



Comment on **essd-2020-399**

Anonymous Referee #2

Referee comment on "Operational implementation of the burned area component of the Copernicus Climate Change Service: from MODIS 250m to OLCI 300m data" by Joshua Lizundia-Loiola et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-399-RC2>, 2021

General Comments:

This manuscript presents the methods and results for creating an operational Sentinel-3-(OLCI) and MODIS-derived burned area product created under the Climate Change Initiative and Copernicus Climate Change Service. The algorithm is an adaptation of the MODIS-based FireCCI51 product. The manuscript's organization is logical, but it requires minor proofreading throughout for both grammar and typography (e.g., consistency with units like km²) that goes beyond the scope of peer review. In general, many methods in the paper refer to external works that reduce the transparency of this work. Particularly, the discussion of the validation methods is far too vague to be understood by the reader, especially considering the study does not implement the short sampling units of the work that it cites. Additionally, the lack of discussion of the dataset's limitations is problematic for a data descriptor manuscript, especially for burned area products that typically demonstrate high omission error rates.

There are several major criticisms of this work as a whole. First and foremost, the C3SBA10 product is not an improvement over its predecessor when considering the accuracy metrics presented by the authors (Dice coefficient is lower in all cases, commission errors are slightly better or the same, omission errors are higher in all cases, and relative bias is higher in all cases). Simultaneously, its dependence on the MODIS sensor means that the product cannot be generated any further into the future than the FireCCI51 product that it replaces. Thus, when the MODIS sensors are decommissioned ca. 2023, C3SBA10 will no longer be functional in the proposed form (and presumably be replaced by another version of the product using a different active fire data source, that will also be submitted for peer review in another manuscript). So, the authors' assertion on line 98 that this work will "ensure that multi-decadal analyses can benefit from both datasets uninterrupted" is untrue and unrealistic as C3SBA10 can never be multi-decadal. C3SBA10, therefore, does not improve the quality of the available burned area data, nor does it extend the length of the burned area archive, making its value very limited.

The presentation of the results and validation raise many flags as well. The authors present the relationship between FireCCI51 and C3SBA10 as a function of the linear regression's slope. The reader is left to assume that the simple linear regression (ordinary

least squares) was used – this is not the correct method because OLS regression assumes the independent variable is free of errors, and the choice of axis will change the result (i.e., making C3SBA the independent variable and FireCCI51 the dependent variable will yield a completely different result). A Deming regression, like total least squares, is more appropriate as it does not assume an error-free independent variable, nor does it assume variable dependence.

The comparison of RMSE and regression slopes are in a gray area given that the products are created using essentially the same algorithm and data that is as similar as possible. Both results could be completely wrong, for the same reason, and show high agreement. Comparison to an independent data source would be more informative.

There is no acknowledgment of any of the proposed methods' limitations and resulting product throughout the work. For example, the dependence on MODIS is a substantial drawback of this method that cannot be easily remedied using alternative data sources (discussed at the end of this review). The C3SBA10 product itself shows lower accuracy than the FireCCI51 product that it replaces, but consistency comparisons are emphasized instead. In general, the fire community should aim to make more accurate products, not less accurate products, and the continued generation of products that perform worse than their predecessors serves only to erode users' confidence in the products. The authors have taken an incremental improvement and versioning approach with the previous FireCCI50 and FireCCI51 products. However, a key distinction is that the FireCCI50 method was novel, and FireCCI51 made a material change to the algorithm and showed an improvement in accuracy – C3SBA10, on the other hand, does not use a novel method and does not improve on previous results. If the C3SBA10 product did improve (significantly) upon the accuracy of the FireCCI51 product, or implemented an active fire product with potential for long-term use, the manuscript would be more compelling.

Specific Comments:

The manuscript's introduction includes an overview of the FireCCI, C3S, and ECV's that goes far beyond what is relevant to the reader – a simpler summary and reference would suffice. Similarly, the discussion of technical specifications of OLCI in section 2.2.1 is rather extensive, but the sensor specifications are not discussed in the context of the results. Even though the results section attributes any discrepancies to sensor differences, little effort is made to identify what properties are responsible for specific discrepancies.

In section 2.4, the authors mention that the product is produced in 10x10 degree tiles like the predecessor product. A recent paper by Liu and Crowley (2021) identifies what were described as "severe tiling artifacts" in the FireCCI51 product. Is this also the case in the C3SBA10 product?

Section 2.6.2 notes that active fire data was used to assess the temporal accuracy of the product. The reader would assume this means the MCD14ML product – but the manuscript does not establish how this can be considered independent of the burned area product generation itself, which is very important given that MCD14ML is a direct input into the burned area product. Why was an independent sensor like VIIRS or SLSTR not used to avoid this issue altogether? Noteworthy – the method that the authors replicate (Boschetti et al., 2010) for the temporal accuracy assessment was designed for the MCD45A1 burned area product that did not use active fire detections as an input.

In section 2.7, the authors note that "C3SBA10 operational product cannot be understood as a unique, independent dataset, but as a continuation of its predecessor FireCCI51" – why not compare to an independent product like MCD64A1?

Section 3 – A 3 year-trend is not a significant trend, as noted by the authors; why not just call this the summary of the annual burned area? With no consideration for fire seasonality, how can the authors be certain that the “trend” is meaningful rather than a slightly shifted fire season in Africa from December through January?

In line 343, it is noted that almost all burned area in tropical forests occurred between 20 N and 20 S – this is, by definition, the only latitude band where tropical forests exist, so this observation is meaningless.

Section 3.2 (and subsequent relevant sections) – The manuscript barely discusses the causes of errors in the product and spends significantly more time on the consistency assessment (comparing the algorithm to itself) than on accuracy assessment (comparing to independent data). The emphasis placed on consistency assessment, combined with the “long units” accuracy assessment, is part of the broader pattern of de-emphasizing any negative aspect of the work and re-framing it in a positive light. There should be a critical and realistic self-evaluation of the product so that users can understand its limitations.

In line 382, do the authors mean “significantly higher for 2019” in the context of statistical significance?

In lines 383-384, the authors state that “more images due to the presence of two S3 satellites can explain why 2019 was the most similar year,” but do not offer any proof for the statement. The algorithm could be run with only S3A as a comparison to test the assertion. Additionally, while the result may be more similar to the FireCCI51 product, it was also the worst of the three years for commission and omission error – why does doubling the number of observations lead to a worse result?

Section 3.2.2. – 10 days is an extremely long detection window following active fire detection. The Boschetti et al. (2010) paper, which the authors referenced in the manuscript, found that the MCD45A1 product was vastly superior in preserving fire timing more than ten years ago (75% of burn dates were within 4 days of the active fire detection, vs. 64.6% within 5 days for the present manuscript). Given that this is a dataset description manuscript, users would benefit from an explanation of why the burn detection dates are so far behind the actual day of burning.

Section 3.3 – It seems as though the underestimation of C3SBA10, relative to FireCCI51, is being “obscured” in how statistics are presented here. For example, the 0.65, 0.56, and 0.28 Mkm² differences noted in the manuscript correspond to differences of approximately 17%, 15.5%, and 8%, respectively (percentages not included in the manuscript). These are large deficits, regardless of RMSE, regression coefficients, and slope, making it hard to believe that the results truly are consistent. The use of RMSE reported in square kilometers does not make sense because the analysis grid is in degrees. The maximum error at the poles approaches zero, while the maximum error at the Equator is approximately 12,000 sq km.

Line 452 – I don’t think the word “traduced” is being used appropriately.

Lines 495-496 – The discussion of long vs. short units highlights why the use of long units is inappropriate. The manuscript states: “The short units’ approach is more affected by the temporal reporting accuracy of the global BA products than when using long units.” This shows that the method is fitting the validation data and methods to match the result. At a fundamental level, there is no certainty that the fire observed in the long sampling unit is the same fire that was observed in the burned area dataset, given that the fire observations can be up to a year apart (as described in 2.6.1). This is especially the case in very fire-prone environments like African Savannas that may burn more than once per year, and minimizing the time between validation scene observations is done to avoid

misrepresenting these errors. By the authors' own writing, the validation method was purposefully constructed to accommodate the temporal errors in the dataset (i.e., recast incorrect classifications as correct classifications).

Lines 524-528 – The discussion about human impacts on fires does not seem relevant to this work.

Lines 529-537 – This paragraph's thesis states that sensor characteristics affect the products, but then the following sentences don't support that thesis concretely. The authors should provide evidence for causation here, the discussion of croplands and shrublands is anecdotal.

Discussion – If this work's goal is to continue the time series for multi-decadal analysis, it makes no sense to rely on the same sensor used in the predecessor work because the time series will effectively be ended once MODIS is decommissioned. The authors note that VIIRS could be a replacement, but this is only partially true because VIIRS does not have a morning overpass, so the effective temporal resolution is worse than MODIS. The authors reference the SLSTR algorithm (Xu et al., 2020) as having good capabilities for small fires, but the algorithm they referenced is currently nighttime-only. SLSTR is poorly suited for daytime detections because of the mid-infrared bands' saturation over surfaces hotter than 38C – is there any indication from the SLSTR algorithm developers that a viable/effective daytime algorithm will be available soon and if so, how would that affect the implementation of the present algorithm? Given both the limitations and longevity of VIIRS and SLSTR, it would make sense for this work to implement or test either of those sensors as they represent the only paths forward in a post-MODIS era. There are, therefore, significant challenges associated with this product in the near future that need to be addressed.

References:

- Boschetti, L., Roy, D. P., Justice, C. O., & Giglio, L. (2010). Global assessment of the temporal reporting accuracy and precision of the MODIS burned area product. *International Journal of Wildland Fire*, 19(6), 705-709.
- Liu, T., & Crowley, M. A. (2021). Detection and impacts of tiling artifacts in MODIS burned area classification. *IOP SciNotes*, 2(1), 014003.
- Xu, W., Wooster, M. J., He, J., & Zhang, T. (2020). First study of Sentinel-3 SLSTR active fire detection and FRP retrieval: Night-time algorithm enhancements and global intercomparison to MODIS and VIIRS AF products. *Remote Sensing of Environment*, 248, 111947.