Integrating Process-Based and Machine Learning Approaches for Estimating the Global Methane Soil Sink

Chris Smith[1,2], Licheng Liu[3], Qing Zhu[4], Gavin McNicol[5], Sparkle Malone[6], Zhenong Jin[3], Youmi Oh[1,2]

(1) Cooperative Institute for Research in Environmental Studies, (2) NOAA/ESRL Global Monitoring Laboratory, (3) Department of Bioproducts and Biosystems Engineering, University of Minnesota, (4) Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, (5) Department of Earth and Environmental Sciences, University of Illinois Chicago, Chicago, (6) Yale School of the Environment, Yale University

Introduction

Natural methane (CH_4) oxidation by microbes in upland soils is the second largest sink in the global CH_4 budget (Saunois et al. 2020), but its magnitude and long-term trend are uncertain (Conrad 2009; Belova et al. 2020).

- The current estimate of the global CH_4 soil sink based on meta-analysis, ~30 $TgCH_4yr^{-1}$, has large uncertainty (Dutaur and Verchot 2007; Smith et al. 2000; Saunois et al. 2020).
- The long-term trends in the global CH_4 soil sink are highly uncertain as well. (Ni and Groffman 2018; Gatica et al. 2020).

Methodology	Reference	Global uptake by soils ($Tg\,CH_4\,yr^{-1}$)
Observation	Dörr et al. (1993)	28.7
Observation	Smith et al. (2000)	29 (7 to >100)
Observation	Dutaur and Verchot (2007)	36 ± 23
Atmospheric inversions	Hein et al. (1997)	30 ± 15
Model (P96)	Potter et al. (1996)	20 ± 3
Model (R99)	Ridgwell et al. (1999)	38.1 ± 1.1
Model	Spahni et al. (2011)	38.9
Model (C07)	Curry (2007)	29.3 ± 0.6
Model	Zhuang et al. (2013)	34 ± 2
Model (MeMo)	Murguia-Flores et al. (2018)	33.5 ± 0.6

Table 1. Global CH_4 soil sink estimated by observation-based meta-analyses and process-based models. Table modified from Murguia-Flores et al., 2018.

Previous studies estimated natural CH_4 using either process-based or machine-learning approaches. Both methods have benefits and limitations.

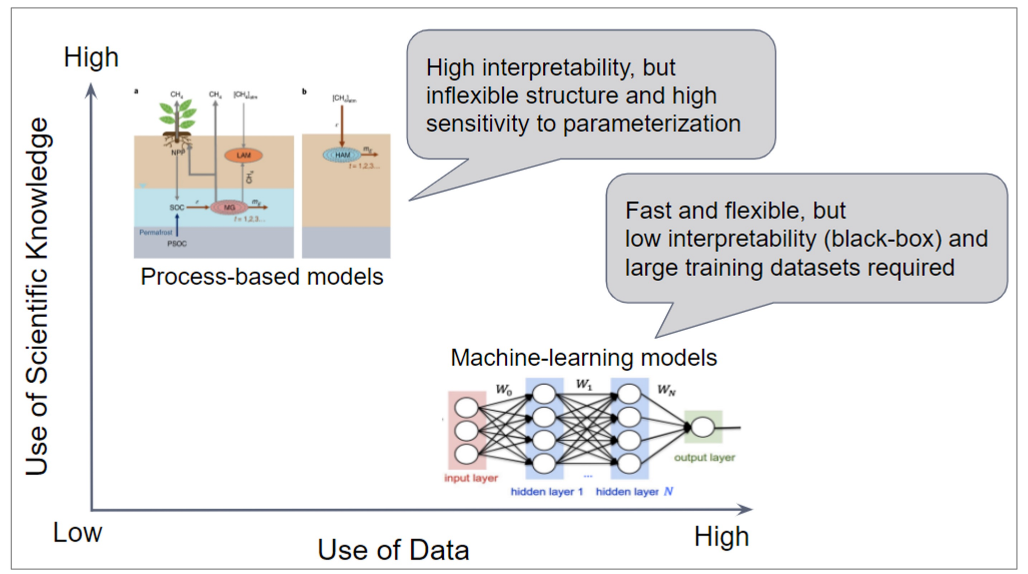

Figure 1. Benefits and limitations of process-based and machine-learning models.

Knowledge-Guided Machine Learning (KMGL)

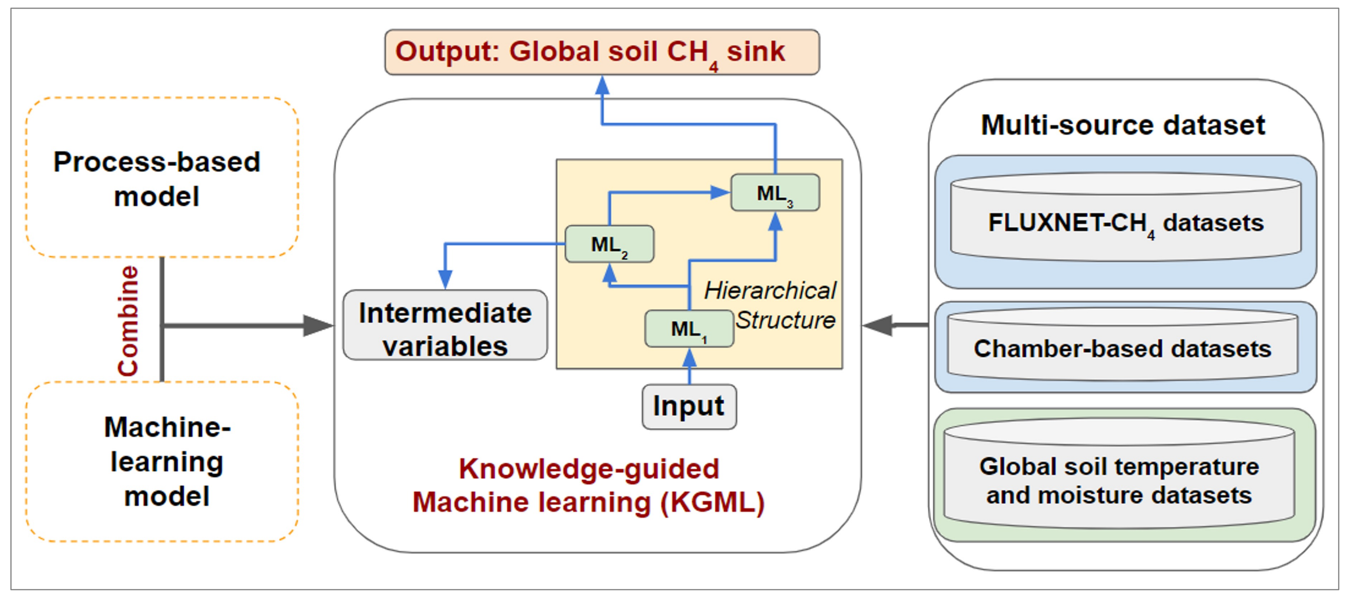

Figure 2. KGML framework for estimating global CH_4 soil sink.

- The synergy of process-based and machine-learning modeling, which combines the strengths of each approach, has gained attention in earth science fields and is called 'Hybrid modeling' or 'KGML' (ElGhawi et al. 2022; Daw et al. 2017).
- Despite the early success of the KGML approach (e.g., Liu et al. 2023), research in combining process-based and machine-learning models in terrestrial biogeochemical ecosystems is still at a nascent stage.
- We are developing a novel KGML framework that synthesizes process-based and machine-learning models, as well as multi-source direct and indirect observations of soil CH_4 oxidation, to quantify the spatial and temporal variability of the global CH_4 soil sink.

Results of Process-based modeling

- We used a model with microbial dynamics including High- and Low-Affinity Methanotrophs (HAM and LAM, respectively) (Zhuang et al., 2004; Oh et al. 2020) and optimized explicit methane oxidation processes for regions with 8 different vegetation types.

No.	Vegetation	Reference for Uplands
1	Alpine Desert	*Jørgensen et al., 2015*
2	Wet Tundra	*D'Imperio et al., 2016*
3	Boreal Forest	*Dinsmore et al., 2017*
4	Temperate Forest	*Castro et al., 1995*
5	Grasslands	*Jones et al., 2017*
6	Shrubland	*Castaldi and Fierro, 2005*
7	Tropical Forest	*Davidson et al., 2008*
8	Woodlands	*Crill et al., 1991*

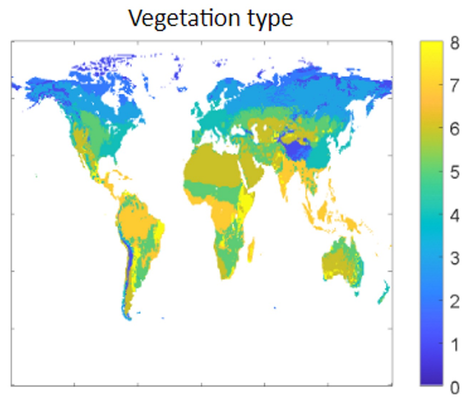

Figure 3. (left) Summary of 8 vegetation types and references where optimized data are from, and (right) a spatial map of vegetation types.

- Methane oxidation was higher in low soil moisture and high temperature conditions, and was especially sensitive to temperature in boreal forests. While there are field studies showing that desert methane soil sink can be up to 7 Tg/yr, additional investigation is needed to improve our simulation for this region.
- Since HAM prefer low SOC and pH conditions and may not be dominant for all upland soils, we conducted sensitivity tests to change the HAM and LAM dominant regions based on different SOC and pH criteria (Fig. 4-5).

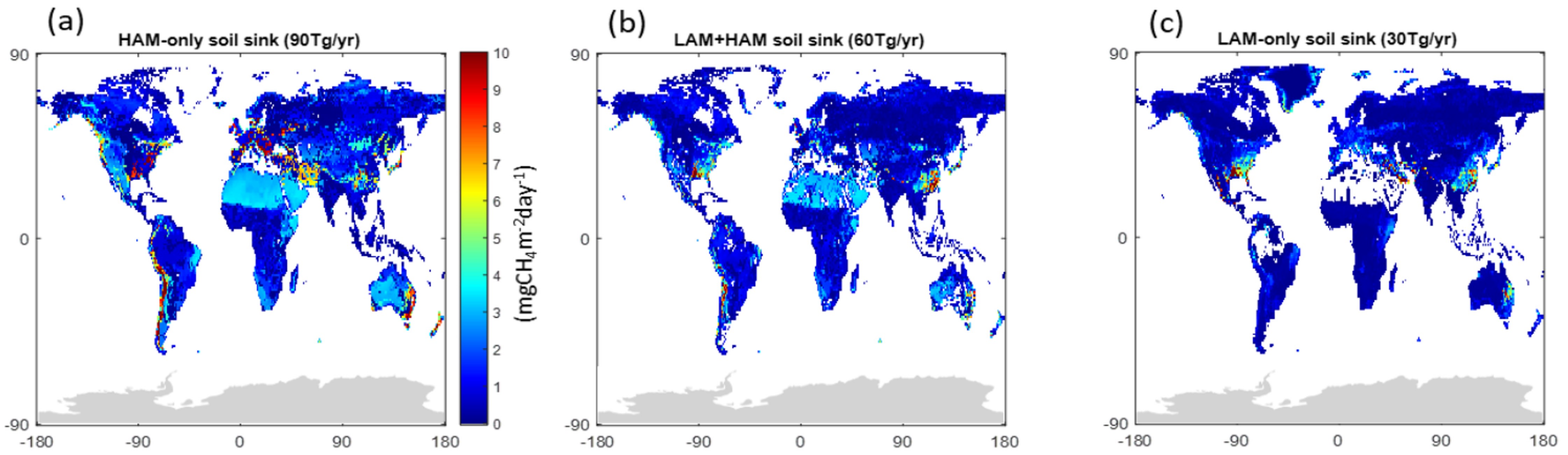

Figure 4. Global map of soil sink for (a) HAM-only, (b) HAM+LAM, and (c) LAM-only scenarios based on the SOC threshold in Fig. 5.

- The preliminary results show that there are large uncertainties in process-based estimation (30-90 $Tg_{CH4}yr^{-1}$) due to parameter optimization and governing microbial processes.

Criterion 1 ($Tg_{CH4}\ yr^{-1}$) – max. SOC threshold	HAM only	SOC 5%	SOC 1%	SOC 0.5%	LAM only
	90	90	73	60	33
Criterion 2 ($Tg_{CH4}\ yr^{-1}$) – max. pH threshold	HAM only	pH 6	pH 7	pH 8	LAM only
	90	76	64	46	33

Figure 5. Sensitivity test of process-based model using soil organic carbon (SOC) and soil pH criteria.

Preliminary results: machine learning pre-training

- We pre-trained a deep neural network (GRU) on data from the process-based model.
- Evaluating the model on held-out simulated data (R^2=0.91), high accuracies were achieved for well-represented vegetation types, but performed worse on some vegetation types.

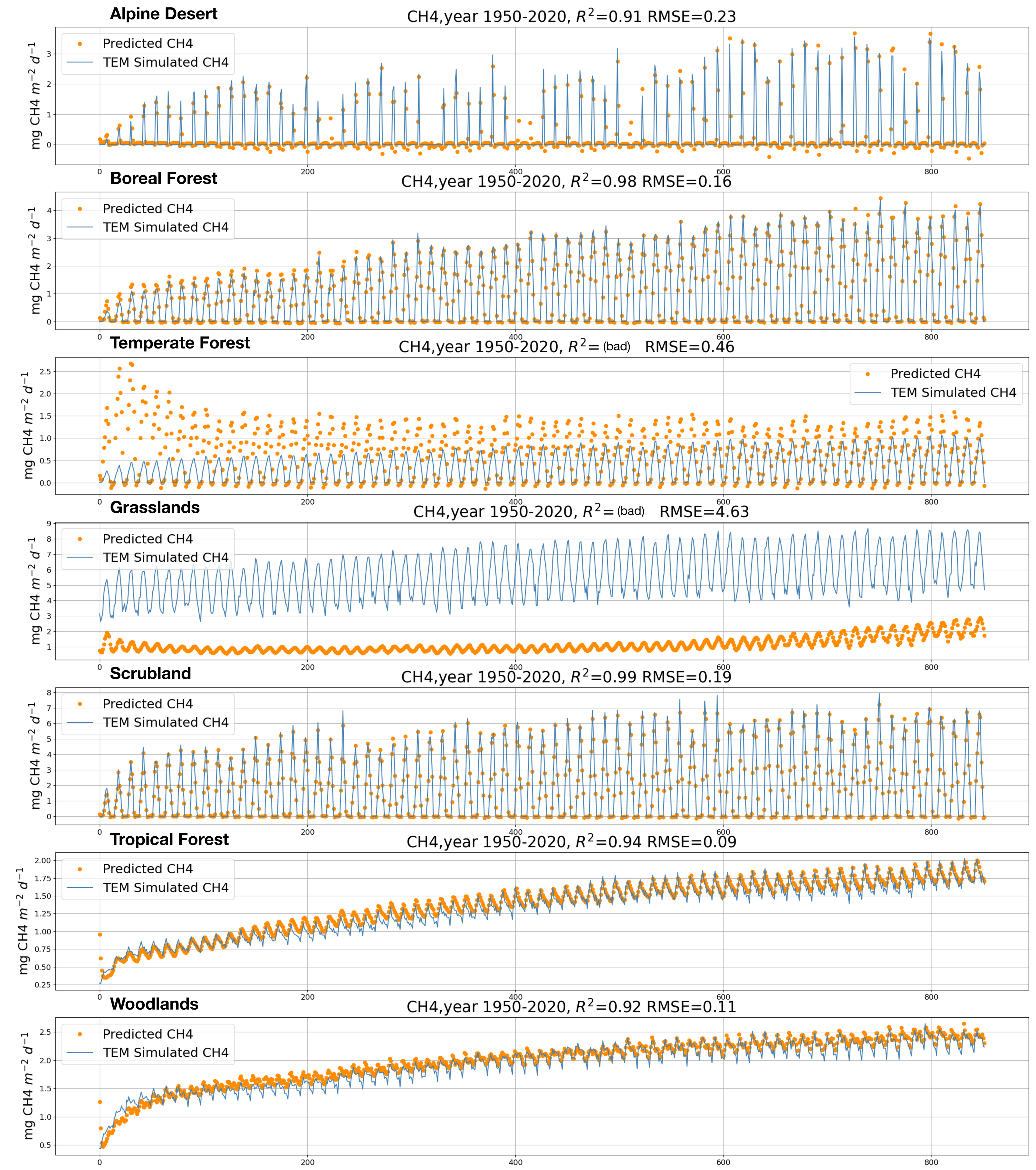

Figure 6. Evaluating pre-trained model on held-out simulated data, for different vegetation types

Preliminary results: between-site differences

- Different geographic sites with the same vegetation type show different performance.

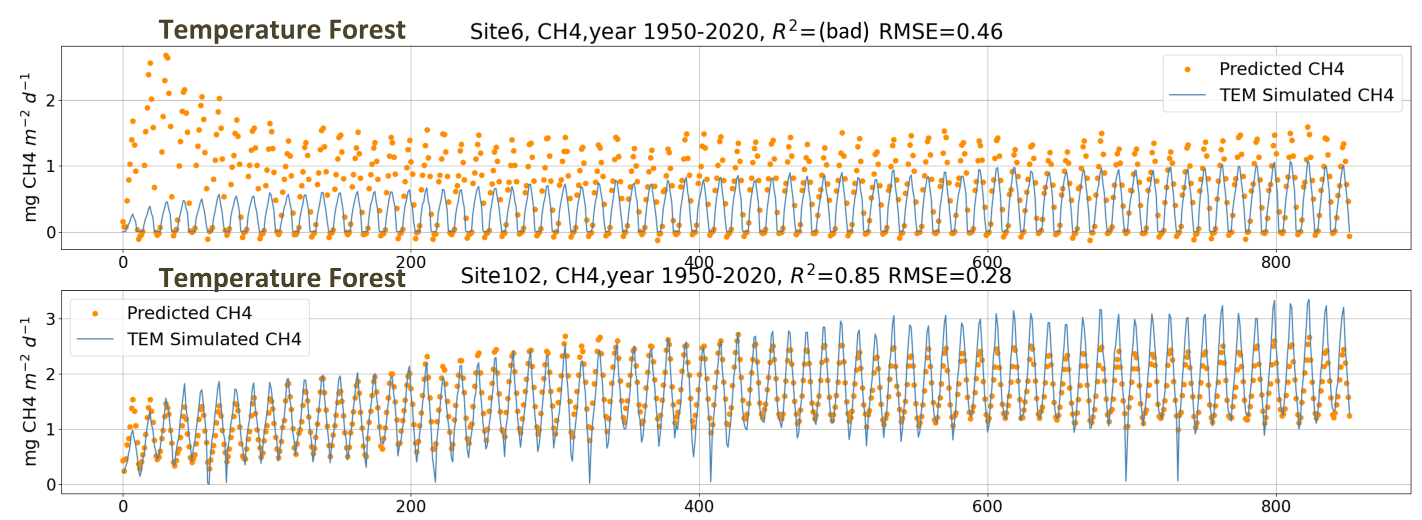

Figure 7. Evaluating pre-trained model on held-out simulated data, for different Temperate Forest sites.

References

- Belova, Svetlana E., et al. "Methane-Oxidizing Communities in Lichen-Dominated Forested Tundra Are Composed Exclusively of High-Affinity USCα Methanotrophs." *Microorganisms*, vol. 8, no. 12, Dec. 2020, https://doi.org/10.3390/microorganisms8122047.
- Conrad, Ralf. "The Global Methane Cycle: Recent Advances in Understanding the Microbial Processes Involved." *Environmental Microbiology Reports*, vol. 1, no. 5, 2009, pp. 285–92.
- Daw, Arka, et al. "Physics-Guided Neural Networks (pgnn): An Application in Lake Temperature Modeling." *Knowledge-Guided Machine Learning*, Chapman and Hall/CRC, 2017, pp. 353–72.
- Dutaur, Laure, and Louis V. Verchot. "A Global Inventory of the Soil CH4 Sink." *Global Biogeochemical Cycles*, vol. 21, no. 4, 2007, pp. 1–9.
- ElGhawi, Reda, et al. "Hybrid Modeling of Evapotranspiration: Inferring Stomatal and Aerodynamic Resistances Using Combined Physics-Based and Machine Learning." *Earth and Space Science Open Archive*, 30 Aug. 2022, https://doi.org/10.1002/essoar.10512258.1.
- Gatica, Gabriel, et al. "Environmental and Anthropogenic Drivers of Soil Methane Fluxes in Forests: Global Patterns and Among-biomes Differences." *Global Change Biology*, vol. 26, no. 11, 2020, pp. 6604–15.
- Liu L, Zhou W, Guan K, ..., Kumar V, Jin Z. (2023). Knowledge-Guided Machine Learning can improve C cycle quantification in agroecosystems. Nature Communications. doi: 10.1038/s41467-023-43860-5
- Murguia-Flores, Fabiola, et al. "Soil Methanotrophy Model (MeMo v1.0): A Process-Based Model to Quantify Global Uptake of Atmospheric Methane by Soil." *Geoscientific Model Development*, vol. 11, no. 6, 2018, pp. 2009–32.
- Ni, Xiangyin, and Peter M. Groffman. "Declines in Methane Uptake in Forest Soils." *Proceedings of the National Academy of Sciences*, vol. 115, no. 34, 2018, pp. 8587–90.
- Oh, Youmi, et al. "Reduced Net Methane Emissions due to Microbial Methane Oxidation in a Warmer Arctic." *Nature Climate Change*, vol. 10, no. 4, 2020, pp. 317–21.
- Saunois, Marielle, et al. *The Global Methane Budget 2000 – 2017*. 2020, pp. 1561–623.
- Smith, Lesley K., et al. "Methane Emissions from the Orinoco River Floodplain, Venezuela." *Biogeochemistry*, vol. 51, no. 2, 2000, pp. 113–40.
- Zhuang, Q., et al. "Methane Fluxes between Terrestrial Ecosystems and the Atmosphere at Northern High Latitudes during the Past Century: A Retrospective Analysis with a Process-Based Biogeochemistry Model." *Global Biogeochemical Cycles*, vol. 18, no. 3, 2004, https://doi.org/10.1029/2004GB002239.

Contact Information Chris Smith, chris.c.smith@noaa.gov