(72-240416-C) **Estimating the Global Methane Soil Sink using Knowledge-guided Machine Learning**

C. Smith[1,2], L. Liu[3], and Y. Oh[1,2]

[1]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO 80309; 720-295-0538, E-mail: chris.c.smith@noaa.gov
[2]NOAA Global Monitoring Laboratory (GML), Boulder, CO 80305
[3]Purdue University, West Lafayette, IN 47907

We are estimating the spatial and temporal variability in global CH4 uptake using a knowledge-guided machine learning (KGML) framework. This framework combines process-based and machine-learning models, and synthesizes multi-source direct and indirect measurements of soil CH4 oxidation to improve model training, interpretability, and accuracy across spatial and temporal scales. Natural CH4 oxidation by microbes in upland soils is the second largest sink in the global CH4 budget, but its importance is not fully understood. The magnitude and long-term trends of global CH4 soil sinks are highly uncertain due to overlooked microbial processes and contradicting model outputs. Accurately quantifying global CH4 soil sinks is extremely important to reduce biases in current and future global CH4 budgets. We use a process-based model as the scientific foundation of the KGML hierarchical structure and to generate millions of synthetic data for pre-training. We build separate machine-learning submodules for soil thermal, hydrological, and biogeochemical processes, and an overarching model structure to link the submodules. The key biogeochemical constraints (e.g. soil CH4 substrate, temperature, and moisture influences) are carefully embedded into the cost function using known principles and empirical functions as knowledge-guided losses. The KGML model will be trained/validated with direct measurements of soil CH4 oxidation fluxes from FLUXNET-CH4 and chamber measurements. Using global soil moisture and temperature data, we further optimize the model to capture temporal and spatial heterogeneity. The well-constrained KGML model will ultimately be extrapolated to the global scale and be used to generate new global CH4 soil sink products at daily and 4-km resolution from 1984 to 2022.