Comment on gmd-2021-83

Matthias Forkel (Referee)

Referee comment on "Building a machine learning surrogate model for wildfire activities within a global Earth system model" by Qing Zhu et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-83-RC2, 2021

Review of Zhu et al. 2021: "Building a machine learning surrogate model for wildfire activities within a global earth system model"

This study presents approach to build a deep learning-based model to better simulate burned area as part of an Earth system model. Although several machine learning and data-driven fire models were developed in the last years, this is a first study that directly aims to implement a deep neural network (DNN)-based fire model with a Earth system model. The paper is well written.

However, I have several questions and concerns.

1 Integration of DNN-based fire model with the Earth system model

The paper is not clear about how the DNN-based model with implemented with the Earth system model (ESM). For the title and abstract, I expect the DNN model was implemented in the ESM. This would allow analyses about how the improved simulation of fire affects the simulated carbon fluxes and stocks in the ESM. But as the paper does not represent such results, I assume that DNN-based fire model was just applied outside of the ESM and that both models were actually not coupled. Hence, I'm wondering how the authors to imagine to couple both models. Especially the final DNN-Fire-GFED setup simulates clearly

different burned area then the original BASE-Fire or DNN-Fire models setups. This implies, that for example a much higher simulated burned area in Africa should result also in a much lower biomass in Africa and hence changes the fuel load variable as input to the DNN-fire models. In the coupled model, the DNN-Fire-GFED model would lead to results that are inconsistent with the feature space that was initially used to train the DNN-Fire model. Ideally, the authors should do a sensitivity analysis in the coupled DNN-Fire-GFED and ESM models to see if the results are still consistent and reliable. If this is not feasible, the authors should at least discuss how they would address such inconsistencies. I assume that only a joint optimization of fire and fuel loads/biomass in the coupled model would solve this issue (Drüke et al., 2019).

2 Training and testing

The authors trained a DNN model for each GFED region. Training the model for different regions is an unfair approach in comparison to process-based fire models as these models are truly global models, maybe with a PFT-dependent parametrization. Hence the authors should provide a good reasoning why they trained the model per GFED region. In addition, it does make sense at all that a fire model is parametrised per GFED region for an application in an Earth system model. As Earth system models are applied to assess future changes, a parametrisation per region will fast lead to useless results. For example, if climate and vegetation conditions change in future, which regional model should be applied in a certain region? Fire should be only simulated as a response to climate, vegetation and socioeconomic conditions. If regional parametrisation is necessary, the parameters should be based on vegetation or socioeconomic conditions.

The monthly burned area data from all grid cells in each regions was splitted randomely in 80% training data and 20% for testing. This is one of the simplest tests as the underlying conditions and statistical distribution of both samples is the same. However, in the context of an Earth system model, we expect non-stationary conditions and hence the model should be tested how well it can predict into 1) different regions, 2) different time periods (was done but the conditions in the two time periods are very similar), and 3) to different environmental conditions (Klemeš, 1986).
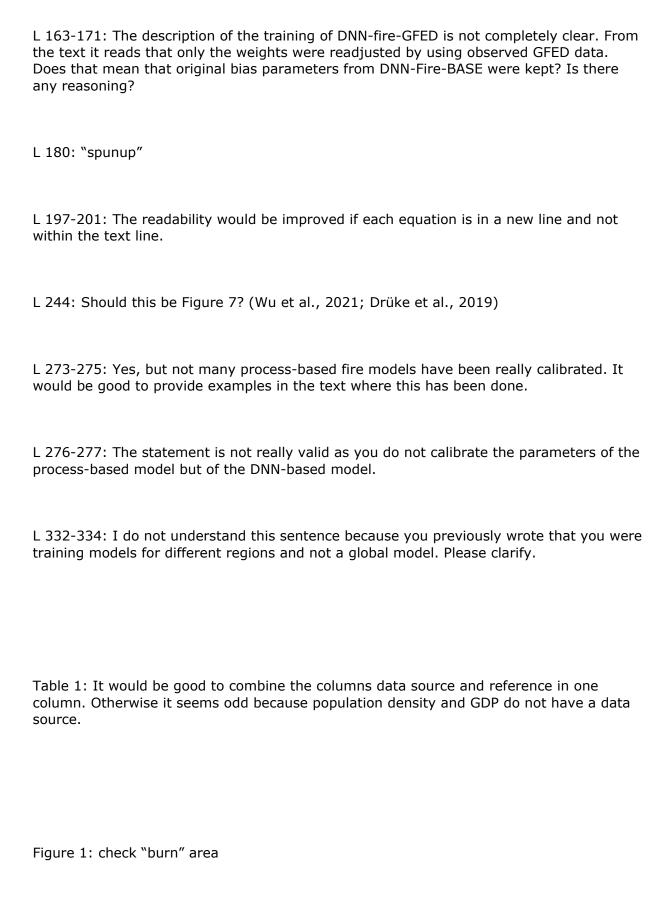
3 Input data

Most of the input data for the DNN model comes from climate, land use or socioeconomic datasets. Information on fuel loads, fuel wetness and temperature, however, was taken from ELMv1 model simulations. I wonder about how good are these simulated variables in

comparison with independent (e.g. Earth observation) data. For example, any biases in simulated biomass will directly affect the simulated burned area. Please compare the simulated biomass and soil moisture with useful datasets. Alternatively, a residual analysis would be also useful to see if any errors in simulated burned area rea related to errors in the simulated input.

Can you please demonstrate that the tree cover from the LUH2 dataset is consistent with the simulated biomass. Are there any areas where the simulated biomass does not correspond to tree cover?

Specific comments

L 26-27: From this statement it is not clear if the DNN is implemented as part of the E3SM or if it is independent of the ESM and just returns the same output. Please clarify

L 30-31: It is not clear what the R2 means. Is it the R2 between the observed and predicted global annual total burned area in 2001 and 2015?

L 41: The statement should be updated with newer estimates, e.g. by (Lasslop et al., 2020)

L 78-93: You should clarify the scale of wildfire models. Fire behaviour models aim to model the spread and intensity of individual fires and are widely used in fire management. Fire models as parts of global vegetation or Earth system models have a different purpose. I assume that you are mainly addressing the second group of models, so please clarify it.

L 102-107: Here you should specify that the first group focus mostly on predicting large-scale regional fire dynamics, whereas the second group focus more on predicting fire in individual grid cells.

Chapter 2.2: The text might be easier to understand if you draw the network structure as a figure including all input variables, the hidden layers, neurons and output.

L 163-171: The description of the training of DNN-fire-GFED is not completely clear. From the text it reads that only the weights were readjusted by using observed GFED data. Does that mean that original bias parameters from DNN-Fire-BASE were kept? Is there any reasoning?

L 180: "spunup"

L 197-201: The readability would be improved if each equation is in a new line and not within the text line.

L 244: Should this be Figure 7? (Wu et al., 2021; Drüke et al., 2019)

L 273-275: Yes, but not many process-based fire models have been really calibrated. It would be good to provide examples in the text where this has been done.

L 276-277: The statement is not really valid as you do not calibrate the parameters of the process-based model but of the DNN-based model.

L 332-334: I do not understand this sentence because you previously wrote that you were training models for different regions and not a global model. Please clarify.

Table 1: It would be good to combine the columns data source and reference in one column. Otherwise it seems odd because population density and GDP do not have a data source.

Figure 1: check "burn" area

Figures 3, 5, 6: I recommend to combine these figures in one figure (with 4 columns per region) in order to directly compare the experiments in one plot. In addition, it would be good to also draw in a same way boxplots or violin plots of monthly burned area in order

to check if the different experiments capture the statistical distribution of fire.

Figure 4: This figure includes a lot of spatial aggregation. Can you draw a density scatter plot of the original monthly data in the used 1.9 x 2.5° resolution?

Figure 7 b: Is this a global averaged seasonal cycle? How do the seasonal cycles look like in different GFED regions?

References

Drüke, M., Forkel, M., Bloh, W. von, Sakschewski, B., Cardoso, M., Bustamante, M., Kurths, J., and Thonicke, K.: Improving the LPJmL4-SPITFIRE vegetation–fire model for South America using satellite data, Geosci. Model Dev., 12, 5029–5054, https://doi.org/10.5194/gmd-12-5029-2019, 2019.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrol. Sci. J., 31, 13–24, https://doi.org/10.1080/02626668609491024, 1986.

Lasslop, G., Hantson, S., Harrison, S. P., Bachelet, D., Burton, C., Forkel, M., Forrest, M., Li, F., Melton, J. R., Yue, C., Archibald, S., Scheiter, S., Arneth, A., Hickler, T., and Sitch, S.: Global ecosystems and fire: Multi‐□□model assessment of fire□□induced tree□□cover and carbon storage reduction, Glob. Change Biol., https://doi.org/10.1111/gcb.15160, 2020.

Wu, C., Venevsky, S., Sitch, S., Mercado, L. M., Huntingford, C., and Staver, A. C.: Historical and future global burned area with changing climate and human demography, One Earth, https://doi.org/10.1016/j.oneear.2021.03.002, 2021.