> *We thank the reviewer for their time and efforts reviewing this manuscript. Their comments were very helpful in improving the paper. The comments were very helpful in determining where the issues were and we have made substantive efforts to improve the the paper.*
>
> *We disagree with the reviewer's thesis that the paper is premature. This addressed in a summary at the end.*
>
> *The reviewer's comments are provided in regular font – our responses are shown in italics. Revised sections of the manuscript are captured in quotes with starting line references in the revised manuscript included in square brackets.*

Anderson et al. set out to create a global version of the Canadian Forest Fire Danger Rating System. Overall, this is an interesting proof of concept; however, I believe the publication of this paper is premature. Please see specific comments below.

Major Comments

Section 3.2: Do the authors have to worry about double counting fires when all the VIIRS sensors are used? Was any preprocessing done on the fire data? Why did the authors not filter out the presumed non-vegetation fire using the type flag? The science quality dataset includes this information.

> *We only used one VIIRS sensor (S-NPP). We used hotspots from 2012-2019, during most of which there was only one VIIRS in orbit. In future, with multiple satellites, double- and multiple-counting will be handled by the "times burned" metric and methodology.*
>
> *Upon investigation, we have found that the type flag mentioned by the reviewer unfortunately is not very reliable. We cross-compared our own industrial sites mask for Canada, to the categories present in the type flag: that is, we have ground truthed data upon which to evaluate the type flag. We found that the hotspots associated with these known industrial sites were classified in the VIIRS type flag as a mixture of type 0 (vegetation fires), type 2 (static), and type 3 (offshore). Most of the hotspots from smaller industrial sites were categorized by the VIIRS type flag as vegetation fires. Meanwhile an oddly large number of hotspots from wildfires are classified as offshore, though they are certainly on land, from their latitude and longitude coordinates. However, it is true that almost all the type 2 hotspots are from industrial sources, and these could have been removed from the analysis.*

Lines 310 – 315: It seems the author's primary justification for using GSI instead of NDVI is the ease of use because remote sensing data requires an extra step to mosaic the data together. How different would the results be using NDVI? Is there a more substantial scientific justification for using GSI instead of NDVI that can be added here?

> *One of the goals of GFFEPS is running in real time or forecasting, for which NDVI would not be available. NDVI would probably be better than GSI for deciduous leaf*

> *phenology [and grass curing]. However, obtaining and processing NDVI data is significantly more difficult. To [re-]run GFFEPS for the 2012-2019 testing period using the VIIRS/NPP and JPSS1 Vegetation Indices 16-day global 1km product would involve downloading and processing some 80,000 files totalling 2TB – not impossible, but a significant undertaking.*
>
> *Note that the referenced GSI paper Jolley . (2005) has an extensive comparison of GSI against NDVI. The GSI is currently being used by the US Forest Service as part of the National Fire Danger Rating System (https://www.firelab.org/project/national-fire-danger-rating-system). We hope this addresses the reviewer's comment.*

Section 3.4: The authors may want to add some additional caveats including the fact that the FAO stats may not be accurate in countries where agricultural burning is illegal but widespread. For example, in Ukraine and Russia. Please see the following paper for further information: https://iopscience.iop.org/article/10.1088/1748-9326/abfc04

> *Thank you for providing the reference on illegal burning in the Ukraine (Hall et al., 2021). We acknowledged that the approach presented may have issues, especially in developing countries and, in the case of Ukraine, there will be additional caveats to our approach.*
>
> *Hall et al. (2021) that indicates "*that cropland BA [in Ukraine] was significantly underestimated (by 30%–63%) in the widely used Moderate Resolution Imaging Spectroradiometer-based MCD64A1 BA product*", yet this would be a small slice of global emissions. Figure 7a shows that on average, Boreal Asia accounts for 28 MgC yr^{-1}, 2% of the 1479 MgC yr^{-1} global average. Also, many of the fires in Boreal Asia are forest fires in Siberia.*
>
> *Regardless, as the reviewer suggests, we have added the following*
>
>> "With that said, the Tier 1 methodology used by the FAO to determine this value may not be rigorous in developing countries (Tubiello et al., 2014) or where illegal agricultural burning is widespread (Hall et al., 2021); nevertheless, its application in GFFEPS seemed a direct and practical solution for real-time smoke forecasting while addressing the small fire issue specific to agriculture activities." [446]

Furthermore, the Global Cropland Burned Area dataset (Hall et al., 2024) was recently released. It represents the cropland burned area within GFED5 (https://essd.copernicus.org/articles/16/867/2024/). I suggest using either this product or another product specifically designed to map agricultural burned area and compare some of the burned area statistics. The above-mentioned paper focused on Ukraine uses VIIRS active fires, so that is more in line with the author's methodology.

> *Thank you for the recent publication on global cropland burning.*

We have now added a sensitivity analysis to the manuscript as Appendix B. In this analysis, we tested the impact of specific input parameters on the GFFEPS results for 2019. In B.2, we replaced our FAO statistical approach (as presented in 3.4), which accounts for small agricultural fires, with a default fuel load of 0.6 kg m^{-2} *and a burned area per hotpot (as described in 4.1) with no accounting for small fires. The results are strikingly similar to those presented in Hall et al (2024), where we say*

> "Examining the regional differences within agricultural areas we find that in Europe, which has a large fraction of agricultural land though a small contribution to total emissions, the FAO approach used by GFFEPS produced 4.7 times the emissions produced using the average fuel load. Similarly, the FAO approach relative to the fixed values generates in TENA 2.9, in CEAS 2.3 and in MIDE 2.1 times the emissions. These are similar to recently published results by Hall et al. (2024), who reported a 2.7-fold increase in annual average cropland burned area (2003–2020) in cropland regions using the new global cropland area burned dataset (GloCAB) over the MCD64A1 product." [1198]

Section 5.1: The GFFEPS model underestimates the burned area in BONA (two areas with large burned area scars) and TENA. The authors then go on to say that the R^2 shows that the BA methodology is appropriate. Surely, the authors require an appropriate accuracy assessment to make this claim.

In the text we stated that the r^2 *values "*suggest[s] the methodology for estimating burned area is appropriate.*" This is not a strong claim as there is uncertainty in the physical mapping of the fires. The detail in the criticisms of the national statistics is indicative of the complexities involved in accepting reported number and why we can't provide an accurate assessment.*

The point of this paragraph is to compare GFFEPS calculation to real-world data. Regretfully, the latter has its own issues.

The portion of the paragraph has been rewritten as

> "This suggests the methodology for estimating burned area used by GFFEPS is appropriate, though with a bias. On the other hand, reported national statistics of burned area have their own sources of error. For example, the level of rigour in mapping varies between Canadian provincial and territorial agencies, where unburned areas within fire perimeters may be captured by some agencies and not by others. This variable quality is then passed onto the national statistics. Similar issues are likely occurring in US statistics. The issue of mapping irregularities was also recognized by Fraser et al. (2004), who indicated the coarse resolution burned-area (approx. 1-km) provided by SPOT VEGETATION and NOAA AVHRR imagery produced burned-area estimates 72 percent larger than the crown fire burned area mapped at 30 m using Landsat TM (11,039 versus 6,403 ha average area). This bias was attributed to spatial aggregation effects. In summary, it is difficult to make clear

> conclusions from national statistics but these indicate the GFFEPS methodology is producing realistic results." [604]

Fraser RH, Hall RJ, Landry R, Lynham TJ, Lee BS, Li Z (2004) Validation and calibration of Canada-wide coarse-resolution satellite burned area maps. Photogrammetric Engineering and Remote Sensing 70, 451–460.

Line 569 and 579: The authors should compare their BA against GFED5 BA before making this claim since GFFEPS does not account for smaller fires and uses FAO stats for agricultural burning. Also, Africa is dominated by small fires in general, not just agricultural fires. Small fires include the smaller burned patches around larger burn scars, not just an actual small fire (which seems to be how the author interprets them based on lines 665 onwards).

Line 569 states "The lower values are largely attributed to the inclusion of daily fire behaviour in the combustion completeness calculations, not accounted for in the other models" *while 579 states* "These are areas dominated by agricultural burning, highlighting the impact of using FAO's crop-burning statistics."

*The reviewer states that "*GFFEPS does not account for smaller fires*". This is not correct: Agricultural Burning (3.4) of the submitted manuscript specifically describes how GFFEPS accounts for small fires. The reviewer correctly states that GFFEPS "*uses FAO stats for agricultural burning*", without describing why this is a concern. The FAO stats are a direct reporting of all agricultural burning, and hence is directly relevant for accounting for emissions from small, undetected fires.*

With that said, the reviewer references African fires, where FAO statistics may not be handled as well as those in developed countries, and where small fires may be occurring in non-agricultural regions. While this may account for the lower values predicted by GFFEPS in the region (Figure 12), it likely goes beyond agricultural fires. This is discussed in the new Appendix B.1, where discrepancies were prevalent in Africa when substituting MCD12C1 for GLC2000 land classification

> "These differences are likely to poor matching of coniferous versus deciduous forests, a distinction not captured in MCD12C1 classifications Savannas and Woody savannas (as previously described). The difference between coniferous and deciduous fuels is critical in the FBP fire behaviour calculations and any misclassification would have an impact on predictions. Also, difficulties mapping fire emissions and land classifications in Africa have been discussed in various papers (Ramo et al., 2021; Nguyen and Wooster, 2020; Zhang et al. 2018), possibly accounting for the discrepancy shown in this comparison." [1163]

As for comparing GFFEPS burned area against GFED5, reference was made to Chen et al 2023 in the original manuscript (line 76) as a method to account for the small fire boost. This was published as our text was being prepared and would require a new analysis (see summary).

Finally, the methodology presented in the submitted manuscript, using FAO data, is the most recent (and we feel, the most accurate) of a series of approaches we examined to address the issue of small fires in GFFEPS. The first approach included the small fire boost documented in Randerson et al. (2012). The second approach assumed that fire size follows a power law (Cumming 2001; Hantson et al. 2016; Reed et al., 2002), using data from the Global Fire Atlas (Andela et al. 2019). Insufficient information was available to give us confidence in the first approach, and the second, while promising, resulted in extrapolations of small fires that often produced unacceptable results. With that said, the second approach pointed to a clear distinction of agricultural verses non-agricultural small fires, allowing us to confidently proceed with our assumption that small fires are inconsequential in non-agricultural regions. This was further supported in the Appendix B.2, where we now say

> "Figure B.4 shows most variation between the methods occurs near the origin, and closer examination reveals this variation occurring primarily in the agricultural regions. Examining the regional differences within agricultural areas we find that in Europe, which has a large fraction of agricultural land though a small contribution to total emissions, the FAO approach used by GFFEPS produced 4.7 times the emissions produced using the average fuel load. Similarly, the FAO approach relative to the fixed values generates in TENA 2.9, in CEAS 2.3 and in MIDE 2.1 times the emissions. These are similar to recently published results by Hall et al. (2024), who reported a 2.7-fold increase in annual average cropland burned area (2003–2020) in cropland regions using the new global cropland area burned dataset (GloCAB) over the MCD64A1 product." [1196]

Andela, N., Morton, D.C., Giglio, L., Paugam, R., Chen, Y., Hantson, S., Van Der Werf, G.R. and Randerson, J.T., 2019. The Global Fire Atlas of individual fire size, duration, speed and direction. Earth System Science Data, 11(2), pp.529-552.

Cumming, S.G., 2001. A parametric model of the fire-size distribution. Canadian Journal of Forest Research, 31(8), pp.1297-1303.

Hantson, S., Pueyo, S. and Chuvieco, E., 2016. Global fire size distribution: from power law to log-normal. International journal of wildland fire, 25(4), pp.403-412.

Reed, W.J. and McKelvey, K.S., 2002. Power-law behaviour and parametric models for the size-distribution of forest fires. Ecological Modelling, 150(3), pp.239-254.

Section 6: Why is there no accuracy assessment/ validation on the burned area product? Since you are using MODIS and VIIRS, the authors can use the BARD dataset (https://edatos.consorciomadrono.es/dataset.xhtml?persistentId=doi:10.21950/BBQQU7). I don't recommend publishing a new product paper without an adequate burned area validation assessment since that is the primary input into the emissions calculations.

> *The GFFEPS system is primarily intended as a near real time forecasting system, building on the Canadian model. To emphasize this, we extended (shown in italics) our motivation statement in the introduction to now say:*
>
> > "The motivation for this work was the recognized need in extending FireWork's current North American air-quality forecasting to the global domain, *thus improving Canadian forecasts by introducing near real time global simulations of smoke emissions external to the original North American domain*." [108]
>
> *Through the historic burned-area per hotspot approach, GFFEPS calculates burned area and emissions in near-real time on a daily basis. Other models such as GFED depend on month-end summaries of burned area, thus providing emissions in an historical context.*
>
> *In Burned Area (4.1), we described our approach of using the historical MCD64A1 data to determine a historic burned area per hotspot, which is then used to predict burned area on a daily basis. This can be summed into global burned-area estimate (essential this is what we are doing in 5.2), yet this would ultimately come back to the original burned-area values from MC64A1.*
>
> *The GFFEPS results are based on MCD64A1 data as its underlying, historical burned area. The MC64A1 burned area product has been assessed/validated in Giglio et al. (2018), which, in turn, was used by GFED4.1s and used to calibrate GFAS1.2; validation and comparative studies have been conducted based on this burned area product (Van der Werf et al., 2017; Pan et al. 2020). This is what we chose as our benchmark for comparing GFFEPS against other models, using common burned area products to allow us the ability to focus our attention on the impact of spatial and temporal variability.*
>
> *Ultimately, MCD64A1 will be replaced by other, newer products. With that in mind, we feel that conducting a formal assessment/validation of various burned area products is best provided by the producers of these new products. Instead, the GFFEPS model, as a design requirement, focuses on providing operational, real-time calculation of fuel consumption, which drives the forecasted emissions predicted in the model. This is the product of fuel load and combustion completeness (FL × CC) of the Seiler and Crutzen equation. This is a component under-evaluated in most global models; where instead, a static value per biome is frequently used.*

Minor comments:

Title: Why is the product called a "Forest Fire" product when the authors are mapping burning in all land cover types?

> *The model's title is in keeping with other models in the Canadian Forest Fire Danger Rating System (CFFDRS), including the Canadian Forest Fire Weather Index (FWI) System and the Canadian Forest Fire Behavior Prediction (FBP) System.*

Line 16: change to "showing"

Done

Line 64: FRE is the time integral of FRP

Corrected. Now sentence reads

> "… measurements of fire radiative energy (FRE, the time integral of fire radiative power (FRP)…" [65]

Lines 74 - 79: The GFED5 Burned Area product has incorporated the GloCAB data (Hall et al., 2024) which provides a cropland burning-specific dataset. https://essd.copernicus.org/articles/16/867/2024/

Thank you for the recent publication on global cropland burning.

GloCAB represents a recently published dataset that is not inline with the methodology using FAO statistics as presented in this manuscript; we've referenced it in our revised manuscript as another methodology of determining cropland burning estimates. It may be of value in future versions of the GFFEPS model, this would require a specific comparison between the approaches which is beyond the scope of this paper (see summary).

Line 115: change to "multiplied"

Corrected

Section 3.1: The justification for using the GLC2000 is weak, especially now that it is almost 25 years old. Have the authors run a sensitivity analysis with other land cover datasets to see how well the GLC2000 dataset has held up in recent years?

A sensitivity analysis is now included as Appendix B.1. It compares GFFEPS emissions for the 2019 using GLC2000 land cover to those using MCD12C1. Although there were regional differences, global results indicated near equality.

> "The scatter plot shows near equality between the two model predictions (a slope of 0.98) when forced through the origin, with an r2 of 0.93. Total annual emissions were 2,957 and 3,028 Mt as predicted by GLC2000 and MCD12C1 respectively." [1145]

Line 269: "sufficiently complete" is quite a strong statement. VIIRS does not include the morning overpass compared to MODIS so the fire location data is already missing a large number of fires. I would remove the last portion of that sentence. It is also worth mentioning that since you are using the active fire product that you only have the afternoon snapshot of fire pixels as opposed to MODIS which has morning and afternoon.

The sentence states "In spite of these limitations, hotspot data from VIIRS provides a picture of global fire activity that is consistent, continuous, and sufficiently complete."

*This was intended as a generalized, self-evident statement. "*Sufficiently complete" *was referring to global coverage. We have rewritten it as follows:*

> "In spite of these limitations, we selected VIIRS data because it is sub-daily, global, readily available, higher resolution than alternative sensors, available in near-real time, and expected to continue well into the 2030s.." [299]

We acknowledge that, lacking a morning overpass, VIIRS hotspot data is less able to characterize diurnal patterns of fire behavior - though not completely unable, since VIIRS also does nighttime detection (as does MODIS). In spite of this limitation, because of its higher resolution and sensitivity, VIIRS detects more fires, smaller fires, and more burned area than MODIS, and the hotspots have better geolocational accuracy (references below). Also, it is well established that fire behaviour peaks in the afternoon (Countryman 1972) and hence, afternoon detections would more likely and reliable. Finally, GFFEPS has a diurnal pattern to model area growth per time step as documented for CFFEPS in Chen et al. (2019).

Schroeder, W., Oliva, P., Giglio, L. & Csiszar, I. 2014. "The New VIIRS 375 m active fire detection data product: Algorithm description and initial assessment." Remote Sensing of the Environment, 143(2014). doi:10.1016/j.rse.2013.12.008

Fu, Y.; Li, R.; Wang, X.; Bergeron, Y.; Valeria, O.; Chavardès, R.D.; Wang, Y.; Hu, J. Fire Detection and Fire Radiative Power in Forests and Low-Biomass Lands in Northeast Asia: MODIS versus VIIRS Fire Products. Remote Sens. 2020, 12, 2870. https://doi.org/10.3390/rs12182870

Figure 6 and Figure 7 captions: It would be helpful for readers unfamiliar with these indices to have a brief description of the meaning of the value in the caption.

Addressed. Captions now read

> "Figure 6. Buildup Index (BUI) for September 1, 2019 as interpolated to the 63,566 hotspot locations observed on that date. The BUI, a principal driver in calculating fuel consumption in the FBP system, is calculated using meteorological data from Environment and Climate Change Canada's Global Environmental Multiscale (GEM) model." [323]
>
> "Figure 7. Growing Season Index (GSI) for September 1, 2019 as interpolated to the 63,566 hotspot locations observed on that date. The GSI provides a method to estimate the greenness of deciduous forests and degree of grass curing, both important factors in fuel consumption. The 21-day average GSI is calculated using meteorological data from Environment and Climate Change Canada's Global Environmental Multiscale (GEM) model." [356]

Line 374: There is a newly released dataset that compiles all the crop-specific emission coefficients from the literature. It is available here: https://doi.org/10.5281/zenodo.7013656

> *Thank you for the updated information. The reported values range from 0.89 to 0.92, with a generic crops value of 0.92. This is very close to the value we used (0.90). New crop specific values will be included in future runs.*

Line 387: Determining the other small fires as "inconsequential" is not an adequate justification for not developing a methodology to improve the representation of small fires, especially given the numerous papers showing how many small fires there are on the landscape. I suggest rephrasing.

> *Line 387 states* "while assuming small fires in other, non-agricultural, landscapes were deemed inconsequential". *The reviewer has not acknowledged that the FAO approach we presented in subchapter 3.4 deals specifically with agricultural fires and presents a methodology that captures the emissions from large and small in agricultural zones. This is further discussed in the new Appendix B.2. This was also answered earlier in response to the reviewer's questions regarding lines 569 and 579.*
>
> *As for non-agricultural fires, our experience in Canada, as referenced on the Canadian Wildland Fire Information System is that* "Fires of all sizes are included in the database, but only those greater than 200 hectares in final size are shown in the map above — these represent a small percentage of all fires but account for most of the area burned (usually more than 97%)." *To support this, we added the following text*
>
> > "The approach does assume small fires in other, non-agricultural, landscapes are inconsequential, which we see as acceptable. This is certainly the case in Canada, where the National Forestry Database (https://cwfis.cfs.nrcan.gc.ca/ha/nfdb, last accessed 2024-05-28) indicates that between 1980-2021, fires less than 1 ha, which constituted 73% of fires, account for only 0.03% of the burned area nationally; that fires less than 10 ha, which account constituted 87% of fires, account for only 0.18% of the burned area." [440]

Line 427: Why not just remove the persistent sources?

> *In some cases it is difficult to determine which persistent sources are fires. Slow-burning fires that burn within the same pixel for more than one day, or areas that are frequently burned for agricultural purposes, could be considered "persistent"; e.g. agricultural burning may take place through transporting the material to be burned to a common site. However, the suggestion is a good one, as there are many locations that are clearly not fires that could have been removed. However, because our method minimizes the emissions from persistent sources, the impact on smoke emissions results would be insignificant.*

Summary

We disagree with the review's thesis is that we are premature in publishing the GFFEPS model; that there is newer data that should be incorporated (NDVI; GFED 5; Global Cropland Burned Area); and that several comparative studies need to be conducted (GSI vs NVDI; Global Cropland Burned Area vs FAO data; GFFEPS versus BARD; GLC2000 vs MODIS landcover datasets). We note that the intended use of GFFEPS in the context of real-time forecasting was not clear to the reviewer, that this purpose governs many of the choices made regarding model inputs, and we have modified the manuscript accordingly to clarify this constraint.

We added a sensitivity analysis as Appendix B, where we carried out comparisons using of MODIS land use types versus GLC2000, using a default crop burning value and burned area per hotspot as opposed to the FAO agricultural burning, using fuel consumption based on fixed combustion completeness for grass, forest and peatland versus fuel consumption based on daily weather and fire behaviour. Other requests such as comparison of the GSI vs NVDI have taken place in the literature – and we note that in the context of real-time forecasting, the former represents a much smaller processing time requirement. As new data become available, they may be incorporated into GFFEPS, provided that they may be used in an operational forecasting environment, where minimizing processing time while generating emissions is a key requirement.

Our position is that we have reached a clear benchmark and feel that the methodology and our results based on common published datasets is worthy of publication. The model is compared to other models (GFAS1.2, GFED4.1s, FINN1.5/2.5), using results that have been published for each. Many of these models (e.g., GFED4.1s) use older burned area products (MC64A1) and it is only fitting that model comparisons should be conducted with similar input. We also note that historical burned area data must be used in the context of real-time forecasting.

This manuscript presents the initial version of GFFEPS. The model will continue to evolve with new methodologies and improved underlying data.