission	LT omission by type	LT omission by intensity
Harvest-high	46	0	1	4	0.10		
Harvest-medium	29	1	0	8	0.24		
Harvest-low	10	1	3	23	0.73		
Harvest-all	85	2	4	35		0.33	
Fire-high	6	0	0	1	0.14		
Fire-medium	15	0	2	2	0.21		
Fire-low	2	2	0	1	0.60		
Fire-all	23	2	2	4		0.26	
Pathogen-medium	4	0	0	2	0.33		
Pathogen-low	1	0	1	6	0.88		
Pathogen-all	5	0	1	8		0.64	
Other-high	2	0	0	0	0.00		
Other-low	1	0	0	2	0.67		
Other-all	3	0	0	2		0.40	
All-high	54	0	1	5			0.10
All-medium	48	1	2	12			0.24
All-low	14	3	4	32			0.74

The seven missed pre-commercial harvest were all described as partial harvests that occurred in the understory, and were thus not evident when viewed from above the canopy. Similarly, the two understory, prescribed burns, were described as ground fires under the forest canopy.

We also learned from this comparison that of the 10 TimeSync-identified harvests not contained in the secondary datasets, 8 could be clearly confirmed by TimeSync (as illustrated in Fig. 10), and simply represent errors in those datasets. The other two were identified by TimeSync as low intensity, and although upon reexamination we confirmed these were real, secondary data to support this claim do not exist. Similarly, although forest clearing associated with road building and/or maintenance is supposed to be contained in these datasets, the three identified with TimeSync were absent from those datasets. Agreement about pathogens was not as good as for other disturbance agents (Tables 10 and 11), which is at least partially due to the challenges of using FHM data for this type of assessment (i.e., FHM data are polygon based and known to have some spatial misregistration problems). Another reason is that pathogen disturbances tend to be of lower intensity, on average, than harvest and fire.

5. Discussion and conclusions

5.1. New perspectives on validation of change maps

Change detection using Landsat imagery is undergoing a major paradigm shift due to the convergence of a need for more temporally detailed information over larger areas, the free availability of data from the US archive, and the emergence of automated Landsat time series (LTS) algorithms. Automated algorithms like LandTrendr (Kennedy et al., 2010) and Vegetation Change Tracker (VCT) (Huang et al., 2010) are designed to exploit the Landsat archive by taking advantage of high temporal densities of data that enhance a relatively low signal to noise ratio commonly associated with comparing lower temporal density datasets. Because the change maps derived from these new algorithms are considerably richer in temporal detail than their predecessor maps, the challenge of calibrating and validating them is significantly increased. We developed TimeSync to address this challenge.

Table 12
Matrix of agreement between TimeSync and MTBS for fire disturbances.

MTBS (Intensity)					
TimeSync	Low	Medium	High	Not disturbed	Agreement
Harvest	1				n/a
Fire	8	10	4	5	081
Not disturbed				360	1
Agreement	0.89	1	1	0.99	0.98

Collecting historic information about forest change can be a difficult and costly endeavor. Moreover, those data do not exist everywhere, nor at a temporal detail fine enough to satisfy the needs of LTS map validation. Using TimeSync, which can ingest the exact same datasets as those used by the automated algorithm, we have the best opportunity to assess the quality of LTS maps using a statistically valid sample design. Because the data are free and already processed using the automated algorithm, the only cost of validation data collection is the time it takes to collect and summarize data for the required number of plots. The remaining challenge is in assessing the quality of the change detection calls made by the TimeSync analyst. For this, we can take advantage of whatever existing historic information there is, in a plot-by-plot comparison, without major concerns about statistical sampling constraints. This strategy of primary map validation with TimeSync, followed by a secondary validation of TimeSync with independent datasets, was successfully demonstrated in this paper.

Although we speak here of validation, it is important to highlight that the truth about historic events can be very difficult to ascertain. Extending photo-interpretation techniques to LTS facilitated visual detection of a large proportion of relevant historic change processes in the forest systems under study here, as use of the secondary validation datasets demonstrated. However, using these datasets we also learned what the limits of photo-interpretation techniques are, including detection of processes that do not sufficiently express themselves in the upper canopy layers (e.g., pre-commercial harvests), or are relatively subtle and immediately follow a major change event (e.g., high intensity wildfire followed by salvage logging). Furthermore, because the extant, secondary datasets we used here had erroneous omissions (e.g., as much as one-half of the harvests identified with TimeSync), we cannot simply use these data as the ultimate truth source to determine the efficacy of TimeSync interpretations. We propose a perspective that recognizes map validation more in terms of its agreement with independent assessments and datasets. Because we know from this study, at least for the forests examined here, that TimeSync interpretations were quite accurate for changes expressed in the upper canopy, it was fair to use those interpretations as the reference against which to compare automated LTS algorithm output. Although useful for helping us understand what the TimeSync analyst could not detect (i.e., false negatives), the secondary datasets had clear limitations for identifying TimeSync-detected disturbances that were not real (i.e., false positives).

Table 13
Matrix of agreement between TimeSync and the integrated FACTS–FOI–FHM dataset for a variety of disturbance types.

Integrated FACTS–FOI–FHM							
TimeSync	Harvest (commercial)	Harvest (pre-commercial)	Understory burn	Insect	Road	Not disturbed	Agreement
Harvest	8	2				10	0.5
Insect				7		1	0.88
Rood						3	0
Not disturbed	1	7	2			143	0.93
Agreement	0.89	0.22	0	1	0	0.91	0.87

5.2. New capabilities for validation of change maps

Dense LTS change products require new methods for assessing their quality. Using TimeSync, we took the first steps across that frontier by examining sample LandTrendr output from several novel viewpoints.

Segment-based assessments describe the degree of over- or under-fitting of trajectories. As LandTrendr results depend on the values of a host of parameters, this is a basic means of determining the sensitivity of results to various parameters of the algorithm. Sensitivity can be assessed across and by categories of change. However, these assessments are in bulk, in that they do not compare specific segments from LandTrendr with specific segments from TimeSync.

Vertex-based assessments are similar to traditional error matrices, in that they compare the timing of identified changes between the automated and human interpreters of LTS. In the past, with interval-based Landsat datasets, matrices consisted of disturbance classes (e.g., harvest, fire) by interval, and a no disturbance class. Here, annual resolution vertices replace intervals, and the no change class becomes the no vertex class. Additionally, because we use annual LTS, we can explicitly focus on recovery and stability as classes.

Match-score analyses were a means of integrating duration of segments into the assessments. This was particularly important because many disturbances are subtle but increase in intensity over many years, and because recovery from disturbance is often a long, slow process. In these cases, it is important to know not just that a disturbance or recovery process has begun, but how long it has lasted.

TimeSync facilitated a labeling of disturbance by both agent type and relative intensity. Agent type was only possible due to the spatial context afforded by photo-interpretation. Relative intensity was somewhat subjective, being based on proportion of existing vegetation canopy disturbed, but was extremely valuable for assessing the degree to which relative intensity affected LandTrendr's ability to detect disturbances. It is important to keep in mind here that expectations for forest change detection mapping errors derived from past studies that focused on stand-replacing (i.e., high intensity)

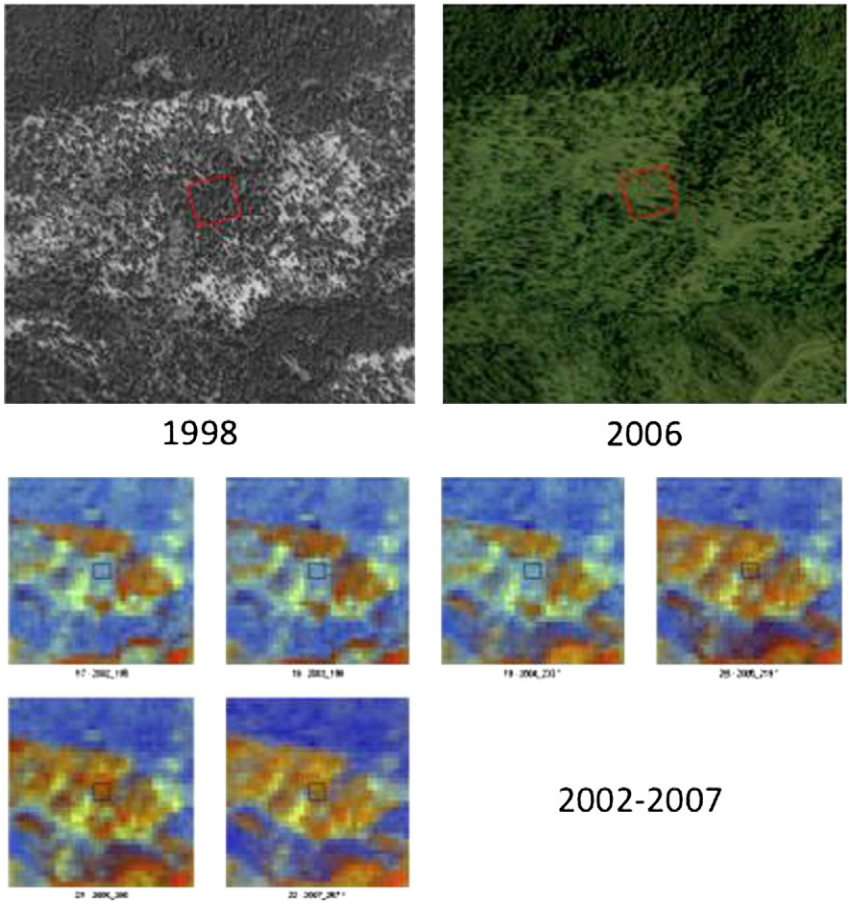


Fig. 10. Example of a harvest identified by TimeSync but excluded from the integrated secondary validation dataset. Shown are the image chips from 2002 to 2007 (bottom) and two screen grabs from Google Earth imagery (top) for the same plot (1998 and 2006). The harvest occurred between 2004 and 2005. Note the plot going from blue (conifer forest) to red (lack of vegetation).

disturbance (Cohen et al., 2002; Healey et al., 2008; Huang et al., 2010) must be revised for lesser intensity disturbances (Royle & Lathrop, 1997; Jin & Sader, 2005). Certainly, we should expect high intensity disturbances to be at least as accurately mapped as before, but the mapping of medium and low intensity disturbances is basically new map information that was not (or only minimally) previously available via the older methods using coarser time densities.

5.3. Other TimeSync uses

TimeSync was developed and used here for change detection algorithm calibration and map validation, but a number of other uses and investigations, several sample- rather than map-based, are possible. For example, using TimeSync, supported by secondary validation datasets, we discovered that fires are the most accurately detectable disturbance type using LTS, in the forests studied here. We might hypothesize that the ability to examine each plot in its spatial context, by zooming out within the chip window, made this possible. However, this was also the case for LandTrendr, which does not consider spatial context. There is apparently an inherently spectral signature associated with fires that is not associated with other disturbance types.

We might also have begun to learn something about recovery in terms of spectral residence time. The average number of years per plot in a recovery state, as characterized by TimeSync was 10.4 years. This might hint at the length of time following a disturbance that spectral recovery is detectable, and is consistent with the literature on this subject (Wulder et al., 2004; Masek et al., 2008). However, this is an average across plots that were both disturbed and not disturbed within the time window examined, plots that were disturbed more than once, and plots that were only partially disturbed. A more direct assessment, which was beyond the scope of this study, would have been to isolate high intensity disturbances that occurred within our time frame, and examine the length of time for these plots to recover to a spectrally stable condition. We might then further examine the effect of disturbance agent and geoclimatic regime on this phenomenon.

Other potential uses of TimeSync using data we collected here include:

- Comparative explorations of the behavior of a suite of spectral indices to a variety of vegetation system perturbations across a variety of vegetation systems. How does NDVI compare to NBR and wetness for forest change detection in needleleaf forests vs. broadleaf forests? What spectral indices are best utilized for change detection in agricultural, shrubland, and grassland systems? Is woody encroachment detectable with Landsat using LTS?
- Forest vs. non-forest masking. How well do different masking algorithms accurately map forest vs. non-forest? With TimeSync, each plot was labeled by land cover type, which would facilitate this assessment.
- Land use change. How much forest or agriculture is lost over time due to conversion to other uses? This would be possible because, with TimeSync, we noted the land cover/use class for the start and end vertices of each observed segment.

Beyond this study, TimeSync could be used for:

- Examination of spectral change with increases in percent impervious surface.
- Bridging MSS and TM datasets. How do different spectral indices behave across time series that include both types of Landsat data? How do these behave after normalization?
- Intra-annual vegetation characterization. Where intra-annual Landsat datasets exist, how well does Landsat data track seasonal progressions of spectral response relative to MODIS?
- Augmentation of plot-based datasets. For forest inventory or long-term ecology plots, chip window snapshots can provide a useful addition to temporally fill those datasets between field measurement dates and to provide spatial context.

Acknowledgements

The development and testing of TimeSync were made possible with the 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. 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.

References

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. A., Alig, R. J., Oetter, D. R., Maiersperger, T. K., & Fiorella, M. (2002). Characterizing 23 years (1972–1995) of stand replacement disturbance in western Oregon forests with Landsat imagery. *Ecosystems*, *5*, 122−137.
Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*, 35−46.
Coppin, P. R., & Bauer, M. E. (1994). Processing of multitemporal Landat TM imagery to optimize extraction of forest cover change features. *IEEE Transactions on Geoscience and Remote Sensing*, *32*, 918−927.
Crist, E. P., & Cicone, R. C. (1984). A physically-based transformation of thematic mapper data—The TM tasseled cap. *IEEE Transactions on Geoscience and Remote Sensing*, *22*, 256−263.
DeFries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., & Loveland, T. R. (2000). A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, *6*, 247−254.
Eidenshink, J., Schwind, B., Brewer, K., Zhu, Z. L., Quayle, B., & Howard, S. (2007). A project for monitoring trends in burn severity. *Fire Ecology*, *1*, 3−21.
Franklin, J., & Dyrness, C. (1988). *Natural vegetation of Oregon and Washington.* Corvallis (OR): Oregon State University Press.
Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., et al. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment*, *83*, 287−302.
Goward, S. N., Masek, J. G., Cohen, W. B., Moisen, G., Collatz, G., Healey, S., et al. (2008). Forest disturbance and North American carbon flux. *American Geophysical Union EOS Transactions*, *89*, 105−116.
Hansen, J., Sato, M., Ruedy, R., Lacis, A., & Oinas, V. (2000). Global warming in the twenty-first century: An alternative scenario. *Proceeding National Academy of Sciences*, *97*, 9875−9880.
Healey, S. P., Cohen, W. B., Spies, T. A., Moeur, M., Pflugmacher, D., Whitley, M. G., et al. (2008). The relative impact of harvest and fire upon landscape-level dynamics of older forests: Lessons from the Northwest Forest Plan. *Ecosystems*, *11*, 1106−1119.
Healey, S. P., Zhiqiang, Y., Cohen, W. B., & Pierce, 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.
Heller, R. C. (1975). *Evaluation of ERTS-1 data for forest and rangeland surveys, Research Paper PSW-112.* Berkeley, CA: USDA Forest Service.
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.
Kaufmann, R. K., & Seto, K. C. (2001). Change detection, accuracy, and bias in a sequential analysis of Landsat imagery in the Pearl River Delta, China: Econometric techniques. *Agriculture, Ecosystems & Environment*, *85*, 95−105.
Kennedy, R. E., Cohen, W. B., & Schroeder, T. A. (2007). Trajectory-based change detection for automated characterization of forest disturbance dynamics. *Remote Sensing of Environment*, *110*, 370−386.
Kennedy, R .E., Yang, Z., & Cohen, W. B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms, *Remote Sensing of Environment*.

Key, C. H., & Benson, N. C. (2005). Landscape assessment: Remote sensing of severity, the Normalized Burn Ratio. In D. C. Lutes (Ed.), *FIREMON: Fire effects monitoring and inventory system, General Technical Report, RMRS-GTR-164-CD:LA1-LA51*. Ogden, UT: USDA Forest Service, Rocky Mountain Research Station.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., et al. (2001). The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change*, *11*, 261−269.

Landsat Science Team (C.E. Woodcock, R. Allen, M. Anderson, A. Belward, R.. Bindschadler, W.B. Cohen, F. Gao, S.N. Goward, D. Helder, E. Helmer, R. Nemani, L. Orepoulos, J. Schott, P. Thenkabail, E. Vermote, J. Vogelmann, M. Wulder, R. Wynne). 2008. Free access to Landsat data, Letter, *Science* 320:1011.

Lunetta, R. S., Johnson, D. M., Lyon, J. G., & Crotwell, J. (2004). Impacts of imagery temporal frequency on land-cover change detection monitoring. *Remote Sensing of Environment*, *89*, 444−454.

Masek, J. G. (2001). Stability of boreal forest stands during recent climate change: Evidence from Landsat satellite imagery. *Journal of Biogeography*, *28*, 967−976.

Masek, J. G., Huang, C., Cohen, W. B., Kutler, J., Hall, F. G., Wolfe, R., et al. (2008). North American forest disturbance mapped from a decadal Landsat record: Methodology and initial results. *Remote Sensing of Environment*, *112*, 2914−2926.

Masek, J. G., & Collatz, G. James (2006). Estimating forest carbon fluxes in a disturbed southeastern landscape: Integration of remote sensing, forest inventory, and biogeochemical modeling. *Journal of Geophysical Research*, *111*, G01006. doi: 10.1029/2005JG000062.

Moeur, M., Spies, T. A., Hemstrom, M., Martin, J. R., Alegria, J., Browning, J., et al. (2005). Status and trend of late-successional and old-growth forest under the Northwest Forest Plan. *General Technical Report PNW-GTR-646*. Portland, OR: USDA Forest Service, PNW Research Station.

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, *421*, 37−42.

Royle, D. D., & Lathrop, R. G. (1997). Monitoring hemlock forest health in New Jersey using Landsat TM data and change detection techniques. *Forest Science*, *43*, 327−335.

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *BioScience*, *54*, 547−560.

Schroeder, T. A., Cohen, W. B., Song, C., Canty, M. J., & Zhiqiang, Y. (2006). Radiometric calibration of Landsat data for characterization of early successional forest patterns in Western Oregon. *Remote Sensing of Environment*, *103*, 16−26.

Singh, A. (1989). Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, *10*, 989−1003.

Skole, D., & Tucker, C. (1993). Tropical deforestation and habitat fragmentation in the Amazon: Satellite data from 1978 to 1988. *Science*, *260*, 1905−1909.

Turner, D. P., Ritts, W. D., Law, B. E., Cohen, W. B., Yang, Z., Hudiburg, T., et al. (2007). Scaling net ecosystem production and net biome production over a heterogeneous region in the western United States. *Biogeosciences*, *4*, 597−612.

Wulder, M. A., Skakun, R. S., Kurz, W. A., & White, J. C. (2004). Estimating time since forest harvest using segmented Landsat ETM + imagery. *Remote Sensing of Environment*, *93*, 179−187.