General comments

1. This manuscript presents GlobMap FFP v1.0, a global 30 m dataset of forest burned area patches derived from the Landsat archive for the period 1984–2022. Although the technical effort required to process the full Landsat archive at global scale is substantial, the study is built on a fundamentally flawed premise: Landsat's temporal resolution is insufficient for reliable global burned area mapping. Burned area is a phenomenon characterized by short-lived spectral signals, often detectable for only a few weeks to a few months, and strongly dependent on observation timing and cloud-free conditions. A sensor with a nominal 16-day revisit cycle, further reduced by cloud cover and data gaps, cannot consistently observe this phenomenon at global scale, particularly in tropical and temperate forests.

The manuscript implicitly acknowledges this limitation by adopting multi-year temporal compositing, typically using five-year windows (and even longer periods prior to 2000). However, this strategy does not solve the underlying problem. Aggregating observations over five-year intervals inevitably suppresses short-lived burn signals and biases detection toward high-severity or slowly recovering fires. As a result, large fractions of real burned area are missed, especially in regions with frequent low-intensity fires or rapid vegetation recovery. An additional unresolved issue concerns areas that burn multiple times within the same compositing window. If I understand the methodology correctly, how are repeated fire occurrences affecting the same pixel over a five-year period handled, and how is fire recurrence represented when only a single observation per pixel is retained in the composite?

Response:

We thank the reviewer for these important comments. We agree that the relatively low temporal sampling frequency of Landsat, together with cloud contamination and uneven observation availability, presents significant challenges for global fire reconstruction, particularly in ecosystems where burn signals are short-lived and vegetation recovery is rapid.

The approximately five-year compositing interval was adopted as a compromise between temporal specificity and observation completeness, rather than to fully eliminate the constraints. Shorter intervals would provide greater temporal precision but would substantially reduce observation availability in many parts of the globe, leading to fragmented spatial coverage and reduced consistency across the time series. Longer intervals would increase observation availability but increase temporal aggregation and the likelihood of merging distinct fire events. The selected interval was therefore designed to balance these competing considerations while maintaining computational feasibility at the global scale.

Regarding repeated burning, as only one observation is retained for each pixel within a compositing interval, multiple fire occurrences affecting the same location during the same interval are be independently reconstructed. In practice, the composite preserves the observation associated with the strongest retained burn signal, and a single burned year is assigned to that pixel. Consequently, fire recurrence within a compositing interval is not explicitly represented. We have clarified this limitation in **Lines 509 – 513 of** the **Discussion**.

We acknowledge that multi-year compositing inevitably suppresses short-lived burn signals. This is an inherent trade-off when using medium-resolution optical sensors to achieve global, gap-reduced time series at decadal scales. We do not consider this a methodological flaw of compositing, but rather a constraint that defines the appropriate interpretation of the dataset, which we have clarified as a spatially explicit fire patch product rather than a complete burned area inventory.

To facilitate appropriate interpretation, we emphasize throughout the revised manuscript that GlobMap FFP should be regarded as a spatially explicit fire patch dataset rather than a complete inventory of all burned area occurrences. Its primary value lies in preserving patch geometry, spatial organization, and long-term fire patch dynamics at 30 m resolution across nearly four decades, while the limitations associated with temporal aggregation and repeated burning are now more explicitly acknowledged.

2. A critical consequence of this approach is evident in the reported burned area estimates. The dataset reports mean global forest burned area values of approximately 7.3 Mha yr^{-1} for the period 2001–2021, whereas established global burned area products report values close to 19–20 Mha yr^{-1} over the same period. This represents an underestimation by a factor of roughly three.

This strong underestimation is particularly pronounced in tropical regions, where the manuscript itself reports the largest divergences relative to existing products. These are precisely the regions where burned signals are short-lived, cloud cover is persistent, and Landsat's sparse clear-sky observations are least capable of capturing fire effects. Therefore, I find it very difficult to reconcile the large disparity in burned area between the GlobMap product and MODIS with the validation metrics reported in the manuscript. According to Table 2, omission errors (24%) and commission errors (13%) are relatively low, the Dice coefficient is high (0.82), and the reported relative bias is relatively small (−11%). These values would normally indicate a product with only moderate underestimation. However, the intercomparison shows that GlobMap detects substantially less burned area (Fig. 7; MODIS ≈ 19.6 Mha yr^{-1} versus GlobMap ≈ 7.3 Mha yr^{-1}). If such a large and systematic discrepancy truly exists, it is difficult to explain how it could coexist with a low relative bias and only moderate omission errors. This inconsistency is not resolved in the manuscript and fundamentally undermines the credibility of the validation results.

Response:

We thank the reviewer for this insightful comment. We agree that the apparent inconsistency between the relatively good agreement metrics reported in Table 2 and the substantially lower burned area estimates relative to MCD64A1 requires clarification.

First, we acknowledge that the omission errors, commission errors, Dice coefficients, and relative bias reported in Table 2 should not be interpreted as measures of burned area completeness. These metrics were derived from sampled Landsat scenes and quantify agreement only for burned scars remained detectable in the available Landsat observations. Consequently, they characterize classification agreement conditional on burn detectability rather than the completeness of burned area reconstruction. Fires missed because of limited observation availability, persistent cloud cover, rapid post-fire vegetation recovery, or compositing effects are not represented in these statistics. Therefore, relatively good agreement within the evaluated samples does not necessarily imply close correspondence in cumulative burned area estimates when products are aggregated across regions and decades. The relative bias reported in Table 2 reflects agreement within the sampled locations and is not directly comparable to differences in total burned area between GlobMap FFP and MCD64A1. To clarify, we have added the following sentences in **Lines 260 – 262, Lines 321 – 324**:

"*Because both datasets were derived from Landsat imagery and relied on related burned area detection procedures, these metrics should be interpreted as measures of internal consistency rather than fully independent estimates of product accuracy.*"

"*Yet, it is worth noting that the reported omission and commission errors characterize agreement between GlobMap FFP and the Landsat-based reference samples within evaluated locations where burned scars remained detectable in the available observations. These metrics quantify classification agreement conditional on burn detectability rather than the completeness of burned area reconstruction at regional or global scales.*"

Second, we agree that the lower burned area estimates of GlobMap FFP relative to MCD64A1 warrant further discussion. In the original manuscript, we primarily attributed these differences to the coarser spatial resolution of MODIS. We acknowledge that this explanation alone is insufficient. In the revised manuscript, we now emphasize that the discrepancy likely reflects the combined effects of multiple factors, including incomplete detection of short-lived fire signals under heterogeneous Landsat observation conditions, particularly in cloud-prone and rapidly recovering tropical forests; differences in image compositing and fire patch delineation procedures among products; and limited representation of repeated burning within compositing intervals. We have expanded the **Discussion (Lines 501–514)** to explicitly address these sources of discrepancy and no longer attribute the differences solely to spatial-resolution effects.

Finally, we have revised the positioning of both the manuscript and the product to avoid misunderstanding. The revised manuscript no longer presents GlobMap FFP as a complete global burned area inventory. Instead, the primary objective of the dataset is to provide a consistent, spatially explicit characterization of forest fire patches at 30 m resolution over nearly four decades. Its principal value lies in enabling analyses of fire patch geometry, spatial organization, patch size distributions, and long-term changes in fire patch structure, while the limitations and uncertainty sources associated with burned area completeness are now discussed more explicitly throughout the manuscript.

3. The validation framework itself further limits the interpretability of the reported accuracy metrics. Although the authors state that they follow the spatial sampling framework of the Burned Area Reference Database (BARD) by selecting Landsat TSAs, they do not use the BARD reference perimeters. Instead, reference burned area within each TSA is generated independently using the same algorithmic approach employed for training sample generation and product development. As a result, the validation does not rely on independent reference data, but rather evaluates internal methodological consistency.

Moreover, while BARD TSAs are part of a global, stratified sampling design explicitly constructed to support statistically rigorous accuracy assessment and uncertainty estimation, this sampling design is not adopted in the present study. The authors do not implement a probability-based sampling scheme adapted to their analysis, nor do they report uncertainty measures (e.g. confidence intervals or standard errors) for the accuracy metrics. Consequently, the reported omission and commission errors, Dice coefficients, and relative bias cannot be interpreted as statistically robust estimates of real-world burned area detection performance.

In addition, the validation methodology lacks essential temporal detail. The manuscript does not clearly specify which years or portions of the 1984–2022 period are actually covered by the validation, nor whether the reported accuracy metrics are representative of the entire time series. It remains unclear whether the validation is dominated by periods with higher observation density (e.g. the Landsat 7 and Landsat 8 eras) or whether earlier periods, characterized by sparser data availability and longer compositing windows, are adequately represented. Furthermore, the additional filtering of validation imagery to scenes with less than 40% cloud cover further reduces the number of usable observations, particularly in cloud-prone regions, compounding uncertainties regarding the representativeness and robustness of the reported metrics.

Response:

We thank the reviewer for this thoughtful and detailed comment. We agree that the assessment does not constitute a fully independent validation. Although the reference dataset was generated separately from the final product and based on independently selected sample locations, both datasets were derived from the Landsat archive and relied on related burned area interpretation procedures. Thus, they share common sources of uncertainty, and the reported metrics should not be interpreted as fully independent estimates of burned area detection accuracy. To avoid overstating the level of independence, we have revised the terminology throughout the manuscript and now describe the assessment as a Landsat-based evaluation of product agreement and internal consistency rather than an independent validation.

We also agree that the use of BARD TSA locations should not be interpreted as adoption of the full BARD validation framework. Here we used the TSA locations to achieve broad geographic coverage across major forest regions. However, we did not used the BARD reference perimeters themselves, or adopt the probability-based sampling design developed for statistically rigorous global accuracy estimation, so that formal uncertainty estimation was not implemented. To avoid misunderstanding, we have clarified this in **Lines 248 – 249** of the **Methods** as below:

"*To evaluate the performance of GlobMap FFP, we constructed a Landsat-derived reference dataset using the TSA units selected to achieve broad geographic coverage following the spatial distribution of the burned area reference database (BARD) (Franquesa et al., 2020).*"

Regarding the temporal representativeness of the evaluation dataset, the sampled scenes were selected across the full study period (1984-2022) and include observations from the Landsat 5, Landsat 7, and Landsat 8 eras. Therefore, the assessment is not restricted to periods with higher observation density. Nevertheless, we acknowledge that the temporal distribution of available Landsat observations is inherently uneven. Earlier portions of the Landsat archive generally contain fewer cloud-free observations and are associated with greater uncertainty than more recent periods. The requirement for cloud cover below 40% was introduced to improve the interpretability and consistency of the reference data, but may reduce the representation of highly cloud-prone regions in the evaluation dataset. We have clarified the temporal coverage of the reference samples in the **Methods (Lines 252 – 255)** as below:

"*In total, we sampled 74 Landsat TSA units across major forest biomes (Fig. 3). Then Landsat scenes with cloud cover below 40% in these units were sampled, spanning the full study period and encompassing observations from the Landsat 5, 7, and 8 archives, resulting in a total of 945 Landsat scenes. The cloud cover threshold was introduced to improve the interpretability and consistency of the reference data, but may reduce the representation of highly cloud-prone regions in the evaluation dataset.*"

4. An additional conceptual limitation concerns the interpretation of the mapped units as 'individual' fire events. In the proposed dataset, burned pixels are aggregated into patches based on spatial proximity and assignment to a single burned year, without explicit information on

ignition timing, fire duration, or intra-annual separation. Under this framework, independent fires occurring at different moments within the same year may be merged into a single patch, while the same fire spreading over extended periods may be inconsistently represented depending on observation availability. As a result, the mapped patches cannot be unambiguously interpreted as fire events. This has direct implications for the analysis of fire size distributions, fire frequency, and fire regime characteristics, and further limits the comparability of the dataset with products that explicitly track fire events using finer temporal information. The manuscript does not sufficiently clarify these limitations or their consequences for downstream analyses.

Response:

We agree that the mapped units should not be interpreted as individual fire events in the strict sense. Because fire patches are delineated using annual temporal attribution and spatial clustering, temporally distinct fires occurring within the same year may be merged into a single patch, while detailed information on ignition timing and fire progression is not retained. To avoid ambiguity, we have modified the writing throughout the manuscript to consistently describe the product as a forest fire patch dataset rather than an event-based fire inventory. We have also expanded the discussion of the implications of annual aggregation for fire frequency estimates, fire patch size, and downstream analyses as below in **Lines 513 – 516**:

"*Furthermore, the annual aggregation may merge temporally adjacent fires into a single fire patch, likely resulting in overestimated patch sizes and underestimated fire frequencies. Consequently, GlobMap FFP should be interpreted as a spatially explicit fire patch dataset rather than a complete inventory of all forest burned area, particularly in frequently burned tropical forests.*"

5. Overall, while Landsat-based burned area mapping can be highly effective at regional scales under appropriate conditions, this manuscript does not demonstrate that such approaches can be straightforwardly generalized to a global product without substantial loss of information. The strong underestimation of burned area, the internal inconsistency between area estimates and validation metrics, the lack of independent validation, and the inability to represent repeated burning indicate that the proposed dataset does not provide a reliable or improved representation of global forest burned area. In this context, it remains unclear what scientific or practical value a new dataset based on higher spatial resolution sensors offers if it does not demonstrably improve the representation of burned area relative to existing products.

Response:

We thank the reviewer for this important comment. We agree that, if GlobMap FFP were interpreted as a complete global burned area inventory, the limitations identified by the reviewer, including burned area underestimation, uncertainties in the evaluation framework, and the inability to fully represent repeated burning, would substantially constrain its utility. We also agree that the original manuscript did not sufficiently articulate the scientific objectives and intended applications of the dataset.

In response, we have revised the positioning of the product throughout the manuscript. The revised version no longer presents GlobMap FFP as a complete reconstruction of global forest burned area or as a replacement for existing burned area products. Instead, the primary objective of GlobMap FFP is to provide a consistent, spatially explicit characterization of forest fire patches at 30 m resolution across nearly four decades of Landsat observations.

The principal value of the dataset lies in its ability to preserve fine-scale fire patch geometry and spatial organization. Many existing global burned area products were developed primarily to quantify burned area extent and temporal dynamics. While highly valuable for monitoring burned area, their spatial resolution limits the representation of fire patch boundaries, internal heterogeneity, patch connectivity, and landscape-scale fire patterns. By contrast, GlobMap FFP was specifically designed to retain individual fire patch structure and associated attributes, enabling analyses of fire patch morphology, patch size distributions, spatial aggregation, landscape fragmentation, and long-term changes in fire regime characteristics across global forests.

We therefore emphasize that the scientific contribution of GlobMap FFP is not solely determined by its ability to reproduce burned area totals reported by existing products. Rather, it provides a complementary perspective on forest fire dynamics by explicitly representing the spatial structure of fire patches over a nearly four-decade period. This type of information is difficult to derive consistently from coarser resolution global products and is particularly relevant for studies of fire regime shifts, landscape ecological impacts of fire, spatial fire patterns, and fire patch scaling relationships.

To better reflect these objectives, we revised the **Introduction, Discussion, and Conclusions** sections to more clearly distinguish between burned area completeness and fire patch characterization. We also expanded the discussion of limitations and uncertainty sources associated with burned area reconstruction, while highlighting the specific applications for which the dataset was designed. We hope these revisions clarify the intended scope, limitations, and scientific value of GlobMap FFP.