**Specific Comments**

Abstract: the abstract mentions only regions where the introduction of the HDI improved the INFERNO model, which is highly misleading. There are many areas in the model where the introduction of the HDI resulted in worse performance, including regions where observed positive trends that were previously reflected in the model are no longer present after the inclusion of the HDI.

Lines 7-8: It should also clarified that the reduction of the global burned area results in a greater global bias than was present in the model version without inclusion of the HDI. Without this clarification this statement creates the appearance that the reductions in burned area were always an improvement.

Line 9: “by more than 100%” is a confusing way of wording this, it implies that these positive biases became negative biases, which is true in some cases, but not all. Maybe the intent is to say that the relative bias is reduced by more than 100 percentage points?

Line 9: “without statistically significant impact to 10 other areas” statistical significance is never discussed in the manuscript when referring to biases in the burned area, so it unclear where this statement comes from. Many of the other areas do include substantial changes. Northern Hemisphere Africa and Southern Hemisphere Africa together have an increase in their biases, after inclusion of the HDI, of 26.86 Mha, which is more than the total burned area in all of the other regions individually, apart from Australia. Australia itself has an increase in its bias of 24.28 Mha, from -8.46 Mha to -32.74 Mha, relative to a total burned area of 39.88 Mha. An argument that this is not statistically significant is false and extremely misleading.

Line 10: The observed burned area trend in Southern Hemisphere Africa, according to Table 2 in this manuscript, is -0.54 Mha/yr, the trend in the model version without the HDI is -0.14 Mha/yr and the trend in the model version with the HDI is -1.94 Mha/yr. Therefore, the model version without the HDI is closer to the correct trend. While it is true that the HDI version of the model shows a closer trend in Northern Hemisphere Africa, the total observed trend in Africa is -2.74 Mha/yr, the total trend in the model version without the HDI is -1.4 Mha/yr, and the total trend in the model version with the HDI is -4.65 Mha/yr. The model version before the introduction of the HDI therefore has a bias of 1.34 Mha/yr across Africa and the model version after the introduction of the HDI has a greater bias of -1.91 Mha/yr across Africa. I am leaving out the parts of the middle east region that are on the African continent, but since that region as a whole only has a burned area of 0.9 Mha over the time period under analysis it should not make a significant difference. Therefore, the statement that the introduction of the HDI “improves the representation of the burnt area trends, especially in Africa. Central Asia and Australia” is false, or at the very least highly misleading.

In addition, the abstract fails to mention that the trend is worse in several regions including Temperate North America and South East Asia, where observed positive trends incorrectly become negative trends. Without this clarification the abstract gives the impression that the introduction of the HDI resulted in a uniform improvement in modeled trends.

Lines 33-36: It is unclear whether this summary is supported by the Andela et al. (2017) paper that the authors cite. Andela et al. (2017) show contrasting relationships between population density and burned area in different regions that cannot be simply summarized by stating, as the authors do, that “areas with low population density are associated with lower burnt areas, and densely populated areas tend to be associated with increased burnt area.” Even though the authors caveat this in the subsequent sentence, they only do so for prosperous regions, which is not supported by the literature. E.g. the findings of Archibald et al. (2009), whom the authors cite in this work.

The subsequent statement that “for heavily populated and prosperous regions, burnt area decline is likely a function of perceived threats to highly valued infrastructure, prompting extensive fire suppression efforts, sometimes involving high monetary costs.” Is somewhat problematic as it implies that suppression efforts are based on strictly economic considerations, and that fire is not suppressed in heavily populated regions that are not prosperous. This neglects the impact of population density on fuel availability as even less prosperous urban areas contain significantly less fuel than non-populated areas. The impact of high population density on reducing burned area even without significant prosperity is visible in the strong negative correlation between population density and burned area in, e.g., Bangladesh in Figure 4 of Andela et al. (2017)

Line 47: This is the only source the authors cite in which the HDI is used to predict fire activity. However, Chuvieco et al. (2021) use the HDI as a predictor of the coefficient of variation of burned area. This is in contrast to this manuscript where the HDI is directly implemented in equations that determine the modeled number of ignitions and amount of fire suppression and the aim is to improve burned area in an absolute sense rather than just its coefficient of variation. Without further analysis this is an insufficient theoretical basis for including the HDI in the INFERNO model in the manner that the authors do.

Line 57: The authors identify a “need for data collection to improve the quantification and modelling of fire activity and human populations.” This statement is strongly supported by their discussion up to this point. However, many datasets are available, including ones that the authors use in this manuscript, that would allow for a data analysis to establish what the relationships between HDI and fire activity are, and to what extent they exist. Without such analysis the implementation of the HDI presented here is somewhat arbitrary.

Line 87: While it is true that local meteorology and topography cannot be resolved by ESMs, is it sufficient to leave out these considerations entirely? For example, Haas et al., 2022 (DOI: 10.1088/1748-9326/ac6a69), show that topography and wind speed have an impact on fire size even when aggregated to a 0.5° grid cell scale.

Equations 2 and 3: What is the justification for this particular functional form? Beyond general arguments that greater development should lead to fewer ignitions being caused and greater resources that can be applied to suppression, the authors do not provide a justification for why the relationships should be linear in particular, and why, in these linear relationships the HDI should be multiplied by -1, rather than any other negative number (e.g. multiplying the equations by 0.5-0.5HDI). Because the authors also do not provide any evidence that the HDI can act as a predictor for the number of ignitions or amount of suppression, these equations do not have a sufficient theoretical basis.

An additional concern is that these equations strictly reduce the number of ignitions and increase the amount of suppression from the previous values, since the HDI is always greater than 0. This implies that, previously, ignitions were always overcounted and suppression was always underrepresented. Please clarify how these equations were derived in the first place and why this derivation may have led to these biases.

In addition, these equations imply that greater development always results in increased fire suppression. This neglects key aspects of fire management and changing practices in many regions, for example in British Columbia as described by Nikolakis and Roberts (2020), whom the authors cite on several occasions. These changing practices in many cases involve fire management policies that do not rely on complete fire exclusion but involve prescribed burning, which is itself a source of ignitions. There is a significant body of evidence that extreme fire exclusion results in fuel buildup that can cause subsequent fires to become larger (e.g. summarized by Plucinski 2019, DOI: 10.1007/s40725-019-00085-4). Therefore, many agencies are moving away from a focus on fire exclusion, and this is in direct contrast to the parametrization in this manuscript.

Finally, the authors' implementation of the HDI to account for fire management impacts the number of fires but does not account for the ability of fire suppression to reduce the final size of spreading fires. By only including the impact of HDI on the number of fires, the authors fail to reflect the findings in Andela et al. (2017), whom they cite, that show a decrease in fire size as well as the number of fires over a similar time period to this study, with this decrease in the number of fires having a significant impact in regions including northern Africa.

Line 153: please provide more details on how these PFT-specific average burned area values are derived from the values in Andela et al. (2019). There is quite a stark contrast between the previous values and the new ones, in some cases the new ones are over triple the previous values.

Line 171: please clarify how this spatial correlation metric is calculated, is it based on the mean burned fractions per grid cell across the time period in question?

Figure 3: the introduction of the HDI appears to cause some sharp boundaries between countries such as Canada and the United States despite this not being visible in the satellite data. This could be worth commenting on, and the figure descriptions in the captions should be expanded in general.

Also, please clarify in the caption that this figure uses a color axis in which the difference in colors does not correspond linearly to differences in burned fraction, and include maps with a continuous, perceptually uniform color scale for better comparisons between different regions in addition to the current plots.

In general, the maps in this paper would be improved by using an equal area map projection. Because the purpose of these maps is in part to compare how much of the world's surface each model version performs well in, a projection such as the Equal Earth projection (Šavrič el al., 2018, DOI: 10.1080/13658816.2018.1504949) would allow for this, whereas the current projection appears to inflate areas farther from the equator.

Figure 4: This point is also addressed elsewhere, but these maps do not appear to show a clear improvement in the model by including the HDI. Especially considering that many of the regions where the bias is lower in the non-HDI model run, such as northern Australia and many parts of Africa have very high burned areas, so biases there can have strong impacts on global fire.

Table 2: For the sake of easier interpretability please highlight which metrics show better performance than their counterpart in the other experiment, e.g., by coloring the cells in the table

In general, more of the metrics in this table, where the two model version show a difference, 72 vs 58 by a cursory count (not double counting mean BA and bias), are in favor of the model version without HDI. This is 4 vs 5 on a global scale, and 68 vs 53 on the regional scales for the non-HDI version vs the HDI version. The authors state on line 72 that they “aim to improve the regional representation of human–environmental coupling for applications at large spatial scales within an Earth System Model (ESM) context.” These metrics appear to show that this aim is not met.

Please also include confidence intervals and $R^2$ values for the trends in this table

As the authors cite Chuvieco et al. (2021) where the HDI is shown to have an impact on the coefficient of variation of burned area, the coefficient of variation should also be included in this table and discussed in the text.

Figure 5: The legend states that GFEDv4 is being used for this plot, whereas GFED 4s was used in the rest of this paper. Is this a typo? Please clarify

Also, please clarify what the -clim experiments are. Do these correspond to the sensitivity analysis described later? If so, why are these shown rather than the others? And is it true that, as these plots appear to show, the burned area in the HDI experiments drops nearly to 0 if only the atmospheric drivers are transient and all other drivers are kept at their 1990 values?

Line 225: these are global numbers, please clarify how they correspond to a regional improvement. The statement that the improvement in global standard deviation is reflective of a regional improvement is extremely problematic since STD/STDGFED4s is better for the non-HDI model version in 9 out of the 14 regions and only improved in 5, in some cases only marginally.

Line 226: “reducing the RMSE” the RMSE is significantly lower for the non-HDI version, only the $RMSE_{UE}$ is reduced.

Figure 6: The inclusion of the HDI appears to exacerbate the negative bias in the trends in many parts of Africa, and, given the values in Table 2, imposes almost uniformly negative trends on the burned area, even in regions where this is not the case in the observations. This is to be expected,

given that the HDI shows an almost global increase over this time period, as shown in Figure A2 and Equations 2 and 3 impose a negative correlation between HDI and ignitions as well as the number of unsuppressed fires. Given the lack of theoretical backing and explicit validation of those equations, this imposition of a negative trend in a uniform manner is somewhat arbitrary, and does not sufficiently support the statement in line 243 that “Overall, including socio-economic factors in INFERNO results in an improvement in burnt area trends.”

Line 256: This section is interesting in general. The only part missing is a discussion of how the drivers that are compared individually correlate to each other. From the standpoint of strictly testing model behavior, varying these drivers individually is instructive, but if broader conclusions are intended it is relevant to discuss if, e.g., a change in population density also corresponds to land use changes in general and, therefore, whether it is realistic to decouple them.

Table 3: please also include mean burned areas as in Table 2, and include $R^2$ values for the trends.

Line 320: Conclusions is not the most appropriate title for this section, discussion would be more fitting.

Line 325: This statement is not supported by the evidence the authors provide, for reasons discussed previously, particularly given that the non-HDI version shows better performance in many areas, most by some metrics

Line 332: Given that the trends are so uniformly negative upon implementation of the HDI, whether this is an improvement is unclear.

Line 337: Same comment as above

Line 367: Is this sentence referring to an impact on the effect that the anthropogenic drivers have? Somewhat unclear, please rephrase

Line 376: it is good that the authors address this issue here, but it is somewhat hidden among the statements claiming that the model has been improved. Given the evidence shown, the issues in this approach should be emphasized to a greater extent.

Line 408-409: please provide a citation, or clarify that this is based on the authors’ experience

Line 434: Given the issues in this paper, and the fact that the HDI version appears to enforce strong decreasing trends in many parts of the model, even where there are none observed, it is somewhat concerning that calculations for such future scenarios may be biased towards less fire activity, leading to an underestimation of the dangers that fire poses under climate change.

**Technical Corrections**

Line 5: should read “the Human Development Index” rather than “a Human Development Index”

Line 10: should be a comma after Africa rather than a period

Line 55: “due to the effects of reflecting the effects” redundant, please rephrase

Line 315: There appears to be a typo in this sentence as the “JULES-INFERNO+HDI (pop and lu)” experiments are referred to as containing both a burned area decrease and increase.

Line 321: Stray section header at the start of this sentence

Line 346: “the inclusion of socio-economic factors reduce the role of temperature in driving trends by reducing the role of temperature in driving trends” redundant, please rephrase

Line 348: should be “that climate drivers” rather than “of climate drivers”

Line 405: remove the word “on” from “impact on the modelled burnt area”

Figure A1: some subplots have y labels while others do not. Since what is represented on the y-axis is the same for all plots, it would be better to just leave them out

Figure A1: the legend obscures too much of the timeseries, please move it for clarity

Figure A2a: labels on colorbar are cut off

Figure A2e: title says “Pupulation” rather than Population

Line 505: This reference appears to be repeated twice