The manuscript "CONFEX: A Database for CONUS Fire EXtent" presents an ambitious and valuable effort to build a wildfire perimeter dataset for CONUS and Alaska using VIIRS 375 m active fire detections. Providing a spatially comprehensive and temporally dynamic dataset is critical for fire management and ecological modeling.

However, after a review of both the manuscript and the released geospatial data, I found that the current version suffers from important methodological limitations, severe geospatial inaccuracies, and a lack of robust uncertainty quantification. While the conceptual pipeline is promising, its execution introduces artifacts and biases that compromise the dataset's scientific reliability. Furthermore, the manuscript frequently substitutes rigorous quantitative validation with speculative language.

In its current state, the dataset and the accompanying manuscript do not meet the rigorous "high-quality data" standards expected for publication in ESSD. I therefore recommend a major revision, and I detail my concerns below.

General Comments

- While the manuscript clearly outlines the conceptual steps of the data processing pipeline, it does not currently specify the programming languages, software, or core libraries used to implement the clustering and perimeter generation. Please add a brief statement outlining the primary software environment and tools utilized to generate the CONFEX database.

- A serious spatial-domain issue must be addressed. Although the manuscript repeatedly describes CONFEX as a dataset for CONUS and Alaska, the released product appears to contain fire-event polygons outside that stated domain. For example, I identified a polygon with year = 2023 and cluster_final = 85586 (centroid: 44.9835°N, 66.9877°W) that lies entirely in Canada, and another with year = 2013 and cluster_final = 41847 (centroid: 27.1205°N, 99.4487°W) that lies entirely in Mexico. In addition, the dataset appears to contain nearly 1,000 Hawaii polygons. These records suggest that the spatial filtering, study-area clipping, or release-version control is not fully consistent with the manuscript's stated domain and must be corrected or clarified.

- A major methodological concern is the apparent reliance on parameter tuning from a single region and year. The key DBSCAN thresholds used in CONFEX (2000 m in space and 48 h in time) were tuned using only the 2020 California FRAP dataset, yet these same settings were applied across all of CONUS and Alaska. Given the drastic differences in fire regimes, fuels, climate, and fire behavior across the full study domain, it is unclear why a California-based calibration should generalize well. The authors must justify this transferability explicitly or provide sensitivity tests showing that the selected parameters remain appropriate outside California.

- The manuscript states that EPSG:5070 (NAD83 / Conus Albers) was used as the unified processing CRS for both CONUS and Alaska to maintain "consistency". This justification is technically unsound. EPSG:5070 is an equal-area projection optimized for the contiguous United States; its accuracy degrades significantly when extended to high-latitude regions like Alaska. The authors should justify why this projection is acceptable for Alaska or provide sensitivity tests showing that the resulting distortions do not materially affect the perimeter area calculations.

- The treatment of low-detection events is unconvincing. In the README, the authors recommend the "CONFEX_ALL_YEARS_min3pts.zip" subset because it provides "more reliable perimeters", implying that events with fewer than three detections have greater geometric uncertainty. Yet, the manuscript does not quantify this uncertainty. This issue is compounded by Figure 3, which shows that DBSCAN was run with "min_samples = 1". Since "min_samples" includes the point itself, setting it to 1 is an extremely permissive choice that allows isolated detections to become standalone clusters rather than noise. The authors should justify this choice carefully and provide a direct uncertainty analysis for single- and two-point events, rather than merely acknowledging their lower reliability in the README.

- The dataset appears to contain at least some non-exclusive fire-event objects. For example, in "VIIRS_final_perims_ALL_YEARS.gpkg", cluster_final IDs 52266 (2022/06/19–2022/07/01) and 52424 (2022/06/20–2022/06/26) overlap both temporally and spatially, with approximately 27 km^2 of shared area. This overlap contradicts the premise of each cluster representing a distinct fire event and suggests that additional topological quality control is needed.

- L298-305: The paragraph discussing the poor performance in the Eastern CONUS suffers from severe analytical spin and a notable geographic error. By reporting 9,084 false positives (FP) and 1,916 true positives (TP), the algorithm yields a precision of merely 17.4%, while 2,703 false negatives (FN) equate to a recall of 41.5%. It represents a fundamental algorithm limitation in the Eastern CONUS, not a "complementary" relationship, and claiming that more than 9,000 FPs simply identify "substantially more fire activity" without robust ground validation is scientifically unsound. Additionally, framing the splitting of a single MTBS fire into multiple fires as a "feature" ignores the artificial fragmentation likely caused by the rigid 48-hour temporal threshold. Compounding these analytical issues is a stark geographic mismatch: the authors use the Soberanes Fire—which occurred in California (Western CONUS)—to justify algorithm behavior in an Eastern CONUS error analysis. I recommend that the authors objectively state that the current clustering pipeline performs poorly in the East rather than spinning a 17% precision score as a strength, and completely remove the logically disconnected Soberanes fire example from this Eastern CONUS discussion.

Specific Comments

- L74: The manuscript incorrectly characterizes FIRED as a strictly "regionally focused effort" for CONUS, overlooking that it has already been expanded into a global fire perimeter dataset via the FIREDpy algorithm (Mahood et al., 2022, Scientific Data). Please update this statement and citation.

- L100-152: The background section would be more effective if the discussion of the historical development of fire remote sensing were shortened and made more concise. At present, a substantial portion of the introduction reviews earlier satellite fire products and monitoring approaches, which is useful context but somewhat longer than necessary for a data description paper focused on presenting CONFEX. I suggest condensing this narrative and instead adding a summary table that compares the main existing datasets with CONFEX in terms of spatial coverage, temporal coverage, spatial resolution, event definition, and major strengths and limitations.

- L202: An important methodological consideration concerns the geometric implication of using the Chebyshev metric in the normalized spatiotemporal feature space. As described in the manuscript, this choice enforces independent hard thresholds in each dimension, which implies an axis-aligned square neighborhood in the spatial plane rather than a circular Euclidean neighborhood. Consequently, two hotspots can be linked even when their Euclidean separation exceeds 2000 m, provided that their separations along both coordinate axes remain within the 2000 m threshold, whereas two hotspots separated by slightly more than 2000 m along a single axis are forced into different clusters. This is a defensible design choice, but it has a clear physical interpretation that should be stated more explicitly. I therefore suggest that the authors either justify why a square spatial neighborhood is appropriate for this application, or provide a sensitivity comparison with an alternative formulation.

- L209-211: Applying a static 2016 NLCD layer uniformly to classify fires from 2012–2024 could introduce classification inconsistencies due to land cover change. The authors should consider using time-matched land cover epochs or discuss the uncertainty introduced by a single baseline year.

- Figure 3 is oversimplified in several important respects. The diagram does not clearly show the branching logic for 1–2 point, 3-point, and ≥4-point clusters, even though these cases are handled differently in the algorithm.

- L247: In the Parameter Tuning Section, the alpha-shape parameter of 1, 0.01, and 0.001 was checked, but there appears to be a geometric mismatch between these values and the 375 m nominal spatial resolution of the VIIRS detections. Based on the formulation where retained triangles must have a circumradius < 1/alpha, testing alpha = 1 and alpha = 0.01 translates to thresholds of 1 m and 100 m, respectively, which are geometrically highly restrictive given the 375 m pixel spacing and would likely reject almost all triangles. Consequently, the selection of alpha = 0.001 (a 1000 m threshold) might appear optimal primarily because the other two values fall well below the spatial limit of the sensor, rather than representing a comprehensive tuning of the algorithm's behavior.

- L273: The statement "it is also possible that these two groups overlap" is too tentative in this context. It would be preferable to verify this directly with a spatial analysis and report the result explicitly, rather than leaving the relationship between the two groups uncertain. Clarifying this point would make the validation discussion more precise and easier to follow.

- L280-282: The discussion of FRAP durations is helpful, but I think the role of FRAP in the validation framework should be clarified. If the authors conclude that FRAP durations are not representative of actual fire spread or acreage, it becomes somewhat unclear how this dataset can still serve as a benchmark for evaluating the temporal performance of CONFEX. I suggest that the authors explain more explicitly which aspects of FRAP remain suitable for comparison and which should be treated with caution.

- L287, 310-311: Please explain why the 2024 data for both MTBS and FRAP were excluded from the validation analyses, given that both are currently available and the CONFEX dataset extends to 2024.

- L337-338: The manuscript describes CONFEX as a wildfire database, but some of the detected large fire events in regions such as Florida may include prescribed burns. Because prescribed fires and wildfires can differ in their drivers and behavior, it would be helpful for the authors to clarify whether the current product represents wildfire only or a broader fire-event database.

- L359-360: The sentence stating that Alaskan fires are "spread across two different ecoregions of boreal and tundra forests" contains two errors. The term "tundra forests" is ecologically incorrect. By definition, the tundra biome is characterized by the absence of trees. Alaska comprises more than two Level I ecoregions. According to the standard Commission for Environmental Cooperation (CEC) framework, Alaska also includes

Marine West Coast Forests, and Northwestern Forested Mountains. Please rephrase to use accurate ecological terminology.

- L378-379: Applying a strict 48-hour "hard threshold" makes the algorithm highly vulnerable in regions known for prolonged cloudiness, such as Alaska. If a fire smolders under cloud cover for three days without a satellite pass, the rigid cutoff will artificially split one continuous fire into multiple separate events. I suggest exploring or discussing the feasibility of a dynamically adjusted temporal threshold to enhance robustness against missing data.

- Figures S1 and S2: Please redraw with appropriate y-axis limits. The tallest CONFEX bars are visibly truncated in Figure S1, preventing full inspection of the data.

- Supplement L67-70: The authors mention in S2 that events with < 3 detections have "associated uncertainty". This is a severe understatement. A single VIIRS 375m pixel covers approximately 35 acres. If an event has only 1 or 2 detections, its true physical footprint cannot exceed 100 acres. If such events are somehow bypassing the 500-acre or 1000-acre filters, it proves that the current algorithm is generating massive, physically impossible areas.

- Supplement L72-73: Throughout the manuscript and in Figure S5, the authors claim that CONFEX consistently detects more fire events than MTBS due to its high-resolution clustering. However, Figure S5 clearly shows a distinct reversal in Nevada (and to a lesser extent, Utah), where MTBS reports more >1000-acre fires than CONFEX. The authors should explicitly explain this discrepancy.

- Supplement L98: The authors use highly speculative language to explain the data discrepancies: "many of which may be near or just above and perhaps smaller, spatially fractured burns...". The authors have the shapefiles for both CONFEX and MTBS. They should perform a simple spatial intersection to quantitatively report exactly how many extra CONFEX polygons are fragments of a single MTBS fire, and how many are genuinely new ignitions.

- Supplement S1.2 and S2.2: The claim that CONFEX registers more events because its resolution "resolves more distinct events" is contradictory. Given the 48-hour temporal truncation issue identified in the main text (which fragmented 16 FRAP fires into 88 pieces), the inflated CONFEX counts are demonstrably driven by algorithmic over-fragmentation, not solely superior detection.

- Data Structure: The released shapefile attributes are difficult to interpret because several field names have been truncated to meet the 10-character limit of the shapefile format. For example, columns such as "centroid_l", and "centroid_1" are no longer self-explanatory, which increases the risk of misinterpretation. I suggest either shortening all exported field names to clear, unambiguous names within the shapefile limit, or preferably using GeoPackage as the primary distributed vector format, since it preserves full field names and avoids this limitation.

- Geospatial Artifacts: Visual inspection of the vector perimeters (e.g., cluster_final=79727 in 2017) reveals severe geometric artifacts. The perimeters exhibit highly fragmented boundaries and deep, orthogonal "stair-step" trenches. These artifacts appear to be caused by using a high-resolution 30 m raster mask (LANDFIRE) to hard-clip perimeters derived from 375 m VIIRS detections, imprinting the arbitrary 30m raster grid onto the vector dataset. The authors should discuss whether additional geometric smoothing or topological cleaning is warranted.

- References: The reference list appears to contain several entries that are not cited in the main text (e.g., Bright et al., 2019; Giglio and Kendall, 2001; Itten et al., 2008; Marks-Block and Tripp, 2021; Urbanski et al., 2018; van Marle et al., 2017; van Wees et al., 2022). All dataset and institutional citations must be reformatted to adhere to the strict ESSD data citation guidelines. Please add missing DOIs where available.

Technical Corrections

- L21: Replace "thirst" with a more formal academic term.
- L59: Replace "nearest infrared" with the standard remote-sensing term "near infrared".
- L71: Remove the redundant closing parenthesis after "2005-2017))".
- L152: "Figure 1" is a tabular representation and should be relabeled as "Table 1".
- L256: Replace the colloquial idiom "hit the sweet spot" with a formal expression.
- L275, 286, 292: Figures and tables should be cited in the main text in the order of their first appearance. At present, the manuscript refers to Figure 9a in Section 3.5 before earlier figures such as Figure 6 are introduced, and some display items, including Figure 5 and Table 4, do not appear to have clear in-text callouts in the relevant discussion. The authors should therefore revise the numbering and citation order so that all figures and tables are introduced sequentially and explicitly referenced at the appropriate places in the text.
- L325, 346: Use km^2 rather than "Km^2" throughout the manuscript.
- L337: Revise "Florida has highest number" to "Florida has the highest number".
- L364-365: Correct the typographical error by changing "reliant" to "reliable".
- L404: Replace the informal phrase "have-it-all" with a more formal academic term.
- Define abbreviations such as FRP, DBSCAN, HDBSCAN, MTBS, FRAP, and VIIRS at first mention.