# Comment on acp-2021-634

Anonymous Referee #1

---

Referee comment on "Interpreting machine learning prediction of fire emissions and comparison with FireMIP process-based models" by Sally S.-C. Wang et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-634-RC2, 2021

---

In this paper, the authors followed their previous study (Wang et al., 2021) and used ML technique for predicting fire emissions using gridded GFED fire emission dataset (as target) and meteorological, land-surface properties, and socioeconomic variables (as predictors). The performance of ML is evaluated and compared against FireMIP process-based models, and is interpreted using SHAP. The paper is clear written, and the scientific findings presented are important and suitable for the journal of ACP. However, there are a few issues need be clarified before the paper is considered for publication. Here the reviewer recommends a "resubmission after major revisions".

General comments:

- It is well known that different fire emission datasets (i.e. GFED, QFED, FINN, etc.) predict biomass burning emissions with large discrepancies, for example, Figure 6 in Liu et al. (2020). Large uncertainties associated with GFED dataset are due to the accumulated errors from burned area, fuel type/condition, and burning condition/fire weather. It is reasonable that ML results agree well with GFED, because it is trained against GFED. However, when validating FireMIP, the authors may consider the other fire emission dataset and examine whether the correlation coefficients are different. I understand that this requires additional work. The authors can ignore this suggestion but add additional discussion in the conclusion.

- The way the ML model is trained guarantee the better performance of ML technique in capturing the interannual variability of fire, because the 10-fold cross-

validation/random sampling method is used. When randomly splitting the whole sampling pool into 10 groups, the interannual variability information is stored in the 9 groups that are used for training. I strongly suggest the authors to examine the performance of ML by using entire or two years' data for validation purpose only. For example, the authors can train the ML using data from 2000 to 2019, and perform the ML model to the data of 2020, and examine the performance of ML model (total PM2.5, spatial distribution, seasonal variability, etc.). I believe by doing so, it can be considered as fair comparison against FireMIP models. In addition, the method will be more suitable for the future prediction. For the case studies, aren't the data of these two cases are included in training ML, unless I miss the text.

- I am wondering, since the authors have identified the top important predictors for ML model, whether it is possible to reconstruct ML model using just these top 20 predictors and what the performance of this new ML model will be. When incorporating this ML model in ESM like E3SM or CESM, it is always desirable to use less parameters.

Specific comments:

Line 49: "same parameters".  To my knowledge, there are some studies that use tuned parameters for different regions, like Zou et al., 2019.

Figure 7 looks too busy to read. Is it possible to improve the readability of this figure, especially b, d, f, h. The authors may consider to increase the thickness of ML results and GFED data

Reference:

Liu, T., Mickley, L. J., Marlier, M. E., DeFries, R. S., Khan, M. F., Latif, M. T., & Karambelas, A. (2020). Diagnosing spatial biases and uncertainties in global fire emissions inventories: Indonesia as regional case study. *Remote Sensing of Environment*, *237*, 111557.

Zou, Y., O'Neill, S. M., Larkin, N. K., Alvarado, E. C., Solomon, R., Mass, C., ... & Shen, H. (2019). Machine learning-based integration of high-resolution wildfire smoke simulations and observations for regional health impact assessment. *International journal of environmental research and public health*, *16*(12), 2137.