

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

A. Johnson (Referee)

johnson.aiden@gmail.com

Received and published: 24 September 2013

What we did

Because we were charged with a dataset review, and not a data analysis, this review differs from a typical scientific manuscript review. We adhered to instructions provided here: http://www.earth-system-science-data.net/review/ms_evaluation_criteria.html.

Instead of focusing on the accuracy and clarity of any one analysis, we focused on the accuracy and clarity of the data itself. The manuscript portion of this product is essentially metadata, explaining the methods of data collection and verification, placing this dataset into the historical context of fire datasets, and providing insight into proper uses and possible biases of the dataset. Therefore, our review addressed the issues

C128

of the usefulness, reliability, and accessibility of the data, with additional comments on how these issues are addressed in the paper. We reviewed the manuscript as a team of reviewers over a period of approximately 6 weeks.

We ran multiple tests on the database:

(i) We accessed the data using ArcGIS, Microsoft Access, and R (using RODBC with the Microsoft Access driver). (ii) We successfully grouped by state and by year. (iii) We plotted the geographic locations of fires for one state in one year (Colorado 2010) and noticed ~6 fires clearly plotted outside the borders of Colorado (>5 km). (iv) We repeated this last exercise for the state of Wyoming and noticed ~11 fires plotting outside the state border. (v) We noticed in the case of these ~17 discrepancies that two types of errors might be contained in the database: - The geographic location datum might be incorrect, in which case the state column might be correct but the coordinates are incorrect. - Alternatively, the state assignment might be incorrect (and the coordinates are, in fact, correct).

We thought the author had done a commendable job of compiling a vast amount of data of national significance into one place; the database and the associated manuscript should be of value to many different types of users. A complete database of wildfires in the US is very important for management and scientific needs, especially since despite the existence of hundreds of reports, collections and databases, none seems complete. We found this work to be a very detailed and well-organized representation of the arduous endeavor of piecing together an indispensable dataset. We appreciated the historical background to efforts to compile a comprehensive database (e.g., Show and Kotok, 1923), as well as all the descriptions of sources and error-checking. The tables and figures were very helpful in illustrating the data conversion processes, but some reviewers felt as if a flowchart describing the quality control process might have been more helpful for readers.

We tested the accessibility of the dataset, and found that the Forest Service

C129

server was down when we attempted to access the link provided in the manuscript (doi:10.2737/RDS-2013-0009). A quick email to the Forest Service yielded <~2 hour turnaround time for the actual database files! That's fantastic, but perhaps mirrored (or replicate) servers should be a consideration for improved database accessibility.

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.

The reviewing team felt the author had done such a commendable job with this important section that we felt it warranted consideration of a shift forward ("front-load") in the manuscript, perhaps to the introduction. The author could also consider making this section more prominent by separating this user-focused discussion from the broader discussion of methods used for data verification.

Greatest concerns with ms

Error checking. Our database evaluation, and especially our identification of fire locations plotting outside of a state's boundaries, suggested a simple and rapid check could be performed that evaluates whether the coordinates associated with a fire match onto the state assigned to the same fire.

Although the manuscript identifies and explains the different procedures used for the spatial verification of fire reports from federal records, our group felt that the potential accuracy gained by critically filtering these records would be more valuable than the loss of the relatively few fires well outside their indicated jurisdiction.

In our limited analysis of spatial errors, we were able to identify fires from federal sources which were mis-located (incorrect coordinates) as well as fires which were mis-identified (incorrect state). Our ability to identify the type of error was dependent on external information for certain well-publicized fires, which is of course not available for all fires.

C130

Because there is no easily automated method for determining the accuracy of these records, we recommend that any spatially uncertain records from federal sources be flagged for further scrutiny by end-users.

Database format. Considering this dataset and manuscript is to be published in an open source journal, we recommend providing the data in an additional format more easily accessible to open source software platforms. The data are currently available as an Microsoft Access 2007 / 2010 database (.accdb), or an ESRI geodatabase (.gdb), both proprietary formats.

The data would be available to a broader range of users if it was also available in an open format, such as a primitive text file (e.g. .csv) or a broadly supported compressed data format (e.g. .netcdf). We acknowledge that a large .csv is opened in Microsoft Excel (and subsequently saved) can be truncated at the record number corresponding to an overall file size of 2Gb. However, the same .csv can be easily used by freely available statistical and spatial analysis programs (e.g. R, GRASS). Although proprietary Microsoft and ESRI file formats are convenient for many institutional users, an open format would allow access by a broader range of users. Perhaps a caution to potential users (e.g., a reminder of how many records should appear in particular files) might be in order to allow for simple verification of dataset completeness.

Spatial usefulness. We appreciated the author's candor about the limitations of the database, but we felt the manuscript could be improved if the author outlined how it *might* be useful, with a special consideration for the spatial limitations of the location field. While we were hopeful the database might include the maximum burn extent (where available) for fires (after all, the manuscript is titled "A spatial database of wild-fires in the United States, 1992-2011"), we recognize the tremendous amount of work compiling this incomplete information might require.

With only an origin and an area, a database user could only plot the areal extent of a fire by assuming a circle (or any polygon actually) of that area centered on the presumed

C131

origin. Again, the manuscript could be improved with a single example of how the combination of origin and area could be useful for a spatial analysis of wildfires during the database period.

Data sources. The reviewer team felt 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. This is particularly true on pages 306 in the Data Sources section.

The reviewing team felt as if restricting the database to only wildfires instead of both wildfires and prescribed burns could limit the usefulness of the database, but we also understand this might have entailed considerable additional effort. We appreciated the consideration that some prescribed fires can transform into wildfires (as the author acknowledges on line #2, page 309).

Although the historical background was interesting, and the error-checking details appeared thorough, the reviewing team felt that many sections of the manuscript could have been more concise without reducing the usefulness of the data product..

Tiki taka

The reviewing team identified numerous small jejune issues or concerns with the manuscript, and some examples are provided below. Our hope is that a final, careful reading along these lines should further improve the manuscript.

Abstract

Lines 8-10. The responsibility of data content, quality, and interpretation always falls on the user. Perhaps the point trying to be made here is that data checking for redundancies is a laborious and time consuming task that could be negated by use of this universal data set? A gentle reminder to the reader might be in order.

Line 12: Nowhere is "scrub" defined. We can guess what this means, but we think this

C132

"slang" belittles the work the author has done.

Line 16: In the ms, there should be an acknowledgement of how many fires were >2.6 km², as this would then require some consideration of what the "fire location" represents.

Lines 19-21 The use of the words "fairly high resolution geospatial analysis" is not specific enough. Any abstract should provide enough detail for one to discern whether data are of value. Stating the range of spatial resolutions found in the data set or the least amount of spatial resolution afforded by the data would be more informative.

Introduction

Page 298 Line 26 'for example' is not needed. Page # 299, Line 12 -15 You cite 12 references as examples. Are all necessary? p. 299 Line 21 should be NICC 2012a since it's the first NICC ref Line 26 Situation Report should precede the corresponding acronym p. 300 Line 22 "Reports for all fires on lands under federal protection" By the end of p. 300, we were feeling a bit of acronym overkill p. 301 Perhaps a graphic laying out federal vs. non-federal sources might help a reader p. 302 There is a demonstration of how the ms could be tightened up at the end of this page... Are lines 25-27 necessary? ("how daunting a task...") p. 303 Perhaps to any reader of a spatial database article it would be clear what a 1-degree grid resolution represents, but why not add this information (at least parenthetically) so that there is greater context with respect to the 2.6 km² cutoff criterion p 305 No reference for Line 11? Page 305 Line 26 Please briefly describe how the mtbs is made. All other mentioned datasets are described at length

This section ("Introduction") generally struck the reviewing team as too verbose, although we acknowledge the importance of considering the many varied audiences who might be interested in this database. Separating the introduction to the database from background information might provide the author with a way of helping readers iden-

C133

tify relevant background information (history of wildfire databases) versus introductory information (specific to the author's efforts).

Methods p. 306 Lines 16-17 is troubling—clearly this represents a monumental effort, but perhaps then specify *which* states and territories or at very least, the number of other states and territories from which “data were not sought directly.” P 307:

2.2 Data processing and quality control .Is this paragraph necessary? My thoughts are no, why do subsections need an introduction or make it longer and more informative so it can stand on its own.

2.2.1 Standardization Page 307 Line 19 don't need to include 'negative' Page 308 Line 10 Why were names converted to uppercase? Page 308 Line 12 Why weren't the names corrected to insure connectivity across sources and accuracy of the data? Page 309, line 16: How are FPU boundaries created in the first place? Page 310: The author acknowledges 'some obvious errors were not included while getting data, but there is no mention of how many of those erroneous data were excluded nor is there any acknowledgement of how additional potential errors (including those identified by this reviewing team) might be accommodated through subsequent versions of this database. p. 311 Line 26 insert “ues do not match exactly”

2.2.4 Removing redundant records Page 312 Lines 14-17 This seemed like a spurious choice, why not keep the records which are the most thorough and/or accurate? Rather than maintaining federal records preferentially. Page 313 Lines 16-20 How could one accurately validate these data without using visual interpretation of aerial imagery for the area immediately proceeding the fire event? Please describe why this method wasn't used. p. 313 Line 22 delete the word in from “flagged in during Steps” p. 314 Line 26 What is protection type #5? p. 316 Line 26 insert a reference for “recent interagency guidelines” p. 316 Line 28 be clear and careful about FireCode system versus FIRE_CODE. I note inconsistencies

2.3 Completeness Evaluation p. 319 Lines 21-23 This is refreshingly honest, but the

C134

level of uncertainty or some measure of data reliability should be published in order for data to be usable with reportable certainties.

Page 321 Lines 14-17 Wording obscures the point, please be more direct about when which source was used. We got a bit lost in the reasoning for relative vs. absolute similarities, for example. Clarify please.

2.3.2 Methods of assessment Page 322 Lines 7-20 What value is this? Why do I care if the state records matched the federal records? What are you trying to show with this comparison? p. 322 Pithify your prose: Lines 23-25 could be edited to “A total of ... were obtained (Table 1). The bulk ...”

Results

Page 323 Lines 15-20: Please make a statement similar to " State level data provided additional information that improved the overall data set quality." Or, at least work to orient the reader with a lead-in or to finish the paragraph reminding us what you've just conveyed. Line 22 The two maps in Figure 2 should be complementary rather than juxtaposed, these two would be expected to match not oppose each other the comparison is of similarity rather than contrast. p. 323 We do not think details from your initial 2010 effort are required in this ms. This would allow for deletion of half the lines in this paragraph.

Section number missing from 'Agreement with national estimates': 3.1?

Discussion p. 325 is nicely set up with reference to historical overview p. 325 replace “disunity” Page 327 Lines 23-26 This should appear early to set-up the paragraph for the reader more effectively. p. 328 line 13 incorporated is misspelled Page 329 Line 21 'fairly well' is not helpful without a number associated with it or describing it. Page 331 Lines 2-3 This seems obvious and has been previously stated in the manuscript. p. 331 line 3 critique should probably be evaluate p. 331, Lines 14-16 Should be moved to the top of the paragraph to answer the question first followed by discussion of

C135

why. p. 332 line 27 delete "to" in front of check p. 334 line 10 fulfill misspelled Page 349, Table 7 This table only illustrates examples of cases of redundant records and is unlikely to be complete. Is it necessary to include?

Tables 9 and 10: It appears 1998 data are mostly missing. It is not clear why.

Figures 4, 8, 9, and 10: How are the different sources of data used? For example, in figure 4, B, the area burned seems mostly consistent between USFS/NICC and FPA FOD. However, in figures 8, 9 and 10, different sources are not consistent or are very different from each other. For the same parameters (like area burned or number of wildfires), why would different data sources yield such differences?

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