

Interactive comment on “A spatial database of wildfires in the United States, 1992–2011” by K. C. Short

K. C. Short

kcshort@fs.fed.us

Received and published: 2 December 2013

I thank the reviewers for their extensive feedback. I have organized my responses into categories, with references to specific comments and their page numbers in the review document.

Data Quality/Error Checking

Pages C129–130: The reviewers correctly note that some fires in the dataset do not map within the boundary of the state named or otherwise interpreted as the fire's location in the fire report. They also correctly note that there may be errors in either the coordinates or the nominal state designations that result in the mismatch of information in those cases. Of the ~1.6 million records in the database, there are approximately

C219

2,000 with this inconsistency. About 1,700 of those 2,000 records are from federal systems of record and were not summarily excluded from the database, because, as explained in Section 2.2.2.: " . . . some agency units span or have fire protection responsibility or cooperative agreements in more than one state (e.g. NPS Appalachian National Scenic Trail, BLM Miles City Field Office) and some nominal state designations were based on the state designation in the unit name, which may not reflect the true state location of the fire. The flagged subset of suspect federal fires was therefore visually inspected and only records with obvious spatial errors (e.g. fires <40 hectares mapping several states away from the expected domain) were excluded from further processing. In some cases, fires did not map within the state expected from the fire report, but did map within the domain of the interstate reporting unit (e.g. fires reported from Dinosaur National Monument mapping in Colorado and Utah, across which the unit spans) or were responded to under cooperative agreement or as a threat to the unit's land. In other cases, fires mapped near enough to the proclaimed state or unit such that the mismatch was ostensibly due to imprecision of the reported location. We did find fires that were clearly mis-located, because they mapped over water, but we retained them if they fell within the expected domain of state or FPU."

There are, however, nearly 300 records from nonfederal or interagency systems of record that, according to our rules to exclude any fires from those systems that mapped outside of the expected state domain (p. 309, lines 19–24), should be removed from the database. Some of the points map so close to the expected states' boundaries that it appears that they were retained when a different version of the GIS layer depicting US state boundaries was used for QAQC purposes, and that layer had slightly different linework. Those fires map so near to the expected state that it seems imprudent to remove them from the published dataset outright; instead, users should be allowed to remove or retain them based on their analysis criteria. A handful of other records that were apparently added at a late stage of dataset development from the interagency ICS-209 system are clearly mis-located and were retained in error. I am grateful to the reviewers for drawing those to our attention, and the approximately 30 records that are

C220

clearly in error will be removed from the second edition of the database.

Page C131: The reviewers “recommend that any spatially uncertain records from federal sources be flagged for further scrutiny by end-users.” To this, I must quote back to them the first half of the sentence in which that recommendation comes, explaining that “there is no easily automated method for determining the accuracy of these records.” One must keep in mind that we are talking about point locations intending to represent the origins of wildfires, and if all of the coordinates in the dataset that are “spatially uncertain” must be flagged, all ~1.6 million records in the dataset will be flagged. These data are inherently “uncertain,” and all should be scrutinized further by potential users. In fact, the reviewers themselves make a point to emphasize, on Page C132, that, “the responsibility of data content, quality, and interpretation always falls on the user.”

Data Access

Pages C129-130: The reviewers explain that the US Forest Service (FS) server was down when they initially attempted to access the data product from the FS Research Data Archive. While such outages should be rare, the FS is working to minimize disruption of services and is considering replication of infrastructure among the possible approaches to improve data access.

Data Format

Page C131: The reviewers recommend “providing the data in an additional format more easily accessible to open source software platforms.” To this end, a version of the database in the non-proprietary SQLite (<http://www.sqlite.org/>) format has been created and, once fully tested, will be submitted to the FS Research Data Archive as an additional format to be included as part of the data product.

Data Utility

Page C131-132: It is unclear what the reviewers mean when they ask for an “outline” of how the dataset “might be useful, with a special consideration for the spatial limitations

C221

of the location field.” And I’m surprised by the request for “a single example of how the combination of origin and area could be used for a spatial analysis of wildfires during the database period.” In the first paragraph of the Introduction, I offer several ways in which these types of data, spatial warts and all, have been used and will continue to be used: national fire danger rating applications, fire potential forecast models, geospatial fire modeling systems. I go on to explain that those applications and systems “are relied upon to generate consistent national data for risk assessment, planning, budget formulation, and decision support at multiple scales. Outside of the operational realm, spatiotemporal analyses of US wildfire activity are used increasingly to characterize local, regional, and national patterns and trends as they relate to factors such as climate, population, land use, and fire policy and to predict how wildfire activity and values at risk may be influenced by changes in those factors.” Every bit of literature cited in that opening paragraph (and there’s quite a bit) provides an example of how wildfire activity data – often at much coarser spatial resolutions – have and can be used. Then, in the Discussion, I specifically walk through “examples of questions to ask before proceeding with any analysis of wildfire activity using the FPA FOD,” including #2 (Page 331, line 18) “What spatial resolution is required for the analysis?”, which I answer in the text as follows:

“The FPA FOD should provide point locations of wildfires at least as precise as a Public Land Survey System section (2.6-km² grid). But many non-federal records that were excluded from the database due to imprecise fire location information could be used directly from the source systems for analyses at, for example, the county level (e.g. per the Cohesive Strategy [WRSC, 2012]). If the analysis does require precise wildfire location information, analysts must bear in mind that the coordinates provided in the FPA FOD may or may not represent actual ignition points, or even fall within the actual burn perimeter, due to reporting inconsistencies and imprecise georeferencing. Moreover, the spatial impacts of large fires, which, by definition, burn far from their ignition points, can be characterized imprecisely at best with this or any point-based reporting dataset. Burned area estimates from the FPA FOD will be necessarily georeferenced

C222

to the contributing wildfires' ostensible points of origin or nominal domains (e.g. state, reporting unit). Without fire footprints and temporal progression information, one cannot assert for a time period of interest that a given area burned, for example, in the state of Nevada; rather, an estimate from the FPA FOD would represent area burned by fires reported as starting (or having been discovered) within the specified time period and spatial domain. The FPA FOD is therefore most useful for characterizing the statistical properties of fires reported as starting at a given place and time. Supplemental information about spatial and temporal impacts of large fires can be found in the ICS-209 records and MTBS dataset using the linkages provided in the FPA FOD."

It seems unnecessary and well outside the scope of this data paper to go beyond this and to hold the reader's hand and walk them through some hypothetical application of the data, as the reviewers seem to recommend.

Page C131: The reviewers explain that "while we were hopeful the database might include the maximum burn extent (where available) for fires . . . we recognize the tremendous amount of work compiling this incomplete information might require." Final fire size is a required element and reported for all fires in the dataset. Based on the next comment, at the bottom of C131, I must interpret their objections to be aimed the lack of fire perimeters in the "spatial" dataset. Mapped fire perimeters do not exist for all reported wildfires, and the effort that would be required to identify, compile, and associate what does exist to the fire reports in the FPA FOD, is far beyond the scope of this work. However, knowing the value of fire perimeters for certain analyses of wildfire activity, perimeters from the MTBS project have been linked to many of the records in the FOD, and that effort alone was a "tremendous amount of work." That link is made via the MTBS ID attribute in the FOD, and because this link has been made, it is untrue to state, as the reviewers assert, that "a database user could only plot the areal extent of a fire by assuming a circle (or any polygon actually) of that area centered in the presumed origin." The truth is that, for any record in the FOD with an MTBS ID populated, that ID can be used to identify and view the corresponding perimeter from

C223

the publicly available MTBS dataset: <http://www.mtbs.gov/dataaccess.html>.

Data Sources

The reviewers contend that the "manuscript could have been improved if it had included more detailed data from all the states instead of just relying upon mostly the already published sources, or perhaps at the very least including some justification for the primacy (with respect to data quality) assigned to federal versus non-federal data." Section 2.1 explains that, in addition to "published" nonfederal datasets, we did, in fact, seek and acquire what should be deemed "unpublished" data from other US states, but we did so as efficiently as possible by capitalizing on the efforts of other projects (e.g., MTBS, SWRA, EFETAC) that had already attempted to compile this information rather than duplicating those efforts and going directly to the states' fire services with redundant requests. In addition to acquiring otherwise "unpublished" nonfederal data from these other projects, we learned from those efforts that data that would meet our location and information requirements simply did not exist from all states. If "missing" data can be acquired in the future, they may be incorporated in a new edition of the FPA FOD. The second half of the comment asks for "some justification for the primacy . . . assigned to federal versus non-federal data," which can be found on Page 312, lines 13-15, as follows: "When we identified a set of federal and non-federal wildfire records that were redundant, we always selected a federal record to retain, because the federal records tend to be more fully attributed (e.g. fire name and cause more consistently populated)."

The reviewers "felt as if restricting the database to only wildfires instead of both wildfires and prescribed burns could limit the usefulness of the database." This work was prompted by a need for a spatial database of US wildfires, not prescribed fires. An entirely different approach would be needed to assemble a similar database of spatially explicit prescribed fire records, to what degree they exist, and such work was far beyond the scope of this effort.

C224

Page C130: The reviewers indicate that the “final pages of the discussion strongly drove home the important role these data can play. We especially appreciated the emphasis on understanding potential limitations and inaccuracies of this dataset.” They suggest shifting this “important section” to an earlier position in the manuscript, “perhaps to the introduction,” and “making this section more prominent by separating this user-focused discussion from the broader discussion of methods used for data verification.” I do not agree that a discussion of the potential uses and limitations of the dataset should precede the description of how it was developed and how it was evaluated, which is information necessarily presented in the Methods section. I will also argue that the Discussion section (Section 4), which is largely “user-focused” in its entirety, is already separated by the “discussion of methods used for data verification”, which is in Section 2.3.2.

Page C132, Re: Abstract, Lines 8-10: Beyond the “standard” issues of data quality and interoperability, the overlapping nature of the multiple federal and nonfederal fire reporting systems puts an additional burden on those seeking to pool data from the multiple sources for analyses of wildfire activity, i.e. “the onus is then on the user to check for and purge redundant records of the same fire . . .”.

Re: Abstract, Line 12: The term “scrub” is used in the manuscript to refer to the process of removing redundant records from the compiled dataset. The term, “data scrubbing” is often used interchangeably with the term “data cleaning” or “data cleansing” (http://en.wikipedia.org/wiki/Data_cleansing). Udechukwu et al. (2005) refer to the process as “de-duplication.” I prefer “scrub” for its brevity, and because it seems no less of a non-technical, or “slang,” term than “clean” or “cleanse”. Moreover, the term, “de-duplication” appears to imply that the problem is more straightforward than it is, given that truly duplicated, or identical, records are much easier to identify and purge from a dataset than are those with redundant but non-identical (i.e. inconsistently formatted) information.

C225

Page C133, Re: Abstract, Line 16: The reviewers appear to be asking for a count of fires in the dataset that are reported as larger than 2.6 square kilometers in size, “as this would then require some consideration of what the ‘fire location’ represents.” I interpret this comment to mean that it is not clear in the abstract that the locations in the database are points generally intended to represent ignition locations (to the degree possible) and not burned area footprints (i.e. fire perimeters). If this is indeed the case, I propose inserting the word “ignition” at the beginning of Page 298, Line 16, for clarification.

Page C133, Re: Abstract, Lines 19-21: The requested detail is provided in the preceding Line 16, which states that the locations are “at least as precise as a Public Land Survey System section (2.6 km² grid).”

Re: Page 298, Line 26: The recommended change will be made in the revised manuscript.

Re: Page 299, Lines 12-15: Yes, multiple references are necessary to support the far-ranging assertion that, “spatiotemporal analyses of US wildfire activity are used increasingly to characterize local, regional, and national patterns and trends as they relate to factors such as climate, population, land use, and fire policy and to predict how wildfire activity and values at risk may be influenced by changes in those factors.”

Re: Page 299, Line 21: It is not the first NICC citation in the References section, which, for papers by the same author in the same year, I have organized alphabetically by title. I will defer to the editor’s direction as to whether I have misinterpreted the manuscript preparation guidelines posted here: http://www.earth-system-science-data.net/submission/manuscript_preparation.html

Re: Page 299, Line 26: The recommended change will be made in the revised manuscript.

Re: Page 300, Line 22: The recommended change will be made in the revised

C226

manuscript.

Re: Page 300, "acronym overkill": A table of abbreviations, acronyms, and aliases and their definitions was originally submitted with the manuscript as Appendix A, not Table 11, as it appears now (on Page 355). I had not caught this change until after reading the reviewers' comment, nor had I noticed that the reference to that Appendix/Table appears to have been edited out of the manuscript. The statement, which should have appeared on Page 301, Line 21 (concluding the paragraph), originally read parenthetically: "Abbreviations, acronyms, and aliases used repeatedly throughout this paper are defined in Appendix A."

Re: Page 301: Sources of federal versus nonfederal wildfire records included in the FPA FOD are distinguished in Table 1. Re-insertion of reference to the Appendix/Table of acronyms at Page 301, Line 21, should help readers keep track of the data sources introduced to this point.

Re: Page 303, Lines 8-10: I propose to edit the statement to read, "They explained that the quality of the location data for some of the records, particularly the "older" subset, "constrains a comprehensive, regional-scale analysis to a 1-degree [approximately 111-km] grid resolution."

Re: Page 305, Line 11: I can offer the following information bulletin for reference: [USD] US Department of the Interior: Fire Program Analysis Application, Bureau of Land Management, Washington, D.C., USA, Information Bulletin No. 2009-040, available at: http://www.blm.gov/wo/st/en/info/regulations/Instruction_Memos_and_Bulletins/national_info_040.html, 2009.

Re: Page 305, Line 26: Generation of the MTBS dataset is detailed in the Eidenshink et al. (2007) publication cited in the statement. To "briefly describe how the MTBS is made," I propose to simply edit Lines 25-27 to read, "Record identifiers from the ICS-209 application and the satellite-derived national Monitoring Trends in Burn Severity

C227

(MTBS) fire-perimeter dataset (Eidenshink et al., 2007) are also included for a subset of the fires, providing, in essence, bridges to those information systems." Or, as an alternative to the term "satellite-derived", I could insert the term "remotely sensed."

Page C134, Re: Page 306, Lines 16-17: While not all state fire-service organizations were contacted directly for data, data from all US states were sought, albeit indirectly. Section 2.1 explains that, in addition to "published" nonfederal datasets (including the database of the National Association of State Foresters, which attempts to acquire wildfire data from all US states), we did, in fact, seek and acquire what should be deemed "unpublished" data from other US states, but we did so as efficiently as possible by capitalizing on the efforts of other projects (e.g., MTBS, SWRA, EFETAC) that had already attempted to compile this information rather than duplicating those efforts and going directly to the states' fire services with redundant requests for data. In addition to acquiring otherwise "unpublished" nonfederal data from these other projects, we learned from those efforts that data that would meet our location and information requirements simply did not exist from all states.

Re: Page 307, Section 2.2: I don't agree that removal of this paragraph would improve readability, nor do I believe it needs to be more verbose to "stand on its own."

Re: Page 307, Line 19: I disagree. Negative longitude is the NWCG data standard (http://www.nwcg.gov/pms/stds/standards/longitude_v1-0.htm), and an important step in our standardization process was to format those values accordingly.

Re: Page 308, Line 10: The letter case used for fire names was highly variable in the source datasets, and conversion to uppercase was the most straightforward way to consistently format the entries. (The NWCG Incident Name standard is not case sensitive.)

Re: Page 308, Line 12: The question implies that there is some master list of "correct" fire names in existence somewhere, which there isn't. Even if all names were "accurate", there's no requirement that fire names be unique, and thus fire names are not

C228

sufficient to “ensure connectivity” across sources.

Re: Page 309, Line 16: FPU's largely follow agency administrative boundaries. For clarification, I propose to edit the statement in question so that it begins, “All US states and territories are fully contained within the collective extent of the FPU boundaries, . . .”

Re: Page 310: Numbers of records acquired and considered for inclusion in the FPA FOD versus those actually retained are reported in the Results section. The goal of this effort was not to present a full accounting of the “accuracy” of the data in the various reporting systems, and records deemed nonviable due to missing critical elements were not further scrutinized for the accuracy of the other potentially relevant elements. In other words, records excluded from further processing because they lacked start dates or final fire sizes were not checked for potentially viable location information. Therefore, it is not possible to provide figures that specifically address inaccuracies in the wider universe of fire reporting data, nor was it ever within the scope of this work to do so.

The FPA FOD described and evaluated in this manuscript has been published in the FS Research Data Archive and assigned a persistent Digital Object Identifier (doi:10.2737/RDS-2013-0009). Any subsequent updates, including corrections and additions, to the dataset will result in a new edition of the complete dataset, which must be assigned a unique DOI. Users attempting to access the FPA FOD using the DOI included in this manuscript will be notified, via the FS Research Data Archive interface, that a new edition is available and will be provided the necessary DOI to link to the updated dataset.

Re: Page 311, Line 26: The recommended change will be made in the revised manuscript.

Re: Page 312, Line 14-17: There is, of course, no perfect knowledge base to reference and use to evaluate the relative “accuracy” of the redundant records. If there was, it

C229

would obviate this entire effort. The reviewers ask, “why not keep the records which are the most thorough?” But that’s exactly what we tried to do, using a general rule: Page 312, Lines 13-16 explain that “we always selected a federal record to retain, because the federal records tend to be more fully attributed (e.g. fire name and cause more consistently populated).”

Re: Page 313, Lines 16-20: Beside from not having superpowers, this “method” wasn’t used because it was not anywhere within the scope of this effort to “accurately validate these data,” which, if one recalls correctly, are 1.6 million wildfire records from a 20-year period. I don’t think this fact needs to be asserted in the manuscript.

Re: Page 313, Line 22: The recommended change will be made in the revised manuscript.

Re: Page 314, Line 26: Protection Type codes and definitions are listed in Table 5, which is referenced on Page 309, Line 5.

Re: Page 316, Line 26: The reference to the NWCG memo that provides the guidance is already included at the end of the sentence.

Re: Page 316, Line 28: I will check again for inconsistencies and make any necessary changes in the revised manuscript.

Re: Page 319, Line 21-23: This comment implies that complete and accurate (i.e., “true”) reference values for numbers of wildfires and area burned exist, but they don’t.

Page C135, Re: Pages 321-322, As explained in the preface to those subsections, on Page 319, Lines 8-10: “We attempted to evaluate, at least nominally, the completeness of the resulting dataset by comparing estimates of annual fire numbers and area burned, by state, from the FPA FOD to other published estimates. . . . Because the published estimates of annual wildfire numbers and area burned can differ considerably among sources due to inconsistencies and errors in measurement and reporting (e.g. see Urbanski et al., 2009), several sources of reference estimates were included

C230

in our assessment. We consider agreement in estimates of the same metrics from the FPA FOD and a given reference source as a proxy for 'completeness' with respect to the latter. How accurately the reference estimates reflect actual wildfire activity is unknown; however, none are presumed to represent the true values, and therefore completeness, in fact, cannot be known by way of this assessment, or, indeed, at all. In other words, agreement of estimates from the different sources implies nothing about their accuracy."

The review comments aimed at the subsections that follow these statements appear to ask for revisions that, again, imply that complete and accurate (i.e., "true") reference values for numbers of wildfires and area burned exist, but they don't. Moreover, the reviewers contend that they "got a bit lost in the reasoning for relative vs. absolute similarities" between the (imperfect) FOD and (imperfect) REF estimates. However, the only reference to similarity in "absolute terms" is in the statement, "Because we were interested in the similarity of estimates (in relative rather than absolute terms), we calculated, for each state and year, the ratio of (1) wildfire numbers and (2) wildfire area burned estimated from the FPA FOD and the same metrics from the reference source."

Re: Page 322, Lines 7-20: This section simply explains how we derived a single "score" for each state, which has value in that it provides a single metric of "agreement" that can be displayed graphically, for example, in Figure 5.

Re: Page 322, Lines 23-25: I do not believe that the suggested edit would improve clarity.

Re: Page 323, Lines 15-20: I don't follow.

Re: Page 323, Line 22/Figure 2: I disagree with the reviewers' apparent definition of "juxtapose" and their assertion that it is misused in the manuscript.

Re: Page 324, Yes, the Section number (3.1.) is missing, and the correction will be

C231

made in the revised manuscript.

Re: Page 325, Replace "disunity" with "lack of unity"?

Re: Page 327, Lines 23-26: This concluding statement (and the preceding sentence) is positioned as-is to "set up," or lead into, the subsequent paragraphs, which provide examples of "trends" in the data (Figures 7-10) that are best explained by changes in non-federal reporting levels.

Re: Page 328, Line 13: The correction will be made in the revised manuscript.

Re: Page 331, Lines 2-3: These are such important points that I would rather err on the side of caution and re-emphasize them, in case they are not "obvious" to all at this point.

Re: Page 331, Line 3: I prefer the stronger term, "critique", which means to "evaluate in a detailed and analytical way."

Re: Page 331, Lines 14-16: I disagree: the concluding sentence does not "answer" the "yes-no question" as directly as the lines preceding it. It is simply expanding on the "If no, . . ." portion of the answer.

Page C136, Re: Page 328, Line 13: The recommended change will be made in the revised manuscript.

Re: Page 334, Line 10: The correction will be made in the revised manuscript.

Re: Page 349, Table 7: Of course it is incomplete and does not need to be complete to help illustrate the process described in Section 2.2.4., which may be otherwise difficult to follow.

Re: Tables 9 and 10: The general lack of reference figures by state for the year 1998 is explained in Section 2.3.1, Sources of Reference Estimates (i.e. Page 320, Lines 1-3).

Re: Figures 4, 8, 9, and 10: The reason for the discrepancies is fully detailed in the

C232

discussion, which explains the reporting biases that Figures 7-10 are intended to illustrate (and which is also evident in 4A). Figure 4B shows area burned, which is not at issue in Figures 7-10, which focus on fire numbers. On Page 332, Lines 19-26, we explain why agreement in area burned can be fairly high despite the discrepancies in fire numbers.

References

Udechukwu, A., Eziefe, C., and Barker, K.: Independent de-duplication in data cleaning, *Journal of Information and Organizational Sciences*, 29, 53-68, 2005.

Interactive comment on Earth Syst. Sci. Data Discuss., 6, 297, 2013.