

Dear Editor and Reviewers,

We thank the editor and four reviewers for their positive assessment of this manuscript, and their valuable comments. We addressed all comments to the full extent which improved the manuscript considerably. We responded to the review comments below, and we implemented the associated modifications in the revised manuscript.

M. Sadegh on behalf of all co-authors

**RC1 (28 Feb 2024):**

Wildfires are increasingly impacting ecosystems, particularly in a changing environment. The FPA FOD version 6 data published in 2022 provide tons of fire observations and records. This study developed an extra attribute dataset for the FPA FOD data, which augments the FPA FOD v6 with nearly 270 attributes that coincide with the date and location of each wildfire ignition in the United States. For each wildfire record, the physical, biological, social, and administrative attributes were collected and added. Major classes of these attributes encompass climate, weather and fire danger, topography, land cover and vegetation, jurisdiction and management, infrastructure, and social context. This publicly available dataset can help to answer numerous questions about human- and lightning-caused wildfires and analyze fire impacts on ecosystems.

This work is important. I suggest accepting this manuscript.

Best of luck.

**Response:** We appreciate the positive assessment of our manuscript by the reviewer. Thank you for endorsing its publication.

**RC2 (28 Feb 2024):**

The manuscript describes the database variables and methodology used and then carries out some cursory analyses that describe patterns, trends and differences between naturally and human-ignited fires. I think FPA FOD-Attributes is a formidable effort and useful resource and offer just a few comments for the authors to consider.

**Response:** We appreciate the positive assessment of our manuscript by the reviewer.

L44-45. I see the point, but written as is it reads as if these changes are permanent or constant. Please rephrase.

**Response:** We appreciate this comment and revised the text to read:

*"Climatic changes in the past several decades have generally decreased the fire season moisture content of living and dead vegetation, lengthened the fire season, and contributed to a marked increase in the number of critical fire danger days across much of the United States with distinct geographical and seasonal trends and patterns (Westerling, 2016; Dennison et al., 2014; Bowman et al., 2011; Alizadeh et al. 2023)."*

L48-49. Shouldn't this be a bit more elaborated? I'm thinking about those with less prior knowledge.

**Response:** We expanded this sentence. Thank you.

L65. It's forest floor moisture, or duff moisture, rather than soil moisture.

**Response:** We changed "Soil moisture" to "Fuel moisture". Thank you.

L140. Why fuel model G, it's not particularly representative across USA (and in the current NFDRS is included in fuel model Y). And why not based on the specific existing fuel model? (an attribute that I suppose LANDFIRE includes but is absent from the database, why?).

**Response:** We appreciate this comment. This selection is driven by the data availability through gridMET. Our product is a derivative dataset, and its variables are hence dependent on the source data. Future users can adopt fuel model agnostic metrics such as 100-hr and 1000-hr dead fuel moisture rather than ERC and BI, when other fuel models are more appropriate. We clarified this point in the revised text:

*"ERC and BI are fuel model dependent, and hence are most aligned with coniferous forest fuel type (Fuel Model G), but 100-hr and 1000-hr dead fuel moisture variables are fuel model agnostic."*

L148. And why not 10-hour and, especially, 1-hour fuel moisture contents? These size classes drive fire spread. VPD has an influence on fine fuel moisture but it's not quite the same, as it is mostly responsive to air temperature, while 1-hour moisture is much more responsive to relative humidity.

**Response:** We appreciate this comment. This selection is driven by the data availability through gridMET. Our product is a derivative dataset, and its variables are hence dependent on the source data. Note that we include relative humidity, which is more directly related to 1- and 10-hr fuel moisture.

L150. Why this particular window? It would make more sense to have them for the duration of the fire, or at least for the duration of its active spread.

**Response:** This particularly designed for fires associated with uncertainty in their start dates. This includes for example, lightning started fires that smolder for a few days before they pick up by wind. We revised the text to indicate this issue.

Fire duration in the FPA FOD dataset is mostly management relevant, and is not a physical phenomenon. The end date does not necessarily relate to weather, but is governed by other considerations. Furthermore, end date is not available for many fires

L441. Florida is not mountainous.

**Response:** We removed Florida from the text.

L552. And burn severity?

**Response:** St Denis' dataset does not support burn severity.

## RC3 (11 Mar 2024):

### Overall comments

Uniqueness: this is a substantial update to prior FPA-FOD releases, but nonetheless is still v6 of an updated dataset, so the uniqueness is at best moderate.

Usefulness: given the amount of effort a typical user would have to go to in order to access and query the 270 attributes, this is a very useful dataset that assists in more harmonized analysis on the very important topic of fire occurrence.

Completeness: the attributes themselves seem to be very complete. I would note however about the completeness (or uniformity) or fire data across some state lines, i.e. Figure 4 shows a strong contrast in natural ignition density at the borders of New York state compared to Pennsylvania and New Hampshire. Is there some bias in tracking methods or thresholds that the user should be warned of? I did not recall reading any such warning in the text, or to a reference that gives more detail to these sorts of administrative contrasts in adjacent jurisdictions.

**Response:** We appreciate the positive assessment of our manuscript by the reviewer.

There are known inconsistencies in the reported fires in the FPA FOD dataset, as described in Short 2014, driven by reporting and record keeping inconsistencies, specially when fires are responded to by local authorities in rural areas with limited staff and resources.

Records are more consistent when federal agencies responded to the incident. In any case, FPA FOD dataset is the most comprehensive record of fires across the US and potentially worldwide. We addressed this point in the revised text:

*A rigorous quality assurance and quality check process was applied to the original FPA FOD dataset, but some uncertainties remain. For example, some smaller fires are overseen by local jurisdictions that may not have reporting standards as strict as those of federal firefighting agencies (Short, 2014).*

### Specific comments

Line 649: please revise the reference to the Strategy to match the (likely) permalink on Frames.gov: <https://www.frames.gov/catalog/14351> “Wildland Fire Executive Council. 2013. The National Cohesive Wildland Fire Management Strategy: Phase III Western Regional Action Plan. 99 p.”

**Response:** Done. Thank you for noticing this.

Table 1: worth differentiating MODIS NDVI (Didan, 2021) vs the Vermote, 2019 AVHRR-sourced NDVI in the “variable category” column or elsewhere on the table. Otherwise, it looks like a duplicate unless one looks at the references.

**Response:** We added the following information to that Table footnote to clarify:

*\*\*NDVI from Didan, 2021 provides monthly mean vegetation health information for the 12 months prior to fire, whereas that from Vermote, 2019 offers NDVI value in the day prior to fire start date as well as daily mean, max, and min NDVI for each month within one year prior to fire.*

Line 161: percentiles should be reported at k<sup>th</sup>, not k%.

**Response:** Done. Thank you!

Overall, I feel that the above points constitute only a minor revision, and I'd be happy with the editor confirming their completion, no need to send it back for reconsideration of the peer reviewers.

**Response:** Thank you for endorsing the publication of this paper!

#### **RC4 (04 April 2024):**

##### General Comments to the Authors

The paper is well written stylistically. I think the intent was to describe this impressive dataset, but I am unsure of the intent of the results as I don't know that there is much that is truly novel in the results. I would have liked more information regarding why you selected the variables that are in the dataset. Also, a discussion of autocorrelation would be good as I think there are likely a number of variables included that are highly correlated. This can lead to overfitting in machine learning models. In addition, why did you leave out some important variables to fire occurrence prediction? The ignition of fires by power lines is mentioned regarding the Camp fire, but that attribute is not in the database. Nor is railways lines, which is also a significant contributor to ignitions. I only mention this because there are so many variables included that I have not seen in the fire occurrence prediction literature and not some of the more common ones.

**Response:** We appreciate the positive assessment of our manuscript by the reviewer.

We changed the title of the “Results” section to “Illustrative Analysis”, since the purpose of this paper is to describe the dataset and not to test any hypothesis. The provided figures and text are only to offer some context on what the dataset entails. Thank you for this comment.

As noted in the revised text, *the FPA FOD-Attributes dataset can support descriptive, diagnostic, predictive, and prescriptive wildfire analytics, including development of machine learning models*. Not all included variables are proper for ML modeling. Even those that are helpful for AI and ML modeling have quite a bit of overlap and correlation. Careful selection of variables is needed for future users to ensure the data is supportive of the intended goals. We addressed this issue in the revised text:

*“We recommend that future users carefully select variables among the wealth of information provided in FPA FOD-Attributes. Specifically for AI-ML modeling, variables have substantial overlap and correlation, which need to be addressed.”*

Note that the original FPA-FOD data contains information on fire cause such that fires ignited by powerlines and rail would be coded as such. However, a dataset of electric grid distribution lines is not publicly available. Hence, we cannot add that information to the FPA FOD-Attributes dataset. Railroad started fires occur linearly along the railroads and are less modulated by human behavior and more controlled with technological advancements that prevent such fires. Since this is a rather well-known phenomenon, it is of less interest for the intended audience of this dataset.

##### Specific comments to the Authors

-Has there been any attempt to quantify the error in the ignition data?

**Response:** This has been addressed in the original data descriptor of FPA FOD dataset by Short 2014. We addressed the temporal and spatial uncertainties in the ignition data in multiple locations in our paper. Although uncertainties remain – which we endeavored to address in our dataset – FPA FOD is one of the most comprehensive sources of information for wildfire analyses in the US. Similar uncertainties, if not more, are omnipresent in other datasets.

Short, Karen C. 2022. Spatial wildfire occurrence data for the United States, 1992-2020 [FPA\_FOD\_20221014]. 6th Edition. Fort Collins, CO: Forest Service Research Data Archive. <https://doi.org/10.2737/RDS-2013-0009.6>

-How useful and available are all these variables in real time operationally i. e. What is the operational feasibility of using this dataset?

**Response:** This data can support operational analyses in a variety of ways, including:

1. Enhanced understanding of the drivers of fire ignitions by cause can help fire managers make operational decisions given environmental conditions and weather forecasts
2. The FPA FOD-Attributes dataset can be used to train AI-ML models, which can in turn be forced with environmental data and weather forecasts to predict the probability of human ignitions

The FPA FOD-Attributes in itself is not an operationally-aligned dataset.

-How does evacuation time influence the probability of ignition of human or lightning caused fires? I can see the linkage with important decisions that are made after the fire has started, but I don't see a direct link to prediction of human or lightning caused fires.

**Response:** As noted in the revised text, *the FPA FOD-Attributes dataset can support descriptive, diagnostic, predictive, and prescriptive wildfire analytics, including development of machine learning models.* Not all included variables are proper for ML modeling. Evacuation time might be more aligned with impact analyses than prediction of human ignitions.

-Section 3.1. Manual comparison; why did you only use 100 points? There are 2.3 million fire records, so that is 0.004%.

**Response:** The dataset was compiled automatically and systematically using Python codes (which are shared along with the dataset). We only needed to test a couple of points to see if what the manual tests return similar values to that of the code. But to ensure our testing is robust, we selected 100 points from various geographical locations and times. This is an arbitrary selection, and to be honest, an overkill.

-Section 3. 2 Seven fires out of 2.3 million also seems a little low, in my opinion. Especially if they are based on the same weather variables. I would expect weather derived indices using the same weather source to be the same. I'm not clear on how this is a validation, maybe I am missing something?

**Response:** Same as above, this is an additional test to ensure our data match those reported in the literature. The variables reported in this section are the same as those in Khorshidi et al. 2020, since their study is our testbed.

-Line 383 why did you use those attributes at the ignition point from Oct to Dec? The active fire front and areas of spread for a large fire will be far beyond the point of ignition months later. How large was the final fire perimeter?

**Response:** Here, same as the two comments above, selections are arbitrary and just to show the reported values in FPA FOD-Attributes dataset match those of the source data.

Figure 2-what are the acronyms, please spell them out.

**Response:** They are all explained earlier in the manuscript, but we spelled them out in the Figure caption. Thank you.

-Why did you use the media, how reliable is this as a validation source? Are there not fire perimeters and fire records for all suppressed fires from the fire management agencies?

**Response:** Same as the previous few comments, this is an additional step to show our records match those reported by experts interviewed in the media. This is an extra, independent source of information to validate our dataset. The data is rigorously tested even without this step.

-Were all validation fires recent fires or did you also include older fires? The dataset goes back to 1992, but most validation fires appear to be from more recent years.

**Response:** The 100 fires used in section 3.1. are selected from all years, including earlier years. However, those used in sections 3.2. and 3.4. are bound by the source of the data (i.e., literature and media reports).

Results; why are there references to your findings? Should that not be discussion?

**Response:** To ensure the flow of the paper, we did not repeat this information in the Discussion section and opted to provide references in the “Illustrative Analysis” section. We realize different authors adopt different writing styles.

Lines 428-437. Since you are also discussing results. The trends in human and lightning caused fires; is this the same as other studies? In Canada we see the opposite.

**Response:** The trends shown in this figure only apply to the US. Not only these trends are different in Canada, but also trends change from one region to another in the US.

Line 428-Where are decadal trends shown? Fig 3 shows annual trends.

**Response:** We removed decadal. Thanks.

Line 433 if human caused fires are increasing how is this related to fire prevention strategies? We can't prevent lightning fires, but we can use fire restrictions etc. to decrease human caused fires.

**Response:** Certain categories of human-started fires are declining (e.g., smoking) and other categories are associated with an increasing trend. Fire prevention strategies are certainly important and they can potentially prevent fires, but climatic changes made fuels more receptive to fire, WUI areas expanded, and population of the US increased in the past several decades which collectively contributed to increasing human ignitions.

Line 439- CONUS is used many times before this; define sooner

**Response:** Done. Thanks

Line 443-What other subcategories of natural fires are there?

**Response:** Volcanos is one of the subcategories of natural fires, however in FPA FOD, all the natural fires are recorded as lightning started fires.

Fig 4 - Nice graphic, I like the addition of histograms

**Response:** Thank you!

Paragraph 468-475 and Figure 6. I'm still not sure of the take-away here. More people, more small fires except in California? We know already that lightning caused fires occur everywhere and often where people are not.

**Response:** Generally smaller fires occur closer to high population density. Larger fires – mainly lightning started – occur farther away from human settlements, and are associated with lower population density. Forty million people live in California with larger population densities compared to the rest of the West (although not necessarily where population is dense), and hence small and large fires in California should only be compared within the state.

Line 517 - How easily accessible are these datasets for fire management, again going back to operational availability or is the intent for research only?

**Response:** We have developed a freely and publicly available portal for this purpose:  
<https://fpafod.boisestate.edu/>

Line 549 - 550 Here it is stated that the dataset can't be used to assess large fire growth, but Section 3.3 is about temporal evolution of fire attributes. I found this really confusing.

**Response:** We specifically state that the FPA-FOD attributes cannot be directly used to be paired with daily fire growth – particularly for fires that may have large growth days several days to weeks after the ignition. The temporal evolution of weather associated with one fire was used as a test case to validate the dataset. It was not focused on fire growth, nor was it intended to indicate that FPA FOD-Attributes support fire growth studies. Rather, we demonstrate that the extreme fire weather conditions on the day of ignition were very likely responsible for the fire resisting containment/suppression efforts.