

# **DOE Whitepaper: Elucidating environmental regulation on microbial-mediated soil methane emission using gene-to-ecosystem level data and artificial intelligence**

**Yang Song, Department of Hydrology and Atmospheric Sciences, University of Arizona**

**Focal Areas:** This white paper is framed around two focal areas (1) The importance of high-potential datasets (omics data) and the combination of multiscale data across spatial and temporal scales may lead to new scientific insights. (2) Key uncertainties and knowledge gaps where new methodology, infrastructure, or technology can advance predictive understanding of the methane cycle.

**Science or Technological Challenges:** Methane ( $\text{CH}_4$ ) is the second most abundant anthropogenic greenhouse gas (GHG) after carbon dioxide ( $\text{CO}_2$ ), accounting for about 20% of global emissions. Although gene-level analysis, ground-level observations, and global-scale earth system mode have separately advanced in interpreting microbial regulation on methane ( $\text{CH}_4$ ) emission and benchmark global methane emission prediction, the accurate estimation of global  $\text{CH}_4$  sources and sinks remains a significant challenge. **Challenge 1:** Soil  $\text{CH}_4$  emission is driven by specific soil microbes and their released enzymes. However, the spatial distribution of  $\text{CH}_4$  emission-associated microbial functions and their interactions with other microbial functions is still unclear. This knowledge gap limits our ability to employ site-level scientific findings to interpret ecosystem-level methane emission uncertainty. **Challenge 2:** Soil  $\text{CH}_4$  emissions highly fluctuate with spatial heterogeneity and temporal change in multiple environmental factors. Interpreting the nonlinear regulation of multiple environmental factors on soil  $\text{CH}_4$  emissions is still difficult, especially considering the potential acclimation and adaptation of soil microbial communities to changing environments. The lack of this knowledge brings significant uncertainty in projecting soil  $\text{CH}_4$  emissions under climate change. **Challenge 3:** Although there are increased efforts to represent microbial-mediated soil  $\text{CH}_4$  emission schemes in the Earth system models (ESMs), the scale difference between the earth system model and mechanistic understanding of the  $\text{CH}_4$  emission process at the gene or lab scale makes it challenging to utilize emerging gene-scale observations to parameterize microbial-mediated soil  $\text{CH}_4$  emission schemes in ESM.

**Rationale:** Overcoming the above challenge requires performing a regional or global-scale investigation of the distribution of  $\text{CH}_4$  emission-associated microbial functions and elucidating how the relative abundances of these microbial functions for  $\text{CH}_4$  emission vary with spatiotemporal change in environmental conditions. Although the outbreak of omics technology has brought data to investigate this question, the spatiotemporal representation of these data is still limited. The interpretation of environmental regulation on omics-informed gene function associated with  $\text{CH}_4$  emission is highly very with the sampling location. In my previous work, we have harnessed the power of artificial intelligence (AI) and omics data to map soil microbial function involved in soil organic matter decomposition<sup>1</sup>. This study highlights the possibility of synthesizing global-scale omics data to identify microbial functions associated with  $\text{CH}_4$  emission and integrating this information with the corresponding environmental information to predict the spatiotemporal dynamics of microbial functions for  $\text{CH}_4$  emission in response to environmental change. Moreover, to enable the utilization of this gene-scale environmental regulation on microbial  $\text{CH}_4$  emission function to advance soil  $\text{CH}_4$  flux simulation in ESM, an

effective scaling methodology is required. Microbial-mediated CH<sub>4</sub> emission is an enzyme-catalyzed process and can be calculated using the Michaelis-Menten equation as a function of enzyme abundance, substrate concentration, and kinetics parameters. Therefore, it's possible to utilize the Michaelis-Menten equation to integrate AI prediction of CH<sub>4</sub> emission enzyme functional information with the process-based CH<sub>4</sub> fluxes simulation in ESM.

**Narrative:** To overcome current challenges for elucidating the uncertainty from soil CH<sub>4</sub> emission, here I propose to develop an integrated research framework that harnesses the power of gene-to-ecosystem scale data, applies AI technology for environment-microbial function prediction and model parameter optimization, and mechanistically advances the representation of soil CH<sub>4</sub> emission in ESM. In detail, this research frame will need to include (1) Identifying the spatial distribution of microbial enzymes associated with CH<sub>4</sub> emission by integrating omics and environmental information across diverse sites to develop an AI prediction; (2) Elucidating environmental regulation on the dynamics of soil enzymes for CH<sub>4</sub> emissions by integrating temporal omics and environmental data to train an AI model for predicting the response of soil enzyme functional composition in response to temporal environmental change. (3) Assessing the implication of the environmental regulation on CH<sub>4</sub> emissions enzyme for soil CH<sub>4</sub> emission by coupling AI-prediction for microbial function for CH<sub>4</sub> emission in response to environmental change with process-based CH<sub>4</sub> fluxes simulation in E3SM land model (ELM). The implementation of this research framework will deliver an integrated dataset that pairs microbial CH<sub>4</sub> function information with the corresponding environmental information at the global scale. Applying this dataset to AI prediction for the environmental feedback of microbial CH<sub>4</sub> function information will enable us to elucidate climate, edaphic, and vegetation regulation of the composition and abundances of microbial enzyme functions involved in CH<sub>4</sub> emission at the regional or global scale. Besides this AI application, a surrogate-based AI model will also be employed to optimize new parameters used in omics-informed soil emission simulation in ELM. Applying coupled AI model and ELM prediction will leverage omics data applications in constraining uncertainty in soil CH<sub>4</sub> emission and provide a deep insight into microbial-mediated soil CH<sub>4</sub> emission under changing environments. The success of this study will advance the DOE BER program by overcoming current technical bottlenecks in gene-to-Earth system prediction for the global methane cycle and advance the DOE E3SM model capacity for predicting soil methane emission under more extreme climate conditions.

## References:

- 1 Fan, C., Song, Y., Mishra, U. & Sagar, G., Mayes M. Harness the power of machine learning and omics to identify microbial functional composition across diverse environments. *JGR Biogeoscience* (In review).