

A Protocol for Collecting Burned Area Time Series Cross-Check Data

Harry R. Podschwit ^{1,2,*}, Brian Potter ³ and Narasimhan K. Larkin ³

¹ College of the Environment Special Programs, Quantitative Ecology & Resource Management (QERM), University of Washington, Seattle, WA 98195, USA

² Missoula Fire Sciences Laboratory, U.S. Forest Service, Rocky Mountain Research Station, W. Broadway Street, Missoula, MT 59808, USA

³ Pacific Wildland Fire Sciences Laboratory, U.S. Forest Service, 400 N. 34th Street #201, Seattle, WA 98103, USA

* Correspondence: harryp@uw.edu

† Former affiliation.

Abstract: Data on wildfire growth are useful for multiple research purposes but are frequently unavailable and often have data quality problems. For these reasons, we developed a protocol for collecting daily burned area time series from the InciWeb website, Incident Management Situation Reports (IMSRs), and other sources. We apply this protocol to create the Warehouse of Multiple Burned Area Time Series (WoMBATS) data, which are a collection of burned area time series with cross-check data for 514 wildfires in the United States for the years 2018–2020. We compare WoMBATS-derived distributions of wildfire occurrence and size to those derived from MTBS data to identify potential biases. We also use WoMBATS data to cross tabulate the frequency of missing data in InciWeb and IMSRs and calculate differences in size estimates. We identify multiple instances where WoMBATS data fails to reproduce wildfire occurrence and size statistics derived from MTBS data. We show that WoMBATS data are typically much more complete than either of the two constituent data sources, and that the data collection protocol allows for the identification of otherwise undetectable errors. We find that although disagreements between InciWeb and IMSRs are common, the magnitude of these differences are usually small. We illustrate how WoMBATS data can be used in practice by validating two simple wildfire growth forecasting models.

Keywords: data cleaning; data collection; InciWeb; uncertainty; missing data; wildfire growth; spread; bias



Citation: Podschwit, H.R.; Potter, B.; Larkin, N.K. A Protocol for Collecting Burned Area Time Series Cross-Check Data. *Fire* **2022**, *5*, 153. <https://doi.org/10.3390/fire5050153>

Academic Editor: James A. Lutz

Received: 19 August 2022

Accepted: 23 September 2022

Published: 29 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rapid wildfire growth can have numerous significant effects on anthropogenic and environmental systems. Fast spreading fire has been implicated in dramatic changes in vegetation structure and composition [1,2], which in turn can cause negative downstream effects on water [3,4] and soil conditions [5]. The sudden emission of large quantities of smoke [6] and the rapidly advancing flames of a wildfire can pose a threat to human [7,8] and non-human health [9,10]. In addition to the associated safety risks, fast-spreading fires are particularly difficult for firefighters to control [11]. Despite fire spread's relevance to these impacts, final fire size is perhaps the most common wildfire characteristic researchers analyze [12]. This methodological bias is unfortunate because, in many cases, the aforementioned impacts occur in fairly moderate-sized fires. The Rattlesnake fire of 1953, the Oakland Hills fire of 1991, and the Cramer fire of 2003 were rapidly spreading events and significant in terms of loss of life [13–15], but not extraordinarily large. The Rattlesnake fire only burned approximately 500 hectares [13], the Oakland hills fire burned about 600 hectares [14], and the Cramer fire burned approximately 750 hectares (see www.mtbs.gov; accessed on 18 September 2022). Final fire size can also be an inappropriate proxy of ecological impacts. Fast-growing fires will often burn an area at high

intensity [1], which can severely affect ecosystem functions such as water and soil quality over a relatively small-but ecologically important-area [16]. Populations of fire-sensitive wildlife limited to small geographic areas can be severely impacted by high-severity fires, even if not particularly large [17,18].

For these reasons, there is a need to not only investigate the drivers and effects of large fire occurrences, but also the effects and drivers of daily variability in fire growth. To that end, estimates of the size of an individual wildfire over regular intervals of time, hereafter referred to as burned area time series (BATS), are immensely useful to researchers seeking a holistic understanding of relationships between wildfire and the environment. BATS are used to answer research questions that cannot be revealed with final wildfire size data alone, and BATS have been critical to modeling how temporal variation in wildfire growth impacts vegetation [1], public health [8], and firefighting effectiveness [11]. BATS data have also been useful in identifying factors that are relevant to predicting fire growth [19–22], estimating fire spread rates [23], and validating predictions from wildfire spread models [24].

Although clearly informative, BATS are notoriously uncommon [25], and data about the location, date, and final size of wildfires are often all that are available. For instance, in the United States, although the Monitoring Trends in Burn Severity (MTBS) project has systematically recorded detailed information about the final burned area and severity using Landsat satellite data going back to 1984 [26], it cannot be used to determine daily variation in fire growth. One source of progression data, and a large amount of other information, is from wildfire case studies. Unfortunately, less than one-tenth of one percent of wildfires are documented in this manner [27], and the reports must be heavily processed before they can be used in scientific studies. Historical administrative records are a more commonly available source of progression data [25], which omit many of the details provided by case studies but are more widely available and provide a structured presentation of information. However, these data frequently require some preprocessing in order to convert raw data into products useful for scientific research [28] and are not available for all wildfires. In the United States, Incident Status Summary (i.e., ICS-209) forms are an example of historical administrative records that can produce BATS [28]. ICS-209s are forms that are submitted throughout most wildfires' lifetimes, which record up to 53 blocks of information, including incident name, location, ignition date, current size, containment levels, and ignition cause [29]. ICS-209s are closely related to Incident Management Situation Reports (IMSRs), which are released near-daily to summarize wildfire activity across the United States from the previous day's evening ICS-209s and other reports [30]. Where ICS-209s are issued daily or more often for a single wildfire, IMSRs are issued once daily to summarize multiple wildfires in a single report. In addition to case studies or historical administrative reports, recent advances in cloud-based computing have made satellite-based data increasingly accessible for research and analysis. Satellite data can produce BATS using data that are available globally and sampled at regular intervals [31]. Still, many of the problems associated with other data sources remain. Satellite data can have missing size measurements [32,33], can require extensive pre-processing to identify individual fires [34] and fill in missing days [32], and may not accurately reflect actual size estimates [35]. In addition to the case studies, historical administrative records, and satellite data, partial BATS can also be produced using information from newspapers [36], photography, personal communications [27], hand-drawn progression maps [37], and web-based information [27,38].

Although BATS can be produced from many data sources, the data are not available for all wildfires and even when available, are often incomplete and contain errors [32,35]. Missing observations are fairly common in historical administrative records and can arise for a number of reasons. For example, if a wildfire is not under full suppression or if little growth is anticipated, then reporting may only be required weekly [29]. A wildfire may be undetected in the early-days of the incident's lifetime or may not be large enough to warrant reporting [29]. Missing data can also arise from changes in incident management [29], such as when an individual wildfire switches to being managed and reported as part of a

complex. In addition to missing data, measurement errors are another common problem. In some cases, measurement errors are obvious and can be identified without consulting multiple data sources. For example, ICS-209 data will often produce a BATS that decreases at some point in time [28]. Since fires cannot unburn the landscape, these errors are easy to detect and often occur when improved mapping corrects initial overestimates of burned area. However, errors within an expected range (e.g., larger than the $t - 1^{th}$ observation and smaller than the $t + 1^{th}$ observation) require more effort to identify [39]. These errors are somewhat ubiquitous and can arise for a variety of reasons. Satellite-derived size estimates are likely to disagree with ground-based observations, particularly in topographically complex locations [35]. Transcription errors can easily produce plausible, but ultimately incorrect size measurements. Wildfire size changes over the course of a day and measuring at irregular intervals (i.e., reporting size in the morning one day and the evening on another) can result in inconsistencies across BATS data sources. Although these inconsistencies may not technically be errors in that they faithfully report the size at the time the measurements were taken, because BATS are presumed to represent a sample of size over regular intervals of time, they can become a source of measurement uncertainty when used to estimate daily size.

These missing and erroneous observations are a relatively common characteristic of BATS data and are problematic for scientific applications [35]. For that reason, various kinds of data cleaning methods have been proposed to complete and correct BATS [39], including simple rule-based corrections like those used to produce ICS-209-PLUS data [28], complex statistical models [32], and hand-cleaning by analysts [40]. Although it is a relatively time-consuming process, cross-checking the BATS against other information has two major benefits. Firstly, cross-check data can explicitly fill in missing observations without making assumptions about wildfire progression. Secondly, cross-checking data can gauge measurement reliability in a way that is impossible in the absence of multiple data sources [39].

Given that (1) there is a dearth of BATS data and (2) consultation of multiple data sources appears likely to improve data completeness and accuracy, we would like to develop a procedure to collect BATS data in a documented and transparent manner [38]. To that end, we describe and apply a method of aggregating wildfire growth information from multiple data sources to build a novel BATS dataset based on ground-based observations. We will use these data to assess measurement uncertainty and demonstrate its application in research contexts. The methods are described in Section 2, which is organized into two subsections: a data collection protocol subsection and a dataset analyses subsection. In the data collection protocol subsection, we describe the data collection methods generally and also demonstrate their specific application to the Saddleridge wildfire. The aforementioned data collection protocol is applied to produce the warehouse of multiple burned area time series (WoMBATS) data, which are a collection of BATS with cross-check data for 514 wildfires in the United States between 2018 and 2020. In the dataset analyses subsection, we present the methods used to gauge the quality of WoMBATS data and a description of the example model validation exercise that is performed to illustrate the application of WoMBATS data in research. In Section 3, the results of the three analyses described in the previous section are presented. In Section 4, we discuss the advantages and limitations of our protocol and highlight potential future research directions. In Section 5, we summarize our overall conclusions and recommendations.

2. Methods

2.1. Data Collection Protocol

2.1.1. Overview

The data collection protocol generates two products: a BATS table and a metadata table. The BATS table reports daily size estimates of an individual wildfire from multiple sources. The metadata table provides information on the data sources used to estimate daily size. These two data products are created through a three-step workflow. In the first step,

InciWeb data are used to create a case list of wildfires and record important information such as the ignition date and location. Second, InciWeb webscrapes, Incident Management Situation Reports (IMSR), and other data are consulted to build individual daily BATS for each wildfire in the case list. In the third step, the BATS are aligned side-by-side in a table (Figure 1), and the source and quality of data are reported in the metadata table.

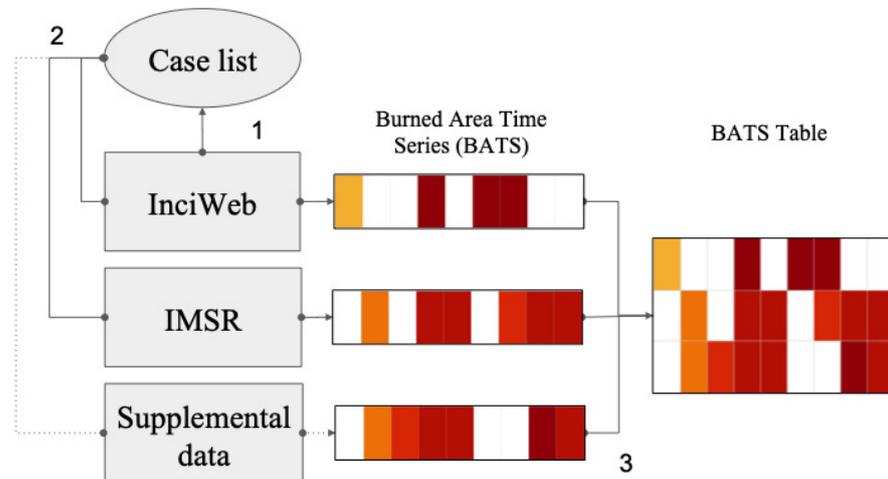


Figure 1. Graphical depiction of three-step workflow of data collection protocol. (1) a case list of wildfires that are to have data collected is produced from InciWeb. (2) burned area time series, represented here as a colored grid, are produced for all wildfires in the case list. (3) the burned area time series are combined into a single table. Increasingly red colors in the grid represent increases in reported fire size and white colors represent missing data. Numbers represent step number of workflow.

2.1.2. InciWeb Case List

Data collected from the InciWeb (www.inciweb.org; accessed between 2 June 2018 and 31 December 2020) website are used to create an initial case list and establish basic information about the wildfire. Information from InciWeb comes from a variety of sources that can overlap with that used to produce IMSRs but there are often instances where the sources are not the same. For example, late-evening infrared measurements may be reported in InciWeb, but not IMSRs. A scheduled webscraper queries the InciWeb website at 5:00 UTC (22:00 PST) each day. The webscraper extracts 12 variables for each wildfire reported on the InciWeb website including size, event category, location name, ignition date, url, latitude, longitude, and name. This web-scraping program was run near-daily between 2 June 2018 and 31 December 2020.

The daily webscrape data are aggregated into a master dataset and a case list of unique events is obtained using the event urls as identifiers. The list was filtered to remove all non-wildfire events; wildfires with no daily size exceeding 405 ha; wildfires with the text “complex” in the name; and wildfires with an unknown ignition date.

2.1.3. Construction of BATS

The measurements reported in the master dataset are the same as if one visited the InciWeb website at the same time every day and recorded the size reported on the website. These data are used to produce an initial partial BATS that is next cross-checked against IMSRs and other data.

Initial BATS cross-check data are obtained by querying IMSRs (<https://www.predictiveservices.nifc.gov/intelligence/archive.htm>; accessed between 5 February 2021 and 27 September 2022). Beginning on the ignition date identified from the InciWeb data, the relevant IMSRs are manually searched to see if the relevant wildfire name is reported. If yes, the size is recorded. Otherwise, the size is recorded as NA. Once the IMSR no longer

reports the wildfire's size, the query is terminated. The partial BATS produced from IMSRs are back-shifted by one day to account for reporting delays [30]. After consulting the IMSR data, it is likely that the resulting BATS will still have days that have missing observations and/or InciWeb data that has not been cross-checked. Additional cross-check data could come from a number of sources [27,31,37–39], but these data will not be included using the systematic structured methods as were used for InciWeb and IMSRs. Instead, we will include these additional data sources as convenient using an unstructured search.

We assume that InciWeb and IMSR data provide point estimates of size since there is, by definition, only one measurement per day, but other data sources may report multiple observations in a day or offer an interval estimate instead. Rather than discard this information, we will report two partial BATS: a daily minimum and a daily maximum BATS. In most instances, this interval estimate will be redundant since-like the InciWeb and IMSR data-there will only be a single measurement associated with a given day. However, when measurement variability is observed, interval estimates can document this in a way that is superior to point estimates.

The BATS table is produced by aligning the individual BATS side-by-side. The URL, access date, and quality category (low, medium, or high) are recorded for each web-based data source in a metadata table. Data quality is defined as “high” if it is a government, land, or fire management agency; “medium” if it is a credible secondary source such as a newspaper; and it is “low” otherwise. InciWeb and IMSR data are classified in the “high” data quality category by default.

2.1.4. Example

The Saddleridge wildfire ignited in Southern California on 10 October 2019, and is used to demonstrate the real-world application of the data collection protocol. The master dataset of records produced from the InciWeb webscraper associated with the Saddleridge wildfire is shown below (Table 1). Note that the wildfire has at least 1 observation exceeding 405 ha, has a known ignition date and location, and is not named as part of a complex. Note also that the BATS produced by InciWeb is incomplete, with observations unavailable for the first five days of the wildfire. The mandatory cross-check data produced from the IMSRs does not fill in any of these missing days, but corroborates most of the InciWeb estimates following 17 October (Table 2).

Table 1. InciWeb master dataset for the Saddleridge fire (2019).

Access date	Ignition Date	Latitude	Longitude	Hectares	URL
15 October 19	10 October 19	34.326	−118.481	3396	http://inciweb.nwcg.gov/incident/6643/
16 October 19	10 October 19	34.326	−118.481	3396	http://inciweb.nwcg.gov/incident/6643/
17 October 19	10 October 19	34.326	−118.481	3396	http://inciweb.nwcg.gov/incident/6643/
18 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
19 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
20 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
21 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
22 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
23 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
24 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
25 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
26 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
27 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
28 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
29 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/
30 October 19	10 October 19	34.326	−118.481	3561	http://inciweb.nwcg.gov/incident/6643/

Table 2. Incident Management Situation Report cross-check data for the Saddleridge fire (2019). Accessed on 27 September 2022.

Date	Hectares	URL
10 October 19	NA	https://www.predictiveservices.nifc.gov/IMSR/2019/20191011IMSR.pdf
11 October 19	NA	NA
12 October 19	NA	NA
13 October 19	NA	NA
14 October 19	NA	NA
15 October 19	NA	NA
16 October 19	NA	NA
17 October 19	3396	https://www.predictiveservices.nifc.gov/IMSR/2019/20191018IMSR.pdf
18 October 19	3561	https://www.predictiveservices.nifc.gov/IMSR/2019/20191019IMSR.pdf
19 October 19	3561	https://www.predictiveservices.nifc.gov/IMSR/2019/20191020IMSR.pdf
20 October 19	3561	https://www.predictiveservices.nifc.gov/IMSR/2019/20191021IMSR.pdf
21 October 19	3561	https://www.predictiveservices.nifc.gov/IMSR/2019/20191022IMSR.pdf
22 October 19	3561	https://www.predictiveservices.nifc.gov/IMSR/2019/20191023IMSR.pdf
23 October 19	3561	https://www.predictiveservices.nifc.gov/IMSR/2019/20191024IMSR.pdf
24 October 19	3561	https://www.predictiveservices.nifc.gov/IMSR/2019/20191025IMSR.pdf
25 October 19	3561	https://www.predictiveservices.nifc.gov/IMSR/2019/20191026IMSR.pdf
26 October 19	3561	https://www.predictiveservices.nifc.gov/IMSR/2019/20191027IMSR.pdf
27 October 19	3561	https://www.predictiveservices.nifc.gov/IMSR/2019/20191028IMSR.pdf
28 October 19	3561	https://www.predictiveservices.nifc.gov/IMSR/2019/20191029IMSR.pdf
29 October 19	NA	https://www.predictiveservices.nifc.gov/IMSR/2019/20191030IMSR.pdf
30 October 19	NA	https://www.predictiveservices.nifc.gov/IMSR/2019/20191031IMSR.pdf

At this point, although the two mandatory sources have been consulted, the size is still unknown for the first five days. Other data sources are sampled to fill in these values and better assess the variability in daily size measurements. Specifically, additional high-quality cross-check data are available in the form of an incident narrative by the Los Angeles Fire Department (LAFD). The utility of interval size estimates becomes apparent here as, on some days, the narrative reports the size of the fire multiple times. These supplemental cross-check data are reported in Table 3 and are available from: <https://www.lafd.org/news/saddle-ridge-brush-fire> (Accessed on 27 September 2022). The aggregation of all three data sources describes the fire progression and measurement uncertainty in a way that no one of the basis sources could (Figure 2).

Table 3. Supplemental BATS data for the Saddleridge fire (2019) from the Los Angeles Fire Department (<https://www.lafd.org/news/saddle-ridge-brush-fire>; Accessed on 27 September 2022). Daily size is reported in hectares. The local time of the narrative are reported in the parenthesis.

Date	Minimum Size (Pacific Standard Time)	Maximum Size (Pacific Standard Time)
10 October 19	24 (22:55)	
11 October 19	647 (00:19)	3052 (17:00)
12 October 19	3056 (10:45)	3223 (19:00)
13 October 19	3223 (8:00)	3223 (18:00)
14 October 19	3223 (7:00)	3396 (21:00)
15 October 19	3396 (7:00)	3396 (21:00)
16 October 19	3396 (9:00)	3396 (19:00)
17 October 19	3396 (7:00)	3396 (19:00)
18 October 19	3396 (7:00)	3561 (19:00)
19 October 19	3561 (19:00)	3561 (19:00)
20 October 19	3561 (19:00)	3561 (19:00)
21 October 19		
22 October 19	3561 (17:00)	3561 (17:00)

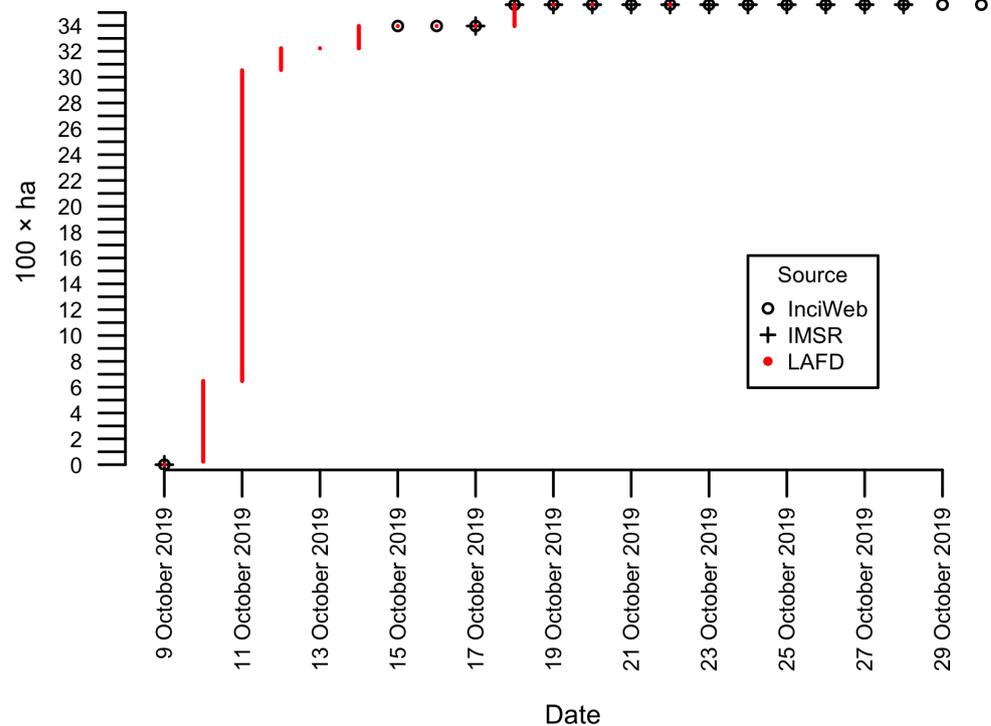


Figure 2. Composite burned area time series from three data sources: InciWeb webscrapes, Incident Situation Management reports (IMSRs), and narratives from the Los Angeles Fire Department (LAFD).

2.2. Dataset Analyses

2.2.1. Overview

The Saddleridge example was just one of 514 wildfires in the filtered case list for 2018–2020, and the data collection protocol was applied to each. The data produced from this procedure are available at: https://figshare.com/articles/dataset/WOMBATS_basic_data_2018-2020/14788206; DOI: 10.6084/m9.figshare.14788206 (Accessed on 27 September 2022). This collection of BATS tables, hereafter the warehouse of multiple burned area time series (WoMBATS) data, were subjected to three analyses. In the external comparison analysis, we compare the distribution of variables derived from the WoMBATS data to those derived from MTBS fire occurrence data. In the internal comparison analysis, we compare the completeness and consistency of BATS produced from the two main sources of WoMBATS data: InciWeb webscrapes and IMSRs. In the application analysis, we illustrate how WoMBATS data can be used in research by validating the predictions from simple fire growth models. All analyses were performed in the R programming language [41].

2.2.2. External Comparison

MTBS data spans from 1984 to 2016, and systematically records fire information across all 50 states and are subject to quality controls (see <https://www.mtbs.gov/faqs>; Accessed on 27 September 2022). Non-wildfires and fires that were not large (i.e., that did not exceed 405 ha) were omitted from further analysis for both the WoMBATS and MTBS data.

Wildfire counts were modeled using parametric and non-parametric approaches. The parametric approach assumes that the wildfire counts follow a binomial distribution and models the space-time distribution of counts conditional on a known total number of fires. For each state i , the total number of large fires nationally was used as the number of trials (n) parameter, and the empirical frequency of large fires was used as the binomial model's success probability (p_i):

$$n = \# \text{ of large fires.} \quad (1)$$

$$p_i = \frac{\# \text{ of large fires in state } i}{\# \text{ of large fires}}; \quad (2)$$

Similarly, for each state i and month t , the number of large fires within the state was used as the number of trials (n_i) parameter, and the empirical frequency of large fires in each month and state was used as the success probability (p_{it}) parameter:

$$n_i = \# \text{ of large fires in state } i. \quad (3)$$

$$p_{it} = \frac{\# \text{ of large fires in state } i \text{ and month } t}{\# \text{ of large fires in state } i}; \quad (4)$$

These parameters were estimated from MTBS data and were used to calculate the central 95th percentile of expected values for WoMBATS-derived wildfire counts. The non-parametric approach uses MTBS data to construct a sample of wildfire counts over a time period analogous to the WoMBATS data and describes the absolute frequency of fires. For each state i , the number of large wildfires between 2 June, T and 31 December, $T + 2$ was calculated for every $T \in \{1984, 1985, \dots, 2014\}$. This same method was applied to each state and month to produce the non-parametric version of the monthly wildfire counts. The distribution of wildfire size were described for each state using empirical cumulative distribution functions. Two-sample Kolmogorov–Smirnov tests were used to compare the empirical distribution of wildfire size derived from WoMBATS and MTBS data, and identify any states where the two distributions differ.

2.2.3. Internal Comparison

In this analysis, we described the level of dependence between the two required sources (InciWeb webscrapes and IMSRs) used to produce the WoMBATS data with three methods. First, we described the typical data completeness of BATS and reported the frequency that size measurements are available from both, one, or none of the sources. Second, we determined if measurements from one source tended to be systematically larger or smaller than the other by reporting the relative frequency of days in the BATS with these characteristics. Third, we measured how close IMSR size measurements were to InciWeb size measurements when both were available.

Each day of each of the 514 wildfires produced from the case list was classified into one of four self-explanatory data availability categories: “Both”, “IMSR-only”, “InciWeb-only”, or “None”. Each day with cross-check data was also classified into one of three bias categories: “IMSR>InciWeb”, “IMSR<InciWeb”, “Exact”. The average proportion of the BATS in each of these categories was recorded. The wildfires with cross-check data were also used to calculate the relative dispersion of size measurements using the average coefficients of variation (CV):

$$CV_{it} = \frac{\sigma_{it}}{\mu_{it}} = \frac{|X_{it1} - X_{it2}|/\sqrt{2}}{(X_{it1} + X_{it2})/2}; \quad (5)$$

Here μ_{it} represents the average of the measurements for individual wildfire i on day t , and σ_{it} represents the standard deviation. Because there are at most two values per day and wildfire—one value from an IMSR and one value from an ICS209—this quantity is equivalent to $\sqrt{2}$ times the range of the burned area measurements divided by the sum of burned area measures. For each wildfire, the daily CV was averaged across the days with cross-check data. In addition to size measurements, a square root data transformation was applied to size measurements to convert them into pseudolinear-spread measurements. The average CV of the pseudolinear-spread measurements were used to classify each wildfire with cross-check data into one of four data quality categories: “Exact”, “Close”, “Adequate”, and “Inadequate”. If there was no disagreement between the InciWeb and IMSR data, it was classified as “Exact”. If the CV was less than 2.5 percent, then it was classified as “Close”. If the CV was less than 35 percent, it was classified as “Adequate”.

If the CV was equal to or greater than 35 percent, it was classified as “Inadequate”. These categories are based on published data quality thresholds [42].

2.2.4. Application

In this analysis, we used WoMBATS data to measure the relative performance of two simple fire growth models that predict wildfire size on day t using the size estimates from day $t - 1$ and day $t - 2$. The areal persistence model predicts that the next day’s size equals the current size plus today’s growth (Equation (6)).

$$\hat{X}_t = X_{t-1} + (X_{t-1} - X_{t-2}) = 2X_{t-1} - X_{t-2}; \quad (6)$$

Here \hat{X}_t is the size estimate on day t , and X_{t-2} and X_{t-1} are the two previous daily size observations. The areal persistence model described here is already used within the BlueSky modeling framework to produce more realistic smoke forecasts than those based on current size estimates [43]. Since the area of a circle increases quadratically with each unit increase in radius, this model of fire growth implicitly assumes that tomorrow’s average fire spread rate is less than today’s. An alternative model of fire growth might instead use radial persistence, where tomorrow’s size estimate is produced assuming a constant radial spread (Equation (7)).

$$\hat{X}_t = (\sqrt{X_{t-1}} + (\sqrt{X_{t-1}} - \sqrt{X_{t-2}}))^2 = (2\sqrt{X_{t-1}} - \sqrt{X_{t-2}})^2; \quad (7)$$

The performance of the models described in (Equations (6) and (7)) were based on a sample of fire size predictions that were calculated from contiguous three-day time windows. This sample was produced from WoMBATS data in a three-step method. First, an average BATS was calculated for all 514 wildfires. Specifically, on days when both data sources were available, half the sum of the IMSR and InciWeb estimates was used as the size estimate. When only one source was available, the available data were used instead. Second, a sample of days within each BATS that are suitable for validation were identified. A day of a BATS was deemed suitable if the previous three days monotonically and strictly increased (Equation (8)).

$$X_t > X_{t-1} > X_{t-2}; \quad (8)$$

Using these final validation data, the mean absolute percent error (MAPE) between the observations and size predictions produced from (Equations (6) and (7)) were calculated. An average of the MAPE estimates were calculated for days two through nine for 10 states in the Western United States: California, Arizona, Nevada, New Mexico, Colorado, Oregon, Idaho, Montana, Utah, and Washington. Sample sizes greater than 1 were not consistently available after the ninth day. For each state and day, a two-sample t-tests was performed to determine if MAPE estimates are significantly different in forecasts produced using areal persistence versus forecasts produced using radial persistence. Significance was defined using an $\alpha = 0.05$ p -value threshold.

3. Results

3.1. External Comparison

WoMBATS fire counts in a number of states (Table 4) were anomalous compared to what would be expected from MTBS data. WoMBATS fire counts were below the binomial distribution’s central 95th percentile (under-represented) in Alaska, Florida, Idaho, Kansas, Kentucky, Minnesota, Oklahoma, South Dakota, and West Virginia. WoMBATS fire counts were above the binomial distribution’s central 95th percentile (over-represented) in Arizona, California, Colorado, Nevada, and Washington. Only Montana, New Mexico, and Oregon had WoMBATS fire counts inside the top half of the binomial distribution’s central 95th percentile. In addition to disproportionately representing wildfire occurrence in some states, a number of WoMBATS-derived fire counts for individual states were anomalously low compared to what would be expected over such a time frame. WoMBATS fire counts

were below the empirical distribution’s central 95th percentile (fewer fires than would be reported from MTBS) in Florida, Idaho, Louisiana, Minnesota, Nebraska, and South Dakota (Figure 3). In no case were the WoMBATS fire counts above the empirical distribution’s central 95th percentile. Arizona and Colorado were the only states where the WoMBATS fire counts fell inside the top half of empirical distribution’s central 95th percentile.

Table 4. Number of wildfires from two WoMBATS and MTBS, dis-aggregated by state and year. The number of large (>405 ha) wildfires are reported in the parenthesis. States divided into Western Continental United States (CONUS), Eastern CONUS, and Outside CONUS. The asterisk * denotes that two long-duration wildfires that ignited in 2017 are included in Oregon 2018 WoMBATS fire counts.

Region	State	WoMBATS				MTBS
		2018	2019	2020	2018–2020	1984–2016
Western CONUS	California	27 (27)	16 (16)	41 (41)	84 (84)	1431
	Arizona	18 (18)	21 (20)	42 (42)	81 (80)	618
	Nevada	24 (24)	9 (9)	19 (19)	52 (52)	819
	New Mexico	11 (11)	18 (17)	10 (10)	39 (38)	649
	Colorado	21 (21)	4 (4)	9 (9)	34 (34)	245
	Oregon *	19(19)	5 (5)	18 (18)	42 (42)	640
	Idaho	18 (18)	4 (4)	12 (12)	34 (34)	1243
	Montana	14 (14)	6 (6)	14 (14)	34 (34)	617
	Utah	13 (13)	6 (6)	4 (4)	23 (23)	554
	Washington	14 (13)	3 (3)	14 (14)	31 (30)	418
	Wyoming	4 (4)	2 (2)	5 (4)	11 (10)	309
Eastern CONUS	Texas	5 (5)	2 (2)	26 (26)	33 (33)	802
	Florida	3 (3)	0 (0)	0 (0)	3 (3)	456
	South Carolina	0 (0)	1 (1)	0 (0)	1 (1)	30
	West Virginia	0 (0)	1 (1)	0 (0)	1 (1)	185
	Oklahoma	0 (0)	0 (0)	0 (0)	0 (0)	390
	South Dakota	0 (0)	0 (0)	2 (2)	2 (2)	177
	Minnesota	0 (0)	0 (0)	0 (0)	0 (0)	170
	Kansas	0 (0)	0 (0)	0 (0)	0 (0)	109
	Kentucky	0 (0)	0 (0)	0 (0)	0 (0)	101
	Louisiana	0 (0)	0 (0)	0 (0)	0 (0)	83
	North Carolina	0 (0)	0 (0)	0 (0)	0 (0)	74
	Nebraska	0 (0)	0 (0)	0 (0)	0 (0)	73
	Georgia	0 (0)	0 (0)	0 (0)	0 (0)	61
	Tennessee	0 (0)	0 (0)	0 (0)	0 (0)	54
	North Dakota	0 (0)	0 (0)	0 (0)	0 (0)	46
	Virginia	0 (0)	0 (0)	0 (0)	0 (0)	45
	Mississippi	0 (0)	0 (0)	0 (0)	0 (0)	44
	Arkansas	0 (0)	0 (0)	0 (0)	0 (0)	43
	Alabama	0 (0)	0 (0)	0 (0)	0 (0)	41
	Missouri	0 (0)	0 (0)	0 (0)	0 (0)	41
	Michigan	0 (0)	0 (0)	0 (0)	0 (0)	22
	New Jersey	0 (0)	0 (0)	0 (0)	0 (0)	18
	Maryland	0 (0)	0 (0)	0 (0)	0 (0)	9
	New York	0 (0)	0 (0)	0 (0)	0 (0)	6
	Pennsylvania	0 (0)	0 (0)	0 (0)	0 (0)	4
	Wisconsin	0 (0)	0 (0)	0 (0)	0 (0)	3
	Indiana	0 (0)	0 (0)	0 (0)	0 (0)	2
Ohio	0 (0)	0 (0)	0 (0)	0 (0)	2	
Delaware	0 (0)	0 (0)	0 (0)	0 (0)	1	
Iowa	0 (0)	0 (0)	0 (0)	0 (0)	1	

Table 4. Cont.

Region	State	WoMBATS				MTBS
		2018	2019	2020	2018–2020	1984–2016
Outside CONUS	Maine	0 (0)	0 (0)	0 (0)	0 (0)	1
	Alaska	0 (0)	7 (7)	1 (1)	8 (8)	984
	Hawaii	1 (1)	0 (0)	0 (0)	1 (1)	17
Total		192 (191)	105 (103)	217 (216)	514 (510)	11,757

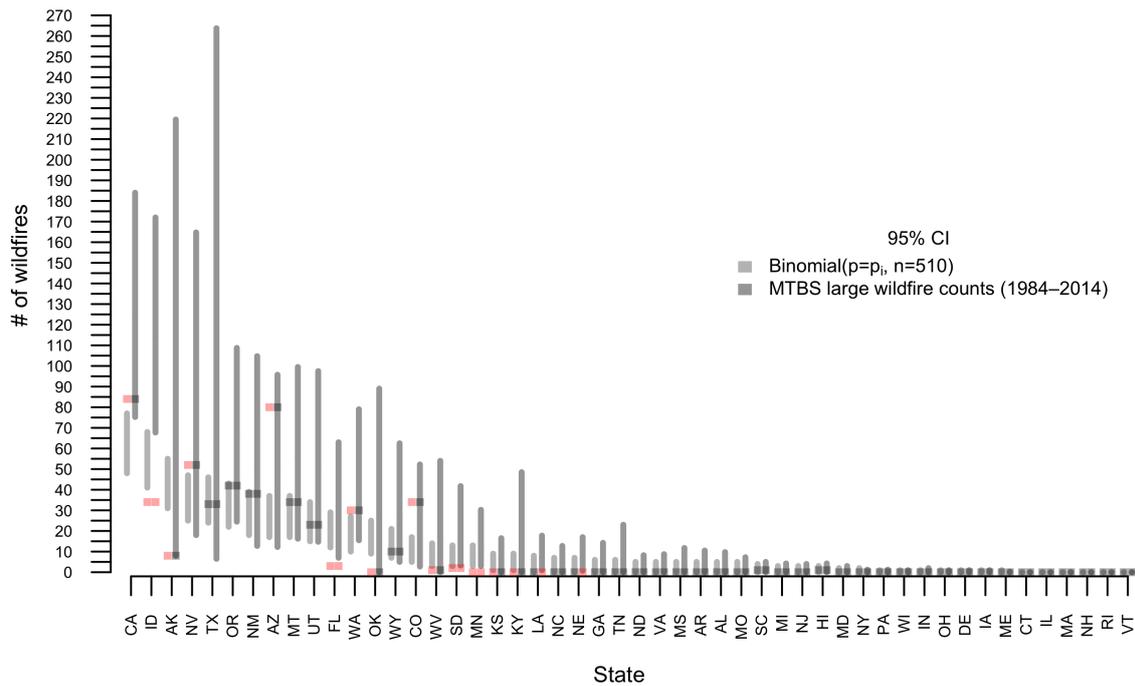


Figure 3. The number of large (>405 ha) fires for each state as reported from WoMBATS data between 2018 and 2020 compared to historical expectations as derived from Monitoring Trends in Burn Severity (MTBS) data. Each square represents the number of large wildfires in the WoMBATS data for each state. Vertical bars are used to represent the central 95th percentile of the expected wildfire counts from the binomial distribution (expected wildfire counts conditional on national counts) and empirical distribution (expected wildfire counts for that state and time-window). Squares that fall outside the central 95th percentile of expected fire counts are colored red.

In addition to revealing spatial biases, the parametric and non-parametric models also identified several temporal biases. Of the ten states considered, monthly wildfire counts fell outside the binomial distribution’s central 95th percentile in California, Arizona, New Mexico, Colorado, Oregon, and Idaho. In most states, these anomalies were restricted to only one or two months. However, in California, monthly intra-annual patterns of WoMBATS fire count proportions showed multiple differences from MTBS-derived proportions. Specifically, May–June wildfires were under-represented in the WoMBATS data and August–September wildfires were over-represented. In eight instances were WoMBATS monthly fire counts outside the empirical distribution’s central 95th percentile, and the WoMBATS fire counts did not show any obvious tendency to preferentially overestimate or underestimate (Figure 4).

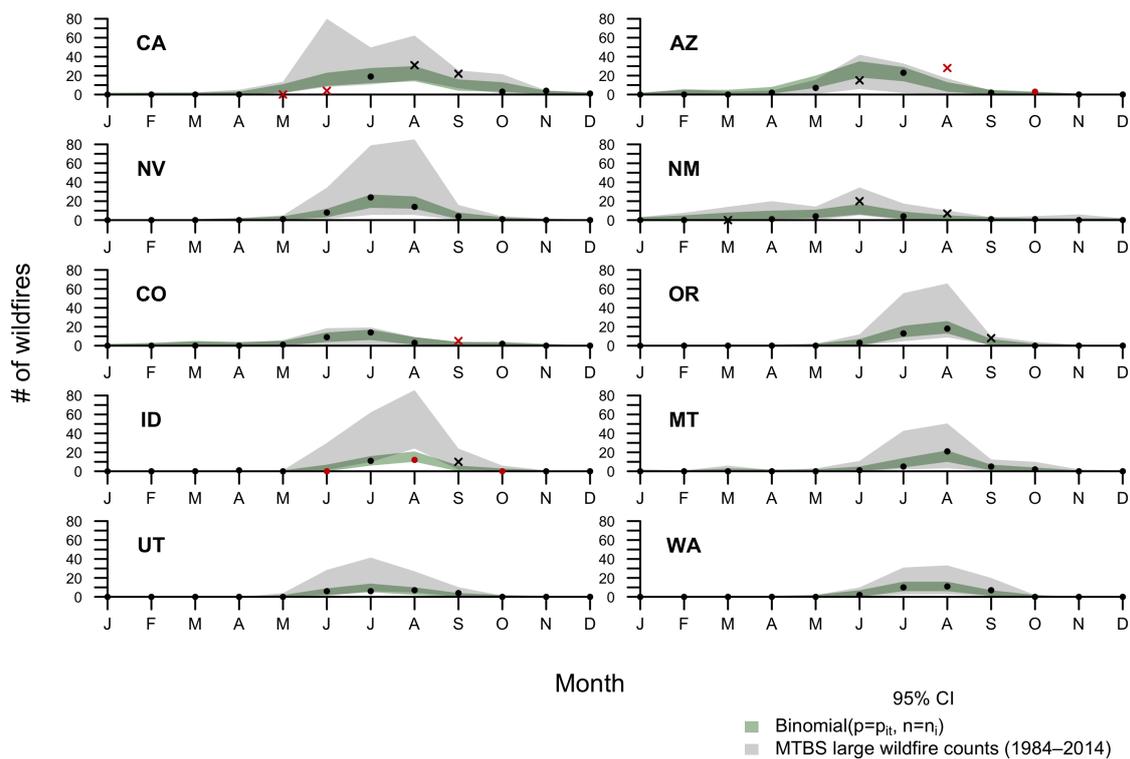


Figure 4. The number of large (>405 ha) wildfires for each state and month as reported from WoMBATS data between 2018 and 2020 compared to historical expectations as derived from MTBS fire occurrence data. Each point represents the number of large wildfires in the WoMBATS data for each state and month. The envelope represents the central 95th percentile of the expected wildfire counts from the binomial distribution (expected monthly wildfire counts conditional on total state counts) and empirical distribution (expected wildfire counts for that state, month, and observation-window based on observations). Points that fall outside the binomial distribution’s central 95th percentile of identified with an “X”. Points that fall outside the empirical distribution’s central 95th percentile are colored red.

Differences in the distributions of wildfire size derived from WoMBATS data and from MTBS data were statistically significant ($\alpha < 0.05$) in five of the 10 states considered. In the cases where a difference could be identified, WoMBATS data tended to include more large fires than what would be expected from MTBS data. The largest differences were observed in Oregon, where the median fire size was nearly four times as large when calculated from WoMBATS data than when calculated from MTBS data (Figure 5).

3.2. Internal Comparison

InciWeb-derived BATs were typically more complete than IMSR-derived BATs. The average InciWeb BATs reported size estimates data on 83 percent of the days in the wildfire’s lifetime. In contrast, the average IMSR BATs only reported size estimates on 51 percent of the days in the wildfire’s lifetime, and a large majority of these days (82 percent) were also reported by InciWeb. There were no missing days—a size estimate was available from at least one of either InciWeb or IMSR data for every day of the wildfire—in 156 of the 514 wildfires in the case list (Figure 6).

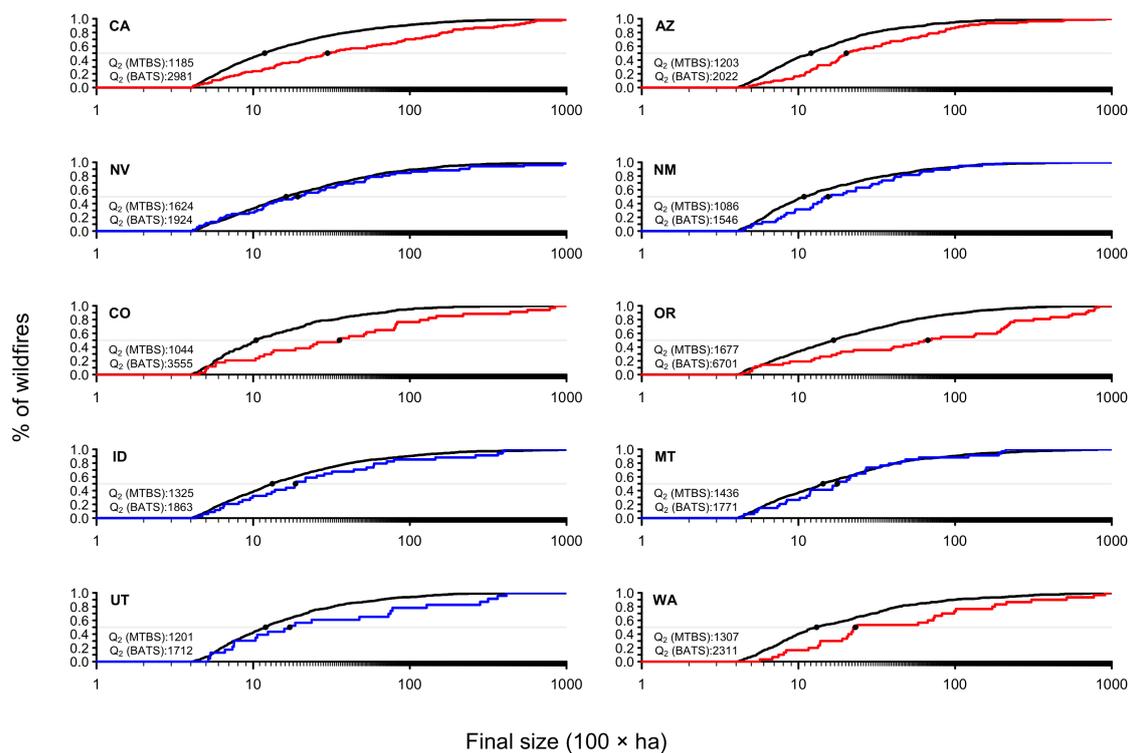


Figure 5. Cumulative distribution functions of wildfire size derived from MTBS and from WoMBATS. The former is colored black, and the latter is colored blue or red depending on its statistical distance from the MTBS data. Specifically, if no detectable difference can be found in the two size distributions, as quantified via a Kolmogorov–Smirnov using a critical p -value of $\alpha = 0.05$, the cumulative distribution function is colored blue. Otherwise it is colored red. Median values from both data sources are listed in the bottom left and are identified with black points.

Size measurements from Inciweb and IMSRs were often close. The average BATS had exactly matching data on 73 percent of days when both data sources available (Figure 6B), and about 21 percent of the wildfires had BATS with InciWeb data in perfect agreement with the IMSR data (in the 494 BATS where both data were shared on at least one day) (Figure 6C). Even when the two data sources disagreed, the difference tended to be small. For wildfires where both data sources are available, the average standard deviation of size measurements was 2.88 percent of the mean, and only about 6 percent of fires had average coefficients of variation in excess of 10 percent. In many cases, InciWeb and IMSRs were in perfect agreement, so the low average coefficients of variation are expected. However, even if wildfires in which no disagreement between data sources was detected were omitted, the average coefficient of variation would remain modest. In wildfires with nonzero coefficients of variation, the average standard deviation of size measurements was 3.63 percent of the mean, and only about 8 percent of fires had average coefficients of variation in excess of 10 percent (Figure 6C). Data coherence also appears high when considering the distribution of data quality classes [42]. Of the 494 relevant wildfires in the case list, 21 percent had ‘Exact’ agreement, 61 percent had ‘Close’ agreement, 18 percent had ‘Adequate’ agreement, and none were at levels considered to be “Inadequate” [42]. When and where differences occurred, IMSRs were the larger measurement about 1.7 times more often than the reverse case.

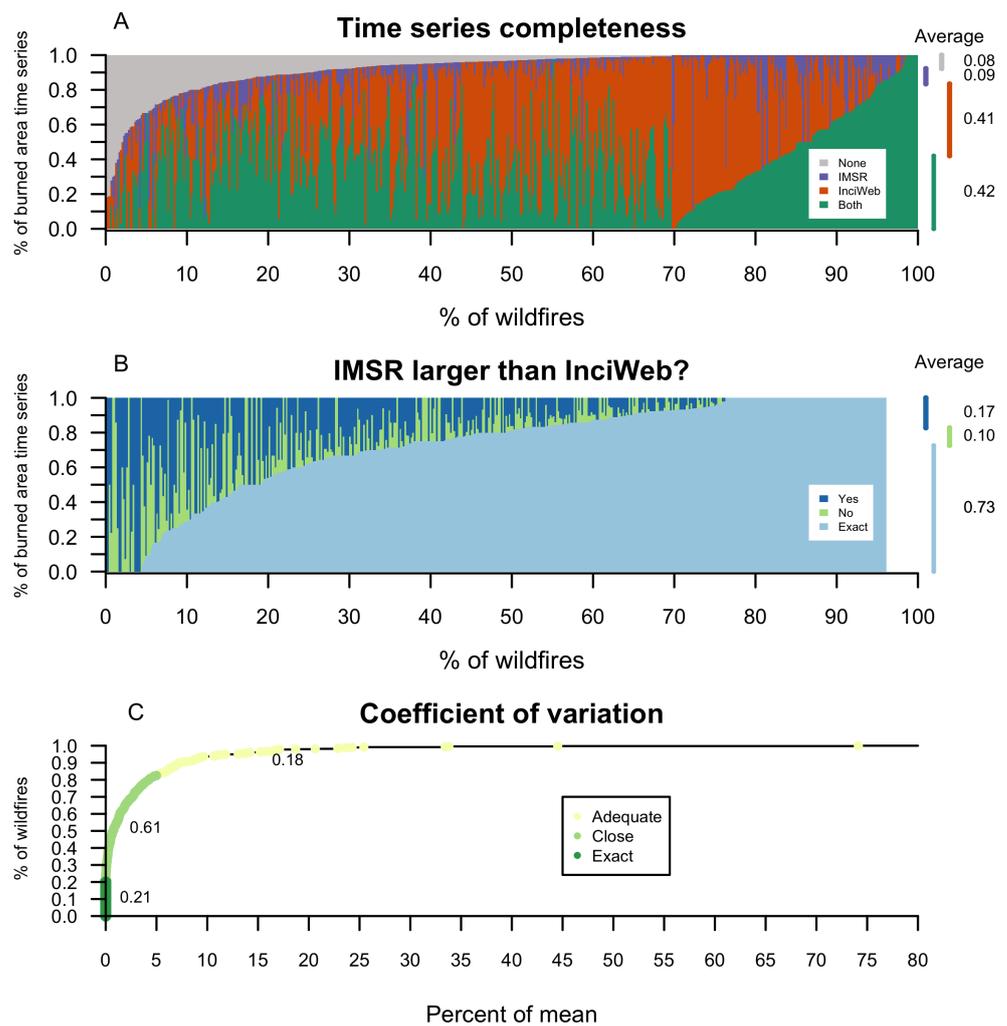


Figure 6. Graphical summaries of burned area time series (BATS) data availability and coherence in the WoMBATS data. Panel (A) describes the data completeness. Each vertical bar represents the proportion of each BATS where one, both, or none of the data sources are available. Panel (B) characterizes the level of agreement between InciWeb and IMSRs when both are available in terms of biases. Each vertical bar represents the proportion of the BATS measurements with cross-check data where IMSR measurements are larger, InciWeb measurements are larger, or are the same. BATS with no cross-check data are represented with white space. In Panels (A,B), wildfires are ordered by BATS completeness and the level of agreement, respectively. Panel (C) shows the empirical cumulative distribution of the coefficient of variation calculated from the 494 fires with cross-check data. Colored points are used to represent which data quality category [42] applies to each wildfire’s BATS.

3.3. Application

The data cleaning processes produced a final validation dataset of 3758 3-day partial BATS from 442 fires. Using these data, we found that the differences in performance between the areal persistence model and the radial persistence model were generally small. In approximately 57 percent of the sample, the MAPE was lower in the areal persistence model than in the radial persistence model. The median absolute percent error (across the entire sample of 3758 time windows) was 3.91 percent for the areal persistence model versus 4.27 in the radial persistence model. The slight preference for the areal persistence model was also seen when the sample was disaggregated by state and day of fire. Of the 10 states examined and eight fire days considered, the MAPE was lower in the radial persistence model than in the areal persistence model in only nine cases, and in no case was the radial persistence model’s MAPE significantly lower than the areal persistence model’s

(Figure 7). Both models appear to produce more accurate predictions as the fire progresses, with the MAPE generally being highest soon after ignition and then steadily decreasing. The performance in New Mexico is somewhat of an exception to this trend, but the sample sizes on days 2 and 3 are too small to accurately estimate the expected MAPE (Figure 7).

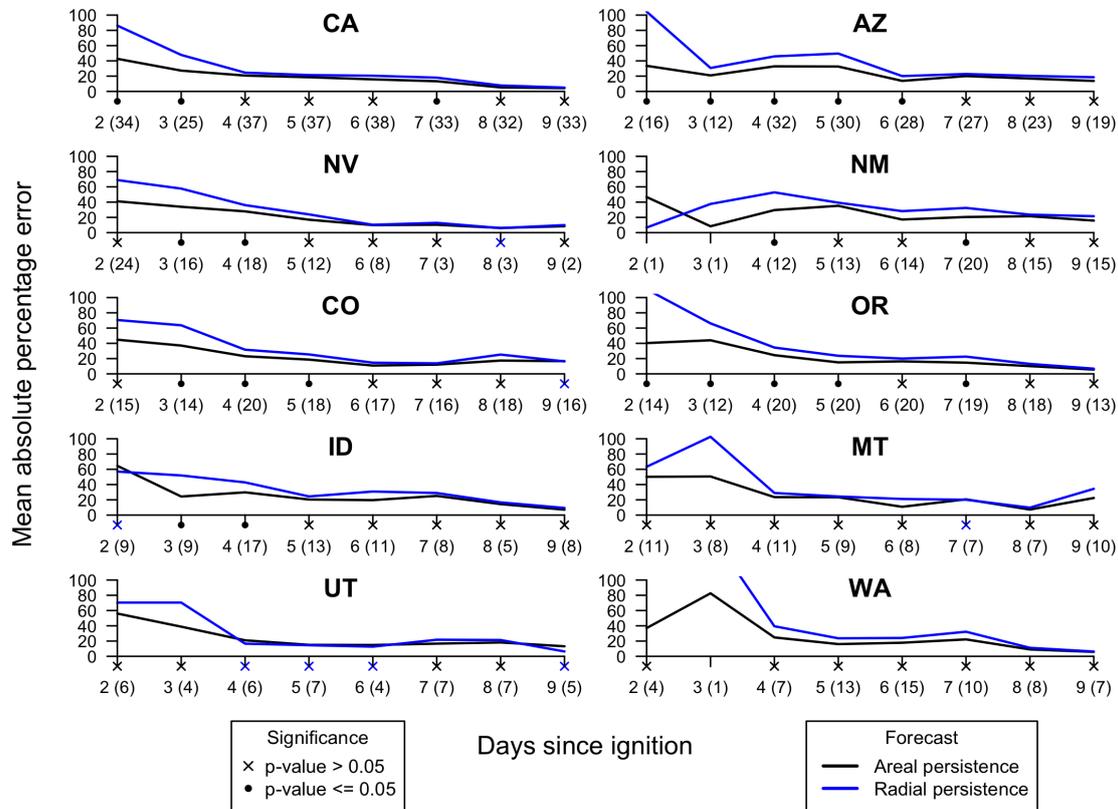


Figure 7. The mean absolute percent error (MAPE) of the areal persistence model and radial persistence model as estimated using all available samples from the WoMBATS data. MAPE estimates are stratified by the state and the days since ignition. Below the x-axis, there are points that are black except for days in which the radial persistence model performed better than the areal persistence model, which are instead colored blue. The points are hollow unless the difference in MAPE estimates in the areal persistence model and the radial persistence model are significant, in which case they are solid. Sample sizes are reported in the x-axis in parenthesis.

4. Discussion

4.1. Advantages and Limitations

Although it is sometimes possible to produce BATS using a single data source [28], the use of multiple data sources has a number of apparent advantages. BATS produced with cross-check data are usually more complete than any of the constituent data sources. InciWeb webscrape data were particularly effective at filling in missing days, and an average of 41 percent of the time series would be missing if InciWeb data were not consulted (Figure 6A). However, it should be noted here that there are other factors to consider besides the number of days that a data source fills in or corroborates. During the end days of the fire, there is little-to-no growth, and the scientific value associated with filling in these days is substantially less than filling in days when there is large amounts of fire growth. As seen in the Saddleridge example, other data sources can be particularly useful for this task, even if they fill in relatively fewer days than InciWeb or IMSR data. Although the LAFD narrative reported size estimates for fewer days of the Saddleridge fire than InciWeb or IMSR data, it was still a particularly important data source because the estimates it reported coincided with the most active days of the fire (Table 3). In addition to filling in missing

data, the use of multiple data sources had the added benefit of allowing us to gauge the reliability of the size measurements. Although exact agreement between the IMSR and InciWeb BATS were uncommon, differences were usually fairly small, suggesting that the reported size data from IMSRs and InciWeb were probably reasonably accurate and robust to data source substitution. These small differences could not have been identified without cross-checking, as many would not have otherwise seemed suspicious. This information about the measurement reliability is not only useful to researchers that want a qualitative sense of the credibility of the BATS data they are using, but also useful for accurately calibrating existing BATS data cleaning methods that assume noisy size estimates [32].

Clearly, consulting multiple data sources has the potential to increase data completeness, accuracy, and credibility. However, the catalog of wildfires produced from the protocol is also quite small relative to existing final burned area data (Table 4) and even other BATS data sources [28]. From our statistical analyses (Figure 3), it is clear that the InciWeb website only reports some of the wildfires that are larger than 405 ha, and that the case list is not exhaustive. If these fire detection probabilities were the spatially uniform, then the use of WoMBATS data for describing the statistical characteristics of fire across the United States may not be excessively problematic. However, there are some states that are more likely to report wildfires to InciWeb than others. In Arizona, we can see that the number of fires reported in WoMBATS data is generally consistent with what would be expected from a sample of MTBS data over the same time window, and that the percentage of all WoMBATS fires occurring here is noticeably higher than what would be expected based on MTBS data. We can also see that in Idaho and Florida, the number of fires is well below what would be reported from MTBS over the same time window, and that the percentage of all WoMBATS fires occurring here is noticeably below what would be expected based on MTBS data. We can infer from these results that the probability that a fire is reported to InciWeb is probably relatively high in states like Arizona compared to states like Idaho and Florida, and that InciWeb fire detection probabilities are not spatially uniform (Figure 3). Monthly wildfire counts could be anomalously low in some contexts too, suggesting that WoMBATS data may sometimes fail to recreate actual seasonal wildfire trends (Figure 4) and that the distribution of wildfire size derived from WoMBATS data tends to be upwardly biased relative to the MTBS-derived distribution (Figure 5). These statistical biases are consistent with the nature of InciWeb data, which are primarily intended to provide relevant information to the public, and fires that are small or do not pose a threat to human populations may not be reported by InciWeb. Analysts should then be cautious when interpreting results derived from WoMBATS data and consider whether the intended application requires correcting for these biases. If these biases are problematic, as might be the case in a national analysis, researchers may consider instead using ICS-209-PLUS data, which may have less accurate day-to-day estimates of fire size, but will have a more complete case list [28]. It is worth mentioning that just because the reported size of a fire is similar between two data sources does not necessarily imply that the reported size is correct. As described in the introduction, there are multiple reasons why errors can arise in fire size estimates, and because there is a fair amount of information sharing, it is possible that two sources can report the same, but incorrect, estimates. Still, this epistemological uncertainty is likely impossible to correct in most contexts, and the protocol that was presented here incorporates more safeguards against this problem than any protocol based on a single data source. Consumers of WoMBATS data should be aware of these measurement errors because the protocol is particularly inclusive about the collection of supplemental data sources and allows consideration of lower-quality size estimates. The data collection protocol permits the inclusion of these low-quality supplemental data sources because it can sometimes provide useful information despite being less reliable [40]. For that reason, analysts should make use of the data quality flags within the metadata table and deliberate whether these supplemental data should be used for the given application.

4.2. Future Research

Given that BATS data can be used to answer many questions that final burned area data cannot, and that BATS are relatively rare, there are clearly a wide range of analyses that could be performed using WoMBATS data. Beyond adopting WoMBATS data in wildfire research to assess questions about day-to-day variation in fire size, we recommend that future work prioritize two aspects of BATS data: cleaning and collection. As mentioned earlier, there are multiple methods of converting a collection of partial BATS into a final “clean” BATS. Since the choice of data cleaning method will influence the final product and is often context-specific [39], it was deliberately avoided in the data collection protocol. Still, data cleaning is an important research area that should be explored further since it is a necessary step for applying BATS data to scientific problems. The InciWeb data used in this paper was collected in a semi-automated fashion, and so long as InciWeb data can be collected to produce case lists and IMSRs are available, the BATS data can be updated at regular intervals like already existing wildfire datasets such as the MTBS [26]. Data collection is also ongoing in terms of the consultation of supplemental data sources. Since exhaustive consultation of all available supplemental sources is impractical, future data collection efforts should search for supplemental data in a systematic and effective manner [39]. For instance, the approximately 8 percent of days with missing data and 50 percent of days without cross-check data could be prioritized, or individual wildfire’s that are missing a large amount of data could be prioritized. Moreover, satellite data are collected regularly on a global scale, and future collection efforts might then focus on including this source as a third cross-check data source. Satellite data have already been shown to be a helpful data source when data are missing and erroneous [32]. Data from the GlobFire project may be a potentially useful cross-check source, as it will have already been clustered into individual wildfires [34], avoiding much of the preprocessing and cleaning that would otherwise be required. Spatially explicit maps will clearly provide more information than simple time series, and the development of an analogous warehouse of burned area progression maps data collection protocol would be an ambitious but immensely valuable future area of research.

5. Conclusions

BATS are a notoriously uncommon but critically important data for developing a holistic understanding of the relationship between wildfire behaviors and the environment. Even when available, BATS are vulnerable to a number of data quality issues, many of which can only be corrected by consulting multiple data sources. In this paper, we described a protocol for collecting BATS cross-check data from Incident Situation Management Reports, InciWeb webscrapes, and other data sources, which we use to create the warehouse of multiple burned area time series (WoMBATS) dataset. In our analysis of WoMBATS data, we found that this data collection protocol can greatly improve data completeness and identify errors that would not have been detected without cross-checking data. However, a comparison of WoMBATS data to long-term quality-controlled fire occurrence data from the Monitoring Trends in Burn Severity project identified noticeable spatial and temporal biases in the WoMBATS-derived distributions of fire occurrence, as well as biases in the WoMBATS-derived distributions of fire size. We demonstrated how WoMBATS data could be used in research applications using a model validation analysis of simple fire growth models as an example. Given the relative rarity of these data, we anticipate that WoMBATS data can be immensely valuable to wildfire researchers, but also highlight some potential limitations. To facilitate the continued use of WoMBATS data, we make it freely available and plan to continue to collect data in accordance with the protocol we have described.

Author Contributions: Conceptualization, H.R.P., N.K.L. and B.P.; methodology, H.R.P. and N.K.L.; software, H.R.P.; validation, H.R.P. and B.P.; formal analysis, H.R.P.; investigation, H.R.P.; resources, N.K.L.; data curation, H.R.P.; writing—original draft preparation, H.R.P.; writing—review and editing, H.R.P., N.K.L. and B.P.; visualization, H.R.P.; supervision, N.K.L.; project administration, H.R.P. and N.K.L.; funding acquisition, N.K.L. and H.R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the United States Forest Service, agreement number 19-JV-11261987-139.

Data Availability Statement: The data produced from this procedure are available at: https://figshare.com/articles/dataset/WOMBATS_basic_data_2018-2020/14788206; DOI: 10.6084/m9.figshare.14788206.

Acknowledgments: The authors would like to acknowledge Ernesto Alvarado, Elizabeth Ashley Steel, Susan O’Neill, and Colton Miller, and the AirFire team for their input on drafts and for their helpful discussions throughout the production of this manuscript.

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

References

- Birch, D.S.; Morgan, P.; Kolden, C.A.; Hudak, A.T.; Smith, A.M. Is proportion burned severely related to daily area burned? *Environ. Res. Lett.* **2014**, *9*, 064011. [CrossRef]
- Hantson, S.; Andela, N.; Goulden, M.L.; Randerson, J.T. Human-ignited fires result in more extreme fire behavior and ecosystem impacts. *Nat. Commun.* **2022**, *13*, 1–8. [CrossRef] [PubMed]
- Johansen, M.P.; Hakonson, T.E.; Breshears, D.D. Post-fire runoff and erosion from rainfall simulation: Contrasting forests with shrublands and grasslands. *Hydrol. Process.* **2001**, *15*, 2953–2965. [CrossRef]
- Wilder, B.A.; Lancaster, J.T.; Cafferata, P.H.; Coe, D.B.; Swanson, B.J.; Lindsay, D.N.; Short, W.R.; Kinoshita, A.M. An analytical solution for rapidly predicting post-fire peak streamflow for small watersheds in southern California. *Hydrol. Process.* **2021**, *35*, e13976. [CrossRef]
- Stefanidis, S.; Alexandridis, V.; Spalevic, V.; Mincato, R.L. Wildfire effects on soil erosion dynamics: The case of 2021 megafires in Greece. *Agric. For. I Sumar.* **2022**, *68*, 49–63.
- Stephens, S.L.; Burrows, N.; Buyantuyev, A.; Gray, R.W.; Keane, R.E.; Kubian, R.; Liu, S.; Seijo, F.; Shu, L.; Tolhurst, K.G.; et al. Temperate and boreal forest mega-fires: Characteristics and challenges. *Front. Ecol. Environ.* **2014**, *12*, 115–122. [CrossRef]
- Page, W.G.; Freeborn, P.H.; Butler, B.W.; Jolly, W.M. A review of US wildland firefighter entrapments: Trends, important environmental factors and research needs. *Int. J. Wildland Fire* **2019**, *28*, 551–569. [CrossRef]
- Moeltner, K.; Kim, M.K.; Zhu, E.; Yang, W. Wildfire smoke and health impacts: A closer look at fire attributes and their marginal effects. *J. Environ. Econ. Manag.* **2013**, *66*, 476–496. [CrossRef]
- Noble, J.C. Behaviour of a very fast grassland wildfire on the Riverine Plain of southeastern Australia. *Int. J. Wildland Fire* **1991**, *1*, 189–196. [CrossRef]
- O’hara, K.C.; Ranches, J.; Roche, L.M.; Schohr, T.K.; Busch, R.C.; Maier, G.U. Impacts from Wildfires on Livestock Health and Production: Producer Perspectives. *Animals* **2021**, *11*, 3230. [CrossRef]
- Finney, M.; Grenfell, I.C.; McHugh, C.W. Modeling containment of large wildfires using generalized linear mixed-model analysis. *For. Sci.* **2009**, *55*, 249–255.
- Doerr, S.H.; Santín, C. Global trends in wildfire and its impacts: Perceptions versus realities in a changing world. *Philos. Trans. R. Soc. B: Biol. Sci.* **2016**, *371*, 20150345. [CrossRef] [PubMed]
- Cliff, E.; Price, J.; Lindh, C.; Mays, L.; Cochran, H. *The Rattlesnake Fire*; USDA Forest Service: Washington, DC, USA, 1953. Available online: <http://wlfalwaysremember.org/images/incidents/documents/1953-07-09-rattlesnake-report.pdf> (accessed on 4 April 2019).
- Trelles, J.; Pagni, P.J. Fire-induced winds in the 20 October 1991 Oakland Hills fire. *Fire Saf. Sci.* **1997**, *5*, 911–922. [CrossRef]
- Viegas, D.; Simeoni, A. Eruptive behaviour of forest fires. *Fire Technol.* **2011**, *47*, 303–320. [CrossRef]
- Keller, E.A.; Valentine, D.W.; Gibbs, D.R. Hydrological response of small watersheds following the Southern California Painted Cave Fire of June 1990. *Hydrol. Process.* **1997**, *11*, 401–414. [CrossRef]
- Brown, D.K.; Echelle, A.A.; Propst, D.L.; Brooks, J.E.; Fisher, W.L. Catastrophic wildfire and number of populations as factors influencing risk of extinction for Gila trout (*Oncorhynchus gilae*). *West. N. Am. Nat.* **2001**, *61*, 139–148.
- Recher, H.F. Impact of wildfire on the avifauna of Kings Park, Perth, Western Australia. *Wildl. Res.* **1997**, *24*, 745–761. [CrossRef]
- Finney, M.A. Modeling the spread and behavior of prescribed natural fires. In Proceedings of the International Conference on Fire and Forest Meteorology, Jekyll Island, GA, USA, 26–28 October 1993; Volume 13, pp. 8–143.
- Billmire, M.; French, N.H.; Loboda, T.; Owen, R.C.; Tyner, M. Santa Ana winds and predictors of wildfire progression in southern California. *Int. J. Wildland Fire* **2014**, *23*, 1119–1129. [CrossRef]
- Potter, B.E.; McEvoy, D. Weather Factors Associated with Extremely Large Fires and Fire Growth Days. *Earth Interact.* **2021**, *25*, 160–176. [CrossRef]

22. Potter, B.E.; McEvoy, D. *Fire Growth and Associated Weather Data for Selected Fires of Unusual Size (FOUS) and Other Fires from 2004 to 2018*; Forest Service Research Data Archive: Fort Collins, CO, USA, 2022. [CrossRef]
23. Abell, C.A. *Rates of Initial Spread of Free-Burning Fires on the National Forests of California*; Research Note PSW-RN-24; US Department of Agriculture, Forest Service, California Forest and Range Experiment Station: Berkeley, CA, USA, 1940; 27p.
24. Anderson, H.E. *Predicting Wind-Driven Wind Land Fire Size and Shape*; US Department of Agriculture, Forest Service, Intermountain Forest and Range: Ogden, UT, USA, 1983; Volume 305.
25. Taylor, S.W.; Woolford, D.G.; Dean, C.; Martell, D.L. Wildfire prediction to inform fire management: Statistical science challenges. *Stat. Sci.* **2013**, *28*, 586–615. [CrossRef]
26. Eidenshink, J.; Schwind, B.; Brewer, K.; Zhu, Z.L.; Quayle, B.; Howard, S. A project for monitoring trends in burn severity. *Fire Ecol.* **2007**, *3*, 3–21. [CrossRef]
27. Alexander, M.; Thomas, D. Wildland fire behavior case studies and analyses: Other examples, methods, reporting standards, and some practical advice. *Fire Manag. Today* **2003**, *63*, 4–12.
28. St Denis, L.A.; Mietkiewicz, N.P.; Short, K.C.; Buckland, M.; Balch, J.K. All-hazards dataset mined from the US National Incident Management System 1999–2014. *Sci. Data* **2020**, *7*, 1–18. [CrossRef] [PubMed]
29. National Interagency Fire Center. ICS-209 Program (NIMS) User’s Guide. Online. 2020. Available online: https://www.predictiveservices.nifc.gov/intelligence/ICS-209_User_Guide_4.0_2020.pdf (accessed on 1 April 2022).
30. Reading the Situation Report. Online. 2018. Available online: https://www.predictiveservices.nifc.gov/intelligence/Reading_the_Situation_Report_2018.pdf (accessed on 1 April 2022).
31. Veraverbeke, S.; Sedano, F.; Hook, S.J.; Randerson, J.T.; Jin, Y.; Rogers, B.M. Mapping the daily progression of large wildland fires using MODIS active fire data. *Int. J. Wildland Fire* **2014**, *23*, 655–667. [CrossRef]
32. Podschwit, H.; Guttorp, P.; Larkin, N.; Steel, E.A. Estimating wildfire growth from noisy and incomplete incident data using a state space model. *Environ. Ecol. Stat.* **2018**, *25*, 325–340. [CrossRef]
33. Gemitzi, A.; Koutsias, N. A Google Earth Engine code to estimate properties of vegetation phenology in fire affected areas—A case study in North Evia wildfire event on August 2021. *Remote Sens. Appl. Soc. Environ.* **2022**, *26*, 100720. [CrossRef]
34. Artés, T.; Oom, D.; De Rigo, D.; Durrant, T.H.; Maianti, P.; Libertà, G.; San-Miguel-Ayanz, J. A global wildfire dataset for the analysis of fire regimes and fire behaviour. *Sci. Data* **2019**, *6*, 1–11. [CrossRef]
35. Kolden, C.A.; Weisberg, P.J. Assessing accuracy of manually-mapped wildfire perimeters in topographically dissected areas. *Fire Ecol.* **2007**, *3*, 22–31. [CrossRef]
36. Donovan, G.H.; Prestemon, J.P.; Gebert, K. The effect of newspaper coverage and political pressure on wildfire suppression costs. *Soc. Nat. Resour.* **2011**, *24*, 785–798. [CrossRef]
37. Callister, K.E.; Griffioen, P.A.; Avitabile, S.C.; Haslem, A.; Kelly, L.T.; Kenny, S.A.; Nimmo, D.G.; Farnsworth, L.M.; Taylor, R.S.; Watson, S.J.; et al. Historical maps from modern images: Using remote sensing to model and map century-long vegetation change in a fire-prone region. *PLoS ONE* **2016**, *11*, e0150808. [CrossRef]
38. De Longueville, B.; Smith, R.S.; Luraschi, G. “OMG, from here, I can see the flames!” a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In Proceedings of the 2009 International Workshop on Location Based Social Networks, Seattle, WA, USA, 3 November 2009; pp. 73–80.
39. Van den Broeck, J.; Argeseanu Cunningham, S.; Eeckels, R.; Herbst, K. Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLoS Med.* **2005**, *2*, e267. [CrossRef] [PubMed]
40. Sana, M.; Weinreb, A.A. Insiders, outsiders, and the editing of inconsistent survey data. *Sociol. Methods Res.* **2008**, *36*, 515–541. [CrossRef]
41. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2013. Available online: <https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing> (accessed on 1 April 2022).
42. Cruz, M.G.; Alexander, M.E. Uncertainty associated with model predictions of surface and crown fire rates of spread. *Environ. Model. Softw.* **2013**, *47*, 16–28. [CrossRef]
43. Strand, T.M.; Larkin, N.; Craig, K.J.; Raffuse, S.; Sullivan, D.; Solomon, R.; Rorig, M.; Wheeler, N.; Pryden, D. Analyses of BlueSky Gateway PM2.5 predictions during the 2007 southern and 2008 northern California fires. *J. Geophys. Res. Atmos.* **2012**, *117*, 1–14. [CrossRef]