JGR Biogeosciences

INVITED ARTICLE
10.1029/2025JG009668

Special Collection:
Biogeosciences Leaders of Tomorrow: JGR: Biogeosciences Special Collection on Emerging Scientists

Key Points:
- Global methane soil sink is 40–45 Tg CH_4 yr^{-1} from process-based and machine-learning approaches, 30%–50% larger than conventional estimates
- Machine-learning models exhibit smoother spatial and temporal patterns due to the limitation of observations used for model training
- A larger soil sink better reproduces observed latitudinal gradients and seasonality of atmospheric methane and carbon isotope changes

Supporting Information:
Supporting Information may be found in the online version of this article.

Correspondence to:
Y. Oh,
youmi.oh@noaa.gov

Citation:
Oh, Y., Liu, L., Lee, J., Bruhwiler, L., Lan, X., Michel, S., et al. (2026). Revising the magnitude and trends of the global methane soil sink with process-based, machine-learning, and atmospheric inversion modeling approaches. *Journal of Geophysical Research: Biogeosciences*, *131*, e2025JG009668. https://doi.org/10.1029/2025JG009668

Received 30 DEC 2025
Accepted 2 JUL 2026

Author Contributions:
Conceptualization: Youmi Oh, Licheng Liu, Jaehyun Lee
Data curation: Youmi Oh, Licheng Liu, Jaehyun Lee

Revising the Magnitude and Trends of the Global Methane Soil Sink With Process-Based, Machine-Learning, and Atmospheric Inversion Modeling Approaches

Youmi Oh[1,2], Licheng Liu[3,4], Jaehyun Lee[5], Lori Bruhwiler[2], Xin Lan[1,2], Sylvia Michel[6], Sourish Basu[7,8], John B. Miller[2], Qing Zhu[9], Sparkle Malone[10], Gavin McNicol[11], and Qianlai Zhuang[12]

[1]Cooperative Institute for Research in Environmental Science, University of Colorado, Boulder, CO, USA, [2]Global Monitoring Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA, [3]Department of Bioproducts and Biosystems Engineering, University of Minnesota, St. Paul, MN, USA, [4]Department of Biosystems Engineering, University of Wisconsin-Madison, Madison, WI, USA, [5]Climate and Environmental Research Institute, Korea Institute of Science and Technology, Seoul, Republic of Korea, [6]Institute for Arctic and Alpine Research, University of Colorado, Boulder, CO, USA, [7]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA, [8]NASA Goddard Space Flight Center, Greenbelt, MD, USA, [9]Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [10]Yale School of the Environment, Yale University, New Haven, CT, USA, [11]Department of Earth and Environmental Sciences, University of Illinois Chicago, Chicago, IL, USA, [12]Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, USA

Abstract Methane (CH_4) oxidation by microbes is the largest biological sink of global methane, yet its magnitude and long-term variability remain uncertain. Here, we combined process-based (PB), machine-learning (ML), and atmospheric inversion approaches to evaluate the global methane soil sinks and its implication for the atmospheric CH_4 budget in this study. Both PB and ML approaches estimated annual global methane soil sink to be 40–45 Tg CH_4 yr^{-1}, 30%–50% larger than conventional estimates. Although the two approaches agreed on total magnitude, they differ in their representation of variability. PB models simulate stronger spatial heterogeneity, seasonality, and long-term increases in CH_4 uptake because environmental sensitivities are explicitly represented through mechanistic equations. In contrast, ML models reproduce site-level observations more closely but exhibit muted spatial and temporal variability due to their limited environmental sensitivities from sparse and discrete observations used for model training. Atmospheric inversions further indicate that incorporating the larger soil sink improves agreement with observed atmospheric CH_4 and its stable carbon isotope changes and requires larger microbial CH_4 emissions. Together, these results suggest that the global methane soil sink may have been underestimated and demonstrate the value of integrating PB and ML modeling, and atmospheric constraints to improve understanding of global methane cycling.

Plain Language Summary Methane is a powerful greenhouse gas, and upland soils remove methane from the atmosphere through methane-oxidizing bacteria, known as methanotrophs. However, the magnitude and long-term changes of this soil methane sink remain uncertain. In this study, we combined process-based, machine-learning, and atmospheric inversion modeling to better quantify global methane uptake by soils. We find that upland soils remove about 40–45 gigatons of methane per year, more than previously estimated. While process-based and machine-learning models produce similar global totals, they differ in how they represent changes across regions and through time. Process-based models simulate stronger spatial and temporal patterns, whereas machine-learning models produce smoother patterns. The larger soil methane sinks are more consistent with large-scale atmospheric methane and carbon isotope patterns from our atmospheric inversion modeling. Together, these results show that soil methane uptake may be more significant than previously recognized, and that combining multiple approaches has improved our understanding of its spatial and temporal variability.

1. Introduction

Methane (CH_4) oxidation by microbes is the largest biological sink of global CH_4, but its importance may have been underestimated (Conrad, 2009; Tveit et al., 2019). Studies found overlooked CH_4 soil sinks in diverse terrestrial ecosystems, such as alpine tundra, forest, grasslands, tropical savanna, and the Antarctic, thought to be due to high-affinity methanotrophs (HAM; Gauci, 2025; Kato et al., 2011; Lau et al., 2015; Ortiz et al., 2021). In

Formal analysis: Youmi Oh, Licheng Liu, Jaehyun Lee
Funding acquisition: Youmi Oh, Licheng Liu
Investigation: Youmi Oh, Licheng Liu, Jaehyun Lee
Methodology: Youmi Oh, Licheng Liu, Jaehyun Lee
Project administration: Youmi Oh
Resources: Youmi Oh, Licheng Liu
Software: Youmi Oh, Licheng Liu
Supervision: Youmi Oh, Lori Bruhwiler
Validation: Youmi Oh, Licheng Liu
Visualization: Youmi Oh, Licheng Liu
Writing – original draft: Youmi Oh, Licheng Liu, Jaehyun Lee
Writing – review & editing: Youmi Oh, Licheng Liu, Jaehyun Lee, Lori Bruhwiler, Xin Lan, Sylvia Michel, Sourish Basu, John B. Miller, Qing Zhu, Sparkle Malone, Gavin McNicol, Qianlai Zhuang

dry mineral soils, the dominant CH_4 oxidizers, HAM, survive and grow at the low concentrations (~2 ppm) of CH_4 observed in the atmosphere (Baani & Liesack, 2008), and are highly temperature-sensitive compared to low-affinity methanotrophs (LAM; He et al., 2023; Juncher Jørgensen et al., 2015; Lau et al., 2015). After considering the microbial dynamics of HAM and methanogens, soil sinks could be twice the current estimate at the pan-Arctic scale and may continue to increase in the future (Oh et al., 2020).

The under-estimated global CH_4 soil sinks may partly compensate for the well-known discrepancy in the global CH_4 budgets estimated by bottom-up process-based models and top-down atmospheric inversions (Saunois et al., 2020, 2025). The bottom-up process-based models estimate CH_4 budgets that are ~100 Tg CH_4 yr^{-1} larger than those from top-down inversions, primarily due to high emissions from natural wetlands and inland freshwaters estimated by bottom-up approaches (Saunois et al., 2025). The current estimation of the global CH_4 soil sink is ~30 Tg CH_4 yr^{-1}, but with a huge uncertainty (7 to >100 Tg CH_4 yr^{-1}) from data-driven meta-analyses (Dutaur & Verchot, 2007; Smith et al., 2000).

Furthermore, there are contradicting studies on long-term trends in global CH_4 soil sinks. A measurement and meta-analysis study showed that CH_4 uptake from global forest soils decreased by 77% from 1988 to 2015, driven by increases in precipitation (Ni & Groffman, 2018). In contrast, another meta-analysis found that the relationship between CH_4 oxidation and precipitation is ecosystem-specific, with higher CH_4 oxidation increasing with precipitation in temperate and sub-tropical systems (Gatica et al., 2020), reflecting competing controls of soil and meteorological drivers on microbial activity and biogeochemical processes that differ across ecosystems. Additionally, process-based models exhibit a long-term increasing trend, driven by rising temperatures and atmospheric CH_4 levels (Murguia-Flores et al., 2021; Zhuang et al., 2013). These contrasting estimates highlight substantial uncertainties in both the magnitude and long-term variability of the global CH_4 soil sink.

Process-based (PB) and machine-learning (ML) approaches offer complementary strengths for addressing these uncertainties but may produce different estimates due to their distinct representations of CH_4 oxidation processes. PB models explicitly represent microbial and biogeochemical mechanisms and therefore exhibit explicit sensitivity to environmental drivers that are highly sensitive to the parameterization and governing equations (Murguia-Flores et al., 2021; Zhuang et al., 2013). In contrast, ML models leverage observational data to capture complex nonlinear relationships but may exhibit muted spatial and temporal variability and limited extrapolation if the training data is sparse and discrete (Feron et al., 2024).

We therefore address three key scientific questions: (a) How big is the global CH_4 soil sink? (b) How do PB and ML approaches differ in their representation of spatial and temporal variability? and (c) Can revised soil sink estimates help reconcile discrepancies between bottom-up and top-down approaches? We further hypothesize that the global CH_4 soil sink is larger than the previous estimate of ~30 Tg CH_4 yr^{-1}, that PB models will exhibit a stronger response to environmental parameters, and that incorporating a larger soil sink will improve agreement with atmospheric constraints.

To test the hypotheses, we combined PB, ML, and atmospheric inversion approaches to evaluate the spatial and temporal variability of global CH_4 soil sinks. For PB modeling, we used a model that incorporates microbial dynamics of methanogens and methanotrophs based on the Terrestrial Ecosystem Model (TEM; Oh et al., 2020; Zhuang et al., 2004). For the ML approach, we collected ~12,000 data points of observed global monthly CH_4 uptake rates from the literature (Figure 1) and ran multiple ML models to better represent the global CH_4 soil sinks. To compare our revised estimates from PB and ML models with atmospheric observations, we used them as inputs to the CarbonTracker-CH_4 atmospheric inversion systems, which optimizes CH_4 flux based on atmospheric measurements of CH_4 and its stable carbon isotopes (denoted $\delta^{13}C$-CH_4; Oh et al., 2023, 2026). Note that our estimates of PB and ML models in this study represent CH_4 uptake by soils rather than net CH_4 soil fluxes.

2. Method

2.1. Process-Based (PB) Modeling of the Global CH_4 Soil Sink

As our reference PB models, we presented the Methanotrophy Model ("MeMo" scenario; Murguia-Flores et al., 2018) and the Terrestrial Ecosystem Model ("TEM" scenario; Zhuang et al., 2013). These models differ in their representation of ecosystem complexity (Xu et al., 2016). Compared to the MeMo model, the TEM model provides a more detailed representation of methanotrophic responses to environmental changes. Specifically, TEM reveals ecosystem-specific parameters for the Michaelis-Menten function, as well as temperature and soil

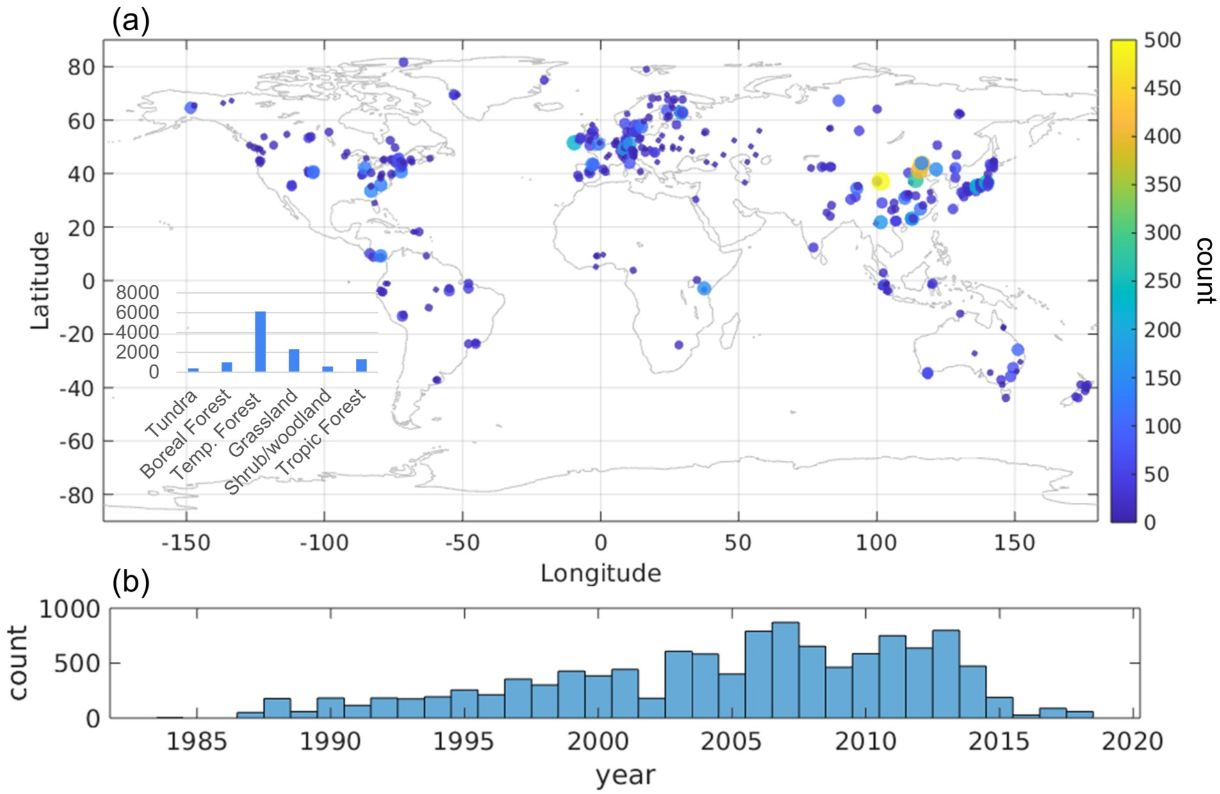

Figure 1. (a) Spatial distribution of site-level CH_4 soil uptake observations compiled for this study. Each circle represents an individual measurement site, with circle size and color indicating the number of observations available at that site. The inset shows the distribution of samples across major vegetation types. (b) Temporal distribution of observations from 1984 to 2020.

moisture-sensitivities (Table S1a in Supporting Information S1; Bender & Conrad, 1993). In contrast, MeMo applies a universal function and fixed parameters for temperature and soil moisture sensitivities, showing a Q_{10} of ~1.95 and a maximum soil moisture threshold of ~0.2 v/v, without using a Michaelis–Menten function (Table S1b in Supporting Information S1). The Michaelis–Menten formulation that MeMo does not use is important for capturing microbial responses to atmospheric CH_4 concentrations, and TEM's half-saturation constant values (10–15 μmol L^{-1}) represent microbial dynamics of LAM (Baani & Liesack, 2008).

To incorporate HAM into the global CH_4 soil sink estimation, we extended the model XPTEM-XHAM that we developed for the pan-Arctic study (Oh et al., 2020; "HAM" scenario). XPTEM-XHAM represents more detailed microbial dynamics than TEM, as it explicitly simulates the microbial biomass changes and the physiology of methanogens and methanotrophs. To capture the capacity of HAM to oxidize CH_4 at low concentrations, this model set the half-saturation constant of the Michaelis–Menten function to 0.11 μmolL^{-1} (Baani & Liesack, 2008). We further optimized parameters related to CH_4 oxidation processes using observations from regions with 8 different vegetation types to represent HAM (Table S2 and Figure S1 in Supporting Information S1; Borken et al., 2006; Castaldi & Fierro, 2005; Castro et al., 1995; Crill, 1991; Davidson et al., 2008; D'Imperio et al., 2017; Dinsmore et al., 2017; Ishizuka et al., 2000; Supporting Information S1). Our optimization results show that XPTEM-XHAM captures the magnitude and seasonality of observed CH_4 soil uptake fluxes, with a mean bias of <0.15 $mgCH_4\ m^{-2}\ day^{-1}$ across ecosystem type (Figure S2 in Supporting Information S1).

The model is then extrapolated globally using spatially and temporally explicit data sets of land cover, meteorology, and soil properties. We then applied the spatial resolution of 0.5° latitude by 0.5° longitude globally for global upland ecosystems with an hourly time-step for microbial dynamics and a daily time-step for other processes and modules during 1984 to 2020. Five years of spin-up were used for biomass equilibrium in soils for XPTEM-XHAM. Uncertainty in model parameters was estimated from ensemble simulations. We excluded CH_4 uptake from hyper-arid areas based on Zomer et al. (2022) (Figure S3b in Supporting Information S1), as microbes in these regions find it hard to form their community due to unstable soil conditions (More details in Section 4.1).

Lastly, since HAM prefer low soil organic carbon (SOC) upland conditions and may not dominate across all global upland soils (Conrad, 2007), we conducted sensitivity tests by varying the SOC threshold to adjust the relative dominance of HAM and LAM regions ("HAMLAM" scenario). Because there is limited literature describing ecosystem-specific distributions of HAM and LAM (Ho et al., 2013; Knief, 2015), we adopted the general definition of upland soils (<5 kg Cm^{-2}; Jobbágy & Jackson, 2000) as the default criterion for the HAMLAM scenario using topsoil total carbon from the Harmonized World Soil Database (Figure S3c in Supporting Information S1; Wieder et al., 2014). More details on PB optimization and upscaling can be found in the Text S1 in Supporting Information S1.

2.2. Data Collection

For the ML estimation of the global CH_4 soil sink, we collected extensive global CH_4 soil uptake data sets from previously published articles (Figure 1). We searched articles published between 1989 and 2022 using the keyword (CH_4 uptake or CH_4 sink) AND (forest OR grassland OR Shrubland OR woodland OR tundra) in Web of Science, Elsevier's Scopus academic database, and Google Scholar. Reported monthly CH_4 uptake rates, the date of measurement, and the coordinates of the study site were extracted from each paper. When the data were presented in the form of figures, they were extracted by Webplotdigitizer (Version 4.5, Pacifica, CA, USA). In the case of field manipulation experiments such as warming, elevated carbon dioxide, and fertilizer addition, the CH_4 uptake rates from only the control plot were collected. This includes data from Ni and Groffman (2018), but we further screen the data to remove manipulated and incorrect data sets.

From the screened data, we compiled two sets of data for model training and comparison. First, for ML model training, we collected ∼12,000 data points of monthly CH_4 soil uptake rates derived from observations across 260 study sites reported in 164 papers (Figure 1). Second, to compare the spatial representation of PB and ML models, we derived 260 site-averaged CH_4 uptake rates from the above database following the same data compilation procedure based on (Lee et al., 2023), with additional inclusion of non-forested ecosystems.

2.3. Machine-Learning (ML) Modeling of the Global CH_4 Soil Sink

We tested multiple tree-based models, including Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB), with different hyperparameter settings, and chose the RF model due to its highest R^2 and lowest root mean square error (RMSE; Figures S4 and S5 in Supporting Information S1). To ensure a balanced and representative training data set across vegetation types, we divided the data separately for each vegetation category. Within each type, 80% of the samples were randomly assigned for model training and the remaining 20% for testing. This stratified split preserved the proportional distribution of vegetation types in both subsets, preventing model bias toward ecosystems with larger sample sizes (Figure S3a in Supporting Information S1). To quantify model uncertainty, we implemented a 5-fold stratified cross-validation framework with 5 random seeds (25 experiments total). For each fold, we repeated the training with 5 different random seeds to account for model stochasticity, and performed global upscaling from 1984 to 2020 for each of the 25 experiments independently. The total uncertainty was further decomposed into data partition (fold) uncertainty and model stochasticity (seed) uncertainty. Input parameters include vegetation types, air temperature, precipitation, soil moisture, soil carbon, pH, and bulk density for both site-level training and global upscaling (Abatzoglou et al., 2018; Carter & Scholes, 2000; Harris et al., 2014; Myneni et al., 2002; Wieder et al., 2014; Zhuang et al., 2003; Figure S6 in Supporting Information S1). Similar to the PB modeling setup, we excluded CH_4 uptake from hyper-arid regions (Figure S3b in Supporting Information S1; Zomer et al., 2022).

To elucidate the sensitivity of ML models for different environmental parameters, we used the RF-generated impurity-based metric to analyze the contribution of each environmental variable in predicting CH_4 uptake (Breiman, 2001; Grömping, 2009). The partial dependence plots (PDP) were also used to analyze the response of CH_4 uptake to each predictor variable (Chen et al., 2022; Zhu et al., 2021). More details on ML data preprocessing, model configuration, experiment setup, and additional ML analysis results can be found in the Text S2 in Supporting Information S1.

2.4. Top-Down Atmospheric Inversion of Global CH_4 Budgets With Revised Global CH_4 Soil Sink

We used NOAA's atmospheric CH_4 inversion system, the CarbonTracker-CH_4, for this study (Oh et al., 2023, 2025, 2026). This system provides quantitative estimates of global CH_4 emissions from microbial, fossil, and pyrogenic sources consistent with observed patterns of atmospheric CH_4 and its stable carbon isotopes (δ^{13}C-CH_4; Basu et al., 2022).

There are 5 scenarios that we compare in this paper. For the baseline scenario ("CTCH$_4$" scenario), the optimized flux and simulation of CarbonTracker-CH_4 v2023 were used (Oh et al., 2023). From here, we revised the input for the global CH_4 soil sink using estimations from both the PB and ML approaches. Since the revised soil sink is larger than the previous input of the soil sink (Zhuang et al., 2013), we increased the prior emissions from wetlands accordingly to retain agreement with observed atmospheric CH_4 growth. Based on these updated priors, the atmospheric inversion subsequently optimizes fluxes by redistributing emissions from microbial, fossil, and pyrogenic sources to best match atmospheric observations. The baseline CarbonTracker-CH_4 uses 21‰ for the fractionation factor during the CH_4 uptake by methanotrophs; thus, our new scenarios use the same fractionation factor ("PB 21" and "ML 21" scenarios). However, the values have a significant uncertainty (15‰–21‰; King et al., 1989). To test the impact of fractionation on optimized CH_4 emissions and atmospheric CH_4 and the isotope changes, we tested two more scenarios that use the fractionation factor of 18‰ ("PB 18" and "ML 18" scenarios). The atmospheric inversion using four new scenarios was conducted during 2000–2019 and analyzed based on optimized emissions and atmospheric CH_4 and δ^{13}C-CH_4 changes.

For data assimilation and validation, we use the comprehensive data set of atmospheric CH_4 measurements from ~350 sites and δ^{13}C-CH_4 measurements from ~70 sites through collaboration with 30 laboratories worldwide (Lan et al., 2023). All data were rigorously quality-checked and intercompared across laboratories, then converted to common calibration scales (Umezawa et al., 2018). A subset of the network air sampling sites predominantly influenced by well-mixed background air is used to construct Marine Boundary Layer (MBL) zonally averaged surfaces using methods developed by Masarie and Tans (1995). For model-observation comparisons, we sampled model results from the same set of MBL sites to compare the global mean, latitude, and seasonal gradient of atmospheric CH_4 and δ^{13}C-CH_4.

3. Results

3.1. Process-Based (PB) Modeling of the Global CH_4 Soil Sink

Our reference PB models (MeMo and TEM scenarios) estimate the global CH_4 soil sink to be 28–35 Tg CH_4 yr^{-1} between 1984 and 2020 (Figures 2a and 2b and green and yellow in Figure 3). Because of its more detailed parameterization and process representation, TEM produces greater spatial variability than MeMo. Relative to MeMo, TEM simulates larger uptake in temperate forests, reflecting its higher temperature sensitivity (Table S1 in Supporting Information S1), but produces smaller fluxes in grasslands (Table S3 in Supporting Information S1). When HAM was incorporated (HAM scenario), our revised estimates of the CH_4 soil uptake increased to 45–60 Tg CH_4 yr^{-1} during 1984–2020 (Figure 2c and purple in Figure 3). Boreal forests, in particular, show higher total uptake due to a stronger temperature sensitivity (Q_{10} of ~3.5; Juncher Jørgensen et al., 2015), accounting for approximately 10 Tg CH_4 yr^{-1} of CH_4 uptake (Table S3 in Supporting Information S1). Lastly, the hybrid approach, incorporating both HAM and LAM contributions (HAMLAM scenario), results in a global CH_4 uptake of 40–45 Tg CH_4 yr^{-1} during 1984–2020, about 30%–50% higher than our reference estimates (Figure 2d and blue in Figure 3). Compared to the TEM scenario, boreal and grassland regions still exhibit stronger uptake, while temperate and tropical regions remain comparable (Table S3 in Supporting Information S1).

Although PB models differ significantly in their estimates of total magnitude and spatial variability, they consistently show strong annual increases and pronounced seasonality (Figure 3 and Figures S7, S8 in Supporting Information S1). Both TEM and MeMo scenarios simulate large increases in soil uptake (200% and 150%, respectively, from 1950 to 2000) driven by rising atmospheric CH_4 concentrations and temperature (Murguia-Flores et al., 2021; Zhuang et al., 2013). This is consistent with our new estimates from HAM and HAMLAM scenarios that showed long-term increases from 1984 to 2020 (Figure 3). All PB models also capture strong seasonality in temperate and polar regions, with peak CH_4 uptake in July-August (Figure S8 in Supporting Information S1), driven by higher temperatures, lower soil moisture, and the absence of winter snow cover. In

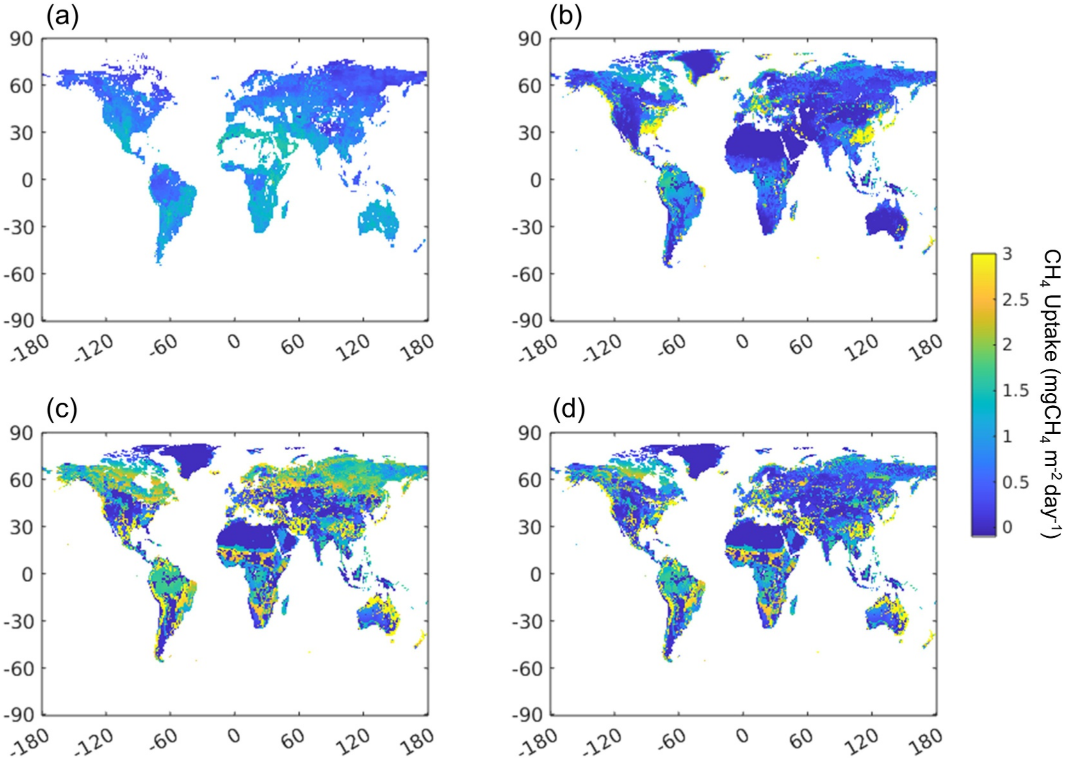

Figure 2. Annual mean CH_4 uptake rate (mg m^{-2} day^{-1}) from process-based models during 1990–2010: (a) MeMo, (b) TEM, (c) HAM (revised TEM with HAM), and (d) HAMLAM (revised TEM with both HAM and LAM) scenarios.

contrast, tropical forests, shrublands, and woodlands exhibit relatively small seasonality, as they lack such physical constraints. For the comparison with ML models in the following sections, we used the HAMLAM scenario as our default "PB" model (Figure 2d and blue in Figure 3).

3.2. Machine-Learning (ML) Modeling and Model-Data Comparison of the Global CH_4 Soil Sink

Our ML models estimate a global CH_4 soil sink of 41.9 ± 0.5 Tg CH_4 yr^{-1}, comparable in magnitude to our HAMLAM scenario, but with substantially lower spatial and temporal variability (Figure 4). Data partition (fold) uncertainty (0.5 Tg CH_4 yr^{-1}) is ~2.5 times larger than model stochasticity uncertainty (0.2 Tg CH_4 yr^{-1}), indicating that the choice of training data has a greater impact on predictions than random initialization. Because the ML model is trained on limited observational data (Figure 1; see Section 4), it reproduces muted spatial heterogeneity compared to the PB scenario (Figures S9 and S10 in Supporting Information S1). The mean CH_4 soil uptake rate from the ML model was ~0.78 $mgCH_4$ m^{-2} day^{-1} and exhibited less variation among ecosystem types than the PB estimates (Table S3 in Supporting Information S1). This muted variability is also reflected in the temporal domain, with relatively stable long-term changes and seasonality from 1984 to 2020 (Figure 4c and Figures S9, S10 in Supporting Information S1). Despite this muted variability, the ML model simulated a minor increase in the global annual CH_4 soil sink from 1984 to 2020 (red line in Figure 4), rather than the pronounced long-term decline reported by Ni and Groffman (2018).

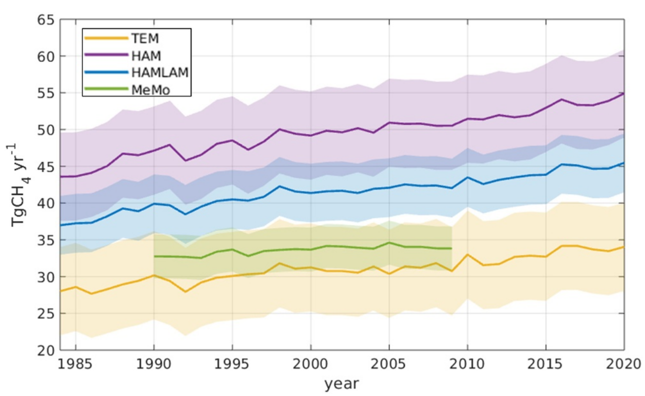

Figure 3. Long-term annual changes in the global CH_4 soil sink in $TgCH_4$ yr^{-1} from process-based models during 1984–2020: MeMo (green), TEM (yellow), HAM (purple), and HAMLAM (blue) scenarios. Shaded regions of TEM, HAM, and HAMLAM scenarios represent uncertainty derived from one standard deviation of ensemble simulations that account for parameter uncertainty. Shaded regions of the MeMo scenario represent uncertainty from the Monte Carlo approach described in Murguia-Flores et al. (2021).

To compare how PB and ML models represent the spatial variability of observations, we used ~260 aggregated CH_4 soil uptake data points described in

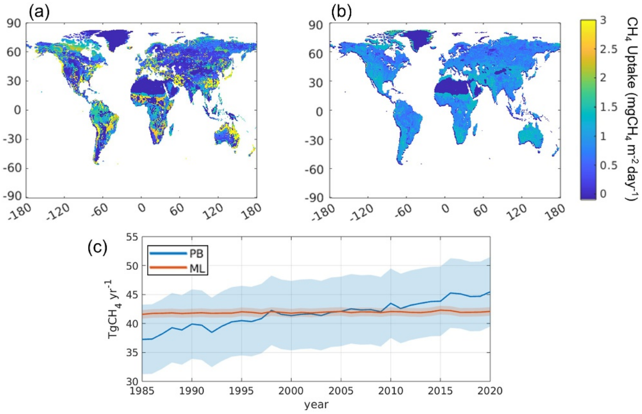

Figure 4. Spatial variability and long-term changes in the CH_4 soil sink from process-based (PB) and machine-learning (ML) models. (a, b) Annual mean CH_4 uptake (mg m^{-2} day^{-1}) from (a) PB model (HAMLAM scenario) and (b) ML model. (c) Long-term changes in the global CH_4 soil sink in $TgCH_4$ yr^{-1} during 1984–2020 from the PB (blue) and the ML (red) model. Shaded regions represent uncertainty derived from one standard deviation of ensemble simulations for the PB model and from a five-fold sampling strategy for the ML model.

Section 2.2 (Figures S11a and S11b in Supporting Information S1). Compared to observations, the ML model showed smaller RMSE (average of 1.00 $mgCH_4m^{-2}day^{-1}$) because the ML model constrain fluxes within a narrow range (0–3 $mgCH_4m^{-2}day^{-1}$), whereas PB models vary more widely (0–8 $mgCH_4m^{-2}day^{-1}$), resulting in larger RMSE (average of 1.32 $mgCH_4m^{-2}day^{-1}$). The linear fit between model and observation shows that the PB model generally over-estimates the CH_4 uptake (slope of 0.54, R^2 of 0.45), whereas the ML model under-estimates the uptake (slope of 1.38, R^2 of 0.38) compared to observations (Figures S11a and S11b in Supporting Information S1). Consistent with this spatial comparison, site-level seasonality analyses across Arctic, temperate, and tropical regions indicate that PB models better capture observed seasonal variability, while ML models show muted seasonality but more accurately reproduce mean uptake, particularly in regions with weak seasonality (Figures S11c–S11e in Supporting Information S1). Overall, PB and ML approaches produced similar totals of the global CH_4 soil sink but differed substantially in their representation of spatial and temporal variability.

3.3. Comparison With Atmospheric CH_4 and $\delta^{13}C$-CH_4 With Revised Inputs of Global CH_4 Soil Sinks From PB and ML Approaches

The atmospheric inversion scenarios show that revised CH_4 soil sinks and associated fractionation factors substantially affect optimized CH_4 emissions and the partitioning of microbial, fossil, and pyrogenic sources (Figure 5).

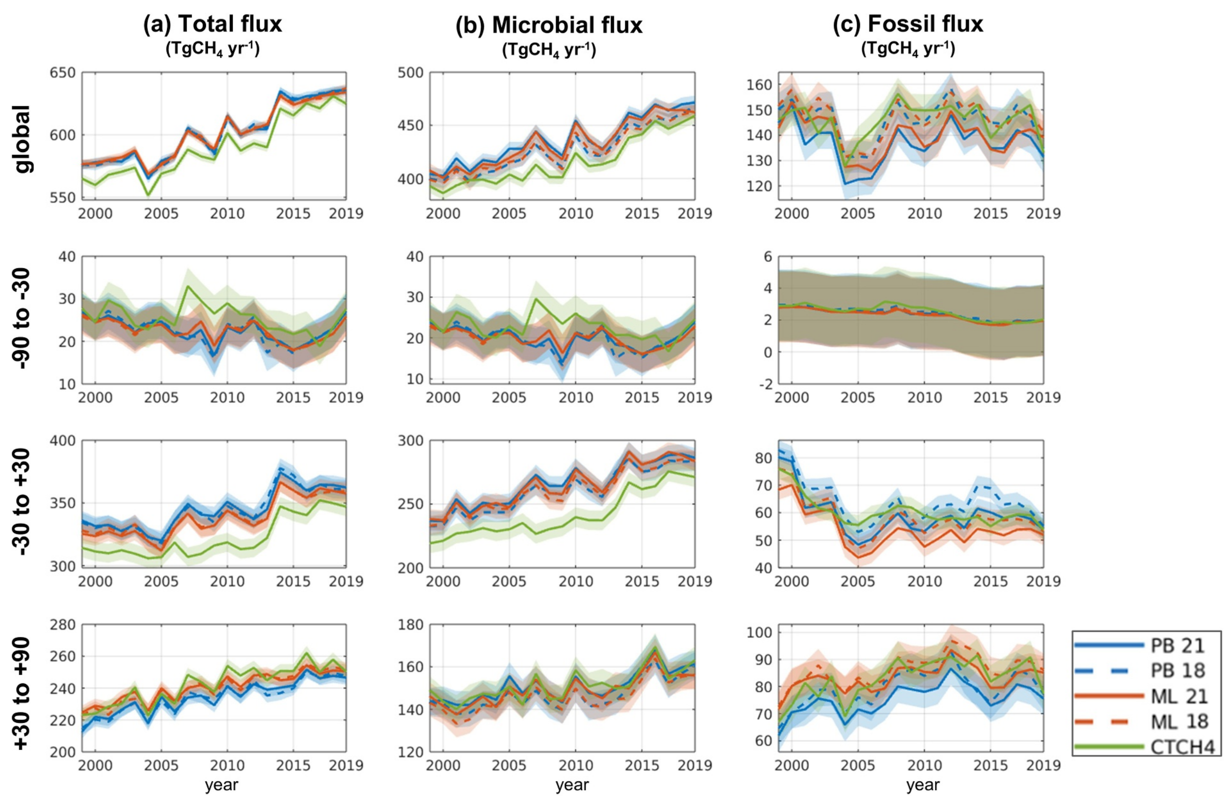

Figure 5. (a) Total and source-specific annual CH_4 emissions from (b) microbial and (c) fossil fluxes in $TgCH_4$ yr^{-1}, shown globally (top row) and for three latitudinal bands (90°S–30°S, 30°S–30°N, and 30°N–90°N; second–fourth rows) during 2000–2019. Results are shown from five atmospheric inversion scenarios with different soil sink estimates: PB model with isotopic fractionation of 21‰ (blue solid) and 18‰ (blue dashed), ML model with isotopic fractionation of 21‰ (red solid) and 18‰ (red dashed), and the estimates from the baseline CarbonTracker-CH_4 2023 (Green solid; Oh et al., 2023, 2026). Shaded regions represent uncertainty derived from one standard deviation of ensemble simulations that account for input uncertainty from observation and prior fluxes.

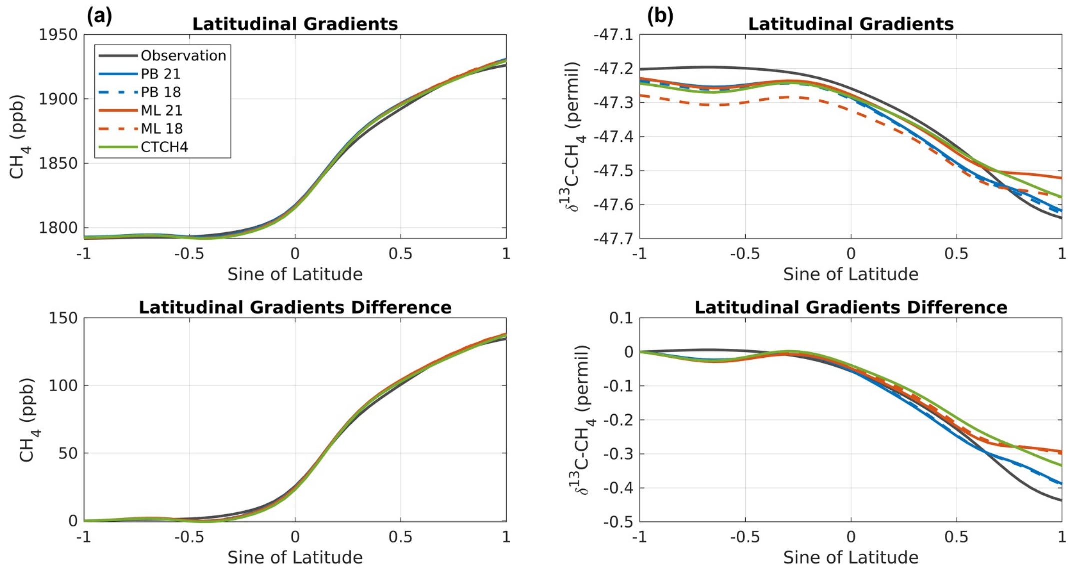

Figure 6. Latitudinal gradients of global (a) atmospheric CH_4 and (b) $\delta^{13}C$–CH_4 at marine boundary layer (MBL) sites from observations (black) and five atmospheric inversion scenarios: PB model with isotopic fractionation of 21‰ (blue solid) and 18‰ (blue dashed), ML model with isotopic fractionation of 21‰ (red solid) and 18‰ (red dashed), and the estimates from the baseline CarbonTracker-CH_4 2023 (Green solid; Oh et al., 2023, 2026). The top panels show simulated and observed atmospheric CH_4 and $\delta^{13}C$–CH_4, while the bottom panels show differences relative to the value at the sine latitude of −1 (90°S).

Globally, the simulation with revised soil sinks from PB and ML approaches (Blue and red lines in Figure 5) yield larger total and microbial emissions compared to our baseline scenario (green line in Figure 5), as larger soil sinks increase emissions from lighter microbial sources such as wetland emissions. Larger fractionation (21‰) for the soil sink causes a greater increase in microbial emissions (solid red and blue lines in Figure 5) compared to smaller fractionation (18‰; dashed red and blue lines in Figure 5), as the larger sink-weighted fractionation factor causes atmospheric $\delta^{13}C$-CH_4 to be heavier. Tropical regions showed the largest increase in microbial emissions (region −30 to +30 in Figure 5) when the revised soil sinks were applied, due to the region's largest contribution to global microbial emissions (Table S3 in Supporting Information S1).

These emission adjustments influence the model's ability to reproduce observed atmospheric CH_4 and $\delta^{13}C$-CH_4 variations (Figure 6 and Figures S12, S13 in Supporting Information S1). Because both PB- and ML-based soil sinks require larger microbial emissions (Figure 5), the resulting increase in simulated CH_4 seasonality improves agreement with observed seasonal cycles (blue and red lines in Figure S13 in Supporting Information S1). For the latitudinal gradients of CH_4, all scenarios reproduce the observed gradient of ∼+130 ppb from −90° to +90°, as atmospheric inversions constrain model–observation biases (Figure 6a). In contrast, $\delta^{13}C$-CH_4 gradients are not fully constrained, making them a useful diagnostic for distinguishing scenarios. The latitudinal gradients show that the larger heterogeneity in the soil sink of PB models helps reproduce the latitudinal gradients of $\delta^{13}C$-CH_4 (Figure 6b). In contrast, ML scenarios underestimate the gradients (red in Figure 6b, respectively), reflecting their muted spatial distribution of soil sinks (Figure 4).

4. Discussion

4.1. Revised Magnitude and Atmospheric Implications of the Global CH_4 Soil Sink

Both our PB and ML approaches estimate the total global CH_4 soil sink at 40–45 Tg CH_4 yr^{-1}, approximately 30%–50% higher than previous estimates (Figure 4; Murguia-Flores et al., 2018). Despite substantial differences in model structure and representation of CH_4 oxidation processes, the agreement of the total magnitude of PB and ML estimates suggests that CH_4 uptake by upland ecosystems may have been underestimated. This larger soil sink estimation is also consistent with recently published data-driven analysis of global CH_4 soil sink (Jiang et al., 2025).

Our atmospheric inversions further provide independent evidence supporting a larger global CH_4 soil sink (Figures 5 and 6 and Figure S13 in Supporting Information S1). Incorporating the revised soil sink estimates into atmospheric inversions resulted in larger optimized microbial CH_4 emissions while improving agreement with atmospheric CH_4 and $\delta^{13}C$-CH_4 observations. The improved representation of atmospheric CH_4 seasonality across both PB and ML scenarios further suggests that a stronger CH_4 sink is more consistent with observed atmospheric variability (Figure S13 in Supporting Information S1). This adjustment could partly compensate for the over-estimation of natural CH_4 emissions by bottom-up CH_4 budgets compared to the top-down estimates (Saunois et al., 2025).

Nevertheless, substantial uncertainties remain in the magnitudes of global biological CH_4 sinks. First, more studies are necessary to distinguish the optimal soil, vegetation, and climate conditions for HAM to be dominant in terrestrial ecosystems, which remains one of the most uncertain parameters in our PB modeling (Figures 2 and 3). Second, we excluded potential CH_4 uptake from hyper-arid regions due to large uncertainties (Figure S3b in Supporting Information S1). However, several studies have reported unexpected uptake rates in these extreme environments, such as the Sahara Desert, the Tibetan Plateau, Inner Mongolia, and Antarctica (Ortiz et al., 2021; Striegl et al., 1992; Wang et al., 2011; Yue et al., 2022), which should be further examined. Finally, recent observations indicate large CH_4 sinks from tree stems (Gauci, 2025; Gauci et al., 2024), suggesting that current estimates of terrestrial CH_4 consumption may still be incomplete. Together, these uncertainties indicate that the global biological CH_4 sink needs to be investigated more for accurate CH_4 budget estimations and realistic climate mitigation strategies (He et al., 2023; Lim et al., 2024; Rani et al., 2024).

4.2. Why PB and ML Models Produce Different Spatial and Temporal Variability

Another key insight from our study is the representation of spatial and temporal variability in PB and ML models. All PB models reproduce substantial long-term decadal increasing trends and strong seasonality and spatial variability (Figures 2 and 3 and Figures S7, S8 in Supporting Information S1), whereas the ML model generates muted variability (Figure 4 and Figures S9, S10 in Supporting Information S1). This contrast is consistent with our hypothesis that PB and ML approaches represent environmental controls on CH_4 oxidation differently.

The stronger spatial and temporal variability in PB models is primarily driven by their explicit representation of environmental sensitivity such as temperature. In the PB models, CH_4 oxidation responds directly to environmental changes through mechanistic equations (Oh et al., 2020), resulting in pronounced increases in CH_4 uptake under long-term warming and strong seasonality and spatial patterns. In contrast, the muted temporal variability in the ML-based estimate, compared with the PB models, reflects a fundamentally weaker learned environmental sensitivity. For example, the effective Q_{10} of the RF model ranges from 1.01 to 1.23 across vegetation types (Table S4 in Supporting Information S1), compared with 1.25–3.50 in the HAM model and 0.80–5.50 in TEM (Tables S1 and S2 in Supporting Information S1). The spatially pervasive low Q_{10} (global mean 1.11 ± 0.13; Figure S14 in Supporting Information S1) confirms that the ML model has learned a nearly flat temperature response. This limitation arises because the RF model must infer temperature sensitivity entirely from observational co-variation, whereas PB models prescribe it through mechanistic equations.

This weak temperature sensitivity of ML model likely reflects the limitations of discrete and sparse observations used for model training (Figures S15–S17 in Supporting Information S1). Most sites in our data set have very short temporal records: 357 of 468 sites (76%) have fewer than 10 continuous monthly measurements (Figure S15a in Supporting Information S1), limiting the model's ability to learn long-term trends from these discrete samples. The temporal estimates from ecosystem-specific RF models remained consistent with the global RF prediction, suggesting that the muted long-term trends are not driven by ecosystem-level confounding (Figures S18 and S19 in Supporting Information S1). In addition, chamber data do not fully capture long-term environmental changes, particularly temperature (Figures S15b–S15g in Supporting Information S1). The sampled temperatures show large year-to-year variability and are biased high in tundra and boreal ecosystems. The discontinuous training data and scale mismatch between 0.5° gridded inputs and point-scale chamber measurements further constrain the model's ability to capture long-term warming-driven trends.

Despite the muted variability, the ML model properly identifies key controls on CH_4 uptake, with soil moisture, precipitation, and temperature as the strongest predictors (Figure S20 in Supporting Information S1). It also showed that uptake generally increases with temperature and precipitation but decreases with higher soil moisture, thereby preserving realistic sensitivities even when variability is muted (Figure S21 in Supporting

Information S1). These findings suggest that current observational data sets are sufficient to constrain the mean spatial distribution of CH_4 uptake but remain inadequate for robustly constraining its large spatial and temporal dynamics.

4.3. Limitations of Current PB and ML Models and Future Opportunities

This study showed that PB models provide an open-box framework that explicitly represents the biogeochemical processes underlying CH_4 oxidation. Yet many environmental factors and processes remain poorly constrained from current formulations, for instance, the Michaelis-Menten function for substrate interactions for CH_4 uptake. Another limitation arises from the parameterization strategy. PB models rely on parameter sets calibrated to specific experimental data sets, which may not be globally representative, introducing systematic bias when scaled to regional or global levels (Figures 2 and 3). ML approaches, by contrast, offer flexibility and high efficiency in capturing nonlinear relationships between CH_4 fluxes and environmental drivers. However, ML models remain heavily dependent on the size, coverage, and quality of the training data (Figure 1). Despite extensive data collection, CH_4 uptake data sets remain sparse compared to other flux data sets, such as wetland CH_4 emissions (Delwiche et al., 2021). Furthermore, ML models are inherently black-box approaches (Hutson, 2022). While they may reproduce mean fluxes well, their mechanistic interpretations are complex and often limited to capture spatial and temporal variability.

Given the complementary strengths and weaknesses of PB and ML models, a promising path forward lies in hybrid approaches, particularly Knowledge-Guide Machine Learning (KGML; Liu et al., 2024; Reichstein et al., 2019). KGML integrates mechanistic understanding into ML frameworks, combining the interpretability and process fidelity of PB models with the flexibility and data-driven strengths of ML. In the context of CH_4 soil uptake, KGML could, for example, constrain ML training with mechanistic equations for enzyme kinetics and temperature sensitivities, while allowing data-driven algorithms to refine parameter values or identify missing environmental predictors. This approach could reduce biases from narrow parameterization data sets, address the black-box nature of ML, and improve predictive power under out-of-sample climate conditions. Moreover, hybrid models could maximize the value of limited observational data by leveraging both mechanistic priors and statistical generalization. This is especially critical for CH_4 uptake, where observational coverage is sparse, and large-scale atmospheric implications hinge on small-scale microbial processes. The development and evaluation of KGML models, however, are beyond the scope of this study and represent an important direction for future research.

5. Conclusion

This study provides a comprehensive assessment of the global CH_4 soil sink by integrating PB, ML, and atmospheric inversion approaches. Both PB and ML approaches estimate a global CH_4 soil sink of 40–45 Tg CH_4 yr^{-1} during 1984–2020, ~30%–50% larger than previous estimates, supporting the hypothesis that CH_4 uptake by upland ecosystems may have been underestimated. Although PB and ML approaches agree on a similar global magnitude, they differ substantially in their representation of spatial and temporal variability. PB models simulate stronger spatial heterogeneity, seasonality, and long-term variability because they explicitly represent environmental sensitivities through mechanistic equations. In contrast, the ML model exhibits comparatively muted variability because sparse and discrete observations limit its ability to represent environmental sensitivities. Atmospheric inversions provide independent support for larger CH_4 soil sinks. Incorporating the revised sink estimates requires larger microbial CH_4 emissions while maintaining consistency with atmospheric CH_4 and $\delta^{13}C\text{-}CH_4$ observations. These results demonstrate that uncertainties in CH_4 soil sinks directly influence estimates of CH_4 emissions and atmospheric CH_4 and $\delta^{13}C\text{-}CH_4$. Future efforts combining the mechanistic representation of PB models with the observational constraints of ML approaches may improve estimates of spatial and temporal variability of the global CH_4 soil sink.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

Global soil uptake data used for training the ML model are available from Lee et al. (2026). Codes for the PB and ML models are available from Oh, Liu, and Lee et al. (2026): https://doi.org/10.5281/zenodo.21361402. The atmospheric CH_4 and $\delta^{13}C$-CH_4 data used for the atmospheric inversion modeling can be downloaded from Lan et al. (2023): https://doi.org/10.15138/T4MZ-2Z29. The CarbonTracker-CH_4 code for performing the atmospheric inversion is publicly available at: https://sourceforge.net/p/tm5/cy3_4dvar/ci/default/tree/.

Acknowledgments

This research was supported by the U.S. Department of Energy Grants DE-SC0024360. This research was supported in-part by the NOAA cooperative agreement NA22OAR4320151. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect the views of NOAA or the Department of Commerce.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, *5*(1), 1–12. https://doi.org/10.1038/sdata.2017.191

Baani, M., & Liesack, W. (2008). Two isozymes of particulate methane monooxygenase with different methane oxidation kinetics are found in Methylocystis sp. strain SC2. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(29), 10203–10208. https://doi.org/10.1073/pnas.0702643105

Basu, S., Lan, X., Dlugokencky, E., Michel, S., Schwietzke, S., Miller, J. B., et al. (2022). Estimating emissions of methane consistent with atmospheric measurements of methane and $\delta^{13}C$ of methane. *Atmospheric Chemistry and Physics*, *22*(23), 15351–15377. https://doi.org/10.5194/acp-22-15351-2022

Bender, M., & Conrad, R. (1993). Kinetics of methane oxidation in oxic soils. *Chemosphere*, *26*(1–4), 687–696. https://doi.org/10.1016/0045-6535(93)90453-c

Borken, W., Davidson, E. A., Savage, K., Sundquist, E. T., & Steudler, P. (2006). Effect of summer throughfall exclusion, summer drought, and winter snow cover on methane fluxes in a temperate forest soil. *Soil Biology and Biochemistry*, *38*(6), 1388–1395. https://doi.org/10.1016/j.soilbio.2005.10.011

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Carter, A. J., & Scholes, R. J. (2000). Spatial global database of soil properties. In *IGBP global soil data task CD-ROM. International Geosphere-Biosphere Programme (IGBP) data information systems. Toulouse, France*.

Castaldi, S., & Fierro, A. (2005). Soil-atmosphere methane exchange in undisturbed and burned Mediterranean shrubland of southern Italy. *Ecosystems*, *8*(2), 182–190. https://doi.org/10.1007/s10021-004-0093-z

Castro, M. S., Steudler, P. A., Melillo, J. M., Aber, J. D., & Bowden, R. D. (1995). Factors controlling atmospheric methane consumption by temperate forest soils. *Global Biogeochemical Cycles*, *9*(1), 1–10. https://doi.org/10.1029/94gb02651

Chen, S., Liu, W., Feng, P., Ye, T., Ma, Y., & Zhang, Z. (2022). Improving spatial disaggregation of crop yield by incorporating machine learning with multisource data: A case study of Chinese maize yield. *Remote Sensing*, *14*(10), 2340. https://doi.org/10.3390/rs14102340

Conrad, R. (2007). Microbial ecology of methanogens and methanotrophs. *Advances in Agronomy*, *96*(7), 1–63. https://doi.org/10.1016/S0065-2113(07)96005-8

Conrad, R. (2009). The global methane cycle: Recent advances in understanding the microbial processes involved. *Environmental Microbiology Reports*, *1*(5), 285–292. https://doi.org/10.1111/j.1758-2229.2009.00038.x

Crill, P. M. (1991). Seasonal patterns of methane uptake and carbon dioxide release by a temperate woodland soil. *Global Biogeochemical Cycles*, *5*(4), 319–334. https://doi.org/10.1029/91gb02466

Davidson, E. A., Nepstad, D. C., Ishida, F. Y., & Brando, P. M. (2008). Effects of an experimental drought and recovery on soil emissions of carbon dioxide, methane, nitrous oxide, and nitric oxide in a moist tropical forest. *Global Change Biology*, *14*(11), 2582–2590. https://doi.org/10.1111/j.1365-2486.2008.01694.x

Delwiche, K. B., Knox, S. H., Malhotra, A., Fluet-Chouinard, E., McNicol, G., Feron, S., et al. (2021). FLUXNET-CH 4: A global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands. *Earth System Science Data*, *13*(7), 3607–3689. https://doi.org/10.5194/essd-13-3607-2021

D'Imperio, L., Nielsen, C. S., Westergaard-Nielsen, A., Michelsen, A., & Elberling, B. (2017). Methane oxidation in contrasting soil types: Responses to experimental warming with implication for landscape-integrated CH4 budget. *Global Change Biology*, *23*(2), 966–976. https://doi.org/10.1111/gcb.13400

Dinsmore, K. J., Drewer, J., Levy, P. E., George, C., Lohila, A., Aurela, M., & Skiba, U. M. (2017). Growing season CH4 and N2O fluxes from a subarctic landscape in northern Finland; from chamber to landscape scale. *Biogeosciences*, *14*(4), 799–815. https://doi.org/10.5194/bg-14-799-2017

Dutaur, L., & Verchot, L. V. (2007). A global inventory of the soil CH4 sink. *Global Biogeochemical Cycles*, *21*(4), 1–9. https://doi.org/10.1029/2006gb002734

Feron, S., Malhotra, A., Bansal, S., Fluet-Chouinard, E., McNicol, G., Knox, S. H., et al. (2024). Recent increases in annual, seasonal, and extreme methane fluxes driven by changes in climate and vegetation in boreal and temperate wetland ecosystems. *Global Change Biology*, *30*(1), e17131. https://doi.org/10.1111/gcb.17131

Gatica, G., Fernández, M. E., Juliarena, M. P., & Gyenge, J. (2020). Environmental and anthropogenic drivers of soil methane fluxes in forests: Global patterns and among-biomes differences. *Global Change Biology*, *26*(11), 6604–6615. https://doi.org/10.1111/gcb.15331

Gauci, V. (2025). Tree methane exchange in a changing world. *Nature Reviews Earth & Environment*, *6*(7), 471–483. https://doi.org/10.1038/s43017-025-00692-9

Gauci, V., Pangala, S., Shenkin, A., Barba, J., Bastviken, D., Figueiredo, V., et al. (2024). Global atmospheric methane uptake by upland tree woody surfaces. *Nature*, *631*(8022), 796–800. https://doi.org/10.1038/s41586-024-07592-w

Grömping, U. (2009). Variable importance assessment in regression: Linear regression versus random forest. *The American Statistician*, *63*(4), 308–319. https://doi.org/10.1198/tast.2009.08199

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/joc.3711

He, L., Groom, J. D., Wilson, E. H., Fernandez, J., Konopka, M. C., Beck, D. A. C., & Lidstrom, M. E. (2023). A methanotrophic bacterium to enable methane removal for climate mitigation. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(35), e2310046120. https://doi.org/10.1073/pnas.2310046120

Ho, A., Kerckhof, F.-M., Luke, C., Reim, A., Krause, S., Boon, N., & Bodelier, P. L. E. (2013). Conceptualizing functional traits and ecological characteristics of methane-oxidizing bacteria as life strategies: Functional traits of methane-oxidizing bacteria. *Environmental Microbiology Reports*, *5*(3), 335–345. https://doi.org/10.1111/j.1758-2229.2012.00370.x

Hutson, M. (2022). Taught to the test. *Science*, *376*(6593), 570–573. https://doi.org/10.1126/science.abq7833

Ishizuka, S., Sakata, T., & Ishizuka, K. (2000). Methane oxidation in Japanese forest soils. *Soil Biology and Biochemistry*, *32*(6), 769–777. https://doi.org/10.1016/s0038-0717(99)00200-x

Jiang, J., Yan, Z., Jian, J., Peng, S., Tian, H., Morris, K. A., et al. (2025). Global soil methane uptake estimated by scaling up local measurements. *Global Change Biology*, *31*(4), e70194. https://doi.org/10.1111/gcb.70194

Jobbágy, E. G., & Jackson, R. B. (2000). Global controls of forest line elevation in the northern and southern hemispheres. *Global Ecology and Biogeography: A Journal of Macroecology*, *9*(3), 253–268. https://doi.org/10.1046/j.1365-2699.2000.00162.x

Juncher Jørgensen, C., Lund Johansen, K. M., Westergaard-Nielsen, A., & Elberling, B. (2015). Net regional methane sink in High Arctic soils of northeast Greenland. *Nature Geoscience*, *8*(1), 20–23. https://doi.org/10.1038/ngeo2305

Kato, T., Hirota, M., Tang, Y., & Wada, E. (2011). Spatial variability of CH4 and N2O fluxes in alpine ecosystems on the Qinghai-Tibetan Plateau. *Atmospheric Environment*, *45*(31), 5632–5639. https://doi.org/10.1016/j.atmosenv.2011.03.010

King, S. L., Quay, P. D., & Lansdown, J. M. (1989). The 13C/12C kinetic isotope effect for soil oxidation of methane at ambient atmospheric concentrations. *Journal of Geophysical Research*, *94*(D15), 18273–18277. https://doi.org/10.1029/jd094id15p18273

Knief, C. (2015). Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on pmoA as molecular marker. *Frontiers in Microbiology*, *6*. https://doi.org/10.3389/fmicb.2015.01346

Lan, X., Schuldt, K., Michel, S., Miller, J., Bruhwiler, L., Oh, Y., & Bash, S. (2023). Database of methane (CH_4) abundance and its stable carbon isotope ($d^{13}C$-CH_4) composition from atmospheric measurements used infor CarbonTracker-CH_4 2023. https://doi.org/10.15138/T4MZ-2Z29

Lau, M. C. Y., Stackhouse, B. T., Layton, A. C., Chauhan, A., Vishnivetskaya, T. A., Chourey, K., et al. (2015). An active atmospheric methane sink in high Arctic mineral cryosols. *The ISME Journal*, *9*(8), 1880–1891. https://doi.org/10.1038/ismej.2015.13

Lee, J., Oh, Y., Lee, S. T., Seo, Y. O., Yun, J., Yang, Y., et al. (2023). Soil organic carbon is a key determinant of CH_4 sink in global forest soils. *Nature Communications*, *14*(1), 3110. https://doi.org/10.1038/s41467-023-38905-8

Lee, J., Oh, Y., & Liu, L. (2026). Global compilation of soil methane uptake measurements from 1984 to 2018. Estimation of Global Methane Soil Sink Using Synthesized Datasets and Knowledge-Guided Machine Learning [Dataset]. *ESS-DIVE repository*. https://doi.org/10.15485/3377679

Lim, J., Wehmeyer, H., Heffner, T., Aeppli, M., Gu, W., Kim, P. J., et al. (2024). Resilience of aerobic methanotrophs in soils; spotlight on the methane sink under agriculture. *FEMS Microbiology Ecology*, *100*(3), fiae008. https://doi.org/10.1093/femsec/fiae008

Liu, L., Zhou, W., Guan, K., Peng, B., Xu, S., Tang, J., et al. (2024). Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems. *Nature Communications*, *15*(1), 357. https://doi.org/10.1038/s41467-023-43860-5

Masarie, K. A., & Tans, P. P. (1995). Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record. *Journal of Geophysical Research: Atmospheres*, *100*(D6), 11593–11610. https://doi.org/10.1029/95jd00859

Murguia-Flores, F., Arndt, S., Ganesan, A. L., Murray-Tortarolo, G., & Hornibrook, E. R. C. (2018). Soil Methanotrophy Model (MeMo v1.0): A process-based model to quantify global uptake of atmospheric methane by soil. *Geoscientific Model Development*, *11*(6), 2009–2032. https://doi.org/10.5194/gmd-11-2009-2018

Murguia-Flores, F., Ganesan, A. L., Arndt, S., & Hornibrook, E. R. C. (2021). Global uptake of atmospheric methane by soil from 1900 to 2100. *Global Biogeochemical Cycles*, *35*(7), e2020GB006774. https://doi.org/10.1029/2020gb006774

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., et al. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, *83*(1–2), 214–231. https://doi.org/10.1016/s0034-4257(02)00074-3

Ni, X., & Groffman, P. M. (2018). Declines in methane uptake in forest soils. *Proceedings of the National Academy of Sciences*, *115*(34), 8587–8590. https://doi.org/10.1073/pnas.1807377115

Oh, Y., Bruhwiler, L., Lan, X., Basu, S., Schuldt, K., Thoning, K., et al. (2023). *CarbonTracker CH4 2023*. NOAA Global Monitoring Laboratory. https://doi.org/10.25925/40jt-qd67

Oh, Y., Bruhwiler, L., Lan, X., Halder, S., Riddell-Young, B., Basu, S., et al. (2026). Isotopic constraints in methane inversions reveal larger trends in wetland emissions with improved linkage to terrestrial water storage. *Nature Communications*. https://doi.org/10.1038/s41467-026-74777-4

Oh, Y., Bruhwiler, L., Lan, X., Schuldt, K., Basu, S., Thoning, K., et al. (2025). *CarbonTracker CH4 2025*. NOAA Global Monitoring Laboratory. https://doi.org/10.25925/HXKS-V755

Oh, Y., Liu, L., & Lee, J. (2026). Meta-data and code for process-based and machine-learning model for global methane soil sink estimation. *Zenodo*. https://doi.org/10.5281/zenodo.21361402

Oh, Y., Zhuang, Q., Liu, L., Welp, L. R., Lau, M. C. Y., Onstott, T. C., et al. (2020). Reduced net methane emissions due to microbial methane oxidation in a warmer Arctic. *Nature Climate Change*, *10*(4), 317–321. https://doi.org/10.1038/s41558-020-0734-z

Ortiz, M., Leung, P. M., Shelley, G., Jirapanjawat, T., Nauer, P. A., Van Goethem, M. W., et al. (2021). Multiple energy sources and metabolic strategies sustain microbial diversity in Antarctic desert soils. *Proceedings of the National Academy of Sciences*, *118*(45), e2025322118. https://doi.org/10.1073/pnas.2025322118

Rani, A., Pundir, A., Verma, M., Joshi, S., Verma, G., Andjelković, S., et al. (2024). Methanotrophy: A biological method to mitigate global methane emission. *Microbiology Research*, *15*(2), 634–654. https://doi.org/10.3390/microbiolres15020042

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Saunois, M., Martinez, A., Poulter, B., Zhang, Z., Raymond, P. A., Regnier, P., et al. (2025). Global Methane Budget 2000–2020. *Earth System Science Data*, *17*(5), 1873–1958. https://doi.org/10.5194/essd-17-1873-2025

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., et al. (2020). The Global Methane Budget 2000 – 2017 (pp. 1561–1623). https://doi.org/10.5194/essd-12-1561-2020

Smith, K. A., Dobbie, K. E., Ball, B. C., Bakken, L. R., Sitaula, B. K., Hansen, S., et al. (2000). Oxidation of atmospheric methane in Northern European soils, comparison with other ecosystems, and uncertainties in the global terrestrial sink. *Global Change Biology*, *6*(7), 791–803. https://doi.org/10.1046/j.1365-2486.2000.00356.x

Striegl, R. G., McConnaughey, T. A., Thorstenson, D. C., Weeks, E. P., & Woodward, J. C. (1992). Consumption of atmospheric methane by desert soils. *Nature*, *357*(6374), 145–147. https://doi.org/10.1038/357145a0

Tveit, A. T., Hestnes, A. G., Robinson, S. L., Schintlmeister, A., Dedysh, S. N., Jehmlich, N., et al. (2019). Widespread soil bacterium that oxidizes atmospheric methane. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(17), 8515–8524. https://doi.org/10.1073/pnas.1817812116

Umezawa, T., Brenninkmeijer, C. A., Röckmann, T., Van Der Veen, C., Tyler, S. C., Fujita, R., et al. (2018). Interlaboratory comparison of $\delta^{13}C$ and δD measurements of atmospheric CH 4 for combined use of data sets from different laboratories. *Atmospheric Measurement Techniques*, *11*(2), 1207–1231. https://doi.org/10.5194/amt-11-1207-2018

Wang, Z., Hao, X., Shan, D., Han, G., Zhao, M., Willms, W. D., et al. (2011). Influence of increasing temperature and nitrogen input on greenhouse gas emissions from a desert steppe soil in Inner Mongolia. *Soil Science & Plant Nutrition*, *57*(4), 508–518. https://doi.org/10.1080/00380768.2011.591283

Wieder, W. R., Boehnert, J., Bonan, G. B., & Langseth, M. (2014). *Regridded harmonized world soil database v1. 2*. ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/1247

Xu, X., Yuan, F., Hanson, P. J., Wullschleger, S. D., Thornton, P. E., Riley, W. J., et al. (2016). Reviews and syntheses: Four decades of modeling methane cycling in terrestrial ecosystems. *Biogeosciences*, *13*(12), 3735–3755. https://doi.org/10.5194/bg-13-3735-2016

Yue, P., Zuo, X., Li, K., Li, X., Wang, S., & Misselbrook, T. (2022). Precipitation changes regulate the annual methane uptake in a temperate desert steppe. *The Science of the Total Environment*, *804*(150172), 150172. https://doi.org/10.1016/j.scitotenv.2021.150172

Zhu, P., Abramoff, R., Makowski, D., & Ciais, P. (2021). Uncovering the past and future climate drivers of wheat yield shocks in Europe with machine learning. *Earth's Future*, *9*(5), e2020EF001815. https://doi.org/10.1029/2020ef001815

Zhuang, Q., McGuire, A. D., Melillo, J. M., Clein, J. S., Dargaville, R. J., Kicklighter, D. W., et al. (2003). Carbon cycling in extratropical terrestrial ecosystems of the Northern Hemisphere during the 20th century: A modeling analysis of the influences of soil thermal dynamics. *Tellus Series B*, *55*(3), 751–776. https://doi.org/10.3402/tellusb.v55i3.16368

Zhuang, Q., Melillo, J. M., Kicklighter, D. W., Prinn, R. G., McGuire, A. D., Steudler, P. A., et al. (2004). 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*, *18*(3). https://doi.org/10.1029/2004GB002239

Zhuang, Q., Xu, K., Tang, J., Saikawa, E., Lu, Y., Melillo, J. M., et al. (2013). Response of global soil consumption of atmospheric methane to changes in atmospheric climate and nitrogen deposition. *Global Biogeochemical Cycles*, *27*(3), 650–663. https://doi.org/10.1002/gbc.20057

Zomer, R. J., Xu, J., & Trabucco, A. (2022). Version 3 of the global aridity index and potential evapotranspiration database. *Scientific Data*, *9*(1), 409. https://doi.org/10.1038/s41597-022-01493-1

References From the Supporting Information

Barthel, M., Bauters, M., Baumgartner, S., Drake, T. W., Bey, N. M., Bush, G., et al. (2022). Low N_2O and variable CH_4 fluxes from tropical forest soils of the Congo Basin. *Nature Communications*, *13*(1), 330. https://doi.org/10.1038/s41467-022-27978-6

Cardenas, L. M., Chadwick, D., Scholefield, D., Fychan, R., Marley, C. L., Jones, R., et al. (2007). The effect of diet manipulation on nitrous oxide and methane emissions from manure application to incubated grassland soils. *Atmospheric Environment*, *41*(33), 7096–7107. https://doi.org/10.1016/j.atmosenv.2007.04.055

Chen, T., & Guestrin, C. (2016). XGBoost. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. KDD'16: The 22nd ACM SIGKDD international conference on knowledge discovery and data mining, San Francisco California USA*. https://doi.org/10.1145/2939672.2939785

Donateo, A., Famulari, D., Giovannelli, D., Mariani, A., Mazzola, M., Decesari, S., & Pappaccogli, G. (2025). Observations of methane net sinks in the upland Arctic tundra. *Biogeosciences*, *22*(12), 2889–2908. https://doi.org/10.5194/bg-22-2889-2025

Duan, Q., Sorooshian, S., & Gupta, V. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology*, *158*(3–4), 265–284. https://doi.org/10.1016/0022-1694(94)90057-4

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gordon, A. D., Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. *Biometrics*, *40*(3), 874. https://doi.org/10.1201/9781315139470

Liu, L., Zhuang, Q., Oh, Y., Shurpali, N. J., Kim, S., & Poulter, B. (2020). Uncertainty quantification of global net methane emissions from terrestrial ecosystems using a mechanistically-based biogeochemistry model. *Journal of Geophysical Research: Biogeosciences*, *125*(6), e2019JG005428. https://doi.org/10.1029/2019jg005428

Liu, Y., Christiansen, J. R., Huang, K., Liu, D., Duan, Y., Liu, G., et al. (2025). Temperature and moisture both control net methane uptake in a temperate forest soil. *Agricultural and Forest Meteorology*, *370*(110574), 110574. https://doi.org/10.1016/j.agrformet.2025.110574

Matthews, E., & Fung, I. (1987). Methane emission from natural wetlands: Global distribution, area, and environmental characteristics of sources. *Global Biogeochemical Cycles*, *1*(1), 61–86. https://doi.org/10.1029/gb001i001p00061

Murguia-Flores, F., Arndt, S., Ganesan, A. L., Murray-Tortarolo, G., & Hornibrook, E. R. (2018). Soil Methanotrophy Model (MeMo v1. 0): A process-based model to quantify global uptake of atmospheric methane by soil. *Geoscientific Model Development*, *11*(6), 2009–2032. https://doi.org/10.5194/gmd-11-2009-2018