

A LandTrendr multispectral ensemble for forest disturbance detection

Warren B. Cohen^{a,*}, Zhiqiang Yang^b, Sean P. Healey^c, Robert E. Kennedy^d, Noel Gorelick^e

^a USDA Forest Service, PNW Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331, United States

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

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

^d College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331, United States

^e Google Switzerland GmbH, Zurich, CH 8002, Switzerland.



ARTICLE INFO

Keywords:

Forest disturbance
Landsat time series
Multispectral change detection
Disturbance signal-to-noise ratio
Stacking ensemble

ABSTRACT

Monitoring and classifying forest disturbance using Landsat time series has improved greatly over the past decade, with many new algorithms taking advantage of the high-quality, cost free data in the archive. Much of the innovation has been focused on use of sophisticated workflows that consist of a logical sequence of processes and rules, multiple statistical functions, and parameter sets that must be calibrated to accurately classify disturbance. For many algorithms, calibration has been local to areas of interest and the algorithm's classification performance has been good under those circumstances. When applied elsewhere, however, algorithm performance has suffered. An alternative strategy for calibration may be to use the locally tested parameter values in conjunction with a statistical approach (e.g., Random Forests; RF) to align algorithm classification with a reference disturbance dataset, a process we call secondary classification. We tested that strategy here using RF with LandTrendr, an algorithm that runs on one spectral band or index. Disturbance detection using secondary classification was spectral band- or index-dependent, with each spectral dimension providing some unique detections and different error rates. Using secondary classification, we tested whether an integrated multispectral LandTrendr ensemble, with various combinations of the six basic Landsat reflectance bands and seven common spectral indices, improves algorithm performance. Results indicated a substantial reduction in errors relative to secondary classification based on single bands/indices, revealing the importance of a multispectral approach to forest disturbance detection. To explain the importance of specific bands and spectral indices in the multispectral ensemble, we developed a disturbance signal-to-noise metric that clearly highlighted the value of shortwave-infrared reflectance, especially when paired with near-infrared reflectance.

1. Introduction

Employing Landsat time series for the characterization and mapping of forest disturbance has received considerable attention over the past decade (Hansen et al., 2013; Kim et al., 2014), since the opening of the image archive when data became freely available in a highly-calibrated format (Roy et al., 2014). Many newer algorithms have used all cloud-free observations, either directly or after data reduction to derive annual composites, before subjecting the time series of data to various sophisticated algorithm functions designed to detect disturbances (Hermosilla et al., 2015). For example, Brooks et al. (2014) identified abrupt disturbances with all available data using residuals from harmonic regression and statistical quality control charts, DeVries et al. (2015) used harmonic regression with a breakpoint seeking method called “moving sums”, Kennedy et al. (2010) subjected annual

composite time series to temporal segmentation with the goal of mapping both abrupt and gradual change, and Huang et al. (2010) highlighted spectral anomalies in moving multi-year windows to characterize disturbances. Prior to this new era of freely available, well-calibrated data, most applications of Landsat time series to map forest disturbance were limited to less dense time series (Cohen et al., 2002; Masek et al., 2008). These applications commonly relied on traditional statistical methods, such as post-classification map comparison, bi-temporal differencing, principal components analysis, and supervised classification (Coppin et al., 2004; Healey et al., 2005).

Denser time series data and more sophisticated approaches facilitate detection of subtler disturbance signals, which has led to a move away from an almost exclusive characterization of stand replacement disturbances (Healey et al., 2008; Wulder et al., 2004) towards the exploration of partial (i.e., non-stand replacement) disturbances

* Corresponding author.

E-mail addresses: wcohen@fs.fed.us (W.B. Cohen), zhiqiang.yang@oregonstate.edu (Z. Yang), seanhealey@fs.fed.us (S.P. Healey), rkennedy@coas.oregonstate.edu (R.E. Kennedy), gorelick@google.com (N. Gorelick).

<https://doi.org/10.1016/j.rse.2017.11.015>

Received 6 July 2017; Received in revised form 6 November 2017; Accepted 16 November 2017

Available online 24 November 2017

0034-4257/ Published by Elsevier Inc.

associated with forest thinning, degradation, and insect and disease activity that unfolds over multiple years (Meigs et al., 2011; Meddens and Hicke, 2014; Cohen et al., 2016; Hughes et al., 2017). Exploring subtler signals within time series data has an attendant risk of false detection of change associated with noise, as indicated by time series studies from other disciplines (Trenberth, 1984; Pohmann et al., 2016). However, as forest management policy has shifted away from stand replacement harvests towards maintenance of healthy forest systems (Moeur et al., 2011), international agreements on forest monitoring have begun to include forest degradation along with deforestation (Kissinger et al., 2012), and recognition that climate change is making forests more vulnerable to mortality associated with increasing physiological stress (Allen et al., 2015; Mildrexler et al., 2016), there are now greater demands on remote sensing to provide a full range of detection capabilities from subtle to dramatic forest disturbance (McDowell et al., 2015).

When attempting to detect low magnitude disturbances with Landsat time series, the signal associated with spectral change due to disturbance may be masked by noise associated with normal temporal variation from imperfect atmospheric and geometric corrections, vegetation phenology, sun angle variations, and sensor degradation. In this regard, Kennedy et al. (2010) found that different spectral indices had varying abilities for accurate detection of subtler disturbance signals in western Oregon, with the normalized difference vegetation index (NDVI) performing less effectively than the normalized burn ratio (NBR) or Tasseled Cap Wetness (TCW). Because many of the newer algorithms employ a limited set of spectral bands or indices to detect disturbance, such as the NBR (Kennedy et al., 2012), Forestness Index (Huang et al., 2010), NDVI and SWIR-NIR (shortwave-infrared, near-infrared) ratio (Vogelmann et al., 2012), or Tasseled Cap Angle (TCA, Brooks et al., 2014), careful consideration of the comparative signal-to-noise (SNR) strengths among spectral indices is important. To address this need, we formalized the derivation of a disturbance SNR (DSNR) metric and used that to test the effectiveness of the six primary Landsat reflectance bands (i.e., TM/ETM + bands 1–5, 7) and a host of common spectral indices for detecting forest disturbance across a wide variety of forest types in the US.

Application of any given algorithm or approach for detecting disturbance requires one or more thresholds and calibration steps to separate disturbance signal from temporal noise. These are usually derived using statistical procedures, but also involve a great degree of heuristics. For example, Huang et al. (2010), Kennedy et al. (2010), Brooks et al. (2014), and Hughes et al. (2017) all describe the complexity of their unique Landsat-based forest disturbance detection algorithms, the in-depth rigorous steps involved in calibration for local conditions, and the hands-on assessments and related cautions regarding potential limits of the calibrated parameters in new forested systems. Given the effort involved to recalibrate complicated, but effective algorithms for new forest systems and conditions, a reasonable question to ask is: Could these algorithms be applied in new forest types or locations using well-tuned parameter sets from a limited set of localized applications, with an additional, bulk statistical calibration from a reference dataset and a secondary statistical classification approach such as Random Forest (RF, Breiman, 2001)? We test this idea of secondary classification here using the LandTrendr (Landsat-based detection of Trends in Disturbance and Recovery) algorithm (Kennedy et al., 2010).

LandTrendr runs on a single band or spectral index (Kennedy et al., 2012), which may unnecessarily limit its value as a forest disturbance detection algorithm. In a recent study (Cohen et al., 2017), multiple algorithms were run on a common Landsat dataset across six diverse forested areas in the US, with each algorithm using different spectral bands and indices. When the maps from those algorithms were compared against each other they were found to be quite different, suggesting that, at least in part, spectral bands/indices used was a factor in the differences among maps. If calibration of LandTrendr through a

secondary classification model is effective, there would also be the opportunity to run the algorithm multiple times, each time using a different band or spectral index, before integrating the results from all runs as a multispectral ensemble using the secondary classification model.

The ensemble integration of maps from a variety of algorithms using RF was recently tested by Healey et al. (in press). In that study, empirical weights among an ensemble of map products were generated through a process called stacking (stacked generalization), in conjunction with reference data acquired through visual interpretation of Landsat time series data using a tool called TimeSync (Cohen et al., 2010). Disturbance mapping errors from the ensemble, relative to the individual maps from each algorithm, were greatly reduced when compared to the reference data. Healey et al. (in press) showed that adding informative, non-overlapping predictor information from different algorithms improved ensemble change detection performance. In this study, we tested the idea that valuable, non-overlapping information can be generated from a single algorithm operating on different parts of the electromagnetic spectrum. Specifically, we tested the stacking ensemble approach using LandTrendr and a combination of the six primary Landsat reflectance bands plus seven commonly used vegetation indices. This is similar to an approach used by Schultz et al. (2016), where maps from different indices derived from the BFAST Monitor algorithm (DeVries et al., 2015) were fused to create a single, improved map of deforestation in the tropics.

Three main objectives were addressed in this study:

- Quantify distributions of forest DSNR values for the original Landsat spectral bands and selected spectral indices, and the relationship between DSNR and disturbance detection error rates;
- Test secondary classification of LandTrendr when run on a single band/index using RF, and determine if there is a relationship between classification error rates and DSNR values; and
- Combine single band/index outputs of LandTrendr in a RF stacking ensemble to understand the power of secondary classification in a multispectral context, and evaluate the complementarity among bands/indices for forest disturbance detection.

2. Methods

2.1. The disturbance signal-to-noise ratio (DSNR) metric

To calculate the DSNR we used a TimeSync (Cohen et al., 2010) reference dataset that was collected from 1800 single pixel-sized plots (300 randomly selected per scene) over six, largely forested Landsat scenes widely dispersed across the conterminous US (Cohen et al., 2017). The forests consisted largely of a variety of needleleaf evergreen and broadleaf deciduous tree species commonly found across the different forested regions of the US (see Table 2, Cohen et al., 2017 for details). Of the 1800 plots, 1303 were forested, as determined by visual interpretation of high spatial resolution images in Google Earth. The Landsat time series data for each plot was temporally segmented by human interpretation using the TimeSync tool, which integrates simultaneous viewing of an annual series of full resolution Landsat image chips, temporal trajectories for each plot in a variety of spectral bands and indices, and the high-resolution images within Google Earth centered on the plot. To temporally segment the time series from 1984 to 2013 for a given plot, multiple spectral bands and indices (along with other tools) and breaks in the trends of spectral values were evaluated and identified (see Fig. 3 in both Cohen et al., 2010 and Kennedy et al., 2010). Using TimeSync, each segment was assigned a label represented by three types of observed forest processes – disturbance, growth, and stable – based on expert opinion. By definition, each segment was at least one year in length and bounded by two break points (start vertex and end vertex). For single segment plots the start vertex was 1984 and end vertex was 2013, whereas for multiple segment plots there were

multiple start and end vertices. The large majority of observed growth and stable segments were more than one year in length and thus had more than two data points (i.e., start and end vertices plus data points in between). Many disturbances (some harvest and most forest health decline events) took multiple years to occur and thus also had more than one year between vertices; however, abrupt disturbances (those associated with fires, wind events, and most harvest events) had only start and end vertices.

Of the 1303 forested plots, 754 had experienced a disturbance and these were used to calculate the DSNR. All disturbed plots had multiple segments (some combination of the disturbance, growth, and stable labels) across the full time-series length, and the spectral values for each segment were fitted with an ordinary least squares regression line where spectral values were fit as a function of time. To assure continuity of fitted segments, they were fitted in successive order, starting with the first segment. To fit the second segment, the fitted value from the end vertex of the first segment was used, and so on for all observed segments. For abrupt disturbance segments no fitting was possible (or necessary), but the fitted end vertex from the previous segment was used in place of the observed spectral value for that segment's start vertex. The fitting was done independently for each spectral band and vegetation index evaluated (Table 1).

The total number of disturbance segments across the 754 disturbed plots was 999 (an average of 1.3 disturbances per disturbed plot). For any given band or index, the DSNR was calculated for each disturbance segment in the whole time-series of observations from 1984 to 2013 at the plot (i.e., pixel) scale. The signal was the spectral band/index difference between the start and end vertices (i.e., the difference across the full segment length of one or more years) of each disturbance segment. Noise was calculated as the root mean square error (RMSE) from the collection of multi-segment regression residuals across the full time-series, as $RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$, where \hat{y} = the predicted spectral value and y = the observed spectral value for all fitted points from 1 – n . The DSNR for a disturbance segment was derived simply by dividing the calculated disturbance signal by the calculated plot-level noise. If there was more than one disturbance segment for a given plot, multiple DSNR values were calculated for that plot.

To evaluate the relative strengths of the bands/indices examined, the distribution of DSNR values across the 999 disturbance segments was displayed for each band/index as a box plot. The signs of the signals were predominantly negative (NIR, TCG, TCW, TCA, NDVI, NBR, and NDMI) or positive (Blue, Green, Red, SWIR1, SWIR2, TCB) depending on the band/index evaluated, with some small proportion of each having signals with the opposite sign than normal. To enable direct visual comparison of DSNR distributions using box plots it was desirable that the median box plot value for each band/index utilized be positive. Thus, prior to graphing the distributions for negative-dominant bands/indices, those distributions were inverted and each signal value multiplied by -1 . This also facilitated the relative rankings among bands/indices in terms of absolute median DSNR values, which are hereafter referred to as median DSNR values without an absolute value qualifier.

Table 1
Spectral indices used in this study.

Index	Formulation	Reference
TCB	$0.2043 * \text{Blue} + 0.4158 * \text{Green} + 0.5524 * \text{Red} + 0.5741 * \text{NIR} + 0.3124 * \text{SWIR1} + 0.2303 * \text{SWIR2}$	Crist (1985)
TCG	$-0.1603 * \text{Blue} + -0.2819 * \text{Green} + -0.4934 * \text{Red} + 0.7940 * \text{NIR} + -0.0002 * \text{SWIR1} + -0.1446 * \text{SWIR2}$	Crist (1985)
TCW	$0.0315 * \text{Blue} + 0.2021 * \text{Green} + 0.3102 * \text{Red} + 0.1594 * \text{NIR} + -0.6806 * \text{SWIR1} + -0.6109 * \text{SWIR2}$	Crist (1985)
TCA	Arctan (TCG/TCB)	Powell et al. (2010)
NDVI	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$	Rouse et al. (1974)
NBR	$(\text{NIR} - \text{SWIR2}) / (\text{NIR} + \text{SWIR2})$	Key and Benson (2005)
NDMI	$(\text{NIR} - \text{SWIR1}) / (\text{NIR} + \text{SWIR1})$	Wilson and Sader (2002)

Table 2

Unique combinations of bands/indices as a function of number of bands/indices used for LandTrendr training, and the median balanced error rates among combinations (Fig. 4).

Number of bands/indices	Combinations	Median error
1	13	0.637
2	78	0.431
3	286	0.388
4	715	0.362
5	1287	0.346
6	1716	0.336
7	1716	0.327
8	1287	0.321
9	715	0.316
10	286	0.311
11	78	0.308
12	13	0.304
13	1	0.296

2.2. Single band/index LandTrendr disturbance detection with and without secondary classification

When no mapping is involved, as in this study, applications of LandTrendr have two main steps: segmentation and filtering (Kennedy et al., 2010). Temporal segmentation is conceptually similar to that described above for TimeSync but is algorithmically complicated, requiring a series of tests to find an optimum value set for a host of parameters for any given band/index and forested area. Filtering is designed to separate signal from noise, by subjecting each segment to a thresholding rule based on percent vegetation cover change predicted from a spectral change model, which also has an optimum parameter set that must be determined. To test the efficacy of secondary classification, as applied here, LandTrendr temporal segmentation was accomplished by running the algorithm with a fixed set of segmentation parameter values (see Table 2 of Kennedy et al., 2012), regardless of band/index used. These values were derived after thorough testing with NBR by Kennedy et al. (2010) in western Oregon, and subsequently used for mapping there with NBR (Kennedy et al., 2012). Also, rather than using a threshold filter, for any band/index we simply labeled as disturbance any segment with a non-zero slope in the direction of disturbance, and all other segments as no disturbance. This was expected to reduce disturbance omissions and increase disturbance commissions relative to a reference dataset, which we expected the secondary classification to correct for.

Random Forest was used for secondary classification of LandTrendr, where the goal was to predict a binary classification label (disturbed or not disturbed, as classified in the reference dataset) independently for every year from 1984 to 2013, for each of the 1303 forested plots (a total of 39,090 annual predictions and/or data records). Of the total number of records in the TimeSync reference dataset, only 1965 (annual disturbance observations from across the 999 disturbance segments) were labeled as disturbance. Because RF should not be trained with a highly imbalanced classification dataset, we therefore could not include all 39,090 reference data records in the training dataset. Instead, we used only the 1965 records of disturbance from the

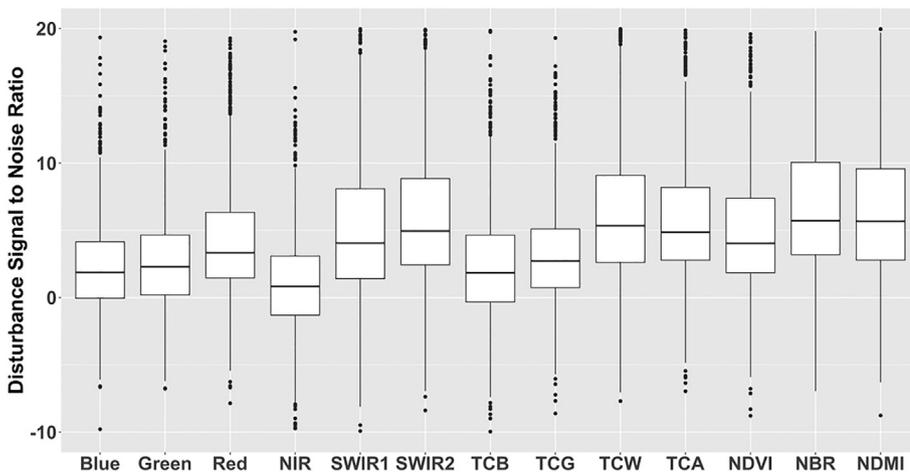


Fig. 1. Distribution of DSNR values for each band/index evaluated. The distributions for NIR, TCG, TCW, TCA, NDVI, NBR, and NDMI were inverted and multiplied by -1 for display (see Section 2.1).

reference dataset and augmented that with records of false positive disturbances from the LandTrendr classification for the given band/index under consideration. Because we did not use a threshold filter with the LandTrendr segmentations, we expected a significant number of records to be falsely predicted by LandTrendr as disturbed (determined by comparison to the reference dataset across all records). This enabled us to use the LandTrendr predictions to augment the selection of the reduced training dataset (1965 records), and thereby better achieve a balance of disturbance and no disturbance labels for training. This had the additional advantage of adding more challenging records to the overall training dataset, thereby enhancing the algorithm's potential to properly classify more difficult cases. We labeled the combined TImeSync disturbance records with the records having false positive disturbance labels from LandTrendr as the union 1 training dataset. Among the bands/indices, the number of false positives (as well as the years and plots for which they were predicted), and hence the sizes of union 1 training dataset, were variable.

For application of the RF model for any band or index, probability of being disturbed for any given year (t), as declared by TimeSync, was predicted using LandTrendr derived information: spectral value from year t , the spectral difference between t and $t - 1$ year, and segment duration for any data record in union 1. RF was run at the annual time-step on all training dataset observations (union 1) with 500 trees and out-of-bag (OOB) errors for the training dataset were reported as a function of RF probability threshold (using default parameters from randomForest version 4.6–12 in R). Changing the threshold shifted the training dataset records among the four cells of the disturbance-no disturbance error matrix, and thus the balance of omission and commission.

Our goal for this study was to report balanced disturbance omission and commission rates (i.e., where omission and commission are equal), for which we could have simply selected a probability threshold that balanced OOB errors. However, those OOB errors were only relevant to the training dataset records. Thus, to obtain an error matrix that included all 39,090 records, the remaining data records (39,090 minus union 1) needed to be added back in. These were the large majority of records that consisted of the union of reference data and LandTrendr predictions where both labels were no disturbance (union 2). Adding these union 2 records in without adjusting the probability threshold would have resulted in an imbalance of omission and commission. Hence, we adjusted the RF probability threshold from union set 1 to one that would achieve balance for the combined union 1 and union 2 records. For visual assessment, we created a box plot to display the distribution of error balance points among bands/indices. From this, we identified the model having the least amount of error from among the 13 bands/indices tested.

2.3. Multispectral LandTrendr disturbance detection with secondary classification

For the multispectral LandTrendr stacking ensemble we used all possible combinations of various bands and indices, 2–13 at a time (Table 2; the 13 individual bands/indices were included for completeness). The classification process described above for the single band/index executions of LandTrendr was repeated using a band/index combination-specific training dataset derived from a new union 1 rule that included any annual time-step observation where either TimeSync or LandTrendr (in any of the bands or indices used, from 2 to 13) declared a disturbance. Likewise, for the multispectral application of the RF model, probability of being disturbed for any given year (as declared by TimeSync) was predicted using the multiple band/index-specific LandTrendr-derived spectral and spectral change values and segment durations. As before, RF was run at the annual time-step on all multispectral (i.e., new) union 1 training dataset observations with 500 trees. Similarly, to achieve a balanced omission-commission error matrix for each combination, new combination-specific union 2 records were added back in and the probability thresholds adjusted accordingly. Box plots were created for the balanced error rates from all band/index combinations of each set from 2 to 13. From the box plots, we identified the model having the least amount of error for each of the sets.

3. Results

3.1. DSNR distributions among bands and indices

DSNR distributions reveal meaningful patterns among bands and indices tested (Fig. 1). The lowest distribution of values (i.e., those with the lowest signal in relation to noise) was associated with the NIR band, whereas the higher values were associated with SWIR bands and SWIR-based indices (NBR, NDMI, and TCW). Among the visible bands, the red band tended to have the highest distribution of DSNR values, but all three visible bands had higher values than the NIR band. Among the indices tested, TCB (Tasseled Cap brightness) and TCG (Tasseled Cap greenness) had the lowest distributions of DSNR values. The NDVI had higher values and, given the low DSNR of the NIR band, undoubtedly indicated the contribution of the red band. TCA values tended to be higher than the SWIR1 values, and were otherwise grouped more closely with SWIR2 and the SWIR-based indices. These observations are supported by the ranking of the median distribution values among the bands/indices, with the NBR having the highest median value and NIR having the lowest (Table 3).

Because both signal and noise contribute equally to the calculation of DSNR (data not shown), it was interesting to note that among the non-SWIR bands, NIR actually had the highest distribution of signals.

Table 3
Median DSNR values and rank for each band/index examined.

Band/index	Median DSNR	Rank
Blue	1.904	11
Green	2.329	10
Red	3.444	8
NIR	0.815	13
SWIR1	4.360	6
SWIR2	5.497	4
TCB	1.873	12
TCG	2.742	9
TCW	5.872	3
TCA	5.139	5
NDVI	4.167	7
NBR	6.581	1
NDMI	6.118	2

Table 4
Balanced error values and rank for each band index shown in Fig. 3.

Band/index	Error balance	Rank
Blue	0.690	11
Green	0.698	12
Red	0.666	10
NIR	0.758	13
SWIR1	0.515	4
SWIR2	0.502	3
TCB	0.639	8
TCG	0.649	9
TCW	0.469	1
TCA	0.555	6
NDVI	0.637	7
NBR	0.500	2
NDMI	0.531	5

But as it also had the highest distribution of noise values it was the weakest performer among the non-SWIR bands. Similarly, the distribution of signals for NDVI was third highest, after only NBR and NDMI (normalized difference moisture index); but given its high noise distribution, NDVI ranked only 7th among the DSNR median values (Table 3).

3.2. Single band/index LandTrendr disturbance classification

Without use of a LandTrendr threshold filter, the two SWIR bands and all three SWIR-based indices had the lowest disturbance omission rates, all below 0.4 (Fig. 2, upper left). The non-SWIR based bands and indices all had omission error rates between ~0.45 (TCA) and ~0.60 (NIR). Commission rates for all bands/indices were quite high (> 0.825), as expected given the lack of a threshold filter. Omission and commission rates were clearly related to DSNR, in that those bands/indices having the highest median DSNR values tended to exhibit the lowest error rates (Fig. 2, lower left and right).

After secondary classification via RF and the training dataset, the balanced omission-commission rates among the bands/indices clearly highlighted the value of SWIR reflectance for forest disturbance detection, with all SWIR bands and SWIR-based indices having the lowest rates (0.469–0.531) and all non-SWIR bands/indices having values between 0.555 and 0.758 (Table 4). In terms of the error balance point, TCW was the best performer and NIR the worst. TCA was grouped more closely with the SWIR-based than the non-SWIR-based data (Fig. 3). The linear relationship between balanced error rates and median DSNR was striking, having a correlation coefficient of 0.94.

3.3. Multispectral LandTrendr disturbance classification

Multispectral secondary classification of LandTrendr was quite effective at reducing error rates, with the median error balance point among single bands/indices dropping from 0.637 to 0.296 when all 13 bands indices included (Table 2). Much of the improvement happens when just 2 (0.431) or 3 (0.388) bands/indices are included in the

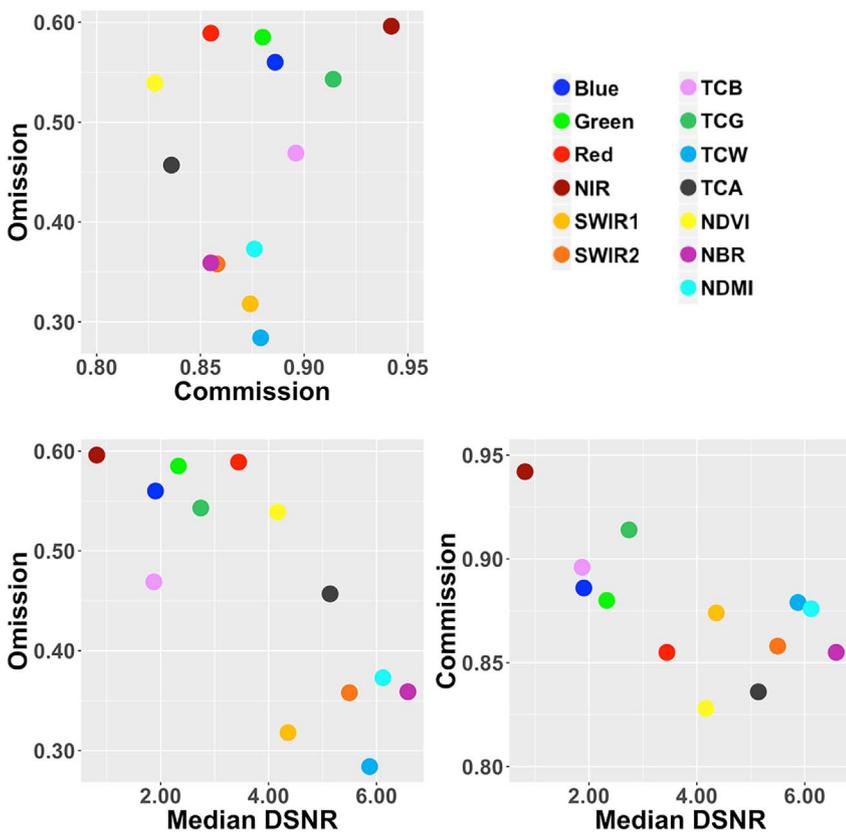


Fig. 2. (upper left) Omission and commission errors for the unfiltered LandTrendr runs using the six original Landsat bands and seven selected spectral indices. LandTrendr (lower left) omission and (lower right) commission errors as a function of band- and index-specific median DSNR values.

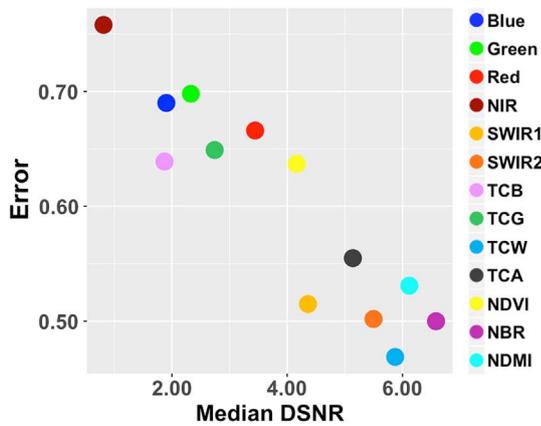


Fig. 3. Omission-commission error balance rates (from the LandTrendr secondary classification) among bands and indices evaluated as a function of median DSNR values.

multispectral classification. The relationship between number of bands/indices and balanced error rates is clearly asymptotic (Fig. 4), with a slow reduction in rates past 4 (0.362) or 5 (0.346) bands/indices, after the initial steeper drop from using a single predictor.

Although TCW exhibited the lowest error (i.e., was the best performing) among the bands/indices used in single variable secondary classification (error balance of 0.469, Table 4), that index was not included again in the best performing set – i.e., those combinations having the lowest error rates in Fig. 4, as number of bands/indices went from 2 to 12 – until six bands/indices were included in the disturbance predictor set (Table 5). The model set having 13 predictors included all bands/indices and was only included in Table 5 for completeness. Noteworthy is that TCW was included in all best performing predictor sets including at least six predictors. The best two band/index predictor model included SWIR1 and NBR, and these were included in every subsequent best performing set from three to 12. TCG was added to the best predictor set list at three predictors and remained in all subsequent best predictor sets. Similarly, TCB was added as the fourth predictor and then remained important for all subsequent sets. The red band was added in the five-variable model, but then dropped in and out of importance as other bands/indices were added. SWIR2 was the last band to be added in any predictor set, not being included until 12 variables were used.

Interestingly, the single best performing two-variable model, SWIR1 and NBR (0.376, Table 5; lowest error indicated on box plot for two bands/indices in Fig. 4) had a lower error balance rate than the median among three-variable models (0.388, Table 2). This pattern was consistent with the addition of every additional predictor through a total of

Table 5

Best performing models containing from 1 to 12 bands/indices, and their error balance rates. The model including all 13 bands/indices was included for completeness.

Number of bands/indices	Band/index set	Error balance
1	TCW	0.469
2	SWIR1, NBR	0.376
3	SWIR1, TCG, NBR	0.345
4	SWIR1, TCB, TCG, NBR	0.325
5	Red, SWIR1, TCB, TCG, NBR	0.316
6	Green, SWIR1, TCB, TCG, TCW, NBR	0.309
7	Green, SWIR1, TCB, TCG, TCW, NBR, NDMI	0.305
8	Green, NIR, SWIR1, TCB, TCG, TCW, NBR, NDMI	0.300
9	NIR, SWIR1, TCB, TCG, TCW, TCA, NDVI, NBR, NDMI	0.299
10	Green, Red, NIR, SWIR1, TCB, TCG, TCW, NDVI, NBR, NDMI	0.298
11	Blue, Green, Red, NIR, SWIR1, TCB, TCG, TCW, TCA, NDVI, NBR	0.297
12	Blue, Green, NIR, SWIR1, SWIR2, TCB, TCG, TCW, TCA, NDVI, NBR, NDMI	0.299
13	Blue, Green, Red, NIR, SWIR1, SWIR2, TCB, TCG, TCW, TCA, NDVI, NBR, NDMI	0.296

12 band/index predictors (Fig. 4). After the best eight-variable predictor set, there is apparently little need to add other bands/indices to that set, as there is minimal additional reduction in error balance.

Counting the number of times each band/index was included in the best predictor set models and then ranking them (Table 6) highlights the relative importance of each band/index in the presence of others for forest disturbance detection. The model set having 13 predictors included all bands/indices and was not included in Table 6. SWIR1 and NBR were included in 11 of the 12 best models and were thus ranked tied for 1st place, as most important bands/indices. SWIR2 was included only in the 12-variable model, suggesting its near total redundancy in the presence of other band/indices. This is very likely because SWIR2 is included in the formulation of NBR. Noteworthy is that although NIR had the lowest DSNR (Table 3), when used in combination with SWIR2 (i.e., within the NBR, Table 1), it was equally as important as SWIR1 when considering the best model sets (Tables 5,6). As its 3rd rank status suggests, TCG was the next most important index when examined in this context, being included in 10 of the 12 best performing models among sets. TCG is essentially a linear contrast between NIR and red (Table 1), like the NDVI, but is clearly more important than NDVI for disturbance detection in the presence of other bands/indices (Table 6). Again, it appears that the NIR band is

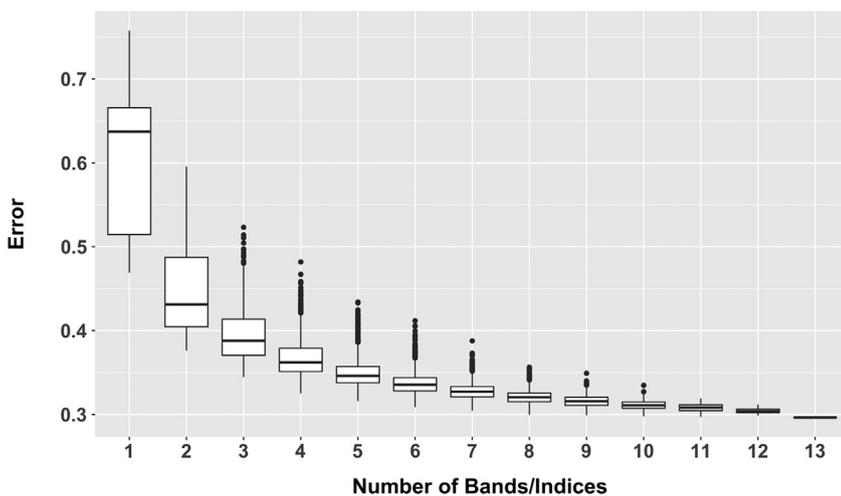


Fig. 4. Distributions of balanced omission and commission errors among all combinations for each number of bands/indices used.

Table 6

Number of times (from zero to N) that a given band/index was in the highest performing model for each set that includes different numbers of bands and indices from 1 to 12 (Set 13 excluded, as all bands were included in that set).

Band/Index	N	Rank
Blue	2	12
Green	6	6
Red	3	10
NIR	5	7
SWIR1	11	1
SWIR2	1	13
TCB	9	4
TCG	10	3
TCW	8	5
TCA	3	10
NDVI	4	9
NBR	11	1
NDMI	5	7

important, but this time when in combination with red reflectance (and small contributions from other bands (Table 1). Aside from SWIR2 when used in isolation, the Blue band had the distinction of also being among the least important spectral regions for forest disturbance detection, being included in only two models and thus ranked 12th among all bands/indices (Table 6).

4. Discussion

In a recent study, Cohen et al. (2017) used seven of the newer Landsat time series change detection algorithms to characterize forest disturbance over six different forested systems widely distributed across the US. The goal was to understand how different were the disturbance maps derived from the seven algorithms using the same Landsat time series stacks. All of the algorithms were previously calibrated for specific conditions (see Table 1 in Cohen et al., 2017), some more generally calibrated than others, and all performed well where tested. None were specifically calibrated for all of the local forest conditions over which they were applied in Cohen et al. (2017). When comparing the disturbance maps against reference data and themselves, it was learned that not only were the maps less accurate than in the earlier studies for which the algorithms were calibrated and tested, but also that the maps were remarkably different from each other. The most interesting outcome was that each map provided only part of the full array of disturbance types and magnitudes contained in the reference data. This raised several questions, including: (1) Can the maps be combined in a way that takes advantage of the strengths and minimizes the weaknesses of each to provide an overall better, single annual disturbance map series? (2) Because each algorithm used different dimensions of Landsat spectral space to derive the maps (and most used only one spectral index), could differences among the maps be, at least in part, a function of how spectral space was exploited? (3) If the answers to (1) and (2) are yes: (3) Could we be more strategic about combining maps while, at the same time, more fully exploiting spectral space when deriving the maps? (4) Given that the vast majority of omitted disturbances from mapping algorithms are associated with low magnitude disturbances (Cohen et al., 2017), how much would map accuracy improve if low magnitude disturbances are removed from consideration?

The basic answer to question 1, was provided by Healey et al. (in press), where the seven map sets from Cohen et al. (2017), in addition to others, were combined using RF and a training dataset based on TimeSync. In that study, a stacking ensemble of the various map products reduced mapping errors by as much as 50% relative to the individual input maps. We now build upon the Healey et al. (in press) study, as well as results from the objectives explored in this study, to answer questions 2–4.

4.1. How different Landsat spectral bands/indices likely impact forest disturbance detection

We developed the DSNR metric as a way to test the effectiveness of different Landsat bands and selected spectral indices for detecting forest disturbance. For that we used a dataset that represented many forest types and disturbance regimes across the US. The primary lesson from comparing band- and index-specific DSNR distributions among observations was the superior performance of SWIR reflectance (Table 3, Fig. 1, and similar to the findings of Schultz et al., 2016). Interestingly, it was not the original SWIR bands, but indices derived from them, that indicated the most promise for accurate disturbance detection. In particular, NBR, NDMI, and TCW. What the first two have in common is that they contrast SWIR against NIR reflectance, with the former relying on SWIR2 and the latter on SWIR1 (Table 1). That NBR exhibited a lower balanced error rate is most likely associated with the fact that SWIR2 had a higher DSNR distribution than SWIR1 (Table 3). TCW contrast both SWIR bands against, primarily, red and green reflectance, with some contribution from the NIR band (Table 1). With respect to the DSNR, the addition of the visible bands and the lesser importance of the NIR band had the effect of somewhat reducing the DSNR performance of TCW relative to the SWIR-based indices that avoided the visible region. It is noteworthy that TCA, which, being a combination of TCB and TCG mostly minimize the contribution of SWIR, but nonetheless performed better than SWIR1 using the DSNR metric. NDVI, by far the most commonly used spectral index in terrestrial remote sensing, was ranked only 7th (out of 13) with respect to DSNR performance.

Consistent with Schultz et al. (2016), we saw from the results of the multispectral ensemble classification of LandTrendr that using more than one spectral band/index can be quite advantageous for forest disturbance detection. To gain additional insight into the likely complementarity of multiple bands and indices when used together to detect forest disturbance in an ensemble setting, it is useful to examine band/index correlations for DSNR values among individual disturbance observations (Fig. 5). NBR (rank 1) and NDMI (rank 2) had the highest median DSNR ranks (Table 3) but, given the superiority of NBR, NDMI was not needed in the best performing multispectral ensemble models until seven variables were included (Table 5). This is likely because the DSNR correlation between NBR and NDMI was quite high (0.81, Fig. 5). TCW had the 3rd highest median DSNR rank and it was the most important index for single variable secondary classification of LandTrendr, but it was not added to the best performing multispectral models until six variables were utilized. Again, this is likely due to relatively high DSNR correlation with NBR (0.70), but also with SWIR1 (−0.83), both of which were included in the first several best performing multispectral ensemble models. NBR and SWIR1 were included in all of the best performing multispectral models in spite of the fact that their DSNR values were moderately correlated (−0.52). Because NBR uses SWIR2 and NIR reflectance, this suggests a strong complementary power for forest disturbance detection among these two Landsat bands and SWIR1. It was not until the three-variable model, and all subsequent models using 4–12 variables, that a visible band became important with the inclusion of TCG, which contrasts NIR against red reflectance. Although TCG was moderately correlated to NBR (0.55) it had a low DSNR correlation with SWIR1 (−0.13). The red band had a relatively low correlation with NBR (0.44) and moderate correlation with SWIR1 (0.62), which may help explain its importance in this context. TCB, TCG, and TCW all had low inter-index DSNR correlations (absolute values between 0.10 and 0.44) suggesting good complementarity among these three primary Tasseled Cap spectral variables for forest disturbance detection. Not surprisingly then, they were added in succession with the three-, four-, and five-variable models and remained important in all subsequent best performing multispectral ensemble models.

In Cohen et al. (2017), of the seven map sets compared, five were based on a single spectral dimension: NBR (LandTrendr), NDMI

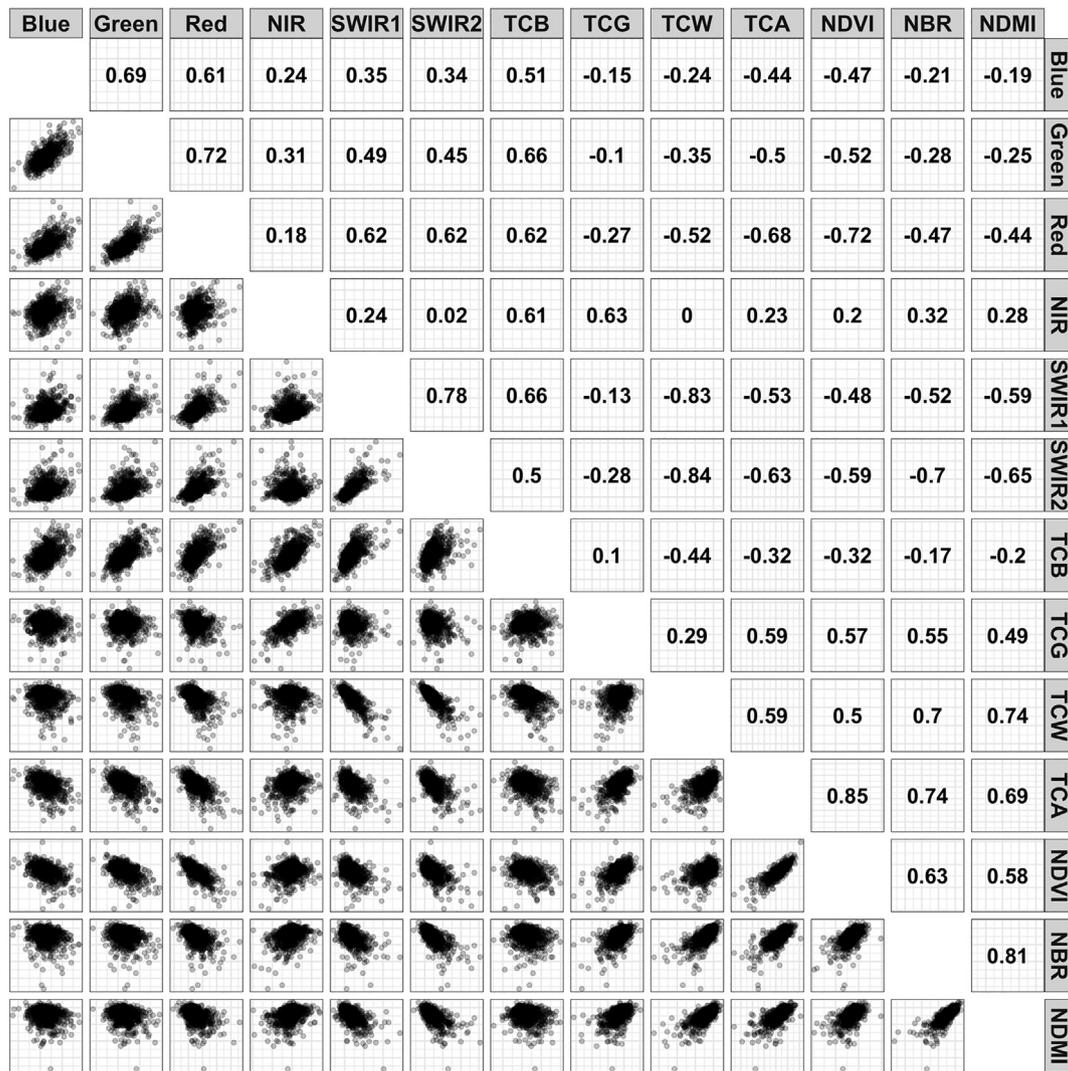


Fig. 5. Correlations of DSNR values among pairs of bands/indices.

(VeRED, Vegetation Regeneration and Disturbance Estimates through Time), NDVI (EWMA, Exponentially Weighted Moving Average and ITRA, Image Trends From Regression Analysis), and Forestness (an index unique to VCT, Vegetation Change Tracker). In light of the different performances among the bands and indices tested in this study, both in terms of DSNR and single band/index LandTrendr secondary classification errors, there can be little doubt that one of the important reasons the maps were so different was that they each employed different spectral dimensions. Although the two other algorithms compared (MIICA, Multi-index Integrated Change Analysis; CCDC, Continuous Change Detection and Classification) used multiple spectral dimensions, they either used multispectral bands/indices in sequence to identify disturbance (MIICA), or combined the results from multispectral statistical tests prior to identifying disturbance (CCDC). Neither of these, nor the single variable algorithms, used secondary classification to improve model performance, which may be key to facilitating cost-effective application of these algorithms in new forested locations.

4.2. Secondary classification as an effective way to exploit spectral space

Healey et al. (in press) demonstrated the value of secondary classification using multiple algorithm outputs with RF and a reference training dataset. In that study, error rates were substantially reduced relative to individual maps, with omission and commission errors balanced at 44% (see Fig. 4, BL outputs, Healey et al. in press). Schultz

et al. (2016) used a multispectral ensemble for secondary classification of maps derived from BFAST Monitor and also realized reduced error rates relative to single index algorithm outputs. In this study, we confirmed the value of a multispectral ensemble based on a single algorithm. However, using the same study areas and reference dataset here as that used by Healey et al. (in press), affords the opportunity to gain insights into the comparative value of a multi-algorithm ensemble vs. a multispectral ensemble for disturbance classification. Compared to the 44% balanced error rate from Healey et al. (in press) based on seven algorithms, the median error balance from a two-band multispectral ensemble with a single algorithm (i.e., LandTrendr) was 43% (Table 2), and for the best performing two-band ensemble omission and commission errors were 38% (Table 5). The addition of more bands/indices to the multispectral ensemble further lowered error rates.

Many of the modern algorithms being developed for and/or applied to the problem of forest disturbance mapping with Landsat time series are complicated, consisting of a logical progression of many steps, each often having a set of parameters that must be tuned to the forest conditions the algorithm is being applied to. While one always has the option of developing, parameterizing, and/or maintaining multiple algorithms to employ within the context of ensemble stacking to improve forest disturbance detections and mapping, that can be a labor-intensive, largely heuristic endeavor for new forest application areas and/or bands/indices utilized. As an alternative strategy, we demonstrated here that an ensemble made up of the outputs of even a single algorithm

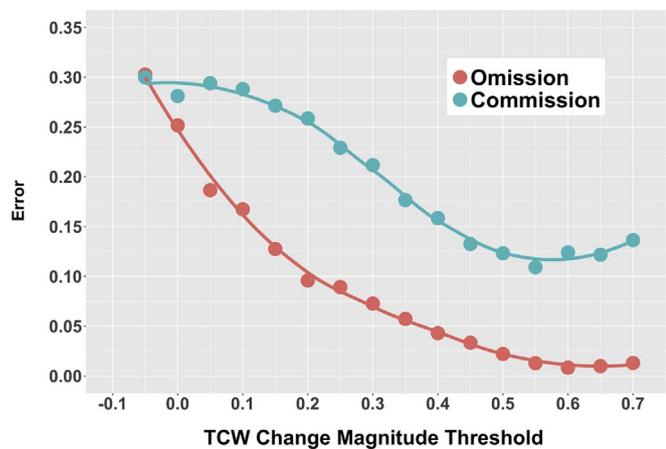


Fig. 6. Omission and commission error rates for different levels of disturbance magnitude trimming.

can produce large increases in accuracy if that algorithm is applied efficiently across the Landsat reflectance spectrum. Given that it is easier to operate a single algorithm than many different algorithms, this results points to a strategic approach to secondary classification that should maximize the local calibration value of stacking while simplifying operational processing.

4.3. Low magnitude disturbances

Given the effect on forests of a changing climate (Mildrexler et al., 2016), a more modern focus on ecosystem management over stand replacement harvests (Moeur et al., 2011), and international agreements including forest degradation in addition to deforestation (Kissinger et al., 2012), low magnitude disturbances are increasingly important to detect with remote sensing (McDowell et al., 2015). However, because these can be challenging to capture using an automated algorithm (Cohen et al., 2016), when targeting the full breadth of disturbance types and magnitudes high omission rates can be expected (Cohen et al., 2017). Fortunately, secondary classification that employs ensemble stacking of map output from numerous algorithms appears to be an effective means of reducing omission errors (Healey et al., in press). Moreover, this study has demonstrated that an efficient multispectral ensemble using LandTrendr alone may provide comparable or even improved performance. Nonetheless, even in this study, there remained a significant portion of low magnitude disturbances that could not be accurately detected ($\sim 30\%$ using 13 predictors, Table 2).

Because not all mapping applications require the full range of disturbances to be detected, for those applications it is possible to trim the low end of the magnitude range as a way to increase map accuracy. As a simple demonstration of this, using the predictions from the 13 band/index model, we progressively relabeled cases in both the TimeSync and predicted LandTrendr datasets from disturbed to undisturbed based on an incrementally higher change magnitude representing disturbance, as measured by annual relative change in TCW. Using an increasingly higher TCW magnitude threshold to successively remove lower magnitude disturbances from consideration (including those that exhibited TCW magnitudes in the “wrong” direction; i.e., those below zero), errors of both omission and commission became lower, but by differing amounts (Fig. 6). Errors of omission dropped to $\sim 10\%$ when all disturbances below 0.2 relative TCW change magnitude threshold were declared undisturbed. No longer balanced using this approach, errors of commission were $\sim 26\%$ with the 0.2 TCW change threshold. Using a 0.5 TCW change threshold (approximately equivalent to stand replacement disturbance) errors were reduced to $\sim 2\%$ (omission) and $\sim 12\%$ (commission). It is important to note that lower magnitude disturbances can be removed from any map using this same threshold

rule, so that there would be a disturbance map to match any degree of magnitude trimming desired.

5. Summary and conclusions

This study focused on the use of a popular Landsat time series algorithm, LandTrendr (Kennedy et al., 2010, to improve the potential for accurate mapping of forest disturbance over six Landsat scenes widely dispersed across the conterminous US. Like many newer Landsat disturbance mapping algorithms, LandTrendr specifies a workflow designed to run using a single spectral band or index, while incorporating numerous logical decision sequences and statistical tests to detect disturbance. Moreover, LandTrendr requires statistical calibration of basic parameter sets to function properly over targeted conditions. Although this requirement ensures good performance over the conditions for which LandTrendr is calibrated, the quality of mapping results was significantly lessened when it was applied in new areas without in-depth recalibration (Cohen et al., 2017).

As demonstrated by Healey et al. (in press), secondary classification using a stacking ensemble as a means to align a reference disturbance dataset with multiple disturbance map datasets from different algorithms can be a powerful way to improve disturbance mapping. In this study, we demonstrated that use of a single algorithm may be a more economical approach to employing ensemble stacking for improved disturbance mapping. This was accomplished by fuller exploitation of spectral dimensionality, running LandTrendr several times, each time using a different spectral band or index and then aligning the derived multispectral output with a disturbance reference dataset using RF.

As part of this study, we developed a disturbance signal-to-noise ratio (DSNR) metric that appears to be a good proxy for disturbance detection error. The DSNR explained the relative performance of LandTrendr run on individual bands/indices and illustrated that using SWIR reflectance or a SWIR-based index was the most important consideration when using a single spectral dimension for forest disturbance detection across the forests of the US. For a two-dimensional LandTrendr multispectral ensemble, we also demonstrated that NIR reflectance is the single most important complementary spectral band, in spite of its tendency to exhibit low DSNR values. Adding additional spectral bands/indices improved detection results, but the relationship between reduced errors and number of bands/indices was asymptotic, with most of the improvement happening with the use of 4–7 spectral dimensions.

Although omission and commission errors remained close to 30% using all 13 bands/indices integrated into the secondary classification, most of those errors were associated with low magnitude spectral change. As not all mapping applications require the full range of disturbance magnitudes to be represented, we demonstrated a means for determining the reduction in error by trimming the target magnitude range. For example, by eliminating the lowest 20% of spectral change from consideration, errors were reduced to $\sim 10\%$ (omission) and 25% (commission). A disturbance map set can be created to match any level of desired trimming, thereby tailoring the map to the required application.

Acknowledgements

This research was funded by the US Forest Service (IRDB-LCMS-NWFP) Landscape Change Monitoring System, the Region 6 Effectiveness Monitoring Program of the US Forest Service, the NASA Carbon Monitoring System (NASA reference number: NNH13AW62I), and the Oregon Department of Forestry (FRXNB216).

References

Allen, C.D., Breshears, D.D., McDowell, N.G., 2015. On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the

- Anthropocene. *Ecosphere* 6, 1–55.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- 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, 3316–3332.
- Cohen, W.B., Spies, T.A., Alig, R.J., Oetter, D.R., Maiersperger, T.K., Fiorella, M., 2002. Characterizing 23 years (1972–1995) of stand replacement disturbance in western Oregon forests with Landsat imagery. *Ecosystems* 5, 122–137.
- Cohen, W.B., Yang, Z., Kennedy, R.E., 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., Huang, C., Meigs, G.W., 2016. Forest disturbance in the conterminous US from 1985–2012: the emerging dominance of forest decline. *For. Ecol. Manag.* 360, 242–252.
- 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. <http://dx.doi.org/10.3390/f8040098>.
- Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., Lambin, E., 2004. Digital change detection methods in ecosystem monitoring: a review. *Int. J. Remote Sens.* 25, 1565–1596.
- Crist, E.P., 1985. A TM Tasseled Cap equivalent transformation for reflectance factor data. *Remote Sens. Environ.* 17, 301–306.
- DeVries, B., Verbesselt, J., Kooistra, L., Herold, M., 2015. Robust monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. *Remote Sens. Environ.* 161, 107–121.
- Hansen, M.C., Potapov, P.V., Moore, R., et al., 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342, 850–853. with supplementary materials. www.sciencemag.org/content/342/6160/850/suppl/DC1.
- Healey, S.P., Cohen, W.B., Zhiqiang, Y., Krankina, O., 2005. Comparison of Tasseled Cap-based Landsat data structures for forest disturbance detection. *Remote Sens. Environ.* 97, 301–310.
- Healey, S.P., Cohen, W.B., Spies, T.A., Moeur, M., Pflugmacher, D., Whitley, M.G., Lefsky, M.A., 2008. The relative impact of harvest and fire upon landscape-level dynamics of older forests: lessons from the Northwest Forest plan. *Ecosystems* 11, 1106–1119.
- Healey, S.P., Cohen, W.B., Yang, Z., Brewer, C.K., Brooks, E.B., Gorelick, N., Hernandez, A.J., Huang, C., Hughes, M.J., Kennedy, R.E., Loveland, T.R., Moisen, G.G., Schroeder, T.A., Stehman, S.V., Vogelmann, J.E., Woodcock, C.E., Yang, L., Zhu, Z., 2017. Mapping forest change using stacked generalization: an ensemble approach. *Remote Sens. Environ.* <http://dx.doi.org/10.1016/j.rse.2017.09.029>. (in press).
- Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2015. An integrated Landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. *Remote Sens. Environ.* 158, 220–234.
- Huang, C., Goward, S.N., Masek, J.G., Thomas, N., 2010. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sens. Environ.* 114, 183–198.
- Hughes, M.J., Kaylor, S.D., Hayes, D.J., 2017. Patch-based forest change detection from Landsat time series. *Forests* 8, 166. <http://dx.doi.org/10.3390/f8050166>.
- 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 Sens. Environ.* 114, 2897–2910.
- Kennedy, R.E., Yang, Z., Cohen, W.B., Pfaff, E., Braaten, J., Nelson, P., 2012. Spatial and temporal patterns of forest disturbance and regrowth within the area of the Northwest Forest Plan. *Remote Sens. Environ.* 122, 117–133.
- Key, C.H., Benson, N.C., 2005. Landscape assessment: remote sensing of severity, the normalized burn ratio. In: Lutes, D.C. (Ed.), FIREMON: Fire Effects Monitoring, and Inventory System. USDA Forest Service, Rocky Mountain Research, Station, Ogden, UT.
- Kim, D.-H., Sexton, J.O., Noojipady, P., Huang, C., Anand, A., Channan, S., Feng, M., Townshend, J.R., 2014. Global, Landsat-based forest-cover change from 1990 to 2000. *Remote Sens. Environ.* 155, 178–193.
- Kissinger, G., Herold, M., De Sy, V., August 2012. Drivers of Deforestation and Forest Degradation: A Synthesis Report for REDD + Policymakers. Lexeme Consulting, Vancouver Canada.
- Masek, J.G., Huang, C., Cohen, W.B., Kutler, J., Hall, F.G., Wolfe, R., Nelson, P., 2008. North American forest disturbance mapped from a decadal Landsat record: methodology and initial results. *Remote Sens. Environ.* 112, 2914–2926.
- McDowell, N., Coops, N.C., Beck, P.S.A., Chambers, J.Q., Gangadagamage, C., Hicke, J.A., Huang, C.-Y., Kennedy, R.E., Krofcheck, D.J., Litvak, M., Meddens, A.J.H., Muss, J., Negrón-Juarez, R., Peng, C., Schwantes, A.M., Swenson, J.J., Vernon, L.J., Williams, A.P., Xu, C., Zhao, M., Running, S.W., Allen, C.D., 2015. Global satellite monitoring of climate-induced vegetation disturbances. *Trends Plant Sci.* 20, 114–123.
- Meddens, A.J.H., Hicke, J.A., 2014. Spatial and temporal patterns of Landsat-based detection of tree mortality caused by a mountain pine beetle outbreak in Colorado, USA. *For. Ecol. Manag.* 322, 78–88.
- Meigs, Garrett, Kennedy, R.E., Cohen, W.B., 2011. A Landsat time series approach to characterize bark beetle and defoliator impacts on tree mortality and surface fuels in conifer forests. *Remote Sens. Environ.* 115, 3707–3718.
- Mildredler, D., Yang, Z., Cohen, W.B., 2016. A forest vulnerability index based on drought and high temperatures. *Remote Sens. Environ.* 173, 314–325.
- Moeur, M., Ohmann, J.L., Kennedy, R.E., Cohen, W.B., Gregory, M.J., Yang, Z., Roberts, H.M., Spies, T.A., Fiorella, M., 2011. Northwest Forest Plan—the first 15 years (1994–2008): status and trends of late-successional and old-growth forests. In: Gen. Tech. Rep. PNW-GTR-853. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR 48 p.
- Pohmann, R., Speck, O., Scheffler, K., 2016. Signal-to-noise ratio and MR tissue parameters in human brain imaging at 3, 7, and 9.4 Tesla using current receive coil arrays. *Magn. Reson. Med.* 75, 801–809.
- Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B., Ohmann, J.L., 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: a comparison of empirical modeling approaches. *Remote Sens. Environ.* 114, 1053–1068.
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. In: Freden, S.C., Mercanti, E.P., Becker, M. (Eds.), Third Earth Resources Technology Satellite-1 Symposium. Volume I: Technical Presentations. NASA SP-351, NASA, Washington, D.C., pp. 309–317.
- Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R.E., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindaschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalsky, V., Lee, Z.P., Lyburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: science and product vision for terrestrial global change research. *Remote Sens. Environ.* 145, 154–172.
- Schultz, M., Clevers, J.G.P.W., Carter, S., Verbesselt, J., Avitabile, V., Quang, H.V., Herold, M., 2016. Performance of vegetation indices from Landsat time series in deforestation monitoring. *Int. J. Appl. Earth Obs. Geoinf.* 52, 318–327.
- Trenberth, K.E., 1984. Signal versus noise in the southern oscillation. *Mon. Weather Rev.* 112, 326–332.
- 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.
- Wilson, E.H., Sader, S.A., 2002. Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sens. Environ.* 80, 385–396.
- Wulder, M.A., Skakun, R.S., Kurz, W.A., White, J.C., 2004. Estimating time since forest harvest using segmented Landsat ETM + imagery. *Remote Sens. Environ.* 93, 179–187.