Title
Differing Sensitivities to Fire Disturbance Result in Large Differences Among Remotely Sensed Products of Vegetation Disturbance

Journal
Ecosystems, 22(8)

ISSN
1432-9840

Authors
Palomino, Jenny
Kelly, Maggi

Publication Date
2019-12-01

DOI
10.1007/s10021-019-00367-9

Peer reviewed

Ecosystems
https://doi.org/10.1007/s10021-019-00367-9

ECOSYSTEMS

Differing Sensitivities to Fire Disturbance Result in Large Differences Among Remotely Sensed Products of Vegetation Disturbance

Jenny Palomino[1] and Maggi Kelly[2,3,4]*

[1]*Earth Lab, University of Colorado, Boulder, Boulder, Colorado, USA;* [2]*Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, California 94720-3114, USA;* [3]*Geospatial Innovation Facility, University of California, Berkeley, Berkeley, California, USA;* [4]*Division of Agriculture and Natural Resources, University of California, Davis, California, USA*

Abstract

Recent advances in high-performance computing (HPC) have promoted the creation of standardized remotely sensed products that map annual vegetation disturbance through two primary methods: (1) conventional approaches that integrate remote sensing-derived vegetation indices with field data and other data on disturbance events reported by public agencies on a year-to-year basis, and (2) ''big'' data approaches using HPC to automate algorithms and workflows across an entire time series. Given the recent proliferation of these annual products and their potential utility for understanding vegetation dynamics, it is important for product end users (that is, practitioners and researchers in domains other than remote sensing) to understand the differences in their representations of disturbance and the conditions under which they report it. We use fire in California as a case study to compare reported disturbance across three widely used vegetation disturbance products—LANDFIRE (representing the conventional approach), Hansen Global Forest Change (GFC), and North America Forest Dynamics (NAFD), the latter two created from automated approaches. Using Google's Earth Engine, we compared their total and annual amounts of fire and non-fire disturbance for 2001–2010 and examined the products' reported disturbance across different environmental and burn conditions. We found that GFC and NAFD reported similar amounts of disturbance that were consistently much lower than LANDFIRE's reported disturbance across all years, regions, and habitats. We also found that despite the differences in amounts of reported disturbance, the products identified disturbance in similar ranges of bioclimatic conditions and habitat types, and thus, differing environmental conditions in areas reported as disturbed were not the drivers of the difference. Rather, we found that lower sensitivity to fire disturbance for GFC and NAFD, as compared to LANDFIRE, was a key driver of the overall differences in the amounts and locations of reported disturbance; both GFC and NAFD reported much lower amounts of fire disturbance than LANDFIRE across all burn conditions. Furthermore, the dif-

Received 24 April 2018; accepted 25 February 2019

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s10021-019-00367-9) contains supplementary material, which is available to authorized users.
Data hosted on Figshare at https://figshare.com/articles/Vegetation_Disturbance_Product_Comparison/7768586.

Author Contributions JP and MK conceived of the work together. JP led the analysis and data curation; MK provided feedback on the analysis. JP wrote the original draft; MK provided review and editing throughout the process.

Corresponding author; e-mail: maggi@berkeley.edu

Published online: 12 April 2019

ference in reported disturbance between LANDFIRE and GFC/NAFD was greater for fire disturbance than for non-fire disturbance; LANDFIRE reported more than double the total amounts of fire disturbance of GFC and NAFD in the study period. Based on our results, we encourage end users to choose the appropriate disturbance product based not only on spatial extent and habitat but also on the disturbance type of interest (that is, fire and non-fire). Overall, rather than focusing on accuracy, our study quantifies the extent to which the products exhibited differences in the amounts and locations of reported disturbance to provide insight into these products' representations of disturbance and help end users evaluate and choose the most appropriate product for their needs.

Key words: vegetation disturbance; fire; Hansen Global Forest Change; LANDFIRE; North American Forest Dynamics (NAFD).

Highlights

- LANDFIRE reported the highest amounts of vegetation disturbance in all years and all habitat types.
- Differences among products are greatest for fire, rather than non-fire disturbance.
- Low sensitivity to fire disturbance for GFC and NAFD was a key driver of differences in reported disturbance.
- Choosing a disturbance product needs to be based on spatial extent, habitat, and disturbance type.

Introduction

Recent advances in high-performance computing (HPC; including distributed, parallel, clustered, and cloud-based methods) have provided new opportunities to analyze ''big'' remotely sensed data across broader spatial scales (for example, global) and finer temporal (for example, annual) resolutions (Plaza and Chang 2007; Lee and others 2011; Kalluri and others 2015; Kang and Lee 2016; Kumar and others 2017). These HPC-based remote sensing analyses are increasingly being used to identify long-term vegetation changes using the Landsat Time Series (LTS) (Hermosilla and others 2016; Soulard and others 2016). Some of these efforts have resulted in standardized maps (that is, products) of annual vegetation disturbance (that is, annual changes in vegetation due to natural or anthropogenic events) across the USA and globally (Hansen and others 2013; Goward and others 2016). These LTS-based annual products of vegetation disturbance have been primarily produced through two methods: (1) conventional approaches that integrate remote sensing-derived vegetation indices with field data and other data on disturbance events reported by public agencies on a year-to-year basis, and (2) ''big'' data approaches using HPC to automate algorithms and workflows across an entire time series (typically all LTS images collected for a given time period and spatial extent).

A key example of the conventional, year-to-year approach is Landscape Fire and Resource Management Planning Tools (LANDFIRE) for the United States (USA). Historically, LANDFIRE focused on providing spatially explicit data of canopy characteristics, such as vegetation height and cover, which were typically derived from modeling and scaling up of sampled field data using satellite imagery (Keane and others 2007; Reeves and others 2009; Rollins 2009; Ryan and Opperman 2013). To create a product specifically focused on annual vegetation disturbance, LANDFIRE gathered location data on vegetation disturbance events (for example, fire, harvest, pestilence) reported by public agencies and integrated these data with HPC-based calculations of vegetation indices from the LTS (such as normalized burn ratios used to identify burned areas) to produce a map of vegetation disturbance for each year between 1999 and 2014.

In contrast to mapping disturbance on a year-to-year basis, ''big'' data approaches analyze satellite data across an entire time period of interest to identify key inflection points that indicate vegetation disturbance. Specifically, these approaches employ automated algorithms and workflows that leverage HPC to identify annual changes in spectral signatures (that is, reflectance) across large multi-temporal stacks of images from LTS. Key examples of this automated approach are North American Forest Dynamics (NAFD) and Hansen Global Forest Change (GFC). NAFD is a collection of standardized, annual maps of vegetation disturbance across North America for 1984–2010 (Goward and others 2016), produced by applying the Vegetation Change Tracker (VCT) algorithm (Huang and others 2010) to the LTS within an HPC environment developed by NASA (NASA Earth Exchange, NEX) (Nemani and others 2011). GFC is the first global forest change product to be produced at the spatial and temporal resolution of LTS (Hansen and others

2013) and was produced through large-scale collaboration between Google and academic researchers that leveraged Google's cloud-based HPC infrastructure to produce standardized, annual maps for 2000–2014. Additional algorithms that have been developed for the LTS, such as LandTrendr (Kennedy and others 2010), Continuous Change Detection and Classification (CCDC) (Zhu and Woodcock 2014), and other examples (Cohen and others 2017; Healey and others 2017) will likely result in future vegetation disturbance products that are also produced from HPC applications of automated algorithms (Pengra and others 2016).

Although there have been several comparative evaluations of the algorithms used by remote sensing experts to map vegetation disturbance, such as VCT, CCDC, LandTrendr and others (Cohen and others 2017; Healey and others 2017), there have been no systematic comparisons (at the time of this publication) of the vegetation disturbance products that are frequently being used by non-remote sensing experts to study the impacts of disturbance on ecosystems (for example, GFC, NAFD, LANDFIRE). Previously published papers evaluating these vegetation disturbance products have focused on the accuracy or validation of an individual product (Krasnow and others 2009; Thomas and others 2011; Zimmerman and others 2013; Hyde and others 2015; Tyukavina and others 2015; McKerrow and others 2016; Gudex-Cross and others 2017; Zhao and others 2018) or on integration of these products (or the algorithms used to create them) to improve the accuracy of disturbance identification (Healey and others 2017; Schroeder and others 2017; Soulard and others 2017).

Given the recent proliferation of these annual products as well as their potential utility for non-remote sensing experts to explore the spatial–temporal impacts of vegetation disturbance on ecosystems, it is important to evaluate these products to understand the differences in their representations of disturbance (for example, amounts and locations) as well as the conditions under which they report it (for example, sensitivity to disturbance across different bioclimatic, habitat, and burn conditions). As each product was created by different initiatives with specific goals and motivations, it is understood that they will demonstrate differences in their locations and amounts of reported disturbance. However, as the amounts and locations of disturbance reported by these products (GFC, NAFD, and LANDFIRE) have not been systematically compared, it is not known to what extent the choice of the map product used to represent disturbance in ecological studies might impact the results and conclusions that have been drawn regarding the impacts of disturbance on ecosystem processes.

In this paper, we provide a first comparative evaluation of GFC, NAFD, and LANDFIRE, which are the three LTS-derived vegetation disturbance products that have overlapping spatial (for example, USA) and temporal (for example, annual from 2001 to 2010) extents at the time of this publication. We use fire in California as a case study to identify where and when these products report disturbance, using two widely used reference datasets of fire across California: Monitoring Trends in Burn Severity (MTBS) and fire perimeters from the California Department of Forestry's Fire and Resource Assessment Program (FRAP) database. As California is a fire-prone state, large wildfires occur annually across the state, resulting in significant changes to forest, shrub and grass (Stephens and others 2007; Moritz and Stephens 2008; Krasnow and others 2017), the three habitat types of focus in this study. Rather than focusing on accuracy (as previously noted publications have already done) or spatial agreement among these products, this paper quantifies the extent to which the products exhibited differences in the amounts and locations of reported disturbance.

Specifically, this paper asked:

1. How comparable were the three vegetation disturbance products in their amounts of reported disturbance across California?
2. How comparable were the environmental conditions (that is, bioclimate and habitat types) in the areas reported as disturbed by the products?
3. To what extent could differences in reported disturbance be attributed to differing sensitivities to fire disturbance?

To help end users of these products better understand how these products were created, we first review the key differences between the automated approaches of GFC and NAFD and the more conventional approach of LANDFIRE and highlight how their methods of creation result in differing thresholds (that is, sensitivities) for reporting disturbance. For our comparative analysis of these products, we employed Earth Engine (EE), a cloud-based, distributed HPC platform created by Google that provides a set of analytical functions for analyzing vector and raster-based geographic data via multiple cloud-based user interfaces (Gorelick and others 2017). Even while limited to California, the disturbance products evaluated in this study are ''big'' data, as approximately 450 million pixels were analyzed for each year based on the LTS spatial resolution of 30 m. As such, we used the JavaScript API Code Editor to leverage the HPC capabilities of EE as well as the built-in functionality such as code-sharing and cloud data storage, which support reproducibility and collaboration (Palomino and others 2017).

We recognize that end users (for example, practitioners and researchers of domains like ecology and conservation biology without remote sensing expertise) are seeking these products to accurately characterize annual vegetation change and disturbance in their work, and this is the first comprehensive study to examine the key differences across these competing annual vegetation disturbance products. We believe our results provide insight into the differences among these products' representations of disturbance and can help end users evaluate and choose the most appropriate product for their needs.

Study Data: Annual, Standardized Vegetation Disturbance Products

The three vegetation disturbance products included in this study are the only annual, standardized products that share an overlapping spatial and temporal extent at the time of this publication. In this section, the goals and workflow of each product are described as well as the key strengths and limitations of the workflow. A summary of the key differences among these products is found in Table 1.

Automated, ''Big'' Data Approach

Hansen Global Forest Change (GFC)

Motivated by the limitations of existing data on global forests and previous workflows that were amalgams of differing datasets, methods, and definitions, the primary goal of GFC was to provide a map of global forest extent and change that could be used to quantify annual forest loss and gain using a systematic and replicatable workflow (Hansen and others 2013). Prior to GFC, there was no spatially and temporally explicit global map at a useful spatial resolution such as that of the LTS, as ''previous efforts have been either sample-based or employed coarse spatial resolution data'' (Hansen and others 2013, p. 850). Using Google's Earth Engine (EE), GFC employed a supervised classification to identify locations of forest loss between 2000 and 2014, using training data of locations prelabeled with known forest loss or no forest loss (that is, discrete identification). As the baseline for forested area, pixels containing tree cover with a height greater than 5 m were identified as forested. In the training dataset, forest loss was represented by pre-identified pixels that had experienced ''stand-replacement disturbance'' leading to a nonforest state for the pixel (Hansen and others 2013). As such, forest degradation that did not result in a new cover type (for example, only reduction in greenness) was not labeled as forest loss. The year of loss was identified through an analysis of a time series of the Normalized Difference Vegetation Index (NDVI; an indicator of greenness calculated from the LTS bands for red and near-infrared); the year with the sharpest drop in NDVI in the time period was identified as the year of loss. The cause or severity of loss was not provided in the data; a pixel-based label of uncertainty is also not provided (that is, uncertainty of loss at a given pixel).

The overall accuracy of the forest loss reported by GFC as well as the producer's and user's accuracies have been evaluated to be greater than 80% (approximately) across all climatic biomes as well as globally (Hansen and other 2013, supplemental instruction). The creators of GFC recommend using their product for regional to global analyses of forest extent and change (Hansen and others 2013), indicating that GFC is not intended for nonforest habitats or for use at local scales. A specific minimal mapping unit (MMU) or analysis resolution scale for use of GFC is not provided.

The key strength of GFC is its unique global coverage of forest cover and loss at the spatial resolution of the LTS. Furthermore, its workflow is simple and easily replicated using standard remote sensing techniques such as supervised classification and time series analysis of vegetation indices; therefore, its creation process is easily understood by end users who are interested in learning more about the product and interpreting its accuracy. However, as the baseline for the forested state is defined at heights of 5 m or greater across the Landsat pixel resolution (30 m), GFC may not be appropriate for all forested habitats, such as early successional forest, mixed vegetation-forest, and sparse or open canopy forest. Another major limitation of GFC stems from its focus on discrete losses of vegetation. Reductions in vegetation cover are not identified until the reduction is significant enough to cause a change in cover type or notable drop in greenness, limiting its identification of certain habitat disturbances (for example, those that result in minimal change in greenness such as low intensity fires) or its applicability to certain

Table 1. Summary of Vegetation Disturbance Products

Disturbance product	Time period	Extent and target vegetation	Definition and identification of disturbance from LTS	Computing environment
Hansen Global Forest Change (GFC)	2000–2014[a]	Global Forest	*Loss of cover (discrete)*: ''stand-replacement disturbance'' leading to a non-forest state for the pixel (Hansen and others, 2013, supplemental material) *Identification method (automated)*: supervised classification of forest loss; NDVI time series analysis to identify year of loss	Google Earth Engine (EE): cloud-based distributed computing platform (proprietary)
North American Forest Dynamics (NAFD)	1986–2010	North America Forest	*Disturbance of cover (continuous)*: annual change in the integrated forest z-score (IFZ), an inverse measure of likelihood that a pixel is forested in a given year *Identification method (automated)*: VCT algorithm applied to LTS, supplemented by dNBR analyses	NASA Earth Exchange (NEX): HPC cluster managed by NASA
LANDFIRE	1999–2014	USA All vegetation	*Loss and disturbance of cover (discrete and continuous)*: depending on data integrated in that year *Identification method (manual integration)*: year-by-year integration of disturbance events reported by public agencies and calculated indices from LTS including NDVI, dNBR, MTBS, VCT algorithm, and Multi-Index Integrated Change Algorithm (MIICA)	Custom multi-node HPC cluster managed by the USGS Earth Resources Observation and Science (EROS)

[a]*First year of identifiable loss is 2001.*

habitats (for example, those that do not demonstrate a clear change in greenness before and after disturbances such as mixed vegetation).

North American Forest Dynamics (NAFD)

Similar to the motivation behind GFC, the creation of NAFD was prompted by the North American Carbon Program (NACP) which recognized that existing monitoring programs did not have data at the appropriate spatial and temporal resolution needed for accurate estimates of carbon fluxes across North America. In addition, they identified that the impacts of forest disturbance on carbon dynamics needed to be better understood, in order to manage ecosystems effectively (Goward and others 2008, p. 105). To address these data gaps, the NAFD project leveraged the spatial and temporal resolution of the LTS to quantify forest change that results from both severe and minor disturbances, ''including phenomena such as partial harvest, thinning, and insect damage, which may not always destroy the whole stand'' (Goward and others 2008, p. 106).

To create the NAFD annual map products (Goward and others 2016), the Vegetation Change Tracker (VCT) algorithm (Huang and others 2010) was employed on NASA Earth Exchange's computing facilities (NEX; Nemani and others 2011) to identify annual forest cover and disturbance between 1986 and 2010. VCT calculates an integrated forest z-score (IFZ), an inverse measure of likelihood that a pixel is forested in a given year, within a time series analysis to identify locations of forest stability and change for a given time period. For NAFD, the IFZ was informed by normalization indices that were calculated from a training dataset of known forest locations. A consistently low IFZ (close to zero) across the time period indicated relative stability in the forest cover, while a marked increase in IFZ indicated a disturbance in forest cover, ranging from partial to total stand disturbance (that is, continuous identification of disturbance). To better incorporate disturbances specifically due to fire, NAFD also integrated differenced Normalized Burn Ratio (dNBR) analyses (that is, the ratio of the difference between the near-infrared and short-wave infrared bands of the LTS over the sum of these bands) that compared the NBR indices between a pair of pre- and post-fire images. Like GFC, NAFD also did not contain information on the cause or severity of the disturbance.

The overall accuracy of NAFD was first evaluated to range from 77 to 88% at an annual scale; for stand-replacement events (including harvesting and fires), the overall accuracy was evaluated to be approximately 92% (Thomas and others 2011). A later accuracy assessment that aggregated disturbance across years into one class reported ''an overall accuracy of 84.5% in representing disturbance that resulted in at least 20% cumulative canopy loss,'' with user's and producer's accuracies of 67% and 62.9%, respectively (Zhao and others 2018, p. 31). Although a pixel-based label of uncertainty was also not provided for NAFD, these robust assessments can be used to improve estimations of disturbed area within a given study area. Regarding its recommended use, NAFD was created using the LTS spatial resolution of 30 m; however, a MMU was applied to produce the final data product (four pixels at a 30 m resolution for disturbed pixels). As such, the producers of NAFD recommend analyzing the data at a coarser resolution, suggesting 60–100 m (Goward and others 2016).

Although NAFD is also produced from an algorithmic approach, the identification of continuous disturbance (that is, reduction in vegetation cover) is a key advantage of NAFD, as compared to GFC which focuses solely on discrete disturbance (that is, stand-clearing or replacement). As NAFD can identify partial disturbances that do not result in land cover change, it potentially ''captures most rapid stand-clearing events (including clearcut harvests and fire), as well as many non-stand-clearing events (partial harvest, thinning, storm damage, insect damage)'' (Masek and others 2013, p. 1089). However, there are some limitations for the use of NAFD that stem from the IFZ disturbance identification process centered on identifying deviations from a persistent forest state. Although the creation process of NAFD is completely reproducible and easily integrated with other analyses due to the portability of the VCT algorithm across platforms, interpretation of the IFZ score makes NAFD less approachable for end users to understand, and therefore, evaluate. Furthermore, smaller disturbances or disturbances in less densely vegetated forest or mixed vegetation-forest could be missed by the IFZ score if they do not result in enough spectral change to be identified. Likewise, it is also possible that some natural processes such as drought could result in similar spectral changes to smaller disturbances such as pestilence and low severity fire, and thus be identified as disturbance by NAFD.

Conventional, Year-to-Year Data Integration Approach: LANDFIRE

Landscape Fire and Resource Management Planning Tools (LANDFIRE) is a multi-agency collaboration between the US Forest Service (USFS) and the US Department of Interior with a goal of providing ''a common ''all-lands'' data set of vegetation and wildland fire/fuels information for strategic fire and resource management planning and analysis'' across the US (LANDFIRE 2018a). Officially launched in 2002, LANDFIRE was motivated by the ''number, severity, and size of wildland fires'' and is widely used by land management agencies for wildfire planning and mitigation purposes (LANDFIRE 2018a). Currently, LANDFIRE provides a suite of over twenty data products at a 30 m spatial resolution, including maps of vegetation canopy characteristics and fuel types as well as fire regimes and disturbance events.

While GFC and NAFD were produced from modern, automated analysis pipelines via HPC, the LANDFIRE disturbance product represents a more conventional, year-by-year approach to mapping disturbance. Specifically, a separate data layer for each year was independently created by combining all known data of reported disturbance in that year: (1) point locations and perimeters of disturbance events provided by public agencies; (2) vegetation and burn indices calculated from remote sensing analyses of the LTS (for example, NDVI, dNBR, Burned Area Reflectance Classification, Rapid Assessment of Vegetation Condition after Wildfire); and (3) other data integrated from MTBS, the VCT algorithm, and the Multi-Index Integrated Change Algorithm (MIICA) (Jin and others 2013). A custom computing cluster managed by the US Geological Survey (USGS) Earth Resources Observation and Science (EROS) was used to process satellite imagery, calculate vegetation and burn indices, and integrate the data into one raster layer for each year (USGS 2016).

Accuracy assessments of the LANDFIRE products have focused on the vegetation characteristics and fuel data products (Krasnow and others 2009; Hyde and others 2015; Gudex-Cross and others 2017), rather than the disturbance data product, and indicate that LANDFIRE should be supplemented with data collected or analyzed at a local scale whenever possible. LANDFIRE has not relied on specific MMU to develop the vegetation and fuel layers (Rollins 2009), and the disturbance data includes all reported disturbances larger than 0.02 acres in size (approximately 81 sq. m). However, LANDFIRE is self-labeled to provide mid-level products, and as such, encourages the evaluation and use of these products at a coarser resolution of 5 acres, or approximately 22 pixels at a 30 m resolution (LANDFIRE 2018b).

Based on its stated mission to cover all vegetation types across the USA, the primary advantage of LANDFIRE is that does not target a specific vegetation type, such as forest which is the target of both GFC and NAFD, and thus, provides broader coverage of vegetation disturbance across the USA. Furthermore, due to its year-to-year data integration approach, LANDFIRE can easily incorporate both discrete and continuous disturbances because it is a manually curated product that aggregates multiple data sources, including both HPC-based remote sensing analyses and location data on disturbance events collected by public agencies across the USA Therefore, LANDFIRE could include smaller or less severe disturbances that may be missed by GFC and NAFD because those events do not result in significant spectral change. The variety of data used in LANDFIRE's conventional approach also supports pixel-based labeling of a disturbance type, severity, as well as uncertainty based on the data source. It must be noted, however, that the uncertainty label provided by LANDFIRE is not based on an uncertainty analysis but rather perceived uncertainty based on the source and/or identification method of the disturbance.

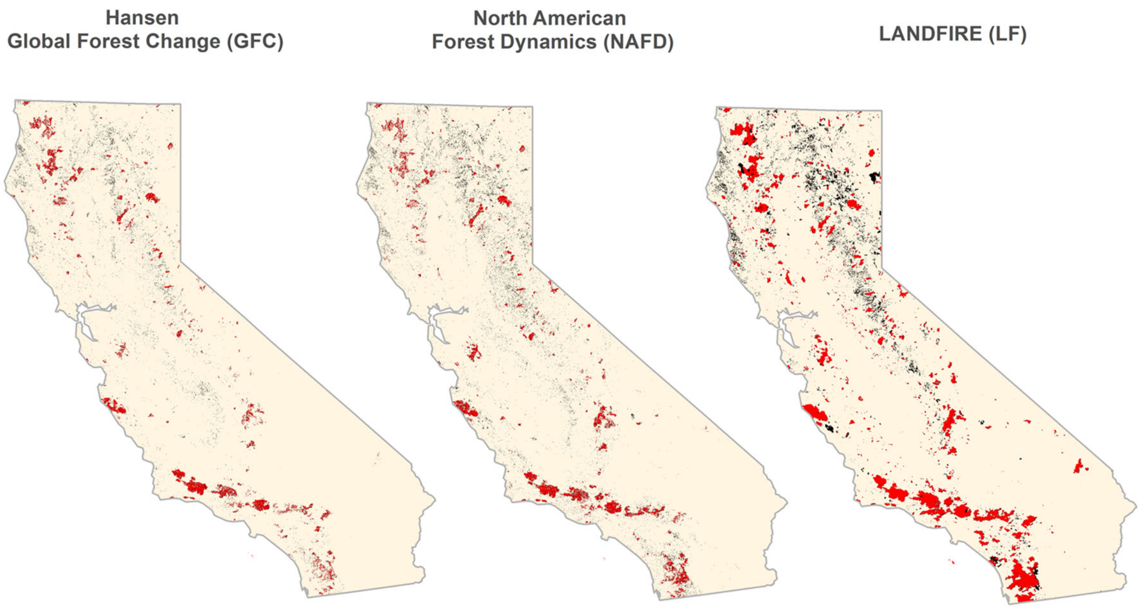
A major limitation of LANDFIRE's approach is that it is neither easily replicatable nor reproducible (because it is not an automated process and aggregates data on a year-to-year basis); this requires end users to devote time querying metadata files and documentation, in order to fully understand how the data are collected and analyzed for a given year or location. Furthermore, there is high potential for compounding data biases and inaccuracies (for example, underestimation or overestimation of disturbance) that are present in the reference datasets that are aggregated into LANDFIRE. For example, disturbance event locations and perimeters reported by public agencies may not always be ground-checked and can be hand-demarcated to include a larger area than the actual footprint of the disturbance. Between years, this curation method could also vary in data quality and accuracy, depending on the data that was received for that year by other public agencies (ranging from local to federal levels).

Methods

Comparison of Reported Disturbance Across California for 2001–2010

To quantify how comparable the three vegetation disturbance products were in their amounts and locations of reported disturbance, we created two sets of comparable raster images from the original GFC, NAFD, and LANDFIRE data using EE to quantify both annual and total reported disturbance. The first set of data contained comparable annual rasters of reported disturbance for each year between 2001 and 2010, while the second set contained comparable aggregated-time rasters of reported disturbance across the study period of 2001–2010. Details on the standardization process used to create these sets of rasters are included in the supplemental material (Online Appendix A). The spatial extent of each product's reported disturbance areas was mapped using the aggregated-time raster created for the product (Figures 1 and 2).

Areas of disturbance attributed to fire were identified by overlaying of the aggregated-time raster created for each disturbance product with a raster of fire occurrence derived from fire perimeters provided by CALFIRE Fire Resource and Assessment Program (FRAP) for the study period (more information on this derived raster of fire occurrence is included in Online Appendix A). We also used the fire occurrence raster derived from FRAP to calculate the total areas of disturbance attributed to fire (m^2/year) for each year and across the study period using the EE function called ee.Image.pixelArea (Figures 1 and 3). This function provides the pixel areas of the two categories in binary images (for example, where pixels both reported as disturbed and overlapping with the fire occurrence are labeled with a value of 1, and all others labeled value of 0). For GFC and NAFD, the sums of these annual values of reported disturbance (both attributable and non-attributable to fire) were equivalent the total area reported as disturbed by each product across 2001–2010 (Figure 1). Due to the annual format of the original LANDFIRE data, pixels could be counted more than once in the sum across the time period (that is, separate disturbances in different years); thus, two sums are provided: the unique area reported as disturbed in the time period (Figure 1) as well as the duplicated total area reported as disturbed (Online Appendix B).



Disturbance 2001-2010	GFC*	NAFD*	LF*
Non-Fire	0.83%	1.64%	2.87%
Fire**	1.71%	2.13%	5.54%
Total	2.54%	3.77%	8.41%

* All values as percent area of California
** Based on overlap with CALFIRE Fire Resource Assessment Program (FRAP) Fire Perimeters

Undisturbed Area
Disturbance (Non-Fire)
Disturbance (Fire)**

Figure 1. Total reported fire and non-fire disturbance across California between 2001 and 2010. Disturbance attributed to fire is based on overlap with fire perimeters from CALFIRE Fire Resource and Assessment Program (FRAP).

We also quantified total reported disturbance by habitat type (for example, how much scrub/shrub GFC reported as disturbed across California for 2001–2010), using the CALFIRE FVEG database to create a raster of four major habitat categories across California: scrub/shrub, forest, grass, and other (for example, desert, agriculture, wetlands, barren, urban). Definitions of the habitat types derived from FVEG are included in Online Appendix A. For these calculations, we applied ee.Image.pixelArea to binary images that combined the aggregated-time raster for each disturbance product and a raster for each of the four habitat types derived from FVEG. These results provided the total pixel area reported as disturbed by each product within each habitat type (Figure 4; Online Appendix B).

Comparison of Environmental Conditions in Areas Reported as Disturbed

Next, we compared the environmental conditions at pixels that were reported as disturbed by each of the three vegetation disturbance products. Based on elevation from the National Elevation Dataset, climate water deficit (CWD) from the California Climate Commons, and mean temperature from the PRISM climate project (see Online Appendix A for more details on these environmental datasets), we produced multiple summary statistics for each disturbance product using the EE functions called ee.Reducer.percentile, ee.Reducer.mean and ee.Reducer.stdDev (Figure 5; Online Appendix B). As a baseline reference, the same statistics were calculated across the total area of California. We also used ee.Image.pixelArea to calculate the total area of each habitat type as a proportion of the spatial extent of the disturbance products (for example, the proportion of the total area reported as disturbed by GFC that was scrub/shrub habitat) (Figure 6; Online Appendix B). As another baseline reference, the areas of each habitat type across California were calculated.

Comparison of Sensitivity to Fire Disturbance

In the final portion of the analysis, we compared the products' sensitivities to fire disturbance by

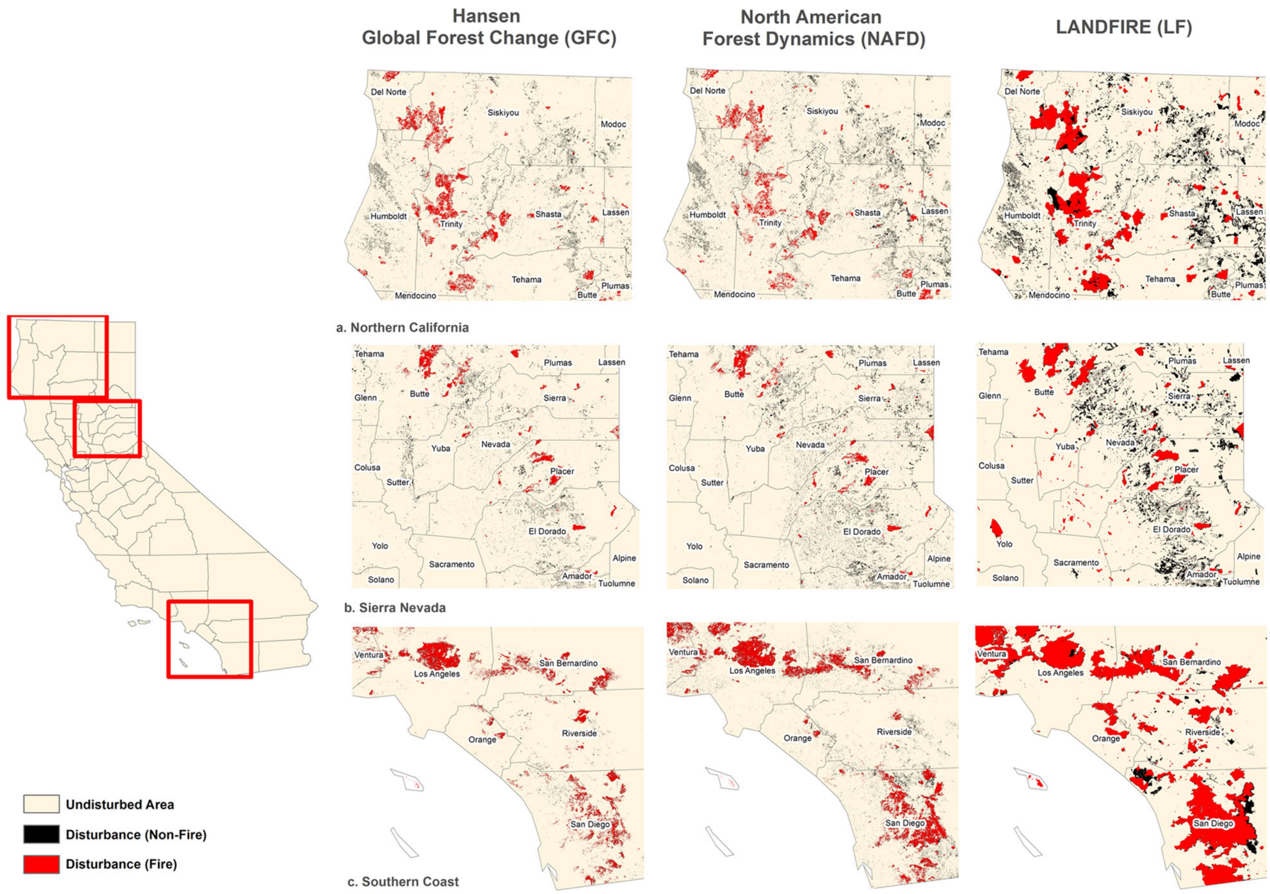


Figure 2. Highlighted areas of spatial differences in reported disturbance for 2001–2010. Spatial differences in reported fire and non-fire disturbance are highlighted across three key regions (Northern California, Sierra Nevada, and Southern Coast).

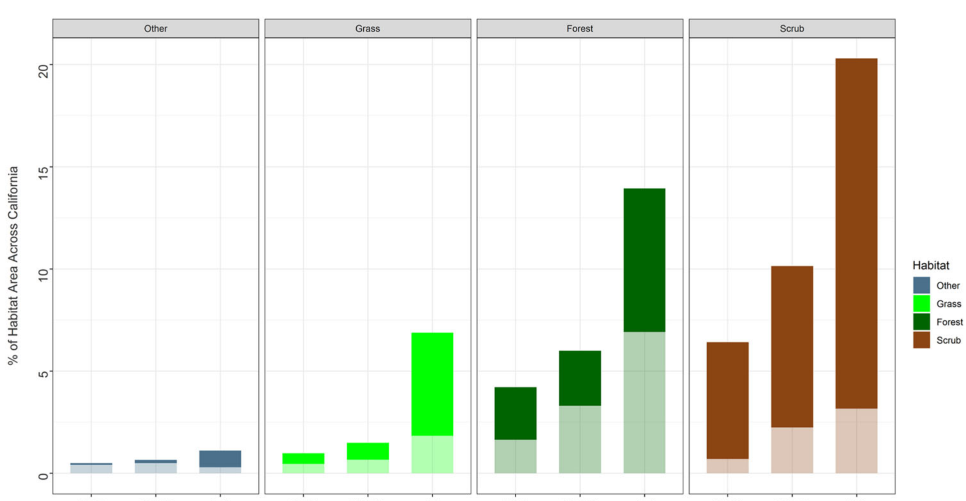


Figure 3. Reported fire and non-fire disturbance by habitat type. Darkest portion of each bar represents proportion of reported disturbed area attributed to fire, based on overlap with FRAP fire occurrence, for Hansen Global Forest Change (GFC), North American Forest Dynamics (NAFD), and LANDFIRE (LF).

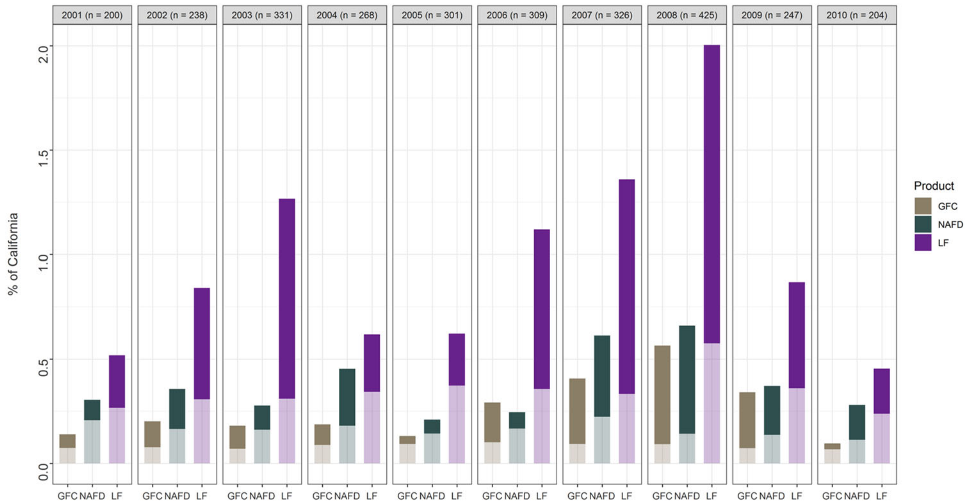


Figure 4. Annual reported fire and non-fire disturbance across California for years 2001–2010. Darkest portion of each bar represents proportion of reported disturbed area attributed to fire, based on overlap with FRAP occurrence, for Hansen Global Forest Change (GFC), North American Forest Dynamics (NAFD), and LANDFIRE (LF). Annual number of fires reported in FRAP database are labeled for each year (for example, n = 200 for 2001).

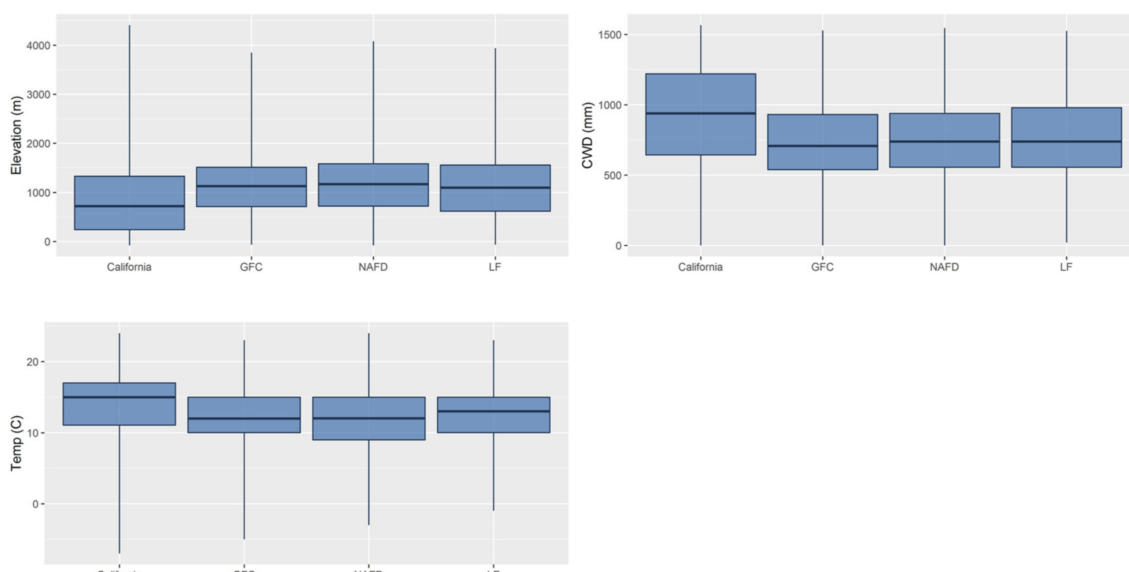


Figure 5. Distributions of bioclimatic conditions across the areas reported as disturbed. For reference comparison, the distributions of bioclimatic conditions across all of California are also reported, alongside distributions for Hansen Global Forest Change (GFC), North American Forest Dynamics (NAFD), and LANDFIRE (LF).

calculating the number of pixels of each product that overlapped with the FRAP fire perimeters and the MTBS maximum burn severity raster (Figures 7 and 8). Since the creation method of LANDFIRE already incorporated versions of the FRAP and MTBS data (see section on Study Data), the primary intention of this analysis was to identify how sensitive GFC and NAFD were to fire disturbance. For a baseline from the fire reference data, we also calculated the total amount of fire disturbance across California (and by habitat type) reported by FRAP (that is, the fire occurrence raster derived from FRAP fire perimeters) and MTBS (that is, raster of maximum burn severity) using ee.Image.pixelArea (Table 2; Online Appendix B). For both fire datasets, we included the year 2000 to account for pixels that may have been reported as disturbed in the first year of the study (2001). As the MTBS data were originally provided as annual rasters, we aggregated them to create a new single

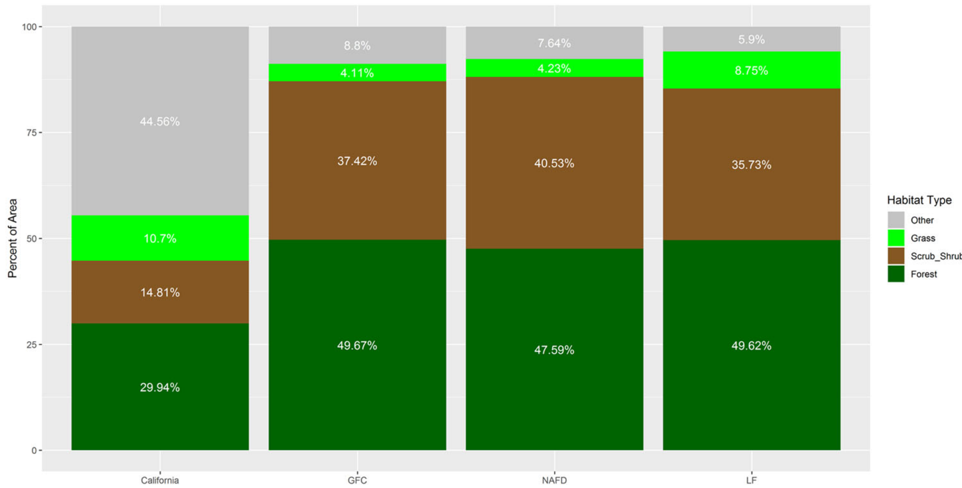


Figure 6. Distribution of habitat type across the areas reported as disturbed. For a reference baseline, total habitat areas as proportions of the area of California are also reported, alongside the proportions for Hansen Global Forest Change (GFC), North American Forest Dynamics (NAFD), and LANDFIRE (LF).

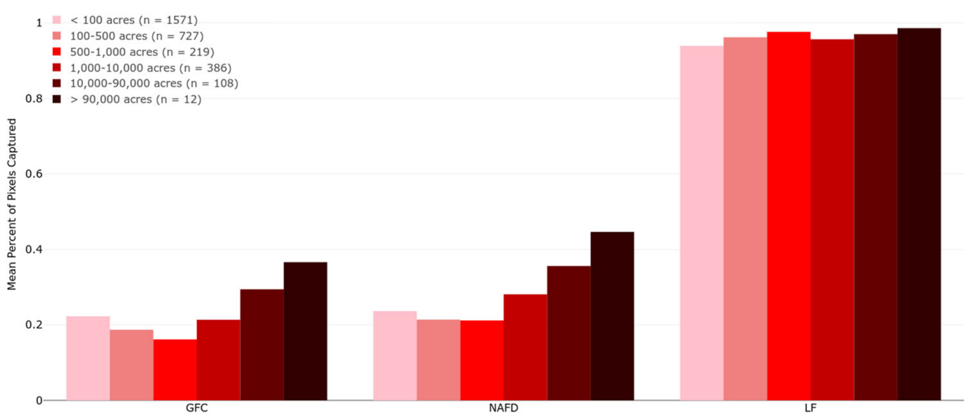


Figure 7. Comparison of sensitivity to fire disturbance across fire perimeter size. Based on overlap with fire perimeters from CALFIRE Fire Resource and Assessment Program (FRAP) for Hansen Global Forest Change (GFC), North American Forest Dynamics (NAFD), and LANDFIRE (LF).

raster that contained the maximum burn severity at each pixel across the study period.

To explore differences in sensitivity to fire disturbance across fire perimeter size, we categorized the fire perimeters into six size classes based on acreage reported by FRAP (that is, less than 100, 100–500, 500–1000, 1000–10,000, 10,000–90,000, greater than 90,000). The overlapping areas between the disturbance products and the fire perimeters were calculated using ee.Reducer.frequencyHistogram, which provided the total number of pixels of each disturbance product contained within each fire perimeter. These pixel counts were converted to percentages by dividing the number of pixels reported as disturbed by each disturbance product by the total number of pixels contained within the fire perimeter. For each size class of fire perimeters, a mean of the percentages was calculated to provide the average percentage of overlap in that size class (Figure 7; Online Appendix B). Last, we used ee.Image.pixelArea to explore differences in sensitivity to fire disturbance across burn severity by calculating the overlap between the disturbance products and each MTBS burn severity level (unburned to low, low, medium, high) across the four habitat types (Figure 8).

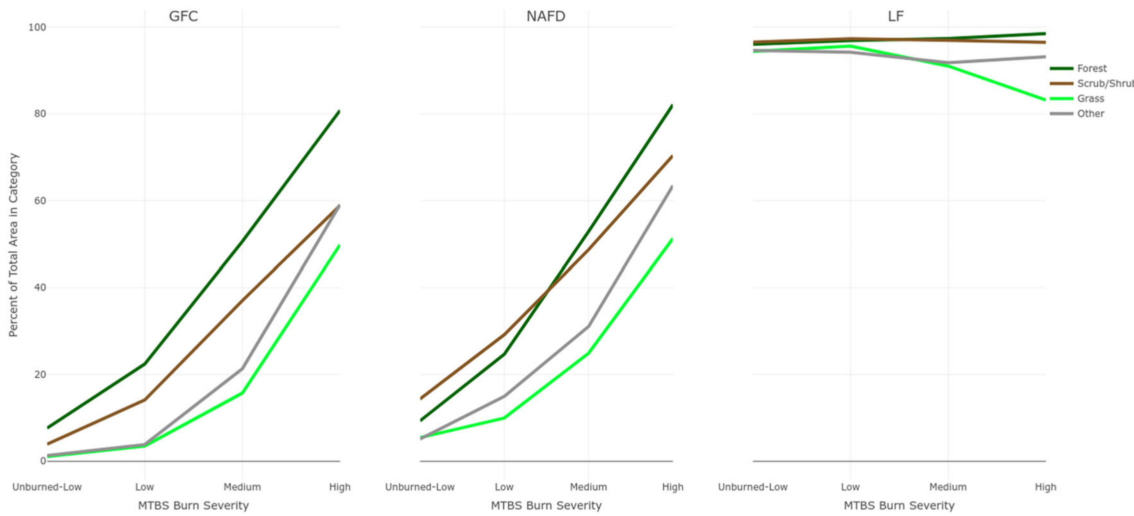


Figure 8. Comparison of sensitivity to fire disturbance across burn severity and by habitat type. Based on overlap with data from Monitoring Trends in Burn Severity (MTBS) for Hansen Global Forest Change (GFC), North American Forest Dynamics (NAFD), and LANDFIRE (LF).

Table 2. Summary of FRAP and MTBS Fire Disturbance Reported for 2000–2010 Across California

	All California (%)	All Scrub/shrub (%)	All Forest (%)	All Grass (%)	All Other (%)
FRAP percent of area burned	5.83	17.87	7.31	5.6	0.89
MTBS percent of area burned	5.77	18.10	7.29	4.99	0.84

Results

LANDFIRE Reported the Highest Amounts of Fire and Non-fire Disturbance Across California

Between 2001 and 2010, GFC and NAFD reported lower totals for fire and non-fire disturbance than LANDFIRE across California (Figure 1). LANDFIRE reported the highest amount of disturbance at 8.41% of the total area of California, while GFC reported the least amount of disturbance at 2.54% of the total area of California. The amount of disturbance reported by NAFD was closer to GFC than LANDFIRE at 3.77% of the total area of California.

The range of values for total disturbance attributed to fire was wider than the range of values for total disturbance attributed to non-fire (Figure 1). For fire disturbance, LANDFIRE reported more than double the disturbance of both GFC and NAFD. Specifically, LANDFIRE reported 5.54% of the total area of California as disturbed by a fire event (determined by overlap with the FRAP fire perimeters), whereas NAFD and GFC reported much less disturbance attributed to fire at only 2.13% and 1.71%, respectively, of the total area of California. For disturbances attributed to non-fire events (determined by no overlap with the FRAP fire perimeters), the products reported more similar totals, with LANDFIRE reporting 2.87% of total area of California as disturbed, and NAFD and GFC reporting 1.64% and 0.83%, respectively.

Spatial differences in reported disturbance among the disturbance products were noticeable across all regions of California (Figure 2). The spatial patterns of fire and non-fire disturbance in the selected regions (Northern California, Sierra Nevada, and Southern Coast) clearly highlighted that more disturbance was reported by LANDFIRE, as compared to GFC or NAFD. The Northern Sierra Nevada was a key area of difference among the products that was attributable to non-fire disturbance, while the Southern Coast was a key area of difference that was attributable to fire disturbance. Northern California was a key area of difference that was attributed to both fire and non-fire disturbance, with LANDFIRE reporting more in both categories as compared to GFC and NAFD. Comparing only GFC and NAFD across the three selected regions, the overall spatial patterns of reported disturbance (Figure 2) reinforced that these two products reported more similar totals of fire disturbance (1.71–2.13%, respectively, as reported in Figure 1), as compared to non-fire disturbance (0.83–1.64%).

The breakdown of total reported disturbance by habitat type also indicated that LANDFIRE reported the highest amounts of disturbance as compared to NAFD and GFC across all habitat types (Figure 3). LANDFIRE reported approximately double the amount of disturbance in scrub/shrub and forest, as compared to both GFC and NAFD (Figure 3). The range of values for total disturbance across the three products was also widest in scrub/shrub (ranging from approximately 6.5–20% of the total scrub/shrub area of California) and in forest (ranging from approximately 4–14% of the total forest area of California) (Online Appendix B). There was also a notable difference in reported disturbance in grass, as GFC and NAFD reported little disturbance in grass (0.98% and 1.49%, respectively), while LANDFIRE reported more than three times these amounts (6.88% of the total grass area of California).

Across all three products, the majority of the reported disturbance in scrub/shrub was attributable to fire (as determined by overlap with the FRAP fire perimeters) (Figure 3). The reported disturbance attributable to fire ranged from 5.7 to 17% of the total scrub/shrub area of California. Reported disturbance in forest was more evenly split between fire and non-fire disturbance (Figure 3; Online Appendix B). For grass, most of the reported disturbance by LANDFIRE was attributable to fire, while the reported disturbances by GFC and NAFD were equally attributable to fire and non-fire disturbance.

Our comparison of total annual disturbance also indicated that LANDFIRE reported the highest amounts of disturbance in each year, whereas GFC and NAFD reported more similar amounts of disturbance across all years (Figure 4; Online Appendix B). NAFD generally reported slightly more disturbance than GFC in each year with the exception of 2006. The widest range of values for annual reported disturbance occurred in 2008, 2003, 2007, and 2006 (in descending order), all years for which FRAP reported the highest annual numbers of individual fire perimeters (between 309 and 425). In these years, LANDFIRE reported much higher disturbance than GFC and NAFD as well as demonstrated the highest overlaps with the FRAP fire perimeters (resulting in higher attribution of disturbance to fire events in these years). GFC and NAFD demonstrated the highest overlaps with the FRAP fire perimeters in 2008, followed by 2007. The smallest range of values for annual reported disturbance occurred in 2010 and 2001, years in which FRAP reported the lowest annual numbers of individual fire perimeters (204 and 200, respectively). In these years, LANDFIRE reported the least amounts of disturbance, and its total disturbance values were the closest to GFC and NAFD, as compared to other years (Figure 4; Online Appendix B).

Minimal Differences in Environmental Conditions Observed Across Areas Reported as Disturbed

To identify the extent to which differences in reported disturbance could be attributed to environmental differences in the areas reported as disturbed, we compared the bioclimatic conditions and habitat types across the areas reported as disturbed by each product. We found that the products reported disturbance across very similar distributions of elevation, climate water deficit (CWD), mean temperature, and habitat type (Figures 5 and 6). In particular, the products reported disturbance at higher elevations, lower CWD, and lower mean temperatures than the California baseline (that is, the full range of conditions across the state) (Figure 5). The distribution of habitat types across the areas reported as disturbed were also similar across the products (Figure 6). As proportions of their total reported disturbance amounts, the products all reported the most disturbance in forest (ranging from approximately 47.5–50% of their total areas) and scrub/shrub (ranging from approximately 36–40% of their total areas). These proportions of disturbed area in forest and scrub/shrub were greater than the overall proportions of those habitat types across California (approximately 30% for forest and 15% for scrub/shrub), indicating that all three products reported disturbance more frequently in those habitat types, as compared to grass.

GFC and NAFD Demonstrated Low Sensitivity to Fire Disturbance Compared to LANDFIRE

As previously noted, the comparison of the products to FRAP and MTBS was primarily intended to evaluate sensitivities to fire disturbance for GFC and NAFD, as LANDFIRE had already incorporated versions of the FRAP and MTBS data and would be expected to display high overlap with both reference datasets. Comparing the reported fire disturbance totals to the baseline summaries of FRAP and MTBS for the study period (Table 2), GFC and NAFD reported notably less fire disturbance than these reference data. The calculated baselines of fire disturbance for the study period from FRAP and MTBS indicated that approximately 5.8% of California was reported as burned (compared to 1.71% for GFC, 2.13% for NAFD, and 5.54% for LANDFIRE) (Online Appendix B). GFC and NAFD also reported less than half of the fire disturbance in scrub/scrub (5.71% and 7.9% of all scrub/shrub habitat, respectively) (Online Appendix B), as compared to FRAP and MTBS, which reported the most fire disturbance in scrub/shrub (approximately 18% of all scrub/shrub habitat across California) (Table 2). Similarly, GFC and NAFD reported less than half of the fire disturbance in forest (2.57% and 2.69%, respectively), as compared to FRAP and MTBS, which reported approximately 7% of all forest habitat across California as burned.

In our direct comparison of each product's overlap with the FRAP fire perimeters, we found that GFC and NAFD excluded more individual fire perimeters (n = 1250 and n = 1083, respectively), than LANDFIRE (n = 141). Despite the difference in the total number of fires excluded, the median fire perimeter size excluded by the products were similar in range (from 35.59 to 41.64 acres). Of the fire perimeters that were overlapped by the products (by at least one pixel), GFC and NAFD demonstrated similar percent overlap with the FRAP fire perimeters that were lower than LANDFIRE across all perimeter sizes (Figure 7). For FRAP fire perimeters larger than 1000 acres, the mean percent of pixels reported as disturbed within the fire perimeters increased with size for both GFC and NAFD, but reached a maximum percent of approximately 40% for the largest fire perimeter size (Figure 7; Online Appendix B).

Similar to its overlap with FRAP, the overlap between LANDFIRE and MTBS was close to 100% across all burn severity levels (Figure 8; Online Appendix B). Both GFC and NAFD demonstrated higher overlap with MTBS (which focuses on fire perimeters larger than 1000 acres), particularly at the medium and high severity categories (Figure 8). For the MTBS data, both GFC and NAFD reported more disturbance as burn severity increased to the medium and high severity classes (Figure 8; Online Appendix B). This pattern was demonstrated across all habitat types but was most noticeable for GFC and NAFD in forest and NAFD in scrub/shrub. At medium severity, both GFC and NAFD demonstrated an approximately 50% overlap with MTBS for forest. NAFD demonstrated a similar level of overlap with MTBS for scrub/shrub at medium severity (approximately 50%). Both GFC and NAFD reported the most disturbance in forest at the highest severity (approximately 80% of the total forest that MTBS reported as burned with high severity).

Discussion

Differing Sensitivity to Fire Disturbance is a Key Driver of Difference in Reported Disturbance

Our comparative evaluation demonstrated that GFC and NAFD reported similar amounts of disturbance that were consistently much lower than LANDFIRE's reported disturbance across all years, regions, and habitats (Figures 1, 2, 3 and 4). We also found that despite these differences in the amounts of reported disturbance, the products identified disturbance in similar ranges of bioclimatic conditions and habitat types; thus, differing environmental conditions in the areas reported as disturbed were not the drivers of the differences in reported disturbance (Figures 5 and 6). Rather, we found that lower sensitivity to fire disturbance for GFC and NAFD, as compared to LANDFIRE, was a key driver of the overall differences in the amounts and locations of reported disturbance. Specifically, both GFC and NAFD reported much lower amounts of fire disturbance across all FRAP fire perimeter size classes and all MTBS burn severity classes (Figures 7 and 8). Furthermore, the difference in reported disturbance between LANDFIRE and GFC/NAFD was greater for fire disturbance; in particular, LANDFIRE reported more than double than amounts of GFC and NAFD across California in the study period (Figure 1).

Although it was expected that LANDFIRE would demonstrate high overlap with FRAP and MTBS due to these types of reference datasets playing a key role in the creation of LANDFIRE, it was not previously known the extent to which LANDFIRE's reported disturbance would differ from that reported by disturbance products created from ''big'' data approaches (for example, GFC and NAFD). Based on the results of this study, it is clear that the automated, time series approaches of remotely sensed products such as GFC and NAFD resulted in much lower reporting of fire disturbance than the year-to-year approach of LANDFIRE that integrates remote sensing-derived vegetation indices with field data and other data on disturbance events reported by public agencies. Furthermore, due to the lower sensitivity to fire disturbance by GFC and NAFD, the differences in reported disturbance between these products and LANDFIRE were greatest in the years with the most fire (Figure 4) and in scrub/shrub habitat (Figure 3) for which both FRAP and MTBS reported the most fire (Table 2).

Comparing only GFC and NAFD, the latter reported more disturbance across all regions and habitats and in most years (Figures 1, 2, 3 and 4). The higher reporting of disturbance by NAFD reflects a key difference in their automated approaches, namely that the approach taken by NAFD is to identify continuous change in vegetation cover (that is, reductions in vegetation), rather than to identify only discrete changes that would cause stand replacement, as is the approach taken by GFC. Notably, this difference in reported disturbance was not observed in the comparison of the products across the burn conditions, as both GFC and NAFD reported similar overlaps with FRAP and MTBS (Figures 7 and 8). This result indicates that for fire disturbance, the differences in continuous and discrete approaches did not play a strong role in its identification, as both products reported equally less fire disturbance than LANDFIRE.

Implications for Use of These Vegetation Disturbance Products

Given their differences in reported disturbance, the lack of comparative evaluations of these products (GFC, NAFD, and LANDFIRE) is a notable omission in the scientific literature, as these products are widely used as sole representations of disturbance in studies on the impacts of vegetation change and disturbance on ecosystem processes such as carbon. For example, GFC has been used to examine the impacts of forest change on carbon dynamics both globally (Tyukavina and others 2015; Arneth and others 2017) and within the USA (Anderegg and others 2016; Woodall and others 2016). NAFD has also been frequently used to explore the impacts of forest disturbance on carbon dynamics within the USA (Gu and others 2016; Williams and others 2016; Dolan and others 2017; Sleeter and others 2018). LANDFIRE has been applied more broadly across landscapes in the USA to explore impacts of past disturbance on hydrology (Boisramé and others 2017), subsequent fire (Parks and others 2014) as well as carbon dynamics, specifically in California (Liu and others 2011; Gonzalez and others 2015). As our analysis indicates that these products reported notable differences in the amounts of disturbance (from GFC at the low end to LANDFIRE at the high end), researchers need to be aware that the choice of the disturbance product can greatly impact the results of their studies aimed at quantifying the ecological impacts of disturbance events.

The applications for remotely sensed disturbance products are not restricted to carbon-related studies; rather, researchers are increasingly highlighting the potential use of these kinds of products to quantify and monitor the impacts of disturbance on ecosystem functions, ecosystem response and resilience, and species abundance and distribution, while also recognizing that disturbances are not limited to changes in land cover (Rose and others 2015; Pettorelli and others 2014). These researchers declare a desire and need for remotely sensed products that can offer standardized time series of a variety of types and magnitudes of disturbance that clearly indicate where disturbance has occurred and how much across multiple types of habitats and environmental conditions. Our analysis of the current state of vegetation disturbance products indicates that more work is needed to quantify uncertainty across these products in a way that can be useful to conservation and resource managers, who want to know which product(s) to use in specific habitat and ecosystem types.

Recommending a product solely on its stated goals and purposes, each product provides a unique spatial and temporal coverage of vegetation disturbance in specific habitats. At the time of this publication, GFC is the only remotely sensed product that maps annual forest change at a global extent with the spatial resolution of the LTS and is the clear choice for studies of forest change at global and continental scales outside of North America. For North American and specifically, USA-centric studies, NAFD provides more nuanced coverage of forest disturbance from low to high magnitude events (that is, from minimal reductions in vegetation cover up to clear-cutting and land use changes). Within the USA, LANDFIRE uniquely provides coverage of disturbance in scrub/shrub and grass, which highlights the need for new products that include coverage of disturbance in these habitats at global and continental scales.

Although these recommendations appear clearly defined—that users should choose the appropriate disturbance product based on its spatial extent and targeted habitat—our analysis clearly identified that the differences between LANDFIRE and GFC/NAFD were greatest for reported fire disturbance, indicating that the choice of disturbance product is of more consequence when the targeted disturbance is fire, rather than non-fire disturbances (for example, clearing, pestilence). Given this key difference, the choice of disturbance product becomes not so straightforward, for example, for research aimed at identifying the impacts of fire-specific disturbance on forest in the USA, as either NAFD or LANDFIRE could be used. Furthermore, as GFC and NAFD reported similar amounts of forest disturbance, the convergence between the two could indicate a more accurate amount of both fire and non-fire forest disturbance, highlighting a potential overestimation of forest disturbance by LANDFIRE (Figure 3).

Furthermore, all three products actually reported the most disturbance in areas dominated by scrub/shrub habitat (Figure 3). Given that GFC and NAFD do not aim to identify disturbance in scrub/shrub, it would seem that their amounts of reported disturbance in scrub/shrub should be lower than the amount in forest. Although scrub/shrub habitats often demonstrate high spectral variability (Hamada and others 2011) which could be interpreted as disturbance by automated products, scrub/shrub was also most frequently reported as disturbed by the fire reference data (FRAP and MTBS), thus indicating the possibility that the automated products (GFC and NAFD) were accurately identifying higher fire disturbance in scrub/shrub as compared to forest.

The fact that the amount of disturbance in scrub/shrub reported by GFC and NAFD was lower than that of LANDFIRE was not surprisingly; however, the difference in reported forest disturbance was unexpected as GFC and NAFD would presumably be more accurate for forest, given their stated foci. One possible explanation is that GFC and NAFD are specifically missing smaller disturbances that do not cause stand replacement (that is, change to a new cover class) or do not cause significant reduction in vegetation cover, which, respectively, would be needed for their workflows to identify the disturbance. However, the size of disturbance does not appear to be a driver of identification for fire disturbance. Although the overall number of excluded fires was high for GFC and NAFD compared to LANDFIRE, the median fire size excluded was similar across the three products, and GFC and NAFD had low overlap with the FRAP fire perimeters across all size classes. This indicates that size of the fire was not a driver of difference in reported fire disturbance among the products; GFC and NAFD were simply less sensitive to fire disturbance overall as compared to LANDFIRE.

However, it must be noted that LANDFIRE's high overlap with FRAP and MTBS does not mean that it is more accurately reporting fire disturbance. As previously discussed, the creation method of LANDFIRE includes ingestion of data directly from reference data sources such as FRAP and MTBS. These data have their own biases and uncertainties that propagate to other workflows that used them as reference data. For example, researchers examining the accuracy of fire reference data have found that manually mapped fire perimeters (such as the vector data provided by FRAP) can overestimate the burned area by an average of 18% (up to 37%), as compared to fire perimeters delineated using remote sensing techniques (Kolden and others 2012; Kolden and Weisberg 2007). Similarly, other researchers have found that independent classification of burned areas resulted in higher accuracy than MTBS data (Meddens and others 2016). Due to LANDFIRE's ingestion of these products, overestimation in reference data such as FRAP and MTBS would also result in higher reported disturbance by LANDFIRE, the extent to which would be unknown without further accuracy and uncertainty analyses.

In addition, the MMU of the reference data can also result in biases and uncertainties in the analyses of the disturbance products. For example, fire disturbance in California grasslands can be widely underreported, due to the MMU of reference data such as the FRAP fire perimeters, which only includes grass fires over 300 acres (FRAP 2018). This limitation of the reference data is supported by our results which indicate grasslands were reported as disturbed less often than forest and scrub/shrub by all of the products, including LANDFIRE (Figures 3 and 8), which is the only one of three products that is actually applicable for mapping disturbance in grasslands in the USA. Interestingly, while the MMU of FRAP in forest (used by both CALFIRE and USFS) is smaller than the MMU for scrub/shrub (10 and 30 acres, respectively), all products reported more fire disturbance in scrub/shrub, indicating that the larger MMU of FRAP in scrub/shrub did not limit the reporting of fire disturbance in scrub/shrub for any of the products. Regarding the MMU of the MTBS data (fires larger than 1000 acres for the Western USA), our results indicated that reported fire disturbance by GFC and NAFD increased with burn severity (Figure 8); however, those results were not matched by our comparison of reported disturbance by FRAP fire perimeter size, for which percent coverage was low, even for fire perimeters larger than 1000 acres (Figure 7). This discrepancy indicates that the analysis of fire disturbance based on fire perimeter size is highly dependent on the input reference data, and thus, more than one reference dataset should be used as the reference point for any evaluation.

Based on our results, we encourage end users to choose the appropriate disturbance product based not only on spatial extent and habitat but also on the disturbance type of interest (that is, fire and non-fire). Furthermore, given the lack of comparative analyses of these products, additional work is needed to quantify and identify the spatial patterns of uncertainty in disturbance across these products with careful consideration to use more than one reference dataset for evaluation. While there has been some research on integrating some of these products (Schroeder and others 2017; Soulard and others 2017) or the algorithms used to create them (Healey and others 2017) to improve the accuracy of disturbance identification (that is, potentially reducing uncertainty), these integration efforts have not been used to quantify uncertainty across study areas and have typically focused only on areas in which at least two products or algorithms identified disturbance. While using data integration to improve the accuracy of identified disturbance contributes greatly to the literature, quantifying spatially explicit measures of uncertainty would provide end users with additional critical information to support their decision-making regarding which product or products would best meet their needs.

One possible method for quantifying spatially explicit uncertainty is through data integration based on spatial agreement. For example, pixels that are reported as disturbed by all products (that is, pixels of highest agreement) and are contained within a fire perimeter could be assigned a low uncertainty for a fire disturbance. Similarly, pixels that are reported as disturbed by all products (again, pixels of highest agreement) but are not contained within a fire perimeter could be assigned a low uncertainty for a non-fire disturbance. On the other hand, pixels that are not reported as disturbed in any product but are contained within a fire perimeter in reference data could be assigned a high uncertainty for fire disturbance (that is, likely unburned area within the fire perimeter), helping to narrow down the true extent of fire events. These kind of spatial agreement metrics could provide a more automated and more objective identification of unburned areas (as compared to dNBR analyses which require defining thresholds for disturbance that vary by vegetation and ecosystem type) as well as highlight areas where individual disturbance products may be overzealous in reporting disturbance (that is, errors of commission).

Integral Role of Spatial Data Science and HPC in Creating Disturbance Products

Previous to this study, comparative evaluations of disturbance products had likely been limited by two related factors: (1) the lack of overlap in products' coverage across space and time (that is, limited and non-overlapping spatial and temporal extents); and (2) the inadequacy of conventional analytical tools to handle and analyze data at increasingly finer resolutions and broader extents. In other words, remotely sensed products of vegetation disturbance (with overlapping spatial and temporal coverage at the spatial resolution of LTS) are recent developments that have been made possible by the creation of computational tools used to create and compare them. In particular, the recent development of HPC tools specifically intended for geospatial analyses (such as EE) has helped to address computational challenges presented by large spatial–temporal data and has been encouraged by CyberGIS researchers interested in expanding tool interoperability and scalability for 'big' spatial data (Yang and others 2010, 2011; Wang 2016). Alongside these technological advances, integration of conventional geospatial methods and modern Data Science techniques (that is, data mining/algorithms, machine learning) have arisen from the development of a Spatial Data Science to support the application of fundamental geospatial analyses to ''big'' spatial–temporal data stacks such as the LTS (Palomino and others 2017).

These developments in Spatial Data Science and HPC have resulted in both the creation of products at finer spatial and temporal resolutions and broader extents as well as an increased ability to compare and evaluate them. Even while limited to the California scale, the disturbance products evaluated in this study are ''big'' data, as approximately 450 million pixels were analyzed for each year to cover California at a 30 m spatial resolution. Just as the processing power and data handling capabilities of HPC were needed to create these disturbance products, a thorough interrogation and evaluation of these products also required the use of HPC to identify patterns across this complex multi-temporal data stack. The platform employed in this study, EE, is an exemplary tool emerging from these advances in HPC and Spatial Data Science, as it supports fundamental geospatial analyses such as raster stack calculations and zonal statistics on data that are not easily handled in conventional desktop tools.

As products of self-contained analysis pipelines or algorithms running autonomously on HPC, GFC and NAFD are the first examples of standardized, annual products that leverage the spatial and temporal resolutions of the LTS and are very likely not the last of their kind. Other algorithms that identify vegetation disturbance using the LTS, such as LandTrendr (Kennedy and others 2010) and Continuous Change Detection and Classification (CCDC) (Zhu and Woodcock 2014), can be leveraged to produce new standardized products. In fact, CCDC is currently under evaluation and validation for the release of a new standardization product to be distributed by the USGS (Pengra and others 2016). Furthermore, these algorithms and others could potentially be expanded to other satellite time series such as Sentinel to produce products with finer spatial resolutions, as compared to the LTS, in the future. In contrast to the lengthy protocols and manual data integration of a product like LANDFIRE, these automated products streamline the identification of disturbance by focusing exclusively on changes in spectral characteristics using a data science approach that does not require a priori knowledge of disturbance events. Moving forward, these modern, automated workflows and the products they create can also enable regional (and possibly global) analyses of fire return intervals and dynamics of burn intensity to identify generalizable trends (Stevens and others 2017), beyond local analyses of individual fires (Collins and others 2007, 2009).

Conclusion

We used Earth Engine to compare the reported amounts of fire and non-fire disturbance for 2001–2010 among three widely used vegetation disturbance products and examined the products' reported disturbance across differing environmental and burn conditions. Overall, GFC and NAFD reported smaller totals for disturbance than LANDFIRE (2.54%, 3.77%, and 8.41% of California, respectively) as well as less disturbed area attributable to fire (1.71%, 2.13%, and 5.55% of California) across the study period of 2001–2010. Despite differences in amounts of reported disturbance, the products identified disturbance in similar ranges of bioclimatic conditions and habitat types. Thus, differing environmental conditions in areas reported as disturbed were not major drivers of difference; rather, lower sensitivity to fire disturbance for GFC and NAFD, as compared to LANDFIRE, was a key driver of the overall differences in the amounts and locations of reported disturbance. In particular, both GFC and NAFD reported much lower amounts of fire disturbance than LANDFIRE across all FRAP fire perimeter size classes and MTBS burn severity classes. Furthermore, the difference in reported disturbance between LANDFIRE and GFC/NAFD was greater for fire disturbance than for non-fire disturbance; LANDFIRE reported more than double the total amounts of fire disturbance of GFC and NAFD across the study period.

Designed and executed within EE, our methodology provides a reproducible framework for comparative analyses of vegetation disturbance products to identify the conditions under which they report disturbance. Rather than focusing on accuracy, this comparative examination of reported disturbance highlights the drivers of differences (that is, uncertainty) in reported disturbance among vegetation disturbance products. As an illustrative example, this paper used fire in California as a case study to help end users of these products understand how their approaches to identifying disturbance impact the amounts and locations of reported disturbance. As we found that the differences among the products were greatest for fire, rather than non-fire disturbance, we recommend that users choose a disturbance product based on spatial extent, targeted habitat as well as disturbance type (that is, fire and non-fire disturbance). Additional work to quantify uncertainty in disturbance across these products and identify spatial patterns in uncertainty is needed to further support end users in choosing the most appropriate product or products for their needs.

Acknowledgements

The authors would like to thank the Geospatial Innovation Facility (GIF), Matthew Potts and Iryna Dronova at the University of California, Berkeley.

References

Anderegg WRL, Martinez-Vilalta J, Cailleret M, Camarero JJ, Ewers BE, Galbraith D, Gessler A, Grote R, Huang C-Y, Levick SR, Powell TL, Rowland L, Sánchez-Salguero R, Trotsiuk V. 2016. When a tree dies in the forest: scaling climate-driven tree mortality to ecosystem water and carbon fluxes. Ecosystems 19:1133–47.

Arneth A, Sitch S, Pongratz J, Stocker BD, Ciais P, Poulter B, Bayer AD, Bondeau A, Calle L, Chini LP, Gasser T, Fader M, Friedlingstein P, Kato E, Li W, Lindeskog M, Nabel JEMS, Pugh TAM, Robertson E, Viovy N, Yue C, Zaehle S. 2017. Historical carbon dioxide emissions caused by land-use changes are possibly larger than assumed. Nat Geosci 10:79.

Boisramé G, Thompson S, Collins B, Stephens S. 2017. Managed wildfire effects on forest resilience and water in the Sierra Nevada. Ecosystems 20:717–32.

Cohen WB, Healey SP, Yang Z, Stehman SV, Brewer CK, Brooks EB, Gorelick N, Huang C, Hughes MJ, Kennedy RE, Loveland TR, Moisen GG, Schroeder TA, Vogelmann JE, Woodcock CE, Yang L, Zhu Z. 2017. How similar are forest disturbance maps derived from different Landsat time series algorithms? For Trees Livelihoods 8:98.

Collins BM, Kelly M, van Wagtendonk JW, Stephens SL. 2007. Spatial patterns of large natural fires in Sierra Nevada wilderness areas. Landsc Ecol 22:545–57.

Collins BM, Miller JD, Thode AE, Kelly M, van Wagtendonk JW, Stephens SL. 2009. Interactions among wildland fires in a long-established Sierra Nevada Natural fire area. Ecosystems 12:114–28.

Dolan KA, Hurtt GC, Flanagan SA, Fisk JP, Sahajpal R, Huang C, Le Page Y, Dubayah R, Masek JG. 2017. Disturbance distance: quantifying forests' vulnerability to disturbance under current and future conditions. Environ Res Lett 12:114015.

FRAP. 2018. FRAP fire perimeters product metadata. http://frap.fire.ca.gov/data/frapgisdata-sw-fireperimeters_download. Last accessed 20 Nov 2018.

Gonzalez P, Battles JJ, Collins BM, Robards T, Saah DS. 2015. Aboveground live carbon stock changes of California wildland ecosystems, 2001–2010. For Ecol Manag 348:68–77.

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens Environ. http://www.sciencedirect.com/science/article/pii/S0034425717302900.

Goward SN, Huang C, Zhao F, Schleeweis K, Rishmawi K, Lindsey M, Dungan JL, Michaelis A. 2016. NACP NAFD project: forest disturbance history from Landsat, 1986–2010. ORNL DAAC, Oak Ridge, Tennessee, USA. http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1290. Last accessed 11 Feb 2017.

Goward SN, Masek JG, Cohen W, Moisen G, Collatz GJ, Healey S, Houghton RA, Huang C, Kennedy R, Law B, Powell S. 2008. Forest disturbance and North American carbon flux. Eos Trans Am Geophys Union 89(11):105–6.

Gudex-Cross D, Pontius J, Adams A. 2017. Enhanced forest cover mapping using spectral unmixing and object-based classification of multi-temporal Landsat imagery. Remote Sens Environ 196:193–204.

Gu H, Williams CA, Ghimire B, Zhao F, Huang C. 2016. High-resolution mapping of time since disturbance and forest carbon flux from remote sensing and inventory data to assess harvest, fire, and beetle disturbance legacies in the Pacific Northwest. Biogeosciences 13:6321–37.

Hamada Y, Stow DA, Roberts DA. 2011. Estimating life-form cover fractions in California sage scrub communities using multispectral remote sensing. Remote Sens Environ 115:3056.

Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, Tyukavina A, Thau D, Stehman SV, Goetz SJ, Loveland TR, Kommareddy A, Egorov A, Chini L, Justice CO, Townshend JRG. 2013. High-resolution global maps of 21st-century forest cover change. Science 342:850–3.

Healey SP, Cohen WB, Yang Z, Kenneth Brewer C, Brooks EB, Gorelick N, Hernandez AJ, Huang C, Joseph Hughes M, Kennedy RE, Loveland TR, Moisen GG, Schroeder TA, Stehman SV, Vogelmann JE, Woodcock CE, Yang L, Zhu Z. 2017. Mapping forest change using stacked generalization: An ensemble approach. Remote Sens Environ. http://www.sciencedirect.com/science/article/pii/S0034425717304418.

Hermosilla T, Wulder MA, White JC, Coops NC, Hobart GW, Campbell LB. 2016. Mass data processing of time series Landsat imagery: pixels to data products for forest monitoring. Int J Digit Earth 9:1035–54.

Huang C, Goward SN, Masek JG, Thomas N, Zhu Z, Vogelmann JE. 2010. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. Remote Sens Environ 114:183–98.

Hyde J, Strand EK, Hudak AT, Hamilton D. 2015. A case study comparison of landfire fuel loading and emissions generation on a mixed conifer forest in northern Idaho, USA. Fire Ecol 11. https://www.fs.fed.us/rm/pubs_journals/2015/rmrs_2015_hyde_j001.pdf.

Jin S, Yang L, Danielson P, Homer C, Fry J, Xian G. 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. Remote Sens Environ 132:159–75.

Kalluri S, Gundy J, Haman B, Paullin A, Van Rompay P, Vititoe D, Weiner A. 2015. A High performance remote sensing product generation system based on a service oriented architecture for the next generation of geostationary operational environmental satellites. Remote Sens 7:10385–99.

Kang S, Lee K. 2016. Auto-scaling of geo-based image processing in an OpenStack cloud computing environment. Remote Sens 8:662.

Keane RE, Rollins M, Zhu Z-L. 2007. Using simulated historical time series to prioritize fuel treatments on landscapes across the United States: the LANDFIRE prototype project. Ecol Model 204:485–502.

Kennedy RE, Yang Z, Cohen WB. 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—temporal segmentation algorithms. Remote Sens Environ. http://www.sciencedirect.com/science/article/pii/S0034425710002245.

Kolden CA, Lutz JA, Key CH, Kane JT, van Wagtendonk JW. 2012. Mapped versus actual burned area within wildfire perimeters: characterizing the unburned. For Ecol Manag 286:38.

Kolden CA, Weisberg PJ. 2007. Assessing accuracy of manually-mapped wildfire perimeters in topographically dissected areas. Fire Ecol 3:22.

Krasnow KD, Fry DL, Stephens SL. 2017. Spatial, temporal and latitudinal components of historical fire regimes in mixed conifer forests, California. J Biogeogr 44:1239–53.

Krasnow K, Schoennagel T, Veblen TT. 2009. Forest fuel mapping and evaluation of LANDFIRE fuel maps in Boulder County, Colorado, USA. For Ecol Manag 257:1603–12.

Kumar U, Ganguly S, Nemani RR, Raja KS, Milesi C, Sinha R, Michaelis A, Votava P, Hashimoto H, Li S, Wang W, Kalia S, Gayaka S. 2017. Exploring subpixel learning algorithms for estimating global land cover fractions from satellite data using high performance computing. Remote Sens 9:1105.

LANDFIRE disturbance product metadata. https://landfire.cr.usgs.gov/distmeta/servlet/gov.usgs.edc.MetaBuilder?TYPE=HTML&DATASET=FA9. Last accessed 6 Jan 2017.

LANDFIRE. 2018a. About web page. https://www.landfire.gov/about.php. Last accessed on 20 Nov 2018.

LANDFIRE. 2018b. Guidelines for evaluating LANDFIRE fuel data. https://www.landfire.gov/documents/EvaluatingLANDFIREFuelsData.pdf. Last accessed on 20 Nov 2018.

Lee CA, Gasster SD, Plaza A, Chang CI, Huang B. 2011. Recent developments in high performance computing for remote sensing: a review. IEEE J Sel Top Appl Earth Obs Remote Sens 4:508–27.

Liu J, Vogelmann JE, Zhu Z, Key CH, Sleeter BM, Price DT, Chen JM, Cochrane MA, Eidenshink JC, Howard SM, Bliss NB, Jiang H. 2011. Estimating California ecosystem carbon change using process model and land cover disturbance data: 1951–2000. Ecol Model 222:2333–41.

Masek JG, Goward SN, Kennedy RE, Cohen WB, Moisen GG, Schleeweis K, Huang C. 2013. United States forest disturbance trends observed using Landsat time series. Ecosystems 16:1087–104.

McKerrow A, Dewitz J, Long DG, Nelson K, Connot JA. 2016. A comparison of NLCD 2011 and LANDFIRE EVT 2010: regional and national summaries. https://pubs.er.usgs.gov/publication/70177839.

Meddens AJ, Kolden CA, Lutz JA. 2016. Detecting unburned areas within wildfire perimeters using Landsat and ancillary data across the northwestern United States. Remote Sens Environ 186:275.

Moritz MA, Stephens SL. 2008. Fire and sustainability: considerations for California's altered future climate. Clim Change. http://www.springerlink.com/index/5411702235mx5432.pdf.

Nemani R, Votava P, Michaelis A, Melton F, Milesi C. 2011. Collaborative supercomputing for global change science. Eos Trans AGU 92:109–10.

Palomino J, Muellerklein OC, Kelly M. 2017. A review of the emergent ecosystem of collaborative geospatial tools for addressing environmental challenges. Comput Environ Urban Syst 65:79–92.

Parks SA, Miller C, Nelson CR, Holden ZA. 2014. Previous fires moderate burn severity of subsequent wildland fires in two large western US wilderness areas. Ecosystems 17:29–42.

Pengra B, Gallant AL, 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 US Geological Survey. Remote Sens 8:811.

Pettorelli N, Laurance WF, O'Brien TG, Wegmann M, Nagendra H, Turner W. 2014. Satellite remote sensing for applied ecologists: opportunities and challenges. J Appl Ecol 4:839.

Plaza AJ, Chang C-I. 2007. High performance computing in remote sensing. London: Chapman & Hall/CRC.

Reeves MC, Ryan KC, Rollins MG, Thompson TG. 2009. Spatial fuel data products of the LANDFIRE project. Int J Wildland Fire 18:250–67.

Rollins MG. 2009. LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment. Int J Wildland Fire 18:235–49.

Rose RA, Byler D, Eastman JR, Fleishman E, Geller G, Goetz S, Guild L, Hamilton H, Hansen M, Headley R, Hewson J. 2015. Ten ways remote sensing can contribute to conservation. Conserv Biol 2:350.

Ryan KC, Opperman TS. 2013. LANDFIRE—a national vegetation/fuels data base for use in fuels treatment, restoration, and suppression planning. For Ecol Manag 294:208–16.

Schroeder TA, Schleeweis KG, Moisen GG, Toney C, Cohen WB, Freeman EA, Yang Z, Huang C. 2017. Testing a Landsat-based approach for mapping disturbance causality in U.S. forests. Remote Sens Environ 195:230–43.

Sleeter BM, Liu J, Daniel C, Rayfield B, Sherba J, Hawbaker TJ, Zhu Z, Selmants PC, Loveland TR. 2018. Effects of contemporary land-use and land-cover change on the carbon balance of terrestrial ecosystems in the United States. Environ Res Lett 13:045006.

Soulard CE, Acevedo W, Cohen WB, Yang Z, Stehman SV, Taylor JL. 2017. Harmonization of forest disturbance datasets of the conterminous USA from 1986 to 2011. Environ Monit Assess 189:170.

Soulard CE, Albano CM, Villarreal ML, Walker JJ. 2016. Continuous 1985–2012 Landsat monitoring to assess fire effects on meadows in Yosemite National Park, California. Remote Sens 8:371.

Stephens SL, Martin RE, Clinton NE. 2007. Prehistoric fire area and emissions from California's forests, woodlands, shrublands, and grasslands. For Ecol Manag 251:205–16.

Stevens JT, Collins BM, Miller JD, North MP, Stephens SL. 2017. Changing spatial patterns of stand-replacing fire in California conifer forests. For Ecol Manag 406:28–36.

Thomas NE, Huang C, Goward SN, Powell S, Rishmawi K, Schleeweis K, Hinds A. 2011. Validation of North American forest disturbance dynamics derived from Landsat time series stacks. Remote Sens Environ 115:19–32.

Tyukavina A, Baccini A, Hansen MC, Potapov PV, Stehman SV, Houghton RA, Krylov AM, Turubanova S, Goetz SJ. 2015. Aboveground carbon loss in natural and managed tropical forests from 2000 to 2012. Environ Res Lett 10:074002.

USGS. 2016. Earth Resources Observation and Science (EROS) User Services. Email correspondence on October 13, 2016.

Wang S. 2016. CyberGIS and spatial data science. Berlin: Springer. https://doi.org/10.1007/s10708-016-9740-0

Williams CA, Gu H, MacLean R, Masek JG, Collatz GJ. 2016. Disturbance and the carbon balance of US forests: a quantitative review of impacts from harvests, fires, insects, and droughts. Glob Planet Change 143:66–80.

Woodall CW, Walters BF, Russell MB, Coulston JW, Domke GM, D'Amato AW, Sowers PA. 2016. A tale of two forest carbon assessments in the eastern United States: forest use versus cover as a metric of change. Ecosystems 19:1401–17.

Yang C, Goodchild M, Huang Q, Nebert D, Raskin R, Xu Y, Bambacus M, Fay D. 2011. Spatial cloud computing: How can the geospatial sciences use and help shape cloud computing? Int J Digit Earth 4:305–29.

Yang C, Raskin R, Goodchild M, Gahegan M. 2010. Geospatial cyberinfrastructure: past, present and future. Comput Environ Urban Syst 34:264–77.

Zhao F, Huang C, Goward SN, Schleeweis K, Rishmawi K, Lindsey MA, Denning E, Keddell L, Cohen WB, Yang Z, Dungan JL, Michaelis A. 2018. Development of Landsat-based annual US forest disturbance history maps (1986–2010) in support of the North American Carbon Program (NACP). Remote Sens Environ 209:312–26.

Zhu Z, Woodcock CE. 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens Environ 144:152–71.

Zimmerman PL, Housman IW, Perry CH, Chastain RA, Webb JB, Finco MV. 2013. An accuracy assessment of forest disturbance mapping in the western Great Lakes. Remote Sens Environ 128:176–85.