Reply to comments of RC1 and RC2

The authors wish to thank the anonymous reviewers for their dedication to improve the manuscript contents. Detailed answers to their comments follow:

RC1

Comment	Reply
Chuvieco et al. describe a a new 250m MODIS burned area product ("Fire_cci v5.0") produced under the auspices of ESA's Climate Change Initiative programme, which the authors note provides the highest spatial resolution among existing burned area (BA) data sets. The authors stated goal in generating this product is to "complement existing BA products generated from the 500m MODIS bands", which consist of NASA's current 500m MODIS MCD64A1 and previous MODIS MCD45A1 BA products, "as well as to improve detection rate of smaller burn patches." In stating this goal I suggest the authors recast the first objective as a means of attaining the second, for merely "complementing" existing operational remote sensing data sets without providing a material improvement is not a very useful undertaking in itself. An aspect of this new data set that is both interesting and concerning is that it is being made at a higher spatial resolution using a less-capable combination of bands (red + NIR) than are used in the existing 500m MCD64A1 and MCD45A1 global BA products. The authors claim that the better spatial resolution makes "this product theoretically more suitable to analyze spatial properties of burned patches than other existing global BA datasets", but in my opinion the authors have not given sufficient attention to the implications/cost of attaining this improved resolution. A source of confusion for me, and I imagine other readers as well, is that the current Collection 6 MODIS BA product seems to map significantly more burned area (11%) than this new product, with the added benefit of having a lower omission error ratio (0.622 versus 0.708) and a much lower commission error ratio (0.353 versus 0.512). Overall I believe the authors do not provide adequate guidance to the community of modelers they seek to serve. With a reported global user's accuracy of just 49%, modelers should be cautioned that less than half of the 250m pixels classified as burned in this data set actually burned according to the authors' own reference data. It would be	The objective of the paper is to present a new global BA product, developed under the ESA CCI programme. The goal for generating the product was to complement existing ones, hopefully to improve them. The analysis of strengths and limitations of our product was not our main objective when sending this paper, and will require diverse testing and the involvement of those that are currently using it. We just wanted to present how the product was generated, its estimated accuracy and relations to existing products. We consider ESSD a good journal for this purpose. On the other hand, we disagree with the statement that creating new products is not relevant if they do not show clear improvements over existing ones. Several papers providing intercomparison between existing global BA datasets have shown that even if one of them provide a higher accuracy at global scale, may have higher errors in particular periods or areas (Humber et al. 2018). Therefore, having several BA products derived from independent sources (either in terms of input datasets or BA algorithms) actually helps the modelling community and reduces uncertainties in global BA estimation. Having a small pixel size provides opportunities to map smaller fires, which is compatible with having higher global errors (because of whole patches not detected or imperfectly delimited, for instance because of having lower fire intensity). We are currently working on an improvement of this algorithm to make it more sensitive to low intense fires. With the spatial approach used for quantifying the spatio-temporal variation of accuracy, it is difficult to analyse the impact of fire patch in actual errors. We should have a dedicated sample to analyse this aspect, but this exceeds the current goals of the paper. This quantification is part of our current research. Actually, we already tried to publish it but we faced problems in the attribution of size to those patches adjacent to no-data areas caused by cloud coverage,

immensely helpful if the authors could quantify this accuracy as a function of patch size. How else are modelers to know if the claimed improvement in the information provided about small patches is legitimate?	sensor failures or the end of the reference data scenes. The section on fire patch size was produced by fire modellers, not by Earth observation scientists, and this analysis is the main basis for claiming that our initial hypothesis – subject to further validation efforts- is certainly in the good direction.
The fire patch analysis is interesting, for here is where a 250m BA product would have something new to offer, but as the manuscript currently stands those results are interpreted exclusively as if the new product is detecting small genuine burned patches as opposed to detecting noise. To properly interpret this analysis the authors must also assess the accuracy of the Fire_cci v5.0 product specifically in this small-patch regime. Without this information it is unclear if the patch results reflect real fire behavior or are instead driven by commission errors in the product.	This analysis is a first approximation to quantify the potentials of fire patch analysis with the new product. It is based on comparison with existing products and reference fire perimeters obtained from national sources. The authors of this section of the paper presents the first results of a dimension that is quite relevant for global modelling of fires. They are in fact fire modellers working on the incorporation of fire patch information to dynamic global vegetation models. They are quite satisfied with the results provided by our product in relation to existing global BA datasets, as explained in the text.
I believe a description of the algorithm used to make the Fire_cci v5.0 product is the one published in Remote Sensing last year (Ramo and Chuvieco, 2017), but this source is oddly not cited.	This is not correct. The algorithm we used here was based on the one developed for the MERIS Fire_cci v4.1 product, which was published by Alonso and Chuvieco, 2015, RSE. Ramo and Chuvieco's paper refers to a Random Forest Algorithm, which was tested in a few global sites with 500 m MODIS data. Now it has been cited in the validation section, as an example of using RF for BA mapping.
P4 L25: It isn't clear how the 5x5 window is related to the 16 250m pixels within each 1000m hotspot.	We located the most likely burning pixel within a window around the central coordinate of the HS by selecting the cloud-free pixel with the lowest NIR value. We choose 5x5 instead of 4x4 to mitigate the impact of reprojection displacements or the impact of off-nadir HS of the MCD14ML product.
P4 L25: 1000m is the size of a MODIS hotspot at nadir. Is the larger size of the non-nadir hotspots taken into consideration?	Yes, we consider it by taken into account a larger window than the estimated 1x1 km.
P6 L2: Imagery from which Landsat platforms were used over the 2003–2014 validation period?	Landsat 5, 7 and 8. Now it is specified in the text.
P6 L4: Which semi-automatic classification algorithm was used to map burned patches in the Landsat scenes?	It was based on the Random Forest Classifier, and further visually verified. Now it is explained in the text.
P6 L7: GFED was used to guide the BA stratification, but this seems problematic because GFED BA was made from the Collection 5.1 MCD64A1 product, which the authors report is highly correlated with the	Our goal was to validate the MODIS Fire_cci v5.0 product, not the ones related to GFED. We used MCD64A1 c5 as a surrogate to global BA distribution.

Collection 6 product (Table 2). Is it reasonable to stratify validation reference data using the data that are to be validated? This is essentially what the authors have done. It seems this could open a door for considerable bias. What if, for example, GFED does not detect any of the true burning in a particular region?	It is true that if MCD64A1 c5 did not detect any of the true burning in a particular region, the resulting sampling size on that region would be small. However, we stablished a minimum sampling size for any stratum, even for deserts. The MCD64A1 c5 BA was used as a surrogate of the variance of the variable we want to infer. As the surrogate quantity is less proportional to the variance, the sample size is less optimal, and the sampling allocation lead to less precise accuracy estimates. This issue is common in sampling allocation analyses and it was described by Hansen et al. (1946), on a survey of business sales. They highlighted that such errors would not introduce bias into the sample estimates, but would diminish the variance reduction achieved by the chosen sample allocation.
P6 L16: Change “that” to “than”.	Done
P7 L14: The rationale for and applicability of the seemingly arbitrary 107 hectares threshold is unclear. On what basis was this threshold selected?	This threshold was selected following the analysis from Laurent et al. (2018). Their dataset provides all fire patches bigger than 5 pixels for MCD64 (corresponding to 5*463m*463m = 107ha), below which we considered the shape and other morphological patch features to be not realistic. We applied this cut because we wanted to compare the same size category of fire patches. We check the similarity between the two products for the same fire size category and not only the ability of MODIS Fire_cci to detect more fire patches. We modified the text to explain clearly the value and the origin of this size cut. However, we realize that using this threshold in Figure 10 hides the fact that MODIS Fire_cci detects significantly more small fires than MCD64. Therefore, we have redone the Figure and removed this threshold. We still discuss the numbers with and without the size cut in the text.
P7 L15: Why “at equator for the MCD64A1 product”? The pixels of the MCD64A1 product have the same area at all latitudes	For the fire patch analysis, we actually used the gridded MCD64 monthly product (tiff format). For this product, pixels do not have the same size at all latitudes (plate carrée projection). This was now clarified.
P8 section 3.3: It is essential to show a confusion matrix for each product as part of the product validation results.	They are now included in Table 2.
P8 L24: For consistency delete “significantly” since the DC difference between the Fire_cci v5.0 and both other products is comparable (0.117 versus 0.113).	Done, thank you.

P8 L23-32: The many accuracy metrics reported here could be much more clearly presented in a table.	They are now also reported in Figure 3.
P8 L23-32: I think it is worth reminding the reader that in comparing the various accuracy metrics here the Fire_cci c4.1 was only validated for the years 2005–2011 whereas the other products were validated for the years 2003–2014 (i.e. five additional years).	Done, thank you.
P8 L31: "From this point of view the MODIS Fire_cci c5.0 product is better equilibrated than the other BA products." This claim seems very weak because 1) the difference in relative bias among all three products is statistically insignificant and 2) the lower bias is the result of having better matched but comparatively high commission and omission error ratios. A product having a matched commission and omission error ratio of 0.99 would by this standard be perfectly "equilibrated".	We actually said that the relB was "slightly less than MCD64A1 c6, 0.415 (0.056), and lower than MERIS Fire_cci v4.1, 0.468 (0.094)". We did not claim that was significantly better, only that it had less bias than the other products. On the other hand, we disagree with the reviewer that bias is not relevant for accuracy assessment. It is obviously a complementary metric to overall accuracy and Oe and Ce, as it provides an interesting estimation of over or underestimation of BA. This was explicitly requested by climate users in our user requirement analysis.
P9 L1-7: This additional validation, which is not described in the methods section, and from which the authors suggest that their validation using 1200 Landsat reference scenes "may be considered a pessimistic estimation of accuracy", requires much more detail. The text cites Chuvieco et al. (2016) in reference to the national fire perimeters, but that source pre-dates the Fire_cci c5.0 product and I am wondering if Ramo and Chuvieco (2017) is perhaps more relevant. This more recent source describes an assessment using national fire perimeters in North Australia, Canada, and California for the years 2006–2008, presumably using the same reference data referred to here. (It is not clear why here the additional assessment is for 2008 only.) In any case, based on the Ramo and Chuvieco (2017) results, I believe the commission error of 0.23 quoted here, which the authors describe as "much lower", is misleading because this value is biased toward North Australia, where BA algorithms typically perform relatively well. In fact, Ramo and Chuvieco (2017) report significantly higher commission errors in Canada and California (0.32 and 0.41, respectively, for 2008) using what appears to be the same RF NIR model used to make the Fire_cci c5.0 product. Moreover, the commission errors were	We did not intend to use this alternative assessment dataset as a strict alternative validation, but only as a complementary estimation of errors. To avoid potential confusion to the readers, we have moved the comparison with national fire perimeters to the intercomparison section. We have included this analysis in the methods section, documenting the sources of those fire perimeters. We have included confusion matrices between our product and national fire perimeters for three additional years as suggested by the reviewer and a new confusion matrix has been calculated. As we say in the text, several papers have used fire perimeters as a first assessment of BA products. We have not intended to do it here as we include a full validation strategy. However, we consider the intercomparison with fire perimeters is still useful, as they provide a comparison with existing national fire databases, and include longer periods of time (full years). We have include a new table to show comparison results in the 3 sites and 4 years. Again, the MODIS Fire_cci v5.0 product is not based on Ramo and Chuvieco's algorithm (2017), neither our results are equivalent to those presented in that paper. The MODIS Fire_cci v5.0 product is based on

higher still (0.38 to 0.56) in both regions for the years 2006 and 2007.	Alonso and Chuvieco's algorithm (2015) and run on 250m MODIS data. Therefore, the errors reported here for the comparison sites are different from those included in Ramo and Chuvieco (2017), which were based on 500 m MODIS data.
P9 section 3.4: Did the comparison with Sentinel-2 BA reference data account for the presence of commission errors in the MODIS BA data sets? In addition, is there a reference for the Sentinel-2 BA data used in this analysis, or is this the first time those data have been used?	We did not use Sentinel-2 data as reference for validation. Only to compare with the fire patch database. We have added a reference to the Sentinel-2 data (Roteta et al. 2018): the paper is on review.
P10 L10: Is this shape index the same as that described at P7 L7?	In the FRY database, different complexity indices were computed. Among them was the so-called Shape Index (SI). We modified the text in page 7 and page 10 to make everything clearer.
Figure 4: Please make color scheme for the different products consistent with Figure 3. Please also plot the 1:1 regression line for comparison.	We have removed this figure, as it was not really adding much to the paper and may bias the reader to think that our main objective was to compare with existing products, while in fact it is to present a new product.
Figure 5: Should note in caption or text why the number of points differs so greatly between the three plots.	We have removed this figure because it provided similar information to what is included in table 3.
Figure 7: It would be very helpful to also show the commission error ratio in this figure.	The figure is included at the end of this document.
Figure 8: The pattern shown in the bottom map, whereby the 500m MODIS product seems to detect more burned patches in many areas, particularly outside the zones of very high fire activity, is surprising. One would think the higher resolution 250m product would more or less consistently map a larger number/density of burned patches.	As explained above, this is due to the fact that we had applied the 107 ha threshold on all survey. This drastically reduced the number of fire patches from the Fire_cci dataset. The figure has been redone without this cut, and now it shows that the Fire_cci product detects significantly more fire patches than MCD64 over the globe. The text has been changed accordingly.
It would be helpful to include a full resolution figure corresponding to one of the reference scenes that shows an example of the small burns detected in the Fire_cci v5.0 product that are missed in the MCD64A1 product.	The greater sensitivity to small patches was not derived from the validation dataset, but from the fire patch analysis. A single figure would not be significant to show the global effect described in the text.

RC2

Comment	Reply
The new global BA product presented in the manuscript is of great interest since it potentially provides time series of global BA maps at 250 m spatial resolution on monthly basis. This sort of information is certainly missing for many applications related to fire monitoring and climate modelling. As a general comment, I think the authors should better and quantitatively prove the contribution of the new product in detecting small burned areas, with respect to existing ones. The authors well review existing products but the inter-comparison and the assessment of the ability in detecting small burned patched is not well developed. This is a central issue since the comparison with other global products shows that, for example, MCD64A1 overall performs better than MODIS Fire cci. To this aim, the authors use BA from Sentinel 2 but little explanation is given and no figure and/or tables are provided for such important results. I think the revision of the manuscript should first focus on this topic. I list below my general and specific comments.	We hypothesised that working with higher spatial resolution would provide additional sensitivity of the final product to smaller fire sizes, but this was not the central message of this paper. We just want to present the technical details of how a new global BA product that is publicly available was generated. A full assessment of this product is beyond the scope of this paper and would require additional processing efforts. We just showed some examples of the potentials for climate users, particularly in the field of fire behaviour modelling. Our validation strategy was based on estimating the thematic accuracy of the product throughout time (first time the temporal stability of accuracy is measured), but our sample is not appropriate to analyse the impact of fire size on accuracy (since only the central part of the 1200 Landsat image pairs was processed). We are working on this issue, which would require an additional dedicated sampling, covering larger sampling sites. The comparison done with S-2 results was centred upon Africa, where a S-2 BA product is being produced. It is not yet ready (only the Northern Hemisphere when writing this paper), but it clearly showed better sensitivity to small fires than global BA products and therefore the comparison with this dataset was relevant. It was included in the fire patch analysis, but a more detailed comparison at continental level is beyond the scope of this paper.
1. The authors claim that one of their goals is to "improve the information available to climate modellers on spatio-temporal patterns of fire occurrence" (Page 10, Lines 26-27) and that "The goal of generating this product was to complement existing BA products (…) as well as to improve detection rate of small burn patches" (Page 3, Lines 14-19). Even if these objectives are realistic thanks to the use of full resolution red/NIR 250 m MODIS bands, in my opinion, little quantitative assessment of the above statements is shown in the validation.	You are right, this was our goal when generating the product, but it is not the main goal of this paper, which is to present the technical details of the product, validate it using a completely new validation dataset and intercompare with some of the existing BA products. We have rewritten the objective section to make this clearer. The ESSD journal is mostly interested on describing new datasets, and for this reason, we consider it is a suitable framework for our paper. Further analysis should check whether in fact the new product is more sensitive to small fires. The validation strategy we followed did not intend to test this, but the

	comparison with S-2 data shows the potentials of the product on this direction. In terms of complementing existing global BA products, the value clearly shown, as the validation and intercomparison analysis shows similar trends to widely used products in spite of being developed from a different dataset and with a different algorithm approach.
2. Further, what do the authors mean with “small burned patches”?	Those that are not currently included in global BA products (below 100 ha).
3. The monthly pixel BA product is least presented and validated. All figures and tables show BA annual synthesis which could "hide" monthly trends within the year by balancing out omission and commission from different months. In general, the temporal resolution (monthly and biweekly) of both products is neglected in presenting results and validation.	The monthly pixel product is the basis of the analysis included in sections 2.5 and 3.4. Therefore, the temporal resolution is fully included in the validation section. We did not aim to estimate how accuracy changes throughout different months. The pixel product is the source for the validation exercise. The Landsat pairs used for reference include a short period of time, which is extracted from the date of detection. The validation metrics include the temporal variation of accuracy, which has never been done before. Following the suggestions of the reviewer, we have included a new figure showing the seasonal patterns of burnings in Africa for a single year, as an example of the temporal information included in the product.
4. The analysis of the quantitative difference between the global BA products (Figure 4) could be integrated with metrics other than the correlation R2; for example, MAE and/or RMSE that better represent difference in quantitative assessment. For example, MODIS Fire cci is systematically lower (higher) than MCD64A1 c6 (GFED 4) and since they have the same trend, R2 is high but the amount of BA is different among the products. Moreover, MODIS thermal anomalies is a common source of information and this could explain the good agreement of temporal trends of the annual estimates (i.e. it is widely accepted that active fires/thermal anomalies are good indicators of fire seasonality). For what concerns section 3, I’d suggest the authors to present first the validation of the MODIS Fire cci BA product and then inter-comparison analysis.	We follow the suggestion of the reviewer and have started the analysis of the results by validation and then include intercomparison. We have removed figure 5 as it was not adding much new information.
5. Figure 6, to my understanding, shows that MCD64 is systematically better than MODIS fire cci, according to all metrics except relB. This is also confirmed at page 8, Lines 23-25, where the authors state that MODIS Fire cci performance, as measured by	Validation results showed that indeed global accuracy of MODIS MCD64A1 was higher than MODIS Fire_cci v5.0’s, but we could not decouple the contribution of fire size versus other factors, as validation analyses were not designed for this purpose.

the Dice coefficient, has lower accuracy than MODIS MCD64A1. If so, how do you explain it? If MODIS Fire cci is supposed to better represent small burned areas and to improve spatio-temporal representation of burned patches, where do the authors see the improvement with respect to MCD64? If the improvement brought by the new product is proved by the comparison with Sentinel 2 (Page 10, Lines 30-33) this topic deserves a central part in the manuscript. May be this point should be better addressed.	Our product is based on just NIR-R reflectances. Not using SWIR (as MCD64 does) probably implies loosing sensitivity for detecting savannas or crop fires. In any case, the reviewer should keep in mind that this product is the collection 6 of an algorithm that was first proposed in 2009 (significant changes were actually included in the last version). Since our product provides the first global BA product based on 250m, the assessment of our product would serve to improve its quality in further processing cycles. On the other hand, a recent paper by Humber et al. (2018) shows that lower average accuracy should not imply rejecting new products that complement existing ones. In fact, the great variety of fire conditions worldwide implies that even lower accurate products may provide higher detection accuracy for certain regions and periods (the example included by those authors include our MERIS Fire_cci v4.1 product). For this reason, it is important to have different global BA products, properly validated and assessed. Again this paper presents the product and offers to potential users the technical guidance to use the product in their own analysis. An example of the interest of developing new BA products is section 3.4 of this paper, where fire modellers are using the pixel product to better understand global distribution of fire shapes. Comparing our product with existing ones and with higher resolution products they showed that our product has higher sensitivity to detect small fires. The improvement brought by MODIS Fire_cci v5.0 is not proved by the comparison with Sentinel 2. This is rather a regional comparison of the S2 and MODIS Fire_cci v5.0. Note also that the Sentinel 2 sample used for this analysis only covers 9 months (Jan-Sep 2016). We think that the comparison with MODIS Fire_cci v5.0 brings some more information, but that it is not the main message of the manuscript.
6. Concerning the description of the methodology, there are some parts of the implemented approach for both BA product development and validation, that deserve better explanation.	The methods have been revised.
7. Compositing daily MODIS reflectance over the month is certainly a suitable approach;	Our method aimed not only to reduce the BRDF effect, but to select the pixel more

however, it should be better proved (or discussed) how the compositing criterion chosen by the authors does really reduce the BRDF effect.	clearly associated to burned signal in case it was affected by the fire. The monthly composite in fact reduces the temporal variability caused by the BRDF, while keeping the change of the burned signal (see an example in the annex of this document).
8. It is not fully clear to me how the authors derive the CDF for burned and unburned areas and how they define adaptive thresholds over the globe. Is this step carried out for each tile?	The processing unit of the algorithm is the MODIS standard tile and it is monthly processed. Therefore, all the thresholds are computed for each tile, which ensures the adaptiveness of the algorithm. The CDF for burned is derived from the hotspots present on the tile and the CDF for unburned is derived using the pixels that are farther than a 41x41 window from any hotspot.
9. In the validation section (2.3), the authors should specify how they derived yearly values of the metrics from the sampled validation units. Alternatively they should provide a valid link to the project's report so that the reader can access all the information.	This is now specified at the end of the section. Years are delimited by groups of strata. Therefore, same formulas applied for the whole population can be applied on each year. The full report is not yet ready. It will be in about 1 month, as it will include other sensors currently being processed. We have cited the web page of the project, where it will be soon available.
10. As I commented above, the role of Sentinel 2 data and BA Sentinel 2 product derived within the Fire cci project seems to be crucial. All information should be provided. What is the algorithm the authors mention on Page 7, Lines 5-6? Is there a reference for this product? This part is very unclear and results are poorly described and discussed. I'd suggest the authors either to clarify this point or to drop this comparison. But in the case of the latter, they should prove the contribution for detecting small burned areas.	We have added a reference to the Sentinel 2 product (Roteta et al. 2018): the paper is on review. As mentioned in the answer to comment 5, we don't think that the Sentinel 2 is crucial, due to its limited temporal and spatial overlap with MODIS Fire_cci. We have rewritten the section 2.5 to make it clearer. We think that the sensitivity to smaller fire patches is already well shown by the combination of Figure 7 (which has been redone) and Figure 8. Figure 7 shows that MODIS Fire_cci detects a higher number of fire patches, and that, at the same time, Figure 8 shows that the beta parameters of the fitted power law are globally higher than for MCD64, meaning that there is a highest relative amount of small fire patches for MODIS Fire_cci v5.0.
11. No results are shown on the fire date products which is one of the user's requirements and certainly a key information for fire regimes assessment.	The validation results are based on the detection date. Fire seasonality is included in the grid products as well. We have added Figure 2 as an example of fire seasonality in Africa.
12. The analysis of fire patches and fire shape is interesting. I wonder whether it is correct to calculate and compare the shape index between two products with different spatial resolution?	This issue is discussed in details in Laurent et al. 2018, where the authors stated that, due to different aliasing biases, it is hard to compare the Shape Index for two products with different resolution. This is why we did not directly show the differences between the two products, as we did for the number

	of patch or the parametrization of the power law distribution. Nevertheless, even if shape index cannot be directly compared between the two products, it shows that, within each product, more complex fire patches are always located in boreal regions and in Australia. We modified the text to specify this issue.
1. Page 4, Line 1. Cloud shadow masking is often critical when the objective is to detect low albedo surfaces. Does shadow masking had an effect on the detection rate of burned areas and burn date? Do the authors evaluate this issue?	At the beginning our intention was not to mask shadows, because we assume that this would lead to increase the omission error. However, we visually noticed that when shadows were not masked, too much noise and artefacts were introduced on the composites. On the other hand, we analysed in several test sites that masking had no overlaps with actual burns.
2. Page 6, Lines 15-24. Please be clearer for the less known metrics, such as the Dice coefficient, even if they are reported in Padilla et al. 2015; the authors could provide formula;	It is now provided in the last paragraph of Section 2.3.
3. Page 6, Line 24. Please do not refer to unpublished material (e.g. in preparation or submitted);	An additional clarification and reference is included at the end of Section 2.3. The unpublished material is a report of the Fire_cci Phase II, which will be made publicly available soon, by the end of the project this September 2018. We have included a reference to the project webpage.
4. For section 2.5 I'd suggest the authors to use a title that better describes the content of this section (i.e. fire size and/or fire shape/patches);	It was an example of the potential uses of the product, not a goal in itself
5. Page 7, Lines 14-15. It is not clear how the authors set a 107 ha threshold value;	The same question was asked by the other referee, so we include the same answer here. This threshold was selected following the analysis from Laurent et al. 2018. Their dataset provides all fire patches bigger than 5 pixels for MCD64 (corresponding to 5*463m*463m = 107ha). We applied this cut because we wanted to compare the same size category of fire patches. We check the similarity between the two products for the same fire size category and not only the ability of MODIS Fire_cci v5.0 to detect more fire patches. We modified the text to explain clearly the value and the origin of this size cut. However, we realize that using this threshold in Figure 10 hides the fact that MODIS Fire_cci v5.0 detects significantly more small fires than MCD64. Therefore, we have redone the Figure and removed this threshold. We still discuss the numbers with and without the size cut in the text.

6. Figure 4. I'd advice the authors to use the same range for the x- and y-axis, draw the 1:1 line and use the same colors as in Figure 3.	We have removed this figure.
7. Table 2. Could it be possible to provide all results and not only average and extreme values?	Done.
8. Figure 6, for the Oe and Ce graphs I'd suggest to use the same y-axis range;	Y-axes for Oe and Ce have now the same range.
Page 8, Line 26. "Global commission and omission error ratios" be consistent with description of accuracy metrics given in 2.3: are the metrics ratios or the errors as computed from the confusion matrix?	Now it is clarified in section 2.3 that accuracy measures are expressed as ratios (as opposite to areas). For example, we referred to them as "omission and commission error ratios".
10. Figures 1 and 2 should have highlighted the areas masked as unburnable. Moreover, why white regions (no burning) apparently cover different areas if they are from the same BA product (cell product)?	The unburnable areas are the same in both maps, but for computing the CV, only cells with burned areas in at least 50% of the years in the time series were considered. We have removed figure 2 and substituted it now for another one that presents the fire seasonality as suggested by the reviewer.
11. Page 10, Line 10. Shape index SI acronym should be introduced before.	We now introduce the Shape Index and its formula sooner in the text (section 2.5 Product assessment).
12. Page 3, Line 10. Replace "ENVISAR" with "ENVISAT"	Done.

References

Alonso-Canas, I., & Chuvieco, E. (2015). Global Burned Area Mapping from ENVISAT-MERIS data *Remote Sensing of Environment, 163*, 140-152.

Hansen, H.M., Hurwitz, W.N., & Gurney, M. (1946). Problems and Methods of the Sample Survey of Business. *Journal of the American Statistical Association, 41*, 173-189.

Humber, M.L., Boschetti, L., Giglio, L., & Justice, C.O. (2018). Spatial and temporal intercomparison of four global burned area products. *International Journal of Digital Earth*, 1-25.

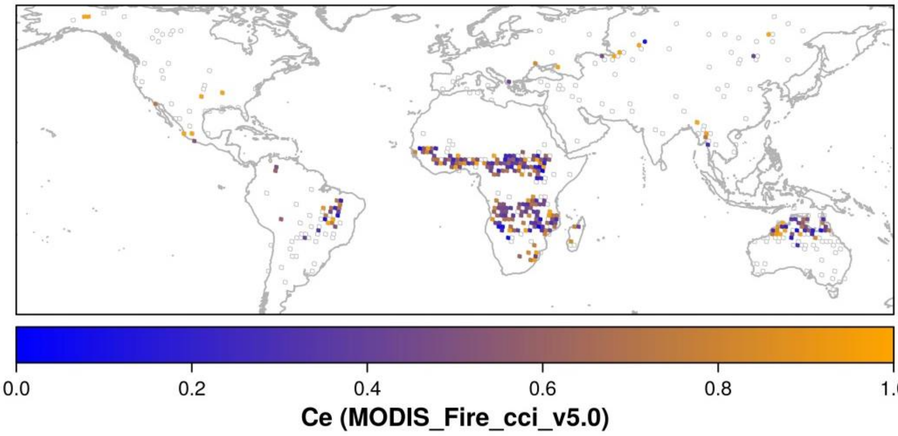
Laurent, P., Mouillot, F., Yue, C., Ciais, P., Moreno, M.V., & Nogueira, J.M.P. (2018). FRY, a global database of fire patch functional traits derived from space-borne burned area products. *Scientific Data, 5*, 180132.

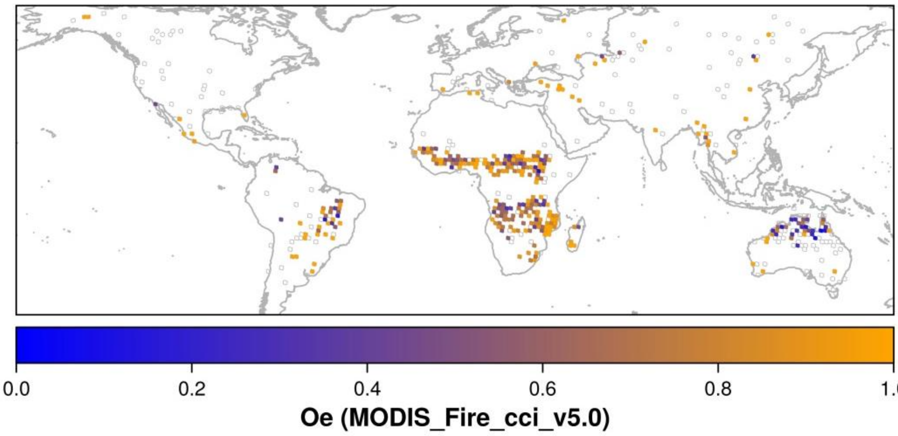
Ramo, R., & Chuvieco, E. (2017). Developing a Random Forest Algorithm for MODIS Global Burned Area Classification. *Remote Sensing, 9*, 1193.

Roteta, E., Bastarrika, A., Storm, T., & Chuvieco, E. (2018). Development of a Sentinel-2 burned area algorithm: generation of a small fire database for northern hemisphere tropical Africa *Remote Sensing of Environment, (in review)*.

Annex

1. Distribution of Oe and Ce for the MODIS Fire_cci product





2. Example of BRDF impact of the daily images and the monthly composite values for the same pixels.