Title: Physics-guided machine learning of wildfire methane emissions

Authors/Affiliations:

Fa Li, and Min Chen
University of Wisconsin-Madison

Qing Zhu, and Kunxiaojia Yuan
Lawrence Berkeley National Laboratory, Climate & Ecosystem Sciences Division

Focal Area(s):

- A solution to a key challenge in implementing advanced statistical approaches as it pertains to the methane cycle.
- Identifying high potential datasets and how advanced statistical and numerical methods can be used to realize new scientific insights.
- How automated or real-time data capture and processing can be used to address issues of spatial and temporal heterogeneity and data sparsity.

Science or Technological Challenge:

Global warming has significantly increased megafire risks, which are important sources of methane (such as the megafire over Indonesia in 1997) (Turner et al., 2019), one of the most important greenhouse gases. However, estimates of wildfire-induced methane emissions from either bottom-up (BU) or top-down (TD) models remain highly uncertain due to the imperfect model structures and limited data constraints. This would bring a big challenge to understanding and projecting the future climate change.

Rationale:

Methane emissions from wildfires are typically estimated using two primary approaches: Bottom-Up (BU) and Top-Down (TD) models. BU models rely on observations of combustion completeness (CC) and emission factors (EFs) for different plant functional types, which are often based on limited site observations. These estimates typically do not account for spatiotemporal variability in CC and EFs across different environments. However, BU models can provide high-resolution estimates of emissions.

On the other hand, TD models rely on atmospheric methane concentration measurements from towers or satellites, which are then used to estimate ground emissions from fires through inverse atmospheric transport modeling. TD models are typically more reliable at coarser spatial and temporal scales. TD models are often combined with BU models to estimate wildfire emissions, as BU models can provide important prior information for the TD approach.

To accurately estimate methane emissions from past and future fires, it is crucial to efficiently integrate data from various sources, such as ground, tower, and satellite observations, to constrain a coupled BU-TD model. Additionally, BU models need to be well parameterized for future projections. Traditionally, BU and TD models are parameterized separately, which requires a large number of time-consuming ensemble runs of the models. Furthermore, data assimilation algorithms such as ensemble Kalman Filter used to constrain these models often assume linear relationships between observations and model state variables (e.g., parameters), which may not be true.

In this white paper, we identified a few points that machine learning could be used to improve estimation and projecting methane emissions due to wildfires by integration with BU and TD models.

Narrative:

Machine learning (ML) help solve the above-mentioned problems in several ways.

First, ML can be used to create surrogate models, which represents a physics-guided machine learning that approximate the behavior of BU and TD models but with grand reduction of the computational cost, making the data assimilation more efficient. For example, Zhu et al., (2022) built a deep neural network (DNN) scheme that surrogates the process-based wildfire model within the Energy Exascale Earth System Model (E3SM) and the surrogate wildfire model successfully captured the observed regional burned area. Such kind of models can be straightforwardly extended for the purpose of simulating methane emission due to fires.

Second, ML can also be used to learn and provide a more accurate representation of the relationship between observations and model variables. Satellite/tower observations usually only provide column methane concentrations at, near or far away from the locations where fires happen, and the relationships between the atmospheric column methane concentration, the fire-emitted methane, the emission factors/combustion completeness are determined by the fire burning and atmospheric transport processes as described in the BU and TD models, which are often highly nonlinear. ML can help identify and model these non-linear relationships, which can improve the accuracy of data assimilation (Abarbanel et al., 2018). ML-based data assimilation is being used in weather forecasting but has not been applied for estimating methane emissions yet.

Additionally, ML can help with the selection and weighting of observations in the data assimilation process (Geer 2021). Traditional data assimilation methods often assume that all observations are equally important, but this may not always be the case. ML can help identify which observations are most important for improving model accuracy and assign appropriate weights to them.

References

Henry D. I. Abarbanel, Paul J. Rozdeba, Sasha Shirman; Machine Learning: Deepest Learning as Statistical Data Assimilation Problems. *Neural Comput* 2018; 30 (8): 2025–2055. doi: https://doi.org/10.1162/neco_a_01094

Geer A. J. (2021) Learning earth system models from observations: machine learning or data assimilation? Phil. Trans. R. Soc. A.3792020008920200089

Turner, A. J., Frankenberg, C., & Kort, E. A. (2019). Interpreting contemporary trends in atmospheric methane. Proceedings of the National Academy of Sciences, 116(8), 2805-2813.

Zhu, Q., Li, F., Riley, W. J., Xu, L., Zhao, L., Yuan, K., ... & Randerson, J. (2022). Building a machine learning surrogate model for wildfire activities within a global Earth system model. Geoscientific Model Development, 15(5), 1899-1911.