# scientific data



# **OPEN** Country-level fire perimeter DATA DESCRIPTOR datasets (2001–2021)

Adam L. Mahood 1,2 , Estelle J. Lindrooth 1,3, Maxwell C. Cook 1,4 & Jennifer K. Balch 1,4

Fire activity is changing across many areas of the globe. Understanding how social and ecological systems respond to fire is an important topic for the coming century. But many countries do not have accessible fire history data. There are several satellite-based products available as gridded data, but these can be difficult to access and use, and require significant computational resources and time to convert into a usable product for a specific area of interest. We developed an open source software package called Fire Event Delineation for python (FIREDpy) which automatically downloads and processes all of the source files for an area of interest from the MODIS burned area product, and runs a spatiotemporal flooding algorithm that converts those hundreds of grids into a single fire perimeter shapefile. Here we present a collection of fire event perimeter datasets for every country on the globe that we created using the FIREDpy software. We will continue to improve the efficiency and flexibility of the underlying algorithm, and intend to update these datasets annually.

# **Background & Summary**

Global fire activity is changing in many areas as temperatures increase and land use intensifies<sup>1-5</sup>. This is sparking an increase in attention given to fire activity and fire ecology. However, the availability of data for spatially delineated fire events is limited or non-existent in many countries<sup>6</sup>, with most global fire data coming from satellite-based active fire detections<sup>7,8</sup> and gridded burned area products<sup>9,10</sup>. The lack of products containing delineated events has led to many global studies about fire ecology that are computationally-intensive, coarse-scale trend analyses<sup>1,4</sup>.

A key advantage of datasets like Monitoring Trends in Burn Severity (MTBS)11 or the Fire Occurrence Dataset<sup>12</sup> lies in their ease of use. Since its inception in 2007 MTBS has been cited 947 times in peer-reviewed studies according to a Google Scholar search at the time of this writing, despite documented limitations for scientific use of some facets of the product<sup>13</sup>. The MTBS dataset is regularly updated, easy to find on the internet, and it is free, fast and easy to download and use. Many environmental scientists and resource managers do not have the computational budget or expertise in big data or remote sensing to deal with the challenges one must overcome to process large fire datasets. This is especially true for cases when all that is needed is a shapefile of fire perimeters that can be used to map fire history. Other global fire perimeter datasets have been produced from satellite-derived burned area products<sup>14,15</sup>, but these are only available in yearly or monthly global shapefiles. Often field-based studies of fire effects require an entire time series over study areas that are only a few hundred km in diameter <sup>16</sup> or a single ecoregion <sup>17</sup>. The end user who wants to understand the fire history for their region would have to download yearly shapefiles with a global extent, clip all of those shapefiles to their area of interest, and then combine them into one shapefile, just to get started. We suspect that the lack of accessible fire perimeter datasets that are easy to download and use contributes to a disparity in research, where fire ecology studies are conducted mostly in developed countries that have either research infrastructure capable of handling big data or longer-term government records, or temperate forested regions that have substantial tree-ring records.

There are two existing global perimeter products, the Global Fire Atlas (GFA) (Andela et al. 14) and the Global Wildfire Information System (GWIS) (Artes et al. 15). Both were created by applying spatiotemporal flooding algorithms to the MODIS MCD64 Burned Area Product. These algorithms assign burned pixels from the MCD64 products using a moving window whose size is defined by spatial and temporal parameters. They are created as monthly or yearly slices of the entire globe, and they can be subsetted. These products are extremely valuable for global scale studies. But when we look at how those products delineate known fire events we see a consistent problem in that they both seem to over-segment events in ways that appear unrealistic. This inconsistent event delineation is not problematic for coarse-scale or regional estimates of burned area or fire seasonality,

 $^{1}$ Earth Lab, Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, USA. <sup>2</sup>Water Resources, USDA-ARS, Fort Collins, CO, USA. <sup>3</sup>Applied Math, University of Colorado Boulder, Boulder, USA. <sup>4</sup>Geography, University of Colorado Boulder, Boulder, USA. <sup>™</sup>e-mail: admahood@gmail.com

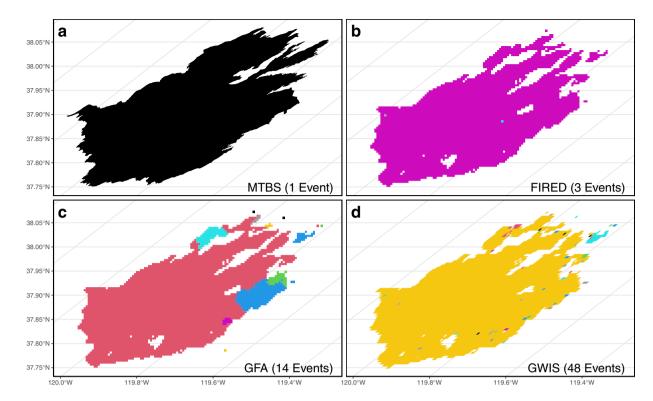
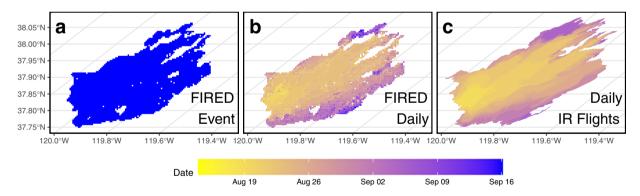


Fig. 1 Comparison of global fire event products performance for the 2013 Rim Fire (a). In the FIRED product (b), the Rim fire was classified as one very large event with two single pixel events. The Global Fire Atlas (GFA, c) and Global Wildfire Information System (GWIS, d) each delineated a very large event, with 13 and 47 smaller events, respectively.



**Fig. 2** The two primary outputs FIREDpy provides are a daily- and event-level product. Panel a shows the default single event polygon. In b, each day has a separate polygon, with associated statistics generated, within each event. Panel c shows the daily perimeters derived from the airborne infrared by the incident management team for comparison.

but can lead to unrealistic estimates for number of fire events and event-level characteristics like fire size and spread rate. In Fig. 1 we illustrate this with an example of the 2013 Rim Fire in California, United States, which was unmistakably a single event that burned about 90,000 ha over the course of three months. Figure 2 illustrates how the day-to-day progression of the Rim Fire was a steady progression from a single ignition in late August. Table 1 shows how the differences in event delineation propagate to calculations of burned area and number of events. In the GFA, the Rim Fire is delineated as one large event of 804.5 km², and 13 additional events totaling 88.7 km². in GWIS it is delineated as one event of 878 km² and 47 additional events totalling 20 km². With FIRED, there is one event of 892 km² and 2 single pixel events totalling less than one km². One cause for potential differences is how one defines a "fire event". Large fires often have multiple ignition sources. The Global Fire Atlas algorithm and others¹9, for example, search for local minima to identify various ignition locations that may begin as small patches, only to later form a large complex and in the end described with a single fire perimeter. The choice of outside sources for optimizing the spatial-temporal parameters, the method of optimization, and the intent of the final product's meaning (defining events as single ignition patches vs contiguous burned area)

Product	Area (km2) (Primary event)	Area (km2) (Secondary events)	Number of Secondary Events	Total Area (km2)
MTBS	1039.8	0	0	1039.8
FIRED	892.3	0.4	2	892.8
GWIS	878.8	20	47	898.8
GFA	804.5	88.7	13	893.2

Table 1. Rim fire comparison.

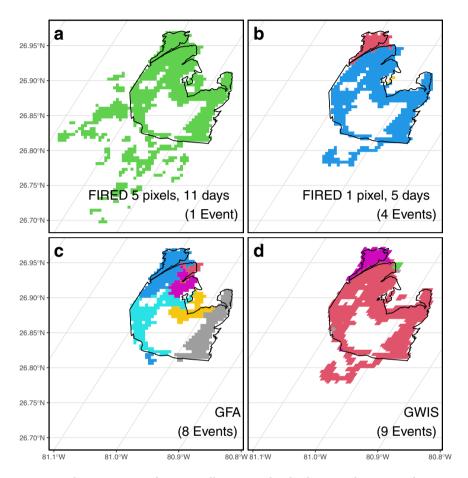


Fig. 3 Product comparison for two small events in Florida, the Moonshine Bay and Sour Orange fires (outlined) that both ignited in February of 2007 and were delineated by MTBS. In b the firedpy product that was optimized for the entire United States with a moving window of 5 pixels, 11 days resulted in aggregation of the two fires delineated by MTBS, but also several smaller fires nearby. In b, it was re-ran with a window of one pixel and five days, for a more realistic result. Delineations by the Global Fire Atlas (c) and the Global Wildfire Information System (d) are shown for comparison.

all lead to different outcomes in the final events that are delineated. Another likely source of this discrepancy is that GWIS and GFA are calibrated to create a single global product. Because different geographical areas have different types of fire regimes, they have fires that grow at different rates and to different sizes, and occur in greater or fewer frequencies, and so the spatial and temporal parameters that work well for defining a fire event in one area may result in over- or under-segmentation in other areas. Here, we decided upon an approach of creating many regional products across the globe, rather than one product for everywhere on earth.

Besides the ease of access and use, the advantage of the FIRED product lies in the user's ability to use the open-source software, FIREDpy, to tailor the spatial and temporal parameters of the moving window algorithm in order to realistically delineate events for their region of interest. In Fig. 3, we illustrate this by comparing the three products for a pair of small fires in Florida. In this case, the FIRED product that was created with a larger moving window (5 pixels and 11 days) over-aggregated the events, but it only required one line of code at command line to recreate the product with a smaller moving window (1 pixel and 5 days) to get more realistic results.

Here, we present a collection of regionally-tailored fire perimeter datasets for every country in the world with significant fire activity<sup>20</sup>, which we created with the open source algorithm, FIREDpy<sup>21</sup>. Each dataset is either a single country or a broader region, depending on the data volume. These datasets differ from other similar

Name	Description	Source
MODIS MCD64A1 Burned Area	Monthly Gridded Burn Date Estimations (463 m)	10
MODIS MOD12Q1 Landcover	Annual Gridded landcover data (463 m)	25
CEC Ecoregions	Ecoregions for North America	23
WWF Ecoregions	World Ecoregions	24

Table 2. Input data.

efforts <sup>14,15</sup> in that each dataset created by FIREDpy is a single file containing a collection of polygons that is generated for the entire time series, rather than monthly or yearly aggregations with a global extent. Furthermore, we have generated the data products at a spatial extent land managers and ecologists would typically use to do regional-scale research, and we adjusted the spatial and temporal parameters for each country to yield realistic event delineations. We also made every effort to ensure that download sizes are reasonable ( < 300 MB). They have a temporal extent from November 2000 to the summer of 2021 at the time of this writing, and they will be actively maintained and updated yearly and upon request. Most importantly the software we developed to generate the datasets is open source and freely available, and so novel fire perimeter datasets can be generated by anyone for any area of interest at any time, using a single command. We hope this will increase the capacity of ecologists and land managers across the world to study fire activity, and incorporate fire history into their work. We invite the broader community to contribute to the continued development of the software package and associated data products.

### Methods

The data were produced using an open source algorithm developed called FIREDpy: fire event delineation for python<sup>21</sup>. Firedpy inputs the gridded Moderate Resolution Imaging Spectroradiometer (MODIS) MCD64A1 burned area product for an area of interest, along with user-specified ancillary data (Table 2), and outputs fire perimeters with their associated characteristics (Tables 3, 4). FIREDpy is available as a software package at <a href="https://github.com/earthlab/firedpy">https://github.com/earthlab/firedpy</a>, and the datasets we have produced are available as a data collection in the University of Colorado's data repository<sup>20</sup>. Balch *et al.*<sup>21</sup> previously used FIREDpy to create fire perimeters for the Coterminous United States, and here we expand upon that work by creating additional country-level products for the rest of the world.

The MODIS burned area product<sup>10</sup> is a gridded product that is produced for every month of the year, starting in November 2000. It is provided to the public in monthly HDF files across a system of globally distributed tiles (Fig. 4). Each monthly HDF file contains five layers: estimated burn date, first day burned, last day burned, burn date uncertainty and a quality control layer. It is also provided in sub-continental windows and as daily shapefiles<sup>22</sup>.

FIREDpy is a data acquisition and processing program that is centered around a spatiotemporal flooding algorithm. It inputs a spatial parameter, temporal parameter, spatial extent and temporal extent. It uses these parameters to create a spatiotemporal moving window that assigns the same event identification number to each pixel that falls within the specified spatial distance and time period from the focal pixel. The spatial parameter is the number of 463.31271653 m pixels and the temporal parameter is the number of days. The spatial extent is an area of interest in the form of a shapefile or MODIS tile names, and the temporal extent is a start and end date. FIREDpy uses the spatial and temporal extent inputs to automatically download all of the monthly MODIS MCD64A1 hdf files within those spatial and temporal extents. It then extracts the burn date layer from each hdf file and creates a spatiotemporal cube of monthly burn dates for the entire time series for the area of interest. It uses the spatial and temporal moving window parameters to assign event identification numbers to each pixel, and then converts the pixels into polygons. The user can choose to have each event as a single polygon, or divided into separate polygons by burn date in order to see the day-to-day progression of the fire (Fig. 2). The user can also specify additional variables to be extracted to each fire event or burn date polygon, such as ecoregions 23,24 and landcover type<sup>25</sup>. See Table 2 for all inputs.

For most countries, we used a spatial parameter of one pixel, and a temporal parameter of five days, following the parameters Artés *et al.* <sup>15</sup> used for their global dataset. For sub-Saharan Africa and Southeast Asia (Cambodia, Vietnam, Thailand, Myanmar and Laos), we used one pixel and two days, based on the analysis of <sup>26,27</sup> and after finding that using one pixel and five days in those regions tended to result in over-aggregation. For the United States and Canada we used 5 pixels and eleven days based on our own validation <sup>21</sup> that agreed with other findings in temperate regions <sup>19</sup>. For smaller countries and regions, we produced both the daily- and event-level polygons and packaged them together in the data repository. For most areas, we only produced the event level polygons due to very large file sizes and data processing times, as the daily polygon files can be an order of magnitude larger than the event files. We encourage those who need the daily-level polygons in those areas to run the algorithm on smaller areas or time periods of interest to yield a manageable data product.

#### **Data Records**

The firedpy github page has links to all of the datasets we have produced (https://github.com/earthlab/firedpy), which are hosted in the Earth Lab Data user collection on CU Scholar<sup>20</sup>. CU Scholar is an "open access institutional repository supporting the research and teaching mission of the University of Colorado Boulder". On CU Scholar, each country or region is listed as FIRED followed by the name of the country or region. Each dataset is stored as a zip file that has several files inside. The first file is a readme text file that serves as the metadata. It

Variable	Description	Data Type	Event, Daily or Both
id	Fire Event ID #	integer	Both
did	Day id #	character	Daily
ig_date	Date of ignition	date	Both
ig_day	Ignition day of year	integer	Both
ig_month	Ignition month	integer	Both
ig_year	Ignition year	integer	Both
last_date	last date a pixel burned	date	Both
event_dur	duration of the event	integer	Both
tot_pix	total pixels burned	integer	Both
pixels	pixel burned that day	integer	Daily
tot_ar_km2	total burned area (km2)	float	Both
dy_ar_km2	area burned that day (km2)	float	Daily
fsr_px_dy	fire spread rate (pixels per day)	float	Both
fsr_km2_dy	fire spread rate (km2 per day)	float	Both
mx_grw_px	maximum growth in pixels	float	Both
mn_grw_px	minimum growth in pixels	float	Both
mu_grw_px	mean growth in pixels	float	Both
mx_grw_km2	maximum growth in km2	float	Both
mn_grw_km2	minimum growth in km2	float	Both
mu_grw_km2	mean growth in km2	float	Both
mx_grw_dte	date of maximum growth	date	Both
ig_utm_x	coordinates of ignition	float	Event
ig_utm_y	coordinates of ignition	float	Event
tot_perim	length of the fire perimeter	float	Event
geom	geometry of the event	list of vertices	Both

**Table 3.** List of primary variables in the FIRED datasets. All variable names are 10 characters or less in order to be compatible with the popular .shp file format.

Variable	Description	Data Type	Event, Daily or Both
eco_mode	numeric ecoregion code	integer	Both
eco_name	name of ecoregion	character	Both
eco_type	ecoregion data source	character	Both
lc_mode	numeric landcover code	integer	Both
lc_name	landcover type name	character	Both
lc_type	source of landcover data	character	Both

**Table 4.** List of secondary variables in the FIRED datasets. These are generated only if explicitly specified by the user. All variable names are 10 characters or less in order to be compatible with the popular .shp file format.

contains a short description of the data that includes the name of all countries included, the temporal extent, the spatial and temporal parameters used for the flooding algorithm, followed by a list of the definitions of each variable that is included in the attribute table. The variables provided for the event and daily products are provided on Tables 3 and 4. The second file is the events shapefile in geopackage (.gpkg) format. Files three to six are the four files that make up an ESRI shapefile, which have the same base name and different extensions: (.shp, .dbf, .prj, .shx). The last file, if included, is a geopackage of the events split into the day to day progression. The default naming convention is fired\_<AOI>\_to<YYYYDOY>\_<events or daily>.gpkg, where AOI is replaced by the name of the shapefile of the area of interest that is supplied as an input, YYYYDOY is the year and day of year of the end of the time series, and the last part of the base file name is whether each polygon is a whole event or day in the progression. An example for Bolivia that ends in July 2021 would be fired\_bolivia\_to2021182\_events.gpkg.

## **Technical Validation**

When we created the fire perimeter datasets for the conterminous United States, we produced 225 iterations of fire perimeters—one iteration for each combination of spatial and temporal parameters between 1 and 15 pixels and days, respectively, that define the size of the moving window. We filtered the dataset to only include fires that overlapped spatially and temporally with fire events from the Monitoring Trends in Burn Severity dataset<sup>11</sup>. We then compared all of the iterations, and selected the spatial temporal combination (5 pixels, 11 days) that

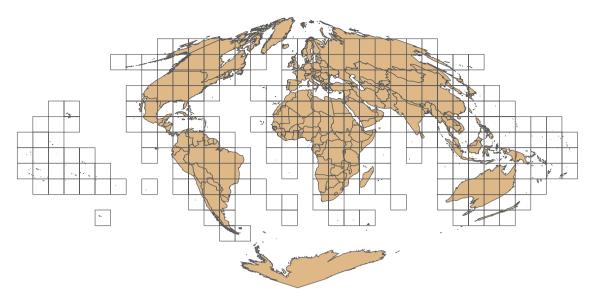


Fig. 4 The tile system used by most MODIS products. MODIS tiles cover the entire globe, but here we show tiles that are included in the MCD64A1 burned area product.

minimized both over- and under-segmentation. In other words, we aimed to find the spatial and temporal parameters that resulted in fire event perimeters that were a one to one match with the Landsat-derived MTBS perimeters. The relationship between MTBS and the spatiotemporally optimized FIRED for burned area is generally strong ( $R^2 = 0.92$  for all events, Balch *et al.*<sup>21</sup>), but it is weaker for smaller fires. Comparing among size classes revealed weaker relationships below 70,000 ha ( $R^2$  values between 0.5 and 0.8). Comparison with MTBS also revealed that the MODIS MCD64A1 burned area product also consistently underestimates burned area for fires below 100,000 hectares, but this is likely an artifact of MTBS not including unburned patches within its perimeters. This can be seen in Fig. 1, where the rim fire perimeter from MTBS has within its boundaries two large unburned areas. These validation methods are described in greater detail in<sup>21</sup>.

For the remainder of the globe we chose more restrictive combinations because we found that using 5 pixels and 11 days often resulted in over-aggregation. Since we were not able to find consistent validation data with which to optimize our parameters outside of the US, we followed the spatial and temporal parameters used by other studies that used the MODIS burned area product to delineate events (see Methods).

#### **Usage Notes**

A typical call to the firedpy command line interface is as follows:

```
firedpy -spatial 1 -temporal 5 -aoi
/home/firedpy/ref/individual countries/bolivia.qpkg -shp type qpkg
```

First, it specifies spatiotemporal parameters, then the area of interest (countries, continents, US states and Australian states that have been reprojected to the MODIS sinusoidal projection are provided in a reference folder) and the output shapefile type. All parameters except area of interest are set to sensible defaults (daily defaults to no, start and end date default to the entire record). The user can also simply type "firedpy", and the software will prompt the user for the input parameters.

Running firedpy in its current form can be computationally intensive. Exactly how intensive it is depends on the spatial and temporal extents, as well as the volume of fire activity experienced by the target region. In particular it has very high requirements for random access memory (RAM). The conterminous United States, for example, can be generated on a personal computer with 16 GB of RAM. Brazil has a smaller spatial extent, but about ten times the fire activity, and needs about 100 GB of RAM to generate. Many countries in sub-Saharan Africa as well as Australia have higher resource requirements. The need for high computational requirements is one of the main reasons we generated datasets for every country, rather than simply providing the software as a tool. We are actively developing the code, and expect the software to be much more efficient in the near future. Users who do wish to create a product on their own with firedpy may wish to use the docker container we created for the purpose of running firedpy on a cloud computing platform like Cyverse (https://cyverse.org/). Instructions for using firedpy in the docker container are on the GitHub page.

# Code availability

All code is open source and available at https://github.com/earthlab/firedpy. The data we produced is available at the Earth Lab Data collection at CU Scholar<sup>20</sup>. Links to each dataset are provided on the front page of the aforementioned github repository, and are provided here in Table S1. The docker container with a custom software environment for running firedpy is at https://hub.docker.com/repository/docker/earthlab/firedpy.

Received: 25 January 2022; Accepted: 19 July 2022;

Published online: 30 July 2022

### References

- 1. Andela, N. et al. A human-driven decline in global burned area. Science 356, 1356-1362 (2017).
- 2. Bowman, D. M. J. S. et al. Fire in the Earth System. Science 324, 481-484 (2009)
- 3. Dennison, P. E., Brewer, S. C., Arnold, J. D. & Moritz, M. A. Large wildfire trends in the western United States, 1984–2011. *Geophys. Res. Lett.* 2928–2933, https://doi.org/10.1002/2014GL059576 (2014).
- 4. Moritz, M. A. et al. Climate change and disruptions to global fire activity. Ecosphere 3, art49 (2012).
- 5. Westerling, A. L. & Swetnam, T. W. Interannual to decadal drought and wildfire in the western United States. *Eos Trans. Am. Geophys. Union* 84, 545 (2003).
- 6. Briones-Herrera, C. I. *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 Mexico. *Remote Sens.* 12, 1–19 (2020).
- Giglio, L., Schroeder, W. & Justice, C. O. The collection 6 MODIS active fire detection algorithm and fire products. Remote Sens. Environ. 178, 31–41 (2016).
- 8. Schroeder, W., Oliva, P., Giglio, L. & Csiszar, I. A. The New VIIRS 375m active fire detection data product: Algorithm description and initial assessment. *Remote Sens. Environ.* **143**, 85–96 (2014).
- 9. Chuvieco, E. *et al.* Generation and analysis of a new global burned area product based on MODIS 250 m reflectance bands and thermal anomalies. *Earth Syst. Sci. Data* **10**, 2015–2031 (2018).
- Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L. & Justice, C. O. The Collection 6 MODIS burned area mapping algorithm and product. Remote Sens. Environ. 217, 72–85 (2018).
- 11. Eidenshink, J. et al. A Project for Monitoring Trends in Burn Severity. Fire Ecol. 3, 3–21 (2007).
- 12. Short, K. C. Sources and implications of bias and uncertainty in a century of US wildfire activity data. *Int. J. Wildland Fire* 24, 883–891 (2015).
- 13. Kolden, C. A., Smith, A. M. S. & Abatzoglou, J. T. Limitations and utilisation of Monitoring Trends in Burn Severity products for assessing wildfire severity in the USA. *Int. J. Wildland Fire* 24, 1023–1028 (2015).
- 14. Andela, N. et al. The Global Fire Atlas of individual fire size, duration, speed, and direction. Earth Syst. Sci. Data Discuss. 1–28, https://doi.org/10.5194/essd-2018-89 (2018).
- 15. Artés, T. et al. A global wildfire dataset for the analysis of fire regimes and fire behaviour. Sci. Data 6, 296 (2019).
- 16. Mahood, A. L. & Balch, J. K. Repeated fires reduce plant diversity in low-elevation Wyoming big sagebrush ecosystems (1984–2014). *Ecosphere* 10, e02591 (2019).
- 17. Rodman, K. C., Veblen, T. T., Saraceni, S. & Chapman, T. B. Wildfire activity and land use drove 20th-century changes in forest cover in the Colorado front range. *Ecosphere* 10 (2019).
- 18. de Santana, M. M. M., Mariano-Neto, E., de Vasconcelos, R. N., Dodonov, P. & Medeiros, J. M. M. Mapping the research history, collaborations and trends of remote sensing in fire ecology. *Scientometrics* 126, 1359–1388 (2021).
- 19. Frantz, D., Stellmes, M., Röder, A. & Hill, J. Fire spread from MODIS burned area data: obtaining fire dynamics information for every single fire. *Int. J. Wildland Fire* 25, 1228 (2016).
- 20. Mahood, A. Earth Lab Data. University of Colorado Boulder https://doi.org/10.25810/vr03-8y36 (2022).
- 21. Balch, J. K. et al. FIRED (Fire Events Delineation): An open, flexible algorithm & database of US fire events derived from the MODIS burned area product (2001–19). Remote Sens. 12, 3498 (2020).
- 22. Giglio, L. et al. Collection 6 MODIS Burned Area Product User's Guide Version 1.3. 34 (2020).
- 23. 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).
- Olson, D. M. et al. Terrestrial Ecoregions of the World: A New Map of Life on EarthA new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience 51, 933–938 (2001).
- Sulla-Menashe, D. & Friedl, M. A. User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product. Usgs 1–18, https://doi.org/10.5067/MODIS/MCD12Q1 (2018).
- 26. Archibald, S., Roy, D. P., van Wilgen, B. W. & Scholes, R. J. What limits fire? An examination of drivers of burnt area in Southern Africa. *Glob. Change Biol.* 15, 613–630 (2009).
- 27. Oom, D., Silva, P., Bistinas, I. & Pereira, J. Highlighting Biome-Specific Sensitivity of Fire Size Distributions to Time-Gap Parameter Using a New Algorithm for Fire Event Individuation. *Remote Sens.* 8, 663 (2016).

### **Acknowledgements**

Special thanks to Lise A. St Denis, Travis Williams Nathan P. Mietkiewicz, Sepideh Dadashi, and Joe McGlinchy for their contributions to the conceptualization and coding of the algorithm. We also are extremely grateful to Ryan Caillet at CU Scholar. In addition, we are grateful for the constructive feedback from the editor and two anonymous reviewers, which greatly improved the manuscript.

#### **Author contributions**

Adam L. Mahood led the writing of the paper, managed the project and created many of the data products. Estelle J. Lindrooth created many of the data products and updated code for the algorithm. Maxwell C. Cook contributed to updating the code for the algorithm. Jennifer K. Balch provided funding and conceptualized the project as a whole.

### **Competing interests**

The authors declare no competing interests.

# **Additional information**

**Supplementary information** The online version contains supplementary material available at https://doi. org/10.1038/s41597-022-01572-3.

**Correspondence** and requests for materials should be addressed to A.L.M.

Reprints and permissions information is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2022