General comments:

This paper presents a comparative assessment analysis of three primary fire datasets across the Southwestern Mediterranean region, including Spain, Portugal, France, and Italy. The paper, its structure, and its analyses have been well designed and presented. The paper was also easy to read. Nevertheless, some minor issues can be addressed to improve the technical and presentation quality of the paper.

We thank the reviewer for the positive comments to our manuscript. We followed the reviewer's suggestion and modified some parts of the manuscript to improve it.

Specific comments:

I think the motivation and the contribution of the paper still are not well presented. So, I invite the authors to clearly explained the novelty of their research and its applications and advantages for the real-work uses.

In order to clarify the novelty and application of our work, we included the following sentences in the manuscript. Please see below:

L 66: In this work, we compared for the first time the three most recent remotely-sensed fire datasets of individual fires (Fire Atlas, FRY and GlobFire) with high-quality fire datasets compiled by regional agencies across the most active fire region in Europe (i.e. Southwestern Mediterranean basin) during the common period of observations (2005 to 2015).

L 68: While most previous studies have evaluated remotely-sensed data on a fire-by-fire basis, this study aggregates individual fires across months and pixels (0.25°) and seeks to estimate how much temporal variability in both fire frequency and burned area are captured by remotely-sensed datasets.

L 250: The low capacity of gridded BA products to detect small-mid fire events (< 100 ha) can be improved by the generation of products based on higher resolution sensors in the range of 10–30m (Roteta et al., 2019). RSD of individual fires derived from finer gridded BA would provide better accuracy in the fire metrics, specifically for NF. In addition, the MCD64A1 product already incorporates the uncertainty of detection as an auxiliary variable of gridded BA data (Giglio et al., 2018). RSD could benefit from this and report similar information at individual fire level.

L 281: In practical applications, our results may provide guidance for end-users regarding RSD limitations at different fire size thresholds.

L 284: Fire agencies may benefit from the spatial and temporal consistency of remotely sensed data to support their operational fire mapping system at regional/national level.

In the abstract, the abbreviation AG has been defined for "fire agencies," but later in a sentence: "Our results show that RS datasets were highly correlated with AG." It is unclear to the reader what the authors do mean here. I would suggest they review these sentences from lines 14 to 17.

In response to this suggestion, we included the abbreviations (i.e. RSD: remote-sensing datasets; AGD: agency dataset) in the manuscript. Please see below:

L 9: Although very promising, these datasets still lack a quantitative estimate of their accuracy with respect to fire agencies (AG) historical records.

L 12: Here, we compared three state-of-the-art RS datasets (RSD; Fire Atlas, FRY and GlobFire) with a harmonized fire agency dataset (AGD) compiled by ground-based monitoring systems across the Southwestern Mediterranean basin (2005-2015).

L 14: We assessed the spatial and temporal accuracy of RSD with respect to both burned area (BA) and number of fires (NF). RSD and AGD were aggregated at monthly and 0.25° resolutions, considering different individual fire size thresholds ranging from 1 to 500 ha. Our results show that both datasets were highly correlated, in terms of monthly BA and NF but RSD severely underestimated both (by 38% and 96%, respectively) when considering all fires > 1 ha.

L 76: The fire agency dataset (AGD) was built from multiple ground-based sources, including records from Portugal, Spain, France and Sardinia in Italy (Table 1).

~~L 85: A harmonized database was constructed from the aforementioned fire agencies (AG) datasets.~~ (deleted)

L 96: We used the most recent remote-sensing datasets (RSD) of individual fires: Fire Atlas (Andela et al., 2019a, 2019b), FRY (Laurent et al., 2018a, 2018b) and GlobFire (Artés et al., 2019; Artés Vivancos and San-Miguel-Ayanz, 2018).

L: 103: All RSD provide fire starting and ending dates, location and the final burned area for each retrieved fire event.

L 115: Table 2. Description of the remote-sensing datasets (RSD) of individual fires, including the digital object identifier (DOI) and reference of each dataset. FA: Fire Atlas; FRY_M05: FRY MODIS (5 days) and GF: GlobFire.

L 119: We compared burned area (BA) and number of fires (NF) estimated by RSD of individual fires, with the ground-based reference AGD (Fig. 2).

L 120: We evaluated the ability of RSD to reproduce observed temporal and spatial patterns of fire activity observed in AGD by fitting ordinary least squares (OLS) linear regressions and using different metrics (OLS slope, R-squared correlation, and bias) to measure RSD accuracy. Only the common period between RSD datasets and AGD has been considered in the following (2005–2015).

L 124: We applied a land cover filter to the RSD using CORINE Land Cover (CLC) to exclude fires located within agricultural or artificial lands that are not reported by the fire agencies.

L 133: Figure 2. The general framework for comparison of RSD burned area and number of fires with AGD ground-based observations across a range of individual fire size thresholds (1 to 500 ha).

L 136: We retrieved the slope coefficient of OLS regressions and the coefficient of determination (R-squared) as a proxy of agreement between RSD and AGD.

L 137: Slope values greater than indicated an underestimation of fire activity as seen by AGD and vice versa.

L 147: Likewise, AGD databases do not provide systematically ignition points.

L 148: Thus, to overcome this limitation, we aggregated both RSD and AGD datasets onto a 0.25° grid (≈ 25 km), setting a common ground for both datasets.

L 150: To examine the spatial agreement between RSD and AGD, we calculated the relative error (Eq. 1) for each grid cell.

L 154: We first analyzed the monthly distributions of RSD and AGD for all fires (>1 ha) aggregated across the whole studied area.

L 156: The best agreement between RSD and AGD occurs mainly during the warm season (May to October; see Fig. 4).

L 172: Table 3. Temporal correlation of monthly and annual burned area and number of fires between RSD and AGD for all fires (>1 ha) between 2005 and 2015 across the study domain.

L 184: Fig. 6 shows the evaluation of RSD through different metrics over the continuum of fire size thresholds.

L 190: Figure 6. Evaluation of RSD through different metrics including the slope (left), R-squared correlation (middle) 190 and relative error (right) for both burned area (top) and the number of fires (bottom) over a range of individual fire size thresholds (1 to 500 ha).

L 195: As expected from previous results, RSD strongly underestimated BA, especially when including smaller fires. However, a few exceptions are seen for fires < 50 ha mainly over eastern Spain, suggesting that RSD detect in that case more fires than AGD. This may be related to a few small prescribed fires that are not reported in AGD.

L 204: Likewise, RSD strongly underestimated NF (Fig. 8), likely disregarding those smaller fires not detected by MODIS.

L 206: Nevertheless, the overall relative error between RSD and AGD decreases when focusing on larger fires for both NF and BA, highlighting the important role of fire size on RSD accuracy.

L 215: Here, we compared remotely-sensed fire data with ground-based datasets across the Southwestern Mediterranean basin to better understand RSD limitations and guide end-users.

L 223: The ability of RSD to determine individual fires depends mainly on two features: the processing algorithm and the underlying reliability of the BA product.

L 233: Hence, differences among RSD are rather expected to be associated with the underlying algorithm used to identify single fire events.

L 235: RSD were found to better estimate BA than NF.

L 240: Lastly, regional features of the fire regime may constrain RSD accuracy.

L 241: These fires do not contribute very much to the total annual burned area but significantly harm the performance of the RSD estimations in terms of NF (Turco et al., 2016).

L 246: Among the analyzed RSD datasets, FA displayed a slightly better performance, with a lower relative error.

L 249: However, uncertainty in MODIS estimations largely outpaces the uncertainties across the RSD.

L 251: According to Turco (2019), the spatial agreement between AGD and RSD increases at lower resolutions, being generally best when aggregating the data onto a 1° grid (approximately 110 km) or beyond.

L 254: Evaluating RSD on shorter timescales and/or finer spatial resolutions would likely deteriorate the agreement with AGD.

L 258: Further studies are still needed to examine RSD spatio-temporal variability at the fire patch level (i.e. assign individual fires from RSD to AGD) in order to more precisely quantify the dataset accuracy at the fire scale.

L 270: Specifically, when considering fires > 100 ha, RSD denoted reasonable agreement with observed AGD.

L 272: Generally, the RSD underestimation of BA and NF for smaller fires is related to the coarse spatial resolution (500 m) of the pixel-based BA product and other observation limitations, preventing the detection of small fires. Features of fire regime at regional scales may also influence the RSD accuracy (e.g. fire size, density, and spread rate).

L 278: Also, RSD were found to better estimate BA than NF. This is rather expected as numerous small fires, which are not detected by satellites, do not contribute very much to the total burned area across the study region.

L 281: A quantitative estimate of uncertainty is crucial to the correct interpretation of RSD and users should take into account their limitations.

L 282: Our findings suggested that global RSD of individual fires can be used for fire modeling, however caution is advised when drawing from smaller fires (< 100 ha) across the Mediterranean region. Future studies using high-quality ground-based fire data in other regions of the world featuring different fire regimes would provide further insights on RSD uncertainties.

The same problem is there again for the sentence 19-20: How RS (Remote Sensing) and AG (fire agencies) can be in agreement!

In response to this suggestion, we modified the following sentences. Please see below:

L 17: The agreement between RSD and AGD strengthens when increasing the fire size threshold, with fires > 100 ha denoting higher correlation and much lower error (BA 10%; NF 35%).

L 19: The agreement between RSD and AGD was also the highest during the warm season (May to October) in particular across the regions with greater fire activity such as the Northern Iberian Peninsula.

In the line 128, the authors have mentioned that investigated different fire size thresholds increasing from 1 to 500 ha. It would be interesting for the reader to know how it has been applied to the original dataset from MODIS that does have a spatial resolution of 500 m.

In response to this suggestion, we included complementary information in the following sentence. Please see below:

L 127: As RSD are prone to omit smaller fires (<25 ha) due to the coarse spatial resolution of MODIS product MCD64A1 (500 m) and other limitations, we investigated different fire size thresholds increasing from 1 to 500 ha.

Again, in equation (1) in section 2.3.1, the authors have used AG as mathematical/physical parameters. However, it was merely an abbreviation for Fire Agency. More clarification here would be helpful for the readers.

We rewrote and moved the equation to section 2.3.

L 122: We calculated the relative error (ε) as:

$$\varepsilon = 100 \times \frac{BA_{RSD} - BA_{AGD}}{BA_{AGD}} \qquad (1)$$

where, BA_{RSD} represents the burned area (BA) detected by remote-sensing datasets (RSD) and BA_{AGD} represents the burned area registered in the fire agencies datasets (AGD) over the study period. The analysis was repeated for the number of fires (NF).

L 139: We also calculated the relative error (Eq. 1) over the study period.

The quality of some graphics and figures can be increased to show more details, e.g., Fig. 7 & 8 are interesting, but the details are not visible.

We increased the graphical resolution of all Figures of the manuscript. For instance, see below Fig 7 and 8:

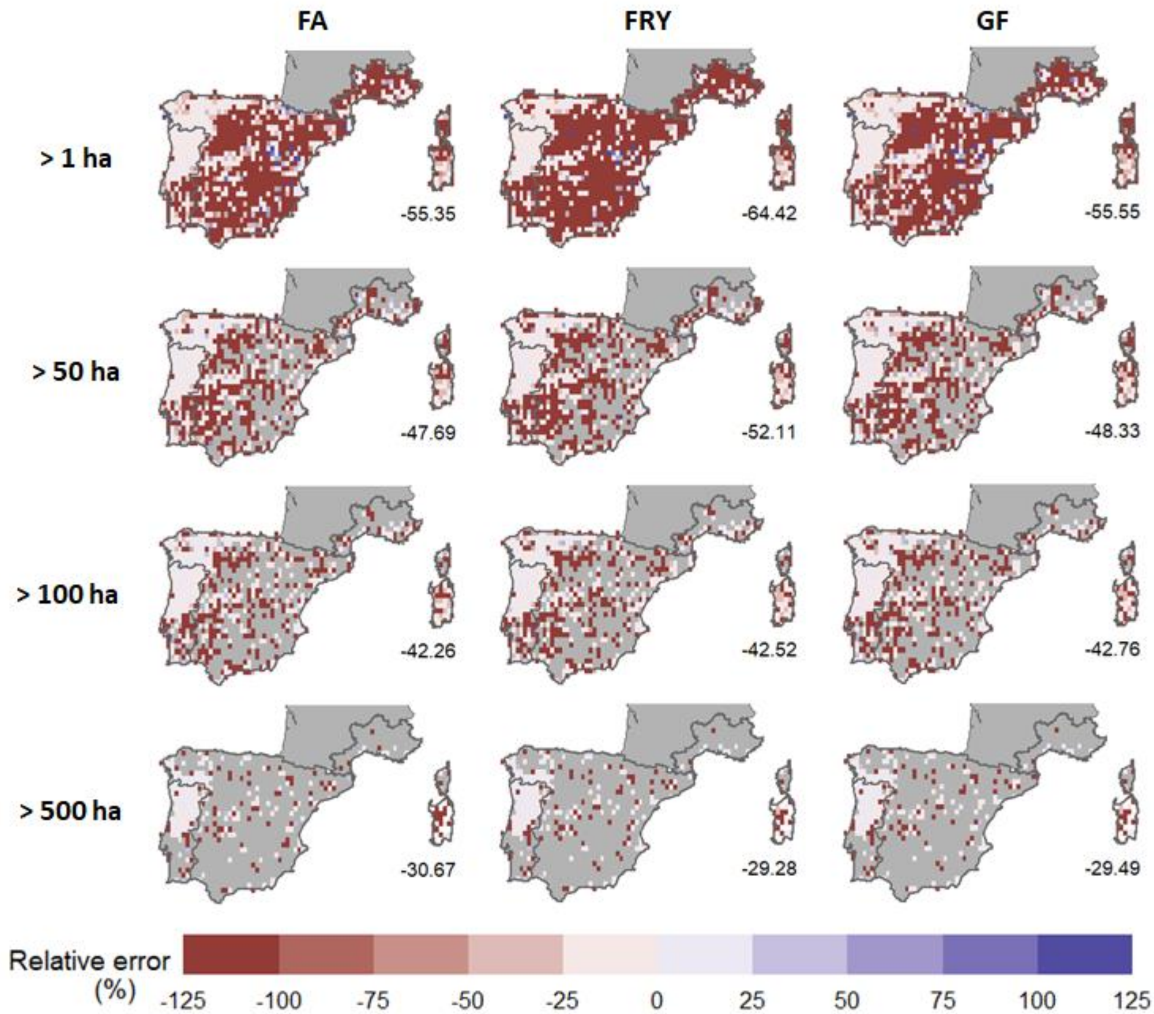

Figure 7. The relative error (ε) of the total BA computed as the relative difference between BA_{RSD} and BA_{AGD} data over different individual fire size thresholds (1, 50, 100 and 500 ha). The overall ε is indicated on each map.

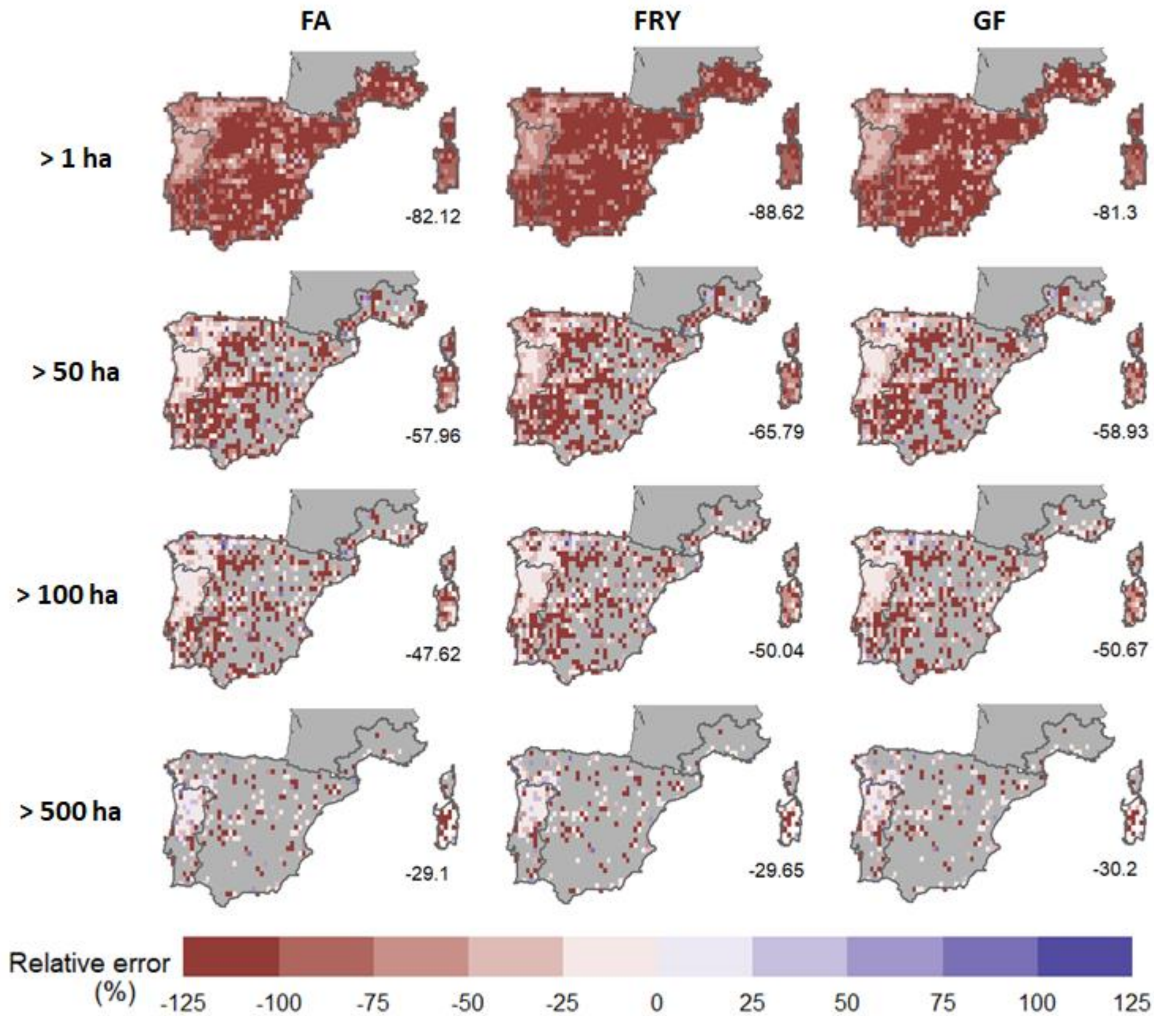

Figure 8. Same as Fig. 7 but for NF.

I think the figures in the supplementary material can also be added to the main manuscript. They will help the readers to see all the information related to study in the same way.

We agree that the Supplementary Material might help the reader to understand the methodology applied in our study, however, they are not critical to support the conclusion of the manuscript. The supplementary information is an outcome related to the pre-processing of the RS datasets, which is a preliminary step in the methodology section to implement the RS x AG comparison properly. In this sense, we think the inclusion of this information in the main text would disrupt the flow of the manuscript.

We however included the Supplementary material as part of the manuscript in Appendix A (Fig. A1 and A2), which are referred to in the main text (L 113 and L 126). In this sense, complementary information can be easily accessed by the reader and may support the understanding of the applied methodology.

Appendix A

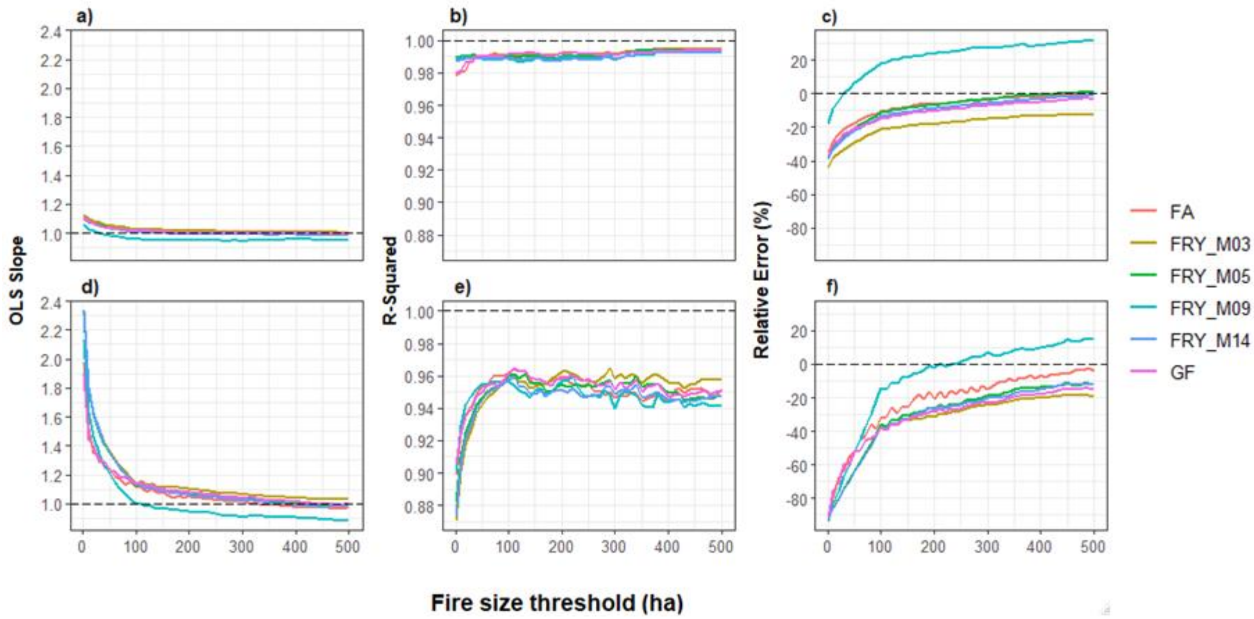

Figure A1. Evaluation of RSD including all FRY cut-off values (3 to 14 days) through different metrics including the slope (left), R-squared correlation (middle) and relative error (right) for both burned area (top) and the number of fires (bottom) over a range of individual fire size thresholds (1 to 500 ha). Dashed lines indicate a perfect fit between RSD and AGD.

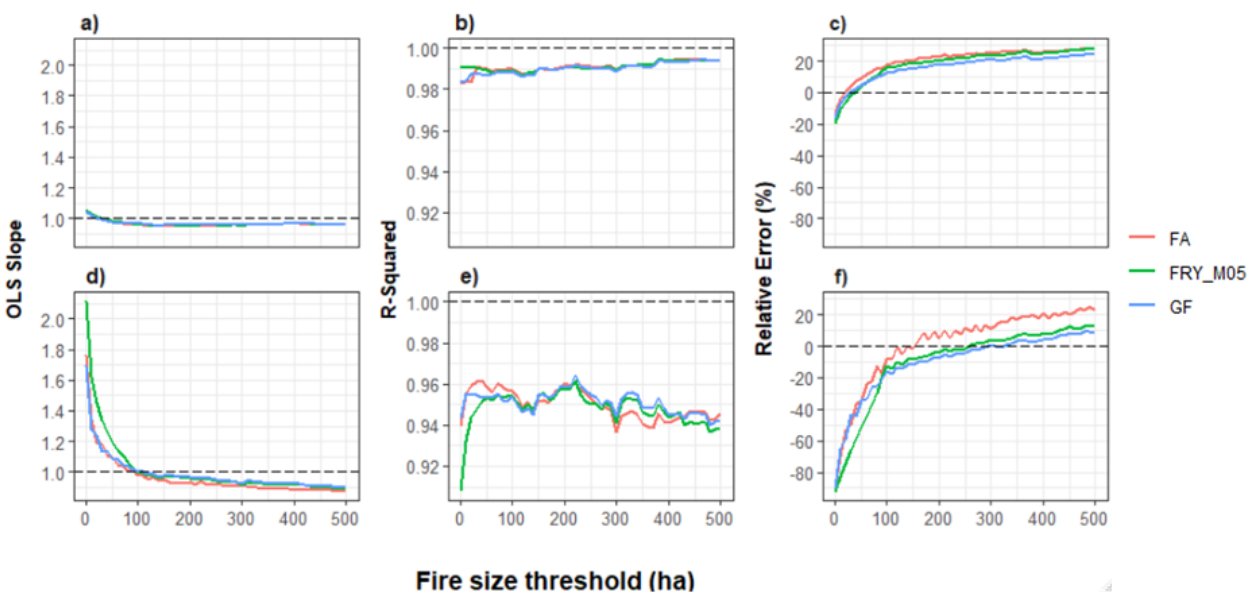

Figure A2. Evaluation of "raw" RSD (i.e. without the land cover filter) through different metrics including the slope (left), R-squared correlation (middle) and relative error (right) for both burned area (top) and the number of fires (bottom) over a range of individual fire size thresholds (1 to 500 ha). Dashed lines indicate a perfect fit between RSD and AGD.

There some long and complex sentences across the manuscript that do the reading and understanding of the text challenging. As a result, final proofreading would help solve these issues.

In response to this suggestion, we have rewritten some sentences to improve the readability. Please see below:

L 11: Although very promising, these datasets still lack a quantitative estimate of their accuracy with respect to ground-based fire datasets.

L 22: Overall, our findings suggest a reasonable agreement between RSD and AGD for fires larger than 100 ha, but care is needed when examining smaller fires at regional scales.

L 25: Over the past four decades, there were an average of 47,766 fires annually and an average of 413,209 hectares burned annually (San-Miguel-Ayanz et al., 2017) causing extensive economic and ecological losses, and even human casualties (Keeley et al., 2011; Molina-Terrén et al., 2019).

L 29: The Mediterranean fire regime is dominated by human-caused ignitions (Ganteaume et al., 2013) with most of the total burned area (BA) linked to a limited number of large fires during the summer (Turco et al., 2016).

L 34: Projecting future changes to fire activity requires modeling efforts across broad geographical scales to better understand processes and mechanisms conductive to fire ignition and spread.

L 81: Table 1. Description of regional fire agencies and reference link to the data used to build the ground-based dataset across Southwest Mediterranean basin.

L 89: The harmonized database contained 95,561 fires including only events that required a firefighting response (i.e., disregarding agricultural and prescribed fires) (see Fig. 1).

L 100: Fires were individualized from different algorithms such as a progression-based algorithm (Andela et al., 2019), a flood-fill algorithm (Laurent et al., 2018), and data mining (Artés et al., 2019) that share a common objective: assemble burned pixels that were adjacent in both space and time to identify and outline individual fire.

L 125: To account for the land cover changes over the study period, we used CLC 2006 as a reference to filter RSD from the 2005-2009 period and CLC 2012 from 2010-2015. Sensitivity analysis to the land-cover filter is shown in the Appendix A (Fig. A2).

L 146: We then sought to examine how the agreement between RSD and AGD datasets varies across space.

L 155: Fig. 3 shows that RSD follow a similar variability in terms of monthly BA but systematically underestimate BA and NF with respect to AGD.

L 166: Table 3 presents the total BA and NF as well as monthly (i.e. including the seasonal cycle) and annual correlation (i.e. excluding the seasonal cycle) for all fires (>1 ha).

L 169: Despite the fact that RSD underestimated the total BA by 38% and the NF by 96% for all fires, they reproduced almost perfectly the temporal variability on both monthly and annual basis.

L 176: The positive slope of the linear trends indicates that RSD generally underestimate both BA and NF when accounting for all fires (> 1 ha).

L 217: Our results demonstrate that agreement between RSD and AGD is strongly dependent on individual fire size.

L 235: This disparity relies on the complexity of extracting individual fires from gridded BA products. Environmental conditions (e.g. topography, cloud/smoke cover) may influence the sensor detection power, resulting in a break in BA continuity thereby increasing the risk of artificially splitting single fires into different fire events.

L 244: The selection of an appropriate fire size threshold depends on the objectives of each analysis. However, in this study we can generally recommend a minimum size of 100 ha, which outstands as a change point in multiple statistics (Fig.6 to Fig.8), with the relative error sharply (dowdily) decreasing in both BA and NF above this threshold.

L 264: The harmonized AGD used here as ground-based reference is available at https://doi.org/10.5281/zenodo.3905040 (Galizia et al., 2020).

L 266: In this work, we built upon previous research and investigated the reliability of three RSD of individual fires (FA, FRY and GlobFire) over a range of fire size thresholds across the Southwestern Mediterranean basin.

L 267: Overall, RSD contain only a small fraction of the total number of fires documented by AGD. However, they capture reasonably well the temporal variability of fire activity across monthly and annual scales.

L 269: Despite the different methodologies used to reconstruct fire patches, all RSD performed similarly and were increasingly accurate when focusing on larger fires.