

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2022-195-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-195

Anonymous Referee #1

Referee comment on "AttentionFire_v1.0: interpretable machine learning fire model for burned-area predictions over tropics" by Fa Li et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-195-RC1>, 2022

General comments:

The authors of "AttentionFire_v1.0: interpretable machine learning fire model for burned area prediction over tropics" developed a novel machine learning model of fire for use in South America and Africa. They use the model to gain insight into the controls of wildfire on the landscape and make future predictions. Overall, this paper is interesting and provides insight into an important process and region of the world. I have some major concerns, however, mostly related to the presentation of the methods that should be addressed to improve the clarity and rigor of the manuscript.

1. Adding more explanation about the model and how it related to other machine learning models in plain terms.
2. Providing more information about the data sets used to conduct this analysis to allow the reader to better understand and assess what was done.
3. Make it more clear how the model was validated and include a test against independent data withheld from any tuning to guard against overfitting.
4. Add more explanation about how the future projections were conducted, what input data sets were used, and if and how they were stepped into the future.
5. Provide methods, background, and discussion for section 3.3 which are missing.

6. Address via analysis or discuss the impact of bias in CESM versus the reanalysis data, the impact of coupling between fire, climate, and biomass, and model/scenario uncertainty on the future projections presented.

Specific comments:

L24-27: I suggest also mentioning anthropogenic drivers as they're included in the model. Currently, only climate is highlighted.

L41: Placing this emissions number in the context of the carbon budget of these regions using published values could better highlight the importance of this work.

L105-108: I suggest adding more detail here, especially for a reader who is not familiar with machine learning models. That is define black boxed and explain why more complex machine learning models are often less interpretable in straightforward terms. The acronym LSTM should be defined here as well.

L120—130/140-179/figure 1: More background is needed in this section, especially for a reader who is less familiar with artificial neural networks. That is I suggest stating that an LSTM is a type of ANN and explaining its practical advantages and disadvantages versus a typical ANN, and the Naïve LSTM in straightforward terms. Maybe this could also take the form of a table. Figure 1 could be better tied into the text with definitions given for more specific terminology used.

L192-193/198-207: The description of the data sets, their time step (i.e. daily, monthly, etc), units, and origin should be included here. I also suggest moving table S2 into the text and editing it to include more information.

L193-195: Please clarify the validation method used here. Is this a leave-one-out cross-validation? Was any model tuning conducted and what data was used to do that? Ideally, the models would be validated against independent data that was withheld from any tuning/testing to guard against overfitting.

L208-215: These future data need to be prefaced and explained a bit more. Was AttentionFire coupled with CESM or are these simply outputs from CESM? How were variables like road network density and livestock projected into the future? If AttentionFire

was not coupled in CESM was bias correction applied to deal with any biases present in the model run, but not the reanalysis? Could these biases impact the results or trends predicted by AttentionFire?

Fig 2: Suggest editing the caption to provide more information about each panel of the figure.

Figure 3: It's unclear from the figure caption what each of these three panels shows as no letter codes are provided.

Figure 3: Was an attempt made to simplify the model by removing low-ranked data sets? This could be beneficial if it eliminates unimportant variables which are uncertain or hard to obtain in the future

Section 3.3: These experiments are not included in the methods, no background for this is included in the introduction and, acronyms are not defined. Substantial background needs to be added here.

Section 3.4: Several points regarding the future projections are not addressed here. First is the possibility that there is bias in CESM which is not present in the reanalysis data. Should a bias correction be applied? Second, fire, climate, and biomass on the landscape are all coupled. Therefore there is a need to address how this could impact the estimates and trends given if the fire model is not coupled with CESM. Finally, if the models are not coupled only a single model run using a strong emissions scenario is presented here. I'd suggest either presenting additional scenarios and including other models or explaining how model and scenario uncertainty could impact the results, their applicability to this region, and the significant trends highlighted.

Minor comments:

L69: Suggest replacing "from climate" with "of climate"

L70: Suggest replacing "up to multiple" with "on the order of"

L81: Suggest replacing "opposing fire" with "opposite fire"

L212: Suggest adding "the" between "2016-2055" and "99th"