

Continuous monitoring of land disturbance based on Landsat time series

Zhe Zhu^{a,b,*}, Junxue Zhang^{a,b}, Zhiqiang Yang^c, Amal H. Aljaddani^b, Warren B. Cohen^{d,e}, Shi Qiu^b, Congliang Zhou^b

^a Department of Natural Resources and the Environment, University of Connecticut, Storrs, CT 06269, United States

^b Department of Geosciences, MS 1053, Science Building 125, Texas Tech University, Lubbock, TX 79409, United States

^c USDA Forest Service, RMRS Research Station, 507 25th Street, Ogden, UT 84401, United States

^d Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, United States

^e Pacific Northwest Research Station, U.S. Forest Service, Corvallis, OR 97331, United States



ARTICLE INFO

Keywords:

COLD
Land disturbance
Change detection
Time series
Landsat
Near real-time

ABSTRACT

We developed a new algorithm for Continuous monitoring of Land Disturbance (COLD) using Landsat time series. COLD can detect many kinds of land disturbance continuously as new images are collected and provide historical land disturbance maps retrospectively. To better detect land disturbance, we tested different kinds of input data and explored many time series analysis techniques. We have several major observations as follows. *First*, time series of surface reflectance provides much better detection results than time series of Top-Of-Atmosphere (TOA) reflectance, and with some adjustments to the temporal density, time series from Landsat Analysis Ready Data (ARD) is better than it is from the same Landsat scene. *Second*, the combined use of spectral bands is always better than using a single spectral band or index, and if all the essential spectral bands have been employed, the inclusion of other indices does not further improve the algorithm performance. *Third*, the remaining outliers in the time series can be removed based on their deviation from model predicted values based on probability-based thresholds derived from normal or chi-squared distributions. *Fourth*, model initialization is pivotal for monitoring land disturbance, and a good initialization stability test can influence algorithm performance substantially. *Fifth*, time series model estimation with eight coefficients model, updated for every single observation, based on all available clear observations achieves the best result. *Sixth*, a change probability of 0.99 (chi-squared distribution) with six consecutive anomaly observations and a mean included angle $< 45^\circ$ to confirm a change provide the best results, and the combined use of temporally-adjusted Root Mean Square Error (RMSE) and minimum RMSE is recommended. *Finally*, spectral changes (or “breaks”) contributed from vegetation regrowth should be excluded from land disturbance maps. The COLD algorithm was developed and calibrated based on all these lessons learned above. The accuracy assessment shows that COLD results were accurate for detecting land disturbance, with an omission error of 27% and a commission error of 28%.

1. Introduction

Land disturbances such as stress, wind, hydrology, debris, harvest, mechanical, and fire, occur on the Earth's surface with high temporal and spatial variability (Cohen et al., 2016; Grimm et al., 2008; Grogan et al., 2015; Hansen et al., 2013; Kennedy et al., 2015; Mildrexler et al., 2009; Senf et al., 2015; Turner, 2010). It is a key component of ecological systems that affects terrestrial ecosystem at a wide range of scales. The spatial and temporal patterns of land disturbances are changing rapidly. For example, large fires are occurring more frequently in many parts of the world (Kasischke and Turetsky, 2006; Marlon et al., 2009; Westerling et al., 2006; Westerling et al., 2011),

even including tundra in Alaska (Hu et al., 2010; Mack et al., 2011; Racine et al., 2004). The hurricane activity in recent years also increased dramatically in the Atlantic (Goldenberg et al., 2001; Saunders and Lea, 2008), with an increasing destructiveness trend (Emanuel, 2005). Beetle infestation has been more severe and extensive than in the past for forests located at higher elevations in western North America, which has changed the forest fragmentation patterns and carbon budgets significantly (Coops et al., 2010; Hicke et al., 2012; Meigs et al., 2015). Mechanical disturbances caused by human activities, such as deforestation, reforestation/afforestation, agricultural expansion/intensification, and urbanization, are increasing at an unprecedented rate, which accounted for 60% of all land changes (Song

* Corresponding author at: Department of Natural Resources and the Environment, University of Connecticut, Storrs, CT 06269, United States.
E-mail address: zhe@uconn.edu (Z. Zhu).

<https://doi.org/10.1016/j.rse.2019.03.009>

Received 3 November 2018; Received in revised form 28 February 2019; Accepted 6 March 2019

Available online 13 March 2019

0034-4257/ © 2019 Elsevier Inc. All rights reserved.

et al., 2018). On the other hand, land disturbances caused by natural disasters threaten human life, property, and can have staggering effects on human well-being and economic costs. Between 1980 and 2004, a total of two million people were killed by around 7000 natural disasters, and the direct economic damage from these natural disasters is estimated around 1 trillion dollars (Stromberg, 2007).

Land disturbance has been defined variously. Here, we define land disturbance as any discrete event that occurs outside the range of natural variability of land surface. Both abrupt changes such as harvest, mechanical, hydrology, fire, wind, and gradual changes (or stress), such as disease, insects, drought are our targeted land disturbance types. Unlike most of the other studies that assume all land disturbances will have an impact of reduced vegetation (Cohen et al., 2018; Healey et al., 2018; Milesi et al., 2003), we do not have this assumption in our definition. This is because some of the land disturbances, such as mechanical changes, wind, hydrology, and fire in non-vegetated areas, do not necessarily reduce vegetation volume. Moreover, regular agricultural practice is not defined as land disturbance, unless it is transformed to some other land use and land cover types or has undergone certain kinds of none repeated agricultural activities, such as changing crop type or lying fallow.

Many algorithms have been developed for monitoring land change based on remotely sensed imagery (Franklin et al., 2015; Hansen and Loveland, 2012; Huang et al., 2018; Zhu and Woodcock, 2014b), but most of them are only focusing on land cover and land use change (target of change), with very few studies on disturbance (agent of change) (Zhu, 2017). Moreover, almost all disturbance studies are focused on forest disturbance (Cohen et al., 2016; Grogan et al., 2015; Hansen et al., 2016; Xin et al., 2013; Zhu et al., 2012), with very limited knowledge on disturbances that occur on non-forested areas. Another major issue is that most of these algorithms do not have the capability of monitoring disturbance in a continuous mode, and the disturbance information derived from these algorithms is less helpful to resource managers and policymakers.

Recently, a Continuous Change Detection and Classification (CCDC; Zhu and Woodcock, 2014b) algorithm has been developed to monitor land cover change using all available Landsat data. The CCDC algorithm can detect a variety of land cover and land use changes continuously with high spatial and temporal accuracies (Pengra et al., 2016; Zhu, 2017), but is reported to have relatively low producer's accuracy for identifying forest disturbance with relatively small change magnitude (Cohen et al., 2017). The CCDC algorithm was originally designed to detect land cover and land use changes that usually result in a large change magnitude. However, for forest disturbances with subtle spectral change (usually still the same forest cover type after the disturbance), such as beetle infestation and selective logging, CCDC is less helpful. On the other hand, many of the changes (or "breaks") detected by CCDC do not necessarily correspond to land disturbances. For example, CCDC identifies a break if it deviates from its predicted statistical intervals, but for breaks caused by gradual condition changes, such as vegetation regrowth, there is no disturbance event associated with them. This kind of error reduces the user's accuracy of land disturbance substantially.

Additionally, though most of the disturbance mapping algorithms were based on Surface Reflectance (SR), there were some studies that used Top-Of-Atmosphere (TOA) reflectance as well (Hansen et al., 2013; Potapov et al., 2012). According to our knowledge, no comprehensive study compares the performance of TOA reflectance and SR time series for mapping land disturbance. Moreover, the United States Geological Survey (USGS) has released the Landsat Analysis Ready Data (ARD) for facilitating time series analysis (Dwyer et al., 2018). The Landsat ARD can provide denser time series observations for "side-lap" regions located between overlapping Landsat scenes than observations from the same Landsat scene. Finally, recent studies demonstrated that the use of multiple spectral bands and indices could provide more accurate forest disturbance results from a supervised classifier (Cohen

et al., 2018). However, there is no comprehensive study on which band/index or selections of bands/indices work best for disturbance detection algorithms that do not rely on training and classification. Therefore, it is of high interest and importance to explore what kind of input data (TOA reflectance vs. SR), which variable or selection of variable, and what temporal frequency (ARD vs. the same scene) would provide the best land disturbance detection results.

In this study, we describe a Continuous monitoring of Land Disturbance (COLD) algorithm that is developed upon the Continuous Change Detection and Classification (CCDC) algorithm (Zhu and Woodcock, 2014b; Zhu et al., 2015b; Zhu et al., 2016), with the objective of providing large-scale, continuous, and accurate detection of land disturbance with Landsat time series. Specifically, we will (i) identify the best time series input for detecting land disturbance; (ii) develop and calibrate the land disturbance detection algorithm based on a well-interpreted reference dataset; and (iii) remove "breaks" that are disturbance irrelevant.

2. Study area and data

2.1. Study area

Our study area includes a total of 180 Thiessen Scene Areas (TSAs) that are defined as the non-overlapping portions of individual frames (Kennedy et al., 2010). These 180 TSAs are selected from 442 TSAs covering the conterminous US based on a stratified sampling strategy designed for selecting a higher proportion of TSAs from forested areas (Fig. 1) (Cohen et al., 2016). Admittedly, our study areas are more forest-oriented, but considering the large amounts of forest disturbance in the US, it is less of an issue. Actually, if we calculate the land cover area proportions of all 180 TSAs based on the 2006 National Land Cover Database (NLCD) cover map (Fry et al., 2011), forest accounts < 40% of the land areas, with other > 60% of land that are covered by agriculture, shrubland, herbaceous, wetland, water, developed, and barren land. Another reason for picking these TSAs as our study area is because a total of 7200 Landsat reference samples across all land types have already been interpreted by experts from United States Forest Service (USFS) to estimate forest disturbance (Cohen et al., 2016). According to our knowledge, this is the only publicly available dataset that provides a high spatial resolution (30-meter) disturbance information at an annual timescale for a relatively large area.

2.2. Reference data

Within each of the 180 TSAs, 40 Landsat plots (30-meter Landsat pixels) were selected based on a simple random strategy, and a sample size of 7200 Landsat plots have been selected as our reference data, in which over half of them are non-forested and have been fully interpreted for cover and change. These datasets were previously interpreted for estimating forest disturbance based on the Landsat time series visualization and disturbance data collection software — TimeSync (Cohen et al., 2010). This tool provides a series of "image chips" (a subset of images) around a Landsat plot of interest (30-meter square), as well as the time series graphical plots, or temporal trajectories to help with interpreting the occurrence of disturbance. Both spectral bands and a variety of indices can be used to view or plot for the time series and the "image chips". For each disturbance event, the causal agent classes such as *Harvest*, *Fire*, *Stress*, *Wind*, *Mechanical*, *Hydrology*, and *Other* (Fig. 2B) were interpreted in TimeSync with help from Google Earth high spatial resolution images and a variety of ancillary data. Cohen et al. (2016) have analyzed the disturbance magnitude for this reference data, in which majority of the disturbances have relative spectral changes less than 50%, and the accumulated tree cover changes from these disturbances are generally lower than 50%.

The Landsat plots were interpreted for each year, and a time range was provided for each observed disturbance. For example, if a

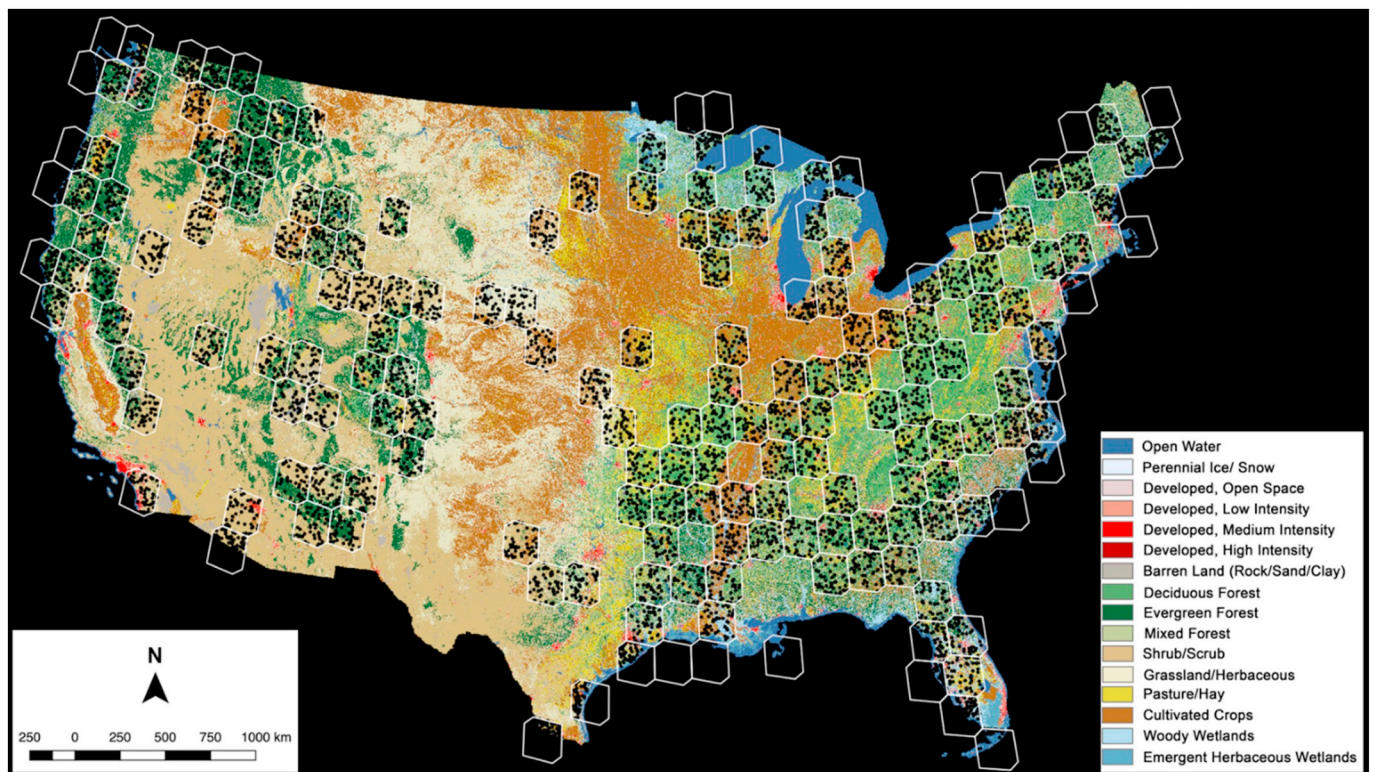


Fig. 1. Study area (white polygons) and the 7200 reference samples (black dots), shown in the context of the 2006 National Land Cover Database (NLCD) cover map.

disturbance is detected in the 2000 growing season image, but not showing up in the previous year, TimeSync will label a disturbance with a time range between 1999 and 2000. In Cohen et al. (2016), only the later year (e.g., 2000) was given as the disturbance year, as it is impossible to know the exact year of disturbance based on the time interval provided in TimeSync and the higher possibility that a disturbance will occur before and during the growing season than after the growing season. However, we found some proportions of disturbances occurred after the growing season (e.g., between July 1999 and December 1999), and if the later year is used as the disturbance year, the labeled disturbance year will be one year later than the actual disturbance year, which will artificially increase both omission and commission errors (a one-year offset was allowed for the comparison in Cohen et al. (2016) to reduce this impact). Therefore, in this study, we do not give a single disturbance year. Instead, we keep the time interval for each disturbance event for training and validation purposes. Most of the disturbances occur between two consecutive years (e.g., fire, wind, harvest, mechanical, and hydrology), and some of the multi-year process disturbances (mostly stress) such as insect infestation and drought can have longer time intervals.

The reference Landsat plots from Cohen et al. (2016) are of high quality. However, we have observed two problems for using them directly for detecting land disturbance. First, though this reference dataset is trying to record all land disturbance types in both forested and non-forested areas, it is less accurate for non-forested areas. This is understandable considering the majority of the interpreters are from USFS. Second, we found this reference dataset is too sensitive to record forest disturbance, as some of the disturbances recorded have no signal in Landsat data (cannot see them in the “image chips” or in the time series plot). Therefore, we have five remote sensing experts looked at all 7200 plots in both TimeSync and Google Earth and corrected these issues in the reference dataset. Basically, we labeled all kinds of land disturbances in the time series, but disturbances that were invisible in Landsat data (time series plot or “image chips”) were changed to the stable category. Moreover, reference pixels do not have enough clear

sky observations (< 24), have large gaps (more than three years without any observations), or difficult to interpret were eliminated from the reference plots. Therefore, a total of 6634 Landsat plots were used as the reference samples for this study.

2.3. Landsat time series from the same scene and Analysis Ready Data (ARD)

Recently, USGS Earth Resources Observation and Science (EROS) Center reorganized Landsat data into a formal tiered data Collection structure, in which Landsat Collection 1 products are categorized into a consistent archive of known data quality, such as Tier 1, Tier 2, and Real-Time. Tier 1 images from Landsats 4–8 and Tier 2 images from Landsat 8 are reported to have high geometry accuracy, and both of them are used to create Landsat ARD for the conterminous US, Alaska, and Hawaii (Dwyer et al., 2018). Landsat ARD are consistently processed to the highest scientific standards and level of processing required for time series analysis, in which compared with Landsat Collection 1 data from the same scene, denser time series observations are expected for places in the adjacent Landsat orbit swath overlap areas (Qiu et al., 2019).

In this study, we compared four kinds of time series observations for monitoring land disturbance, including TOA reflectance Collection 1 data from the same Landsat scene (Scene-TOA), TOA reflectance from ARD (ARD-TOA), SR Collection 1 data from the same Landsat scene (Scene-SR), and SR from ARD (ARD-SR). All available Landsat data were used as the input for both same scene Collection 1 data and ARD data (including images with 100% cloud cover). We have used seven major spectral bands within Landsats 4–8, including the Blue, Green, Red, Near Infrared (NIR), Short-wave Infrared 1 (SWIR1), SWIR2, and Thermal Infrared (TIR) bands. Note that we extracted all Landsat observations directly from Google Earth Engine (Gorelick et al., 2017), and as the USGS Landsat ARD are not included in the current Google Earth dataset, we reprojected the Collection 1 data and created our Landsat ARD based on the same procedure for generating the USGS

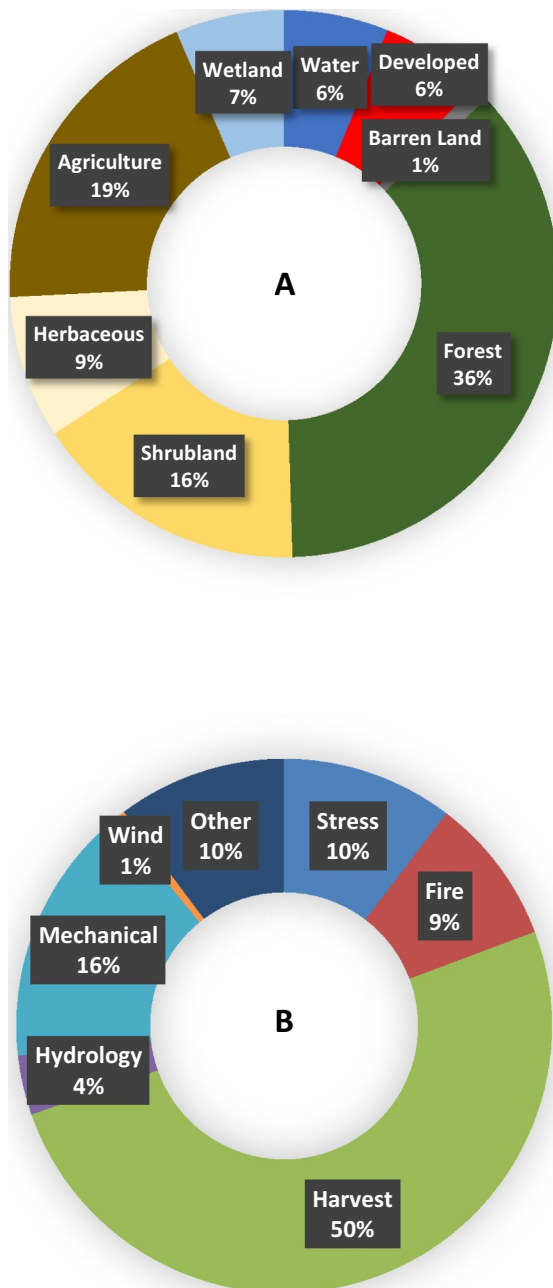


Fig. 2. The area proportions of reference data attribution for the 180 Thiessen Scene Areas. A: Reference data land cover category based on NLCD 2006; B: reference data disturbance type between 1985 and 2011.

Landsat ARD internally in Google Earth Engine (Dwyer et al., 2018).

3. Methods

3.1. The COLD algorithm

The CCDC algorithm works well for detecting land cover and land use changes (Zhu and Woodcock, 2014b), but for land disturbance with small change magnitude, it has large commission errors (Healey et al., 2018). The COLD algorithm is developed based on Zhu and Woodcock (2014b) and Zhu et al. (2015b) but improved and calibrated to better monitor all kinds of land disturbance. It consists of three major processes, including data preparation, model initialization, and continuous monitoring (Fig. 3).

3.1.1. Data preparation

In the data preparation process, there are three main steps, which include Fmask, initializing the model window, and Tmask (blue rectangle in Fig. 3). First, all the available Landsat time series for a location are downloaded, and cloud, cloud shadow, and snow are screened based on the Fmask algorithm (Zhu et al., 2015a; Zhu and Woodcock, 2012) that are provided in the Quality Assessment (QA) band. Initializing the model window in the second step specifies the time window required to initiate a time series model, which should satisfy three criteria including: (i) contain greater than or equal to 12 clear (hereafter clear refers to observations that are not cloud, cloud shadow, or snow) observations; (ii) the total length of the time series should be at least one year; and (iii) the largest data gaps should be less than a year. If all three criteria are met, COLD applies the Tmask algorithm (Zhu and Woodcock, 2014a) to all observations within the initialized model window to exclude cloud, cloud shadow, and snow that are missed in the Fmask algorithm. Note that the COLD algorithm has modified the Tmask algorithm for better monitoring land disturbance (see Section 3.3.2.1 – Tmask modification for details).

3.1.2. Model initialization

The clear observations are digested by the COLD algorithm to build the initial time series model for all spectral bands. Next, COLD tests the stability of the newly initialized models to make sure the model is stable enough to detect change. See Section 3.3.3 for details on how COLD tests the stability of the initialized models. If the model fails to pass the stability test, COLD moves to the next observation by including a new clear observation and removes the earliest clear observation, and this process continues until the model stabilized (see the purple arrow in Fig. 3). Finally, for each newly initialized time series model, it is used as the basis for looking back to the observations ahead of the time series model to modify the start time of the times series model. To determine whether these observations should be included in the current model, COLD uses the same change detection approach used in the continuous monitoring procedure (see Section 3.3.5 for details). COLD searches for all clear observations that are ahead of the current model but are not included in previous and current time series models until it finds a confirmed change, or the search goes to the last observation. For observations that deviate substantially from model prediction but not detected in a consecutive manner, COLD identifies them as non-disturbance related ephemeral changes caused mainly by cloud, cloud shadow, snow or soil moisture changes, and are removed in future time series analysis (left side orange dashed arrow in Fig. 3 and see Section 3.3.2.2 – Non-consecutive Outlier Removal for details). In this way, COLD is able to initialize a time series model that includes all clear observations collected earlier than the start time of the model and follow the predicted trajectory of the current time series model. If there is a change confirmed ahead of the current model and the current model is initialized for the first time for a pixel, COLD estimates a new model based on the observations before the change, and this new model, as well as the confirmed change, is recorded in the time segments database as described in Table 1. Otherwise, COLD will use the current time series model to monitor land disturbance which is described in the next section.

3.1.3. Continuous monitoring

In the continuous monitoring process, there are five major steps including updating model, looking forward, detecting change, recording time segments, and extracting disturbance. First, all the newly collected clear observations will be used to update the current time series model for better predicting future observations. The time series model contains three sets of harmonic components with different temporal frequencies, which are triggered by the number of available clear observations (Eq. (1)). Basically, the more clear observations, the higher frequencies of harmonic components will be included (Zhu et al., 2015b). For clear observations between 12 and 17 ($k = 1$), the four

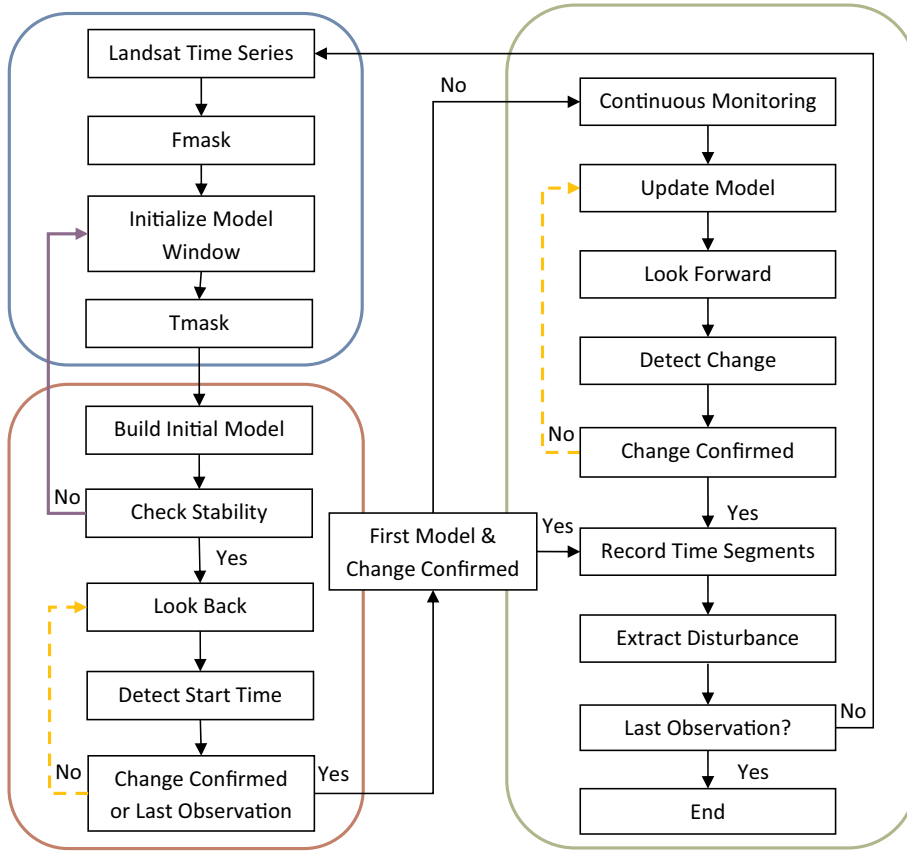


Fig. 3. The process chain of the COLDF algorithm. The blue, red, and green rectangles represent the data preparation, model initialization, and continuous monitoring process respectively. The dashed orange arrows represent the process of include or exclude single new observation for updated models. The purple arrow presents the process of moving to the next observation by including a new clear observation and removes the earliest clear observation in the initialized model window. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

coefficients model is used. For clear observations between 18 and 23, the six coefficients model is used ($k = 2$), and eight coefficients model is used for models with greater than or equal to 24 clear observations ($k = 3$). The reason for picking a maximum of eight coefficients is explained in Section 3.3.4.1 – Number of coefficients. Next, COLDF looks forward to the new clear observations collected after the updated model. For each newly collected clear observation, COLDF compares the model predicted value (from the updated model) with the actual clear observation to calculate the magnitude of spectral change. If the change magnitude is within the expected range, the next observation is considered stable and is used to update the time series model iteratively. In Section 3.3.4.2 – Update frequency, we will discuss the details on the optimal way for model updating. Similarly, for observations that deviate substantially from model prediction but not detected in a consecutive manner (for both magnitude and direction), COLDF identifies them as non-disturbance related ephemeral changes, and they are removed in future time series analysis (right side orange dashed arrow in

Fig. 3). Otherwise, if the change is observed consecutively, it is confirmed (see Section 3.3.5 for details). In the fourth step, all the spectral and temporal information created for each time segment is recorded in a data structure format. A single pixel with N “breaks”, will have $N + 1$ time segments, and for each time segment, there are a total of 77 variables recorded in different fields as described in Table 1. Finally, we can extract land disturbances from all the previously detected “breaks” and progress to the last clear observation (see Section 3.3.6 for details).

$$\hat{\rho}_{i,x} = a_{0,i} + \sum_{k=1}^3 \left\{ a_{k,i} \cos\left(\frac{2\pi}{T}x\right) + b_{k,i} \sin\left(\frac{2\pi}{T}x\right) \right\} + c_{1,i}x \quad (1)$$

where,

- x : Julian date
- i : The i th Landsat Band ($i = 1, 2, 3, 4, 5, \text{ and } 7$)
- k : Temporal frequency of harmonic component ($k = 1, 2, \text{ and } 3$)
- T : Number of days per year ($T = 365.25$)

Table 1
Variables recorded in each COLDF time segment.

Variable name	Size (rows × cols)	Descriptions
t_{start}	1	Date when the time series model starts.
t_{end}	1	Date when the time series model ends.
t_{break}	1	Date when the first “break” is identified
$coefs$	56 (8×7)	Coefficients for each time series model (8 coefficients) for each spectral band (7 bands)
$rmse$	7 (7×1)	Root mean square error from the model fit for each spectral band (7 bands)
pos	1	Position of each time series model (x/y location in the stacked images)
$change_prob$	1	Change probability (0–1). Change probability equals to the number of consecutive anomaly observations divided by N (N : number of consecutive anomaly observations to confirm a change)
num_obs	1	Number of clear observations used for model estimation
$category$	1	Quality of the model estimation, where the tens digit specifies what model fit procedure is used (e.g., 5: a perennial snow pixel, 4: Fmask failed, 3: a reserved value, 2: model fits at the end, 1: model fits at the beginning, and 0: normal procedure for model fit) and the unit digit specifies what model is used (e.g., 4, 6, and 8: models with 4, 6, and 8 coefficients respectively).
$magnitude$	7 (7×1)	Magnitude of spectral band differences between model prediction and observation for each spectral band (7 bands).

- a_0, i : Coefficient for overall value for the i th Landsat Band
 $a_{k, i}, b_{k, i}$: Coefficients for intra-annual change for the i th Landsat Band
 c_1, i : Coefficient for inter-annual change (slope) for the i th Landsat Band
 $\hat{\rho}_{i,x}$: Surface reflectance for the i th Landsat Band at x Julian date from model prediction.

3.2. Algorithm performance evaluation metrics

Every disturbance is identified as an isolated event. For disturbances detected by COLD that are within the stable period of the reference samples, they are identified as commission errors. For disturbance segments in the reference plot, if no disturbance is identified by COLD, they are identified as omission errors. Note that the agreement between COLD detection and the reference samples means they are correct not only in space but also in time. For example, if a disturbance is detected at the time after (or before) the reference disturbance segment, both omission and commission cases will increase by one in the confusion matrix. The period between 1985 and 2011 (27 years) was used for this evaluation. Considering there are almost seven thousand Landsat plots, we have a total of approximately 189,000 evaluating cases (for each plot in each calendar year).

We randomly selected 50% of the reference samples (3317 Landsat plots) for algorithm development and calibration. The remaining 50% of the reference samples were used for validating of the COLD algorithm. Three accuracy metrics, including omission error, commission error, and F1 score, were selected for accuracy assessment. Similar as Healey et al. (2018), the commission error and omission error were selected to evaluate the chances that stable land observations were falsely identified as land disturbance observations, and land disturbance observations were falsely identified stable land observations respectively (Eqs. (2)–(3)). Moreover, as we expected to have unbiased mapping results, algorithms with balanced omission and commission error are preferred (Healey et al., 2018; Zhu et al., 2012). Therefore, we selected the F1 score (or F-measure) as the third accuracy metric (Eq. (4)), as it also provides a measure of balance between precision (commission error) and recall (omission error). For each algorithm performance evaluation, we will calculate the omission error, commission error, and F1 score for five different change probability thresholds (e.g., 0.90, 0.95, 0.99, 0.999, 0.9999). For each threshold, the locations of markers are based on their omission and commission errors in the omission-commission plane, and the colors of the markers represent the values of F1 score. After connecting the markers from the five change detection thresholds in the omission-commission plane, the line that is closest to the origin and with the highest F1 score is considered to have the best performances.

$$\text{commission} = \frac{\text{detected disturbances disagree with reference}}{\text{total number of detected disturbances}} \times 100\% \quad (2)$$

$$\text{omission} = \frac{\text{reference disturbances disagree with detection}}{\text{total number of reference disturbances}} \times 100\% \quad (3)$$

$$\begin{aligned} \text{F1 score} &= \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 200\% \\ &= \frac{(1 - \text{commission}) \times (1 - \text{omission})}{2 - \text{commission} - \text{omission}} \times 200\% \end{aligned} \quad (4)$$

3.3. COLD algorithm development and calibration

We explored six major components for better detecting land disturbance, including input data, outlier removal, model initialization, model fit, change detection, and disturbance extraction (Fig. 4). Methods and thresholds with the best performance for monitoring land disturbance were selected for the COLD algorithm. We have also

provided a full comparison between COLD and CCDC in all these aspects in Table 2. Note that though CCDC has multiple versions released at different places online, there is only one paper that well-explained the algorithm in the literature (Zhu and Woodcock, 2014b). Therefore, we mainly compared COLD with this version of CCDC.

3.3.1. Input data

We explored three science questions regarding what is the best input data for monitoring land disturbance: 1) ARD or same scene data; 2) TOA or SR, and; 3) which variable or selections of variables? To compensate the inconsistency of temporal density caused using ARD data, we developed a new approach that adjusts the change probability and the number of consecutive anomaly observations required to confirm a change based on the overall density of the time series (hereafter, we name it the adjusted approach). See Section 1.1 in Supplementary materials for details.

3.3.1.1. ARD vs. scene & TOA reflectance vs. SR. We observed that if no adjustment was made for the inconsistency of temporal density, detection accuracy from SR was substantially higher than that from TOA reflectance (both ARD and same scene images). Another important discovery is that the use of ARD (for both TOA reflectance and SR) would find more changes than using images from the same scene, which leads to lower omission error but higher commission error. This can be troublesome for monitoring land disturbance using Landsat ARD, as the difference in temporal density between overlap and non-overlap areas would present inconsistent spatial patterns in the disturbance maps.

On the other hand, if the adjusted approach is applied to reduce the impact of the inconsistency of temporal density in the input data, SR is still better than TOA reflectance. However, ARD (including SR and TOA reflectance) with the adjusted approach demonstrate substantially better performance than ARD without adjustment, in which the adjusted approach achieved the similar omission error but much lower commission error (~10%). On the other hand, the adjusted approach achieved almost the same detection accuracy as the approach without any adjustment for data from the same scene. This result suggests that the adjustment will not artificially change the disturbance detection results for the non-overlap area, and for overlap areas, the adjusted approach can detect land disturbance with even higher accuracies. This also demonstrates the importance of using dense time series, as the higher the temporal density, the better the detection of land disturbance. Considering the omission and commission errors were reduced at a similar level, the spatial pattern of land disturbance between overlap and non-overlap should be very similar. Therefore, we will use Landsat ARD SR as the input data type, and the adjusted approach will be employed at the same time.

3.3.1.2. Variable selection. A recent study suggested that the combination of bands and indices works much better in detecting forest disturbance than just using a single variable (Cohen et al., 2018). Therefore, we explored many different combinations of spectral bands and indices to detect land disturbance of all possible kinds.

For scenarios with a single spectral band, the two SWIR bands showed similar (SWIR2 is slightly better) but much better performance than the other spectral bands if used alone. The Red band was ranked the third, but its performance was much lower than the two SWIR bands. The other spectral bands were almost useless if they are used alone, and the TIR, NIR, and Blue bands are the three variables with the worst performance. Another observation is that the use of a single index, such as Normalized Difference Vegetation Index (NDVI), Normalized Burn Ratio (NBR), and Tasseled Cap Wetness was not as helpful as expected. For example, the indices that heavily involved with SWIR bands, such as NBR and Wetness, did not perform as well as just using a single SWIR band, and the widely used NDVI index performed similarly to the Red band.

For scenarios with two or more spectral variables, the combined use



Fig. 4. Overview of the six components for algorithm development and calibration. ARD: Analysis Ready Data; TOA Ref: Top-Of-Atmosphere Reflectance; SR: Surface Reflectance; CCDC: Continuous Change Detection and Classification; COLD: Continuous monitoring of Land Disturbance; LASSO: Least Absolute Shrinkage and Selection Operator; RMSE: Root Mean Square Error.

Table 2

Comparisons of CCDC and COLD in the six major components explored for better monitoring land disturbance. ARD: Analysis Ready Data; SR: Surface Reflectance; CCDC: Continuous Change Detection and Classification; COLD: Continuous monitoring of Land Disturbance; OLS: Ordinary Least Square; LASSO: Least Absolute Shrinkage and Selection Operator; RMSE: Root Mean Square Error.

	CCDC (Zhu and Woodcock, 2014b)	COLD
Input data	Blue, Green, Red, NIR, SWIR1, SWIR2, and TIR SR; Landsat images from the same scene	Green, Red, NIR, SWIR1, and SWIR2 SR; Landsat ARD with the adjusted approach
Outlier removal	0.04 TOA reflectance (Tmask); non-consecutive change larger than the empirical change threshold	Dynamic threshold (Tmask); non-consecutive change probability larger than 99.999%
Model initialization	CCDC approach	COLD approach
Model fit	4 coefficients; updated for every new observation; observations from all available stable years; OLS fit	8 coefficients; updated for every single observation; observations from all available stable years; LASSO fit with a lambda of 20
Change detection	3 consecutive anomaly observations; empirical change threshold normalized by RMSE	6 consecutive anomaly observations; a change probability of 99% normalized by temporally-adjusted RMSE and minimum RMSE; mean included angle < 45°
Disturbance extraction	No	Yes

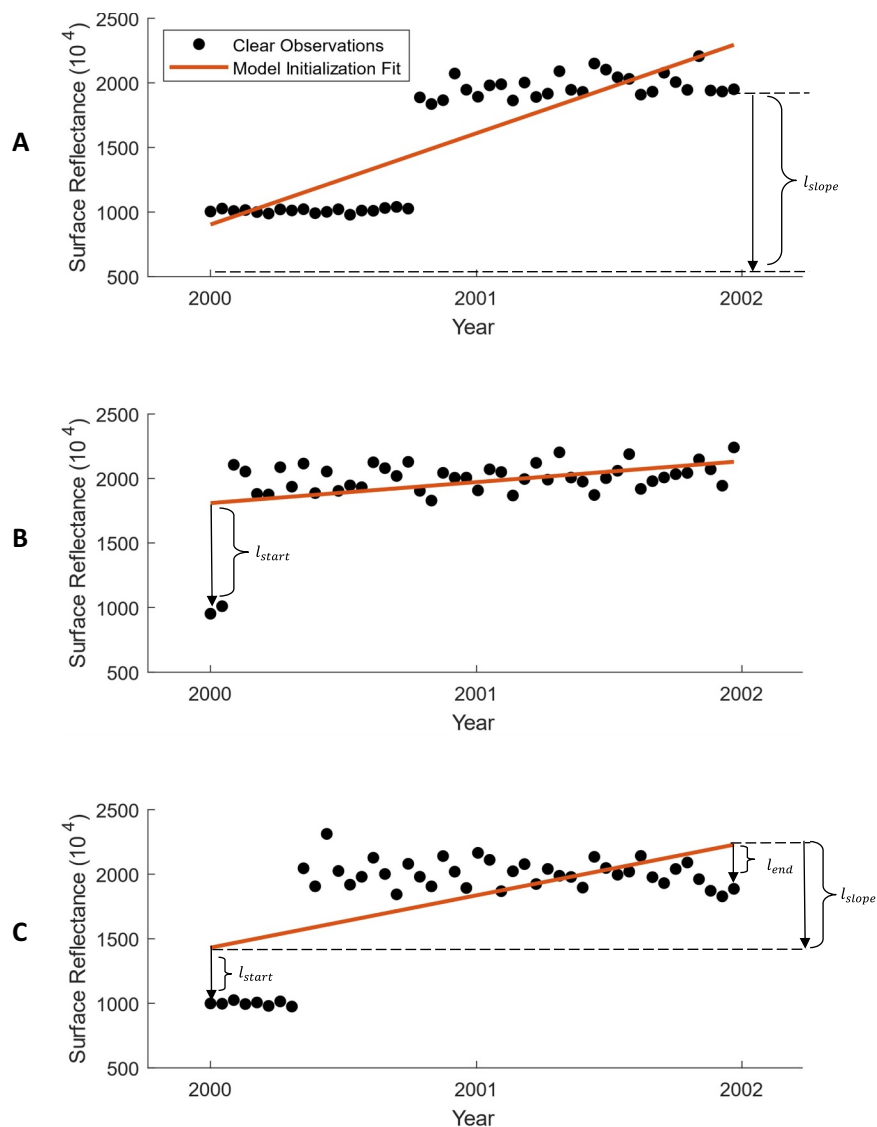


Fig. 5. Three possible model initialization fits. A: Change signal influenced model initialization fit significantly; B: change signal (occurred at the start of the model) has little influence on model initialization fit, and; C: change signal has some degree of influence on model initialization fit.

of the Red and NIR bands achieved similar performance as the use of an NDVI index, and the combined use of the NIR and SWIR1 bands or the combined use of the Red and SWIR1 bands achieved similar performances as the use of a single SWIR1 band. However, the combined use of the Red, NIR, and SWIR1 bands achieved slightly better performance than the use of a single SWIR1 band, and the combined use of the Green, Red, NIR, SWIR1, and SWIR2 bands achieved best results (used for the COLD algorithm), in which omission error was substantially reduced. We think the reduction of omission error is mainly contributed to the use of all the informative spectral bands, as for a different kind of land disturbance, the direction and magnitude of change can be quite different. Moreover, we also explored the scenarios by including other spectral bands and indices in addition to the five bands used by COLD, and the results suggested that the inclusion of the Blue, TIR, Blue and TIR bands or other indices did not offer better performance. We think this is mainly because 1) the Blue and TIR bands signals are more influenced by atmosphere than land disturbance (Zhu et al., 2015b); and 2) if all the useful spectral bands are already used to detect land disturbance, the signal in the vegetation indices will be redundant.

3.3.2. Outlier removal

Cloud, cloud shadow, and snow were first screened by the Fmask

(Function of mask) algorithm for all available Landsat time series (Foga et al., 2017; Zhu et al., 2018; Zhu et al., 2015a; Zhu and Woodcock, 2012). Considering Fmask is not perfect and any outlier in the time series can be detrimental, we also applied the Tmask (multiTemporal mask) algorithm (Zhu and Woodcock, 2014a) to further screen cloud, cloud shadow, snow, and other non-disturbance related outliers. As the Tmask algorithm may also exclude some ephemeral changes, we only applied the Tmask algorithm during the model initialization period. After the model was initialized, observations that did not show as anomaly consecutively were also removed, as land disturbance are more consistent in time than outliers.

3.3.2.1. Tmask modification. In Tmask algorithm, a fixed threshold of 0.04 in Green and SWIR1 bands (TOA reflectance) was used to detect any cloud, cloud shadow, and snow in the time series. This threshold was chosen based on visual assessment of the cloud, cloud shadow, and snow mask created for Landsat images located in one New England scene, but it has not been tested and calibrated for other parts of the world and other kinds of landscapes. Moreover, for places with large natural variations (e.g., urban and agricultural areas), a fixed threshold will likely remove many clear observations, leaving very few observations for model fit. Therefore, we proposed a dynamic

thresholding approach for Tmask based on a specific type of variogram called lag-1 madogram (see Section 2.1 in Supplementary materials for details).

We explored the use of a variety of dynamic thresholds based on a series of probabilities (e.g., 0.99, 0.999, 0.9999, 0.99999 and 0.999999). The dynamic thresholding approach showed better performance than the method with fixed thresholds, and a change probability of 0.99999 (normal distribution) achieved the best performance (see Section 2.2 in Supplementary materials for details).

3.3.3.2. Non-consecutive outlier removal. During change detection, we can calculate change probability for every newly collected clear observation (see Eq. (8) for details). Outliers would show up with large change probability but are more ephemeral in time (e.g., show up once and disappear in the next observation). If the next a few consecutive observations do not indicate a change, but the first observation in these consecutive observations is larger than the change threshold, we will exclude this observation if it has a relatively high change probability. We explored a series of change probability (e.g., 0.99, 0.999, 0.9999, 0.99999, and 0.999999) for removing this kind of non-consecutive outliers. The best result was achieved with a change probability of 0.99999 (chi-squared distribution). Note that if the change probability is < 0.99999 , this observation will be used by COLD to update model fit (the dashed orange arrows in Fig. 3).

3.3.3. Model initialization

3.3.3.1. CCDC vs. COLD initialization. Changes that occur within the time of model initialization can influence model prediction and impact change detection afterward. Therefore, the CCDC algorithm designed a stability test by examining the magnitude of the model slope and the differences in model prediction and observation for the first and last observations (Zhu and Woodcock, 2014b). For example, if a change occurs during the time of model initialization, the slope will respond to the change, and we can use the change magnitude from the slope (l_{slope} in Fig. 5A) to determine whether it is a stable model. Alternatively, if a change occurs at the start (or the end) of the model initialization, they may not have enough changed observations to influence the magnitude of the slope, but they will show a large difference between model predictions and actual observations (l_{start} in Fig. 5B). Therefore, in CCDC, the model will pass the stability test if l_{slope} , l_{start} , and l_{end} are all smaller than the change probability. However, the CCDC initialization approach ignores one possibility of model fit that may fail the stability test. This occurs when the change time is not located at the very start (or end) of the time series or in the middle of the time series. For example, if it occurs in the middle of 2001 (Fig. 5C), the model slope is partially influenced by the observations that have changed. In this case, l_{slope} , l_{start} , and l_{end} are likely to be all smaller than the change probability. To fix this issue, we modified the model initialization based on Eq. (5), where the COLD model stability test is based on the combined absolute value from l_{slope} and the larger absolute values between l_{start} and l_{end} .

$$\text{COLD Stability Test} \quad |l_{slope}| + \max\{|l_{start}|, |l_{end}|\} < \text{Change Probability} \quad (5)$$

We compared algorithm performance by using CCDC model initialization and the new approach we designed for COLD. Both approaches achieved a similar omission rate when change probability is larger than 0.90. However, the COLD initialization approach reduced commission rate by almost 10% with a change probability of 0.99. This demonstrates the importance of model initialization for change detection.

3.3.4. Model fit

3.3.4.1. Number of coefficients. Harmonic (Fourier) models have been widely used for modeling satellite time series (Zhu, 2017). Previous studies suggested that harmonic models with three harmonic terms (annual, bimodal, and trimodal changes) work best for satellite time series (Verbesselt et al., 2012; Zhu et al., 2015b), but there is no evidence on this choice of harmonic terms. For example, the CCDC approach suggests the use of one harmonic term (annual change) can already provide high change detection and classification results. Therefore, we explored the use of time series models with a maximum of 4, 6, and 8 coefficients (Eq. (1)). The models with a maximum of 6 coefficients achieved better results than with 4 coefficients, but models with a maximum of 8 coefficients showed the best result. The increase of model coefficients has greatly reduced omission error and maintained a similar commission error. We think the reduced omission error is mostly contributed from the better modeling of land surface, and the use of more coefficients did not lead to large overfit, which is mainly controlled by the Least Absolute Shrinkage and Selection Operator (LASSO) fit (see Section 3.3.4.4 – LASSO fit). Therefore, we chose the time series model with a maximum of 8 coefficients. Note that this conclusion is solely based on LASSO fit and it may be different if another regression approach such as ordinary least squares regression is used.

3.3.4.2. Update frequency. To provide a more accurate model estimation, all newly collected clear observations were digested iteratively to update the time series model. However, updating the model too frequently can cost a huge amount of computing time, and the benefit can be limited compared with less frequent updating. Therefore, we explored seven different updating scenarios, including updating for a fixed time (e.g., observations accumulated for one, two, or three years) and for a fixed proportion of previous time (e.g., observations accumulated 1/5, 1/4, 1/3 or 1/2 of time range of previous model fit), and compared them with the scenario of updating for every single observation. If there are 20 clear observations per year, updating for every observation is 20 times more computationally expensive than the single year updating approach, and the two to three years update time interval will further reduce the computational time linearly. If we update model based on 1/N of time length of previous model fit window (e.g., if the previous fit takes T days for model fit, the model needs extra T/N days for the next round of model fit), the update frequency will be reduced exponentially, and can be very computational efficient when the time series model spans a long period. Based on the algorithm performance comparisons, updating model for every single observation achieved the highest detection accuracy. When the updating frequency changed from 1 to 3 years, accuracy dropped substantially. On the other hand, if the model is updated based on 1/N of the time range of previous model fit, the smaller the N, the better the performance, but none of them showed better results than it is updated for every single observation. Therefore, the COLD algorithm updates model for every single observation.

3.3.4.3. Moving window. Theoretically, the model fit will be more accurate with more clear observations available, and therefore the CCDC algorithm used all the available historical stable pixels in model fit. However, all algorithms make mistakes, and it is possible that some subtle changes are not detected by the algorithm and the changed pixels will be included in the model fit. This may cause problems in model fit, as a single time series model may not fit for all these observations. One way to reduce this impact is to limit the maximum window size for model estimating and move the fitting window along with the new observations (moving window approach). In this way, the impact of missed changes will fade away when the window moves forward. Note that if the available stable observations are less than the length of the

required moving window, we will use what we have for model fit. We explored this moving window approach in model fit by changing the window size from two years to all available year (for the stable period only). It turned out the use of three years moving window can already achieve good accuracies, but with the number of years increases, the accuracy also increases, and the all available stable years showed the best results. Therefore, we used all the available stable years for model fit.

3.3.4.4. Least Absolute Shrinkage and Selection Operator (LASSO) fit. Though the inclusion of more harmonic terms will generally provide a more accurate estimation of the temporal trajectory of the land surface, this can also cause overfitting because of the extra freedom in the model. The LASSO regression approach (Tibshirani, 2011) is known for reducing the overfitting by minimizing the residual sum of squared errors and the sum of the absolute values of the coefficients (Eq. (6)). One of the most critical variables in the LASSO fit is the penalty parameter (λ), in which the higher the value, the more penalty for overfitting. If λ is set to zero, LASSO fit will be the same as the OLS fit. We explored the LASSO fit with a variety of λ values (from 0 to 100) for 8 coefficient time series models, and the LASSO fit achieved better detection results than the OLS fit did, and the best detection result was achieved when λ equals to 20 (we selected this value for COLD).

$$\min_{\beta \in R^p} \left\{ \sum_{i=1}^n (Y_i - X_i\beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (6)$$

3.3.5. Change detection

3.3.5.1. Number of consecutive anomaly observations. Satellite time series contains rich information but is also inherently impacted by noise from a variety of sources (Cohen et al., 2017). Fortunately, unlike noise that appears randomly throughout the time series, land change signal is usually more consistent. Therefore, the majority of the false positive change contributed to noise will be eliminated if we use a few consecutive anomaly observations to confirm a change. This approach has been widely used in recent change detection studies. For example, the CCDC algorithm confirms a change based on 3 consecutive clear observations (Zhu and Woodcock, 2014b); Hansen et al. (2016) requires 4 consecutive clear observations to alert forest loss; and

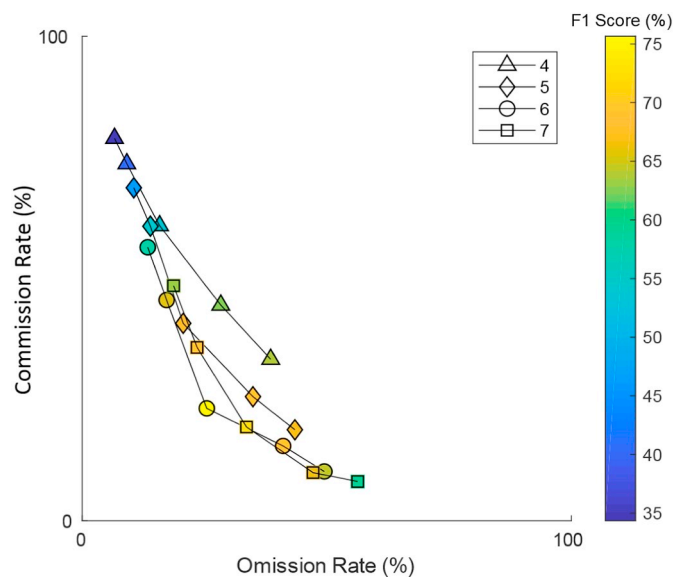


Fig. 6. Change detection with different number (e.g., 4, 5, 6, and 7) of consecutive anomaly observations to confirm a change.

Brooks et al. (2014) detects forest disturbance by the year in which at least four images signal a change. Most of these algorithms are only tested for relatively small areas, and there are no experiments on the optimum number of consecutive anomaly observations needed to confirm a change. Therefore, we explored the different algorithm performances by using four to seven consecutive anomaly observations to confirm a change (Fig. 6). We found that the detection accuracy increased when the consecutive anomaly observations increased from four to six and decreased when consecutive anomaly observations increased to seven. Therefore, we selected six consecutive anomaly observations to confirm a change. Though the increase of the number of consecutive anomaly observations will remove more false positive changes, some short-term real surface changes may also disappear which increases the omission error. It is worth noting that for specific disturbance types such as grassland fires and floods if they present on the land surface for less than six consecutive anomaly observations, the COLD algorithm will not be able to label them as change. Moreover, for places with heavy cloud cover, the COLD algorithm will wait a much longer time to collect a total of six clear observations to confirm a change.

3.3.5.2. Change probability. The threshold used for identifying land change is the most critical variable for change detection. In the CCDC algorithm, an empirical threshold of three times of Root Mean Square Error (RMSE) is used to detect land cover change (Eq. (7)), and in a later version, it is lower to two times of RMSE to capture more subtle changes (Zhu et al., 2015b). In Eq. (7), the RMSE normalized differences between model predictions and observations are averaged for all Landsat spectral bands. This can be problematic for detecting land disturbances that only show change in one of the spectral bands, as the difference found in this band will be averaged for all spectral bands. Therefore, in the COLD algorithm, a normalized change vector magnitude is used as the threshold for change detection (Eq. (8)), and by using the sum of squared deviates, change in any one of the spectral bands will show large deviations even it is averaged with other bands. Considering each of the normalized change deviates follows the standard normal distribution, the sum of the square of the deviates follows the chi-squared distribution, with a degree of freedom equal to the number of spectral bands used in change detection. Therefore, we can create change thresholds based on change probability derived from the chi-squared distribution, in which a change probability of 0.99 with five spectral band works best.

$$\frac{1}{k} \sum_{i=1}^k \frac{|\rho_i - \hat{\rho}_i|}{3 \times RMSE_i} > 1 \quad (7)$$

$$\sum_{i=1}^k \left(\frac{\rho_i - \hat{\rho}_i}{RMSE_i} \right)^2 \sim \chi^2(k) > \chi_{0.99}^2(5) \quad (8)$$

where,

- i : the i th Landsat band
- k : Number of Landsat bands used for change detection
- ρ_i : Observed value for the i th Landsat band
- $\hat{\rho}_i$: Predicted value for the i th Landsat band.

3.3.6. Change angle

Unlike environmental and systematic noise that is completely random, the disturbance signals are more consistent not only in terms of absolute change magnitude but also change vector angle. It is possible that noise or other non-disturbance signals may exist in the time series for a few consecutive times; it is unlikely that all of them are changing in the same direction. Therefore, we can use the mean included angles between each pair of neighboring change vectors (after normalization) to confirm a real land disturbance (Fig. 7 and Eq. (9)). The optimal result was achieved with a threshold of 45° of mean included angles. In

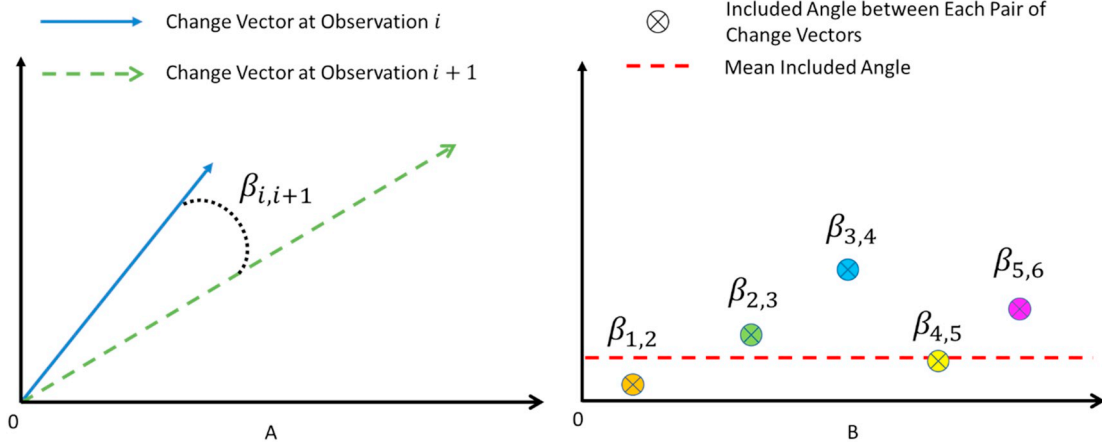


Fig. 7. A: An included angle between a pair of neighboring change vectors. B: Mean included angle from five pairs of neighboring change vectors.

short, COLD confirms a change if the mean included angle between pairs of neighboring change vector smaller than 45° .

$$\frac{1}{k} \sum_{i=1}^{k-1} \beta_{i,i+1} < 45^\circ \quad (9)$$

where,

- i : the i th consecutive anomaly observation
- k : Number of consecutive anomaly observations to confirm a change.

3.3.6.1. Root Mean Square Error (RMSE). RMSE from each spectral band in model fit is used to normalize the change vector. Therefore, the larger the unexplained variations, the less the weight in contributing to the final change probability. This approach works well for many of the land cover types, as the unexplained variations are usually the same throughout the years. However, for some land cover types such as agriculture, semi-arid shrublands, and grassland, there is considerable year-to-year variability in the time of green-up that is related to human activities and interannual climate variability. Therefore, for a certain time of year, the variation of the data can be much larger than other time of year, making a fixed RMSE less ideal for this kind of situation.

Similar to Zhu et al. (2015b), we proposed to use the temporally-adjusted RMSE as the denominator for normalizing the change vector. The temporally-adjusted RMSE uses the nearest (day of the year) 24 observations to calculate the RMSE when the total number of clear observations is > 24 . The other important component for RMSE calculation is that for spectrally dark pixels or models with a limited number of clear observations, sometimes RMSE can be extremely small (see Fig. 21 in Zhu and Woodcock, 2014b). Considering the change vector magnitude is normalized by RMSE, a minor change in spectral bands can cause a large difference in the final change probabilities. Therefore, it is important to provide a predefined minimum value of RMSE, and if the calculated RMSE value is less than this value, the minimum value will be used (see Section 5.2 in Supplementary materials for details).

We explored four different scenarios of using RMSE for normalizing change vector magnitude. The scenarios of using RMSE and RMSE with minimum RMSE showed similar performances. The use of temporally-adjusted RMSE showed better performance, and the combined use of both temporally-adjusted RMSE and minimum RMSE achieved the best results (used in COLD).

3.3.7. Disturbance extraction

The change detection process identifies “breaks” contributed from spectral changes that deviate from model prediction, in which some of them may not related to land disturbance, such as breaks from

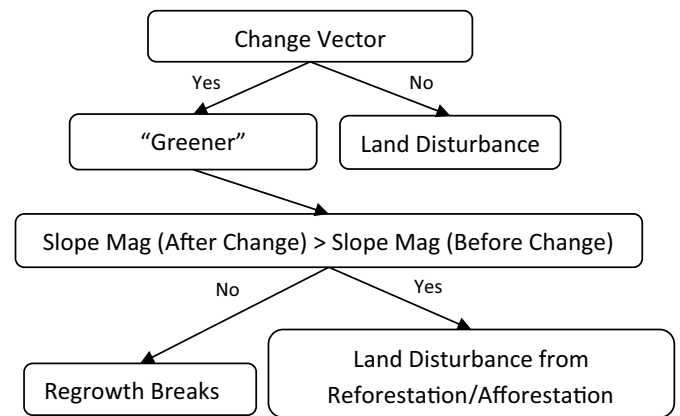


Fig. 8. Extraction of land disturbance based on change vector direction and model slopes.

vegetation regrowth (e.g., the break between the young forest and mature forest). To exclude regrowth breaks from land disturbance, we developed a generic-rule-based land disturbance extraction approach as illustrated in Fig. 8. Unlike most of the other disturbance algorithms that extract disturbance based on a threshold of a certain vegetation index (e.g., NDVI, NBR, or Tasseled Cap Wetness), we assume all previously identified “breaks” are real land disturbances, and only remove “breaks” from the vegetation regrowth. In this way, land disturbances that are not related to gain or loss in vegetation amount will also be included in the results. We first used the change vector from the three important spectral bands (Red, NIR, and SWIR1) to separate all

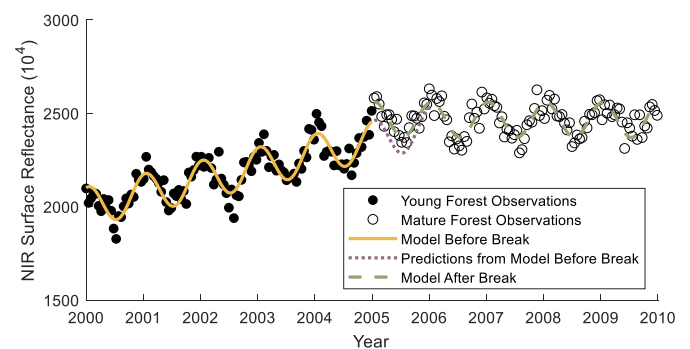


Fig. 9. The “greener” direction in NIR band for a regrowth break. Note that the mature forest NIR observation values are higher than the predicted values from the model before the break.

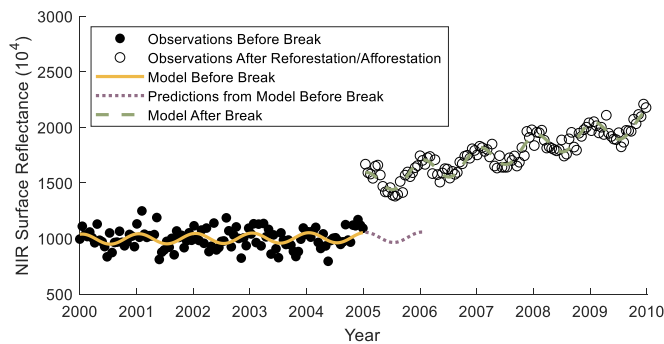


Fig. 10. The slope magnitude in NIR band for a reforestation/afforestation break. Note that the NIR slope magnitude for the model after the break is positive and larger than the slope magnitude for the model before the break.

“breaks” into two categories: “greener” direction breaks (higher NIR band, and lower Red and SWIR1 bands after the “break”; Eq. (10)) and land disturbance. Unlike land disturbance that can have change vector in all directions, the regrowth breaks will only show in the “greener” direction (Fig. 9). All breaks that are not in this “greener” direction will be labeled as land disturbances. This would identify almost all kinds of land disturbances, except for reforestation and afforestation, which also share the same “greener” change vector direction with regrowth breaks. Fortunately, the slopes of time series models before and after the change are quite different between the regrowth breaks and the breaks from reforestation and afforestation. For example, for breaks from vegetation regrowth, the magnitude of the slope before the breaks are usually larger than it after the break and it is the opposite for breaks from reforestation and afforestation. Moreover, as the slope directions for time series models after the reforestation or afforestation are already determined (positive in NIR band and negative in Red and SWIR1 bands), we can create a few rules based on the magnitude and direction of slopes to extract reforestation and afforestation breaks out from the regrowth breaks (Eq. (11)) (Fig. 10).

“Greener” direction:

$$\Delta Red < - Threshold \ \& \ \Delta NIR > Threshold \ \& \ \Delta SWIR1 < - Threshold \tag{10}$$

Slope magnitudes before (S_{Bef}) and after change (S_{Aft}):

$$S_{Red, Aft} < - |S_{Red, Bef}| \ \& \ S_{NIR, Aft} > |S_{NIR, Bef}| \ \& \ S_{SWIR1, Aft} < - |S_{SWIR1, Bef}| \tag{11}$$

3.3.7.1. Thresholds for disturbance extraction. We defined the “greener” direction is where NIR increased, but Red/SWIR decreased. However, how much increase or decrease is considered as the “greener” direction has not been assessed. Therefore, we evaluated algorithm performances by changing the threshold for defining the “greener” direction from -0.04 to 0.04 and included the scenario without using the disturbance extraction process (identify all breaks as land disturbance). The use of disturbance extraction achieved much better results than without using it, and a threshold of -0.02 provided the best results (used by COLD).

4. Results

We evaluated the COLD disturbance detection results both quantitatively and qualitatively. The 50% unseen reference Landsat plots (not used for algorithm development and calibration) were used to validate the COLD land disturbance detection accuracy, and accuracy metrics such as producer's accuracy (or omission error), user's accuracy (or commission error), and F1 score were used for this assessment. We also applied the COLD algorithm to an area that covers the Berkeley-Oakland

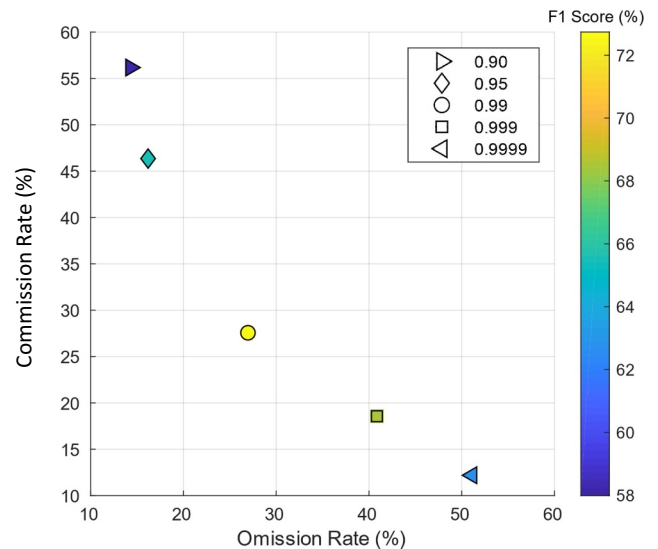


Fig. 11. Omission and commission error when change probability threshold changes from 0.90 to 0.9999.

area in California to examine the detection results qualitatively. We choose this urban area mainly because it has experienced a variety of land disturbances (e.g., earthquakes, fire, house development, and transition of urban land use), and they are well-documented.

4.1. Quantitative evaluation of detection accuracy

We tested the COLD algorithm against the 3317 reference Landsat plots based on a series of change probabilities (e.g., 0.90, 0.95, 0.99, 0.999, and 0.9999) (Fig. 11). Optimum detection accuracy was achieved when change probability equals to 0.99. COLD has reached a producer's accuracy of 73% (27% of omission error), a user's accuracy of 72% (28% of commission error), and an F1 score of 73%. It is worth noting that this accuracy is achieved based on reference samples with a large number of small magnitude disturbances and a variety of land disturbances events that have not been used in algorithm development and calibration. We also evaluated COLD algorithm performance against different land disturbance types. COLD performed well for all seven disturbance types, in which most of the *Harvest* (75%), *Mechanical* (74%), *Wind* (74%), *Fire* (76%), and *Other* (72%) disturbances were correctly identified, and more than half of the *Stress* (60%) and *Hydrology* (63%) disturbances were correctly detected.

4.2. Qualitative evaluation of land disturbance maps in the Berkeley-Oakland area

The Berkeley-Oakland area has a quite complex landscape, where common land cover and land use categories include urban, water, and various kinds of vegetation types (Fig. 12E). This urban and suburban area has been frequently disturbed by a variety of different events that include the 1989 Loma Prieta earthquake (Fig. 12A), the Oakland firestorm of 1991 (Fig. 12B), the decommission of Oakland Army Base (Fig. 12C), and Wilder housing projects (Fig. 12D).

To demonstrate the algorithm's capability of monitoring land disturbance in a continuous mode, we created the accumulated land disturbance maps during the time of each previously mentioned land disturbance event (Fig. 13). The first demonstration is from the Loma Prieta earthquake. The 1989 Loma Prieta earthquake occurred on October 17 in the San Francisco Bay Area heavily damaged the Cypress Viaduct, leading to the instant death of 41 people. The COLD algorithm detected this land disturbance event accurately in a very short time, and we can see a precise shape of the collapsed viaduct in the December 10,

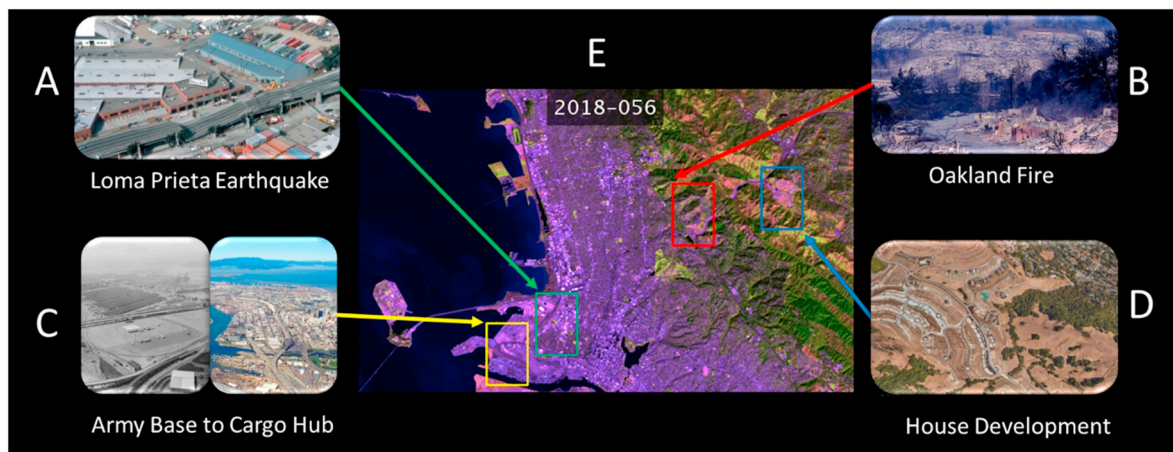


Fig. 12. Historical land disturbance detected by Landsat time series for the Berkeley-Oakland area. A: Aerial view of collapsed sections of the Cypress Viaduct of Interstate 880 (photo adapted from H.G. Wilshire, U.S. Geological Survey, 1989). B: The 1991 Oakland-Berkeley firestorm which devastated 1500 acres in the hills (photo adapted from California Governor's Office of Emergency Services, 1991). C: (Left) View of Oakland Army Base, Maritime Street at West Grand Avenue, Oakland, Alameda County, CA (photo adapted from Ordenana, Adrian, 1994. Institution: Library of Congress Prints and Photographs Division Washington, D.C.). (Right) Aerial view of the port of Oakland, California, USA (photo adapted from Robert Campbell, U.S. Army Corps of Engineers Digital Visual Library). D: Wilder luxury home community in Orinda, California (photo adapted from Google Maps). E: False color composite Landsat image (SWIR-NIR-Red bands) for this area, captured on the 56th day of 2018. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

1989, accumulated disturbance map (Fig. 13A). The second demonstration is from the Oakland firestorm. In the weekend of October 19–20, 1991, Oakland experienced a severe fire over the hills in northern Oakland and southern Berkeley. This firestorm was the worst fire in East Bay history. It ultimately killed 25 people, injured 150 others, and a large number of dwellings and apartments were destroyed. This event was also accurately captured by the COLD algorithm in the November 14, 1991, accumulated land disturbance map (Fig. 13B). The third demonstration is from the urban land use transition from an army base to cargo hubs. The Oakland Army Base served from 1944 to 1999 on the Oakland waterfront, south to the eastern entrance to the San Francisco-Oakland Bay, and later this region started its gateway and port development. The COLD algorithm also detected this land use related disturbance in a timely fashion, and we can see an apparent patch in the yellow box of the June 13, 1999, accumulated land disturbance map (Fig. 13C). The fourth demonstration is from wilder housing development. There were several house development projects in the western of Orinda, California, started approximately in 2007. These homes in the community were constructed to serve people working in downtown San Francisco but living away from crowded cities. The housing development related land disturbance was also well captured by the COLD algorithm, and we can see the spatial pattern in the blue box of the February 14, 2008, accumulated disturbance map (Fig. 13D). The COLD algorithm monitored this area continuously until February 25, 2018, and the final accumulated disturbance map is illustrated in Fig. 13E. Many small scale land disturbances have occurred in this busy urban area, with a majority of them related to urban modifications and intensifications.

5. Discussions and conclusion

Continuous monitoring of land disturbance is difficult in remote sensing. The availability of dense Landsat time series provides the possibility. In this study, we developed a new algorithm for continuous monitoring of land disturbance at high spatial (30-meter) and temporal (up to every four days) resolutions. By using all available Landsat ARD, this approach creates time series model for pixels during the stable periods, and by comparing model predictions with Landsat observations, “breaks” (or spectral changes) are identified if they show change for a few consecutive times. The COLD algorithm flags all the “breaks” as potential land disturbances, and later excludes the non-disturbance

“breaks”. The reference data revealed that the COLD results were accurate for monitoring land disturbance, with both omission error and commission error < 30%. The COLD algorithm is programmed in the MATLAB environment and is provided online as an open-source package that includes both source code and standalone software.

We fully explored and verified various kinds of input data and time series techniques for better detection of land disturbance. Recommendations in six major aspects were provided, including input data, outlier removal, model initialization, model fit, change detection, and disturbance extraction. The best input data for land disturbance are Landsat ARD SR. Moreover, the combined use of spectral bands is better than using a single spectral band or index, and the integrated use of five bands including Green, Red, NIR, SWIR1, and SWIR2 provided the best result. For outlier removal, a probability threshold of 0.99999 (based on normal distribution or the chi-squared distribution) provided the best result. Moreover, model initialization is also essential for change detection, and the COLD initialization approach is recommended. Furthermore, a good model fit provides a better prediction of observations, which will also benefit the detection of land disturbance. We recommend time series models with a maximum of eight coefficients, updated for every single observation, based on all available stable years, and using LASSO regression ($\lambda = 20$). For change detection, the use of a change probability of 0.99 (chi-squared distribution), a total of six consecutive anomaly observations, and a mean included angle < 45° to confirm a change provides the best result, and the combined use of temporally-adjusted RMSE and minimum RMSE is also recommended. Finally, we developed a new generic-rule-based approach to exclude non-disturbance related “breaks”, which substantially reduced the commission error. It is worth noting that though many of the time series techniques can only provide limited improvements, the impact of combined use of all proposed techniques can be substantial.

The COLD algorithm has many advantages. First, unlike many other disturbance algorithms that rely heavily on the availability of training data or inputs from several complicated algorithms (Cohen et al., 2018; Healey et al., 2018), COLD detects land disturbance fully automated with Landsat time series as the only input data. Though we used some reference samples for algorithm development and calibration, most of the rules and techniques are based on ecological and biophysical processes that can be generalized for large areas and different kinds of environments. With some adjustments (e.g., region-based thresholds), the COLD algorithm should have the capability of providing land

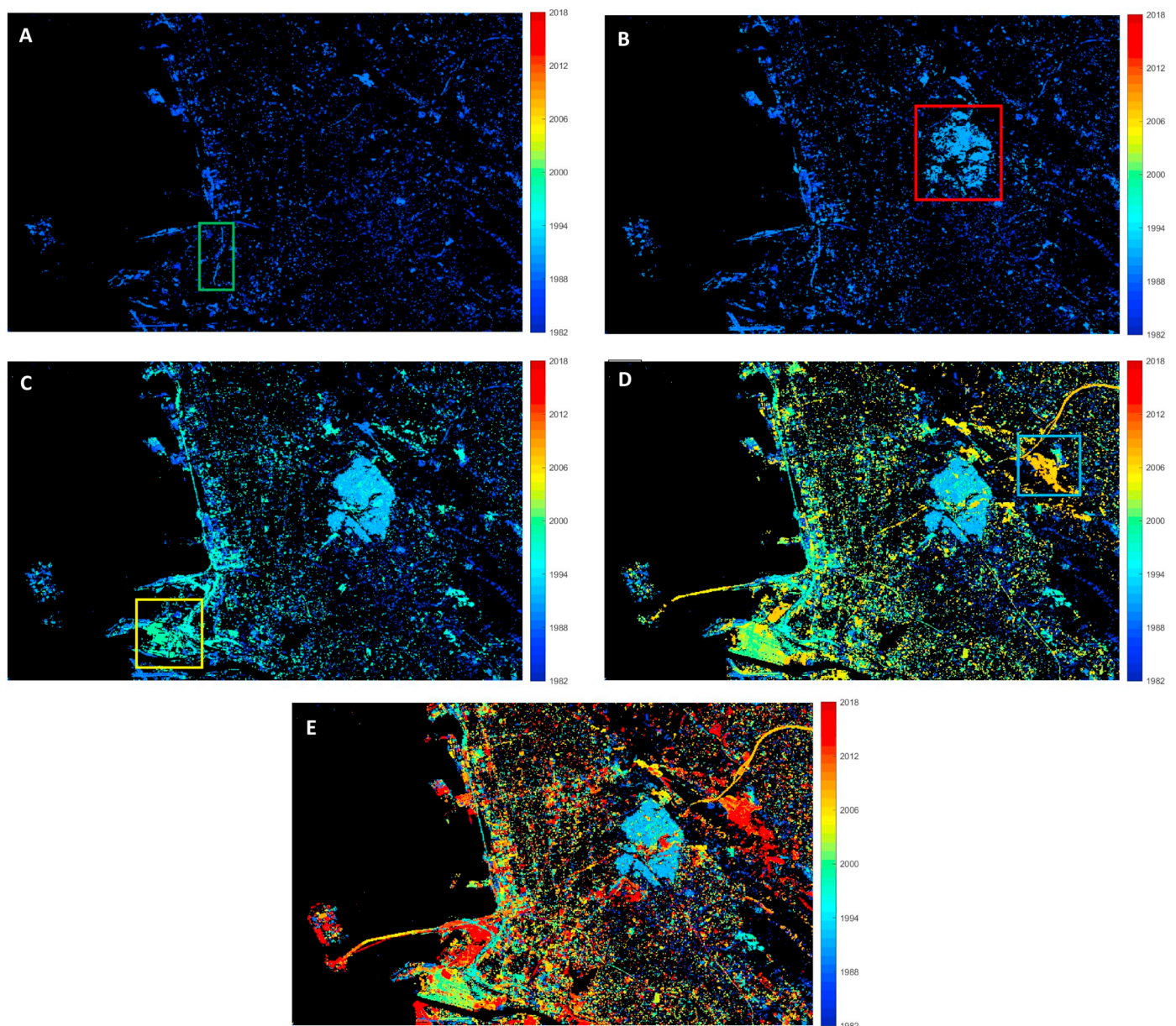


Fig. 13. Accumulated land disturbance map during the time of four major land disturbance events, with colors denoting the disturbance time. A: Accumulated land disturbance map from 1982 to December 10, 1989. Within the green box, the collapse of the Cypress Street Viaduct is captured along Interstate 880. B: Accumulated land disturbance map from 1982 to November 14, 1991. Within the red box, a large patch of land disturbance caused by the Oakland fire is also well captured. C: Accumulated disturbance map from 1982 to June 13, 1999. Within the yellow box, the change of the Oakland Army Base to a cargo hub is captured in the Port of Oakland. D: Accumulated disturbance map from 1982 to February 14, 2008. Within the blue box, the land disturbances from the housing development are well detected. E: Accumulated land disturbance maps from 1982 to Feb 25, 2018. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

disturbance maps at continental or even global scales. It is worth noting that though COLD can create disturbance maps without any training data, for optimal results, the use of training data to calibrate the algorithm parameters for each specific location is highly recommended. Second, by digesting newly collected Landsat ARD continuously, the COLD algorithm can provide land disturbance maps as soon as the land surface is being disturbed. Currently, we only used Landsat ARD, which provides new images in less than eight days (four days in overlapped areas) when both Landsat 7 and 8 are used. This approach can be easily adapted to include Sentinel-2 images, and by combining Landsat 7 and 8, and two Sentinel 2A and 2B, there would be typically ten observations per month (new observations every three days), making near real-time monitoring of land disturbance at high spatial resolution possible. Third, the COLD algorithm is the first disturbance mapping algorithm

that includes many kinds of land disturbance types occurred on all land surface, including disturbances that are not occurring in vegetated areas. This timely generic disturbance information can be pivotal for many different purposes, such as disaster response, urban planning, resources management, etc.

The COLD algorithm also has limitations. First, it is computationally expensive and requires a large amount of data storage. The time required to process one Landsat ARD tile (1–3 thousand images) would still cost 1000–3000 computing hours, and it takes around 0.5–2 Terabytes to store Landsat time series in one ARD tile. Second, the COLD algorithm works well for places that are relatively stable, but for places that are frequently disturbed (e.g., agriculture and urban areas), the COLD algorithm struggles to find a stable period to initialize time series models and may miss those disturbance events. Finally, the

current COLD algorithm can only identify disturbance event but cannot provide more detailed information on disturbance type. Recently, the USGS and USFS are working together to collect independent reference samples for validation and characterization of annual land cover products for the Land Change Monitoring, Assessment, and Projection (LCMAP) project (Pengra et al., 2016; Zhu et al., 2016), in which a total of 25,000 Landsat plots will be collected. This dataset provides both land cover and disturbance information for every pixel in every calendar year, and as it is created based on simple random sampling, it is not biased to forests. We can validate (or re-calibrate) the COLD algorithm and develop methods to classify different kinds of disturbance agent when these reference samples are available.

In conclusion, we developed a new algorithm called COLD for continuous monitoring of land disturbance. It can provide accurate land disturbance maps fully automated at high spatial (30-meter) and high temporal (up to four days) resolutions for large areas.

Acknowledgments

We gratefully acknowledge the support of the USGS-NASA Landsat Science Team (LST) Program for “Toward Near Real-time Monitoring and Characterization of Land Surface Change for the Conterminous US” (#G17PS00256), and USGS Great Plains Cooperative Ecosystem Studies Unit (CESU) Program for “Algorithm Improvement, Near Real-time Monitoring, and New Change Product Designed for the LCMAP Initiative” to Zhe Zhu (#G17AC00057). We also gratefully acknowledge the support of the NASA’s Terrestrial Ecology Program and Carbon Cycle and Ecosystems Focus Area to Warren Cohen (#NNH11AR291).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2019.03.009>.

References

- 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 Transactions on Geoscience and Remote Sensing* 52 (6), 3316–3332. <https://doi.org/10.1109/TGRS.2013.2272545>.
- Cohen, W.B., 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 Sensing of Environment* 114 (12), 2911–2924. <https://doi.org/10.1016/j.rse.2010.07.010>.
- Cohen, W.B., Yang, Z., Stehman, S.V., Schroeder, T.A., Bell, D.M., Masek, J.G., ... Meigs, G.W., 2016. Forest disturbance across the conterminous United States from 1985–2012: the emerging dominance of forest decline. *Forest Ecology and Management* 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., ... Zhu, Z., 2017. How similar are forest disturbance maps derived from different landsat time series algorithms? *Forests*. <https://doi.org/10.3390/f8040098>.
- Cohen, W.B., Yang, Z., Healey, S.P., Kennedy, R.E., Gorelick, N., 2018. A LandTrendr multispectral ensemble for forest disturbance detection. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2017.11.015>.
- Coops, N.C., Gillanders, S.N., Wulder, M.A., Gergel, S.E., Nelson, T., Goodwin, N.R., 2010. Assessing changes in forest fragmentation following infestation using time series Landsat imagery. *Forest Ecology and Management* 259 (12), 2355–2365. <https://doi.org/10.1016/j.foreco.2010.03.008>.
- Dwyer, J.L., Roy, D.P., Sauer, B., Jenkerson, C.B., Zhang, H.K., Lymburner, L., 2018. Analysis ready data: enabling analysis of the landsat archive. *Remote Sensing*. <https://doi.org/10.3390/rs10091363>.
- Emanuel, K., 2005. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*. <https://doi.org/10.1038/nature03906>.
- Foga, S., Scaramuzza, P.L., Guo, S., Zhu, Z., Dilley, R.D., Beckmann, T., ... Laue, B., 2017. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2017.03.026>.
- Franklin, S.E., Ahmed, O.S., Wulder, M.A., White, J.C., Hermosilla, T., Coops, N.C., 2015. Large area mapping of annual land cover dynamics using multitemporal change detection and classification of Landsat time series data. *Canadian Journal of Remote Sensing* 41 (4), 293–314. <https://doi.org/10.1080/07038992.2015.1089401>.
- Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C.G., Limin, Y., ... Wickham, J., 2011. Completion of the 2006 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* 77 (9), 858–864.
- Goldenberg, S.B., Landsea, C.W., Mestas-Núñez, A.M., Gray, W.M., 2001. The recent increase in Atlantic hurricane activity: causes and implications. *Science*. <https://doi.org/10.1126/science.1060040>.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2017.06.031>.
- Grimm, N.B., Faeth, S.H., Golubiewski, N.E., Redman, C.L., Wu, J., Bai, X., Briggs, J.M., 2008. Global change and the ecology of cities. *Science*. <https://doi.org/10.1126/science.1150195>.
- Grogan, K., Pflugmacher, D., Hostert, P., Kennedy, R., Fensholt, R., 2015. Cross-border forest disturbance and the role of natural rubber in mainland Southeast Asia using annual Landsat time series. *Remote Sensing of Environment* 169, 438–453. <https://doi.org/10.1016/j.rse.2015.03.001>.
- Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment* 122 (September), 66–74. <https://doi.org/10.1016/j.rse.2011.08.024>.
- Hansen, M.C.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A. a, Tyukavina, A., ... Townshend, J.R.G.R.G., 2013a. High-resolution global maps of. *Science* 342 (November), 850–854. <https://doi.org/10.1126/science.1244693>.
- Hansen, M.C., Krylov, A., Tyukavina, A., Potapov, P.V., Turubanova, S., Zutta, B., ... Moore, R., 2016. Humid tropical forest disturbance alerts using Landsat data. *Environmental Research Letters* 11 (3), 034008. <https://doi.org/10.1088/1748-9326/11/3/034008>.
- Healey, S.P., Cohen, W.B., Yang, Z., Kenneth Brewer, C., Brooks, E.B., Gorelick, N., ... Zhu, Z., 2018. Mapping forest change using stacked generalization: an ensemble approach. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2017.09.029>.
- Hicke, J.A., Allen, C.D., Desai, A.R., Dietze, M.C., Hall, R.J., Hogg, E.H.T., ... Vogelmann, J., 2012. Effects of biotic disturbances on forest carbon cycling in the United States and Canada. *Global Change Biology*. <https://doi.org/10.1111/j.1365-2486.2011.02543.x>.
- Hu, F.S., Higuera, P.E., Walsh, J.E., Chapman, W.L., Duffy, P.A., Brubaker, L.B., Chipman, M.L., 2010. Tundra burning in Alaska: linkages to climatic change and sea ice retreat. *Journal of Geophysical Research: Biogeosciences*. <https://doi.org/10.1029/2009JG001270>.
- Huang, B., Zhao, B., Song, Y., 2018. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2018.04.050>.
- Kasischke, E.S., Turetsky, M.R., 2006. Recent changes in the fire regime across the North American boreal region - spatial and temporal patterns of burning across Canada and Alaska. *Geophysical Research Letters* 33 (9). <https://doi.org/10.1029/2006GL025677>.
- Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr - temporal segmentation algorithms. *Remote Sensing of Environment* 114 (12), 2897–2910. <https://doi.org/10.1016/j.rse.2010.07.008>.
- Kennedy, R.E., Yang, Z., Braaten, J., Copass, C., Antonova, N., Jordan, C., Nelson, P., 2015. Attribution of disturbance change agent from Landsat time-series in support of habitat monitoring in the Puget Sound region. USA. *Remote Sensing of Environment* 166, 271–285. <https://doi.org/10.1016/j.rse.2015.05.005>.
- Mack, M.C., Bret-Harte, M.S., Hollingsworth, T.N., Jandt, R.R., Schuur, E.A.G., Shaver, G.R., Verbyla, D.L., 2011. Carbon loss from an unprecedented Arctic tundra wildfire. *Nature*. <https://doi.org/10.1038/nature10283>.
- Marlon, J.R., Bartlein, P.J., Walsh, M.K., Harrison, S.P., Brown, K.J., Edwards, M.E., ... Whitlock, C., 2009. Wildfire responses to abrupt climate change in North America. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0808212106>.
- Meigs, G.W., Kennedy, R.E., Gray, A.N., Gregory, M.J., 2015. Spatiotemporal dynamics of recent mountain pine beetle and western spruce budworm outbreaks across the Pacific Northwest Region. USA. *Forest Ecology and Management* 339 (1), 71–86. <https://doi.org/10.1016/j.foreco.2014.11.030>.
- Mildrexler, D.J., Zhao, M., Running, S.W., 2009. Testing a MODIS global disturbance index across North America. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2009.05.016>.
- Milesi, C., Elvidge, C.D., Nemani, R.R., Running, S.W., 2003. Assessing the impact of urban land development on net primary productivity in the southeastern United States. *Remote Sensing of Environment*. [https://doi.org/10.1016/S0034-4257\(03\)00081-6](https://doi.org/10.1016/S0034-4257(03)00081-6).
- Pengra, B., Gallant, A., Zhu, Z., Dahal, D., 2016. Evaluation of the initial thematic output from a continuous change-detection algorithm for use in automated operational land-change mapping by the U. S. Geological Survey. *Remote Sensing* 8 (10), 811. <https://doi.org/10.3390/rs8100811>.
- Potapov, P.V., Turubanova, S.A., Hansen, M.C., Adusei, B., Broich, M., Altstatt, A., ... Justice, C.O., 2012. Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM+ data. *Remote Sensing of Environment* 122, 106–116. <https://doi.org/10.1016/j.rse.2011.08.027>.
- Qiu, S., Lin, Y., Shang, R., Zhang, J., Ma, L., Zhu, Z., ... Zhu, Z., 2019. Making Landsat time series consistent: evaluating and improving Landsat analysis ready data. *Remote Sensing* 11 (1), 51. <https://doi.org/10.3390/rs11010051>.
- Racine, C., Jandt, R., Meyers, C., Dennis, J., 2004. Tundra fire and vegetation change along a hillslope on the Seward Peninsula, Alaska, U.S.A. *Arctic, Antarctic, and Alpine Research*. [https://doi.org/10.1657/1523-0430.\(2004\)036\[0001:TFAVCA\]2.0.CO;2](https://doi.org/10.1657/1523-0430.(2004)036[0001:TFAVCA]2.0.CO;2).
- Saunders, M.A., Lea, A.S., 2008. Large contribution of sea surface warming to recent increase in Atlantic hurricane activity. *Nature*. <https://doi.org/10.1038/nature06422>.
- Senf, C., Pflugmacher, D., Wulder, M.A., Hostert, P., 2015. Characterizing spectral-temporal patterns of defoliator and bark beetle disturbances using Landsat time series.

- Remote Sensing of Environment 170, 166–177. <https://doi.org/10.1016/j.rse.2015.09.019>.
- Song, X.-P., Hansen, M.C., Stehman, S.V., Potapov, P.V., Tyukavina, A., Vermote, E.F., Townshend, J.R., 2018. Global land change from 1982 to 2016. *Nature*. <https://doi.org/10.1038/s41586-018-0411-9>.
- Stromberg, D., 2007. Natural disasters, economic development, and humanitarian aid. *Journal of Economic Perspectives*. <https://doi.org/10.1257/jep.21.3.199>.
- Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>.
- Turner, M.G., 2010. Disturbance and landscape dynamics in a changing world. *Ecology*. <https://doi.org/10.1890/10-0097.1>.
- Verbesselt, J., Zeileis, A., Herold, M., 2012. Near real-time disturbance detection using satellite image time series. *Remote Sensing of Environment* 123 (Turner 2010), 98–108. <https://doi.org/10.1016/j.rse.2012.02.022>.
- Westerling, A.L., Hidalgo, H.G., Cayan, D.R., Swetnam, T.W., 2006. Warming and earlier spring increase Western U.S. forest wildfire activity. *Science*. <https://doi.org/10.1126/science.1128834>.
- Westerling, A.L., Turner, M.G., Smithwick, E.A.H., Romme, W.H., Ryan, M.G., 2011. Continued warming could transform Greater Yellowstone fire regimes by mid-21st century. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1110199108>.
- Xin, Q., Olofsson, P., Zhu, Z., Tan, B., Woodcock, C.E., 2013. Toward near real-time monitoring of forest disturbance by fusion of MODIS and Landsat data. *Remote Sensing of Environment* 135, 234–247. <https://doi.org/10.1016/j.rse.2013.04.002>.
- Zhu, Z., 2017. Change detection using Landsat time series: a review of frequencies, pre-processing, algorithms, and applications. *ISPRS Journal of Photogrammetry and Remote Sensing* 130, 370–384. <https://doi.org/10.1016/j.isprsjprs.2017.06.013>.
- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment* 118, 83–94. <https://doi.org/10.1016/j.rse.2011.10.028>.
- Zhu, Z., Woodcock, C.E., 2014a. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: an algorithm designed specifically for monitoring land cover change. *Remote Sensing of Environment* 152, 217–234. <https://doi.org/10.1016/j.rse.2014.06.012>.
- Zhu, Z., Woodcock, C.E., 2014b. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment* 144, 152–171. <https://doi.org/10.1016/j.rse.2014.01.011>.
- Zhu, Z., Woodcock, C.E., Olofsson, P., 2012. Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment* 122, 75–91. <https://doi.org/10.1016/j.rse.2011.10.030>.
- Zhu, Z., Wang, S., Woodcock, C.E., 2015a. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment* 159, 269–277. <https://doi.org/10.1016/j.rse.2014.12.014>.
- Zhu, Z., Woodcock, C.E., Holden, C., Yang, Z., 2015b. Generating synthetic Landsat images based on all available Landsat data: predicting Landsat surface reflectance at any given time. *Remote Sensing of Environment* 162, 67–83. <https://doi.org/10.1016/j.rse.2015.02.009>.
- Zhu, Z., Gallant, A.L., Woodcock, C.E., Pengra, B., Olofsson, P., Loveland, T.R., ... Auch, R.F., 2016. Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. *ISPRS Journal of Photogrammetry and Remote Sensing* 122, 206–221. <https://doi.org/10.1016/j.isprsjprs.2016.11.004>.
- Zhu, Z., Qiu, S., He, B., Deng, C., 2018. Cloud and cloud shadow detection for Landsat images: the fundamental basis for analyzing Landsat time series. In: *Remote Sensing Time Series Image Processing*. CRC Press, pp. 25–46.