Remote Sensing of Environment 238 (2020) 111482

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

eDaRT: The Ecosystem Disturbance and Recovery Tracker system for monitoring landscape disturbances and their cumulative effects

Alexander Koltunov[a,b,*], Carlos M. Ramirez[b], Susan L. Ustin[a], Michèle Slaton[b], Erik Haunreiter[b]

[a] Center for Spatial Technologies and Remote Sensing (CSTARS), University of California, Davis, One Shields Avenue, Davis, CA, 95616, USA
[b] USDA Forest Service, Pacific Southwest Region, Remote Sensing Laboratory, 3237 Peacekeeper Way, Suite 201, McClellan, CA, 95652, USA

ARTICLE INFO

Keywords:
Remote sensing
Landsat
Forest disturbance
Disturbance detection
Change detection
Time series
Landscape monitoring
Dynamic detection model
Ecosystem disturbance and recovery tracker
eDaRT

ABSTRACT

The worldwide demand for timely and accurate information about ecosystem dynamics at Landsat spatial scale is growing and as of today still exceeds the availability of information. The diversity of required disturbance metrics and trade-offs between sensitivity, reliability, timelines of information generation, and flexibility toward potential customizations suggests that a single system is not likely to fill such demand in the near future. To address this challenge, the scientific community has been developing and improving various Landsat-based algorithms for land change monitoring. We describe the Ecosystem Disturbance and Recovery Tracker (eDaRT) version 2.9 — a highly automated prototype system in continuous development, which has been operated since 2012 by the USDA Forest Service Pacific Southwest Region to generate most current disturbance maps at Landsat scale and provide customized information services and inputs to science and land management applications in the Region.

The eDaRT processing system utilizes all three dimensions of dense Landsat image time series: spectral, temporal, and spatial. Two anomaly detection algorithms are sequentially applied, one estimating pixels' disturbance *status* metrics in every processed image and the other detecting disturbance *events*, the primary output of eDaRT. The first algorithm initially estimates change relative to a user-defined fixed baseline time period, using a stratified version of the Dynamic Detection Model (DDM; Koltunov et al., 2009) applied to Landsat bands and vegetation indexes that reflect canopy greenness, abundance, and moisture content. Using the model residuals and a probabilistic context analysis, the detected anomalies are further classified as disturbed, cloud/snow, or recovered. The resulting residuals, classification maps, and the associated disturbance confidence values provide the most rapid preliminary snapshot of the current cumulative effect of disturbance and regeneration. The second algorithm detects discrete disturbance events as regime changes in the dense time series of the residuals for each pixel. First, the residuals are compared against a recent baseline window and classified to find candidates for disturbance events. Next, candidate-events are accepted based on temporal consistency of their detection. The standard outputs from this algorithm include disturbance event timing (down to 8–16 day precision) and a detection confidence as a proxy for event magnitude.

We initially evaluated eDaRT performance with high-resolution imagery and airborne LiDAR data in a test area in California for several types of annual disturbance events at 30-m scale. These tests modeled detection probabilities as functions of canopy cover loss, which estimated detection rates of 96%, 87% and 92% respectively for fire, harvest, and tree mortality events, when canopy loss values follow a uniform prior distribution. The error of commission varied between 10- 20% for most forest types (12% on average). Following ongoing optimizations and extended validations, eDaRT will expand beyond Landsat instruments to improve ecosystem monitoring, as an independent system and potentially as a part of an ensemble method.

1. Background and motivation

Land managers and ecologists have been using remote sensing technologies for mapping land cover changes since the 1930s, starting with aerial photography, which served as the primary remote sensing image data source for land management agencies through the 1970s

* Corresponding author. Center for Spatial Technologies and Remote Sensing (CSTARS), University of California, Davis, One Shields Avenue, Davis, CA, 95616, USA.
E-mail address: akoltunov@ucdavis.edu (A. Koltunov).

https://doi.org/10.1016/j.rse.2019.111482
Received 7 April 2018; Received in revised form 23 September 2019; Accepted 16 October 2019
Available online 18 November 2019
0034-4257/ © 2019 Elsevier Inc. All rights reserved.

and 1980s. With the launch of the Landsat 1 in 1972, a new way of mapping and monitoring vegetation dynamics was begun, which established the Landsat Program's prominence in ecological remote sensing applications (Cohen & Goward, 2004; Kennedy et al., 2014). Land management agencies, such as the USDA Forest Service (USFS) in cooperation with the California Department of Forestry and Fire Protection (CAL FIRE) recognized that Landsat-based monitoring provided a cost-effective way to assess forest change, which led to the establishment of the now defunct Landcover Mapping and Monitoring Program in the 1990s and early 2000s.

Data accessibility was challenging for users in the first 36 years of the Landsat program, since individual Landsat tiles had to be ordered and substantial preprocessing/delivery costs were incurred by end users. This changed significantly in 2008 with the opening of the Landsat archive (Woodcock et al., 2008; Wulder et al., 2012). Free access to Landsat data, along with advances in computer hardware, set the stage for a proliferation of new techniques and algorithms for monitoring ecosystem disturbances from regional to global scales. These techniques vary in the types of inputs (spectral bands, indexes) and temporal coverage of images used. For example, some methods rely on annual comparisons between two images from the same season, while others require and benefit from a dense image time series up to all available images, which propagates to differences in the targeted disturbance types, detection quality and temporal precision, and availability of outputs to potential users. The methods' outputs are diverse, too, ranging from characterization of periods (sometimes called segments) in ecosystem development, to focusing on events (breakpoints) indicating change in ecosystem development, to categorizing disturbance by types or agents. Also, there are differences in how various methods define "disturbance", whether explicitly or implicitly, i.e. by algorithm design. For some methods a long-term negative trend in a vegetation index is evidence for disturbance, while others identify disturbances as deviations from trends. Each view has its own practical value. Furthermore, some algorithms are constrained or fine-tuned to monitoring forests or their regional subtypes, or are limited to specific disturbance types, e.g. burn severity (Miller & Thode, 2007; Miller and Quayle, 2015), while others are generally applicable to a broader range of ecosystems and land change events. Some of the most prominent disturbance mapping algorithms and processing systems that demonstrated their operational and/or methodological value include: Vegetation Change Tracker (Huang et al., 2010), LandTrendr (Kennedy et al., 2010; Cohen et al., 2018), Break detection For Additive Season and Trend (BFAST, Verbesselt et al., 2010, 2012), Continuous Change Detection and Classification (CCDC, Zhu & Woodcock, 2014), Image Trends from Regression Analysis (ITRA, Vogelmann et al., 2012), Exponentially Weighted Moving Average Change Detection (Brooks et al., 2014), Multi-Index Integrated Change Analysis (MIICA, Jin et al., 2013), ShapeSelectForest (Moisen et al., 2016), Vegetation Regeneration and Disturbance Estimates through Time (VeRDET, Hughes et al., 2017), a disturbance causality mapping method by Schroeder et al. (2017), and the Landscape Change Monitoring System (LCMS) ensemble technique (Healey et al., 2018). Many of these and other published methods are discussed in detail in a review paper by Zhu (2017).

As Landsat-based monitoring methodologies have evolved and diversified, so have the user community, user requirements, and applications of the end products. Different strengths, focuses, and limitations of various techniques with respect to a given application may not always be fully evident from the published materials and tests, albeit they can have profound implications for the end user. One symptom of these differences is the disagreement among algorithms' outputs when applied to the same input. High levels of disagreement of one of the most basic types of output — disturbance events with temporal accuracy of ± 1 year — was convincingly demonstrated by Cohen et al. (2017) in a sample of seven of the most well-known advanced algorithms.

As a result, it is becoming increasingly easy to argue that a static, one-size-fits-all solution to the problem of providing users with Landsat-scale land change information may not be feasible in the foreseeable future; and thus algorithm development groups should continue enhancing and expanding the portfolio of Landsat image processing tools to meet diverse, dynamic, and growing demands. Furthermore, the complexity of automated processing of Landsat image time series increases dramatically in an operational environment at a regional, let alone, larger spatial domains. To this end, computational resources, labor, and the human personnel skillsets required to operate, modify, and re-deploy a processing system in a timely manner have become significant factors determining the availability of land change information — at the right time, place, and in the right form.

To address the unmet need for timely and accurate information about frequently changing ecosystem status in California and the limited operational availability of such information, the *Center for Spatial Technologies and Remote Sensing* (CSTARS) at University of California, Davis and the USFS Pacific Southwest Region *Remote Sensing Laboratory* initially developed and, since 2012, have enhanced a regionally operating Landsat processing system prototype called the Ecosystem Disturbance and Recovery Tracker (eDaRT; Koltunov and Ramirez, 2012; Koltunov et al., 2015). In addition to cost constraints that are typical for small-scale projects in an academic environment, the main priorities that initially determined eDaRT algorithm development and implementation path included:

- Quickly develop a flexible and multi-purpose tool ready to be used by experienced remote sensing professionals for providing on-demand information and collaborative science services.
- Place initial emphases on regional-scale applications and most requested type of information: ecosystem disturbance events.
- Shorten the path toward sustained and informed operational use and allow participating users to benefit from eDaRT experimental products prior to the completion of the research phase and comprehensive vetting.
- Use algorithms and implementations that can benefit from parallel processing on a powerful application server, but do not require supercomputers to run the system at the regional scale.
- Automate the most time-consuming processing steps; however, also allow an operator to easily adjust control parameters, based on the specific end user requirements.
- Maintain incremental development using early input from operational end user partners based on their small-scale project-level tests and continue to elevate the system readiness.

The eDaRT system and algorithms described in this paper represent a snapshot of our progress toward balancing the above priorities, as of August 2018. The primary target objects mapped by eDaRT are disturbance events.

1.1. Normal evolution, disturbance events, status, and cumulative effects

Because the term "disturbance" can have different valid interpretations, we feel it is important to initially define and discuss what is meant by disturbance at a Landsat pixel in the context of this paper. First, we emphasize the differences between two related but distinct notions: disturbance *status* at pixel s and time t and a disturbance *event* at pixel s and time t.

Let us denote $\Delta V(s, t, t_0)$ the disturbance *status* (that we also term *cumulative effect*) at a pixel (s, t) relative to a reference time t_0 and define $\Delta V(s, t, t_0)$ as the difference between the actual vegetation conditions $V(s, t)$ and its conditions $V_0(s, t)$ at time t under normal development since t_0. Thus, $\Delta V(s, t, t_0) = V(s, t) - V_0(s, t)$. Useful examples of vegetation properties denoted by V include canopy cover fraction and canopy water content. $V_0(s, t)$ is often referred to as a "background" signal, and we use this term here as well. The equality $\Delta V(s, t, t_0) = 0$ implies that either vegetation has not been disturbed since t_0, or that the effect of possible past disturbances is currently unobservable, e.g.

because of complete recovery. When it is reasonable to assume that under normal development the vegetation conditions at this pixel must stay the same at all times, then one can simply set $V_0(s, t) = V(s, t_0)$. However, in general, the "static world" assumption is not valid even if V represents actual vegetation properties, let alone their proxies measured by remote sensing. Therefore, a major challenge faced by a forest monitoring system is routine estimation of the background signal.

Assume, without loss of generality, that for disturbed pixels the sign of $\Delta V(s, t, t_0)$ is negative. Then for a chosen baseline time t_0, a disturbance *event* at a pixel can be simply defined as the earliest discrete point in time at which the cumulative effect (status) $\Delta V(s, t, t_0)$ is negative. One can allow t_0 to depend on t and s, thus assigning a possibly different, moving baseline to different pixels. For example, for every pixel s and at a detection time t, one chooses $t_0(s, t)$ to be the time of the most recent observation since the previous disturbance event. In this case, a disturbance event is recursively defined as the earliest time at which the vegetation properties are abnormally different from their properties during the most recent known period of development without events.

The above definition of disturbance events, we believe, is convenient and useful because it includes a wide range of disturbance types and agents that are meaningful from ecological and management perspectives and that have been most often of interest for the user community of the system presented in this paper. These disturbances include canopy removal or damage by human intervention (harvest, mastication, thinning, selective logging, and prescribed burns) and natural causes (blow-down, landslide, flood, wildfire). Also this definition includes tree mortality (i.e. live leaves/trees are replaced with dead leaves/trees irrespective of phenology) beyond the so-called "background mortality" (i.e. trees dying at low rate in the absence of severe stressors such as drought (Das et al., 2016)). These disturbance types can be considered discrete events at the 30-m scale and temporal scales of a few weeks, and thus Landsat observations with 8–16 day repeat are useful for monitoring at these scales.

The disturbance events as we defined them may not include gradual ecological processes resulting in long-term negative trends in forest biochemistry and/or structure that could be associated with long-term climate impacts or represent a natural multiannual cycle in ecosystem development. These types of changes can be considered disturbances by some users and applications (Cohen et al., 2016); however, they characterize time series *segments* and not discrete points in time, and therefore they are not considered disturbance events in the context of this paper, except in the cases when the trend changed after the reference time t_0. The eDaRT algorithm is not designed to directly characterize processes (i.e. temporal segments as a single analysis unit), but is focused on disturbance cumulative effects and events, as defined above in this section.

1.2. Paper objective and outline

The primary objectives of this paper are to inform readers about eDaRT capabilities, provide the first detailed description of the eDaRT algorithm suite version 2.9, including its main components and products (section 2), and briefly address the system implementation in an operational environment. In section 3, we present an initial validation experiment in a test area in California, using reference datasets derived from high-resolution imagery and bitemporal airborne LiDAR data. With these data we assess eDaRT performance with respect to detection of several forest disturbance event categories that are highly meaningful to forest management. Section 4 briefly discusses the system advantages, operation, current limitations, and recent and ongoing applications for forest management and science. Our concluding remarks are offered in section 5.

2. eDaRT system and algorithms

The system represents a suite of advanced algorithms that interact in multiple ways and are used in different combinations, variants, and sequences. Therefore, the presentation in this section follows a top-down approach. Sections 2.1-2.3 provide the reader with a general overview, define terms, notations and overall objectives, and briefly overview the major blocks and stages. The subsequent sections 2.4-2.7 offer a more detailed discussion of the individual algorithms and their interaction in the system. The forward-looking and backward references are extensively provided to help the reader revisit earlier material, as needed. Additional technical details are given in Appendix 1.

2.1. eDaRT primary target objects are disturbance events

Although eDaRT outputs include disturbance status snapshots at every processed image, the primary outputs of the system are disturbance events, as they were defined in section 1.1. The disturbance event output is also the main focus of this paper. Currently, eDaRT does not differentiate between clouds, fog, snow, ice, and smoke and assigns all of these into a single class "non-clear".

2.2. Ground scenes monitored as single units

In this paper, we use the term "eDaRT scene" (or simply *scene* hereafter) as an analyst-defined *real-world region of interest monitored by* eDaRT. It is natural and convenient to define the scene boundary as a rectangle in the image coordinate system (sample, line) of the input Landsat imagery for that scene (see examples in Figs. 5 and 13). We emphasize that unlike many other publications involving Landsat, this paper never uses "scene" for "image" or in a narrow sense referring to Landsat image tiling with respect to a World Reference System (WRS) path/row notation. Each eDaRT scene is processed as a single unit and independently of other scenes, whereas *within the scene*, the results at different pixels are not always independent, as will be evident from sections 2.4-2.7.

2.3. System operation and output product overview

The eDaRT system version 2.9 analyzes Tier 1 terrain-corrected Landsat Collection 1 imagery (USGS, 2018) from three Landsat satellites (5, 7, and 8), using the same algorithm, regardless of the source satellite. Bands that are unique to Landsat 8 are not currently utilized. For each processed image, eDaRT uses the following bands, indexes, and transformations, generically termed "eDaRT bands", or simply *bands*, hereafter:

- 'B1','B2', 'B3', 'B4', 'B5', 'B7' — at-sensor reflectances derived from the corresponding Landsat 5/7 bands or from their respective spectrally similar Landsat 8 bands (USGS, 2019),
- 'TIR' — brightness temperature in a Landsat 11–12 μm band (in degrees K),
- 'NBR', 'NDII', 'NDVI' — the well-known vegetation indexes (VI): the Normalized Burn Ratio, the Normalized Difference Infrared Index, and the Normalized Difference Vegetation Index.
- 'TC1', 'TC2', 'TC3' — first three Tasseled Cap (TC) transformation bands (Huang et al., 2002), and
- 'MAGN' — magnitude (Euclidean norm) of the Landsat six-dimensional at-sensor reflectance vector.

For a given eDaRT scene, the input image sequence (a temporal stack) is automatically compiled by clipping all available images in the Landsat archive to the scene extent, usually subject to two conditions: (i) all observations at the same pixel represent the same Landsat swath (i.e. WRS path), which provides temporal invariance of viewing angle and spatial invariance of expected frequency of observations; and (ii)

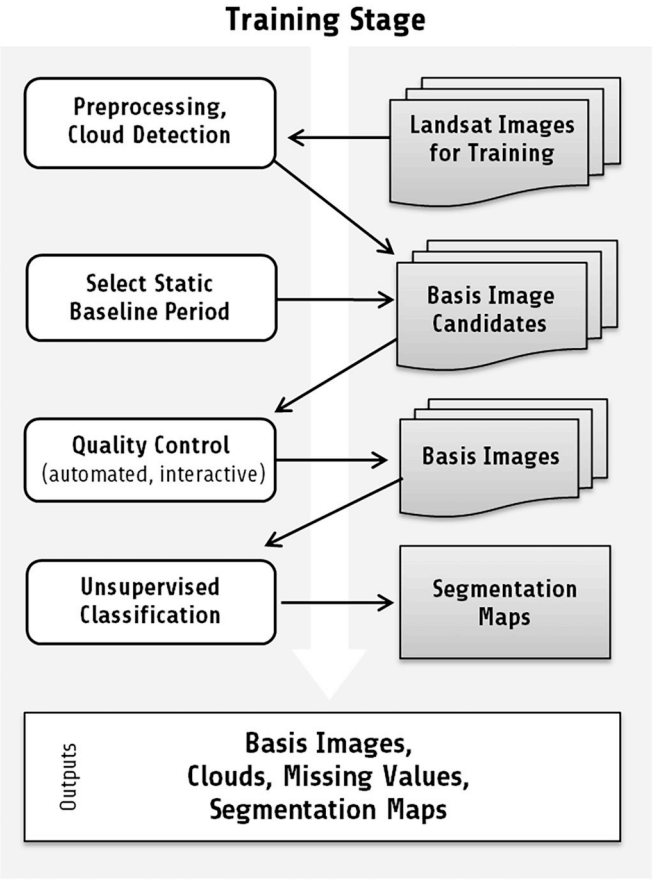

Fig. 1. Schema of the eDaRT Training stage operation (see text in sect. 2.3 for details).

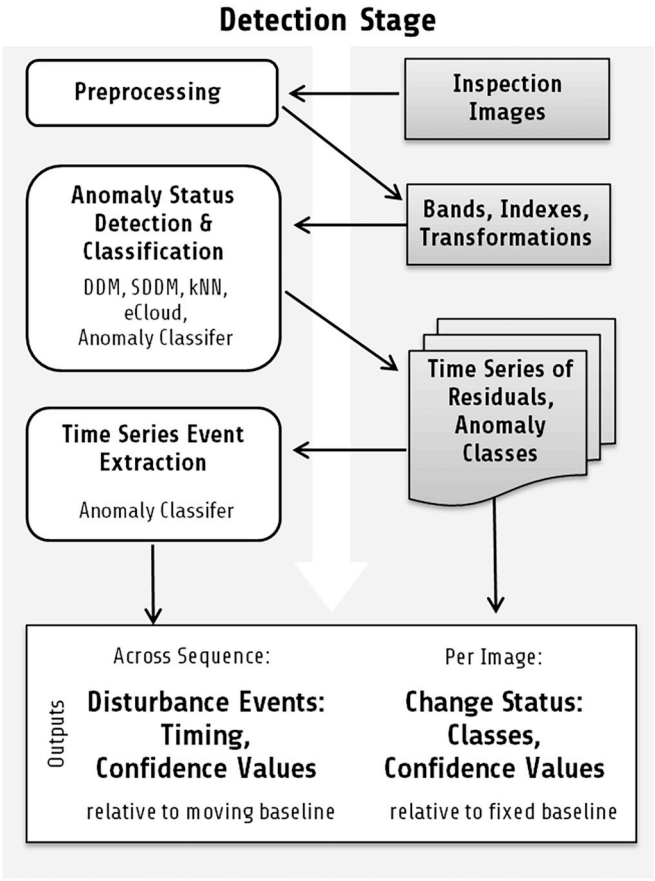

Fig. 2. Schema of the eDaRT Detection stage operation (see text in sect. 2.3 for details).

the clear-view pixels exceed an analyst-chosen fraction of non-missing pixels in the scene. The system operates in two main stages, *Training* stage and *Detection* stage, and initially requires an analyst to split the image sequence into two parts:

- the *inspection* images, which this paper calls images to be automatically inspected during the Detection stage to detect disturbances, and
- the *training* images, i.e. a pool of historic images typically spanning 1–4 years before the Detection stage.

2.3.1. Training stage

The Training stage (Fig. 1) first masks non-clear pixels in the training images, using the Retrospective Cloud Detection algorithm of the *eCloud* module (sect. 2.5 and Appendix 1) and then assimilates and prepares several static datasets for the Detection stage, including:

- **A subset of training images to serve as *basis images*.** The term "basis images" (Koltunov et al., 2009, 2016; Koltunov & Ustin, 2007) stands for images combined by the anomaly detection methods described in section 2.4 to model background signal in pixels of the inspection images and estimate anomalous change relative to the time period and vegetation conditions spanned by the basis images.
- **Segmentation maps** to stratify anomaly detection by vegetation type, usually obtained using Gaussian mixture clustering in the bi-temporal Landsat spectral space formed by two seasonally most distinct basis images. In this way, significant clusters tend to represent vegetation with distinct combinations of canopy cover, species composition, and phenological types.
- **Optional ancillary masks** and maps of land cover types, such as a forest mask or a National Land Cover Dataset (NLCD) land cover map.

2.3.2. Detection stage

Using the above datasets, the Detection stage illustrated by Fig. 2 proceeds in two steps. First, each inspection image of the scene is processed *independently* by the Anomaly Detection and Classification (ADC) module. The ADC estimates the background signal in several eDaRT bands using the Dynamic Detection Model and Segmented Dynamic Detection Model algorithms (sect. 2.4) and analyzes the models' residuals to classify pixels as: no-anomaly, non-clear, missing-value, disturbed, or recovered. These residuals and the classification map can be immediately delivered to the user as a near-real-time snapshot of the current disturbance status (cumulative effect) relative to the vegetation conditions defined by the basis images. Although these outputs can also be used to detect disturbance *events*, notably only one disturbance event in the entire sequence of inspection images can be detected for a given pixel (as the status transition from "never been disturbed" to "disturbed").

In the second step, the Time Series Event Extraction module (TSEE, pronounced "TC") analyzes temporal trajectories of the residuals and classification maps for individual pixels to extract disturbance *events* comprising three elements: location, timing, and a confidence score. In the following sections we discuss the individual algorithms and modules of the system and their use during Training and Detection stages in more detail.

2.4. Change status monitoring algorithms

2.4.1. Anomaly detection with Dynamic Detection Model

The Dynamic Detection Model (DDM, Koltunov et al., 2009, 2016; Koltunov & Ustin, 2007) in eDaRT detects anomalous changes (e.g. disturbances, clouds, other non-clear-view pixels) and suppresses image changes due to normal evolution of undisturbed landscapes and sensor response characteristics. The DDM model can be viewed as a special type of constrained multiple regression methods. To detect anomalies in an image $W(t)$ at time t and pixel location s ($s = 1, \ldots, N$), a DDM estimates the unknown background ("no anomaly") value $W(s, t)$ as a location-invariant function of values that were observed at pixel s previously, at P past time moments $t_1, \ldots, t_P$, termed *basis times* (Koltunov et al., 2009). The eDaRT system uses a linear function:

$$W(s,\ t) = \beta_0(t) + \beta_1(t)W(s,\ t_1) + \ldots + \beta_P(t)W(s,\ t_P) \quad (1)$$

Because the unknown parameter vector $\beta(t) \overset{def}{=} \{\beta_k(t)\}_{k=0}^{P}$ is the same for all N pixel locations, eq. (1) effectively states that an inspection image $W(t)$, as an N-dimensional vector, is a linear combination of past images $W(t_1), \ldots, W(t_P)$. These images are called *basis images*, because as vectors, they form a basis for the vector space of all possible "no-anomaly" images of the monitored scene.

The coefficients $\beta(t)$ are estimated in the least-squares sense, using a random set of training pixel locations that are *a priori* not known to have anomalies or have any missing values at any time $t, t_1, \ldots, t_P$. With these coefficients, a forward application of the DDM to a given eDaRT band (defined in sect. 2.3) for each pixel predicts the background value and produces a standardized residual (also called a z-score):

$$Z = (W_{\text{observed}} - W_{\text{predicted}}) / \sigma, \quad (2)$$

where σ is an estimate of the model r.m.s.e. The z-scores for different bands are used to detect and classify changes in the scene relative to the time period and vegetation conditions spanned by basis images.

Although DDM detects temporal changes, it is a not a classic time-series approach wherein each pixel is modeled independently, as in most, if not all, other known Landsat based monitoring algorithms. This is because indirectly, through the least-squares estimate of $\beta(t)$, the predicted background value at the pixel depends on the training pixel values in the basis image and in the inspection image. In this way, training pixels empirically and indirectly "teach" the model what the current image conditions are and how they are supposed to affect other pixels under the "no-anomaly" hypothesis (Koltunov et al., 2009). This also implies that DDM targets anomalies defined as changes (or lack thereof) that are uncommon among the training pixels, regardless of whether the current conditions are typical of this pixel's historic trajectory. Therefore a disturbance does not require a decrease in a VI value in order to get detected. In fact, landscapes that show a net increase in a VI can have a negative residual of the VI prediction by the DDM if the growth is weaker than the model predicted; thus the pixel can potentially be classified as disturbance. Conversely, apparent reductions in vegetation indices are not flagged as disturbance if DDM finds that such changes are consistent with the pixel values currently observed at likely undisturbed forests in the scene.

Another notable characteristic feature of DDM is that time (as a physical variable) is not an independent predictor. It contributes only indirectly, through pixel values in the corresponding images. Hence, an image on the left-hand side of (1) can be a past image modeled using chronologically subsequent basis images on the right-hand side of (1), or vice versa. This property is used by eDaRT for retrospective detection of non-clear pixels (sect. 2.5, and Appendix 1) during Training stage.

2.4.2. Basis image selection and missing-data implementation for DDM

According to (1), basis images must empirically represent seasonality-driven variability of normally developing landscapes across the entire scene. If during the Training stage all Landsat images from a multiyear training period (sect. 2.3) are accepted as basis images and are also free of non-clear observations, then the seasonal variability in the scene is best represented. However, this can be a redundant representation (Koltunov et al., 2009), i.e. some images may be unnecessary. Thus, to improve computational efficiency, a smaller set of the basis images is selected during the Training stage by aiming to accumulate only one valid observation per calendar month for as many land pixels as reasonable, using as few images as possible. Toward this goal, for each month to be processed during the Detection stage, the training images with a low fraction of non-clear pixels (typically below 25%) are included as basis images, preferring images with fewer non-clear and missing values and from a more recent year. As a recommended optional step, the training images and their non-clear pixel masks are also interactively reviewed, in particular, to make sure that the undetected non-clear pixels and other possible artifacts are minimal on land. As a result, an analyst may choose to add or reject a basis image candidate, or leave some months unrepresented. A typical eDaRT scene in California uses 10–20 basis images selected in this way.

The forward application of the model (1) to a given pixel assumes that the object in the pixel's field of view is the same object in every basis image (Koltunov et al., 2009). Therefore, a single set of basis images can model the entire scene only if all basis images are completely free of non-clear pixels and other anomalies. This is not a typical scenario, unless the number of basis images in the scene-wide model is small. Hence, prior to model estimation at the Detection stage eDaRT analyzes missing- and non-clear pixel masks for various subsets of the basis times $t_1, \ldots, t_P$ and determines which pixels can be modeled for each subset, as follows. First, the basis time subsets are grouped by their cardinality (i.e. number of elements), with at most three subsets per group. The groups are then searched in a descending order of cardinality until the current list of subsets can model all but an acceptably small number of land pixels, or until there is only one basis time in the current subset. For further discussion and illustration of the missing-data implementation of the DDM the reader is referred to a paper by Koltunov et al., (2016), where DDM estimates background brightness temperatures for early detection of wildfires in geostationary satellite images.

2.4.3. Segmented Dynamic Detection Model (SDDM)

To increase the DDM model complexity and accuracy, eDaRT estimates the model independently for each land cover class in a segmentation map created at the Training stage. We call this method the Segmented Dynamic Detection Model (SDDM). In this method, apparent vegetation changes that are common within the land cover class, but not necessarily over the entire scene, are considered normal for that class. Because of non-clear observations, disturbances, and missing values a class can have too few (less than $20 \cdot P$) valid training pixels in the inspection image, and thus SDDM estimation is deemed unreliable for this class. In this case, an alternative segmentation map is used that has more general classes and thus is more likely to have enough valid training pixels. In rare situations, when none of the segmentation maps allow model estimation, the global DDM is applied to these pixels. In eDaRT 2.9, up to two segmentation maps can be used. In the test experiments discussed in section 3, the second, "backup" source of stratification was an NLCD-2006 land cover map.

2.5. Detection of "non-clear" pixels by eCloud module

Depending on specific circumstances and the processing step, eDaRT applies the following three detection algorithms ordered by the increasing complexity and computational demand:

1) Single-date Cloud Detection (SCD),
2) Multitemporal Cloud Mapping (MCM), and
3) Retrospective Cloud Detection (RCD).

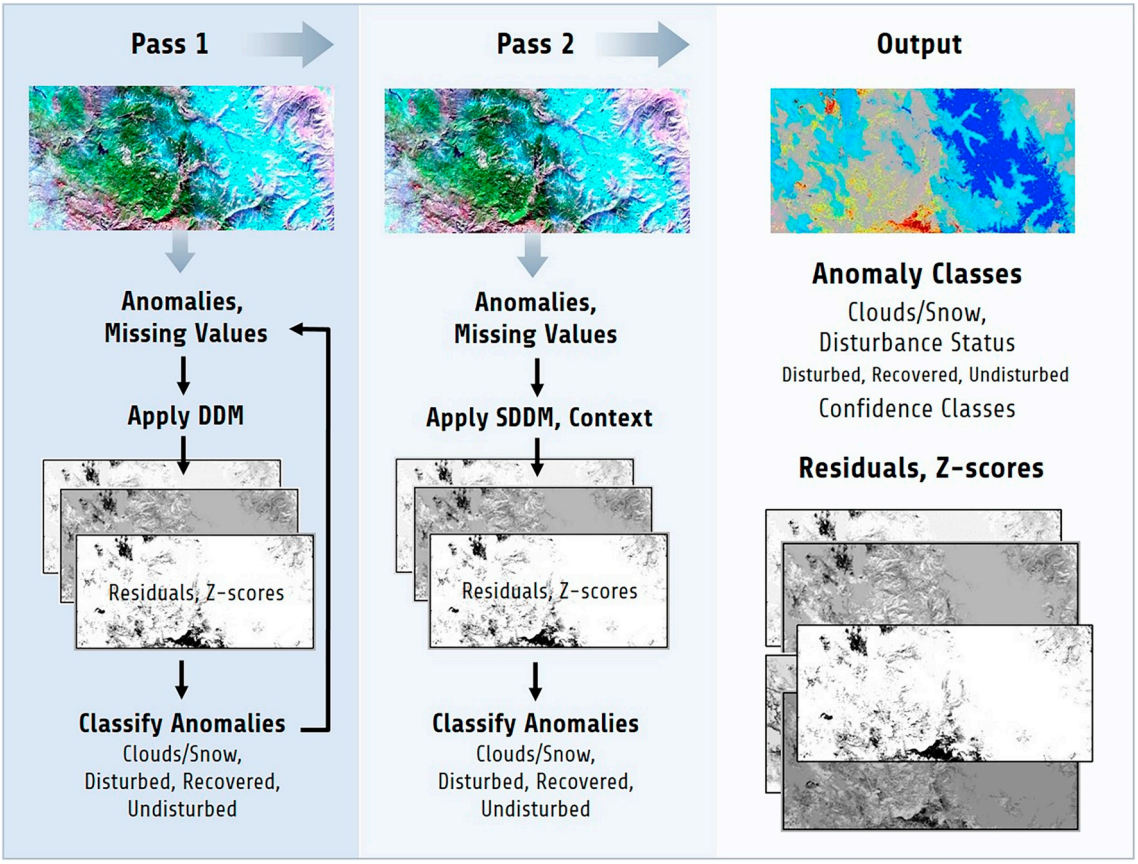

Fig. 3. Schema of the eDaRT Anomaly Detection and Classification (ADC) module (see text in sect. 2.6 for details).

These algorithms have different inputs, constraints, and quality requirements. The SCD algorithm is used during Training and Detection stages for fast and necessarily conservative prescreening that does not require data from other dates and thus can be applied before a multitemporal technique. In contrast, the multitemporal algorithms MCM and RCD are based on temporal change detection by the Dynamic Detection Model, which in turn initially relies on the SCD-detected "non-clear" pixels to mask out a potentially large number of known outliers. Further details are provided in Appendix 1.

2.6. eDaRT Anomaly Detection and Classification (ADC) module

The ADC module is activated at the Detection stage and consists of two passes, as shown in Fig. 3. The primary goal of Pass 1 is to identify pixels whose surface properties have significantly changed relative to the baseline status defined by the basis images, e.g. "non-clear" pixels, disturbances, vegetation regeneration, and other anomalies. These pixels will not be used during Pass 2 for determining the coefficients $\beta(t)$ in the background models (1) employed by DDM and SDDM. As a trade-off between robustness and processing time, Pass 1 proceeds in three iterations or until the class membership does not change for more than 95% of pixels, whichever comes first. In each of these iterations, the following steps are performed:

- Detect "non-clear" pixels using SCD algorithm and ignore those pixels in the following step.
- Apply DDM to predict background values in the following eDaRT bands: 'B1', 'B2', 'TIR', 'NBR', 'NDII', 'NDVI', 'TC1', 'TC2', and 'MAGN', resulting in the corresponding z-scores (standardized residuals).
- Accept SCD detections and detect additional "non-clear" pixels with the MCM algorithm of eCloud module (sect. 2.5, Appendix 1).
- Classify remaining pixels as background, disturbance, or regeneration, using the rules of Anomaly Classifier described below in section 2.6.1.

Pass 2 conducts the following steps:

- Apply SDDM to the same bands as DDM, with the regression using only pixels confidently classified as "no anomaly" by Pass 1.
- Classify pixels as background, disturbance, or regeneration, using the rules of the Anomaly Classifier.

Images with more than 90% of non-clear view pixels over land are excluded from further analysis.

2.6.1. Anomaly Classifier

The Anomaly Classifier calculates spectral (S_{sp}), contextual (Z_{nn}), and the total (S_{tot}) disturbance confidence scores and the regeneration confidence score (S_{reg}) for pixels that are not already assigned to classes "non-clear" or "missing value". These scores are thresholded for class assignment. The inputs for Anomaly Classifier include standardized residuals of SDDM-based prediction of pixel values for bands 'B1', 'B2', 'NBR', 'NDII', 'NDVI', and 'MAGN'.

2.6.1.1. Spectral change analysis. The first step is to correct the NDVI z-score for potential negative bias at darkened pixels, primarily due to cloud shadows:

$\hat{Z}_{NDVI} = Z_{NDVI} - \delta Z_{SH}$, where $Z_{SH} = Z_{NDVI} - Z_{NBR} - Z_{NDII}$ and δ is a binary indicator variable given by:

$$\delta = \begin{cases} 1, & \text{if } (Z_{B1} < 0 \text{ or } Z_{B2} < 0 \text{ or } Z_{MAGN} < 0) \text{ and } Z_{NDVI} < Z_{NBR} + Z_{NDII} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Next, the disturbance spectral score S_{sp} is calculated. For pixels with either one of the individual z-scores Z_{NBR}, Z_{NDII}, $Z_{NDII} > 0$, the value of S_{sp} is set to 0. For the remaining pixels these z-scores are combined using a Z-transform, also known as Stouffer's Z-statistic (Stouffer et al., 1949; Mosteller & Bush, 1954):

$$S_{sp} = -(Z_{NBR} + Z_{NDII} + \hat{Z}_{NDVI})/\sqrt{3}, \quad if \quad Z_{NBR}, Z_{NDII}, Z_{NDII} \leq 0 \tag{4}$$

Positive values of S_{sp} generally indicate candidates for disturbances.

2.6.1.2. Spatial context analysis. Disturbance detection by SDDM in eq. (4) can be viewed as a statistical significance test. As with any statistical test, the effect size (in our case, the actual magnitude of disturbance on the ground) must exceed a certain threshold, in order for the test statistic S_{sp} to fall in the region of rejection of the null-hypothesis. Therefore, actual smaller-magnitude disturbances can go undetected if we rely solely on the spectral change information captured by SDDM. To find additional information, we assume that for many disturbance types, actually disturbed 30-m pixels in the image tend to exhibit spatial association. It can have different forms and occur at various spatial scales and distances, e.g. compact patches of group selection cuts or clusters of non-connected pixels of tree mortality or silvicultural selective thinning treatments. Non-disturbance type anomalies, too, can have spatial associations with each other and perhaps also with disturbed pixels. However, it is reasonable to assume that within-group spatial relationships for disturbed pixels can be probabilistically different from spatial relationships between disturbed and undisturbed. Importantly, SDDM-based detection provides a large number of confident samples from both groups, thus allowing one to model spatial relationships and their inter-group differences using a set of chosen metrics of spatial association. With these models, we can evaluate whether or not a test pixel's proximity to already detected disturbances provides additional, context-based evidence for disturbance.

Thus, to increase the sensitivity of disturbance detection eDaRT analyzes spatial context for pixel-candidates that do not pass a minimal detectability threshold, but bear weak evidence of disturbance, as defined by the condition: $0.5 < S_{sp} < 2.0$. First we define two groups, "disturbed" and "undisturbed", comprising unclassified pixels for which $S_{sp} > 2.0$ and $|S_{sp}| < 1.0$, respectively. For these pixels we calculate a set of distance metrics $\{d_k\}_k$, for $k \geq 1$. For pixel s, distance $d_k(s)$ is defined as the distance to the k-th nearest neighbor (kNN) that is a member of group "disturbed". Currently, eDaRT uses two kNN distances d_2 and d_{10}, thus emphasizing two distinct situations: 1) high proximity to a cluster (small or large) of confidently detected disturbances and 2) the presence of a relatively large (10 or more pixels) cluster (dispersed or compact) at moderate distances. By definition, the group-conditional cumulative distribution function (c.d.f.) for d_k represents the probability for a group member to have the k-th nearest disturbance closer than d_k. These c.d.f.'s are estimated for each group and each distance variable by the corresponding empirical c.d.f.'s in a straightforward manner.

Using the group-conditional c.d.f's for d_2 and d_{10}, for a given pixel-candidate we apply two likelihood-ratio tests corresponding to d_2 and d_{10}, where the c.d.f. for group "undisturbed" is considered the null-distribution. The resulting p-values of the two tests, p_2 and p_{10} are combined using the "minimum" principle: $p_{NN} = \min(p_2, p_{10})$, which has been recommended for situations in which the evidence for disturbance could be fully concentrated in only one of the tests (Loughin, 2004). Finally, the combined p-value is converted into a composite kNN z-score Z_{nn} in a straightforward way, by inverting the c.d.f. of the standard normal distribution. By construction, negative values of Z_{nn} provide evidence for disturbance.

Importantly, our context analysis does not assume any prior knowledge of the disturbance types (e.g. fire, harvest, stress, mortality, or all their subpixel mixtures) for the training pixels. Likewise, we do not need to know what types of phenomena the undisturbed pixels in the training sample represent.

2.6.1.3. Classification based on spectral change and spatial context. The total disturbance confidence score is determined by the following rule in eq. (16). In this rule, negative Z_{nn} values are non-linearly combined with S_{sp} to ensure that Z_{nn} -based evidence does not dominate the spectral score and that Z_{nn} is not used as a sole source of evidence against the "no-anomaly" hypothesis:

$$S_{tot} = \begin{cases} S_{sp}, & \text{if } Z_{nn} \geq 0 \\ S_{sp} + (-S_{sp}Z_{nn})^2, & \text{if } Z_{nn} < 0 \text{ and } -S_{sp}Z_{nn} \leq 1 \\ S_{sp} + (-S_{sp}Z_{nn})^{1/2} & \text{if } Z_{nn} < 0 \text{ and } -S_{sp}Z_{nn} < 1 \end{cases} \tag{5}$$

Note that $S_{tot} \geq S_{sp}$ for all pixels. In other words, pixels lacking spatial association with disturbances detected without using context retain their DDM and SDDM-based evidence given by S_{sp} and thus can be classified as disturbances when S_{sp} is sufficiently high.

The regeneration score in eDaRT 2.9 uses a simple, non-optimized condition: $S_{reg} = \max(Z_{NBR}, Z_{NDII}, Z_{NDVI})$. Pixels for which $S_{tot} > \gamma$ are classified as disturbed and pixels with $S_{reg} > \gamma$ are classified as re-covered/regenerated, for $\gamma > 2.0$.

The Anomaly Classifier algorithm applies to the above standardized residuals from anomaly tests without making assumption regarding the tests themselves. Therefore, it can be applied to z-scores obtained by another anomaly detection method, or to anomalies with respect to a different baseline conditions, or both. This universal property of the Anomaly Classifier is utilized to extract disturbance events by the TSEE module described below.

2.7. Time Series Event Extraction (TSEE)

To generate disturbance events, TSEE module analyzes outputs from ADC, including: 1) temporal trajectories of SDDM z-scores for the bands used by the Anomaly Classifier and 2) the associated anomaly classification maps for each image. In contrast to previous stages of processing in which the outputs for a given pixel are affected by data at other pixels, the TSEE algorithm applies to individual pixels independently.

Fig. 4 illustrates the TSEE application to a pixel in which a tree mortality event with estimated 10% tree canopy cover loss is visible in high-resolution image pair between May 30, 2014 and June 30, 2016. Subplots a)-c) in this figure show scaled time series of NDII, NDVI, and NBR bands for June–October images each year during 2006–2017 for that pixel. The standardized SDDM residual trajectories for these bands are shown in Fig. 4d)–g) as filled circles. The vertical red lines in all subplots indicate the first detection of the mortality event by TSEE.

When applied to pixel (s, t), the TSEE algorithm can be viewed as a combination of backward and forward temporal filtering. The backward filter is a computationally affordable moving-median filter. The filter's temporal window spans at least m_A but no more than M_A most recent available valid (i.e. clear-view and non-missing) observations following the latest disturbance event at that pixel. These valid observations indirectly define the baseline period relative to time t. Note that only valid observations are shown in Fig. 4d)–g). To estimate the standard error σ inside the window we use a conventional robust estimator $\sigma = MAD/0.6745$, where MAD is the median absolute deviation. The difference between the z-score at pixel (s, t) and the window median is divided by σ to yield the confidence that (s, t) is significantly different from the baseline. For the sake of convenience and consistency with the notations of the previous sections, this confidence value will be also called a z-score and denoted by 'ZB'. Thus, the outputs of backward filtering for (s, t) include ZB(s, t) for bands 'B1', 'B2', 'NBR', 'NDII', 'NDVI', and 'MAGN'. In Fig. 4d)–f) the ZB values are marked with a diamond symbol. These z-scores are further analyzed by the Anomaly Classifier as described in section 2.6.1, resulting in the disturbance

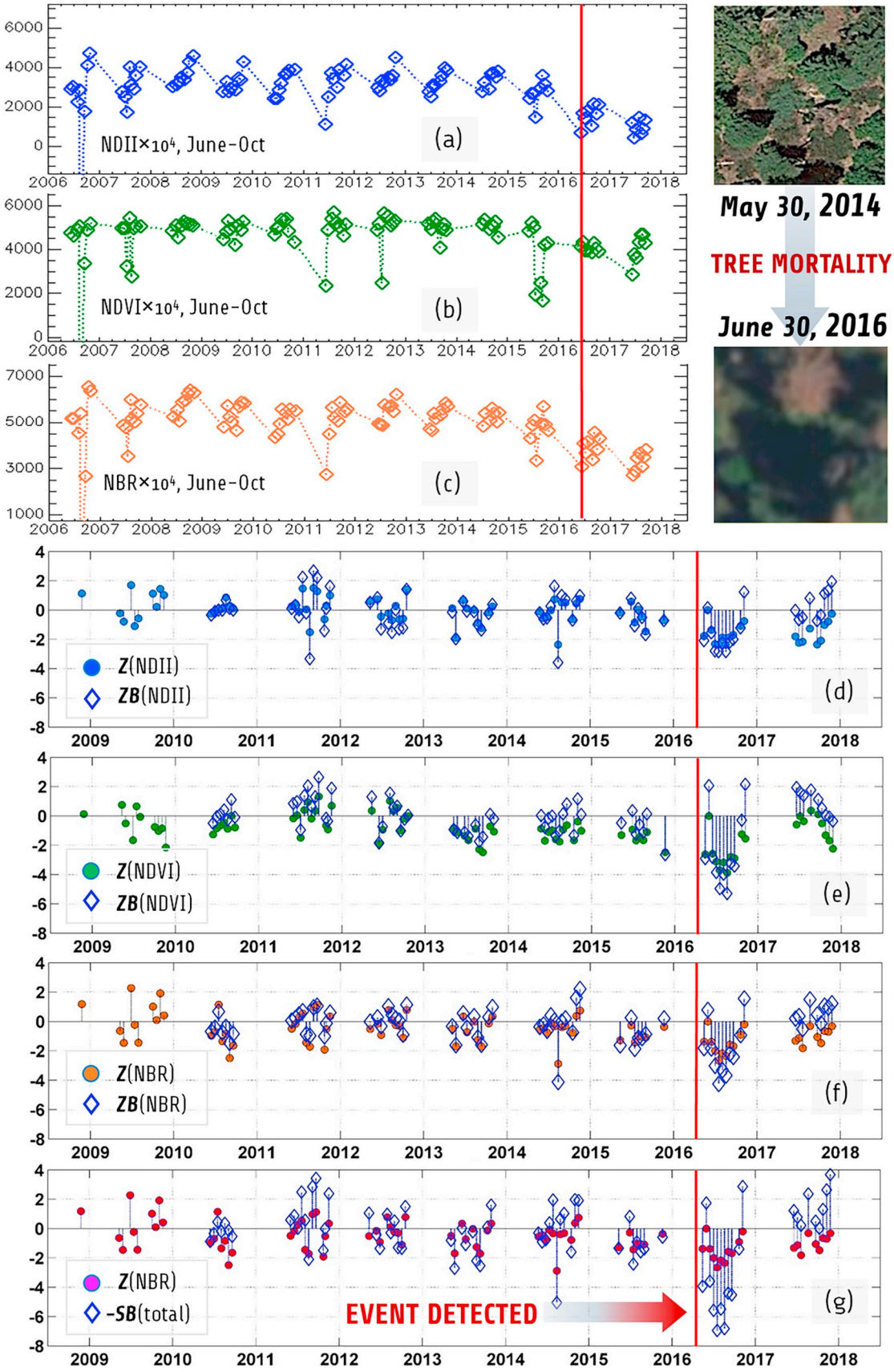

Fig. 4. Illustration of Time Series Event Extraction (TSEE) module detecting a mortality event (see sect. 2.7 for details).

event score $SB_{tot}(s, t)$. For viewing convenience, in Fig. 4g) we also mark SB_{tot} with the diamond symbol and show SB_{tot} with the opposite sign (i.e. $-SB_{tot}(s, t)$), and therefore negative values in this subplot and all other subplots point to a possible disturbance. If for (s, t) the disturbance event score SB_{tot} is significantly positive ($SB_{tot} > \gamma$, for $\gamma > 2.0$), then (s, t) is flagged as a disturbance event candidate. Otherwise, backward filtering is repeated for $(s, t+1)$, with the temporal window of size M_A sliding forward by one image.

Once a candidate-event is found at (s, t), e.g. the first observation of 2016 in Fig. 4 (t = 2016/05/12), the forward filter checks consistency of the ZB(s, t) signal by fixing the baseline window and repeating the backward filtering process for up to $(K-1)$ additional valid observations $(s, t+1)$, …, $(s, t+K-1)$. The candidate (s, t) is accepted as a confirmed event if the candidate-event is re-detected at least $N_{redetect}$ times out of K observations. The overall event confidence S_{event} is defined as the median of the $SB_{tot}(s, t+k)$ for values of k corresponding to re-detections.

In the validation experiments presented in this paper, the TSEE control parameters were set as follows: $m_A = 10$, $M_A = 20$, $N_{redetect} = 7$, and $K = 12$, which is the default option in operational applications. When satellite observations at a pixel are available at a 16-day step, parameters m_A, M_A, and K translate into 160, 320, and 192 days, respectively. In Fig. 4g) the candidate-event was re-detected 8 times in the 11 available valid observations of 2016 and thus the event was accepted and assigned a timestamp t = '2016/05/12'.

Generally, the choice for the parameter pair, $N_{redetect}$ and K, indirectly reflects *a-priori* assumptions about the relationship between the expected duration of the cumulative effect for disturbance events targeted by TSEE, and temporal persistence of the apparent effect of potential false positive events (e.g. melting snow, undetected smoke, leaf loss due to untypically early tree senescence or understory phenology). Because these parameters are formulated *solely* in terms of the number of available valid, clear-view observations, more frequent imaging generally leads to a higher detection probability and temporal accuracy, but also additional false positives. In general, eDaRT does not require temporal regularity or a specific frequency of the imagery and can use any or all of the three Landsat satellites and all swaths overlapping with the scene. However, during years when two Landsat satellites are operational, the average frequency of imaging doubles for some pixels, whereas it only slightly increases in Landsat-7 missed scan areas. As a result, the balance between detection sensitivity and reliability can become systematically inconsistent across different years and pixels. Allowing the algorithm to use both swaths of the same satellite in their overlap zone would have a similar effect, while also introducing additional sources of noise. Hence, to maximize the space-time consistency of performance the standard operational application of eDaRT 2.9, including the experiment presented in section 3, uses a single Landsat swath per scene and preferentially uses Landsat 5 and 8 (when available) in lieu of Landsat 7 during the Detection stage, unless most recent disturbance information is requested. A more flexible and more computationally demanding approach, e.g. to use time in addition to $N_{redetect}$ and K, is left for a future version of eDaRT.

A practical limitation of all time-series change detection algorithms is that change detection is impossible until a sufficient number of valid baseline observations are accumulated, which takes time. This limitation also applies to the TSEE filtering. Indeed, following a detected event at a pixel, eDaRT requires at least m_A valid post-event observations as the new baseline for detecting the next disturbance event. Therefore eDaRT can detect no more than one event per m_A images at the same pixel. For example, when tree mortality at a pixel is followed by a wildfire within the next 160 days, we may not be able to decouple the two events, which could bias analyses attempting to link pre-fire tree mortality and fire effects. The actual length of the post-detection blackout period depends on the number of valid observations in the images following the event detection.

The same problem arises when a true disturbance event is among the first m_A observations in the inspection subsequence (note circles in Fig. 4d)–g) that during 2009 do not have the corresponding diamond markers). These early disturbances cannot be detected by the backward filter. However, as mentioned in sect. 2.3.2, anomaly classification maps from ADC module, on their own, offer an independent means of identifying *the first* disturbance event at a pixel. Thus eDaRT can detect disturbance events, starting from the very first inspection image. In practice, it is often appropriate to choose basis images from an earlier period during the Training stage, thus enabling TSEE application to earlier images.

3. Performance evaluation experiments

The performance assessment experiments presented in this paper were focused exclusively on discrete disturbance *events* (not the status relative to a fixed baseline defined by the basis images) and included comparison of eDaRT events against reference disturbance and non-disturbance events compiled using high-resolution imagery, bitemporal airborne LiDAR data, and in a limited way, Landsat data itself. These experiments had two primary objectives:

1. Assess disturbance detection performance in a test scene that is representative of conditions in which eDaRT has been most used in the last few years, in terms of probability of disturbance detection and probability of false positives.
   - Estimate detection probabilities for major ecologically meaningful disturbance categories: harvest, fire, and tree mortality.
2. Quantify detection sensitivity of eDaRT algorithm and estimate detection probability as function of vegetation canopy cover loss, which is the major factor influencing detection.

3.1. Test scene

Our test scene (Fig. 5) or simply "Scene" hereafter, represents a 1754x4468 pixel subset of Landsat WRS tile with path/row = 42/34. The Scene is located in the central Sierra Nevada, California and spans 6317 km^2, including large portions of the Sierra National Forest and Sequoia-Kings Canyon National Park. The Scene contains vegetation types, environmental settings, and disturbance regimes common in California, especially in the Sierra Nevada, where operational applications of eDaRT products are most common to date. Indeed, conifer forest is the predominant vegetation type in the Scene (44.9% of pixels), with shrubland (19.0%), hardwood (5.4%), and grassland (5.2%) also common (LANDFIRE, 2013). Forests range from ca. 1000–3000 m in elevation, including Ponderosa pine (*Pinus ponderosa*) and mixed conifer forests at lower to mid-elevations, and red fir (*Abies magnifica*), lodgepole pine (*Pinus contorta*), and subalpine forests at higher elevations. Hardwood and shrubland habitats include both evergreen and deciduous components, as well as oak stands (*Quercus* spp.), riparian forests, and montane chaparral. A small proportion of the Scene is occupied by agricultural and urban settings, with water bodies and barren rock outcrops relatively common at mid to high elevations, where snow cover occurs in winter months.

Thirty-nine wildfires were recorded in the Scene during 2009–2016 (Fig. 5a), including three fires exceeding 45 km^2 (2013 Aspen, 2014 French, and 2015 Rough fires (CAL FIRE, 2018);). Large areas are visibly affected by drought/insect-induced tree mortality. Vegetation management practices, including timber harvest, thinning and prescribed fire occur regularly on slopes < 35° in forested areas, with ca. 1890 ha affected annually (USFS, 2018).

3.2. Reference data development method

Because reference data for Landsat-based disturbance mapping methods are limited both spatially and temporally, algorithm validation parameters and datasets used represent a trade-off between two often

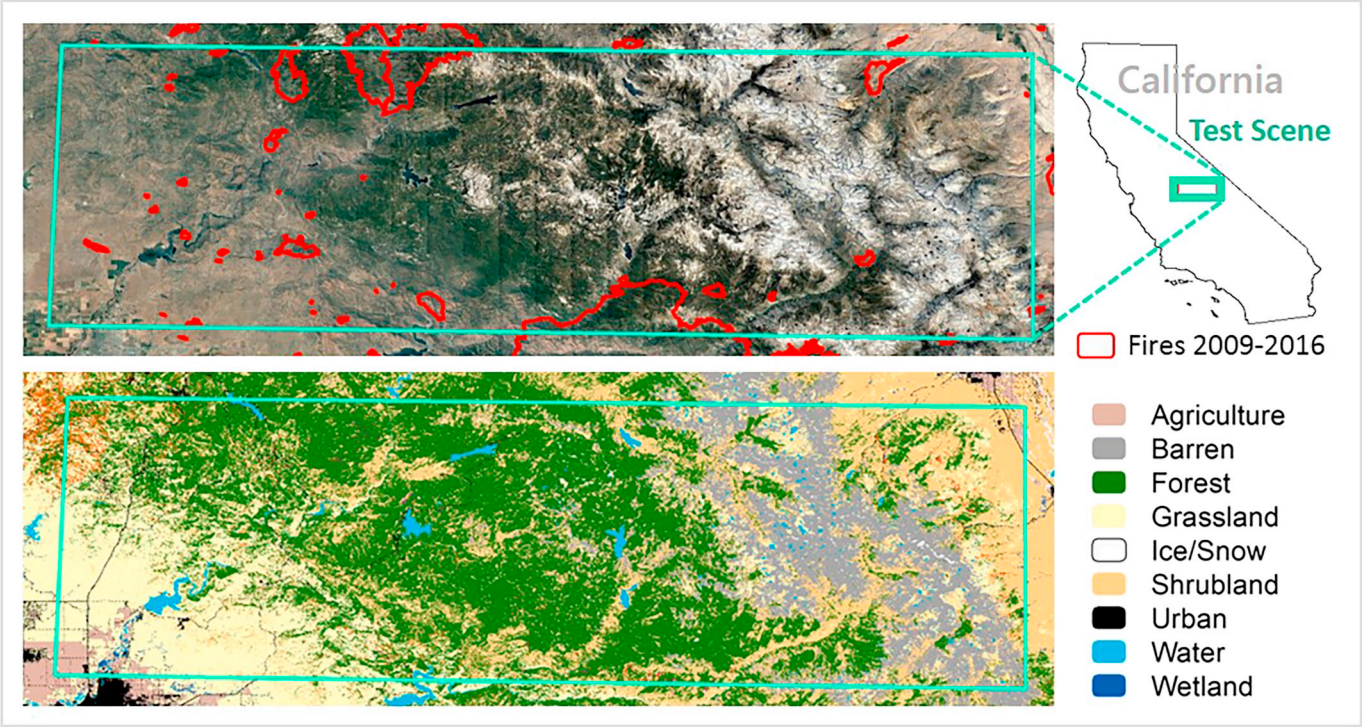

Fig. 5. Test Scene in California, U.S.A. (a) A natural color image with wildfire perimeters for 2010–2016. Three largest incidents, French, Aspen, and Rough fire, are labeled (see sect. 3.1 for further details). (b) Major land cover types from the NLCD-2006 map. . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

conflicting goals: a) greater representation of surface phenomena and operation conditions, and b) minimizing labor and potential biases and uncertainties in developing the reference data, given practical constraints on availability of data and personnel time. Typically, the reference datasets are developed using some combination of human interpretation of multidate higher-resolution imagery, the Landsat data itself, and ancillary sources. This approach was adopted by Kennedy et al. (2010), Zimmerman et al. (2013), Zhu and Woodcock, 2014, Cohen et al. (2017), Healey et al. (2018), and many others; and with modifications, also for the tests presented herein. Let us first briefly discuss the rationale behind our choice of a validation approach.

Every single-source or multi-source analysis method has its own limitations and advantages. Indeed, the multidate higher-resolution imagery can be blurry, oblique-angled, represent different phenological stages, misaligned, infrequent (and sometimes all of the above, while any single one can confuse even the most experienced image analysts), or simply be unavailable for the time periods of interest. Even when a perfect image pair is available, some phenomena, e.g. canopy water loss, are unobservable (even at high magnitudes) in high-resolution images, unless there are also visible changes in canopy cover, structure, and/or pigment. However, as an important advantage of high resolution, an experienced analyst can assign major disturbance types (harvest, mortality, burn) and estimate canopy loss regardless of Landsat image availability and of the apparent shapes of VI trajectories, which directly determine outputs of the algorithm being evaluated. In contrast, the availability and quality of high-resolution imagery during the test period do not affect the algorithm outputs, and thus with due care uncertain cases can be safely excluded from a test sample. To extrapolate performance estimates to time periods lacking high-quality imagery or into the future, one must rely on the assumption that algorithm performance does not directly depend on the chosen multiyear test period or assume that a potential *direct* influence of time is insignificant, compared to other factors, such as canopy cover loss or forest type.

Relying primarily on visual examination of Landsat temporal trajectories in various VIs at a pixel can at times be a highly reliable and effectively independent method, when the disturbance magnitude is high. However, the uncertainty naturally increases when the visually perceived changes in time series are modest or confounded by noise or unaccounted influence of seasonality. In these situations, discrimination between disturbances versus noise or multi-annual scale periodicity in ecosystem development may require statistically sound multivariate modeling and a sophisticated algorithm. Simultaneous visual interpretation of multiple VIs can mitigate this problem; however, in our experience, many types of true disturbances and likely false positives have very similar cross-VI correlations, thus exposing an interpreter to notorious challenges and often counter-intuitive nature of inference in high-dimensional spaces. Given that a visual interpretation of a time series can be seen as another type of algorithm, the validation results can be seriously skewed if our visual interpretation algorithm happens to mimic the logic behind the one being validated (e.g. due to human subconscious biases) or if it implicitly defines the term “true” disturbance as a certain type of Landsat trajectories. Automated software packages, most prominently TimeSync (Cohen et al., 2010), can simplify this process and reduce the above-mentioned risks by conveniently displaying multiple Landsat VI trajectories, multiscale spatial context, and Google Earth images. Nevertheless, even with the proven added value of TimeSync, what determines the actual output are the weights with which these different sources of information are combined by the analyst in various situations and the chosen thresholds. To our knowledge, detailed and quantitative protocols describing these choices have not been published.

Based on the above considerations, we made the following choices:

1. Limit the test period to a period when high quality high-resolution images are available over the test scene and assume that algorithm performance does not depend on time directly.
2. Maximize the use of all available high-resolution image sources, which for our Scene included Google Earth, orthorectified 2-m resolution WorldView 2/3 imagery during 2010–2016 (over ~90% of

the Scene), and National Agriculture Imagery Program (NAIP) at 1-m resolution for years 2010, 2012, 2014, and 2016.
3. Focus validation on disturbance types that are observable and distinguishable without the aid of Landsat data. For example, harvest, fire, mortality, wind, mudslide can be identified using high-resolution imagery, and/or bitemporal LiDAR data, and are often corroborated by other sources, such as wildfire incident records and forest management activity databases.
4. Use available bitemporal high-density airborne LiDAR data as reference source for evaluating harvest detection.
5. Use a visual review of Landsat VI trajectories *exclusively* for improving accuracy of disturbance timing for a reference disturbance event identified *solely* based on a visible change in a high-resolution image pair. In other words, visual analysis of Landsat neither creates nor dismisses candidates for disturbance events, nor was Landsat used to create or dismiss events of type "no-disturbance".
6. Discard all candidate-events for which available high-resolution images are not informative (e.g. blurry, grayscale, or misaligned), thus minimizing the proportion of potentially subjective or uninformed decisions by the image analysts.

The most significant potential limitations of the above choices stem from treating invisible disturbances as non-disturbance events in reference data and from the limited spatial and temporal coverage of the reference data.

3.3. Test period, disturbance classes, sampling, and assessment

The eDaRT control parameters and settings were chosen to closely resemble a typical current operational application. During the Training stage 15 basis images of 2007–2008 were selected, including Landsat 5 and Landsat 7 images. The Detection period spanned 2009–2016. For reasons explained in section 2.7 we used Landsat 7 images for detection only during 2011–2012 when it was the only operational satellite. Only images acquired during May–November of each year with "non-clear" pixels less than 60% in the Scene were processed, typically 8–11 images a year. The eDaRT performance was evaluated over the eight year period of 2010–2016 referred to as the *Test period* hereafter.

3.3.1. Sampling units and classes

Consistent with eDaRT outputs, the unit of sampling and assessment in all of our experiments is an *event* (either a disturbance event or the event of "no-disturbance") at an individual pixel. Five classes of events in space and time were initially defined and named: "No disturbance", "Fire", "Harvest", "Mortality", and "Other". Class definitions are provided in Table 1. In addition to simple labeling, these classes (with the exception of class "Other") are meaningful from ecological and forest management perspectives and defined independently of eDaRT methodological features. It can be seen from Table 1 that events of class 'Other" are largely unobservable by our chosen reference data assessment method. Therefore, evaluation of eDaRT performance with respect to this type of disturbance is not included in this paper and is left for future studies.

3.3.2. Reference data sampling and assessment

The sample events for class "Harvest" were derived from bitemporal LiDAR data and thus locations and timing for the sampled events were beyond our control. Section 3.5 encapsulates a detailed account of the LiDAR data and "Harvest" detection assessment.

The reference events for classes "Mortality", "Fire", and "No Disturbance" were developed based on multitemporal high-resolution imagery, which gave us more flexibility in choosing the sampling method. Although random sampling would be appropriate for our tests, a very large number of samples may be required to ensure sufficient representation of less common types of events. Therefore, we used a two-stage approach that also proved effective in previous studies (e.g. Zimmerman et al., 2013). First, a few large primary sampling areas (PSA, $> 10^5$ pixels) were selected based on *a priori* knowledge of the scene vegetation structure and disturbance history, and then in the second step individual pixels were chosen at random from each PSA. The entire Scene was also used as a PSA. Only forested pixels were sampled, based on the NLCD-2006 map. The PSA for class "Fire" included 5-km buffered final perimeters for three large fires: French, Aspen, and Rough (Fig. 5).

Consistent with previously published validations of Landsat scale disturbance products (e.g. Kennedy et al., 2010; Zimmerman et al., 2013; Cohen et al., 2017; Healey et al., 2018), the temporal scale of reference events was set to one calendar year, i.e. the reference event must characterize the corresponding annual differences. For each sampled pixel, the analysts reviewed the available high-resolution images and either confidently identified one annual event of type "No Disturbance", "Fire", "Harvest" or "Mortality" or discarded the pixel. For disturbance events, the analysts further recorded the disturbance class, the pre- and post-disturbance high-resolution image dates, estimated canopy cover differences, and for all event types assessed the pre-event land form type composition at the pixel: conifer (C), hardwood (H), shrub (S), and barren (B), up to two dominant types for most pixels. Class barren represents a broad category that may include soil, rock, grass, and other potentially sparsely vegetated subtypes which could resemble bare soil in the high-resolution imagery. To assign a land form type class, we first assigned a general dominant vegetation or cover type and a second type if it comprised > 5% of the pixel area, including both the overstory and understory. For example, if a pixel comprised 30% conifer cover, 50% shrub cover, and 20% barren, we would assign it a cover type of 'SC' (shrub/conifer). In cases where the overstory was dominated by one type, and the understory was not visible, we assigned a single cover type.

The years of disturbance events were determined differently for classes "Fire" and "Mortality". For wildfires, the date was readily available from the respective incident activity records. For all "Mortality" events, the differences between the calendar years for the high-resolution images used were either one or two years. In the case of a two-year difference, the disturbance year was determined by examining trajectories of Landsat vegetation indexes (Tasseled Cap Angle, NBR, and NDVI) at the pixel and selecting the year of more significant and consistent apparent change in these trajectories. Although an analyst's interpretation of trajectories is always somewhat subjective

Table 1
Classes of disturbance and no-disturbance events (see sect. 3.3.1 for details).

Class Name	Description
Fire	Observed changes in pixel (s, t) are dominated by biomass burned, which may or may not have affected tree canopies
Harvest	Observed changes in pixel (s, t) are dominated by mechanical or natural removal or disappearance of live vegetation canopy, which could include mastication, clear-cut, partial/selective logging, blow-down (wind), mudslide
Mortality	Observed changes in pixel (s, t) are dominated by completely or partially dead tree canopies (i.e. the trees are present, not removed or burned)
Other	A disturbance event is observed at (s, t); however it does not belong to the above disturbance classes, e.g. loss of vegetation canopy cover, greenness, and moisture, that are anomalous with respect to a reference time t_0, but could not be confidently assigned to any other single disturbance class or event
No disturbance	No disturbance event (s, t, t_0) is observed at pixel s and time t, relative to the baseline time t_0, which may include normally developing, intact, or regenerating forest

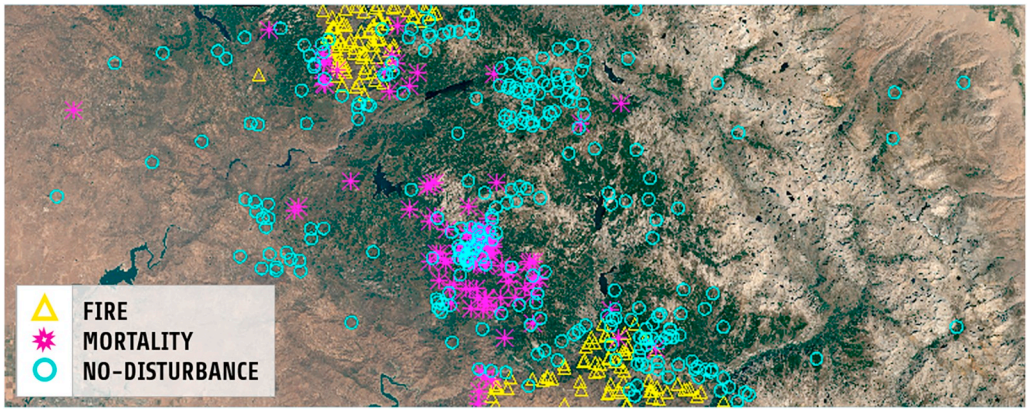

Fig. 6. Pixel locations in the test Scene for reference Fire, Mortality, and No-Disturbance annual events during 2010–2016 (see sect. 3.3.2).

and depends on the analyst's knowledge and expertise, the effect of potential errors on the validation results was reduced because the analyst had to choose between only two years, and not for all samples. When the choice was not obvious from both the Landsat trajectories and the high-resolution image dates, the event was discarded. It is important to emphasize again, that we used the Landsat data not to determine whether or not a disturbance occurred, but only to improve the disturbance timing estimate for theses disturbances.

The resulting reference sample included: 120 “Mortality” events, 159 “Fire” events, and 411 “No-disturbance” events. Fig. 6 shows the geographic distribution of the sampled pixels in the Scene. Fig. 7 provides examples of reference disturbance and no-disturbance events and illustrates the decisions made by the analysts.

3.4. Detection of “fire”, “mortality”, and “no-disturbance” events

To match eDaRT events with the reference “Fire”, “Mortality”, and “No-disturbance” annual events we downscaled the eDaRT event timing outputs to one year resolution. An eDaRT event was considered matching the reference event if the year of detection was accurate up to one year margin, which is similar to previously published evaluations of other disturbance monitoring algorithms (Cohen et al., 2017; Healey et al., 2018). Both disturbance and non-disturbance reference events were matched using the above same temporal margin to ensure equal treatment of both types of events. Adding margins relaxes constraints on temporal precision of matching and usually inflates both true positive and false positive rates, as compared to settings in which temporal accuracy of outputs must meet higher standards. Considering correct only same-year detections (i.e. 0-margin) would provide an overly conservative accuracy assessment, especially in areas where late year fires and mortality events are dominant. For example, during the Test period at least one fire (Rough incident, started on July 31, 2015) was a protracted incident, and large portions of the Scene were affected by heavy smoke and/or late autumn cloudiness in many post-ignition images, thus rendering many disturbances of autumn 2015 unobservable until spring 2016 or apparent recovery. In general, different temporal matching rules may be preferred by different applications.

3.4.1. “Mortality” and “fire” detection performance

The assessment of “Mortality” detection performance is summarized by Fig. 8. The top panel in this figure shows the histograms of the estimated canopy cover loss (ΔCC) and the land form types for the 120 reference “Mortality” events. As our sample was not stratified by vegetation type, it is dominated by conifers (78%), likely due to the actual mortality primarily affecting conifers in this area. According to the analysts’ estimates, no mortality in the sample exceeded the magnitude of 50% cover loss. Low magnitude disturbances, with canopy cover loss of 5%, 10%, 15%, and 20%, represented by 22%, 20%, 21%, and 18% of events, respectively; and together comprised 80% of the reference “Mortality” events.

The bottom panel in Fig. 8 graphs detection probability P as function of ΔCC. The circles are the actual detection rates per 5%-wide bin of ΔCC. The solid black line is the generalized linear model (6) fitted to these data, with log-log link function:

$$\ln(-\ln(P)) = a \cdot \Delta CC + b, \quad (6)$$

with $a = -0.0855$, $b = -0.1105$; and the two red dotted lines are the model's upper and lower 95% confidence bounds. According to this model, for a hypothetical sample of mortality events in which any value of ΔCC is equally likely to occur, eDaRT is expected to detect 92% of those events, as estimated by the expected value of the conditional probability distribution modeled by eq. (6). Lower magnitude mortality tends to be more difficult to detect. For example, for ΔCC = 5% the detection probability is 0.56; and a 30% canopy loss can be detected with probability 0.93.

Fig. 9 (top panel) depicts the distributions of 159 “Fire” events with respect to ΔCC and land form types. The bottom panel plots detection rates (as circles) for 5%-wide ΔCC bins and the corresponding detection probability model eq. (6) with coefficients $a = -0.1145$, $b = -0.6927$. The figure shows that eDaRT is expected to detect 96% of burned forest pixels, when all levels of canopy cover loss are equally represented. For ΔCC = 5% the detection probability is 0.67, and 30% canopy loss has a 98% expected chance of detection.

3.4.2. “No-disturbance” events and the upper bounds on error of commission

Fig. 10 presents the false-positive rate assessment experiment in which 411 “No-Disturbance” reference events were matched to eDaRT events and grouped by the land cover type. The bottom panel in this figure shows a 9.7% false positive rate for conifers, the most dominant group in the sample. Hardwood-dominated pixels and hardwood-conifer mixtures tend to have higher commission error; however, many of these classes are represented by a small number of reference events (as shown by the histogram in Fig. 10 and the standard error bars). The overall error of disagreement amounted to 12.2%. It should be noted that by construction and by definition, our reference “No-Disturbance” events may have included disturbances that are unobservable in high-resolution imagery, such as stress due to low canopy water content or loss of less than 5% live canopy. Therefore, an argument can be made that the actual false positive rates may be lower.

3.5. Harvest detection in bitemporal LiDAR coverage areas

To evaluate eDaRT sensitivity to subtle changes in vegetation canopy cover (CC) and estimate probability of harvest detection as a function of CC loss we used bitemporal airborne LiDAR data available

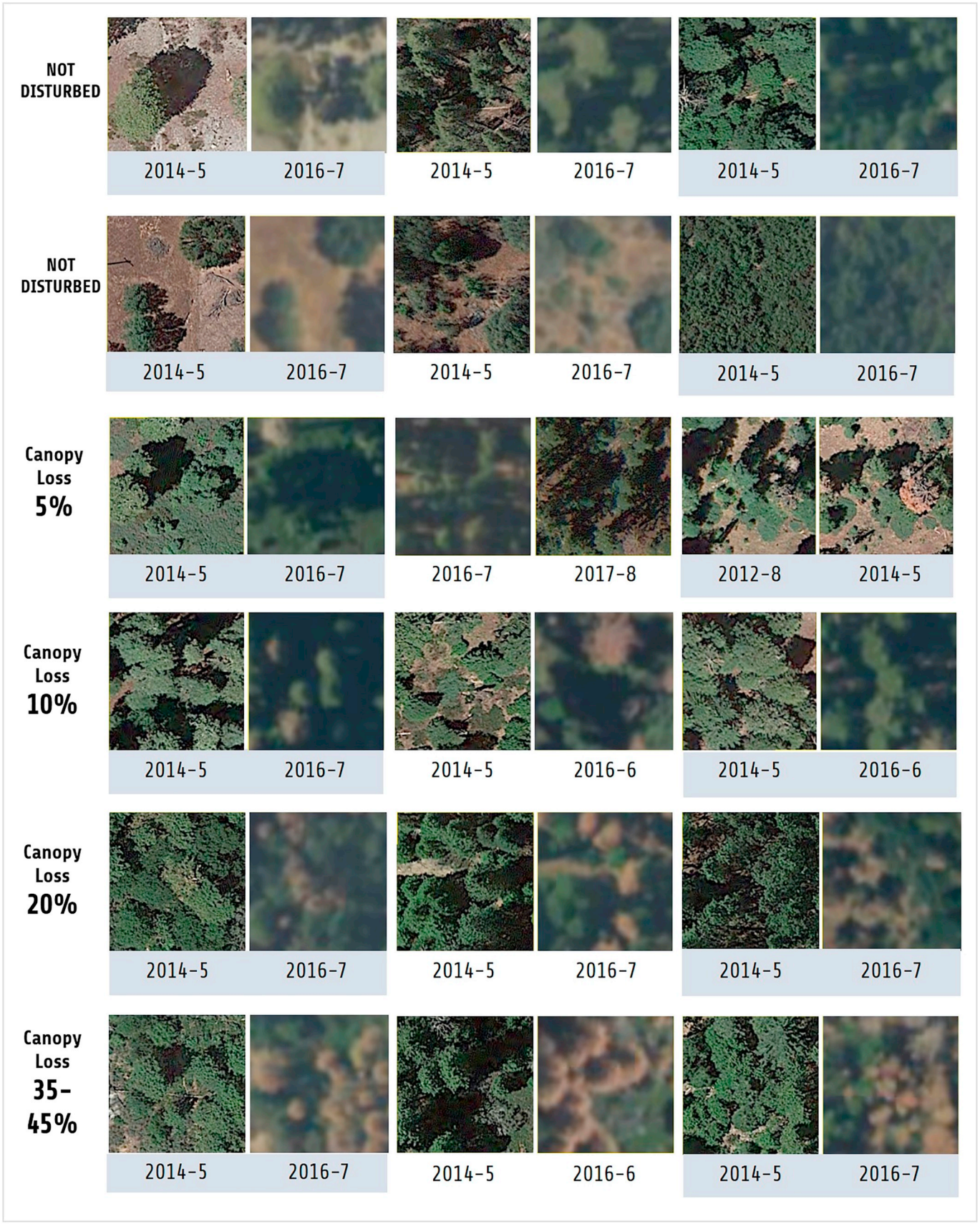

Fig. 7. Examples of reference "No-Disturbance" and "Mortality" events developed using interpretation of high-resolution imagery (sect. 3.3.2), with the analyst-estimated canopy cover loss. The pre-event and post-event images clipped from Google Earth, NAIP.

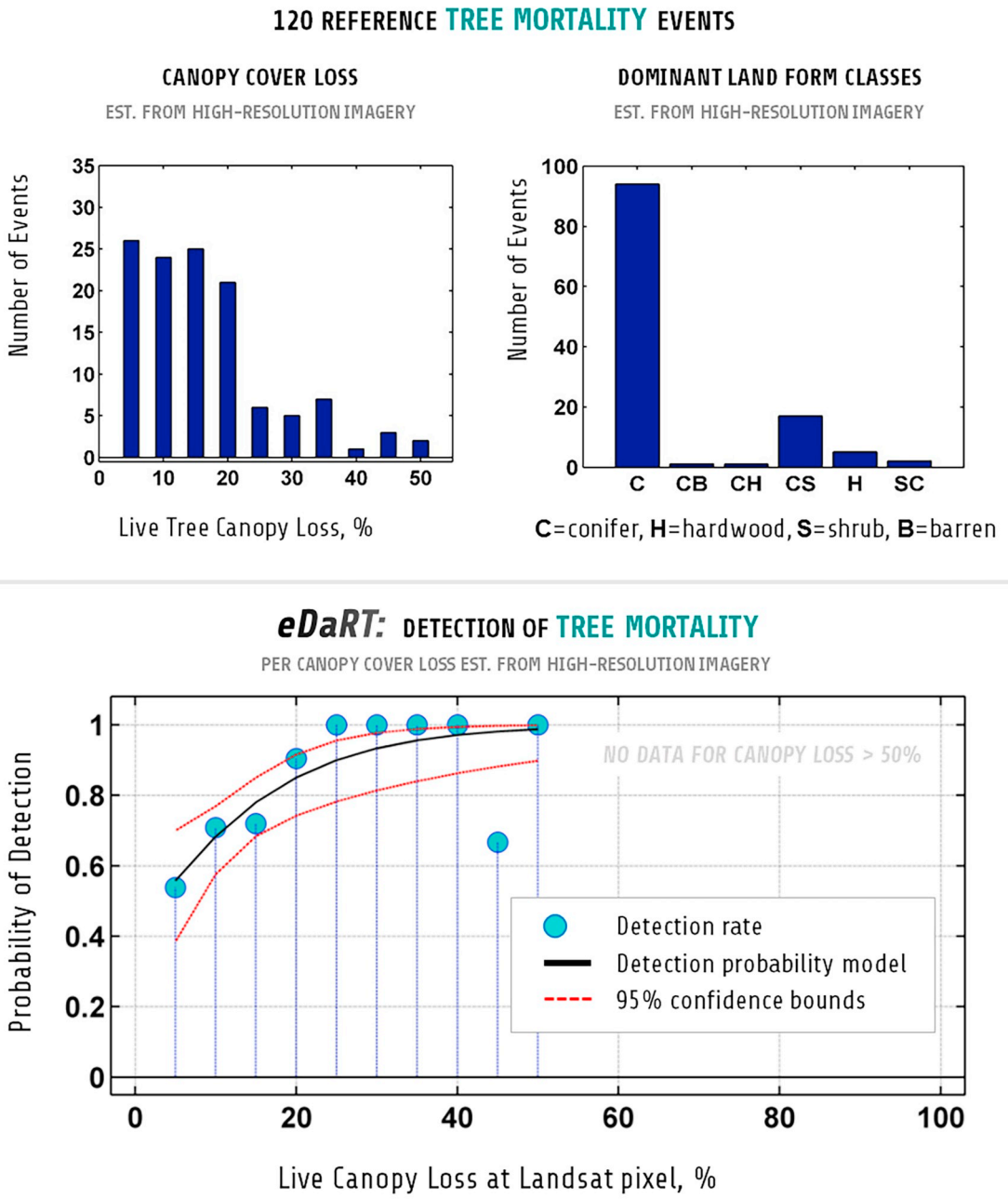

Fig. 8. Reference tree mortality events and their detection by eDaRT.

for some areas of silvicultural treatments in the Scene. These treatments were designed to improve forest structural heterogeneity and thus included a large number of very subtle changes that would be detectable by LiDAR-derived canopy cover, using a minimal height threshold of 1 m above the bare earth surface. The LiDAR datasets were acquired in October 2010 and November 2012 at a minimum density requirement of 8 pulses/m^2. The bitemporal coverage area is shown in Fig. 11a).

The LiDAR point cloud data was processed in the FUSION software package (McGaughey, 2016) using the classified point cloud and a bare earth digital elevation model. To estimate the reference canopy cover loss $\Delta CC_L = CC_{2010}\text{-}CC_{2012}$ at Landsat pixels, we initially calculated canopy cover above 1 m at 1-m resolution for each LiDAR acquisition and then aggregated the 1-m cells to 30-m Landsat pixels. This process did not differentiate between the loss of live and dead canopy, nor did it attempt to account for stem area or other non-photosynthetic components of canopies. Therefore, it is reasonable to assume that ΔCC_L is consistently overestimating the actual loss of live canopy, which can inflate omissions rates, because removal of dead biomass is not defined or detected by eDaRT as a disturbance. The ΔCC_L values at 30-m scale were binned into one percent wide intervals and the negative ΔCC_L values, i.e. canopy cover gain, were zeroed.

The map of the resulting ΔCC_L is shown in Fig. 11b) and the canopy loss histogram for 35,141 test pixels with $\Delta CC_L > 0$ is displayed by Fig. 12 (top panel). Both the histogram and the map appear consistent with the objectives and method of the treatments, as the overwhelming majority of pixels show very low intensities of harvest. Indeed, every 1% bin where $\Delta CC_L < 20\%$ is represented by hundreds to thousands of pixels, whereas intensities $\Delta CC_L > 30\%$ have only a few pixels per bin, if any.

For the sake of consistency with other experiments presented

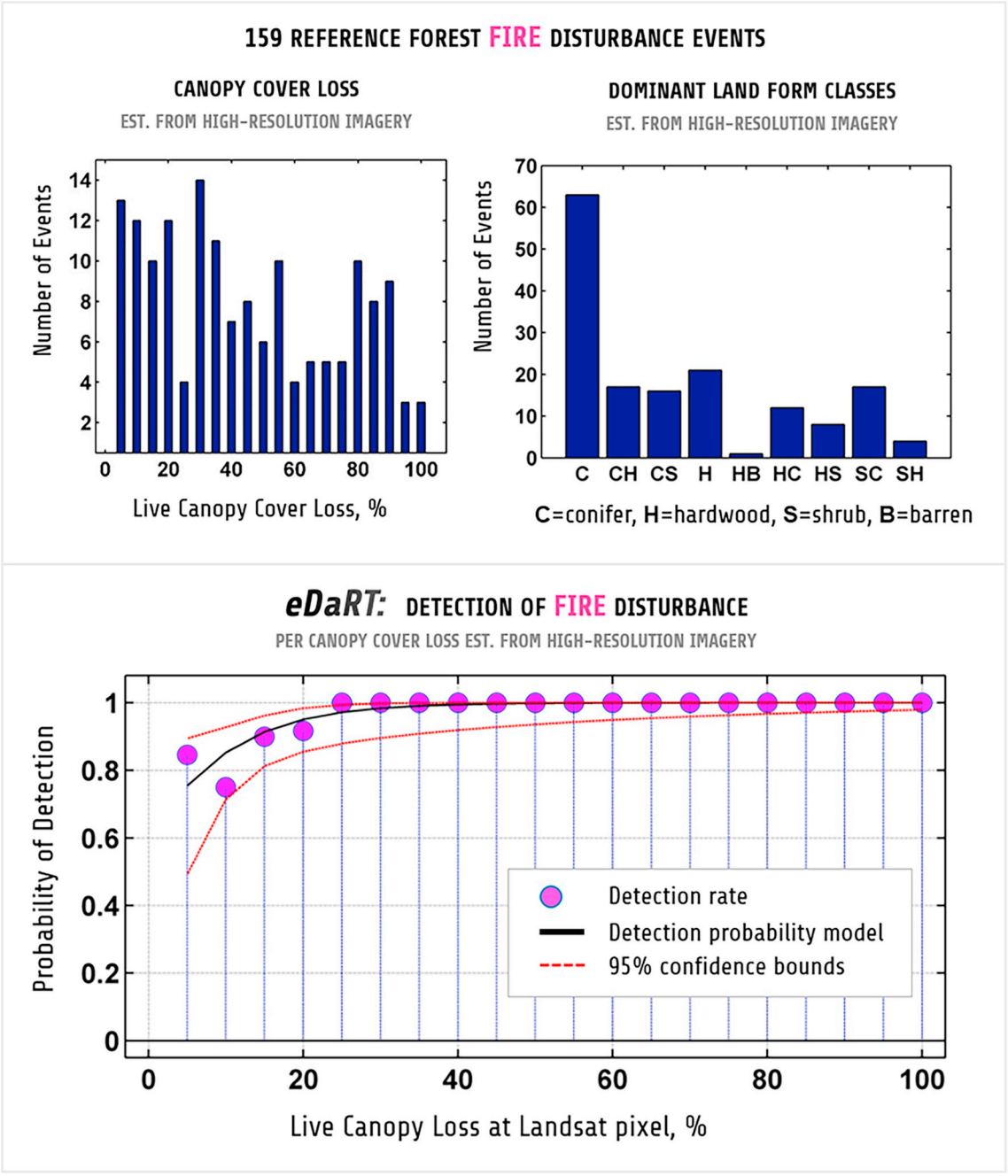

Fig. 9. Reference "Fire" disturbance events and their detection by eDaRT.

herein, a reference pixel with positive ΔCC_L was considered matched by an eDaRT event if the event time was between the LiDAR data collections up to one year margin on the right to accommodate for potential lack of Landsat observations for treatments occurred in late 2012. The detection rates for the 1% bins of ΔCC_L are shown in Fig. 12 (bottom panel) as circles. The two dashed lines represent 95% confidence bounds for the detection probability model (6) with the estimated coefficients: $a = 0.1051$, $b = -0.6155$. The large number of samples for bins of $\Delta CC_L < 30\%$ and highly consistent performance of eDaRT made the confidence bounds tight and so the lines appear nearly identical. The lower performance points in the ca. 30–60% range of ΔCC_L, where we normally had just a few pixels per bin, are extreme outliers to the model. As evident from Fig. 12, eDaRT is routinely capable of detecting as little as 1% change in canopy cover (with prob. 0.19), loss of 10% canopy has 52% chance to get detected, and for a 20% loss the probability of detection is 0.80. When the distribution of canopy losses across harvested pixels is uniform between 1% and 100%, eDaRT is expected to detect ~87% of harvest events.

4. Discussion

4.1. General remarks

Presently, eDaRT is still a prototype algorithm requiring further development, optimization, and validation. Nearly every module of the algorithm can be significantly improved. Below, we briefly mention a few aspects related to eDaRT advantages and limitations.

- As a member of the family of algorithms designed to process dense time series of Landsat images, eDaRT is highly suitable for

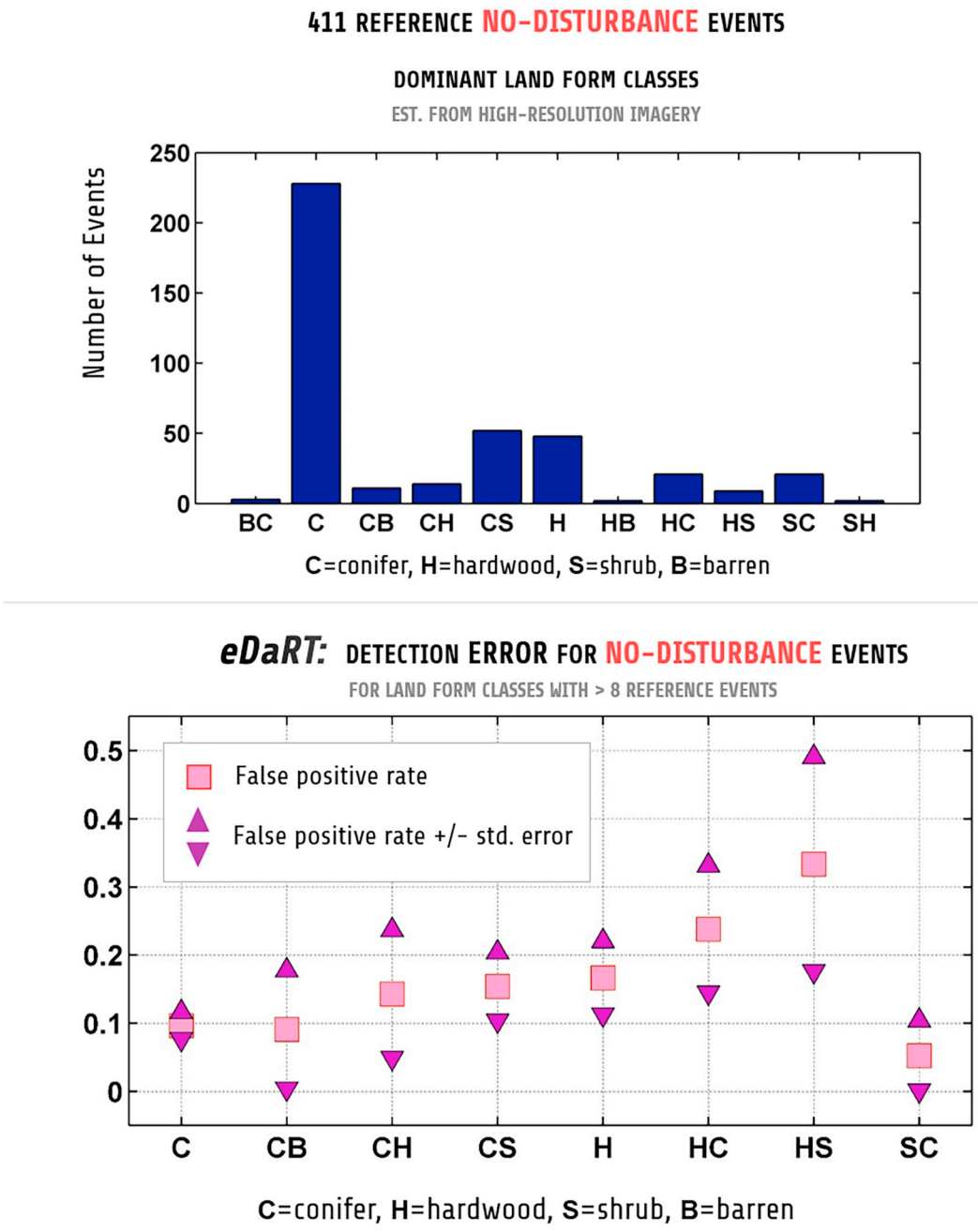

Fig. 10. Reference "No-Disturbance" events and probability of commission error by eDaRT with respect to these data.

retrospective studies, year-around continuous monitoring, and time-critical applications. Using dense time series of images generally provides a greater temporal accuracy (up to 8–16 days) of disturbance detection and a higher sensitivity to subtle disturbances. However, as discussed in section 2.7, the balance between detection sensitivity and reliability is expected to vary across seasons and also significantly depend on the actual time step of observations for a given pixel during the Detection stage (but not the Training stage), because the algorithm control parameters' thresholds are based on the number of available clear-view observations. We believe this problem exists in any dense time series algorithm (including those mentioned in sect. 1) that defines window sizes by the number of observations in them, without also accounting for variable frequency of observation. To mitigate this problem for operational users, we compile image sequences from a single Landsat swath per eDaRT scene and avoid using Landsat 7 during the Detection stage, unless other satellites are unavailable or quick production is a high priority. Our ongoing work is aimed at enhancing the event extraction algorithm to address this issue.

- Should a dense time series algorithm always try to utilize all available satellites and images? An immediate "yes" is a tempting answer; however, there are practical considerations that may suggest a less extreme approach. One of such considerations was discussed in the previous paragraph. Another consideration is that all algorithms rely on their built-in or image provider's capabilities in deciding whether or not an observation is a valid (clear-view and

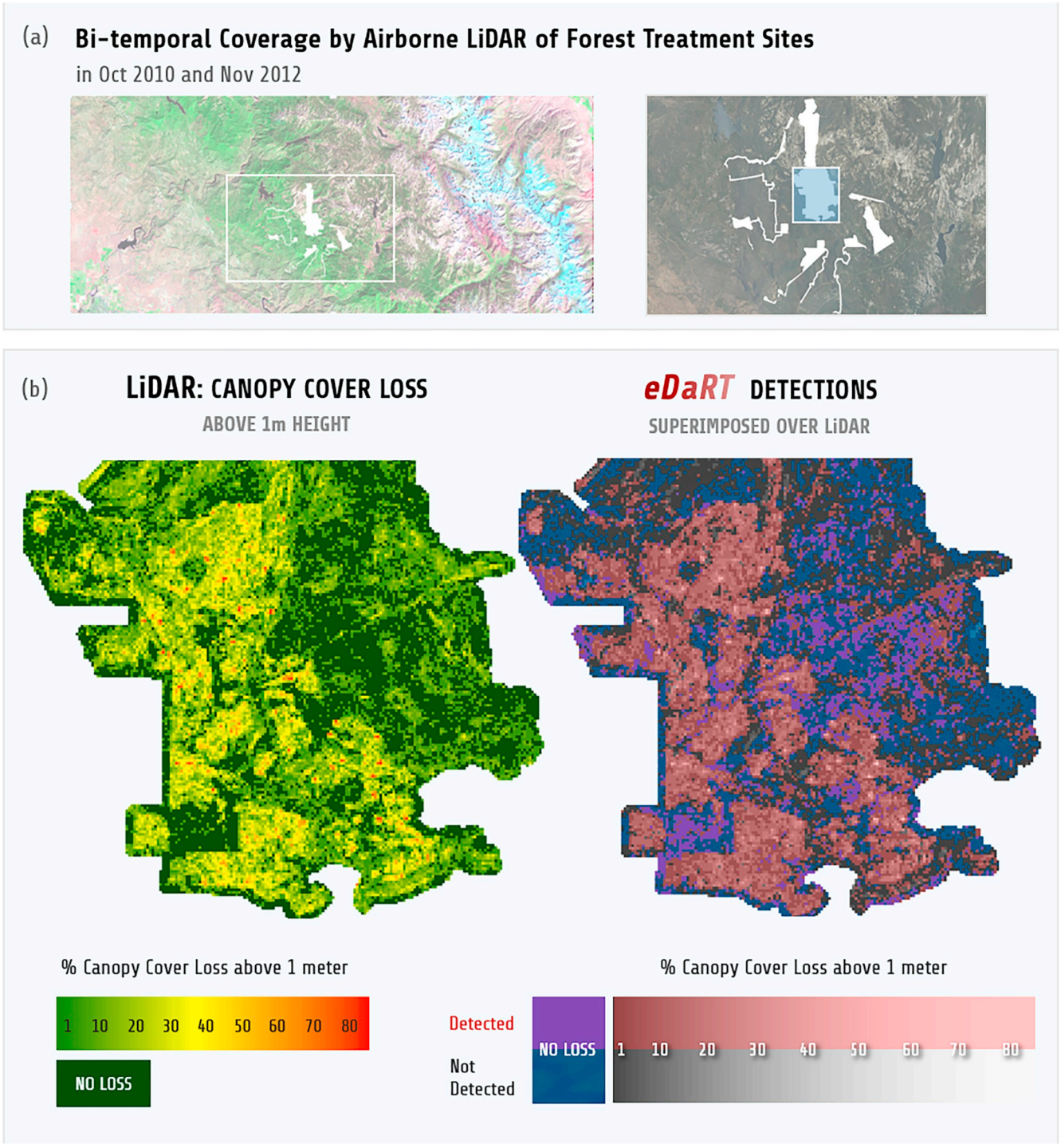

Fig. 11. Bitemporal LiDAR coverage and eDaRT detection of harvested areas.

non-contaminated) observation. What if those capabilities are routinely insufficient, say, during periods of protracted and spatially extensive cloudiness and snow cover? In this case, the potential benefits of occasionally including additional valid observations for some pixels could be outweighed by the massive influx of undetected temporally and spatially correlated noise in the time series for other pixels. We are not aware of published dense time series algorithms that have proven themselves immune to performance deterioration during cloud/snow seasons, and eDaRT is not an exception. Thus a reasonable potential alternative to processing *all* available images is to exclude high-cloud seasons and images, making the time series overall more reliable albeit less dense. Following this approach, the validation experiment in this paper was conducted in seasonally limited settings, and thus the observed high performance should not be unconditionally extrapolated to complex observation conditions in winter.

- False positives produced by eDaRT cloud detection tend to represent bright and cold surfaces, such as sunlit rock outcrops, some urban areas, and occasionally, some clear-cuts in winter conditions. False clouds on clear-cuts can delay disturbance event detection until a warmer or less cloudy season and can introduce inaccuracy of disturbance timing, typically by a few weeks or months. False clouds on

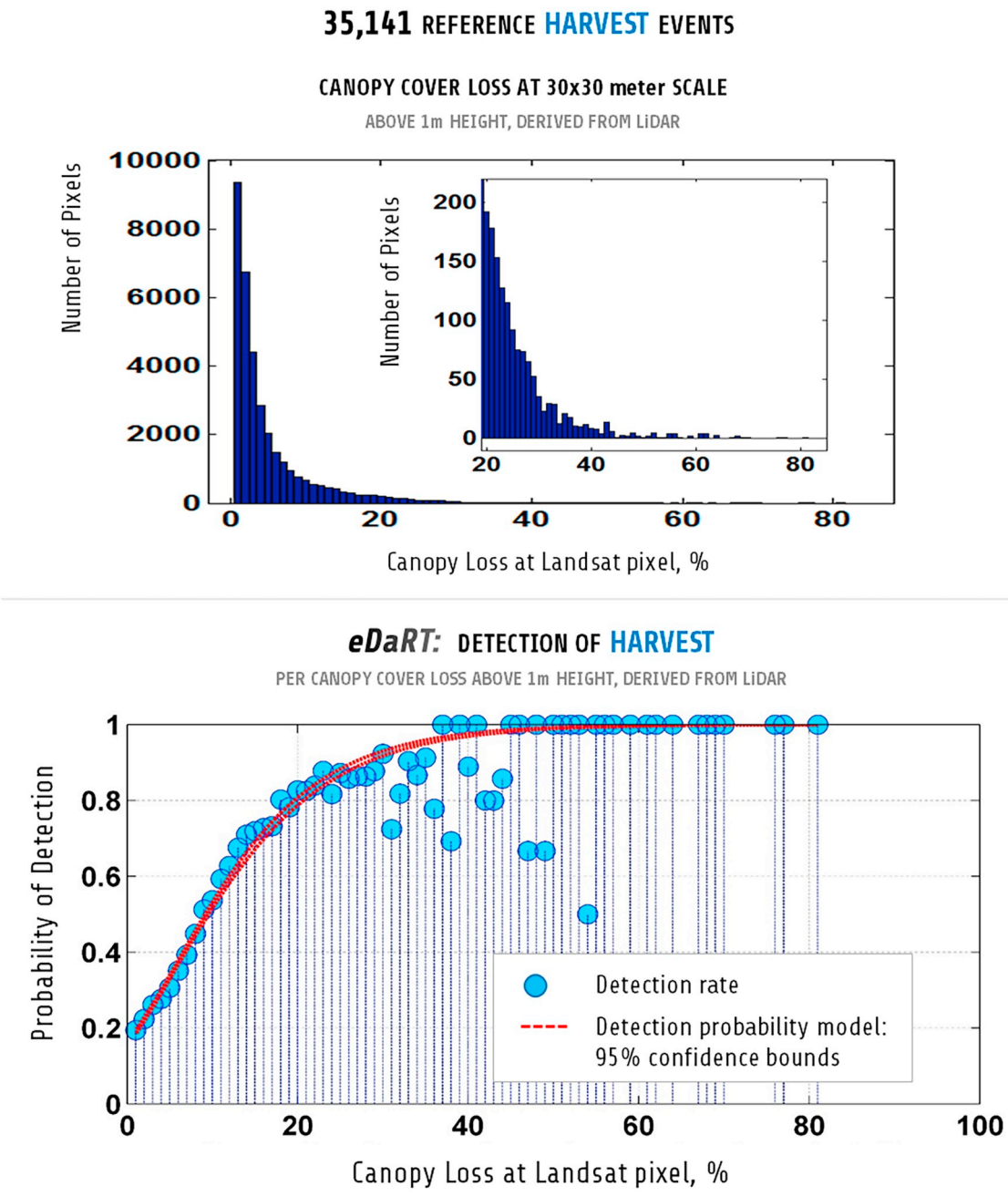

Fig. 12. LiDAR-derived reference "Harvest" events and their detection by eDaRT.

the other types of surfaces are of little importance to mapping vegetation changes in natural ecosystems (forests, shrublands and grasslands), which is the scope of eDaRT. In contrast, false negatives (primarily on semi-transparent clouds, subpixel level snow, and thin smoke) can cause disturbance mapping errors of both omission and commission, and therefore work is underway to improve eCloud algorithms. Another potential direction for improvement could be to utilize operational masks available from the Landsat archive as additional or alternative source of information about contaminated pixels.

- The anomaly detection by DDM looks for anomalies that are relative to the *actually observed* values at training pixels in the inspection image. In this way the DDM provides eDaRT with built-in relative cross-date normalization capabilities. Therefore, issues such as cross-satellite calibration, which can impact the effectiveness of a classic time series method, are mostly immaterial to eDaRT. Because the model is universal enough to apply to reflective bands, indexes, as well as to temperatures, thus eliminating the need for physical modeling or fine-tuning reflectance or temperature thresholds that are notoriously variable in space and time. In this way, eDaRT methodological basis can become the foundation for a future comprehensive multi-satellite forest monitoring system.
- In every image eDaRT attempts to find anomalies by estimating the change in the band values that cannot be explained by normal development of similar landscapes in a given scene. As an advantage, the error due to common natural phenological variations on sub-

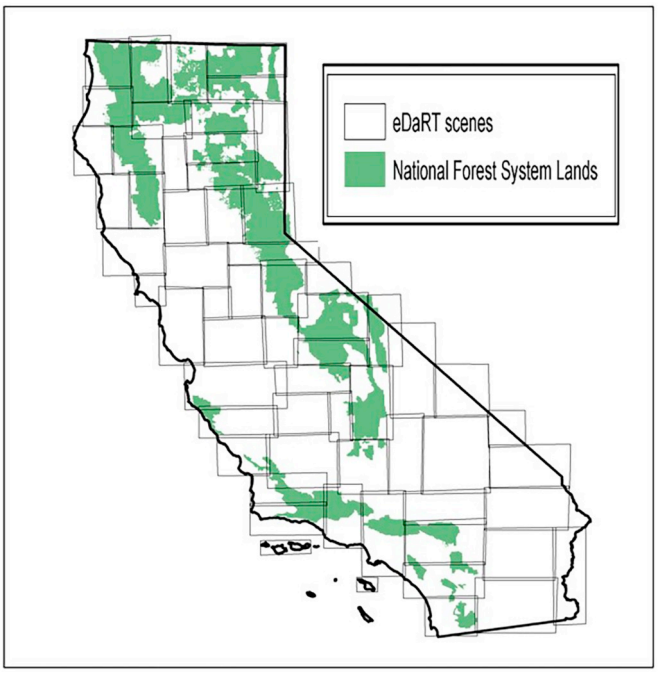

Fig. 13. California scenes independently monitored by eDaRT in regional applications.

annual (e.g. senescence) and multi-annual (e.g. early senescence) scales is mitigated, too. However, when *most* pixels of a given land cover type experience prolonged period of gradual decline and thus can be considered as disturbed in some applications, these pixels cannot be considered anomalies because they are the very pixels used in the inspection image to empirically characterize normal, non-anomalous change. To minimize the use of disturbed pixels as indicators of normal spectral response in a given image, eDaRT relies on algorithm's cascades, iterations, and robust regression.

4.2. eDaRT application overview

In a production environment, as currently in the USFS Pacific Southwest Region (Region 5), eDaRT can process a large area, e.g. California or the western US, by dividing it into smaller scenes to be independently (serially or in parallel) monitored by separate instances of the eDaRT system software (see Fig. 13 for an example).

Since 2012, when the system was first used to map disturbance history in sensitive species habitat (Tempel et al., 2016), its products have been increasingly utilized in forest management and research applications, including:

- **Monitoring tree mortality in California.** The massive scale of this event that peaked in 2016 spurred the USFS and CAL FIRE to expedite and expand monitoring, activity planning, and public information sharing. These efforts relied on the long-term Aerial Detection Survey (ADS) program, an operational airborne campaign of experienced mappers that delineate mortality through visual assessment and hand-drawing. Tree mortality maps derived from eDaRT complemented the ADS data where weather or other factors limited airborne surveys or where higher frequency or higher spatial resolution data were needed, e.g. the lower montane belt of the Sierra National Forest, where severe mortality was coupled with complex wildland-urban interface (Axelson et al., 2019).
- **Forest restoration management and planning.** Targeted vegetation management toward a reference condition (North, 2012) typically occurs at a project scale of 10s to 100s of hectares, and therefore the 30-m scale of canopy status and change information is often appropriate to support these activities. For example, on the USFS Sierra National Forest, silviculturists and vegetation managers combined eDaRT disturbance age and intensity outputs with slope, soil type, accessibility, and other environmental variables to develop initial reconnaissance maps, prioritize locations, and design silvicultural prescriptions for forest re-establishment and growth. Similarly, eDaRT data helped revise an Environmental Assessment on the Inyo National Forest in 2017 and address objections that the assessment did not adequately account for potentially ubiquitous recent mortality in the project area. Using eDaRT, we estimated that at the time only ~0.6% of the project area had undergone significant canopy loss at 30-m scale beyond background mortality levels.
- **Wildfire incident management support.** At the outbreak of a significant fire, teams rapidly assemble available landscape and fuel data to inform fire behavior modeling and assignment of personnel and resources to the incident. For example, standing dead trees present hazards to firefighters. eDaRT-customized disturbance history maps and intermediate layers, such as relative change metrics derived from the DDM residuals, contributed to time-critical decisions during many California fires, including large incidents (name/year): King/2014, Rough/2015, Schaeffer/2015, Cedar/2016, Pier/2017, Empire/2017, Detwiler/2018, and Ferguson/2018.
- **Targeting systematic updates to vegetation maps.** Historically, USFS Region 5 Existing Vegetation datasets have been updated on an as-needed basis, using limited and qualitative information to assess if changes warrant updates and then requiring extensive ground survey or high resolution image interpretation to conduct such updates. During 2016–2017, eDaRT-derived disturbance maps identified over 32,000 ha of forest in the southern Sierra Nevada where updates were high priority due to tree mortality and wildfire.

Key general lessons we have learned in these applications include:

1. Users in forest management and research communities tend to benefit tremendously from close collaboration with disturbance product developers, who can give unique insight into products' strengths and limitations with respect to the users' goals, suggest most effective and tenable uses, or advise against using their product. Early engagement pays off further down the road.
2. Early feedback from users is valuable to prioritize further developments. Most frequently, users request that eDaRT provide disturbance types and direct measures of canopy cover loss, both for discrete events and time intervals.
3. Definition of a “meaningful disturbance” varies greatly between different users and applications, and therefore the algorithm and processing system should be flexible enough to enable customized products based on users' priorities.

5. Summary and conclusion

This paper presented the algorithm (version 2.9) and discussed the operation and initial validation of the Ecosystem Disturbance and Recovery Tracker system developed in response to the growing and diversifying need for timely and accurate information about ecosystem status and disturbance at Landsat spatial and temporal scales. eDaRT represents a newer class of Landsat time series algorithms designed to detect disturbances in all available/usable images and also combines groups of previously published and novel image analysis sub-algorithms. These algorithms, however, have room for optimization and enhancement. The most notable current functional limitation is that eDaRT does not yet categorize disturbances by their type or cause. The disturbance classifier module is currently under development.

The presented initial tests in central Sierra Nevada, California, at the scale of 1x1 pixel for annual change events suggest strong detection capabilities for major ecologically meaningful classes of disturbance: wildfire burn, tree mortality, and forest harvest/treatments. Since

detection outcome is most directly influenced by the effect magnitude, we have provided detailed statistics describing how detection performance depends on the canopy cover loss on the ground. One of these experiments, detection of subtle silvicultural treatments, used more than 35,000 reference Landsat pixels in which low levels of canopy cover loss were accurately measured by LiDAR down to ~1% accuracy. Furthermore, based on more than 400 reference "no-disturbance" samples (potentially including invisible true disturbances), we observed the overall false positive rate as low as 12.2%, with lower rates for conifers and higher rates for hardwood-dominated forests. By relying on stratified relative comparisons to detect disturbance, the system, in effect, calibrates itself at three levels: per scene, landcover type, and pixel. This observation further supports our working hypothesis that the encouraging initial results presented in this paper should approximately repeat in similar application settings across California. However, further experiments are needed to fully quantify and reduce the remaining uncertainties regarding the system's performance and its consistency across space and time.

The current implementation of eDaRT system balances the flexibility of a research prototype that enables continuous enhancement and performance efficiency necessary for regional-scale operational application. As a result, while many of the eDaRT blocks are still in development, land managers across California and the science community increasingly use the system's outputs. In most cases, these outputs are bundled with scientific support by the algorithm development team members who help ensure an informed and appropriate use of eDaRT data. We hope that continuing development, validation, and applications of eDaRT will further complement the efforts by other Landsat algorithm development teams, help improve our understanding of ecosystem processes, and support science-based land management practices in California and beyond.

Acknowledgments

This work was supported by USDA Forest Service (USFS) as the main sponsor, through Region 5 Forest Inventory and Monitoring Program, via Cost Share Agreement with University of California Davis 18-CS-11052003-047, and in part by the Disturbance Assessment and Services Program at the USFS Geospatial Technology and Applications Center (GTAC) and University of California Davis, under Cost Share Agreement 10-IA-11130400-009. We thank Tanya Kohler, Kirk Evans, Laura Young, Nathan Amboy, and Carol Clark (USFS, Pacific Southwest Region, Remote Sensing Lab) for their technical and geospatial assistance and data management support. We are grateful to Ricardo Lopez and Jayne Handley (USFS, Pacific Southwest Region, Information Management) and Brad Quayle (USFS GTAC) for continued support of algorithm development, and to Ramiro Rojas, Sheri Smith and Kayanna Warren (USFS, Pacific Southwest Region, State & Private Forestry) for helpful discussions of algorithm applications. We also thank Zhiqiang Yang (Oregon State Univ.) for providing data that helped design validation experiments.

Finally, we are grateful to the anonymous reviewers and the Guest Editors who gave a large number of helpful, constructive suggestions and comments that significantly improved the manuscript.

Appendix 1. eCloud Algorithms Details

This Appendix provides details about the algorithms comprising eCloud module briefly discussed in section 2.5. These algorithms map "non-clear" observations due to clouds, snow, fog, and smoke. Using the band names introduced in section 2.3, we define the eDaRT Cloud-Snow Index (ECSI) that can be viewed as an enhanced temperature-to-albedo ratio:

$$\mathrm{ECSI} = \mathrm{TIR} \,/\, (a{\cdot}\mathrm{TC1} + b{\cdot}\mathrm{B1} + c{\cdot}\mathrm{TC2} + d) \tag{7}$$

where the coefficients $a = 388$, $b = 1900$, $c = 15000$, and $d = 95$. These values are based on the least squares minimization of the error in cloud/snow detection over a set of Landsat training pixels in California forests.

The SCD algorithm utilizes only information from the inspection image and conservatively detects opaque high altitude clouds or snow if any of the following three tests returns true:

$$\mathrm{TIR} < \mathrm{T_{high}} \text{ AND } \mathrm{TC1} > \mathrm{A_{bright}}, \tag{8}$$

$$\mathrm{TIR} < \mathrm{T_{low}} \text{ AND } \mathrm{TC1} > \mathrm{A_{dark}}, \tag{9}$$

$$\mathrm{ECSI} < \mathrm{C_{min}}, \tag{10}$$

where thresholds $\mathrm{A_{dark}} = 0.08 \cdot \Sigma\tau_k$ and $\mathrm{A_{bright}} = 0.27 \cdot \Sigma\tau_k$ are defined in terms of the fraction of the sum of the TC transformation coefficients τ_k for the TC Brightness band (Huang et al., 2002) to represent low and high albedo respectively; $\mathrm{T_{low}} = 233\mathrm{K}$, $\mathrm{T_{high}} = 273\mathrm{K}$, and $C_{\min} = 0.3$. Because the values of these absolute thresholds are only used for initial prescreening to detect the highest-confidence clouds and snow, in the preliminary experiments we found them satisfactory for the purpose of ecosystem disturbance monitoring over diverse regions in the Continental US and across all seasons.

The MCM algorithm automatically accepts the SCD detections and selects a pool of additional candidates by requiring:

$$\mathrm{TIR} < 278 \text{ AND } \mathrm{TC1} > \mathrm{A_{dark}} \tag{11}$$

Then the algorithm analyzes DDM residuals for ECSI and for eDaRT bands at these pixels (sect. 2.3) and thresholds the corresponding z-scores. A pixel is flagged "non-clear" if any of the tests M1-M3 below returns true for that pixel.

Test M1 returns a high-confidence "non-clear" pixel if:

$$\begin{cases} Z_{\mathrm{TIR}} < 3.0 \text{ and } Z_{\mathrm{B1}} > 0.5 \\ \text{or} \\ Z_{\mathrm{TIR}} < -1.5 \text{ and } Z_{\mathrm{B1}} > 1.5 \text{ and } Z_{\mathrm{ECSI}} < -2.5 \text{ and } Z_{\mathrm{TC1}} > 1.0 \end{cases} \tag{12}$$

Test M2 flags pixels "non-clear" with medium confidence if:

$$\mathrm{Z_{TIR}} < -2.0 \text{ AND } \mathrm{Z_{B1}} > 0.5 \text{ AND } \mathrm{TIR} < 268 \tag{13}$$

OR

$$\mathrm{Z_{TIR}} < -0.5 \text{ AND } \mathrm{Z_{B1}} > 0.5 \text{ AND } \mathrm{TIR} < 263 \tag{14}$$

Test M3 finds additional non-clear observations using condition:

$$Z_{TIR} < -1.5 \text{ AND } Z_{B1} > 2.0 \text{ AND } Z_{ECSI} < -1.5 \text{ AND } Z_{TC1} > 1.5 \quad (15)$$

OR

$$Z_{TIR} < -1.0 \text{ AND } Z_{B1} > 1.5 \text{ AND } Z_{ECSI} < -1.0 \text{ AND } Z_{TC1} > 1.0 \text{ AND } S < -3.5, \quad (16)$$

where $S = Z_{TIR} + Z_{ECSI} - Z_{TC1}$.

Detections of the previous tests are further augmented with lower-confidence detections for which

$$Z_{TIR} < -0.5 \text{ and } Z_{B1} > 0.5 \text{ and } Z_{ECSI} < -0.5 \text{ and } Z_{TC1} > 0.5 \text{ and } S\text{-}Z_{B1} < -2.5, \quad (17)$$

if these pixels are also confirmed by a contextual analysis in a 11x11 window centered at each low-confidence "non-clear" pixel. If at least two higher-confidence detections are found in the window, the center is flagged as "non-clear", too.

The Retrospective Cloud Detection (RCD) algorithm is applied only during the Training stage. This algorithm cannot be applied in the near real-time mode, because for a given image, it takes advantage of the entire input sequence, including both previous and subsequent images. The RCD algorithm proceeds iteratively as follows:

1) Initially assign set *S* to be the entire input sequence.
   a. Define the set of basis images *B* for DDM as the images from *S* with lowest fraction of non-clear pixels.
2) Define the set of inspection images *I* as non-basis images in *S*.
3) Detect clouds in the inspection images:
   a. Estimate background values using DDM method (sect. 2.4.1) with known non-clear pixels and other anomalies masked out for the background estimation purposes.
   b. Apply tests M1-M3 of the MCM algorithm (Appendix 1).
4) Use the resulting maps to detect non-clear pixels in the basis images:
   a. Assign: $S \leftarrow I$, then $I \leftarrow B$, and repeat step 3).
5) Repeat steps 1)-4) using the updated maps.

References

Axelson, J., Battles, J., Bulaon, B., Cluck, D., Cousins, S., Cox, L., Estes, B., Fettig, C., Hefty, A., Hishinuma, S., Hood, S., Kocher, S., Mortenson, L., Koltunov, A., Kuskulis, E., Polini, A., Ramirez, C., Restaino, C., Slaton, M., Smith, S., Strom, B., Tubbesing, C., 2019. The California Tree Mortality Data Network – enhanced communication and collaboration among scientists & stakeholders. Calif. Agric. 73 (2).

Brooks, E.B., Wynne, R.H., Thomas, V.A., Blinn, C.E., Coulston, J.W., 2014. On-the-fly massively multitemporal change detection using statistical quality control charts and Landsat data. IEEE Trans. Geosci. Remote Sens. 52 (6), 3316–3332.

CAL FIRE, 2018. Fire and Resource Assessment Program, Fire Perimeters Version 16_1. California Department of Forestry and Fire Protection Available at. http://frap.fire.ca.gov/data/frapgisdata-sw-fireperimeters_download, Accessed date: 19 March 2018.

Cohen, W., Goward, S., 2004. Landsat ' s role in ecological applications of remote sensing. Bioscience 54, 535–545.

Cohen, W., Yang, Z., Kennedy, R., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync - tools for calibration and validation. Remote Sens. Environ. 114, 2911–2924.

Cohen, W.B., Yang, Z., Stehman, S.V., Schroeder, T.A., Bell, D.M., Masek, J.G., et al., 2016. Forest disturbance across the conterminous United States from 1985-2012: the emerging dominance of forest decline. For. Ecol. Manag. 360, 242–252. https://doi.org/10.1016/j.foreco.2015.10.042.

Cohen, W.B., Healey, S.P., Yang, Z., Stehman, S.V., Brewer, C.K., Brooks, E.B., Gorelick, N., Huang, C., Hughes, M.J., Kennedy, R.E., Loveland, T.R., Moisen, G.G., Schroeder, T.A., Vogelmann, J.E., Woodcock, C.E., Yang, L., Zhu, Z., 2017. How similar are forest disturbance maps derived from different Landsat time series algorithms? Forests 8, 98.

Cohen, W.B., Yang, Z., Healey, S.P., Kennedy, R.E., Gorelick, N., 2018. A LandTrendr multispectral ensemble for forest disturbance detection. Remote Sens. Environ. 205, 131–140. https://doi.org/10.1016/J.RSE.2017.11.015 10.1016/J.RSE.2017.11.015.

Das, A.J., Stephenson, N.L., Davis, K.P., 2016. Why do trees die? Characterizing the drivers of background tree mortality. Ecology 97, 2616–2627. https://doi.org/10.1002/ecy.1497.

Healey, S.P., Cohen, W.B., Yang, Z., Brewer, C.K., Brooks, E.B., Gorelick, N., et al., 2018. Mapping forest change using stacked generalization: an ensemble approach. Remote Sens. Environ. 204, 717–728.

Huang, C., Wylie, B., Homer, C., Yang, L., Zylstra, G., 2002. Derivation of a Tasseled cap transformation based on Landsat 7 at-satellite reflectance. Int. J. Remote Sens. 23, 1741–1748.

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 Sens. Environ. 114 (1), 183–198.

Hughes, M.J., Kaylor, S.D., Hayes, D.J., 2017. Patch-based forest change detection from Landsat time series. Forests 8. https://doi.org/10.3390/f8050166.

Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., Xian, G., 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. Remote Sens. Environ. 132, 159–175. https://doi.org/10.1016/J.RSE.2013.01.012.

Kennedy, R. Z. Yang, Cohen, W., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. Landtrendr - temporal segmentation algorithm. Remote Sens. Environ. 114, 2911–2924.

Kennedy, R.E., Andréfouët, S., Cohen, W.B., Gómez, C., Griffiths, P., Hais, M., Healey, S.P., Helmer, E.H., Hostert, P., Lyons, M.B., Meigs, G.W., Pflugmacher, D., Phinn, S.R., Powell, S.L., Scarth, P., Sen, S., Schroeder, T.A., Schneider, A., Sonnenschein, R., Vogelmann, J.E., Wulder, M.A., Zhu, Z., 2014. Bringing an ecological view of change to Landsat-based remote sensing. Front. Ecol. Environ. 12, 339–346.

Koltunov, A., Ustin, S.L., 2007. Early fire detection using non-linear multitemporal prediction of thermal imagery. Remote Sens. Environ. 110 (1), 18–28.

Koltunov, A., Ben-Dor, E., Ustin, S.L., 2009. Image construction using multitemporal observations and Dynamic Detection Models. Int. J. Remote Sens. 30 (1), 57–83.

Koltunov, A., Ramirez, C., 2012. Toward near real time automated monitoring of anomalous sub-annual scale changes in natural ecosystem development. In: ForestSAT 2012 Conference CD-ROM. September 11-14, 2012, Corvallis, OR, USA.

Koltunov, A., Ramirez, C., Ustin, S.L., 2015. eDaRT: the Ecosystem Disturbance and Recovery Tracking system prototype supporting ecosystem management in California. In: 2015 NASA Carbon Cycle and Ecosystems Joint Science Workshop, April 19-24, 2015, College Park, MD, USA.

Koltunov, A., Ustin, S.L., Quayle, B., Schwind, B., Ambrosia, V.G., Li, W., 2016. The development and first validation of the GOES Early Fire Detection (GOES-EFD) algorithm. Remote Sens. Environ. 184, 436–453.

LANDFIRE, 2013. LANDFIRE existing vegetation type layer; (last updated June 2013). U.S.Department of Interior, Geological Survey. [Online] Available: http://landfire.cr.usgs.gov/viewer/, Accessed date: 8 May 2013.

Loughin, T.M., 2004. A systematic comparison of methods for combining p-values from independent tests. Comput. Stat. Data Anal. 47 (3), 467–485.

McGaughey, R.J., 2016. FUSION/LDV: Software for LIDAR Data Analysis and Visualization. USDA Forest Service, Pacific Northwest Research Station.

Miller, J.D., Thode, A.E., 2007. Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR). Remote Sens. Environ. 109, 66–80.

Miller, J.D., Quayle, B., 2015. Calibration and validation of immediate post-fire satellite-derived data to three severity metrics. Fire Ecology 11 (2), 12–30.

Moisen, G.G., Meyer, M.C., Schroeder, T.A., Liao, X., Schleeweis, K.G., Freeman, E.A., Toney, C., 2016. Shape selection in Landsat time series: a tool for monitoring forest dynamics. Glob. Chang. Biol. 22. https://doi.org/10.1111/gcb.13358.

Mosteller, F., Bush, R.R., 1954. Selected quantitative techniques. In: In: Lindzey, G. (Ed.), Handbook of Social Psychology, vol. 1. Addison_Wesley, Cambridge, Mass, pp. 289–334.

North, M. (Ed.), 2012. Managing Sierra Nevada Forests. Gen. Tech. Rep. PSW-GTR-237. U.S. Department of Agriculture, Forest Service, Albany, CA, pp. 184 Pacific Southwest Research Station.

Schroeder, T.A., Schleeweis, K.G., Moisen, G.G., Toney, C., Cohen, W.B., Freeman, E.A., Yang, Z., Huang, C., 2017. Testing a Landsat-based approach for mapping disturbance causality in U.S. forests. Remote Sens. Environ. 195, 230–243.

Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A., Williams Jr., R.M., 1949. The American Soldier. Adjustment during Army Life, vol. 1 Princeton University Press, Princeton.

Tempel, D.J., Keane, J.J., Gutiérrez, R.J., Wolfe, J.D., Jones, G.M., Koltunov, A., Ramirez, C.M., Berigan, W.J., Gallagher, C.V., Munton, T.E., Shaklee, P.A., Whitmore, S.A., Peery, M.Z., 2016. Meta-analysis of California Spotted Owl (*Strix occidentalis occidentalis*) territory occupancy in the Sierra Nevada: habitat associations and their implications for forest management. Condor 118, 747–765.

USFS, 2018. Forest Service Activity Tracking System. S_USA.Activity_TimberHarvest 12 Mar 2018. Available at. https://data.fs.usda.gov/geodata/edw/edw_resources/meta/S_USA.Activity_TimberHarvest.xml, Accessed date: 19 March 2018.

USGS, 2018. “Landsat collections”, U.S. Geological Survey Fact Sheet 2018–3049 2. https://doi.org/10.3133/fs20183049.

USGS, 2019. What are the best Landsat spectral bands for use in my research?". Available at. https://www.usgs.gov/faqs/what-are-best-landsat-spectral-bands-use-my-research?qt-news_science_products=0#qt-news_science_products, Accessed date: 1 September 2019.

Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. Remote Sens. Environ. 114, 106–115. https://doi.org/10.1016/J.RSE.2009.08.014.

Verbesselt, J., Zeileis, A., Herold, M., 2012. Near real-time disturbance detection using satellite image time series. Remote Sens. Environ. 123, 98–108.

Vogelmann, J.E., Xian, G., Homer, C., Tolk, B., 2012. Monitoring gradual ecosystem change using Landsat time series analyses: case studies in selected forest and rangeland ecosystems. Remote Sens. Environ. 122, 92–105. https://doi.org/10.1016/J.RSE.2011.06.027.

Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P.S., Vermote, E.F., Vogelmann, J., Wulder, M.A., Wynne, R., 2008. Free access to Landsat imagery. Science 320, 1011–1012. https://doi.org/10.1126/science.320.5879.1011a.

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: how free data has enabled the science and monitoring promise of Landsat. Remote Sens. Environ. 122, 2–10. https://doi.org/10.1016/j.rse.2012.01.010.

Zhu, Z., 2017. Change detection using Landsat time series: a review of frequencies, preprocessing, algorithms, and applications. ISPRS J. Photogrammetry Remote Sens. 130, 370–384.

Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171. https://doi.org/10.1016/J.RSE.2014.01.011.

Zimmerman, P.L., Housman, I.W., Perry, C.H., Chastain, R.A., Webb, J.B., Finco, M.V., 2013. An accuracy assessment of forest disturbance mapping in the western Great Lakes. Remote Sens. Environ. 128, 176–185.