Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 110 (2007) 370–386
Remote Sensing of Environment
www.elsevier.com/locate/rse

Trajectory-based change detection for automated characterization of forest disturbance dynamics

Robert E. Kennedy *, Warren B. Cohen, Todd A. Schroeder

USDA Forest Service Pacific Northwest Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331, United States

Received 5 December 2006; received in revised form 12 March 2007; accepted 17 March 2007

Abstract

Satellite sensors are well suited to monitoring changes on the Earth's surface through provision of consistent and repeatable measurements at a spatial scale appropriate for many processes causing change on the land surface. Here, we describe and test a new conceptual approach to change detection of forests using a dense temporal stack of Landsat Thematic Mapper (TM) imagery. The central premise of the method is the recognition that many phenomena associated with changes in land cover have distinctive temporal progressions both before and after the change event, and that these lead to characteristic temporal signatures in spectral space. Rather than search for single change events between two dates of imagery, we instead search for these idealized signatures in the entire temporal trajectory of spectral values. This trajectory-based change detection is automated, requires no screening of non-forest area, and requires no metric-specific threshold development. Moreover, the method simultaneously provides estimates of discontinuous phenomena (disturbance date and intensity) as well as continuous phenomena (post-disturbance regeneration). We applied the method to a stack of 18 Landsat TM images for the 20-year period from 1984 to 2004. When compared with direct interpreter delineation of disturbance events, the automated method accurately labeled year of disturbance with 90% overall accuracy in clear-cuts and with 77% accuracy in partial-cuts (thinnings). The primary source of error in the method was misregistration of images in the stack, suggesting that higher accuracies are possible with better registration.

Keywords: Change detection; Landsat; Forest; Clear-cut; Partial-cut; Curve-fitting; Oregon

1. Introduction

Satellite remote sensing has long been used as a means of detecting and labeling changes in the condition of the land surface over time (Coppin & Bauer, 1996; Mouat et al., 1993). Satellite sensors are well-suited to this task because they provide consistent and repeatable measurements at a spatial scale appropriate for capturing the effects of many processes that cause change, including natural and anthropogenic disturbance (Jin & Sader, 2006; Muchoney & Haack, 1994; Royle & Lathrop, 2002), climate change (Silapaswan et al., 2001), and urbanization (Kwarteng & Chavez, 1998). Accordingly, the field of change detection in remote sensing is rich with case studies, methods, and applications in a wide range of practical situations.

Characterizing change in forested areas is of particular interest. With large stores of carbon in live vegetation and soil, forests play an important role in the global carbon cycle (Houghton et al., 2001). Because the magnitude of carbon loss during and after disturbance is large relative to the yearly carbon flux in undisturbed forest stands, spatially integrated net carbon flux in forests at any given time is largely determined by the spatial extent of disturbance and by the rate of regrowth of forested vegetation (Cohen et al., 1996; Harmon, 2001; Körner, 2003; Law et al., 2004). Land management in forests can also affect biodiversity, hydrology, and economics (Bengtsson et al., 2000; Brown et al., 2005; Drever et al., 2006; Gottschalk et al., 2005; Jules et al., 1999; Kline et al., 2004; Miller et al., 2003; Moore & Wondzell, 2005; Sallabanks et al., 2000; Stephens et al., 2004; Sun et al., 2001; Thornton et al., 2000; Tyler & Peterson, 2004).

As resource managers, policy makers, and the public at large demand knowledge of forest change over increasingly large

* Corresponding author. Tel.: +1 541 750 7498.
E-mail address: robertkennedy@fs.fed.us (R.E. Kennedy).

0034-4257/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2007.03.010

spatial and temporal extents, methodologies for change detection and labeling must continue to improve. Two goals for improvement stand out. First, change detection methods must better characterize long term trends and baseline rates of change. Long term datasets, such as that from the Landsat family of sensors, have the potential to lay this foundation (Cohen & Goward, 2004), but change detection methods that can exploit this archive are generally lacking. Most change detection methodologies documented in the literature focus on only two images at a time (Coppin et al., 2004), preventing extraction of longer-term trends. Second, change detection techniques need to move toward automation if changes are to be monitored over large areas. For two-date change detection, most spectrally-based methods require user designation of a threshold separating real change from spectral changes caused by geometric misregistration, variability in illumination, and vagaries of seasonality and image date (Lu et al., 2004). This manual intervention adds significant cost to efforts to expand change detection to large areas.

In this paper, we describe tests of a method that moves towards automation in tracking long time series data. The central premise of the method is the recognition that many phenomena associated with changes in forests have distinctive temporal progressions both before and after the change event, and that these lead to characteristic temporal signatures in spectral space. Rather than search for single change events between two dates of imagery, we instead search for these idealized signatures in the entire temporal trajectory of spectral values. If an area fits the idealized trajectory according to a simple least-squares measure of goodness of fit, it is likely to have experienced the phenomenon described by that trajectory. Because the entire trajectory is considered, the method can utilize the depth of rich image archives, and because detection is based on the fit of a curve, thresholding is based on a statistical

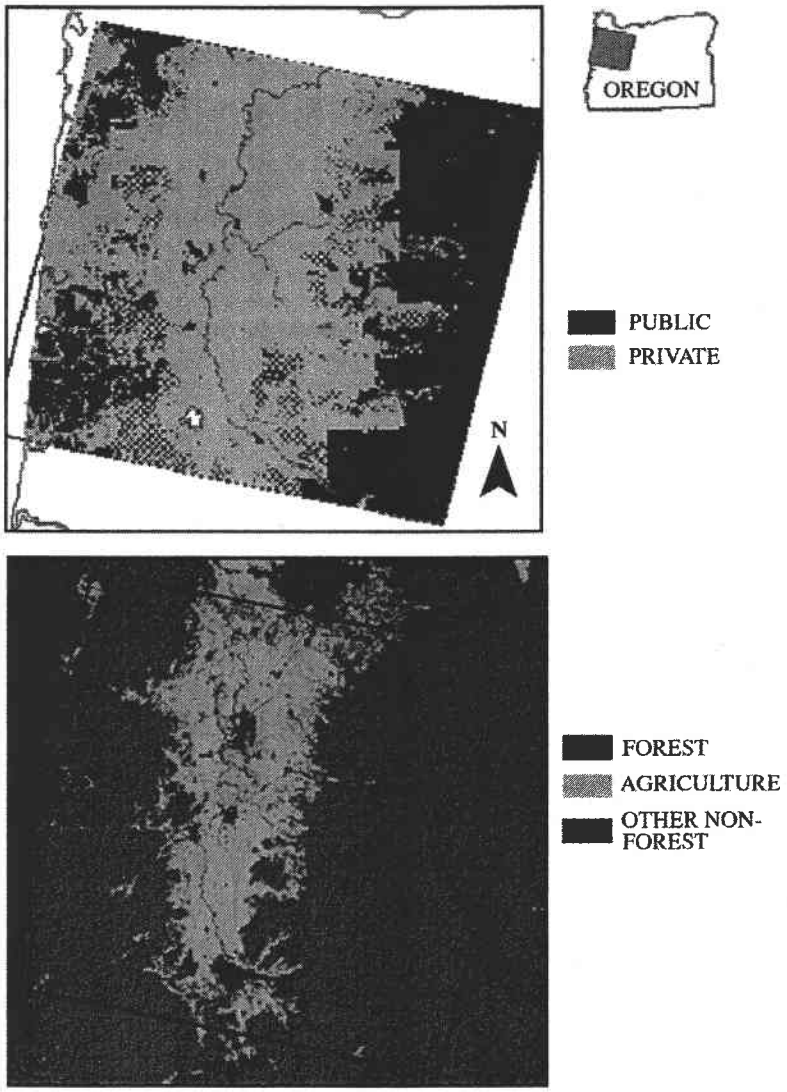

Fig. 1. The Landsat TM scene from path 46 row 29 (46/29) was used for this study. The scene is located in Oregon, U.S.A., and is composed of nearly 75% private lands and 25% public lands. Much of the central north–south running Willamette River Valley is in an agricultural land use.

metric that is internally calibrated to each pixel, avoiding the need for manual interpretation or intervention.

2. Methods

2.1. Data and study area

We focus on a series of Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) images for a single scene in western Oregon, U.S.A. The TM sensor was included on both the Landsat 4 and 5 satellites, the former launched in 1984 and the latter in 1986 (Cohen & Goward, 2004). By including two spectral bands in the short-wave infrared regions and a single band in the thermal region, and by operating with a nominal grain size of 30 m for the non-thermal region, these TM sensors represented a significant improvement in information content over the multispectral scanner (MSS) sensors that began recording imagery in 1972. Improvements in radiometric quality and the inclusion of a 15-m panchromatic band marked the arrival of the ETM+ sensor in 1999. As a medium spatial resolution sensor, all of the TM and ETM+ sensors are well-suited for mapping changes in land cover caused by many anthropogenic factors (Cohen & Goward, 2004; Coops et al., 2007; Franklin et al., 2002).

Land cover and land management methods in the study area are diverse. The central portion of the scene includes the Willamette River Valley, a north–south running agricultural zone planted in row and field crops, as well as in a wide array of specialty crops and orchards (Oetter et al., 2000). Four distinct urban areas in the Willamette Valley are fully included in the scene (Albany, Corvallis, Eugene/Springfield, and Salem), with some suburbs of the southern Portland metropolitan area included at the northern edge of the scene (Fig. 1).

Conifer dominated forests occupy the eastern and western portions of the scene and are the focus of the change detection.

Table 1
Dates of Landsat TM and ETM+ images used in the analysis

Date	Type[a]
8/4/1984	TM
8/26/1986	TM
7/12/1987	TM
8/31/1988	TM
9/3/1989	TM
7/7/1991	TM
8/10/1992	TM
8/29/1993	TM
7/31/1994	TM
8/19/1995	TM
8/21/1996	TM
7/23/1997	TM
8/11/1998	TM
8/22/1999	ETM+
8/16/2000	TM
7/26/2001	ETM+
7/29/2002	ETM+
8/25/2003	TM
7/26/2004	TM

[a]TM: Landsat Thematic Mapper; ETM+: Enhanced Thematic Mapper Plus.

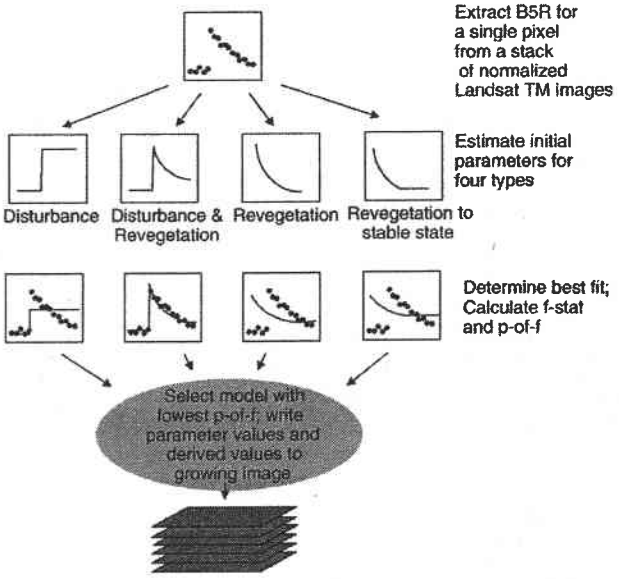

Fig. 2. A conceptual flow of the competing models approach. For each pixel in the image, band 5 reflectance (B5R) is extracted for all 19 years (dates shown in Table 1). Four separate hypothesized models (described in Fig. 3) are then fit to this data, first through an initial estimate of fitting parameters, and then through non-linear iterative curve-fitting. After each of four models is fit to the data, the model with the best fit is selected as describing the pixel, and its descriptive parameter values added to a growing summary image.

As reviewed in Healey et al. (2007), these and many other temperate forests undergo a variety of silvicultural treatments, including clear-cutting (complete or near-complete removal of all trees in a stand), partial stand removal (resulting in patchy residual tree islands), and thinning (removal of a portion of trees throughout a stand). Ownership of lands in the footprint of the scene is approximately 74% private and 26% public, with U.S. federal government forests dominating the public ownership portion.

2.2. Image preprocessing

Landsat TM images were acquired for Path 46 Row 29 (WRS II grid system) for the years between 1984 and 2004 (Table 1; see also Schroeder et al. (2006)). Acceptable images were confined seasonally to the period between July 1 and September 15, the period of Oregon's dry season, and had estimates of cloud cover area below 10% of the area of the image. For 1985 and 1990, images in the Landsat archive (http://earthexplorer.usgs.gov) had more than 40% of the area completely cloud covered, and as such no images were used from those two years.

Geometric and radiometric normalization are critical for any change detection technique where spectral values are compared across time (Garcia-Haro et al., 2001; Lu et al., 2004). Schroeder et al. (2006) describe the steps used for geometric and radiometric processing of these images. Briefly, a procedure for automated detection of image tie points (Kennedy & Cohen, 2003) was used to identify >100 tie points between each scene and the base scene, here the 1987 image, and then used in a

polynomial reprojection to align images across years (RMSE of position generally <0.5 pixels for all scene pairs). All images were then radiometrically normalized using the multivariate alteration detection (MAD) algorithm of Canty et al. (2004), which provides a quick and effective statistical approach to relative normalization of satellite images (Schroeder et al., 2006).

2.3. Fitting temporal trajectories

The overall process of fitting temporal trajectories is shown in Fig. 2. For each pixel in the stack of geometrically and radiometrically normalized images, the full 20-year time series (1984–2004) of band 5 reflectance (B5R) is extracted. For each of four hypothesized temporal trajectories (see below), initial estimates of trajectory shape parameters are made, and then sent to a fitting function that adjusts these initial parameters to find the best fit of the hypothesized trajectory to the observed trajectory. The discrepancy between the best fit hypothesized and the observed trajectories is summarized in terms of a standard *f*-statistic, and the probability of that *f*-statistic (*p*-of-*f*) is calculated. The model with the lowest *p*-value is selected and written to an output image. Additionally, the parameters describing that model are written to separate layers of the output image, as are the statistics (*f*-statistic and *p*-of-*f*) describing the fit of the best model. The model parameters written to the output image describe key aspects of the disturbance regime, including year of disturbance, change in spectral value at disturbance (a proxy for intensity of disturbance), and rate of recovery of spectral values (a proxy for revegetation rate). This process is repeated for every pixel in the image.

Hypothesized trajectories were based on short-wave infrared band reflectance (band 5 of TM and ETM+). The short-wave infrared region has been increasingly recognized as a useful tool in both characterizing vegetation and detecting changes in forested regions (Brown et al., 2000; Chuvieco et al., 2004; Healey et al., 2006; Trigg & Flasse, 2001) and is a key spectral region for inferring severity of burns (van Wagtendonk et al., 2004; White et al., 1996). In western Oregon, B5R is low in mature forests and high in exposed soil. The contrast between the near-infrared (TM band 4) and the short-wave infrared bands (TM bands 5 and 7) is the core of the tasseled cap wetness index (Crist, 1985), which has been shown to be useful in mapping forest structure and condition in coniferous forests (Cohen et al., 2001; Cohen & Spies, 1992; Healey et al., 2006; Lefsky & Cohen, 2003; Wulder et al., 2005).

We hypothesized four trajectories of change in B5R: simple disturbance, disturbance followed by revegetation, ongoing revegetation, and revegetation to stable state (Fig. 3). By developing appropriate functions to describe hypothesized trajectories, the parameters describing those functions themselves capture the key characteristics of disturbance and regrowth. Simple disturbance results in a stairstep function in band 5 (Fig. 3a). Mature conifer forests have low B5R, but soil and non-photosynthetic components of vegetation (bark, branches) have high B5R. When conifer vegetation is removed through disturbance, B5R jumps from an initially low value to a high value, and, in the absence of subsequent revegetation, remains high. The parameter values describing this shape are the end year of the disturbance interval (parameter p_0), the pre-disturbance mean reflectance (p_1), and the post-disturbance mean (p_2).

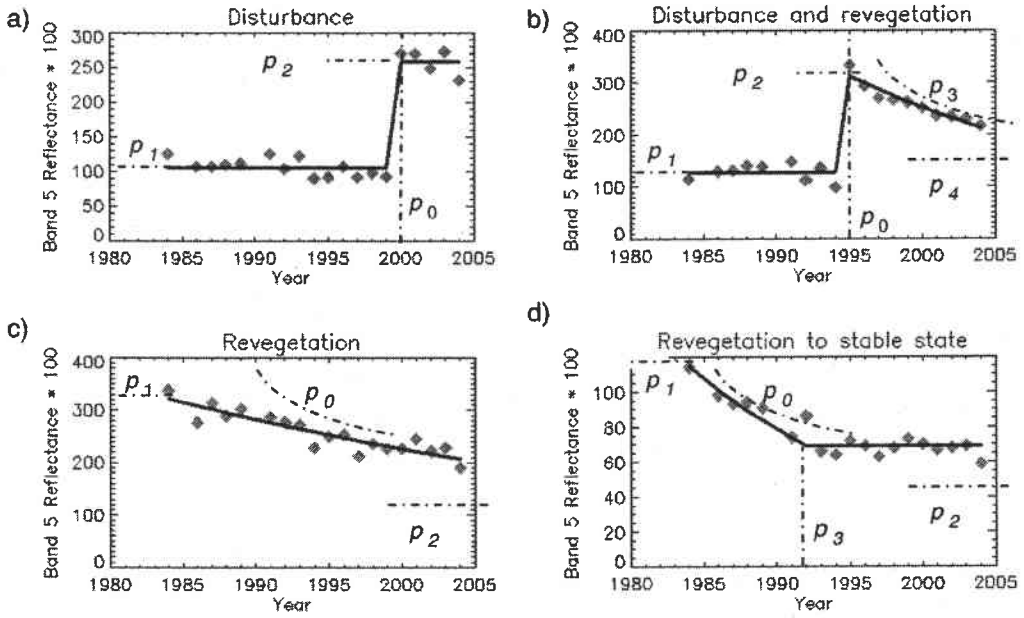

Fig. 3. A detailed view of fitted values from four pixels representing the four hypothesized models of disturbance or recovery. a) Simple disturbance. b) Disturbance followed by exponential revegetation. c) Revegetation from disturbance prior to observation record. d) Revegetation from prior disturbance, reaching a stable point during the observation period.

When disturbance is followed by revegetation, B5R initially spikes, but decreases over time (Fig. 3b). The trajectory from high to low B5R is presumed to follow an exponential decay curve. Two parameters describing this trajectory are identical to those from the simple disturbance: the pre-disturbance mean reflectance p_1 and the year after disturbance p_0. Parameters p_2, p_3, and p_4 describe the exponential function that captures return of spectral values to lower B5R. The exponential function considered here is:

$$r_{b5} = \left((p_2 - p_4) \cdot e^{-p_3' t}\right) + p_4 \quad (1)$$

where r_{b5} is the observed B5R, p_3 is an exponential decay constant, p_4 and p_2 are lower and upper bounds, respectively, on the exponential function (in units of reflectance), and t is the time since disturbance (in years). The parameter p_2 is analogous to the mean reflectance immediately after disturbance in the simple disturbance function, while the parameter p_4 describes the asymptotic level of reflectance if the exponential function were to continue indefinitely (this value is not explicitly linked with p_1). The parameter p_3 describes the rate of decrease of spectral reflectance from the high value p_2 to a value at infinity p_4. The higher the value of p_3, the faster the inferred recovery of vegetation. Units for p_3 are t^{-1}. Solving for t, if p_3 were 0.25, reflectance would recover half the spectral distance between p_4 and p_2 in 2.77 years ($t = \ln(0.5) \cdot (-p_3^{-1})$) and would recover 95% of that distance in 11.98 years. The exponential parameter was constrained to range from 0 to 1.0. The use of the exponential function by itself to model recovery rates from spectral data has precedence (Viedma et al., 1997).

When disturbance happens before the first year of the record (before 1984 for the image stack used in this study), the signal of revegetation may still be evident, and thus the ongoing "revegetation" temporal trajectory was developed (Fig. 3c). The signal is that of exponential recovery alone, without an observed disturbance event. Eq. (1) is used to describe this trajectory, except that parameter meanings are altered, such that t represents the time since the beginning of the record, and that p_2 represents the reflectance for the first year of the record.

A fourth model builds on the ongoing revegetation trajectory, but assumes that the reflectance stabilizes to an essentially unchanging flat line (Fig. 3d). Eq. (1) is used as in the ongoing revegetation trajectory, but an additional parameter is added that corresponds to the year at which stable reflectance is achieved.

Once the four trajectories have been defined, the core of the method is a nonlinear least-squares fit of each hypothesized trajectory to the spectral trajectory observed in each pixel. For fitting, we used MPFIT, an implementation of the Levenburg–Marquardt algorithm adapted for IDL (Interactive Data Language, ITT Visual Information Solutions, Inc.) by Craig Markwardt from the NETLIB MINPACK-1 package. The code is available from http://cow.physics.wisc.edu/~craigm/idl/idl.html. The program allows the user to develop functions of any arbitrary form that define how a set of arbitrary parameters converts a set of X values into a set of Y values. Initial estimates of parameter values are sent to MPFIT, which returns the adjusted parameter set that best fits the observed trajectory.

Initial estimates of the fitting parameters must be reasonable. For the two disturbance trajectory types, the year of disturbance is fixed at the point where the increase in band 5 reflectance from one year to the next is greatest. For simple disturbance, estimates of pre- and post-disturbance reflectance are taken as arithmetic means of reflectance in the years before and after the estimated year of disturbance. For all three trajectories involving exponential decay, the initial estimate of the decay constant is calculated as the difference of the log of the reflectance immediately after disturbance (or at the beginning of the record for ongoing revegetation and the revegetation to stability types) and the log of the reflectance in the following year. If this value is not mathematically defined, an initial estimate of zero is used. Bounding constraints of the exponential functions are reflectance at the year of disturbance (for p_2) and zero (for p_1). Finally, for the "revegetation to stability" type, initial estimates of stability level and onset of stability were estimated as the reflectance at the end of the record and the last year of the record, respectively.

The next phase is that of competing models. For each of the four hypothesized trajectories of change, initial estimates of the parameters are sent to the fitting function, which adjusts them until the best fit is found for that type. Predicted values are calculated and compared with observed values in an analysis of variance paradigm (Wackerly et al., 1996). A standard f-statistic (mean square $\text{error}_{\text{model}}$/mean square $\text{error}_{\text{residual}}$) is calculated, with degrees of freedom for the model equal to the number of parameters in the fitting model and for the residual equal to the number of years minus the number of parameters minus 1. High f-values indicate that the hypothesized trajectory describes the observed trajectory well. The probability of the calculated f-value (p-of-f) being due to chance is taken from standard f-statistic tables, with low p-of-f indicating better agreement between the hypothesized trajectory and the observed trajectory. This process is repeated for each of the four hypothesized trajectories. The four types are considered competing models for the observed conditions. The model with the lowest p-of-f value wins. The parameters of the winning model fully describe the disturbance and recovery dynamics for the pixel, and are written as separate layers to the growing summary image. Layers in the output image are described in Table 2.

No explicit model of "no change" is developed. Rather, the no-change condition is the implicit null-hypothesis against

Table 2
Layers in the summary image created in the competing models phase

Layer	Description
1	Trajectory type (1 through 4)
2	End of disturbance interval (1986 through 2004)
3	Spectral change at disturbance (proxy for disturbance intensity)
4	Decay constant in exponential revegetation (proxy for revegetation rate)
5	Recovery transition year (for recovery-to-stability type only)
6	f-statistic of winning model
7	Mean spectral value before disturbance
8	p-of-f
9	Mean square error of model
10	Mean square error of residuals
11	Spectral value at infinity (for exponential models only)

Table 3
Types of cover change interpreted directly from year-to-year difference in tasseled-cap spectral color

Type	Inferred change	Starting condition	Visual description	Ending condition	Visual description
1	Clear-cut conifer	Conifer	Light to dark blue/cyan	Soil	Bright or dark red
2	Partial cut conifer	Conifer	Light to dark blue/cyan	Any non-conifer	Any orange, red, or yellow, mottled or smooth
3	Treatment (herbicide, fire)	Any non-conifer	Any orange, red, or yellow, mottled or smooth	Soil	Bright or dark red

which all four models are compared. Although each pixel is assigned the model with the lowest *p*-of-*f*-statistic, pixels without change will have high *p*-of-*f* values even for that best model. Thus, no-change pixels are simply those pixels where the best model has a *p*-of-*f* value higher than 0.05.

2.4. Application of trajectory-based change detection to image stack

Temporal curve-fitting was applied to the stack of Landsat TM band 5 images listed in Table 1, resulting in a summary image with the 11 layers listed in Table 2. Although all images in the stack were relatively free of clouds, small areas of residual "popcorn" cumulus clouds remained. For the years 1984, 1986, 1989, 1991, 1992, 1993, 1997, and 2003, cloud masks were developed using post-classification labeling of a simple unsupervised classification of tasseled cap transformations of the images (Crist, 1985). During curve fitting of each pixel, only values from non-clouded years were considered. In the summary image, pixels with a *p*-of-*f* value less than 0.05 were subsequently labeled as "change." Note that this filtering on the *p*-of-*f* value was the only filtering conducted on the algorithm output; no filtering of urban, agricultural, or waterbody areas was conducted.

2.5. Validation

Ideally, any map based on remotely-sensed imagery should be validated with an entirely independent dataset, usually one step "closer" to reality than the imagery used to make the map (Congalton & Green, 1999). Validation of change-detection maps is often more challenging than is validation of single-date maps, because independent reference sources must be available for both ends of the change interval, and because the range of potential changes is often much greater than the range of classes in single-date maps (Coops et al., 2007). This is particularly true in this study, where changes must be validated for near-yearly periods spanning two decades. Therefore, we chose two complementary approaches to evaluate the performance of the trajectory-based change detection algorithms.

Because no independent datasets were available for every year of the record for the entire area, the only consistent source for evaluation was the imagery itself. Cohen et al. (1998) showed that direct human interpretation of tasseled-cap transformed Landsat imagery (Crist & Cicone, 1984) was as accurate for mapping forest clear-cuts as both independent photo and field-based datasets, even though the imagery used for interpretation was the same source used to build the change-detection map. The same approach was used in a broader forest-disturbance project described in Healey et al. (2005). Mapping of forest disturbance through direct image interpretation allows incorporation of spatial pattern and geographic context in the decision-making process, and these are not easily incorporated into automated algorithms. In essence, the error matrix that compares human interpretation of satellite imagery with results of an automated algorithm is a test of agreement between that algorithm and the model of reality used by the human interpreter. Because a primary goal of this study was to develop robust algorithms that eliminated manual intervention, such a comparison is appropriate as one means of evaluating the method. Therefore, the accuracy assessment approach of Cohen et al. (1998) and Healey et al. (2005) formed the basis for our first type of validation.

Two improvements on the approach of Cohen et al. (1998) were implemented in this study. Because change occurs rarely in any given time interval, most of the images are not changed, making detection of false negatives statistically rare when the model output of change is used to draw samples (Morisette & Khorram, 2000). To avoid this problem, we fully interpreted 3 by 3 km square areas without reference to the algorithm change maps. Samples of change and no-change could then be drawn in a balanced design. A second improvement relates to labeling the observed change. Disturbance type was labeled using interpretation rules shown in Table 3. Clear-cuts in conifer forest (Type 1 disturbance) were labeled as distinct from partial cuts in conifer forests (Type 2 disturbance) and from vegetation removal in non-conifer conditions caused by post-harvest treatments such as fire or herbicide application (Type 3 disturbance). For each disturbance label, the interpreter quantified a confidence score based on agreement with five "confidence statements" shown in Table 4. The first two statements correspond directly to the spectral properties before and after disturbance, while the latter three statements correspond to spatial and temporal context. A key consideration is whether the land use around a patch is appropriate for the type of change observed. For example, forest disturbance can be more confidently labeled if it occurs in a land use system of active forestry rather than a land use system of crop agriculture. A single

Table 4
Rules for assigning interpreter confidence scores to digitized disturbance polygons

Confidence statement	Component score range[a]
"Starting condition is spectrally unambiguous"	0, 1, 2
"Ending condition is spectrally unambiguous"	0, 1, 2
"Land use around the patch is appropriate"	0, 1
"Shape and size of patch is appropriate"	0, 1
"Temporal trajectory after disturbance is appropriate"	0, 1
Range of possible scores, 7 is best	0 to 7

[a] Each confidence statement is evaluated for each disturbance polygon, with the total score being the total of component scores across all confidence statements.

interpreter digitized all polygons, with spot quality checking by a second interpreter.

All interpretation occurred within "interpretation boxes" selected from a regular 3 km by 3 km grid overlain on the image area. The population of grid cells fully contained in the area common to all images in the stack ($n=2803$) was ordered into a random list. Interpretation occurred on boxes in sequential order from this random list until several examples of cuts in each year were captured and before project budget was overstepped. In the end, 67 "interpretation boxes" were interpreted (Fig. 4).

For the entire 9 km^2 area in each of these interpretation boxes, polygons were screen-digitized around all forest disturbances in all years. Each year of imagery was loaded into an image viewing platform, as well as year-to-year difference images. Difference images were used as an initial visual guide for identifying potential disturbance, but the final interpretation and digitization of the disturbed area was based on the original tasseled cap imagery. Each disturbance patch was attributed with disturbance year, disturbance type, and interpreter confidence score. With 67 interpretation boxes covering 9 km^2 each for 19 separate image years, 11,457 km^2 of imagery was interpreted.

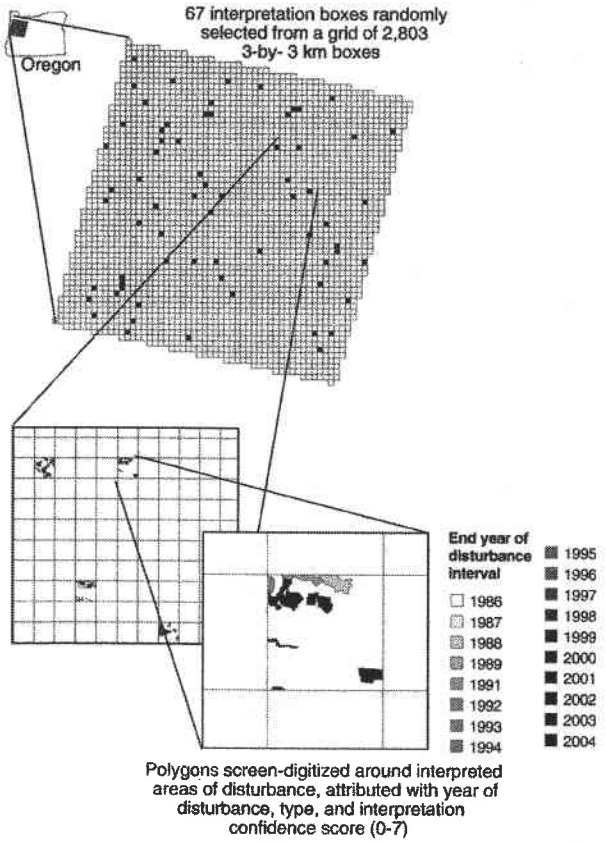

Fig. 4. The grid of potential sample plots within the footprint of Landsat TM scene 46/29, with 67 randomly selected plots used for on-screen digitized shown in shaded tone. Each plot was 3 by 3 km in size. Within each 3 by 3-km box, all disturbances noted by direct interpretation of tasseled-cap imagery were digitized and labeled according to disturbance year, type, and interpreter confidence in change label.

The screen digitized disturbance polygons formed one validation dataset for accuracy assessment. Polygons were converted to a raster layer with 30 m cell size labeled with the year of disturbance. Within the area covered by disturbed-area polygons, 500 samples were selected. Selection for each sample began with a single randomly seeded point, around which a 9-pixel contiguous region was grown (9 pixels=0.81 ha), with the constraint that all 9 pixels remain within the change polygon. This was the first sample point. A second seed point was randomly chosen and a 9-pixel area grown around it, with additional constraint that its boundaries be no closer than three pixels from the existing sample, to limit the effects of spatial autocorrelation among samples. This continued until 500 samples were selected. An identical process was applied to the areas not labeled as change by the interpreter, inferred to be "no-change," resulting in 500 no-change samples. These 1000 change/no-change samples were then compared to the year-of-disturbance layer from the summary image created by the trajectory-fitting algorithms. A sample was labeled "correct" if the majority of the 9 pixels in the human-interpreted map were labeled the same in the algorithm-derived image. Errors were tabulated for both the change/no-change label and for the year of disturbance label. Standard error matrices and summary statistics were then calculated (Congalton & Green, 1999). Accuracy assessment was conducted on the entire set of interpreted disturbances, and then repeated separately for disturbances labeled as Type 1 and Type 2 (Table 3).

In addition to calculating error matrices at the 9-pixel (0.81 ha, i.e. plot) level, agreement between interpreter and algorithm was evaluated at the patch level. First, the year of disturbance layer from the algorithm summary image was filtered to a minimum mapping unit of 9 pixels. Pixels from this filtered image were then counted within each of the 829 screen-digitized polygons. The algorithm was considered accurate for a patch if more than half of the screen-digitized area of the patch was in agreement. The plot level and patch level comparisons between the algorithm and the human interpretation of the imagery were the first form of evaluation of the method.

The second form of evaluation of the method used digital orthophoto quads (DOQs) to evaluate change detection maps for one time period in the 20-year record. While not as temporally dense as the evaluation based on yearly imagery, this method could be considered a traditional accuracy assessment because the source was entirely independent of the imagery used to build the change detection maps, and was an order of magnitude finer spatial resolution than the Landsat imagery. For 20 interpretation boxes selected at random from the 67 boxes described above, DOQs for the years 1995 and 2000 were acquired from the State of Oregon Geospatial Enterprise Office (http://www.oregon.gov/DAS/EISPD/GEO/data/doq.sthml) and used for interpretation. DOQs were black and white with a nominal spatial grain of 1 m. All removals of forest between 1995 and 2000 were screen-digitized based on visual interpretation of the DOQs for the full 3 km by 3 km area, and were then labeled using the same three types shown in Table 3. The fine grain of the photos allowed interpretation based on texture and land-use labeling not possible with the satellite imagery alone. Polygons digitized

from the DOQs were compared for agreement with polygons digitized from the satellite imagery alone, to provide a sense for the utility of the satellite-interpreted dataset for time periods not covered by the DOQs. Polygons from the DOQ interpretation were also compared for agreement directly with the trajectory-based change detection map.

3. Results

3.1. Interpretation

Within the 67 randomly distributed 9-km^2 interpretation boxes, 829 patches of disturbance were located and digitized from the tasseled-cap imagery. The first noted disturbance was Type 1 (clear-cut) in 503 patches and Type 2 (thinning/partial cut) in 312 patches. Confidence scores generally ranged from five to seven, with 700 patches having confidence scores of six or seven. The minimum, median, and maximum patch digitized patch sizes were 0.2, 5.29, and 231.7 ha, respectively. Because patch boundaries were truncated at the edge of interpretation boxes, it is possible that digitized patches may be smaller than the actual size of the disturbed area, and thus these size estimates cannot be used to characterize average patch sizes for the study area.

Within the area of the 20 interpretation boxes randomly selected for DOQ-based validation, 160 distinct polygons of forest disturbance were located. Use of the high-resolution DOQ allowed capture of events with much smaller size than possible with the direct interpretation of the satellite imagery. The minimum, median and maximum patch size were 0.056, 1.51, and 78.1 ha respectively. To provide an indication of the relative size of patches digitized from the DOQs and the satellite imagery, 129 of the 160 DOQ-digitized polygons were smaller than 5.29 ha, the median size of polygons digitized from the satellite imagery.

3.2. Trajectory-based change detection

The trajectory-based change detection resulted in a summary image whose layers corresponded to fitting parameters for those pixels whose trajectory matched one of the four hypothesized trajectories at the $p=0.05$ level or better. Figs. 5 and 6 illustrate how the layers corresponding to end-year of disturbance interval and to disturbance intensity compared visually with interpreter digitized polygons.

3.3. Plot level agreement

At the plot level (9-pixel plots), overall change/no-change agreement between the algorithm and the interpreter was 90%, with a kappa value of 0.72 ($n=1000$, Table 5). Overall agreement on year of disturbance was 84%, with a kappa value of 0.77

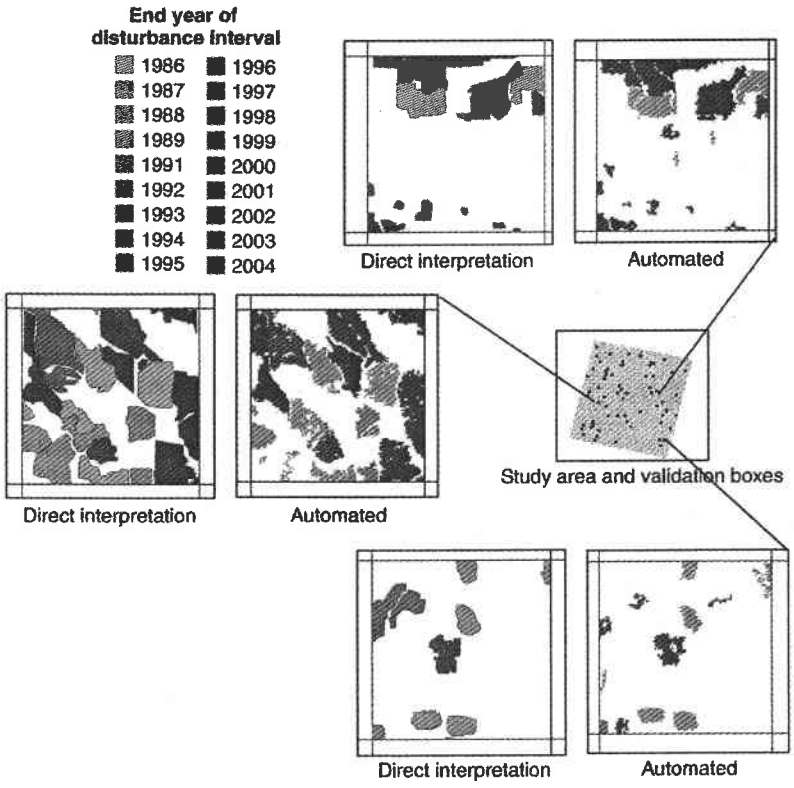

Fig. 5. A comparison of end-year of disturbance interval labeled by the interpreter and by the trajectory-based change detection algorithm. Shown are three of the 67 interpreted sample plots from Fig. 4. Algorithm-based images have been filtered to a minimum map unit of 9 pixels (0.81 ha). For each image pair, the left-hand image is from direct interpreter digitization and the right-hand image is from automated curve-fitting algorithms.