

Response to Reviewer 2

Our response to the comment is given in italics and any additional text in blue.

The manuscript in its current state is ambiguous with regard to the RPD's relationship with the Global Palaeofire Database. I mean this both in a direct sense (i.e. how much of the data in the RPD was directly pulled from the GPD, both earlier versions and the most up to date web version?) as well as in a logistical sense (was the Global Palaeofire Working Group involved in the creation of the RPD?). Statements about these relationships are ambiguous (see specific comments below regarding L98-99, 266-269, and SI Table 1). For example, the current presentation of the RPD does not allow the viewer to tell from where each dataset came. Despite the assertion that the RPD "is a community effort," it is rather unclear how members of this community (e.g., the Global Palaeofire Working Group) were engaged and involved in the process. More broadly, I am very curious as to why the RPD is not being directly integrated into the existing community framework of the Global Palaeofire Database and Working Group?

The Global Palaeofire Working Group and the Global Charcoal Database were originally created by the lead author under the auspices of the IGBP Cross-Project Initiative (2003-2006), who continued to lead (and fund) this effort until 2015. Jenn Marlon, who is an author of the current manuscript, is the co-leader of the second phase of the GPWG and several other authors are part of the steering group of the GPWG2. Most of the authors of the current manuscript have contributed to the GCD and are therefore part of the GPWG community. However, since the GPWG2 has no direct funding for database construction, GCD holdings are out of date, and as we state in the manuscript there are discrepancies between different versions, missing metadata and dating information. Furthermore, there is no effort to update existing age models. It was the recognition that this situation is preventing detailed analyses of palaeofire histories that lead us to create the new database in order to be able to undertake new analyses. As we state in the response to Dan and Jack, we are happy for the data we have compiled to be incorporated into Neotoma and indeed to be used to update the GCD. We have no intention of the RPD becoming a permanent data repository.

As a final comment, the RPD does indeed contain information about the source of the individual records in the entity table. In response to Dan Gavin's comments, we have undertaken a thorough review of the records including replacing entries originally derived from various versions of the GCD with records derived from Neotoma and other repositories. We currently have 532 entities provided by the authors of the manuscript, 64 entities derived from Neotoma, 35 derived from Pangaea, 16 derived from NOAA NCEI, 55 derived from the European Pollen Database, 49 derived from IMPD, and 14 derived from the Arctic Data Centre. There are still 836 entities which were derived from the GCD and have not been updated, although we have updated metadata and dating information for 57 of the original GCD records.

Despite the merits of this manuscript and its associated RPD, I feel that if released in its current form (i.e. as a self-standing database) and without a plan for community integration, then this work could have a detrimental impact on the broader palaeofire field and the willingness of its members to share data. By downloading multiple community-driven databases that embody the spirit of open data, making improvements and expansions, and then creating a separate and less accessible database (see comment below), I would argue this is a step in the wrong direction. Why go through all of the work of improving community

databases to then refuse to return the database to these same communities and in these same frameworks? Despite its flaws, the GCD's web interface is significantly more 'available' than an SQL database (see comment below). The same would be true for Neotoma. As Drs Gavin and Williams have previously expressed in their comments, interfacing the RPD with existing community databases would achieve greater impact and utility for the palaeofire community.

Again, we reiterate that the RPD is not meant to be a public repository or to replace Neotoma or the GCD. It was created in order to provide a reliable database for further analyses. We are happy for these data to be integrated into public repositories. As we have said in response to Dan and Jack's comments, once the RPD paper is published, we will encourage each individual contributor to lodge their data in Neotoma and will work with Neotoma to facilitate this.

Although I understand that SQL is open source and theoretically available to all. I assert that providing the data only in a SQL format poses an equity and accessibility issue. Use of the SQL database requires downloading a large program to read the files. To then use or access the data requires knowledge of a programming language to construct queries. I argue that this is an undue burden on accessibility. For anyone without knowledge of SQL to access the data would require potentially hours of instruction and learning. As a test case, it took nearly two hours for me to download and install MySQL, and then to import the data. Barring integration with community databases, I think the simplest way to address this issue would be to also provide text or csv files of the tables that make up the SQL database. This way, anyone (even those without knowledge of this database style and language) could assess the data freely and easily. I feel that ignoring the hurdle that the SQL format poses to potential users of the RPD would represent undue gatekeeping in direct contradiction to ESSD's aim of "furthering the reuse of high-quality data of benefit to Earth system sciences." Alternatively, I reiterate the suggestions of Drs Gavin and Williams to integrate the RPD with either Neotoma or the Global Palaeofire Database.

The advantage of SQL is that it can be queried in order to extract sub-sets of records for subsequent analysis. However, we agree that some people might find SQL difficult to work with and thus we have also lodged the individual tables as separate csv files in the Reading repository.

Manuscript main text – Specific Comments:

L24-30: References? These statements should be supported by relevant literature.

We have added appropriate references here, as follows:

recent years have seen wildfires occurring in regions where they were historically rare (e.g. northern Alaska, Greenland, northern Scandinavia: Evangeliou et al., 2019; Hayasaka, 2021) and an increase in fire frequency and severity in more fire-prone regions (e.g. California, the circum-Mediterranean, eastern Australia; e.g. Abatzoglou and Williams, 2016; Dutta et al., 2016; Williams et al., 2019; Nolan et al., 2020).

We have added these new references:

Abatzoglou, J. T., and Williams, A. P.: Impact of anthropogenic climate change on wildfire across western US forests, *Proceedings of the National Academy of Sciences*, 113, 11,770–11,775, <https://doi.org/10.1073/pnas.1607171113>, 2016.

- Dutta, R., Das, A., and Aryal, J.: Big data integration shows Australian bush-fire frequency is increasing significantly, *Royal Society Open Science*, 3, 10.1098/rsos.150241, 2016.
- Evangelou, N., Kylling, A., Eckhardt, S., Myroniuk, V., Stebel, K., Paugam, R., Zibtsev, S., and Stohl, A.: Open fires in Greenland in summer 2017: transport, deposition and radiative effects of BC, OC and BrC emissions, *Atmospheric Chemistry and Physics*, 19, 1393-1411, 10.5194/acp-19-1393-2019, 2019.
- Hayasaka, H.: Rare and extreme wildland fire in Sakha in 2021, *Atmosphere*, 12, 1572. <https://doi.org/10.3390/atmos12121572>, 2021.
- Nolan, R. H. Boer, M. M., Collins, L., Resco de Dios, V., Clarke, H., Jenkins, M., Kenny, B., and Bradstock, R. A.: Causes and consequences of eastern Australia's 2019–20 season of mega-fires, *Global Change Biology*, 26: 1039-1041, doi:10.1111/gcb.14987, 2020.
- Williams, A. P., Abatzoglou, J. T., Gershunov, A., Guzman-Morales, J., Bishop, D. A., Balch, J. K., and Lettenmaier, D. P.: Observed impacts of anthropogenic climate change on wildfire in California, *Earth's Future*, 7, 892–910, <https://doi.org/10.1029/2019EF001210>, 2019.

L33-36: Same comment as above.

We have added appropriate references here, as follows:

However, wildfires exhibit considerable local to regional variability because of the spatial heterogeneity of the various factors controlling their occurrence and intensity (Bistinas et al., 2014; Andela et al., 2019; Forkel et al., 2019). Thus, it is useful to use information that can provide a picture of regional changes through time. Charcoal, preserved in lake, peat or marine sediments, can provide a picture of such changes (Clark and Patterson, 1997; Conedera et al., 2009).

We have added these new references:

- Andela, N., Morton, D. C., Giglio, L., Paugam, R., Chen, Y., Hanson, S., van der Werf, G. R., and Randerson, J. T.: The Global Fire Atlas of individual fire size, duration, speed, and direction, *Earth System Science Data*, 11, 529–552, <https://doi.org/10.5194/essd-11-529-2019>, 2019.
- Clark, J. S., and Patterson, W. A.: Background and local charcoal in sediments: Scales of fire evidence in the paleorecord, in: *Sediment Records of Biomass Burning and Global Change*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 23–48. https://doi.org/10.1007/978-3-642-59171-6_3, 1997.
- Conedera, M., Tinner, W., Neff, C., Meurer, M., Dickens, A.F., and Krebs, P.: Reconstructing past fire regimes: methods, applications, and relevance to fire management and conservation, *Quaternary Science Reviews*, 28, 555–576, <https://doi.org/10.1016/j.quascirev.2008.11.005>, 2009.
- Forkel, M., Andela, N., Harrison, S. P., Lasslop, G., van Marle, M., Chuvieco, E., Dorigo, W., Forrest, M., Hantson, S., Heil, A., Li, F., Melton, J., Sitch, S., Yue, C., and Arneeth, A.: Emergent relationships with respect to burned area in global satellite observations and fire-enabled vegetation models, *Biogeosciences*, 16, 57–76, <https://doi.org/10.5194/bg-16-57-2019>, 2019.

L66: I think a clear statement as to the logistical relationship between RPD and the Global Palaeofire Database as well as the GPWG is needed here. Was the RPD as created by these

authors done so independently from the GPWG? Was there community involvement and input during the creation of the RPD?

Please see response to the general comment above. We will add a statement about the purpose of the RPD (lines 73 to 74) as follows:

The RPD is designed to facilitate regional analyses of fire history; it is not designed as a permanent repository.

L98-99: Does this mean that any/all charcoal datasets which were not previously publicly available were provided by one of the authors of this article? Relating to my query regarding L66 above, I think a statement is needed to explain who the authors of this article are and what constituted a 'contribution' to this article, especially as the bulk of the RPD is derived from the GCD and GPWG. Were original authors of datasets in the GCD version not included as co-authors by virtue of their having made their data publicly available?

Please see response to general comment above about the relationship between the authors of the current manuscript and the GPWG. Our statement here reflects the fact that the authors have produced and contributed new charcoal records, some of which are unpublished, and that these were used as a basis for expanding the geographic and temporal coverage.

Authorship here is not because of the data contribution per se but because the authors were actively engaged in quality control of individual records and deriving age models.

L103-108: Although the download of MySQL and the importation of the RPD was relatively straightforward and aided by the documentation provided by the authors, I wonder if this format poses an equity issue and accessibility issue. Namely, could the tables also be provided as csv or text documents so that those less technologically inclined could still view the RPD without using SQL queries?

As we state above, we have also lodged the individual tables as separate csv files in the Reading repository.

L126-127: However, in many cases, I suspect that the accuracy of GPS coordinates for sites are not this great. How can trailing nought values in these latitudes and longitudes be differentiated as being reflective of accurate GPS location versus merely artefacts of non-exact GPS?

We agree that the accuracy of some of the coordinates is unlikely to be to within a metre. In some cases, the coordinates were given to with an even higher precision in the original sources so our decision here simply reflects the need to standardise the reporting. We note that giving coordinates to six decimal places is common in many public repositories e.g. Pangaea, Neotoma.

L159-160: Charcoal measurements are not always made in terms of volume (e.g., by dry mass basis). How were these types of measurements integrated into the RPD? Or were they simply omitted?

The statement here refers to sample volume because together with sample depth this indicates whether the record was sampled contiguously or not. The measurement types are given in the entity table. We did not exclude measurements if they were not made on a volume basis.

L266-269: This sentence is misleading as both versions 3 and 4 are several years old and many datasets have since been added to the online version of the GPD. E.g. as of 16 November 2021, the GPD contains 1231 cores. This is not a fair comparison. A more direct measure of the RPD's value would be the number of new records (not sourced from the GPD or any earlier versions of the GCD).

The earlier versions of the GCD have been used for numerous publications and have been quality controlled to some extent. Although the online version of the GCD contains a larger number of cores, some of these are duplicates of records that were in one of the earlier versions, some only provide the age of samples and not the depths (or dating information), and some are missing crucial metadata. However, we agree that it is important to be more specific about the sources of the records in the RPD, including the number that have been taken from other repositories, the number of entirely new records, and the number of records that have been updated.

This first version of the RPD contains 1676 individual charcoal records from 1480 sites worldwide. This represents a 101% increase compared to the number of records in version 3 of the Global Charcoal Database (GCDv3: Marlon et al., 2016; 736 sites), a 58% increase compared to version 4 (Blarquez, 2018; 935 sites) and a 37% increase compared to the online version of the GCD (1223 sites). The RPD includes 840 records that are not taken from a version of the GCD, and provides updated or corrected information for a further 485 records that were included in the GCD. Raw data are available for 14% of the entities and concentration for 67% of the entities; influx based on the original age models is given for 16% of the entities. The original age models for 67 (4%) of the records included in the RPD were derived solely by layer counting, U/Th or Pb dates, or isotopic correlation and therefore are already expressed in calendar ages. New age models have been run on 22 (33% of these records (33%). Additionally, new age models are available for 807 (50%) of the remaining charcoal records where the original chronology was based on radiometric dating.

L326: Here again, the ‘community’ needs to be defined (similar to my comment regarding L66 and L98-99).

Please see response to general comment above. We have modified this statement to read: The Reading Palaeofire Database (RPD) is an effort to improve the coverage of charcoal records that can be used to investigate palaeofire regimes

L329: It is very difficult to judge whether there is actually expanded coverage (as per my comments regarding L266-269). More direct comparison or quantifications are needed to assess the validity of this statement.

We have modified the text describing the database (see response above) and indeed this demonstrates expanded coverage.

Tables 1, 2 , and 4: For the fields which were ‘Selected from predefined list’, it would be prudent to provide the choices contained within these lists. Upon reading the SI, I see there are tables containing these. It might be good to note this in the main text.

We agree that it would be useful to refer to these pre-defined lists given in the Supplementary (Table S2) in the main text and have now added a reference to this Table to all the appropriate sections.

Figure 4: The legend text is fairly small and hard to read, please consider making this text larger.

We have provided a new version of this figure and have improved the readability of the legends.

Supplementary Information – Specific Comments:

SI Table 1: An extremely useful and important field that is missing from this table is the

source of each site (i.e. which version of the GCD did each come from, was it a new addition by the authors or this article, etc.).

This information is given in the database and we think that the table would be unwieldy if we added this information (at the level of detail required to be useful) to the table in the Supplementary.

SI Table 1: I believe that all of the datasets coming from the NOAA database are inaccurately cited. As per the NOAA database site: “Please cite original publication, online resource and date accessed when using this data. If there is no publication information, please cite Investigator, title, online resource and date accessed.” For example, Bass Lake Kandiyohi County should be attributed to Marlon and Umbahowar.

We have checked the references for sites derived from NOAA and from other sources and updated them where necessary.

SI Table 1 : There appears to be duplicates of the Wild Tussock site.

Citation: <https://doi.org/10.5194/essd-2021-272-RC2>

This was a typo and has now been removed.