Title: FIRED (Fire Events Delineation): An open, flexible algorithm & database of US fire events derived from the MODIS burned area product (2001-19)

Jennifer K. Balch, Lise A. St. Denis, Adam L. Mahood, Nathan P. Mietkiewicz, Travis Williams, Joe McGlinchy, and Maxwell C. Cook

1 Earth Lab, CIRES, University of Colorado Boulder, Boulder, CO 80309 USA;
2 Department of Geography, University of Colorado Boulder, Boulder, CO 80309 USA;
3 National Ecological Observatory Network (NEON), Boulder, Colorado 80301 USA

* Correspondence: jennifer.balch@colorado.edu

Received: date; Accepted: date; Published: date

Abstract:
Harnessing the fire data revolution, i.e., the abundance of information from satellites, government records, social media, and human health sources, now requires complex and challenging data integration approaches. Defining fire events is key to that effort. In order to understand the spatial and temporal characteristics of fire, or the classic fire regime concept, we need to critically define fire events from remote sensing data. Events, fundamentally a geographic concept with delineated spatial and temporal boundaries around a specific phenomena that is homogenous in some property, are key to understanding fire regimes and more importantly how they are changing. Here, we describe FIRED, an event-delineation algorithm, that has been used to derive fire events (N = 51,871) from the MODIS MCD64 burned area product for the coterminous US (CONUS) from January 2001 to May 2019. The optimized spatial and temporal parameters to cluster burned area pixels into events were an 11-day window and a 5-pixel (2315 m) distance, when optimized against 13,741 wildfire

Remote Sens. 2020, XX, x; doi: FOR PEER REVIEW
www.mdpi.com/journal/remotesensing
perimeters in the CONUS from the Monitoring Trends in Burn Severity record. The linear
relationship between the size of individual FIRED and MTBS events for the CONUS was strong
($R^2 = 0.92$ for all events). Importantly, this algorithm is open source and flexible, allowing the
end user to modify the spatio-temporal threshold or even the underlying algorithm approach as
they see fit. We expect the optimized criteria to vary across regions, based on regional
distributions of fire event size and rate of spread. We describe the derived metrics provided in a
new national database and how they can be used to better understand US fire regimes. The
open, flexible FIRED algorithm could be utilized to derive events in any satellite product. We
hope that this open science effort will help catalyze a community-driven, data-integration effort
(termed OneFire) to build a more complete picture of fire.

Keywords: data harmonization; event-builder algorithm; fire regimes; open fire science; satellite
fire detections

1. Introduction

What is a fire? Defining the spatial and temporal boundaries of fire events is critical for
understanding the drivers and trends in fires [1], ecological consequences [2], and adaptation
implications [3]. Answering this question is fundamental to defining fire regimes, or the spatial
and temporal characteristics of fire events in a strict sense [4–6], i.e., size, frequency, intensity,
severity, seasonality, duration, and rate of spread. Remote sensing has increased our capacity to
quantify some of these characteristics at large spatial scales, such as frequency, intensity, size,
and severity [7–9]. However, there is even greater potential to inform our understanding of
changing fire and resilience of ecosystems and society if we are able to delineate events in
remote sensing fire products that preserve the temporal characteristics. We can then better
understand whether ecosystem state transitions depend on fire intensity and speed or how
communities in the wildland-urban interface (WUI) may be vulnerable to rapid fire spread.
There are generally three classes of information that satellite sensors capture about fire behavior: active fires based on thermal threshold exceedance \([10–12]\), fire radiative power as a metric of heat flux \([13–15]\), and burned area derived from a change detection algorithm \([16–18]\), sometimes also informed by active fire detections \([19]\). These fire properties are estimated at the pixel level, which ranges in size for these products from 10s to 1000s of meters. In order to explore fire behavior patterns these pixel-level detections are aggregated in some way, necessitating the assumption of homogenous fire characteristics across that pixel. Global burned area products tend to underestimate total burned area due to missing small fires \([20]\) and within-fire burned area due to underestimation of burned areas within an event \([21]\). Further, global scale studies explore total burned area summed across larger units or the density of hot pixels as a metric of fire frequency \([8,22–24]\), which leaves understanding of actual events missing. Given the abundance of satellite fire data (e.g., Table 1), and that they do not “see” the same aspects of fire \([12,31,32]\), we fundamentally need landscape-scale event delineation to integrate across products and build greater understanding of how fire regimes vary at regional and global scales \([25]\). With event-level delineations we can then also calculate a critical, but less understood property of fire regimes—fire spread. The MODIS-based burned area products \([26,27]\) use sub-daily images to estimate the date a pixel burned. As such, they are uniquely suited to provide estimates of fire spread rate and duration, but only if we can say which pixels are all part of the same event. There have been some attempts to characterize fire spread using active fire products, but the code and resulting data products are not publicly available \([28]\).

Defining events from the MODIS-based products enables capturing fire events, from small to large events at a global-scale, providing key metrics on fire regimes and how they are changing.

There are several different approaches for delineating fire events based on proximity of burned area or hot pixels in space and time. Some studies have clustered the MODIS active fire hotspots (MODIS MOD14) to delineate events in Europe and northern Africa \([29]\) and Indonesian tropical rainforests \([26,27]\) to understand what drives large fires. Others have used clustering of MODIS burned area (MODIS MCD64) pixels \([7,8,30,31]\). Most studies require pixel adjacency...
(Table 1), but a more relaxed spatial criteria facilitates exploring fires that have unburned patches within their perimeters—critical refugia that are necessary for regeneration [21]. This approach is also less likely to over-segment events that are imperfectly detected, due to low fire severity or cloudiness, for example.

Given the number of studies that use the MODIS burned area product (e.g., studies in Table 1) and emerging new fire data products [12, 17, 25–27, 32] that conduct some sort of event delineation as part of the processing, there is a great need to develop an open and well-documented algorithm for defining fire events from remotely sensed detections of fire. Moreover, event delineation enables joining of different data products to build a more complete picture of regional and global fire. Better delineation of the boundaries of events could lead to better estimates of total burned area, as well as exploration of derived spatiotemporal metrics around events that constitute the fire regime (e.g., event size, event shape, ignition point, unburned refugia within a fire, and fire spread rate). Many of the algorithms that have been developed previously were used and optimized for one specific analysis and the code was not published for further development and reuse [7, 33–35]. Furthermore, decisions were often made that lessened the computational cost, but relied on assumptions that are often not universally applicable. Most notably, data were often aggregated into a single yearly layer, which results in the artificial aggregation of pixels that burned multiple times in one year, and the artificial segmentation of events that started in one year and ended in the next [34–36].

Further, there is a need to better validate the temporal and spatial thresholds, as this selection can substantially alter the number of detected fire events. Fire metrics can be sensitive to how boundaries are delineated [37]. Moreover, we expect the optimum temporal and spatial thresholds to vary based on size distribution and spread differences that will vary across ecoregions (e.g., fast, large grassland fires vs. small, slow temperate forest fires) and land use types (e.g., agricultural fires vs. deforestation fires). But even so, ground-based delineations of fire perimeters also have their challenges, incident command delineations may overestimate wildfire perimeters, as delineating unburned patches is difficult on the ground. Also, multiple
Fire patches may start independently and in proximity (e.g., when a lightning storm starts multiple events), which then merge into one fire complex.

Table 1: Studies using spatio-temporal moving window algorithms for delineating fire events, first utilized by Archibald et al. [33].

<table>
<thead>
<tr>
<th>Study</th>
<th>Purpose</th>
<th>Satellite fire product</th>
<th>Spatial criteria</th>
<th>Temporal criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archibald et al. 2009</td>
<td>Examined environmental and anthropogenic drivers of fire in South Africa</td>
<td>MODIS MCD45*</td>
<td>Adjacency</td>
<td>8 days</td>
</tr>
<tr>
<td>Balch et al. 2013</td>
<td>Tested whether cheatgrass occurrence increases fire activity</td>
<td>MODIS MCD45, RMGSC^5</td>
<td>2 pixels (1000 m)</td>
<td>2 days</td>
</tr>
<tr>
<td>Hantsen et al. 2015</td>
<td>Explored global fire size distribution</td>
<td>MODIS MCD45</td>
<td>Adjacency</td>
<td>14 days</td>
</tr>
<tr>
<td>Frantz et al. 2016</td>
<td>Aggregated raster grids from burn date to event objects</td>
<td>MODIS MCD64^+</td>
<td>10 pixels (5000 m)</td>
<td>5 days</td>
</tr>
<tr>
<td>Andela et al. 2017</td>
<td>Examined global fire activity</td>
<td>GFED4s®; MODIS MCD64</td>
<td>Local spread rate x distance</td>
<td>Spatially varying fire persistence threshold</td>
</tr>
<tr>
<td>Laurent et al. 2018</td>
<td>Derived patch functional traits and other</td>
<td>MODIS MCD64, MERIS</td>
<td>1 pixel (500 m)</td>
<td>3, 5, 9 and 14 days</td>
</tr>
</tbody>
</table>
Here, we: i) develop an open, refined, and adaptable algorithm for defining events; ii) derive events and companion metrics for fires in the CONUS from the MODIS MCD64 burned area product, based on the optimum spatial and temporal thresholds; iii) validate the MODIS-derived events against the Monitoring Trends in Burn Severity (MTBS) product, which is manually derived from Landsat imagery [38]; and iii) demonstrate how defining events enables us to explore additional metrics of the fire regime across the US. Here, we define an event [39] as a geographic concept with delineated spatial and temporal boundaries around a specific phenomena that is homogenous in some property and distinct from adjacent areas. The algorithm is designed in a way that makes it adaptable to data source, regional context, and even event type: the spatiotemporal criteria can be altered, and it could be used with newer burned area products (e.g., Fire_cci based on MODIS images at 250 m resolution [26] or VIIRS [12]), or even different types of phenomena (e.g. bark beetle outbreaks, floods, etc.).

2. Materials and Methods

a. Study area and data acquisition and processing

The study area was CONUS. We chose this study area because of the availability of other fire datasets like MTBS [38] which we were able to use to gauge the accuracy of our aggregation of burned pixels to events from the MCD64 dataset. We used the MODIS Collection 6 MCD64 burned area product [27] [available at ftp://fuoco.geog.umd.edu/MCD64A1/C6/] These data contain five layers at 500-m resolution: burn date, first day, last day, a quality assessment, and...
error. The data are available worldwide, via a sinusoidal projection that is divided into 648 tiles
(268 of which are terrestrial), each with 2400 rows and columns at 463-m resolution. We
downloaded the entire monthly time series available for each tile that overlaps with CONUS,
and extracted the burn date layer.

b. Accounting for pixels that burn more than once per year (intra-annual reburns)

Some other studies that have aggregated pixels into fire events from the MODIS burned area
product have aggregated the input data to a yearly time-step [34,35], taking either the earliest
or latest burn date in the case of pixels that burn twice in one year. This assumes a minimal
occurrence of pixels that actually burn twice in one year (e.g. the land burns first in spring and
then again in fall). Aggregating the monthly data to yearly time steps makes the processing of
the data much less complex and computationally costly (i.e., it allows for a 2-dimensional
moving window). However, aggregation at a yearly timescale presents two problems. First, the
occurrence of pixels that burn more than once within a year would result in separate events
being collapsed, resulting in an underestimate of burned area for the study area. Second, fires
that burn from one year to the next become arbitrarily split into two events.

Prior efforts have justified ignoring intra-year or intra-season reburns based on an occurrence of
around 1% [34,35]. However we found that when we examined the study area tile by tile, some
areas experienced rates of intra-year reburns much greater than 1%. To investigate whether
reburned pixels would have a confounding effect on our data, we examined the occurrence of
pixels that burned multiple times per year for each of the tiles overlapping CONUS for each
year. We converted each monthly tile in CONUS to binary (1 for burned, 0 for unburned),
summed each monthly pixel per year and calculated the percentage of pixels that burned more
than once per year, per tile. For 2001 - 2018 for all of CONUS except the tile that contains
Florida, there were a total of 12,676 pixels that burned more than once in a given year, or about
0.48% of pixels. The tile that includes Florida (h10v06) had a rate of 5% (sd 2.3%) of pixels that
burned multiple times per year (Table 2). We suspect that this high reburn occurrence is due to
the year-round growing season combined with year-round occurrence of lightning strikes and
human ignition pressure. Intra-year reburns would present a problem if this algorithm were expanded globally, because there are many ecosystems, especially in the tropics, with year-round growing seasons combined with year-round anthropogenic ignition sources.

Because of the relatively high reburn occurrence, and also due to concern over segmenting winter fires into multiple events, we decided not to aggregate the input rasters by year or fire season. Instead, we created a space-time cube for each monthly tile for the entire time series, where the julian day of the year for each pixel in each month layer was converted to a number along a continuous series starting on January 1, 1970.

Table 2. Number of reburned pixels per year, per tile calculated for 2001-2018 from the monthly MODIS MCD64 Burned area product.

<table>
<thead>
<tr>
<th>Tile</th>
<th>Mean Reburn %</th>
<th>Std Reburn %</th>
</tr>
</thead>
<tbody>
<tr>
<td>h08v04</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>h08v05</td>
<td>0.35</td>
<td>0.27</td>
</tr>
<tr>
<td>h08v06</td>
<td>1.35</td>
<td>1.05</td>
</tr>
<tr>
<td>h09v04</td>
<td>0.36</td>
<td>0.30</td>
</tr>
<tr>
<td>h09v05</td>
<td>0.23</td>
<td>0.19</td>
</tr>
<tr>
<td>h09v06</td>
<td>0.73</td>
<td>0.47</td>
</tr>
<tr>
<td>h10v04</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>h10v05</td>
<td>0.67</td>
<td>0.29</td>
</tr>
<tr>
<td>H10v06 (Florida)</td>
<td>5.12</td>
<td>2.31</td>
</tr>
<tr>
<td>h11v04</td>
<td>0.35</td>
<td>0.33</td>
</tr>
<tr>
<td>h11v05</td>
<td>0.32</td>
<td>0.35</td>
</tr>
<tr>
<td>h12v04</td>
<td>0.35</td>
<td>0.61</td>
</tr>
<tr>
<td>h13v04</td>
<td>0.32</td>
<td>0.29</td>
</tr>
<tr>
<td>Total (excluding h10v06)</td>
<td>0.48</td>
<td>0.55</td>
</tr>
</tbody>
</table>

c. Defining events with a flexible, fast algorithm

We created an algorithm that automatically downloads, processes, defines events and calculates summary statistics for the entire CUS in 30 minutes on a normal laptop. To define events, we used a 3-dimensional moving window to aggregate burned pixels into distinct events. The
algorithm takes as input a spatial variable, representing the number of pixels, and a temporal variable, representing the number of days, within which to group burn detections. It then aggregates by assigning each burned pixel an event identification number.

The data processing script downloads the entire time series of HDF files from the ftp server, extracts the burn date layer from each monthly tile, and adds them to a 3-dimensional netCDF data cube. We used this data structure to maximize efficiency and speed. The event perimeter script reads the netCDF file for each tile, where each band represents one month, and for each burned pixel the date of fire detection is represented as the number of days since January 1, 1970. The netCDF file is converted into a three-dimensional array and the moving window traverses the array. To avoid unnecessary computation, we did not check cells in which there was no burned area assignment throughout the study period.

For each cell where at least one fire detection occurred, the program creates a mask identifying all burned pixels that fall within the spatial and temporal range of the current cell. If the current cell is already part of an existing event, any new burned pixels are assigned the event ID for that event. If it is a new event, the current cell and all overlapping cells are given the next sequential event ID. If there are multiple event IDs within the mask, two perimeters have grown together and they are merged into the first event ID. After the event perimeters are delineated within each tile, all event perimeters that potentially overlap with an adjacent tile are flagged. After all tiles are processed, the flagged events are partitioned and those that overlap spatially and temporally are merged. Finally, events across all tiles are merged into a final dataset and given a new sequential event ID.

**d. Sensitivity analysis: identifying the optimal spatiotemporal parameters for delineating fire events**

In order to find which combination of spatial and temporal variables outputs best defined fire events for CONUS, we assessed how well the FIRED outputs matched fire perimeters from MTBS [38]. MTBS is a dataset of fire perimeters from 1984-2016 derived from Landsat satellite data. It has a minimum size threshold of 404 ha in the western US and 202 ha in the eastern US.
It documents 21,673 fire events throughout the entire US, and 13,741 in the overlapping study area and timeframe, beginning in 2001. One problematic feature of the MTBS data for this comparison is that fire complexes are not dealt with uniformly. Fire complexes are “two or more individual incidents located in the same general area which are assigned to a single incident commander or unified command [40].” In some cases each fire patch is assigned its own ID number and is represented as a single perimeter, and in other cases these complexes are lumped into a multipolygon with a single ID number. To address this issue, we split all multipolygons into single polygons, assigned unique ID numbers to each polygon, and then calculated the area for each individual polygon. This way, our sensitivity analysis would objectively assess how individual polygons matched, without the confounding factor of aggregated multipolygons.

We ran the fire event classifier for all spatiotemporal combinations between 1-15 days and 1-15 pixels (463 - 6,945 m), resulting in 225 spatiotemporal combinations for CONUS. For each combination we matched the FIRED events that were >404 ha in the west and >202 ha in the eastern US to the associated MTBS wildfire perimeter. An accuracy assessment was conducted for each spatiotemporal combination of the FIRED events, based on how well they matched the MTBS events. For each unique fire polygon in the MTBS database, we extracted the ID numbers for each FIRED event overlapping the MTBS polygon. Then, for each unique FIRED event, we extracted each MTBS ID that overlapped. We then calculated the ratio of the number of unique MTBS events that contained a FIRED event divided by the number of unique FIRED events that contained at least one MTBS event, with the optimum value being one. We used this ratio to approximate the spatio-temporal combination that minimized both over- and under-segmentation of the FIRED events based on known MTBS fire perimeters.

Based on the ratio that minimized both over- and under-segmentation, we estimated an optimal combination for the US of 5 pixels (2315 m) and 11 days. We calculated commission and
omission errors for both the FIRED events and the MTBS events.

e. Calculating statistics for each event, and daily statistics within events

Once the optimal spatial-temporal aggregation level was identified, we created two vector products for CONUS: one where individual pixels were aggregated to polygons representing each fire event, and one where individual pixels were aggregated to each date within each event. For the event-level vector product, we calculated ignition location and date, duration, spread rate (burned area/duration), burned area, date of maximum growth, area burned on the dates of maximum and minimum growth (the date with the highest burned area per event), and the mean daily area burned for each event. We also extracted the mode of the International Geosphere-Biosphere Programme land cover classification from the MODIS MCD12Q1 landcover product for the year before the fire [41], and the Community for Environmental Cooperation’s level 1-3 ecoregions [42], for each event (Table 3). For the daily-level vector product, we calculated the daily burned area, cumulative burned area per day, days since ignition, mode landcover per day, and mode ecoregion per day, in addition to the metrics calculated for the event-level product (Table 4). In addition, the algorithm has a third output: a table with each burned pixel as a single row, with coordinates, burn date, and the event identification number derived from the algorithm. This raw output is provided so the end-user can use and manipulate the raw data in any way they see fit.

Table 3. Attributes included in the event-level FIRED product.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ignition date, day of year, month, year, location</td>
<td>date, day of year, month, year, location</td>
</tr>
<tr>
<td>Duration</td>
<td>days</td>
</tr>
<tr>
<td>Burned Area</td>
<td>km², ha, acres, pixels</td>
</tr>
<tr>
<td>Fire Spread Rate</td>
<td>pixels/day, km²/day, ha/day, acres/day</td>
</tr>
<tr>
<td>Maximum, minimum, and mean growth rate</td>
<td>km²/day, ha/day, acres/day, pixels/day, date (max only)</td>
</tr>
<tr>
<td>Attribute</td>
<td>Units</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>Daily Burned Area</td>
<td>km², ha, acres, pixels</td>
</tr>
<tr>
<td>Daily Landcover</td>
<td>mode land cover classification / day</td>
</tr>
<tr>
<td>Daily Ecoregion</td>
<td>mode ecoregion, Levels 1-3</td>
</tr>
<tr>
<td>Cumulative Burned Area</td>
<td>km², ha, acres, pixels</td>
</tr>
<tr>
<td>Ignition Date (of the whole event)</td>
<td>date</td>
</tr>
<tr>
<td>Last Burn Date (of the whole event)</td>
<td>date</td>
</tr>
<tr>
<td>Duration (of the whole event)</td>
<td>days</td>
</tr>
<tr>
<td>Event Day</td>
<td>days from ignition date</td>
</tr>
<tr>
<td>Percent Total Area</td>
<td>percent (%)</td>
</tr>
<tr>
<td>Percent Cumulative Area</td>
<td>percent (%)</td>
</tr>
<tr>
<td>Fire Spread Rate (of the whole event)</td>
<td>pixels/day, km²/day, ha/day, acres/day</td>
</tr>
</tbody>
</table>

Table 4. Attributes included in the daily-level FIRED product

In order to understand how well the FIRED algorithm delineated event size, we compared the estimates of burned area from FIRED events to the estimates of burned area for MTBS events for the subset of events that were captured by both products. Because MTBS does not account for unburned patches within a fire perimeter when they calculate burned area, many burned area estimates reported by MTBS are likely overestimations. Thus, comparing the area burned by the two products represents a trade-off between imperfect satellite detection from MODIS and imperfect burned area reporting in the perimeters that drive selection by the MTBS product. With those caveats in mind, we co-located those events captured by both products (i.e. they overlapped in space and time), and compared estimated area burned at the event level using
two approaches. First, to compare all fire events, we created a linear regression model where the
FIRED-determined area burned predicted MTBS-determined area burned. Second, to
understand how that relationship varied with size class, we binned the fire events into 50 equal
size classes, and created a linear model on each subset. The expectation was that FIRED-based
burned areas would be consistently less than the MTBS-based burned areas. In addition, due to
lower burn detection by MODIS for smaller fires [32], we expected the models at smaller size
classes to explain less of the variation than for large sizes. We also acquired the total yearly
burned area and fire counts from the National Interagency Fire Center (NIFC) for CONUS to
understand how FIRED and MTBS products compared to the aggregation of all reported
wildfires (note, NIFC does not include intentional land use fires or prescribed burns).

Data and code availability
Code for the python command line interface used to download data, classify events, calculate
event- or daily-level statistics, and write tables and shapefiles is available as the “firedpy”
python package at www.github.com/earthlab/firedpy. R code for the analysis presented here is
available at https://github.com/earthlab/modis-fire-events-delineation. R code for the
sensitivity analysis is available at www.github.com/admahood/fired_optimization. Data is
available at CU Scholar [DOI: https://doi.org/10.25810/3hwy-4g07].

3. Results

Classification accuracy assessment
The MODIS-derived events had a 55% omission and 62% commission error, compared to the
MTBS reference dataset, based on a confusion matrix that compares when FIRED and MTBS
identify the same events (Table 5). An additional 24,163 events were detected below the MTBS
size thresholds and were not included in these calculations.

\[
CE = \frac{\text{FIRED}_{\text{true}} \cdot \text{MTBS}_{\text{false}}}{(\text{FIRED}_{\text{true}} \cdot \text{MTBS}_{\text{false}} + \text{FIRED}_{\text{true}} \cdot \text{MTBS}_{\text{true}})} = \frac{11,412}{(11,412 + 7,054)} = 0.62
\]
Table 5. Confusion matrix for the MODIS MCD64-derived events. The MTBS event-size threshold is 404 ha in the western US, 202 ha in the eastern US.

<table>
<thead>
<tr>
<th></th>
<th>MTBS True</th>
<th>MTBS False (Commission)</th>
<th>MTBS False (Commission)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIRED True</td>
<td>7,054</td>
<td>11,412 (over threshold only)</td>
<td>24,163 (under threshold only)</td>
</tr>
<tr>
<td>FIRED False (Omission)</td>
<td>8,721</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 1. A comparison of the spatial distribution of fire events from the FIRED and MODIS products from 2001-2016 shows a similar distribution of fire events and burned area in general, but the FIRED algorithm picks up many more events and burned area in the midwest, southeastern US and eastern Washington.

b. Comparison to MTBS:

There were approximately 3.3 times more wildfire events and 65,000 km$^2$ (18%) more burned area captured in the FIRED product compared to MTBS. The FIRED burned area represents 97%
of the National Interagency Fire Center (NIFC) reported totals from 2001-2016 (Table 6). The
relationship between area burned for the FIRED events and the MTBS events was strong ($R^2 =
0.92$, Figure 2A), and the area reported by MTBS was always higher than the FIRED events (the
points are all above the 1:1 line in Figure 2A) at the event level. As event size increased, the $R^2$
improved from below 0.6 for fires below 50,000 acres, to above 0.8 for fires over 70,000 acres
(Figure 2b). The MODIS MCD64A1 burned area product consistently underestimated burned
area reported by MTBS for fires below 100,000 hectares. This consistent underestimation is not
necessarily a flaw with the FIRED product, rather it is partially due to the fact that MTBS does
not account for unburned patches within a fire perimeter when they calculate burned area, and
burned area is consistently overestimated by MTBS. The burned area captured by MODIS
MCD64A1, and thus FIRED, was much closer to the NIFC totals (Table 6). This is likely because
the MCD64A1 product captures many more small fires than MTBS. However, the MCD64A1
product does not generally capture the smallest fires, below 12.6 ha [32]. There is a dramatically
larger count of individual events reported by NIFC, which includes many fires as small as 0.4
ha.

Table 6: Fire events and burned area by level one ecoregion, 2001-2016. National Interagency
Fire Center statistics compiled from https://www.nifc.gov/fireInfo/fireInfo_stats_totalFires.html

<table>
<thead>
<tr>
<th>Level 1 Ecoregions</th>
<th>MTBS Events</th>
<th>Burned Area (km²)</th>
<th>FIRED Events</th>
<th>Burned Area (km²)</th>
<th>NIFC Events</th>
<th>Burned Area (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Temperate Forests</td>
<td>5,644</td>
<td>47,116</td>
<td>20,556</td>
<td>103,615</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Great Plains</td>
<td>3,350</td>
<td>94,068</td>
<td>11,818</td>
<td>112,907</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Marine West Coast Forest</td>
<td>22</td>
<td>379</td>
<td>249</td>
<td>978</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mediterranean California</td>
<td>368</td>
<td>17,971</td>
<td>1,432</td>
<td>21,251</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>North American Deserts</td>
<td>1,739</td>
<td>80,430</td>
<td>5,689</td>
<td>72,012</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Northern Forests</td>
<td>134</td>
<td>2,130</td>
<td>141</td>
<td>2,086</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Northwestern Forested</td>
<td>1,614</td>
<td>81,189</td>
<td>3,815</td>
<td>68,006</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mountains</td>
<td>159</td>
<td>5,494</td>
<td>260</td>
<td>4,459</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Remote Sens. 2020, XX, x; doi: FOR PEER REVIEW
www.mdpi.com/journal/remotesensing
<table>
<thead>
<tr>
<th>Ecoregion</th>
<th>MTBS (hectares)</th>
<th>MODIS (hectares)</th>
<th>R²</th>
<th>1:1 line</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highlands</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperate Sierras</td>
<td>431</td>
<td>19,374</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Tropical Wet Forests</td>
<td>266</td>
<td>4,818</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conterminous US</td>
<td>13,727</td>
<td>352,967</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: A comparison of burned area for individual fire events delineated by both products. Panel A shows the relationship between area burned for MTBS and FIRED events. While the relationship is generally strong (R² = 0.92 for all events), it is weaker for smaller fires. For panels B and C we binned the data into 50 equal size classes (each bin spans ~ 5000 hectares), and ran a linear regression (MTBS burned area predicted by MODIS burned area) on each bin. Panel B shows the R² values, which do not consistently stay above 0.8 until about 70,000 hectares. Panel C shows the relationship between the slope of the regression line for each size class bin, illustrating that the MODIS MCD64A1 burned area product consistently underestimates burned area for fires below 100,000 hectares.

d. Ecoregion comparisons between FIRED and MTBS

One of the primary differences between the two products is the detection of small fires, which is...
a function of the ~200-ha and ~400-ha cut-off for the eastern and western US in the MTBS product [38]. In the east and central US, where fires are generally smaller, FIRED captured 37,724 fires while MTBS captured 11,008 fires (Figure 1, Table 5). There were several ecoregions where FIRED captured more events, but less burned area (e.g., in North American Deserts; Table 5). This is either due to the lack of smaller events in the MTBS dataset, or that MTBS does not delineate unburned patches within its fire perimeters, and so can overestimates burned area for many fires (e.g., see Figure 3).

Ecoregions with the highest maximum fire spread rates were those with large areas of grasslands - the Great Plains and desert ecoregions (Table 7). However, the three ecoregions with the highest mean fire spread rates were all forested ecosystems - the temperate Sierras, southern semi-arid highlands, and northern forests, and these ecoregions also had the highest variability in fire spread rates.

Table 7. Summary statistics of fire spread rate by ecoregion.

<table>
<thead>
<tr>
<th>Level 1 Ecoregions</th>
<th>Fire Events</th>
<th>Fire Spread Rate (ha/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Max</td>
</tr>
<tr>
<td>Eastern Temperate Forests</td>
<td>20,556</td>
<td>2,756</td>
</tr>
<tr>
<td>Great Plains</td>
<td>11,818</td>
<td>13,584</td>
</tr>
<tr>
<td>Marine West Coast Forest</td>
<td>249</td>
<td>301</td>
</tr>
<tr>
<td>Mediterranean California</td>
<td>1,432</td>
<td>5,883</td>
</tr>
<tr>
<td>North American Deserts</td>
<td>5,689</td>
<td>14,620</td>
</tr>
<tr>
<td>Northern Forests</td>
<td>141</td>
<td>2,442</td>
</tr>
<tr>
<td>Northwestern Forested</td>
<td>3,815</td>
<td>3,878</td>
</tr>
<tr>
<td>Mountains</td>
<td>260</td>
<td>1,755</td>
</tr>
<tr>
<td>Southern Semi-Arid Highlands</td>
<td>447</td>
<td>6,365</td>
</tr>
<tr>
<td>Temperate Sierras</td>
<td>1,394</td>
<td>1,220</td>
</tr>
</tbody>
</table>

Remote Sens. 2020, XX, x; doi: FOR PEER REVIEW
www.mdpi.com/journal/remotesensing
18
Figure 3. Comparison of A) FIRED and B) Global Fire Atlas delineated events for the Sour Orange fire (started February 9, 2007), the Moonshine Bay fire (started February 24, 2007), and a third unnamed event, FIRED event #29790 (started December 28, 2007, and continued into March of 2008). The FIRED product joins the two intra-year burns (#25211) and delineates a third event (#29790) that reburns some of the same pixels. The dark outlines, bold and dashed, show the MTBS fire perimeters for the Sour Orange and Moonshine Bay fires. Note that MTBS does not include unburned patches within perimeters. Panel B) shows the Global Fire Atlas (with an abridged legend showing 3 of 57 colors), which segments the same MODIS burned area pixels into 57 events and no delineation of overlapping reburns.

4. Discussion

Remote sensing has fundamentally changed our ability to quantify fire, and has consequently challenged how we define fire events. The active fire, burned area, and fire radiative power and
severity products [12,14,15,17,18,27,38] have fundamentally changed how we can conceptualize fire regimes. Key to translating this wealth of information is defining fire events in space and time so that we can understand how modern fire regimes are changing. Parallel efforts such as the Global Fire Atlas (based on the MODIS MCD64 product [27]) have converged on identifying the same need, with a key motivation to improve global fire modeling [30]. We argue that the need is more profound, that in order to understand how fire regimes are changing at regional to global scales we need an open, and flexible methodology to identify events and integrate fire data across sources based on these events. This event-based approach could be utilized to derive events in any satellite product to build a more complete picture of fire.

There are several beneficial aspects of our approach that yield more appropriate delineation of multi-year events, small fires, complexes, and intra-annual reburns, while also providing key output metrics, e.g., fire spread and pre-fire landcover. The primary difference between FIRED and other algorithms is that FIRED uses the entire monthly time series as a space-time cube input, upon which a 3-dimensional moving window is applied, compared to aggregating fire seasons or years into one layer upon which a 2-dimensional moving window is applied. This enables proper identification of intra-year reburns (Table 1) and ensures that fires at the end or beginning of months or years are not arbitrarily split into multiple events (Figure 3). Second, because the FIRED database is based on the MODIS MCD64 product, it includes fire events theoretically as small as 4m$^2$, albeit these are rare detections (~90% omission error) [32]. Small fire events greater than 12.6 hectares are more likely the events that are captured in the MODIS MCD64 product (10% omission error) at the size of a MODIS pixel (~500 m) [32], and therefore in FIRED. The MTBS database, in contrast, has a minimum threshold of 202 ha east and 404 ha west of the 97th meridian. Having small fires expands our ability to understand how fire size and burned area are changing, beyond just the large events [43]. Smaller events are difficult to capture systematically but we know these events can be incredibly important in the US, contributing large additional burned areas and emissions [20,44]. Third, the daily-level product preserves the daily-scale information (i.e., daily polygons and ensuing metrics) for the larger
events. This elucidates whether large fire events are actually complexes of smaller
independently ignited fire patches, or if the large event is truly the product of a single ignition
location (e.g., the Rim Fire in figure 4). This also allows users to link daily-level burned area
data within a defined event to daily or even sub-daily covariates (e.g., climate variables).

Fourth, this product provides several attributes that are new pieces of information, refined
across CONUS. For example, fire spread rate is a unique attribute, derived from events, which is
a critical piece of information not easily accessed in other datasets (e.g., MTBS or ICS-209s).
FIREd also provides the landcover for the year before the fire for each event, a coarse metric of
fuels information, and critical for understanding ecosystem impacts and resilience. This annual
landcover information could enable exploration of when fire precipitates rapid vegetation
transitions, particularly as woody plant-dominated systems may lose their resilience to fire
against a backdrop of warmer and drier climates [45,46]. Last, FIREd is also the only
automated, satellite-derived product we are aware of that captures intra-annual reburns. Intra-
annual reburns will perhaps become more prevalent in the future as the decline of resilience in
some ecosystems leads to an acceleration of disturbance regimes [47,48], particularly if novel
ecosystems result from invasive, flammable plants [7,49].

Another key advantage of this approach is that the algorithm is open and flexible; we hope for
community input and we expect it to be improved over time. The spatio-temporal criteria can
be altered based on other information, regionally-specific fire perimeters such as Canada’s
National Burned Area Composite (https://cwfis.cfs.nrcan.gc.ca/datamart), or known
delineations of intentional land use fires or prescribed burns. Further, we anticipate that this
algorithm has wide applicability to other fire products and other efforts to build events based
on any geospatial data that has both spatial and temporal information. Previous studies,
including this team’s previous efforts [7], have not made their workflow and code publicly
available, limiting the potential to facilitate community development of an integrated, evolving
global fire database.

With the plethora of remote sensing data about fire and fire effects, there is a great need to
delineate events at large regional and global scales. There are at least three other recent studies that have created fire events from the MODIS burned area product (Table 1), two of which [34,36] have created global fire event databases. In addition to the global efforts, Frantz et al. [35] created an algorithm based on a study area in sub-Saharan Africa which uses a top–down multilevel segmentation strategy that starts by defining potential ignition points and gradually refines the individual object membership. All three efforts use an approach that starts by identifying potential ignition points and grows objects from the ignition point using only adjacent pixels. The code for the algorithm created by Andela et al. [34] is not publicly available and the code created by Frantz et al. [35] is available upon request. Laurent et al. [36] created a publicly available database and the code is also available upon request. Their output data contains what they term fire patch functional traits, including patch area and other morphological features, but does not preserve daily fire spread information or polygons containing the perimeter shapes of the derived events. Our approach differs in that we use a spatiotemporal window that can capture isolated burned pixels that may be part of the same event, but may be isolated because of the inability of the MODIS sensor to detect burned area in the area between patches due to cloudiness, low vegetation density, low severity, or unburned patches (i.e., fire refugia) that are important elements of an event. It is worth noting that the spatial-temporal thresholds we derived (i.e., 11-day window and a 5-pixel distance) are much greater than those used in most previous studies (e.g. [12,34] but see Frantz et al. [35]), leading to less artificial truncation, or oversplitting, of events. For example, the Rim fire which occurred in California in 2013 was delineated into more than 10 separate events by the Global Fire Atlas algorithm, whereas our algorithm delineated a single event that more closely matches the MTBS delineation (Figure 4). Future improvements could include: i) validation with smaller events, such as those contained in the US-based National Incident Feature Service dataset, formerly Geomac [50] or others; ii) estimates of uncertainty around start and end dates of the fire events; iii) regionally-varying thresholds based on fire regime characteristics; and iv) development of an optimization process that does not rely on already existing fire perimeter polygons. In the current study, we were able to use the MTBS database to define the optimum spatial and temporal parameters for delineating events in CONUS. Unfortunately, these types of data do
not exist for many parts of the world. We attempted to scale the FIRED product to the entire globe and found that our spatial and temporal parameters were inappropriate, particularly for the savanna biome where very high proximity of fires in space and time led to severe aggregation of events. This highlights a substantial need for global fire perimeter data [51], or development of an optimization approach that does not rely on these external data.

Figure 4: The 2013 Rim Fire, which lasted over a month and was more than 100,000 ha in total size according to incident reports, as delineated by the A) FIRED event product; B) global fire atlas C) FIRED daily event product; and D) MTBS. The optimized spatial-temporal criteria we used allowed us to correctly classify it as a single event, while the global fire atlas has segmented the Rim Fire into 14 separate events. The FIRED ignition point is estimated as the average location of all pixels occurring on the first day of the event.

This is a unique moment in the history of fire science, given the abundance of fire data across Remote Sens. 2020, XX, x; doi: FOR PEER REVIEW www.mdpi.com/journal/remotesensing
spatial scales, that requires the fire science community to better coordinate efforts on fire data harmonization challenges and opportunities. We see great potential to build a community-driven, fire data infrastructure that we term OneFire. OneFire is a coordinated architecture that would enable a community of researchers and stakeholders to use, repurpose, and contribute to fire data, code, and workflows. The vision for OneFire is that it will be a coordinated, community-inspired data architecture that connects and integrates the many global, national, and regional fire databases. This is no small task, but integrating these datasets is key to unlocking a transformation in fire science and rapidly accelerating new discoveries about why fire regimes are changing and how societies and ecosystems are vulnerable. There is an enormous amount of data and work relevant for fire science that could be leveraged, if only it was open, reproducible, and scalable. For example, we anticipate that a newly published ICS-209-PLUS dataset that is an integrated database of over 120,000 incident command reports could be connected to MODIS FIRED events to join physical attributes with social impact and response on a daily scale [52]. Social media information around wildfires could also be leveraged, and provide a view of social response that before would not have been possible [53,54]. Additional satellite sensors and their derived products, e.g., active fire, could be leveraged to expand the detections per event and add other key properties like fire radiative power. Key elements of a vision for OneFire include: i) identified fire events across many datasets utilizing the FIRED event-builder algorithm or other approach the delineates events in space and time; ii) integration workflows that then connect those same events across data sources to build a fuller suite of attributes around commonly identified events; iii) data and computational infrastructure that allows for community contributions of data, code, and compute environments; iv) formal linkages to other important climate, environment, and social data sources that provide insights into driving forces or responses; and v) support for community building, engagement, and training that facilitates large, diverse team science.

Ultimately, no single sensor is going to provide all the information we need about fires, and we will never anticipate all the ways that such an integrated source of fire information would get used. OneFire would help us build a fuller, global picture of fire.
5. Conclusions

There is a clear need to derive events from remotely sensed detections of fires, as event perimeters are a key tool for exploring how the spatio-temporal properties of fire regimes are changing [55–57] and how resilience to fire is changing [49,58,59]. Further, there are dozens of fire products available, for the US and globally (Table 1), that could, if combined and harmonized, shed new insights on the drivers and consequences of changing fire. Delineating fire events is key to this process, and we argue that this US database and algorithm offer the opportunity to begin to build OneFire, a community data-integration effort for fire science. No one research group can predict the variables that will be needed for all studies, and there is no one satellite that captures all the needed information about fire. We envision that our algorithm will be optimized at different scales to best capture regional fire size distributions. We also envision that this algorithm can be used across any satellite-based fire product, from active fire detections to burned area products, and particularly new efforts, such as the BAECV product or VIIRS. Moreover, this algorithm can be used with any spatiotemporal data and is not constrained to fire data. As other efforts are built to understand natural hazards, these efforts may help to better delineate the spatial and temporal dimensions of floods, hurricanes, disease outbreaks, and other events. The fire science community can better harmonize fire observations for a larger network of researchers and practitioners who need this information to better help society more sustainably live with fire.

Author Contributions: Conceptualization, JB, AM and NM; methodology, JB, LS, AM, NM and TW.; software, LS, AM, NM, TW, JM, and MC; algorithm development, LS; validation, LS, AM and NM.; formal analysis, LS, AM and NM.; investigation, JB, LS, AM and NM; data curation, AM, NM and TW; writing—original draft preparation, JB, LS and AM; writing—review and editing, JB, LS, AM, NM and JM; visualization, AM and NM.; supervision, JB and LS.; project administration, JB; funding acquisition, JB. All authors have read and agreed to the published version of the manuscript.

Acknowledgements: The authors are grateful to Bethany Bradley, Maxwell Joseph, and Sepideh Dadashi for conversations that informed this work. The authors also thank four anonymous reviewers and the
editor for their constructive comments on a previous version of the manuscript.

**Funding:** This research was funded by the National Aeronautics and Space Administration Terrestrial Ecology Program (grant number NNX14AJ14G) and by CIRES and the Grand Challenge Initiative at the University of Colorado, Boulder.

**Conflicts of Interest:** The authors declare no conflict of interest.
References


40. USDA Forest Service Fire Terminology Available online: https://www.fs.fed.us/nwacfire/home/terminology.html.

41. Friedl, M.; Sulla-Menashe, D. MCD12C1 MODIS/Terra+ Aqua Land Cover Type Yearly L3 Global 0.05 Deg CMG V006 [Data set]. NASA EOSDIS Land Process. DAAC **2015**.

42. Cooperation, C. for E.; Commission for Environmental Cooperation *Ecological regions of North America – Levels I, II, and III: Montreal, Quebec, Canada, Commission for Environmental Cooperation, scale 1:10,000,000*; 2006;

doi:10.1071/WF15039.


51. Briones-Herrera, C.I.; Vega-Nieva, D.J.; Monjarás-Vega, N.A.; Briseño-Reyes, J.; López-Serrano, P.M.; Corral-Rivas, J.J.; Alvarado-Celestino, E.; Arellano-Pérez, S.; álvarez-González, J.G.; Ruiz-González, A.D.; et al. Near real-time automated early mapping of the perimeter of large forest fires from the aggregation of VIIRS and MODIS active fires in...


patterns of post-fire conifer refugia in ponderosa pine-dominated forests of the Colorado