Remote Sensing of Environment 115 (2011) 19–32

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Validation of North American Forest Disturbance dynamics derived from Landsat time series stacks

Nancy E. Thomas [a,*], Chengquan Huang [a], Samuel N. Goward [a], Scott Powell [b], Khaldoun Rishmawi [a], Karen Schleeweis [a], Adrienne Hinds [a]

[a] Department of Geography, University of Maryland, College Park, MD 20742, United States
[b] Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717, United States

ARTICLE INFO

Article history:
Received 13 November 2009
Received in revised form 21 July 2010
Accepted 23 July 2010

Keywords:
North American Forest Disturbance (NAFD)
Landsat time series stacks (LTSS)
Vegetation change tracker (VCT)
Forest disturbance
Time series
Accuracy assessment

ABSTRACT

The North American Forest Dynamics (NAFD) study is a core project of the North American Carbon Program (NACP). The NAFD project is evaluating forest disturbance patterns and rates of disturbance by integrating U.S. Department of Agriculture (USDA) Forest Service Inventory and Analysis (FIA) field observations with temporally dense time series Landsat imagery. In Phase I of NAFD forest disturbance history was derived for 23 U.S. sample locations over the time period 1984 to 2005 from biennial Landsat time series stacks (LTSS). This study evaluates the accuracy of these Phase I NAFD disturbance history maps for 6 selected sample locations. We evaluate the disturbance maps using 2 reference datasets: 1) a design-based approach incorporating visual analysis of the LTSS in tandem with high resolution imagery and 2) the USDA FIA field observations. Overall accuracy for the NAFD disturbance product assessed at the individual time step level range from 77% to 86%. We examine the success rates of the mapping approach for capturing different types of disturbance and find that 82% of stand clearing events were detected. When we aggregate the data into change and no change categories the accuracy of stand clearing disturbance samples improved to over 92%. The majority of error in the disturbance maps was due to misclassification of partial disturbance as unchanged forest. We analyze the resulting errors of commission and omission as related to both reference datasets for each LTSS and present examples to illustrate the strengths and weaknesses of Phase I NAFD approach. In addition, we discuss the map biases observed in this work and what this may imply for estimating national forest disturbance rates with this approach.

1. Introduction

Uncertainties regarding North American forest dynamics, including disturbance and regeneration, contribute to large uncertainty surrounding estimates of continental carbon fluxes. The first State of the Carbon Cycle Report (SOCCR) estimated that North American forests are currently carbon sinks that offset nearly 13% of U.S. fossil fuel emissions, the equivalent of sequestering 0.21 petagrams C/year (CCSP, 2007). However, the uncertainty of this estimate is ~50%. The underlying forest dynamics—including extent, rate, and magnitude of change events such as fire, harvest, insect damage, and disease—are not currently well understood. Without a better understanding of these underlying dynamics, estimating how forest carbon sources and sinks might vary in the future will be nearly impossible.

Within the North American Forest Dynamics (NAFD) study, a core project of the North American Carbon Program (NACP), we are estimating national rates of forest disturbance and recovery from a combination of Landsat observations and U.S. Forest Service (USFS) Forest Inventory and Analysis (FIA) measurements. The NACP is an interagency, interdisciplinary research program that seeks to improve understanding of carbon sources and sinks in North America (Wofsy & Harris, 2002). NAFD is directed to improve our understanding of disturbance processes as a factor in these sources and sinks.

NAFD activities have been underway since 2003 when we began a prototype study in the U.S. Mid-Atlantic region to evaluate forest disturbance detection using Landsat time series stacks (LTSS). In Phase I (2005–2008) of NAFD 23 LTSS were compiled for a selection of United States sample sites for the purpose of providing an estimate of national forest disturbance rates (Fig. 1). These sites were selected using an unequal probability sampling method that incorporated a number of factors, including forest type, forest area, spatial dispersion, and preferential inclusion of stacks already compiled and available through other projects (Kennedy et al., 2006). An additional 7 LTSS were generated for prototyping and as study areas of particular interest identified by FIA. Because forest change features can rapidly become obscured due to vigorous regrowth (Lunetta et al., 2004; Masek et al., 2008), we compiled dense LTSS consisting of images from approximately biennial time steps for the time period 1984 to 2005. We developed a highly automated vegetation change tracker (VCT) algorithm to map forest disturbance history for each of the NAFD sample site LTSS (Huang et al., 2010).

---

* Corresponding author.
E-mail address: nthomas1@umd.edu (N.E. Thomas).

0034-4257/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2010.07.009

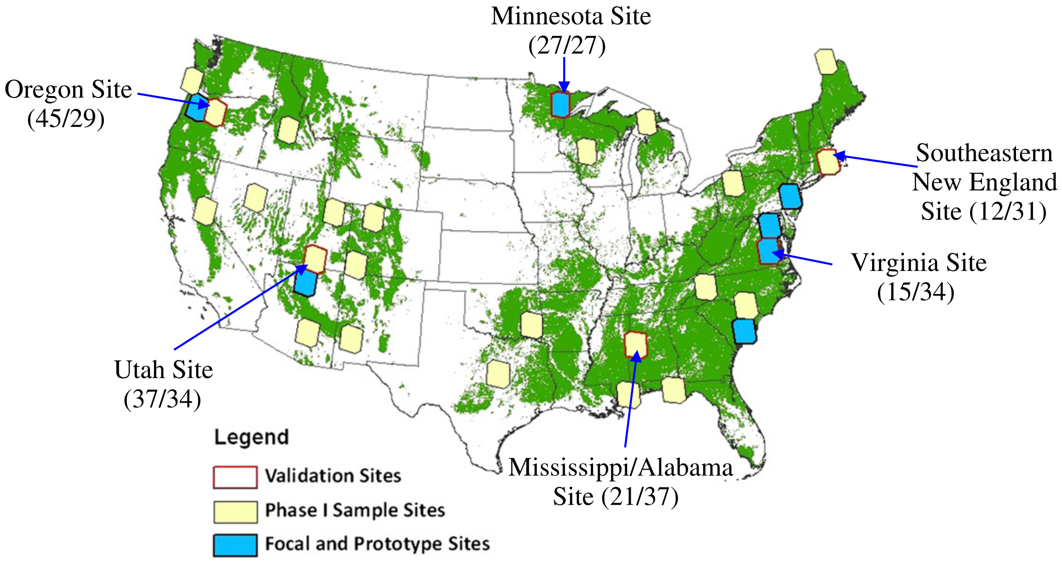


Fig. 1. Locations within U.S. where NAFD disturbance map products are being produced from Landsat time series stacks (LTSS). Validation stacks as discussed in this paper are shown. Acquisition and processing of Phase II stacks is in progress. Additional Phase II stacks are being acquired in Mexico and Canada.

The goal of this research is to examine the validity of disturbance products derived from the LTSS-VCT mapping approach. Understanding the uncertainties in these NAFD disturbance products is needed for any later applications assessing U.S. carbon dynamics. Map validation requires high resolution imagery and/or ground verified independent reference information (Congalton & Green, 1999) which can be difficult to obtain because of availability, accessibility, and costs (Congalton, 1991; Stehman & Czaplewski, 1998). Validation of NAFD products is further complicated due to the lack of conventional validation sources at the required biennial temporal frequency (Lu et al., 2004).

The NAFD project plan originally intended to validate the disturbance products employing USFS FIA inventory as the primary reference data source. As we began to explore these data we found that the density of suitable FIA plots located in forest change areas was insufficient to validate the NAFD biennial disturbance products at an individual time step level. To address this challenge we also developed a design-based approach incorporating visual analysis of the full LTSS for deriving reliable reference data for a greater number of sample locations in disturbed forest then was possible using the FIA plot data. We could then derive statistically unbiased accuracy estimates of the NAFD disturbance products for the selected sample LTSS at the individual disturbance time step level in addition to evaluating NAFD results with an independent ground-based reference dataset (FIA).

Both the design-based and FIA validation approaches have been applied to 6 NAFD sites to evaluate the accuracy of the NAFD disturbance mapping approach. This report describes these two approaches and their outcome for validating the NAFD disturbance products.

2. NAFD disturbance product development

Key aspects of the NAFD project have been described in detail in previous papers (Goward et al., 2008; Huang et al., 2009a,b, 2010). We briefly address the NAFD methodology here and refer the reader to appropriate publications for additional information.

2.1. Landsat time series stacks

During NAFD Phase I, each Landsat scene employed was purchased from the USGS EROS Landsat archive. As a result, we limited the LTSS temporal coverage to approximately biennial time steps to keep costs for the project under control. Most of the selected images were acquired during the summer peak green season (June–September) and had minimum (<10%) or no cloud cover. However, in some cases either the seasonal and/or the cloud conditions could not be met in specific biennial years. In such cases, the temporal interval between consecutive LTSS images can be 1 or 3 years (Huang et al., 2009a).

The selected LTSS images were processed using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al., 2006) to achieve high levels of geolocation accuracy and radiometric integrity. The LEDAPS system starts with USGS L1G Landsat imagery and carries out further image preprocessing including orthorectification; radiometric calibration; and atmospheric adjustments (Masek et al., 2006; Gao et al., 2009). Further details on the algorithms and procedures for producing the LTSS is provided by Huang et al. (2009a).

2.2. Vegetation change tracker analysis

The vegetation change tracker (VCT) algorithm was designed specifically for mapping forest change using LTSS or LTSS-like data sets that consist of temporally dense satellite acquisition (Huang et al., 2010). The VCT algorithm consists of two major steps: 1) individual image analysis and 2) time series analysis (Huang et al., 2010). The VCT outputs disturbance year maps, which identify three static classes — persisting forest, persisting nonforest, and water — in addition to flagging the year of disturbance for all pixels where forest change was detected. Mapped classes include:

- Persistent forest — pixels that remained forested throughout the time series.
- Persistent nonforest — pixels that were never forested during the entire observing period of the time series.
- Persistent water — pixels that were water pixels throughout the observing period are defined as persisting water. The persistent water class was combined into the persisting nonforest class for the validation work described in this report.
- Forest disturbance — pixels that are not classified as one of the persisting land cover classes. The pixel label corresponds to the time step in which the disturbance event occurred.
- Pre-series disturbance — pixels that are classified as nonforest during time 1 of the series but change to forest at some point during

the observation period. Both forest regrowth and afforestation processes could be included in this category.

The disturbance year map product summarizes forest cover changes that have occurred during the observation period (1984–2005) (Fig. 2). For the static classes and pixels where no more than one disturbance occurred during the entire observing period of the LTSS, these classes can be summarized using a single map layer. However, we observed that multiple disturbances can be detected during this observation period due to rapid forest regrowth. To ensure that multiple disturbances were recorded, the disturbance year map was designed to have two layers. The first map layer corresponds to the initial time of disturbance, and the second layer corresponds to the last disturbance occurrence. The two layers have the same values for the static classes and pixels where only one disturbance was detected during the entire observing period of the LTSS.

2.3. Minimum mapping unit filter

For the final NAFD map product, a moving window filter was applied to reduce speckle where individual pixels or small patches consisting of just a few pixels were mapped as change. While some of this speckle might capture real change, most are likely the result of sensor point spread function properties, image-to-image misregistration produced from orbital variations and/or ortho-rectification imprecision (Knight & Lunetta, 2003). Time series analysis are particularly sensitive to errors due to misregistration (Townshend et al., 1992). To minimize the impact these data artifacts might have on disturbance analyses, a minimum mapping unit (MMU) was applied to the VCT results (Lillesand and Kiefer, 1994).

Different MMU were chosen for static classes (persistent non-forest, persistent forest, and water) versus disturbed forest classes. The three static classes were generally considered more reliable because static pixels reflect a consistent signal throughout the entire observing period (12 + time steps) of an LTSS, while disturbance classes may only be detected during a minimum of 2 time steps (Huang et al., 2010). Because forest regrowth can rapidly decrease the disturbance signal, disturbed pixels may return to a forested signal within a few time steps. To reflect these different confidence levels, we applied an MMU of 2 contiguous pixels (0.16 ha) for the static classes and an MMU of 4 contiguous pixels (0.36 ha) for the disturbance classes.

3. Validation methods

Collecting adequate reference data for validating land cover and change products typically requires substantial resources (Congalton & Green, 1999). Due to NAFD project constraints both the design-based and FIA validation approaches have been applied to 6 NAFD sites selected from the 30 NAFD LTSS to evaluate the accuracy of the disturbance analysis approach. These validation sites were selected to be geographically dispersed as well as representative of the various forest ecosystems and disturbance regimes across the U.S. (Fig. 1, Table 1).

3.1. NAFD design-based assessment

Identification of whether or not a forest disturbance event occurred in a particular year is relatively straightforward, as long as field data or high resolution imagery can be acquired immediately before and after the occurrence of that disturbance. However, existing datasets do not provide the required spatial and temporal characteristics to validate the NAFD disturbance products. Typical field plot data collected through the FIA program are available at 5–10 year time intervals, with nationally consistent plot data only available since the late 1990s. High resolution digital aerial photography such as USGS National Agricultural Imagery Program (NAIP) and earlier USGS digital ortho quarter quads (DOQQs) have historically been acquired at 5 year time intervals (http://www.apfo.usda.gov). High spatial resolution spacecraft observatories such as GeoEye IKONOS have only been in orbit since 1999 and thus would not provide relevant information on past disturbances.

Alternatively, the Landsat images in an LTSS can provide pre- and post-disturbance observations for disturbances that occurred during the time period of that LTSS. The spectral change signals of most forest stand disturbances can be identified reliably by experienced image analysts through visual examination of Landsat images acquired both before and after a particular disturbance event (Cohen et al., 1998; Masek et al., 2008; Huang et al., 2009b). Based on this observation, a design-based accuracy assessment method was developed for validating the NAFD disturbance year product, with a goal of obtaining unbiased accuracy estimates for each of the 6 validation scenes.

3.1.1. NAFD sampling design

There are many methods for sample selection in accuracy assessment (Foreman, 1991). For the NAFD project, the main issue

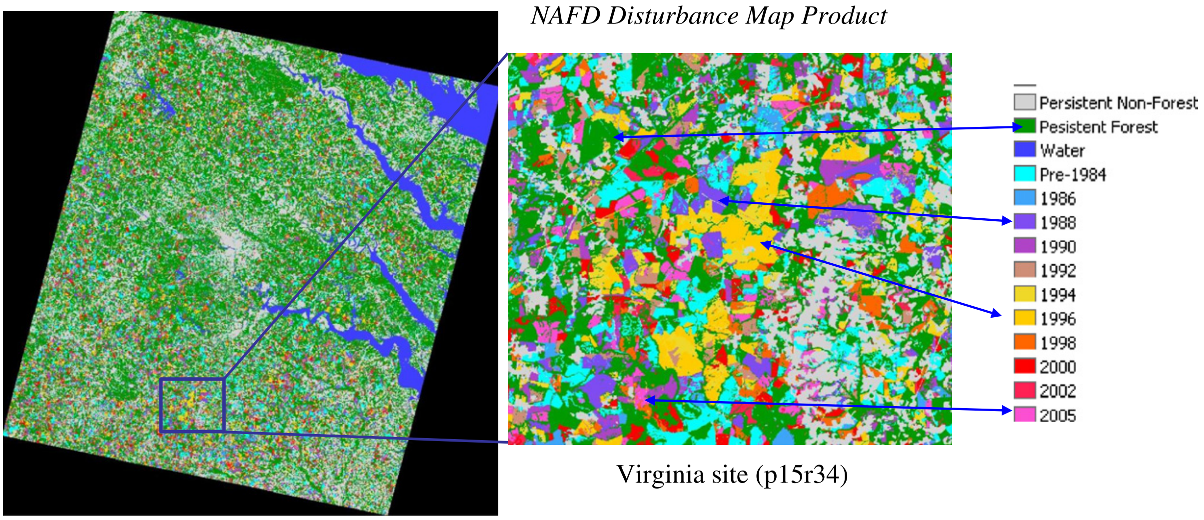


Fig. 2. The legend at right details the map classification system. The first three map categories are static classes which are consistent throughout the time series: persistent nonforest, persistent forest, and water. Forest change pixels are classified according to the year in which change occurred. Actual disturbance year classes vary according to the image dates present in each individual LTSS.

Table 1
Characteristics of the 6 NAFD sites where the disturbance products were evaluated using a design-based accuracy assessment method and FIA field inventory data.

WRS2 path/row	Location	Land cover and forest characteristics	Major disturbances
12/31	South Eastern New England	Mostly temperate deciduous forests, agriculture, urban	Urbanization, harvest
15/34	Virginia	Pine plantation, deciduous or mixed forests, agriculture	Urbanization, harvest
21/37	Mississippi/Alabama	Pine plantation, deciduous or mixed forests, agriculture	Harvest
27/27	Minnesota	Temperate deciduous and mixed forests, agriculture, wetlands	Wind throw, ice damage, harvest
37/34	Southern Utah	Semiarid, mostly shrub and grassland, pinion/juniper forests are typically short and sparse	Fire
45/29	Oregon	Temperate evergreen forests to the west, dry grass and shrubs in the middle and to the east	Fire, harvest, fuel treatment

addressed is that the class proportions are highly unbalanced. Because forest disturbance is a localized event, the area disturbed in any given year is typically much smaller than the areas of persisting forest or persisting nonforest. To ensure that the accuracies of individual disturbance year classes were derived with adequate precision we employed stratified random sampling (Cochran, 1977). A preliminary version of the VCT disturbance map prior to MMU filtering was used to define the scene strata for each validation scene. In each VCT map, all classes, including the individual disturbance year classes and the persisting classes, are considered strata. Only the map layer corresponding to initial disturbance was used to define strata. For each stratum, the inclusion probability of the samples in that stratum was the ratio of the number of samples selected within that stratum over the total pixels of that stratum (Stehman et al., 2003). Known inclusion probabilities allowed for design-based inference on the accuracy of the NAFD disturbance products.

In order to work within our available resources and to achieve satisfactory precision with the individual year estimates we targeted a maximum number of overall samples for each site equal to 50 samples per class, with a minimum of 30 samples for rare change classes (Richards, 1993). The total number of validation points selected for each site ranged from 645 to 750, depending on the number of time steps that comprise each individual LTSS, which varies between 13 and 15 for the six sites (Table 2). Because single pixels may be difficult to co-locate precisely on reference data, whether from the field or high resolution imagery (Congalton & Green, 1999), each sample was a 3 × 3 TM pixel block, centered at the sample pixel location. This block size is slightly larger than the MMU of the disturbance map product (Section 2.3).

3.1.2. NAFD response design

Response design is the method used to designate reference labels for each validation sample (Stehman & Czaplewski, 1998). For the design-based samples, we first visually assessed the high spatial resolution imagery to determine local land cover and use conditions. Land cover type labels correspond to the National Land Cover Data set (NLCD) 1992 project's modified Anderson Level 1 classification scheme (Vogelmann et al., 2001). The Landsat images were inspected in sequence from the earliest to most recent data in ArcMap to determine whether and when disturbances occurred at each sample location.

Table 2
Number of reference samples used by the two validation methods. FIA data used for this study are annual inventory single-condition plots intersecting with the NAFD LTSS.

Path/row	Assessment using FIA data				Design-based accuracy assessment			
	Nonforest	Old forest	Young forest	Total	Nonforest	Old forest	Young forest	Total
12/31	58	82	2	142	198	219	280	697
15/34	201	133	96	430	127	131	392	650
21/37	236	219	220	675	104	102	494	700
27/27	408	823	129	1360	102	188	460	750
37/34	263	167	3	433	245	195	205	645
45/29	192	246	9	447	180	140	380	700

The high spatial resolution image source acquired was primarily 1-m DOQQs from TerraServer (http://www.terraserver.com). If a DOQQ was not available, was of poor quality, or was captured prior to a forest disturbance event, other sources were used including USGS NAIP (http://datagateway.nrcs.usda.gov) and Google Earth (http://earth.google.com) imagery. Depending on the site, aerial photography might be available as panchromatic, natural color (red, green, and blue), or color infrared imagery, with Google Earth imagery available as natural color.

Information recorded for each sample location included: acquisition date of high resolution imagery; land cover class at the first and last time steps; disturbance class (corresponding to the classification system in Fig. 2); and comments. For each sample interpreted as forest disturbance through visual analysis, disturbance magnitude (partial clearing vs. stand clearing) and disturbance type were also recorded. These disturbance characteristics were determined based on both spectral and spatial image information, including landscape pattern, context, texture, shape, and location. Recent Google Earth images were invaluable in determining up-to-date land cover. Four types of disturbance were identified using the LTSS and high resolution image visualization including 3 stand clearing categories and 1 non-stand clearing category as detailed in Table 3. Although subtle changes in forest canopy could often be identified through visual analysis, in many cases it was not possible to determine if a non-stand clearing change was caused by thinning, other management practices, or storm, insect, or disease (Fig. 3). When possible, available ancillary data was reviewed to help assess change type.

For the Mississippi/Alabama sample site (21/37), where substantial multiple disturbances were observed within the 1985–2005 time period, an additional attribute for year of second disturbance was also recorded. While multiple disturbances in the same location could occur in all of the 6 validation sites, only information on the initial disturbance was used to compare to the initial disturbance map layer for the other 5 sites. Information on subsequent disturbance events was noted in the comments attribute for all sites.

Points that could not be confidently labeled were re-evaluated by the project manager (Thomas) and other project staff. In addition, if a validation sample was located on an edge between differing land cover types (such as forest and nonforest) as identified in the high resolution imagery, the validation sample was relocated within 3 TM pixels from the original location, to avoid confusion caused by misregistration. If a forest change patch was within 3 pixels from the original location, the validation pixel was moved to the disturbed patch. If no change patches were present, the sample pixel was moved to the nearest homogenous static class patch. Although avoiding mixed land cover patches will bias the sample toward homogenous regions and under-represent error at land cover edges, we choose this approach to reduce location error. We do not expect this bias to be significant as a fairly small number of sample locations were affected (<5%).

3.2. Assessment using FIA plot data

The FIA program was designed to provide information about U.S. forest resources at the national scale using field data collected at plot locations distributed across the U.S. (Smith, 2002). FIA field inventory

Table 3
General characteristics of disturbance types identifiable through visual time series analysis.

Disturbance type	Description	Spatial	Temporal	Spectral
Non-stand clearing	Partial removal of biomass. Includes variety of events: forest management such as thinning or understory burn, defoliatation due to insects, disease, or climate	Variable patch size, could be difficult to distinguish from surrounding forest	Commonly characterized by quick return to forest (within 1–3 time steps)	Minor change from pre- and post-disturbance spectral signal. Individual spectral characteristics variable and dependent on change type.
Stand clearing				
Harvest	Clear-cut harvest (0–10% tree cover remaining)	Clearly defined patch size and shape (usually rectangular) with smooth and regular texture.	Stand removal harvest has slower return to forest, dependant on region, management practices, and site index.	Immediately following harvest, bright (high reflectance) across all spectral bands.
Conversion	Forest removed and landscape changed to other, nonforest land use (including urban, agriculture, bare ground, etc.)	Commonly regular patch size distinct in pattern from surrounding forested areas	Does not return to forest during time series	Spectral characteristics variable and dependent on change type: initial conversion commonly high across all bands. Urban conversion high in Blue band and low in NIR.
Natural	Tree mortality caused by environmental effects such as severe fire or storm damage. Note that environmental effects can be human-caused (fire)	Usually irregular patch shape and size	Commonly slow return to forest	Fires often characterized by largest changes in NIR and Mid-IR bands: NIR reflectance drops and Mid-IR increases from pre-fire conditions. For other natural events, response will vary.

has been carried out at a national scale since its inception in the early 20th century. However, substantial regional methods variations existed in early FIA data. Beginning in the late 1990s, the FIA program implemented new strategies designed to improve reporting cycles and to achieve better spatial and temporal consistencies (Bechtold & Patterson, 2005). Because of changes in plot design, location, and field methods over time, it is difficult to assess disturbances at the plot level across a range of observation dates. The FIA plot data used in this validation effort were collected following the implementation of the new annual plot design strategies in the late 1990s. FIA field data were collected in coordination with FIA personal from the USFS Northern Forest Research Station following FIA security protocols.

While the FIA does not collect field data for each plot at temporal or spatial frequencies that match those of the LTSS, FIA inventory data are the most reliable source of independent, ground-based information on U.S. forests. Therefore we also explored the use of FIA plot data for validating the NAFD disturbance products.

3.2.1. FIA sampling design

FIA plot locations are selected by dividing the U.S. into equal-sized non-overlapping grid cells of 2500 ha (or 5 km by 5 km). One plot is selected within each grid cell for field data collection. Each plot consists of 4 subplots, with each having a radius of about 7.3 m. The FIA design is intended to obtain unbiased estimates of forest attributes at the state, regional, and national levels. FIA plots are surveyed every 10 years in the west and 5 years in the east in subcycles during which 10%–20% of the plots are targeted each year. When NAFD acquired the FIA plot data in spring 2008, not all subcycles had been completed for each of the states intersecting with the 6 validation sites.

The FIA reference data set for this study was comprised of both forested and nonforested FIA plots falling within the 6 LTSS sites. To minimize the impact of mixed pixels on the derived accuracy estimates, we excluded multiple condition plots. Multiple conditions can refer to different land covers (i.e. forest and nonforest) within a single plot, and also can refer to multiple forest conditions within a single plot, such as varying stand ages or stand densities (USDA, 2007). The impact of multiple condition plots can be further complicated by residual geolocation errors with both the Landsat images and the plot data. For forested plots, only plots containing live tree information were used.

3.2.2. FIA response design

For each FIA plot, the field crew surveys all trees within the 4 subplots that have a diameter at breast height (DBH) of 12.7 cm or larger. Key attributes recorded for each tree include species, DBH, and height. A subplot is defined as forest or nonforest according to the following definitions:

- Forest land — at least 10% stocked by trees at the time of field visit with a minimum area of 1 acre and at least 120 feet in width (USDA, 2007). In addition, plots that had been 10% stocked in the past (and presumably will be again in the future) are also considered forest land. The FIA forest land definition varied by forest type; in some woodland species such as pinyon pine and juniper, 5% crown cover is considered forest. Subsequent to this study, FIA has redefined forest land to be consistent throughout all U.S. regions.
- Nonforest land — not meeting the definition of accessible forest land. Nonforest land includes areas subject to land uses which would prevent natural tree regeneration, including recreation, mowing, or grazing activities. Urban areas which have over 10% tree cover may be defined as nonforest.

The FIA data provides some information on disturbances, including damages due to insects, disease, fire, and weather (USDA, 2008). However, the information recorded in those columns may be incomplete and is often only recorded if the damage event occurs at the same time as the field visit. Additionally, forest management such as thinning and harvest has not been included in the FIA disturbance category, but are included in the NAFD definition. As FIA continues to collect data within the annual cycle design, information on the cause of tree mortality will become available for remeasured plots.

Our approach to exploiting the FIA observations was to categorize the field measures in a structure that directly relates to our NAFD classification. For comparison we aggregated both the FIA plot data and the NAFD disturbance maps into nonforest, old forest, and young forest (equivalent to the NAFD disturbed) classes. Because the FIA disturbance attributes were not well suited to match the NAFD change category, we used stand age as an indicator of the occurrence of disturbance. Stand age represents the average age of the dominant live overstory trees within the plot. Assuming forest growth starting soon after a stand clearing disturbance, the age of a newly generated forest stand is roughly the difference between the year of field survey and disturbance year. The inferred year of disturbance calculated from FIA stand age was compared to NAFD disturbance year for validation.

NAFD disturbance products contain no information on disturbances that occurred before the first acquisition year of the concerned LTSS ~(1982–1984 for most LTSS), so we divided the forest plots into a “young” forest group (≤23 years) and an “old” forest group (>23 years) according to stand age. The old forest class corresponds

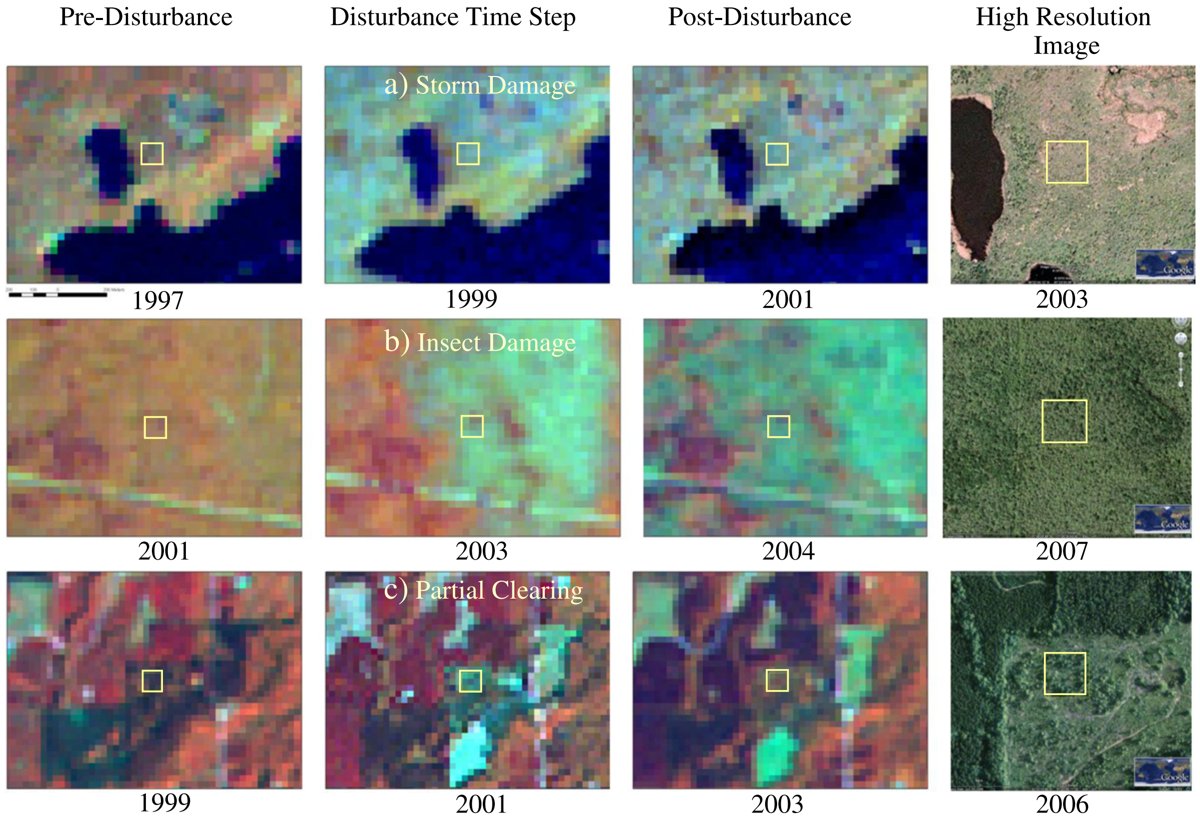


Fig. 3. The validation sites (3 × 3 pixel blocks) shown here illustrate visual identification of subtle disturbance types. Although we can identify that a disturbance has occurred, it may be difficult to confidently identify the type of disturbance without ancillary information. The blowdown event in a is located within the Boundary Waters Canoe Area Wilderness (BWCAW) and corresponds to major storm damage that occurred prior to this image on July 4th, 1999 (27/27). We corroborated our label of insect damage (b) with USDA Forest Health Monitoring (FHM) aerial survey data, which identified Forest Tent Caterpillar defoliation at this spatial and temporal location (12/31). We identified the partial clearing in c through visual analysis (21/37). The Landsat imagery is shown in Bands 4,5,3 combination as red, green, and blue.

to the persisting forest class in the NAFD products, and the young forest class corresponds to the disturbance classes. The NAFD disturbance map products were similarly aggregated into 3 classes to match the FIA reference data: persistent nonforest, persistent forest, and “young” or disturbed forest. Pre-series disturbance pixels were included in the “young” forest category.

4. Results and discussion

The reference data sets derived through the design-based method and FIA data were compared to the final MMU filtered disturbance maps. For each validation scene, an error matrix was created by applying appropriate weights to the reference samples, where the weight of each sample was adjusted based on its inclusion probability following Stehman et al. (2003). Accuracy measures were calculated according to Stehman and Czaplewski (1998) and Congalton (1991), including overall accuracy, kappa coefficient, and per class user's and producer's accuracy. Both the overall accuracy and kappa coefficient are measures of overall agreement between a disturbance year map and reference data. User's and producer's accuracies are related to commission (or false positive) and omission errors as follows (Janssen & van der Wel, 1994):

- Commission error (%) = 100% − User's accuracy (%)
- Omission error (%) = 100% − Producer's accuracy (%)

The results and discussion are organized into four sections. We present the error matrices and analyze accuracy measures from both the NAFD design-based and FIA assessments in Sections 4.1 and 4.2 respectively. Possible sources of error are identified by examining the class specific user's and producer's accuracies. In Section 4.3 we examine how accurately the LTSS-VCT approach captures various types of disturbance and address local disturbance patterns within each site. Lastly, we provide an assessment of potential biases in the NAFD national estimates based on this validation work (Section 4.4).

4.1. NAFD individual time step disturbance accuracies

The goal of the design-based sampling approach was to derive accuracy estimates at the individual time step level. Overall accuracy for the NAFD disturbance products ranged from 77% to 86% (Table 4a–d). Two of the 6 error matrices have recently been published in Huang et al. (2010) so are not reprinted here (Virginia 15/34, and Utah, 37/34). Overall map accuracy is calculated by summing the values in the primary diagonal and dividing by the number of samples. The kappa values show good agreement between the mapped and the reference data for 5 of the 6 sites, ranging from 0.67 to 0.76. The exception is the Utah site (37/24), which had a moderately low kappa value of 0.43. Results of the comparison with NAFD reference data are discussed in detail below, beginning with the static map classes.

4.1.1. Persistent nonforest

The NAFD products mapped the persistent nonforest class consistently well. Both producer's and user's accuracies for this category ranged from 85% to 99% for 5 out of 6 of the validation sites (Table 4a–d). The 6th site (Minnesota, p27r27) had a producer's accuracy of 77% and user's accuracy of 95% for nonforest samples. The majority of omission from nonforest in this site was due to misclassification of herbaceous wetland as persistent forest, with some wetland samples misclassified as disturbed forest.

Table 4

Error matrix for individual time step maps. Class codes are: PNF = persistent nonforest, PF = persistent forest, and P-SD = pre-series disturbance. Pre-series disturbance denotes samples that are not forested in time 1 but become forested by the last date of the time series. Additional labels correspond to year of disturbance (86 = 1986). Results are shown as area percentages such that a cell value of 10 refers to 10% of the LTSS. Note that the majority of the results reside in the primary diagonal where the reference label matches the map label, with smaller percentage values residing in the off-diagonals. The majority of error within forest change categories stems from confusion with the static classes (persistent nonforest and persistent forest). The off-diagonal elements in the matrix most frequently occur below the primary diagonal and are reflective of disturbances that are captured later in the time series or multiple disturbances, where a non-stand clearing disturbance such as partial harvest or storm/insect damage is later cleared.

(a) Southeastern New England (p12r31)

	Reference Data																
NAFD Map	PNF	PF	P-SD	87	89	91	93	95	96	98	00	01	03	04	Total	*n*	*User's*
PNF	**47.33**	0.33		0.99		0.09									48.74	155	97.11
PF	2.21	**29.28**	0.34	0.26	0.46	0.45	0.36	0.56	0.04	0.06	0.38	0.04	0.44	0.09	34.95	286	83.75
P-SD	3.20		**1.10**												4.30	20	25.68
87	0.09			**1.47**											1.56	12	94.21
89	0.11			0.66	**0.62**	0.04									1.43	21	43.24
91	1.22			0.36	0.12	**0.88**									2.58	29	33.96
93	0.33				0.04	0.04	**0.49**								0.90	20	54.65
95	0.33					0.02		**0.63**							0.97	22	64.35
96	0.02	0.01					0.03	0.02	**0.48**	0.01					0.55	25	86.39
98	0.33				0.02	0.02			0.02	**0.67**					1.06	26	63.32
00	0.33					0.02				0.04	**0.74**				1.13	26	65.56
01						0.02		0.02	0.02	0.04	0.06	**0.34**	0.02		0.52	19	65.79
03							0.01			0.02		0.03	**0.99**	0.04	1.09	29	90.74
04		0.04											0.03	**0.15**	0.21	7	69.14
Total	55.50	29.65	1.44	3.74	1.25	1.56	0.89	1.23	0.57	0.83	1.18	0.41	1.47	0.27	100.00		
n	198	219	12	22	27	30	21	27	25	30	30	14	34	8		697	
Producer's	85.28	98.73	76.57	39.34	49.62	56.16	54.97	51.18	84.31	80.68	62.84	83.11	66.97	54.11		*Overall*	**85.16**

(b) Mississippi/Alabama Site (p21r37)

	Reference Data																
NAFD Map	PNF	PF	P-SD	87	88	90	91	93	95	97	99	01	03	05	Total	*n*	*User's*
PNF	**26.37**		0.68	0.26		0.05									27.37	107	96.36
PF		**20.01**	1.00	0.97	1.88	1.74	0.90	2.33	1.50	0.60	0.70	0.91	1.07	0.21	33.82	180	59.17
P-SD			**7.56**	0.33											7.89	47	95.82
87		0.14	0.26	**5.03**	0.36				0.26						6.04	40	83.24
88	0.26		0.08	0.15	**1.60**	0.08									2.16	26	73.93
90			0.05	0.14		**2.24**	0.05								2.48	34	90.18
91				0.04	0.04		**1.18**								1.26	29	93.16
93			0.19		0.09		0.09	**2.18**							2.56	27	85.19
95					0.14		0.07		**2.18**	0.07					2.46	31	88.49
97	0.17				0.06	0.11	0.17	0.09		**2.22**					2.82	47	78.67
99			0.09		0.17	0.26	0.09	0.09	0.09		**2.39**				3.16	37	75.68
01			0.09			0.09	0.18	0.27		0.09		**1.80**		0.09	2.60	29	68.97
03				0.14	0.22	0.06	0.11	0.06	0.06				**1.37**	0.19	2.20	33	62.22
05		0.19	0.22	0.10	0.10	0.19		0.29	0.10	0.19		0.10		**1.71**	3.17	33	54.01
Total	26.79	20.34	10.22	7.15	4.67	4.81	2.84	5.30	4.18	3.16	3.09	2.80	2.44	2.21	100.00		
n	104	102	64	49	46	53	43	44	40	47	33	27	26	22		700	
Producer's	98.42	98.40	74.01	70.29	34.30	46.50	41.40	41.15	52.22	70.00	77.27	64.10	56.14	77.61		*Overall*	**77.83**

(c) Minnesota Site (p27r27)

	Reference Data																	
NAFD Map	PNF	PF	P-SD	86	89	90	91	93	95	97	98	99	01	03	06	Total	*n*	*User's*
PNF	**14.40**			0.43	0.21						0.06					15.10	72	95.35
PF	2.12	**44.41**	1.37	0.25	1.21	0.36	0.02	0.31	0.43	1.03	0.31	0.66	4.16	1.04	1.64	59.34	268	74.85
P-SD	0.11	0.22	**3.02**	0.22	0.21											3.77	34	80.03
86	0.03			**1.60**												1.63	27	98.04
89	0.25				**0.85**	0.03										1.13	29	75.31
90					0.09	**0.83**										0.92	17	90.41
91						0.02	**0.91**	0.24								1.17	28	78.03
93		0.03		0.03		0.03		**1.17**	0.16				0.21			1.62	33	72.09
95		0.03			0.03	0.03		0.03	**0.90**							1.02	27	88.71
97	0.37	0.09							0.06	**1.91**	0.06					2.48	42	76.87
98	0.34	0.06			0.06			0.03	0.12	0.12	**1.01**					1.73	27	58.47
99	0.06	0.06								0.03	0.06	**2.54**				2.75	37	92.16
01	0.69	0.10							0.05	0.06			**1.73**	0.15		2.78	43	62.22
03	0.03								0.07				0.34	**1.41**		1.85	35	76.48
06	0.26												0.09	0.18	**2.17**	2.69	31	80.39
Total	18.65	44.99	4.39	2.53	2.67	1.30	0.93	1.76	1.79	3.15	1.50	3.19	6.53	2.79	3.81	100.00		
n	102	188	36	32	42	22	26	31	40	42	22	38	55	39	35		750	
Producer's	77.19	98.71	68.82	63.38	31.96	64.03	97.75	66.36	50.45	60.53	67.34	79.42	26.51	50.69	56.85		*Overall*	**76.71**

(continued on next page)

Table 4 (*continued*)

(d) Oregon Site (p45r29)

	Reference Data																
NAFD Map	PNF	PF	P-SD	86	88	90	92	94	96	98	00	02	04	06	Total	*n*	*User's*
PNF	**50.84**	3.75	0.00	0.62	0.33	0.30	0.30		0.02						56.16	191	90.52
PF	1.29	**26.00**	1.34	1.04	0.84	0.04	1.07	0.04	0.05	1.06	0.02	0.41	0.06	0.01	33.27	173	78.16
P-SD	0.43	0.86	**1.73**	0.02											3.03	21	56.94
86		0.14	0.02	**0.54**	0.03										0.73	27	73.80
88		0.03			**0.91**										0.93	26	97.29
90					0.22	**0.94**	0.04			0.04	0.04	0.00	0.02		1.30	31	72.10
92		0.30			0.04		**0.82**		0.30						1.46	31	56.12
94						0.04	0.01	**0.43**	0.01						0.50	33	87.38
96					0.01				**0.18**						0.19	22	95.45
98					0.02				0.38	**0.14**					0.55	34	25.81
00					0.01					0.15	**0.27**				0.44	26	62.04
02	0.02	0.02				0.02	0.02		0.01	0.04	0.04	**0.26**	0.06		0.51	25	51.45
04					0.06	0.02	0.04			0.02			**0.49**		0.63	32	77.92
06					0.02		0.01					0.01	0.01	**0.25**	0.31	28	82.14
Total	52.58	31.10	3.09	2.23	2.49	1.36	2.31	0.48	0.95	1.46	0.38	0.68	0.65	0.27	100.00		
n	180	140	20	33	53	26	43	33	40	31	28	16	33	24		700	
Producer's	96.69	83.60	55.96	24.06	36.41	68.92	35.55	90.91	18.84	9.66	71.83	38.67	76.33	95.83		*Overall*	**83.80**

In addition, some nonforest pixels, such as agriculture and mixed urban forest pixels, were misclassified by the VCT as persistent forested pixels. These errors were most common in the New England (12/31) and Virginia (15/34) sites. Residual misregistration errors contributed to some of this confusion. Through LEDAPS orthorectification the average geolocation error of pixels within a single scene was less than 1 TM pixel (30 m). However, for multiple dates in a LTSS the registration error can be as high as ±1 pixel away from each other or ±30 m. These temporal registration errors, along with the impacts of sensor point spread function (Huang et al., 2002) and cubic convolution pixel resampling, contribute to class confusion at edges between classes such as forest and heterogeneous areas such as low density residential area (e.g. path 12/row 31, S.E. New England).

4.1.2. Persistent forest

This class generally had higher producer's accuracies than user's accuracies. Producer's accuracies for the persistent forest class ranged from 84% to 99% for 5 of the 6 sites. Persistent forest user's results for all validation sites varied from 57% to 84%. Most of the error in the persistent forest class resulted from omission from disturbed forest. For all of the sites there were disturbances detected in the reference data but misclassified as persisting forest in the NAFD products (the “persisting forest” row of the error matrices). These errors were generally caused by partial or non-stand clearing disturbance events such as selective logging, understory fire, or defoliation due to insect or storm damage. The VCT algorithm correctly identified these pixels as forest but failed to detect the partial disturbance.

Such disturbances typically resulted in partial removal of tree canopy in which the spectral change signal rapidly weakens with time. Depending on disturbance intensity and the rate of vegetation recovery processes, the canopy gaps resulting from a non-stand clearing disturbance can be filled in 1 or 2 years. This rapid regrowth is difficult to capture with the NAFD Phase I biennial time step approach, because the VCT algorithm requires that a pixel be flagged as disturbed for a minimum of two subsequent time steps to be detected as forest change (Huang et al., 2010). This requirement minimizes false positive errors due to cloud cover and phenological change but also may fail to detect partial disturbances that return to a forested signal within 4 years.

The Utah site (p37r34) had a persistent forest producer's accuracy of 46%. Only a small portion of the Utah site is forested and the forests are mostly short and sparse, which typically appeared much brighter than typical dark and dense forests found at other validation sites. As a result, substantial amounts of the sparse forests were mapped as persisting nonforest in the NAFD products. Additional factors which contributed to the errors at the path 37/row 34 site included changes in solar angles coupled with a rugged terrain.

Semiarid and sparsely vegetated regions as exemplified by the 37/34 site have temporal forest signatures influenced by understory vegetation in addition to the tree cover. Understory vegetation (primarily shrub and herbaceous cover) in semiarid regions is strongly affected by rainfall and seasonality. This scene condition variability can result in areas of permanent forest being misclassified as disturbed forest. The Oregon site (45/29) also experiences this type of error, but to a lesser degree, because only the central-eastern part of the area was in a semiarid environment.

4.1.3. Pre-series disturbance

The pre-series disturbance (P-SD) class was more difficult to define and assess then other NAFD classes. P-SD is identified when a pixel location is classed as nonforest in time step 1 and subsequently changes to forest (based on Forest Index threshold) later in the time series. A P-SD pixel could indicate regrowth from a disturbance occurring in or before the first year of the LTSS, or a conversion of nonforest to forest. For a particular P-SD pixel, however, whether it indicates a conversion or regrowth cannot be determined definitively using the LTSS.

4.1.4. Individual time step disturbance classes

The user's and producer's accuracies of the individual time step disturbance classes varied substantially from one year to another and among the 6 validation sites (Table 5). This is due in part to the relatively small sample sizes (extremely small in some year classes for

Table 5
Overall accuracy and Kappa Statistic are calculated for each site. We also include the average user's accuracy for the disturbed forest classes, calculated from the error matrices (Table 4). Average user's accuracy was also calculated from an additional set of error matrices, where ±1 time step is allowed as a correct match for each disturbance year. The pre-series disturbance class is not included here as a change class.

Path/ row	Overall accuracy	Kappa	Average user's accuracy for forest change classes	Average user's accuracy for forest change classes ±1
12/31	85.16	0.76	66.49	75.84
15/34	80.28	0.75	78.21	85.56
21/37	77.83	0.74	77.61	81.08
27/27	76.71	0.67	79.1	86.67
37/34	85.83	0.43	55.37	64.27
45/29	83.8	0.73	71.05	85.74

Table 6
Error matrix for Mississippi/Alabama site where multiple disturbances are incorporated. Average user's accuracy for forest disturbance classes is 90%. For this site, the reference data includes one label for first disturbance and an additional label for second disturbance occurring at the same sample location. This matrix differs from the individual time step results by also considering sample points a correct match if the VCT map label matches the reference data label for second disturbance. Samples showing multiple disturbances are frequently characterized by forest management activities such as thinning or understory burning early in the time series, followed by stand clearing harvest later in the series. The VCT algorithm captures the stand clearing harvest, but may miss the earlier partial disturbance.

Mississippi/Alabama Site (21/37)

NAFD Map	PNF	PF	P-SD	87	88	90	91	93	95	97	99	01	03	05	Total	*User's aa*
PNF	**26.37**		0.68	0.26		0.05									27.37	96.36
PF		**20.01**	1.00	0.97	1.88	1.74	0.90	2.33	1.50	0.60	0.70	0.91	1.07	0.21	33.82	59.17
P-SD			**7.56**	0.33											7.89	95.82
87		0.14	0.26	**5.03**	0.36										6.04	83.24
88	0.26		0.08	0.08	**1.68**										2.16	77.45
90						**2.43**	0.05								2.48	97.83
91					0.04		**1.22**								1.26	96.58
93			0.19				0.09	**2.28**							2.56	88.89
95									**2.39**	0.07					2.46	97.12
97	0.17							0.09		**2.56**					2.82	90.77
99									0.09		**3.07**				3.16	97.30
01			0.09							0.09		**2.33**		0.09	2.60	89.66
03					0.06	0.06	0.06	0.06					**1.78**	0.19	2.20	81.16
05		0.19												**2.98**	3.17	94.00
Total	26.79	20.34	9.86	6.67	4.02	4.35	2.32	4.75	4.24	3.31	3.78	3.25	2.85	3.47	100.00	
Producer's aa	98.42	98.40	76.71	75.43	41.76	55.78	52.51	47.86	56.48	77.10	81.38	71.94	62.54	85.78	*Overall*	**81.69**

the Utah site) in disturbed forest as compared to sample sizes for persistent forest and nonforest classes (see "n" values in Table 4a–d). On average these individual time step disturbance classes had a user's accuracy of 55% at the Utah site (37/34), 67% at the southeastern New England site (12/31), and over 70% at the other 4 validation sites. The average producer's accuracies for these classes were slightly lower. This suggests that although disturbances at each individual time step were typically rare (up to 1%–3% of total area per disturbance year) as compared to the persistent forest and nonforest classes, on average the NAFD disturbance products were able to capture more than half of those disturbances with relatively low levels (i.e., <30% for 4 of the 6 validation sites) of commission errors.

4.1.4.1. Biennial time step uncertainty. There were samples for each of the sites where the disturbance year differed by one time step between the NAFD products and the reference data. These errors point to inconsistencies between the image analyst and the VCT algorithm in determining the exact year of a disturbance event, primarily where selective logging occurred in the year prior to stand clearing harvests. Conversion events from forest to other land cover types may also take place over more than one time step and thus be difficult to identify as a single date. Occasional cloud cover also can confuse identification of the precise time step of disturbance.

To better understand the effects of these errors on the accuracy results, the error matrices (Table 4a–d) were recalculated to allow ±1 time step from strict agreement between reference and map data to be counted as a correct match. Average user's accuracy for forest disturbance classes increases for all validation sites (Table 5). Note that the user's accuracies increase by an average of 9% indicating that if annual rather than biennial time series stacks had been used in this analysis the results would have been incrementally improved.

4.1.4.2. Multiple disturbances within LTSS time period. On average, the LTSS time period covers a 21 year period. More than 1 forest disturbance can be observed during this period, particularly in the southeastern United States. We observe that disturbed forest omission errors are primarily located below the prime diagonal of the error matrices (Table 4a–d). A re-examination of the misclassified samples at these sites confirmed that most of them had multiple disturbances, where a non-stand clearing disturbance in an early year (such as thinning or fire treatment) was followed by a major disturbance (harvest) in a later year.

This suggests that some early-year disturbances observed in the reference data were not mapped by VCT on this earlier date but the more significant later disturbance was recorded in the NAFD product. Multiple disturbance error is most common in the Mississippi/Alabama site (21/37) (Table 4c) and also evident but less obvious in the Oregon site (p45/29) (Table 4d). An additional error matrix was generated to identify how much map error was due to multiple disturbances for the MS/AL site (Table 6) where the VCT result is assumed correct if it corresponds to either the first or second disturbance as recorded in the reference data. In this case the results show an average disturbance class time step accuracy of 90% versus 78% for the single disturbance assessment. This comparison also suggests that reference datasets generated for any future work should be designed to account for 2 or more possible disturbance events at any single location.

4.2. Comparison with FIA plot data

For each of the 6 validation sites, we derived results using both FIA plot and design-based reference datasets at the 3-class level (Table 7a–f). Overall agreements between NAFD disturbance maps and FIA plot data were 67% at the Utah site (37/34) and between 79% and 84% at the other 5 sites. At the same classification level, the overall agreement between the NAFD products and the design-based reference data ranged from 82% to 87%. The overall accuracies resulting from comparison with the two separate reference data sets are similar for the validation sites, with the exception of the Utah site.

We also see general agreement between producer's and user's accuracies in forest and nonforest classes between the two assessments for these sites. The majority of disagreement between the two sets of results can be attributed to differences in sampling design and class definition between the two assessment methods. Although the original sampling designs of both of these datasets (FIA and design-based) allow derivation of unbiased estimates, the FIA plot data used for this study excludes multiple condition plots, which comprise ~35% of all FIA plots in a given location (see Section 3.2). Because not all plots are included, the estimates derived using the FIA data in this study are not unbiased, resulting in area proportion results (column totals of error matrices) that diverge between the 2 validation methods (Table 7a–f). Additionally, the southeastern New England site (12/31) contains a significantly higher proportion of nonforest in the NAFD disturbance map, because the LTSS footprint includes ~30%

Table 7
Error matrices showing results from FIA assessment (left column) and design-based assessment (right column) for each validation site. Reference and map data have been aggregated to 3 classes for comparison: nonforest, forest, and disturbed forest. VCT and design-based labels are PNF (persistent nonforest) and PF (persistent forest) . FIA disturbed forest class corresponds to FIA plots where the stand age attribute (minus field measurement year) ≤23 years. Design-based assessment disturbed forest class includes all individual time step disturbance classes. The water class has been grouped with nonforest for both assessments. Note that areal proportions do not necessarily match because not all FIA plots from the equal probability sample are not included.

(a) Southeastern New England site (12/31)

	FIA Reference				
NAFD Map	NF	PF	DF	Total	*User's*
PNF	**23.24**	0.70	0.00	23.94	97.06
PF	8.45	**56.34**	0.00	64.79	86.96
DF	9.15	0.70	**1.41**	11.27	12.50
Total	40.85	57.75	1.41	100.00	
Producer's	56.90	97.56	100.00	*Overall*	**80.99**

	Design-based Reference				
NAFD Map	PNF	PF	DF	Total	*User's*
PNF	**47.33**	0.33	1.08	48.74	97.11
PF	2.21	**29.28**	3.47	34.95	83.75
DF	5.96	0.04	**10.30**	16.31	63.18
Total	55.50	29.65	14.85	100.00	
Producer's	85.28	98.73	69.40	*Overall*	**86.91**

(b) Virginia Site (15/34)

	FIA Reference				
NAFD Map	NF	PF	DF	Total	*User's*
PNF	**36.50**	0.44	0.22	37.17	98.21
PF	3.10	**26.77**	2.65	32.52	82.31
DF	4.87	5.09	**20.35**	30.31	67.15
Total	44.47	32.30	23.23	100.00	
Producer's	82.09	82.88	87.62	*Overall*	**83.63**

	Design-based Reference				
NAFD Map	PNF	PF	DF	Total	*User's*
PNF	**29.23**	0.65	0.71	30.59	95.56
PF	0.78	**27.83**	10.23	38.84	71.65
DF	1.97	1.20	**27.41**	30.57	89.65
Total	31.97	29.68	38.35	100.00	
Producer's	91.41	93.77	71.48	*Overall*	**84.47**

(c) Mississippi/Alabama site (21/37)

	FIA Reference				
NAFD Map	NF	PF	DF	Total	User's
PNF	**31.70**	0.30	2.96	34.96	90.68
PF	1.19	**26.22**	6.96	34.37	76.29
DF	2.07	5.93	**22.67**	30.67	73.91
Total	34.96	32.44	32.59	100.00	
Producer's	90.68	80.82	69.55	*Overall*	**80.59**

	Design-based Reference				
NAFD Map	PNF	PF	DF	Total	*User's*
PNF	**26.37**	0.00	1.00	27.37	96.36
PF	0.00	**20.01**	13.81	33.82	59.17
DF	0.42	0.33	**38.06**	38.81	98.07
Total	26.79	20.34	52.87	100.00	
Producer's	98.42	98.40	71.99	*Overall*	**84.44**

(d) Minnesota site (27/27)

	FIA Reference				
NAFD Map	NF	PF	DF	Total	*User's*
PNF	**18.50**	0.35	0.35	19.21	96.32
PF	6.21	**55.44**	4.45	66.10	83.87
DF	4.10	4.10	**6.50**	14.69	44.23
Total	28.81	59.89	11.30	100.00	
Producer's	64.22	92.57	57.50	*Overall*	**80.44**

	Design-based Reference				
NAFD Map	PNF	PF	DF	Total	*User's*
PNF	**14.40**	0.00	0.70	15.10	95.35
PF	2.12	**44.41**	12.80	59.34	74.85
DF	2.13	0.58	**22.85**	25.56	89.38
Total	18.65	44.99	36.35	100.00	
Producer's	77.19	98.71	62.85	*Overall*	**81.66**

(e) Utah site (37/34)

	FIA Reference				
NAFD Map	NF	PF	DF	Total	*User's*
PNF	**60.74**	32.56	0.46	93.76	64.78
PF	0.00	**5.77**	0.23	6.00	96.15
DF	0.00	0.23	**0.00**	0.23	0.00
Total	60.74	38.57	0.69	100.00	
Producer's	100.00	14.97	0.00	*Overall*	**66.51**

	Design-based Reference				
NAFD Map	PNF	PF	DF	Total	*User's*
PNF	**80.00**	6.18	2.26	88.43	90.47
PF	2.59	**5.28**	1.43	9.30	56.76
DF	1.28	0.16	**0.83**	2.27	36.61
Total	83.87	11.61	4.52	100.00	
Producer's	95.39	45.46	18.36	*Overall*	**86.11**

(f) Oregon site (45/29)

	FIA Reference				
NAFD Map	PNF	PF	DF	Total	*User's*
PNF	**42.51**	14.77	0.45	57.72	73.64
PF	0.22	**35.12**	0.22	35.57	98.74
DF	0.22	5.15	**1.34**	6.71	20.00
Total	42.95	55.03	2.01	100.00	
Producer's	98.96	63.82	66.67	*Overall*	**78.97**

	Design-based Reference				
NAFD Map	PNF	PF	DF	Total	*User's*
PNF	**50.84**	3.75	1.57	56.16	90.52
PF	1.29	**26.00**	5.98	33.27	78.16
DF	0.45	1.35	**8.77**	10.57	83.00
Total	52.58	31.10	16.32	100.00	
Producer's	96.69	83.60	53.76	*Overall*	**85.62**

coastal waters for this site, which are not included in the FIA sample data.

As noted in Section 3.2.2, FIA inventory data labels and corresponding definitions differ from the classification system used for NAFD VCT disturbance maps. For example, we found several plots in the FIA data labeled as nonforest which are also correctly labeled in the NAFD maps as disturbed forest. This can occur when forested plots have been converted to a nonforest land use prior to FIA field visit.

Temporally incompatible labels violate commonly accepted standards of mutually exclusive class labels, where each sample can only have one correct label (Congalton & Green, 1999). More significantly, at the time the field data was assembled, FIA defined forest as having a minimum of 5% tree cover in certain woodland species such as pinyon pine and juniper, typical of the Utah (p37r34) site. Here, overall accuracy is significantly lower in the FIA assessment than design-based assessment (67% vs. 86%). For the design-based assessment, analysts applied a 10% forest cover (as estimated from high spatial resolution imagery) as the threshold for forested class across all regions. A change of tree cover from 10% to 5% in the definition of forest cover can result in different land cover labels for many plots. While the results of the design-based and FIA assessments are not equivalent, they are encouragingly similar given the differences in temporal domain, sampling method, and classification rules.

4.2.1. Disturbance year and stand age

Also significant is the difference between the definitions of disturbed forest in the FIA and NAFD schemes. As discussed in Section 3.2.2, we used FIA stand age as an indicator of disturbance. Therefore, only stand clearing disturbances where forest stands are replanted soon after disturbance are directly comparable with NAFD disturbance classes. NAFD detected partial disturbances will be labeled as persistent or "old" forest in the FIA reference dataset.

Out of the 6 validation sites, 3 locations (12/31, 37/34, and 45/29) had less than 10 FIA samples each in the young forest category (Table 2). For the other 3 sites the relationship of FIA stand establishment date and VCT-derived disturbance year is strong (Fig. 4a–c). An exact match may not always be possible due to the difficulty of identifying the exact age of young trees in the field and the biennial time step in VCT analysis. Visual assessment of young forest plots revealed that outliers (greater than 5 year difference in stand establishment between the FIA and NAFD derived dates) often occurred on edge plots located in between different land cover types or forest stands.

Although the FIA sample design does not always adequately capture forest disturbance events at a local scale because of the plot density and uniform sample approach employed, visual assessment of separate land cover types and disturbance results show that where FIA field measurements record disturbances, the VCT approach produces a reliable assessment of when and where these events occur.

4.3. Disturbance type and mapping accuracy

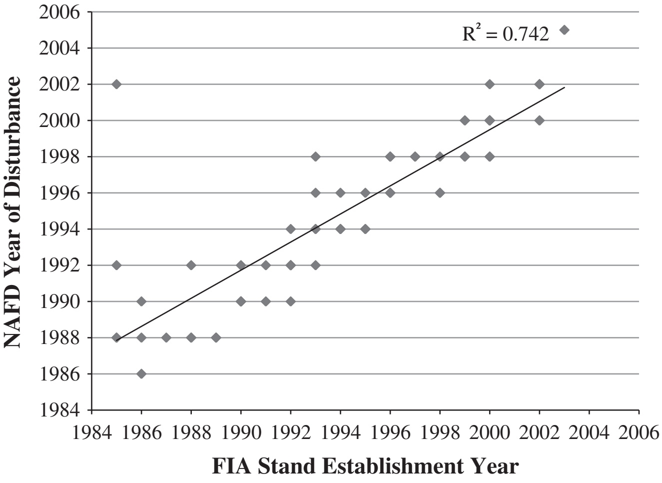
We characterized the success rates of the NAFD disturbance maps by disturbance type identified from the design-based study to examine omission from the disturbed forest classes (Fig. 5). At the individual time step level combined over all 6 validation sites, VCT correctly mapped over 85% of the stand clearing harvest and over 71% of stand clearing natural disturbances (mainly fire) and conversions from forest to other land covers.

As discussed in Section 4.1.2, the LTSS-VCT technique is less successful in capturing non-stand clearing events. The NAFD mapping approach identifies the correct disturbance time step in 38% of non-stand clearing disturbances. When the data is aggregated to the 3-class level and individual disturbance years are grouped into one change class, the detection accuracy of disturbed non-stand clearing disturbance improves to over 60%, The improvement is due to multiple disturbance locations (see Section 4.1.4.2). Stand clearing disturbance detection accuracy increased to over 92% at the aggregated 3-class level.

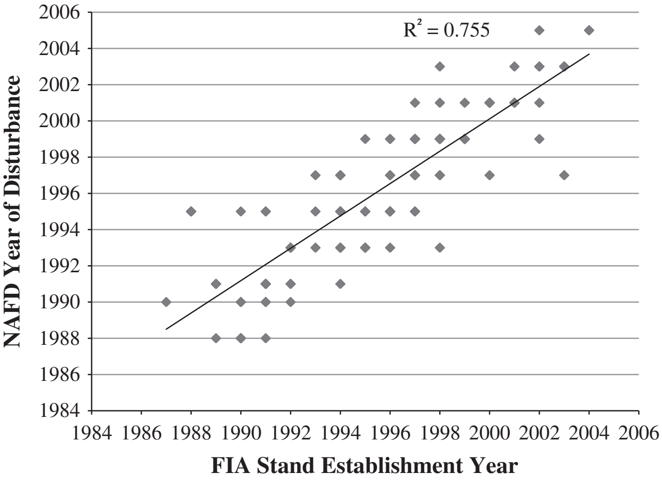
4.3.1. Cumulative disturbed area

Cumulative area of forest disturbance varies considerably among the 6 sites, as does the influence of major disturbance drivers (Fig. 6). The proportion of different land cover types within each validation site, in addition to the proportion of different change types identified within disturbed forest, are estimated from the final NAFD disturbance maps. In the Virginia (15/34) and MS/AL

(a) Virginia Site (15/34)



(b) Mississippi/Alabama Site (21/37)



(c) Minnesota Site (27/27)

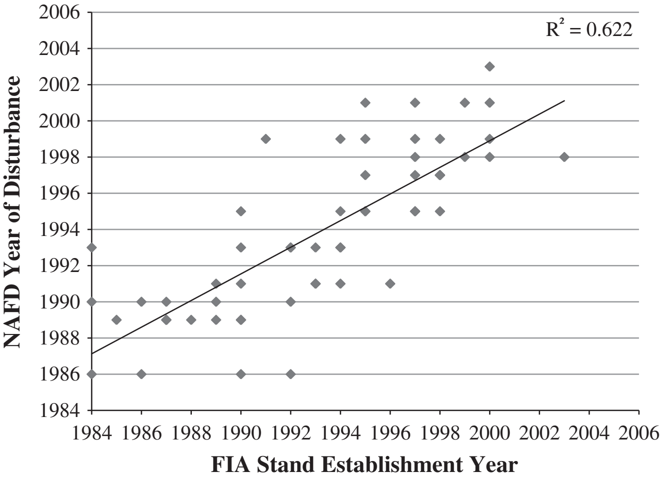
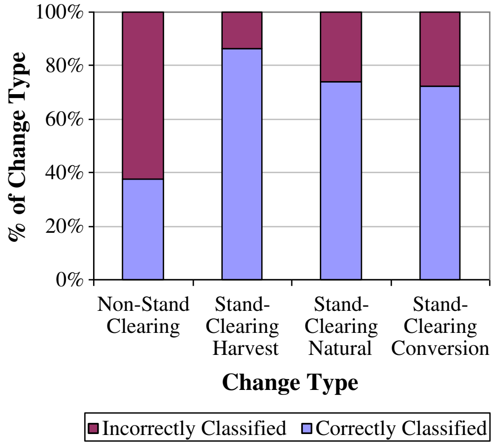


Fig. 4. FIA stand age establishment date plotted against VCT-derived year of disturbance, for all FIA single-condition forested plots where stand age ≤23 years and disturbance map product shows disturbed forest.

(a) Individual year results



(b) Aggregated 3-class results

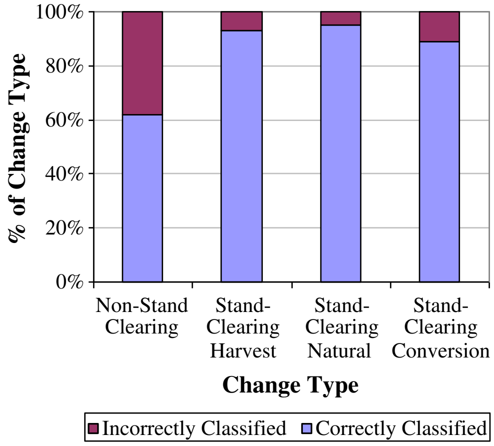


Fig. 5. Design-based assessment results showing VCT detection accuracy for different change types for the six validation sites combined. Non-stand clearing disturbance can have a wide variety of causes , such as thinning and partial damage due to storm, disease, or insect defoliation. We identified stand clearing disturbances as belonging to one of three categories: harvest, stand clearing natural disturbances (including severe fire and storm damage), and conversion from forest to other land use categories. Differences between a and b highlight the fact that the VCT technique may not identify disturbance at the same year as analyst-derived reference data but can often capture changes at subsequent time steps.

(21/37) sites over 20% of the land area is disturbed forest over the time period (1984–2005). This land area increases to 27% (15/34) and 36% (21/37) if we include pre-series disturbance within the disturbed forest category.

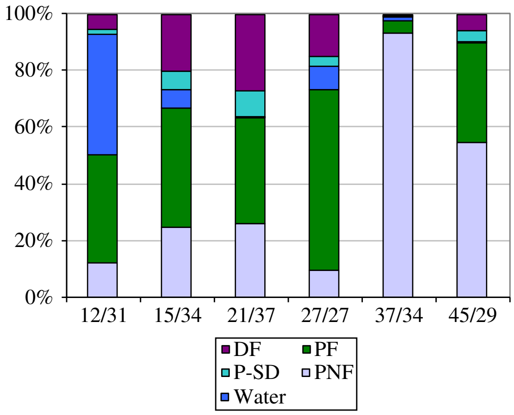
Disturbed forest constitutes a small percentage of the overall landscape in some sites, particularly Utah (37/34), with less than 1% of land area (Fig. 6a). This is also the case although to a lesser degree for the southeastern New England (12/31) and Oregon (45/29) sites with 5% and 6% disturbed forest land area respectively. It is important to note that the southeastern New England LTSS (12/31) footprint extends into coastal water, causing the large water area proportion in that scene.

4.3.2. Local disturbance drivers

The disturbed area by disturbance category shows that for these 6 sites, the majority disturbance categories are harvest and non-stand clearing (Fig. 6b). Stand clearing harvest is the most prevalent type of forest disturbance in four out of six of the validation sites (15/34, 21/37, 27/27, and to a lesser degree 45/29). The Virginia (15/34) and Mississippi/Alabama sites are similar in many characteristics including average user's accuracy for disturbed forest (78%). Intensive forest management occurs throughout both study areas, primarily on private land. Multiple disturbances as discussed in Section 4.1.4.2 are prevalent in these sites. Urban conversion is a common but less intensive disturbance driver in the Virginia and MS/AL sites, with the majority of suburbanization occurring in the greater Richmond and Birmingham areas. Additional examples of forest converted to suburban developments are present within both sites.

The primary disturbance forces within the Minnesota (p27r27) site include forest management and suburbanization, similar to 15/34 and 21/37. This site was selected by FIA as a scene of particular interest because of a major windstorm event that occurred in northern Minnesota on July 4th, 1999. FIA analysts are studying blowdown from this event within the Boundary Waters Canoe Area Wilderness (BWCAW) (Nelson et al., 2009). The LTSS includes an image acquired soon after the event (July 24, 1999), and the subsequent time step was imaged on July 5, 2001. Missed disturbance errors (omission from disturbed forest) are relatively high for this site in 2001 (4.16% of area). The majority of missed disturbance was related to the 1999 windstorm and appeared from visual assessment to be partial disturbances, due either to storm related damage or

(a) Proportion of land cover classes



(b) Disturbed forest change type

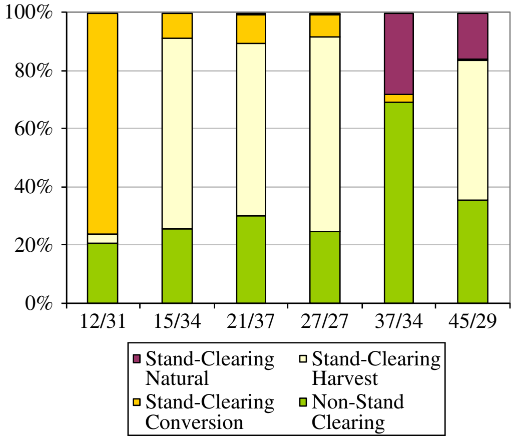


Fig. 6. Proportion of land cover classes as calculated from the final NAFD disturbance map product (a) and disturbance types from reference data (b) within each LTSS. The DF (disturbed forest) category in a is evaluated in b so that each forest disturbance is attributed to a change driver. The proportion of disturbed forest includes all mapped disturbances over the entire time series (1984–2005). Note that the S.E. New England (12/31) LTSS includes coastal water, which accounts for the large proportion in that category. In the validation work, the water class was included in PNF.

possibly to salvage logging. Section 4.2.1 discusses errors due to misclassified wetland, either as forested or as disturbed forest.

The Oregon site (45/29) contains portions of the central Cascades as well as the semiarid region east of the Cascades and as such exhibits a wide range of environmental conditions. The western half of this scene is primarily coniferous forest where disturbance is dominated by harvest and natural disturbances (fire). Large fire events as identified in the Monitoring Trends in Burn Severity (MTBS) dataset occurred during the observation period, including the Simnasho fire in 1996 which burned approximately 47,739 ha, mostly in shrubland, and the Link fire in 2003 (41,306 ha) (http://www.mtbs.gov). Visual analysis of the validation points within the boundary of the Link fire show this event well characterized by the NAFD disturbance product. Misclassified samples are generally due to missed minor disturbance that occurred earlier in the time series, primarily from multiple burns. The eastern portion of this LTSS is in the rainshadow of the Cascades and is similar in characteristics and results to the Utah site. In this region, the most common classification error is shrubland labeled as either persistent or disturbed forest.

The majority of disturbance in the New England (12/31) validation site was conversion of forest to suburbia. Most map errors are coincident with locations identified by the analysts as persistent nonforest, commonly occurring in suburban or exurban areas with mixed tree and urban pixels. Much of this error is due to residual misregistration, as discussed in Section 4.1.1. Commission errors with disturbed forest occur where the NAFD map mislabels persistent nonforest (usually mixed urban and treed pixels) as forest disturbance. In a few samples, herbaceous wetland and agricultural fields were misclassified as forest change.

The Utah (37/34) site had an average forest disturbance user's accuracies of 55%. As discussed in Section 4.1.2, sparsely forested areas may be misclassified by VCT as nonforest. In addition, nonforest or persistent sparse forest pixels may be misclassified as disturbed forest because of significant inter-annual variations in vegetation phenology due to changing precipitation patterns. According to the U.S. Drought Monitor (http://drought.unl.edu/dm), southern Utah experienced abnormally dry conditions in the summer of 2000 and severe, extreme or exceptionally dry conditions (dependent on location) during the summers of 2002, 2003, and 2004. Drought conditions contributed to subtle forest change that was flagged by image analysts but not always captured by the VCT algorithm.

4.4. Estimation of disturbance rates

Mapping errors in the NAFD disturbance products, such as underestimation of partial disturbances, may introduce biases to disturbance rates calculated from these products. One of the main objectives of this research is to understand these biases to improve estimates of disturbance rates for downstream applications. The error matrices derived using the design-based assessment method show the proportion of disturbed area at the individual time step level (Table 4a–d) and over the entire observing period of each LTSS (Table 7a–f, right column). The NAFD estimates of disturbance rates generally track those derived using the reference data (reflected in the Grand Total column and the Grand Total row, respectively).

However, NAFD disturbance products underestimate the proportion of total disturbed area over the entire observing period of each LTSS in 5 of the 6 validation sites (Table 8). As discussed in Sections 4.1.2 and 4.3, the majority of missed disturbances are non-stand clearing events that are difficult to capture using biennial LTSS (Fig. 5). The underestimation of stand clearing disturbances by the NAFD products will be much lower than those shown in Table 8. For the southeastern New England site (12/31), the overestimate of disturbance rate in the NAFD product was due to confusion in mixed urban and treed areas.

Table 8
Map bias calculated for disturbed forest by subtracting the reference data (column total) from the NAFD disturbance product data (row total). Values correspond to row and column totals shown in Table 7 from the design-based reference data.

Path/row	NAFD map	Design-based reference	Map − reference
12/31	16.31	14.85	1.96
15/34	30.57	38.35	−7.78
21/37	38.81	52.87	−14.06
27/27	25.56	36.35	−10.79
37/34	2.27	4.52	−2.25
45/29	10.57	16.32	−5.75

While similar comparisons on disturbance rates can be made at the individual time step level using the reference samples derived through the design-based assessment, one should be cautious in analyzing the differences, because those estimates were calculated using small numbers of samples (roughly 20–50 for each disturbance time step) and can have high levels of variance. Cumulative forest disturbance areas are best estimated from the final NAFD disturbance map products, using individual disturbance map years to ensure inclusion of multiple disturbances. Similarly, disturbance rates, as produced by NAFD for national estimates (Kennedy et al., in preparation) are estimated using individual time step disturbance maps.

5. Conclusions

NAFD disturbance products were validated using two complimentary approaches: 1) a design-based accuracy assessment method (Stehman, 2000) using high spatial resolution imagery in conjunction with visual analysis of the LTSS imagery and 2) comparison with FIA ground measurements. Because the two datasets differ substantially in key aspects such as class definitions and temporal coverage, we were unable to integrate them into one validation assessment. However, we found that the results are quite similar despite these differences. We incorporated both reference data sets in this research to provide a more comprehensive assessment of the NAFD disturbance products then would be possible with only one reference source.

We have found that the disturbance mapping approach developed in NAFD Phase I is generally successful although significant error terms remain. The results from this validation revealed that at the individual disturbance time step level the NAFD disturbance products at 5 of the 6 validation sites had overall accuracies ranging from 77% to 86%, with kappa values ranging from 0.67 to 0.76. The lowest accuracies were found at the 6th site (Utah, 37/34), where sparse forest cover contributed to map error (Section 4.1.2). The average user's accuracy for disturbed forest over all 6 sites is from 55% to 79%. Because individual time step VCT results find that for most sites on average 1%–2% of land area is disturbed in any given year, the accuracies reported here are <0.2% error in the estimated disturbed area in any given year.

Stand clearing disturbances, whether from harvest, conversion, or natural stand clearing disturbances such as fire, were well characterized across all regions. VCT correctly classified over 90% of the stand clearing harvest and over 88% of land cover conversions at the aggregated 3-class level. The majority of remaining map errors result from less effective mapping of non-stand clearing disturbances, such as thinning and partial damage from natural disturbance events.

NAFD has been funded to pursue a Phase II element that is currently underway. Several aspects of the NAFD approach will be improved in Phase II including adding ~27 site locations. Most critically, we will move to annual image stacks to improve the detection of partial disturbances. We found a significant increase (4%–15%) in user's accuracy for forest change classes when we allowed ±1 time step (Table 5). We anticipate an even greater increase in accuracy after we have moved to the annual time step in Phase II. Previously compilation of annual Landsat time series stacks was cost prohibitive. With the USGS Landsat data policy now in place this is no longer the case. We are also developing an automated cloud-clearing methodology that permits LTSS being compiled from nearly cloud-free images within the mid-summer growing season at an annual time step. We expect Phase II analysis, including validation, to be available in early 2011.

Acknowledgements

NAFD is a core project of the North American Carbon Program. The NAFD project and this study are supported by grants from NASA's Terrestrial Ecology, Carbon Cycle Science, and Applied Sciences Programs and with funding from the U.S. Geological Survey. The authors greatly thank Elizabeth LaPoint of USFS FIA National Spatial Data Services for assistance in working with FIA plot data. Kurtis Nelson from USGS/EROS and Andrew Lister from FIA kindly provided ArcMap tools to improve retrieval of DOQQ and NAIP imagery, and Stephen Howard from USGS/EROS facilitated access to MTBS/MRLC images. In addition, we'd like to thank our two anonymous reviewers for their insightful comments that greatly improved this paper.

References

Bechtold, W. A., & Patterson, P. L. (2005). *The enhanced forest inventory and analysis program — National sampling design and estimation procedures.* Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.

Climate Change Science Program. (2007). The first State of the Carbon Cycle Report (SOCCR): The North American carbon budget and implications for the global carbon cycle. In A. W. King, L. Dilling, G. P. Zimmerman, D. M. Fairman, R. A. Houghton, G. Marland, A. Z. Rose, & T. J. Wilbanks (Eds.), *A report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research*. Asheville, NC, USA: National Oceanic and Atmospheric Administration, National Climatic Data Center 242 pp.

Cochran, W. G. (1977). *Sampling techniques.* New York: Wiley.

Cohen, W. B., Fiorella, M., Gray, J., Helmer, E. H., & Anderson, K. (1998). An efficient and accurate method for mapping forest clear-cuts in the Pacific Northwest using Landsat imagery. *Photogrammetric Engineering and Remote Sensing*, *64*, 293−300.

Congalton, R. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*, 35−46.

Congalton, R. G., & Green, K. (1999). *Assessing the accuracy of remotely sensed data: Principles and practices.*Boca Raton: Lewis Publishers (2004).

Foreman, E. K. (1991). *Survey sampling principles.* New York: Marcel Dekker, Inc.

Gao, F., Masek, J., & Wolfe, R. (2009). An automated registration and orthorectification package for Landsat and Landsat-like data processing. *Journal of Applied Remote Sensing*, *3*, 033515.

Goward, S. N., Masek, J. G., Cohen, W., Moisen, G., Collatz, G. J., Healey, S., Houghton, R., Huang, C., Kennedy, R., Law, B., Turner, D., Powell, S., & Wulder, M. (2008). Forest disturbance and North American carbon flux. *EOS Transactions, American Geophysical Union*, *89*, 105−106.

Huang, C., Goward, S. N., Masek, J. G., Gao, F., Vermote, E. F., Thomas, N., Schleeweis, K., Kennedy, R. E., Zhu, Z., Eidenshink, J. C., & Townshend, J. R. G. (2009). Development of time series stacks of Landsat images for reconstructing forest disturbance history. *International Journal of Digital Earth*, *2*, 195−218.

Huang, C., Goward, S. N., Schleeweis, K., Thomas, N., Masek, J. G., & Zhu, Z. (2009). Dynamics of national forests assessed using the Landsat record: Case studies in Eastern U.S. *Remote Sensing of Environment*, *113*, 1430−1442.

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.

Huang, C., Townshend, J. R. G., Liang, S., Kalluri, S. N. V., & DeFries, R. S. (2002). Impact of sensor's point spread function on land cover characterization: Assessment and deconvolution. *Remote Sensing of Environment*, *80*, 203−212.

Janssen, L. L. F., & van der Wel, F. J. M. (1994). Accuracy assessment of satellite derived land-cover data: A review. *Photogrammetric Engineering and Remote Sensing*, *67*, 1067−1075.

Kennedy, R. E., Cohen, W. B., Moisen, G. G., Goward, S. N., Wulder, M., Powell, S. L., Masek, J. G., Huang, C., & Healey, S. P. (2006). A sample design for Landsat-based estimation of national trends in forest disturbance and regrowth. *NASA joint workshop on biodiversity, terrestrial ecology, and related applied sciences. August 21–25, 2006, College Park, MD*.

Kennedy, R. E., Cohen, W. B., Moisen, G. G., Goward, S. N., Wulder, M. A., Powell, S. L., et al. (in preparation). A Landsat-based sampling design for estimating three decades of forest disturbance dynamics in the contiguous United States.

Knight, J. F., & Lunetta, R. S. (2003). An experimental assessment of minimum mapping unit size. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 2132−2134.

Lillesand, T. W., & Kiefer, R. W. (1994). *Remote sensing and image interpretation.*New York: Wiley pp. 157.

Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing*, *25*, 2365−2407.

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., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., Feng, G., Kutler, J., & Teng-Kui, L. (2006). A Landsat surface reflectance dataset for North America, 1990–2000. *IEEE Geoscience and Remote Sensing Letters*, *3*, 68−72.

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

Nelson, M. D., Healey, S. P., Moser, W. K., & Hansen, M. H. (2009). Combining satellite imagery with forest inventory data to assess damage severity following a major blowdown event in northern Minnesota, USA. *International Journal of Remote Sensing*, *30*, 5089−5108.

Richards, J. A. (1993). *Remote sensing digital image analysis: An introduction.* Berlin: Springer-Verlag.

Smith, W. B. (2002). Forest inventory and analysis: A national inventory and monitoring program. *Environmental Pollution*, *116*, S233−S242.

Stehman, S. V. (2000). Practical implications of design-based sampling inference for thematic map accuracy assessment. *Remote Sensing of Environment*, *72*, 35−45.

Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment*, *64*, 331−344.

Stehman, S. V., Wickham, J. D., Smith, J. H., & Yang, L. (2003). Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: Statistical methodology and regional results. *Remote Sensing of Environment*, *86*, 500−516.

Townshend, J. R. G., Justice, C. O., Gurney, C., & McManus, J. (1992). The impact of misregistration on change detection. *IEEE Transactions on Geoscience and Remote Sensing*, *30*, 1054−1060.

USDA. (2007). *Forest inventory and analysis national core field guide, version 4.0.* United States Department of Agriculture Forest Service.

USDA. (2008). *The forest inventory and analysis database: Database description and users' manual version 3.0 for phase 2.* United States Department of Agriculture Forest Service.

Vogelmann, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K., & Van Driel, J. N. (2001). Completion of the 1990's National Land Cover Data set for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, *67*, 650−662.

Wofsy, S. C., & Harris, R. C. (2002). The North American Carbon Program (NACP). *Washington D.C.: NACP committee of the US interagency carbon cycle science program, global change research program* (pp. 82).