

Title: Accelerated trait-based modeling of biogenic methane dynamics using physics guided machine learning

Authors/Affiliations: Jinyun Tang, Qing Zhu, William J. Riley, Eoin Brodie
Lawrence Berkeley National Laboratory, Climate & Ecosystem Sciences Division.

Focal Area(s): (1) Key uncertainties and knowledge gaps where new methodology, infrastructure, or technology can advance predictive understanding of the methane cycle. (2) A solution to a key challenge in implementing AI approaches (e.g., improving uncertainty quantification, federated learning) across the biological and environmental science domains as it pertains to the methane cycle.

Science or Technological Challenge:

Global methane (CH_4) emissions are dominated by biogenic sources resulting from the interplay between production by methanogens and consumption by methanotrophs. While process-based models exist and have been applied for a long time, they frequently fail to accurately capture the response of net CH_4 emissions to variations in environmental factors such as temperature, moisture, and pH. The explicit representation of microbial dynamics has been suggested to improve these models. However, determining how much complexity should be represented in these microbial models is difficult because both CH_4 production and oxidation are carried out by diverse groups of microbes that interact and compete with each other. Trait-based modeling approaches have been proposed to represent the diversity of microbes within a microbial community and their effects on CH_4 biogeochemistry. However, this approach becomes challenging due to the large computational costs for parameterization when more microbes are represented. Moreover, the high computational costs make it challenging to incorporate empirical observations, constrain model parameterization, and quantify modeling uncertainty conditioned on current knowledge in measurements and modeling.

Rationale:

The high computational costs associated with trait-based models are largely due to the high computing cost of the numerical solvers used to integrate the differential equations over space and time. This high computational cost, in turn, makes the process of improving model parameterization through model-data fusion more challenging, as it often requires numerous iterations of calibration with large ensemble simulations. Machine learning (ML) has demonstrated the potential to significantly speed up forward model simulations in areas such as weather and climate modeling and computational fluid dynamics (e.g., Scher, S., Messori, 2019; Weyn et al., 2019; and Kochkov et al., 2021). By creating high-fidelity surrogates of trait-based models using ML, we can accelerate both forward and calibration simulations, allowing for efficient quantification and reduction of parametric uncertainties. Furthermore, the ease of computing derivatives with respect to parameters makes it easier to fine-tune the ML-based surrogate models by incorporating a wider range of data. Finally, by building surrogates of trait-based models with different levels of complexity, we can quantify the relationship between

model complexity and predictive uncertainty, and determine the optimal level of model complexity needed to predict future biogenic CH₄ dynamics.

Narrative:

Our objective is to create a framework that combines (1) a synthetic database of CH₄-related biochemical variables, generated by the microbial modules of EcoSIM (the land model being developed for BioEPIC, originally based on *ecosys* (Grant et al. 2017)) with varying levels of parameterization complexity in microbial dynamics; (2) machine learning surrogates trained using simulations from each complexity configuration; and (3) a model-data fusion framework that incorporates various observations to refine model parameters through the surrogates. To maintain interpretability, we will use a physics-guided machine learning approach, as demonstrated in our recent studies (Liu et al., 2022; Yuan et al., 2022). By repeatedly integrating these three components, we can continuously improve the microbial module of EcoSIM and its surrogates, and assess the impact of observations on model predictions. Finally, the resulting observationally-constrained surrogates will be used for ensemble extrapolation in various scenarios, quantifying uncertainty across different levels of complexity, and determining the optimal complexity for robust CH₄ dynamics predictions.

Reference

- Kochkov, D., Smith, J.A., Alieva, A., Wang, Q., Brenner, M.P., Hoyer, S., 2021. Machine learning-accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences of the United States of America* 118.
- Liu, L.C., Xu, S.M., Tang, J.Y., Guan, K.Y., Griffis, T.J., Erickson, M.D., Frie, A.L., Jia, X.W., Kim, T., Miller, L.T., Peng, B., Wu, S.W., Yang, Y.F., Zhou, W., Kumar, V., Jin, Z.N., 2022. KGML-ag: a modeling framework of knowledge-guided machine learning to simulate agroecosystems: a case study of estimating N₂O emission using data from mesocosm experiments. *Geoscientific Model Development* 15, 2839-2858.
- Scher, S., Messori, G., 2019. Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground. *Geoscientific Model Development* 12, 2797-2809.
- Weyn, J.A., Durran, D.R., Caruana, R., Cresswell-Clay, N., 2021. Sub-Seasonal Forecasting With a Large Ensemble of Deep-Learning Weather Prediction Models. *Journal of Advances in Modeling Earth Systems* 13.
- Yuan, K., Zhu, Q., Li, F., Riley, W.J., Torn, M., Chu, H., McNicol, G., Chen, M., Knox, S., Delwiche, K. and Wu, H., 2022. Causality guided machine learning model on wetland CH₄ emissions across global wetlands. *Agricultural and Forest Meteorology*, 324, p.109115.