Reviewer 2

Williams et al. introduce WUMI2024a, a harmonized, quality-controlled database of 22,464 large wildfire events across the 11 western U.S. states from 1984 – 2024. The product merges seven public government datasets and provides, for each event, a start date, location, final size, a perimeter (observed where available, circular when missing), and a 1-km grid of fractional area burned. About 46% of events inherit 30 m burned-area maps and perimeters from MTBS; another 24% have perimeters from non-MTBS sources; only 30% lack observed perimeters. Given the importance and lack of good quality data for wildfire regime assessment, the work of Williams et al. is highly relevant and useful for the community.

We thank the reviewer for their thorough and constructive review of our paper. We address each comment below (in blue font). In addressing the reviews we discovered some minor errors in our code and also performed additional manual inspections of the fire lists and burned area maps that led to a revised dataset with approximately 1% fewer wildfires. This revision did not cause notable changes to the regional patterns of fire frequency or burned area. The paper and the WUMI2024a database archived online have been revised to reflect these changes.

Major comments:

1. The paper introduces an alternate event list that replaces MTBS "parent" fires with linked non-MTBS "sub-fires," and even reassigns sizes so sub-fires sum to the parent (Sec. 2.4). It is not clear what decision logic was used to separate single events vs. multiple events. How are overlaps across days/IDs handled? What are the implications for trend analysis of counts vs. burned areas when switching between parent and sub-fire lists?

This comment and set of questions motivated us to expand in section 2.4 (Summary of final dataset) on how sub-fires were treated. We describe the text addition below, but first we'd like to clarify to the reviewer that sub-fires were identified using the basic methods we developed to identify linkages between MTBS and non-MTBS fires, described in the 3rd paragraph in section 2.2.2 (Dataset merging). That is, a 'parent' fire is identified as any MTBS fire for which we found multiple linkages to non-MTBS fires. In these cases, the multiple non-MTBS fires linked to the same parent MTBS fire are called sub-fires. We have now added a sentence to section 2.2.2 that indicates to the reader that a description will be provided in section 2.4 regarding how these cases of multiple sub-fires being part of the same parent fire are handled. We also revised section 2.4 to better clarify:

First, at the beginning of section 2.4 immediately after indicating that a number of MTBS fires are composed of multiple non-MTBS fires, we have added a sentence that we feel will clarify to a reader what we are referring to:

"*For example, some fire complexes, when multiple fires that began independently but merged, are represented by MTBS as single events.*"

More generally, this reviewer comment motivated us to thoroughly revise section 2.4 to better describe the logic behind our treatment of sub-fires and the implications. After explaining that we adjust the ignition locations and fire sizes of sub-fires to be consistent with the burned-area maps of the parent fires, we have added (L274-283):

"*We realign the sizes and locations of sub-fires to be consistent with the MTBS parent because the MTBS maps of area burned are satellite-based and rigorously quality controlled. In contrast, the fire sizes and ignition locations reported in other datasets are often approximate, as exemplified by cases in which the same fire is reported by different agencies but with non-identical sizes and ignition locations. On the other hand, adjustment of fire size and ignition location introduces errors for cases in which one or more of the sub-fires was misidentified (did not actually contribute to the burned area mapped by MTBS), not all sub-fires were identified, or two or more sub-fires were actually the same event but not identified as such. That said, we find that, among the 119 cases in which a parent fire is composed of sub-fires, the sum of the areas burned by the sub-fires consistently agree well with the total area of the parent fire (Fig. S1), giving us confidence that any unintended errors caused by adjustments to sub-fire sizes and ignition locations are generally minor.*"

In the above text we now also reference a new Supplementary Figure S1, which we provide below, that shows that the sums of the sizes of sub-fires agree very well with the sizes of their associated parent fires, which serves as evidence that our method of matching parent fires with their associated sub-fires works well.

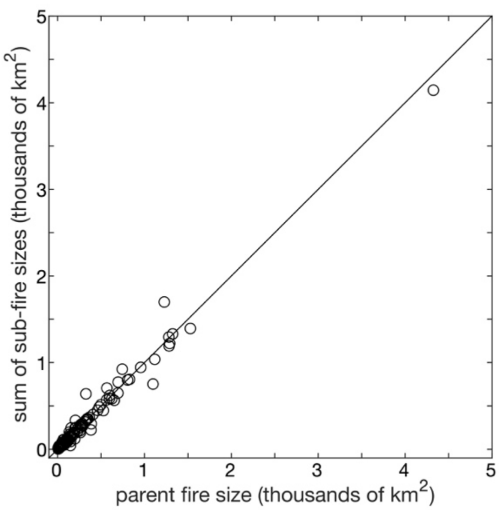

Figure s1. Scatter plot comparing the sums of sub-fire sizes to the sizes of their corresponding parent fires. Dots represent each of 119 cases in which a single MTBS parent fire was found to be composed of two or more non-MTBS sub-fires. The diagonal black line is a one-to-one line.

We then go on to more thoroughly describe cases in which consideration of sub-fires would be desirable over parent fires and vice versa, clarify that these instances of parent and sub-fires are relatively rare and thus will not majorly impact analyses at the large scale of the western US, and also assure the reader that the WUMI2024a also includes the pre-adjustment sizes and locations of sub-fires (L285-294):

"*Thus, the WUMI2024a includes two lists of wildfire events. The list that prioritizes sub-fires over parent fires may be preferable when the goal is to use the most accurate records of ignition dates and fire frequencies. However, perimeter data are not available for some sub-fire events, and MTBS is the only dataset that provides maps of burned versus unburned areas within the fire perimeter. Therefore, prioritization of parent fires over sub-fires may be preferable in* *applications requiring the most accurate maps area burned. Importantly, the 119 parent fires and their 413 sub-fires are rare, constituting <2% of all fires in the dataset, meaning that analyses of fire frequencies or areas burned at the large scale of the western US will not be strongly sensitive to whether parent fires or sub-fires are prioritized. Additionally, for each parent fire the WUMI2024a database includes a text file that provides the pre-adjustment ignition locations and sizes of sub-fires. Table S1 provides descriptions of the wildfire attributes included in the WUMI2024a lists of wildfires and sub-fires.*"

2. In terms of the dataset quality-control, did you quantify how sensitive the final counts and areas are to the chosen thresholds (e.g., L160 "we treat them as duplicates if they are within 25 km, 3 days, and 10% of the larger of the fire sizes")? How were these thresholds chosen?
We did perform tests in developing our methods to assure that the methods we use in the automated portions of the quality control and dataset merging processes are effective, but we also found it unavoidable that ultimately there was no one-size-fits-all set of automations that satisfactorily did the job without leaving the need for additional intensive non-automated work to identify cases of duplicate entries.

That is, each time we refined the methods by, for example, reducing the distance threshold used to identify same-named fires as duplicates from 200 km to 100 km (which we did in response to a comment from Reviewer #1), a renewed and very time consuming effort was needed to manually check the data to identify duplicate events that were not caught by the revised automation method. The extraordinary amount of time and work involved in the manual quality control prevents us from carrying out a formal sensitivity analysis of how the final dataset is affected by the specific threshold we chose. On the other hand, it's the extraordinary care that we put into the manual portion of the quality control process that gives us confidence that, ultimately, the final dataset is not highly sensitive to the threshold choices used. We have added a paragraph to the end of section 2.2.2 better highlights the importance of, and rationale behind, the intensive manual quality control work that went into developing this dataset (L224-232).

"*Importantly, we found it infeasible to automate a one-size-fits-all set of rules that effectively detects duplicate fires within or between datasets without need for intensive additional scrutiny. For example, the same fire is often assigned different names by different government agencies; the same name may be spelled or misspelled in a range of ways; and dates, locations, and sizes are not always reported consistently across agencies, sometimes with large differences. Therefore, at every step of the quality-control and dataset-merging process we conduct rigorous visual inspections of fire lists and perimeter maps to identify additional duplicate fires between or within datasets as well as fires that were misidentified as redundant by our automated routine. Due to the intensive nature of the visual inspection and manual correction process we are confident that the quality of the final WUMI2024a dataset, while certainly not without remaining errors, is not highly sensitive to the specific rules implemented in the automated portion of the process.*"

3. Section 2.2.2 would benefit from a table or flowchart to outline the merging logic.
We thank the reviewer for this suggestion and we have added a flowchart as suggested. We felt a flowchart focusing exclusively on section 2.2.2 without including information about steps described in the preceding or following sections would be a bit limiting, so the new flowchart

provides an overview of the methods we used to develop the dataset, beginning with the datasets (section 2.1), then quality control and merging (section 2.2), then production of the maps and iterative quality control as needed (section 2.3). The flow chart has been added as the new Figure 1 and is also pasted below:

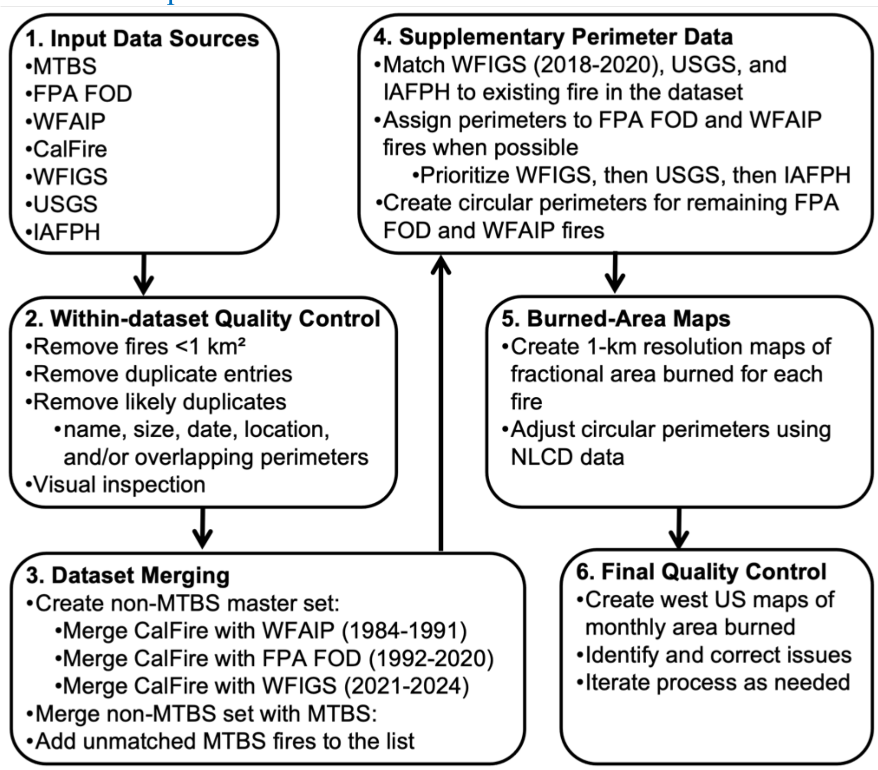

Figure 1. Flowchart outlining the approach to develop the WUMI2024a.

4. The manuscript carefully notes under-representation of pre-1992 non-forest fires and potential artifacts post-2020. Based on these weaknesses, could you add a paragraph/section that outlines "best practices" and makes recommendations for potential users?

Thank you for this suggestion. In addition to deepening our discussion of the under-representation of pre-1992 fires in non-forest areas in section 3 (Data quality and dataset intercomparison), which includes a new state-by-state assessment of how well WFAIP agrees with FPA FOD during 1992–2001 and an associated supplementary figure (Fig. S3), we have added a paragraph that summarizes the dataset weaknesses and our associated recommendations to the Conclusions section. The new paragraph reads (L489-509):

"*The WUMI2024a also has caveats. First, it relies on publicly available government records of wildfire incidents, which can be more accurately characterized as records of fire response and management rather than as purely records of wildfire occurrence. This means the WUMI2024a is missing any wildfires not recorded by a fire management agency and incorporated into a publicly available database. For example, some wildfires may have never been detected and reported, and others may have been extinguished without receiving official attention from a fire management agency. Thus, the database's completeness and accuracy are subject to temporal and geographic inconsistencies related to changes in practices related to fire detection, management, and data archival. These limitations are well exemplified by the fact that the*

WUMI2024a indicates zero wildfires in Wyoming and New Mexico in 1984 (Fig. S2). The lack of large 1984 wildfires in two western US states is likely an artifact and we suggest excluding 1984 in assessments that include the interior Western US. We further presented evidence that fire frequencies in the WUMI2024a are highly likely to be artificially low prior to 1992 (the first year of the FPA FOD) in non-forested areas of the interior West because many wildfires in these areas are addressed by non-federal agencies (ST/C&L), which are not represented in the WFAIP database (Figs. 4, S3a). The negative biases in pre-1992 frequencies are likely to be largest in the non-forest areas of New Mexico, Arizona, Colorado, and Wyoming, as these are the areas where ST/C&L fires are most common in the FPA FOD, so we caution against analyses that rely on consistent fire reporting from the 1980s through 1990s in these areas. Our database does comprehensively represent non-federal fires in California, on the other hand, via data from CalFire. Finally, we were unable to find perimeter data for 29% of the wildfire events in the WUMI2024a. While these fires are relatively small, only accounting for less than 5% of area burned in the dataset, and we do adjust the circular perimeters and maps of burned area to exclude open water, ice, and barren ground, we warn against use of the circular fire features for applications that require accurate maps of area burned. These derived features are provided as resources for applications where an approximated map of area burned is better than none at all (e.g., as inputs for some hydrological or smoke-emission modelling exercises), but should be used with caution."

Minor comments:

1. The abstract cites 22,464 events, while Sec. 2.4 first lists 22,228 parent events and then 22,464 with sub-fire replacement. Could you clarify where the difference comes from?

The difference is that often times fires represented by MTBS as individual events are actually composed of multiple smaller fires. In some cases these are official fire complexes in which multiple smaller fires merged to become a single larger fire, and in other cases what the MTBS team mapped as single event based on satellite imagery was actually constituted of multiple individual events. For example, MTBS fires often consist of several disconnected burned areas and sometimes these disconnected burned areas are distinguished by other data sources as independent fire events. In section 2.4 we specify that in cases like the ones described above we refer to the large burned area represented by MTBS as a 'parent' event and to the smaller events identified as within that parent event as sub-fires.

Thus, the larger number of events indicated in the original submission (22,464) was the total number of fires when parent fires are replaced by their sub-fires and the smaller number (22,228) represented the number of fires when parent fires are considered in place of their sub-fires. Note that these numbers have changed slightly, as identification of a small coding error and additional quality controlling reduced the number of fires by ~1%.

Importantly, we did describe the difference behind these two numbers in the first paragraph of section 2.4, which is the section that the reviewer refers to above. We have now revised that section to more clearly describe the difference between the two sets of fires representing differing numbers of events as well as why one dataset may be preferrable over another depending on the goal of the analysis. The new text is provided above in our response to this reviewer's first major comment.

2. L180 – 185 Where ignition points fall just outside the CalFire perimeter and are "snapped" to the nearest boundary, mention whether snapping is recorded as a flag and provide displacement statistics (median/95th percentile).

This comment refers to our method of adjusting ignition locations for some FPA FOD or WFAIP fires that are linked to a CalFire fire. Specifically, when a FPA FOD or WFAIP is linked to a CalFire fire, but the ignition location for the FPA FOD or WFAIP fire falls somewhere outside the perimeter of the CalFire fire (which is expected on occasion because records of ignition location are often imprecise), we adjust the FPA FOD or WFAIP ignition location to the nearest point along the CalFire perimeter. While this reviewer comment is specific to CalFire perimeters, we also perform this adjustment for FPA FOD and WFAIP fires that are linked to non-CalFire perimeters (MTBS, WFIGS, USGS, IAFPH). This was noted in section 2.2.3 about merging with supplemental datasets that include fire perimeters, but we failed to indicate that this is also done when we merge the non-MTBS fires with MTBS fires so we have added a sentence to the end of the paragraph in section 2.2.2 that describes merging non-MTBS fires with the MTBS dataset.

Importantly, instances of altered ignition locations for FPA FOD and WFAIP fires are relatively rare, and when they occur the adjustments are minor. The adjustments are rare because usually the ignition locations reported in the FPA FOD and WFIAP datasets already exist within the perimeters of the fires they are linked to. Of all 9,434 FPA FOD or WFAIP fires that appear in the final dataset, including sub-fires, 1,585 (16.8%) had adjusted ignition locations. Of these, the median adjustment distance was 406 m and 95% of distances were ≤6.6 km.

We have added a short paragraph to section 2.4 (Summary of the final dataset) describing this and how interested users can access the pre-adjusted ignition coordinates (L295-303):

"*As mentioned above, the final ignition locations for some FPA FOD and WFAIP fires are different from those originally reported because, when we found a linkage between an FPA FOD or WFAIP fire and a fire from an alternative dataset that provides perimeter data, we adjusted FPA FOD or WFAIP ignition locations to the nearest point along the perimeter if the originally reported ignition location fell outside the perimeter. Notably these adjustments were relatively rare and minor. Of the 9,434 FPA FOD or WFAIP fires in the final dataset, ignition locations are adjusted for 16.8%. Of these, the median adjustment distance is 406 m and 95% of adjustments were within 6.6 km. Users can retrieve the originally reported ignition locations from the original lists of FPA FOD and WFAIP fires that we provide as part of the WUMI2024a archive by cross-referencing the unique FPA FOD and WFAIP fire identifiers provided in the final lists of WUMI2024a fires with those in the original lists.*"