Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2021-83-RC1, 2021

Comment on gmd-2021-83

Joe Melton (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-RC1, 2021

Zhu et al. develop a machine learning (ML) burnt area model that can be used in place of a process-based algorithm in ELM. This approach was first used to surrogate the fire model of Li et al. which was in CLM (and then now ELM). The ML approach uses a deep neural network to reproduce the process model result (they call it Base). Then by altering the parameters they tuned it to match GFED4 burned area. The paper is clearly written and results are generally well presented. I found the work interesting as this is an important problem. Present process-based fire models are not overly skillful. Much of this stems from the many complexities of fire modelling - especially anthropogenic influences. I am optimistic this paper can be published but I would like to see some careful consideration of my comments below. At present the manuscript is what I would consider an absolute bare minimum of what can be published and there are many opportunities to make this paper into a much better resource to the community. This particular approach could be valuable but I think it needs some expansion to demonstrate how useful imbedding ML approaches in process models can be. As a result I would like to see some expansion of the work to better demonstrate the utility of the approach.

Main comments:

- The DNN-Fire model was subsequently tuned to match GFEDv4 but this is not the only burned area product available (e.g. Chuvieco et al. 2019). Indeed there are many other products now available and they don't agree so well (e.g. Padilla et al. 2015, Humber et al. 2019). I worry that by tuning the model to reproduce one dataset you may get a result closer to that dataset but at the expense of adopting its same biases and thereby potentially not getting as admirable advances in accuracy at it seems. Why not consider all of the available burned area products to produce a burned area estimate that could then be less biased by a single dataset? As, in reality, we are most interested in increasing our predictive skill - not just reproducing an observation.
- By surrogating Base-Fire, the DNN-Fire then integrates/assumes the biases and issues apparent in ELM's simulations (e.g. too much/little biomass, too dry/wet soil, etc.) and

produces a model that aims to get the right result (burned area matching GFED) potentially for the wrong reasons (based on biased inputs). Why not run an ensemble approach with different forcing datasets (e.g. met forcing of CRUJRA in addition to GSWP3, or a different land cover (if using prescribed), etc.) to try and give at least a measure of the uncertainty in these inputs to the DNN? We have found for our model (run in normal process-based mode) the results can be surprising and have some strong impacts for certain variables. Gitta Lasslop looked at this too and found a large impact upon fire, primarily due to the wind speed differences (e.g. Fig 3 in Lasslop et al. 2014). Alternatively using an observation-based product of one of the ELM variables (Table 1) like soil wetness or above ground biomass as another means to look at the influence of input bias.
- Around line 188 you describe the training/testing split. This approach of doing it randomly makes me wonder if the influence of spatial autocorrealtion will result in an overly optimistic error estimate. Especially as fire is likely autocorrelated. There are many papers in the literature discussing the dangers of random sampling on spatially correlated data (e.g. Roberts et al. 2017; Meyer et al. 2019; Ploton et al. 2020; Kühn and Dormann, 2012). I would suggest an alternate strategy be employed. It also wasn't clear how this test/train split results were integrated. I think it was just in the model score?
- What is the impact of training on such a short timeseries of fire observations when some regions have fire return intervals of >100 years? Also how representative are those years chosen? Would it matter if you instead trained on 2006 - 2015 and tested on 2001 - 2005?

Small comments:

- Figure 7 is the same as the years trained upon so there is little interesting information here. Basically this is showing that it can do an ok job when tested over the same training region. Why not expand this out beyond the satellite era? How does this do from say 1900 on? Yes there is no satellite data but there are other means to check results (see e.g. Arora and Melton 2018)
- Didn't GFEDv4 offer some uncertainy bounds?
- Fig 8 to make a stronger demonstration that this is a signifcant improvement, what about plotting the models of FireMIP as further reference points? E.g. Hantson et al. 2020.
- line 41, a more up to date reference would be Lasslop et al. 2020 as it was done with more advanced models
- l 90: A good reference could be Rabin et al. 2017 as there are some figures showing explictly how the models differ.
- l 186 - to be clear, the 14 submodels were combined to produce the global estimates right? Would there be benefit from doing even more sub-regions? What about 20, 50, etc? Where are the diminishing returns here?
- L 276 - was this talking about the speed of creating DNN-Fire or DNN-Fire-GFED? Several minutes on a laptop? HPC?
- Fig 8 - it seems that DNN-Fire-GFED might be less variable than GFEDv4, is that correct? Is this due to the inputs to the ML or is it a result of the ML approach itself?

Lit. cited:

Chuvieco, E., Mouillot, F., van der Werf, G. R., San Miguel, J., Tanase, M., Koutsias, N., García, M., Yebra, M., Padilla, M., Gitas, I., Heil, A., Hawbaker, T. J., and Giglio, L.: Historical background and current developments for mapping burned area from satellite Earth observation, Remote Sens. Environ., 225, 45–64, 2019.

Padilla, M., Stehman, S. V., Ramo, R., Corti, D., Hantson, S., Oliva, P., Alonso-Canas, I., Bradley, A. V., Tansey, K., Mota, B., Pereira, J. M., and Chuvieco, E.: Comparing the accuracies of remote sensing global burned area products using stratified random sampling and estimation, Remote Sens. Environ., 160, 114–121, 2015.

Humber, M. L., Boschetti, L., Giglio, L., and Justice, C. O.: Spatial and temporal intercomparison of four global burned area products, Int J Digit Earth, 12, 460–484, 2019.

Lasslop, G., Thonicke, K., and Kloster, S.: SPITFIRE within the MPI Earth system model: Model development and evaluation, J. Adv. Model. Earth Syst., 6, 740–755, 2014.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F. and Dormann, C. F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, Ecography , 40(8), 913–929, 2017.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S. and Pélissier, R.: Spatial validation reveals poor predictive performance of large-scale ecological mapping models, Nat. Commun., 11(1), 4540, 2020.

Kühn, I. and Dormann, C. F.: Less than eight (and a half) misconceptions of spatial analysis, J. Biogeogr., 39(5), 995-998, 2012.

Meyer, H., Reudenbach, C., Wöllauer, S. and Nauss, T.: Importance of spatial predictor variable selection in machine learning applications - Moving from data reproduction to spatial prediction, Ecol. Modell., 411, 108815, 2019.

Hantson, S., Kelley, D. I., Arneth, A., Harrison, S. P., Archibald, S., Bachelet, D., Forrest, M., Hickler, T., Lasslop, G., Li, F., Mangeon, S., Melton, J. R., Nieradzik, L., Rabin, S. S., Prentice, I. C., Sheehan, T., Sitch, S., Teckentrup, L., Voulgarakis, A., and Yue, C.: Quantitative assessment of fire and vegetation properties in historical simulations with fire-enabled vegetation models from the Fire Model Intercomparison Project, Biogeosciences, https://doi.org/10.5194/gmd-2019-261, 2020.

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. Chang. Biol., https://doi.org/10.1111/gcb.15160, 2020.

Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., Kaplan, J. O., Li, F., Mangeon, S., Ward, D. S., Yue, C., Arora, V. K., Hickler, T., Kloster, S., Knorr, W., Nieradzik, L., Spessa, A., Folberth, G. A., Sheehan, T., Voulgarakis, A., Kelley, D. I., Prentice, I. C., Sitch, S., Harrison, S., and Arneth, A.: The Fire Modeling Intercomparison Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions, Geoscientific Model Development, 10, 1175–1197, 2017.

Arora, V. K. and Melton, J. R.: Reduction in global area burned and wildfir emissions since 1930s enhances carbon uptake by land, Nat. Commun., 9, 1326, 2018.