# CHAM-net: A Contrastive Hierarchical Adaptive Meta-network for Robust Global Methane Flux Prediction

**Rongchao Dong**[1] , **Yiming Sun**[1] , **Shuo Chen**[2] , **Youmi Oh**[3,4] , **Licheng Liu**[5] , **Yiqun Xie**[6] and **Xiaowei Jia**[1]
[1]University of Pittsburgh, Pittsburgh, PA, USA
[2]Purdue University, West Lafayette, IN, USA
[3]University of Colorado, Boulder, CO, USA
[4]NOAA Global Monitoring Laboratory, Boulder, CO, USA
[5]University of Wisconsin–Madison, Madison, WI, USA
[6]University of Maryland, College Park, MD, USA
{rongchaodong, yis108, xiaowei}@pitt.edu, chen4371@purdue.edu, youmi.oh@noaa.gov, licheng.liu@wisc.edu, xie@umd.edu



## Abstract

Methane is a potent greenhouse gas that significantly contributes to global warming. However, accurately estimating global methane emissions and consumption remains challenging due to the complex interactions among environmental drivers that may vary across spatial and temporal scales. Prior data-driven methods often overlook the inherent spatiotemporal heterogeneity of ecosystems, failing to explicitly capture site-specific characteristics and cross-year evolutionary dynamics. To address these issues, we propose the **C**ontrastive **H**ierarchical **A**daptive **M**eta-network (**CHAM-net**), a novel framework that explicitly learns from historical context to capture site-specific dynamics. CHAM-net employs a hierarchical encoder–decoder architecture, in which the encoder captures site-specific characteristics from historical data and then dynamically conditions the decoder to generate the final prediction. Experimental results demonstrate that CHAM-net consistently outperforms all baseline methods on both simulation and observational datasets for methane emission and consumption, achieving nRMSE values as low as 0.43 and 0.88 with corresponding $R^2$ scores up to 0.97 and 0.68 for emission prediction.[1]

## 1 Introduction

Methane ($CH_4$) is the second most significant greenhouse gas contributing to global warming after carbon dioxide ($CO_2$) and is responsible for about 30% of the increase in global temperature since the industrial revolution [Stocker, 2014]. Unlike carbon dioxide, methane is chemically reactive in the atmosphere and therefore has a relatively short atmospheric lifetime of about 9 years [Saunois *et al.*, 2025]. This short lifetime means that reducing methane emissions can deliver rapid climate benefits, including slower warming rates, reduced climate extremes, and improved air quality [Mar *et al.*, 2022], making methane mitigation one of the most effective near-term strategies for protecting human health, ecosystems, and vulnerable communities [Ocko *et al.*, 2021].

Traditional approaches primarily rely on process-based biogeochemistry models [Zhuang *et al.*, 2004; Tian *et al.*, 2015; Zhang *et al.*, 2017] to simulate and estimate the natural methane cycle. By incorporating the theoretical understanding of methane ecosystem dynamics with the key environmental drivers (e.g., soil features and temperatures), these models can be extended to enable global methane prediction and subsequent budget estimation. However, they are often limited by rigid and extensive parameterization, leading to biased prediction and substantial computational requirements when applied over large regions and long time periods. Recently, data-driven machine learning (ML) methods [Kim *et al.*, 2020; Irvin *et al.*, 2021; Luo *et al.*, 2023; Saha *et al.*, 2025] have emerged as a promising alternative, demonstrating strong capability in capturing non-linear relationships between environmental drivers and methane flux. Existing works have further explored knowledge transfer (e.g., pretraining and fine-tuning) from simulated data to sparse real observations (e.g., collected from eddy covariance towers) to improve the prediction accuracy [Sun *et al.*, 2025].

However, existing ML methods typically assume diverse methane sites across space are governed by a shared set of global parameters, and neglect the spatiotemporal heterogeneity. In particular, methane emission and consumption patterns are highly site-specific because different sites can respond differently to similar input drivers due to the underlying variations in microbial communities and soil properties. For example, Figure 1 (a) compares annual emissions at three sites (US.Myb [Matthes *et al.*, 2018], DE.Hte [Koebsch and Jurasinski, 2018], and US.ORv [Bohrer and Morin, 2015]) in 2013. These sites exhibit clearly different emission patterns and magnitudes. Without fully leveraging site-

[1]Extended version:https://arxiv.org/abs/2606.00338

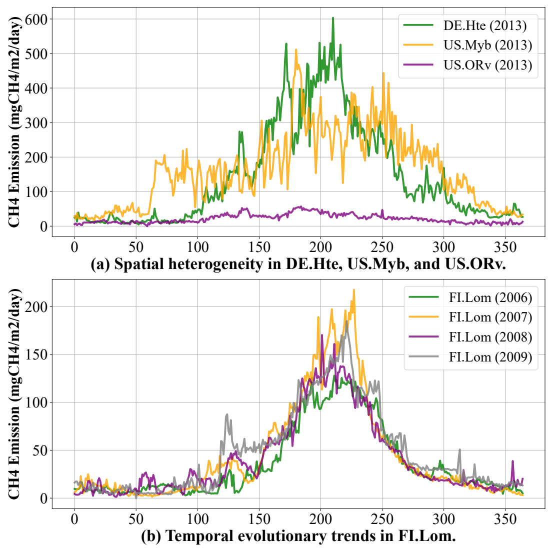


Figure 1: spatiotemporal heterogeneity within the datasets.

specific context, existing models tend to predict averaged values, which could underestimate high-valued regions and overestimate low-valued regions (e.g., the significant differences between sites DE.Hte and US.ORv in Figure 1 (a)).

Additionally, current ML methods mostly utilize short-term data (e.g., data from the current year) and focus on capturing short-term temporal dynamics (e.g., seasonal changes of precipitation). However, they are not designed to capture the impact of many long-term processes (e.g., slow changes in plant cover, soil composition) that also affect methane dynamics over years. As shown in Figure 1 (b), for site FI.Lom [Lohila *et al.*, 2010], we can observe a sustained decline in total methane emissions from 2006 to 2009, reflecting a site-specific cross-year temporal evolution pattern.

To address these limitations, we propose the **C**ontrastive **H**ierarchical **A**daptive **M**eta-network (**CHAM-net**), a novel framework that explicitly learns site-specific dynamics from multi-year historical data. While many site-specific characteristics (e.g., microbial communities and soil properties) are not directly observable, their influence is often manifested in the long-term trend shown in each site's historical methane record. CHAM-net therefore leverages each site's historical methane data to capture these latent characteristics and improve current-year prediction. Specifically, CHAM-net adopts a hybrid meta-learning mechanism [Hospedales *et al.*, 2021] in which the inner model encodes multi-year environmental and methane dynamics into learnable representations to summarize site-specific characteristics (e.g., temporal patterns and scales), while the outer model leverages these learned representations to condition the current-year prediction process. This design shifts prediction from a global model to a site-aware estimation. Additionally, optimizing the outer loop helps shape the inner-loop learning task, which enables effective extraction of site-specific information.

**Connections to the social good.** By applying our advanced AI architecture to resolve the spatiotemporal heterogeneity of methane ecosystems, this work significantly reduces uncertainties in natural methane budgets and improves process-level understanding, thereby enabling more effective and accurate methane mitigation strategies in support of the Global Methane Pledge [Malley *et al.*, 2023]. These advances directly contribute to multiple UN Sustainable Development Goals (https://sdgs.un.org/goals), including Climate Action, Good Health and Well-Being, and Life on Land, by informing near-term climate mitigation pathways and air-quality prediction, and supporting sustainable wetland management.

This paper is conducted in active collaboration with domain experts from NOAA Global Monitoring Laboratory, University of Wisconsin-Madison, University of Maryland, and Purdue University, who are internationally recognized experts in global methane observations, process understanding, and atmospheric data assimilation. These scientists contribute domain knowledge, observational constraints, and evaluate the real-world impact of these novel methodologies throughout model development, training, and interpretation.

We summarize our technical contributions as follows:

- We identify inherent site heterogeneity as a key factor underlying the inaccurate predictions of the current models, and show that historical data encodes critical site-specific information essential for effective site prediction.
- We propose CHAM-net, an encoder–decoder hybrid meta-learning framework that dynamically leverages historical data to calibrate site-specific predictions for the current year.
- We evaluate CHAM-net on extensive methane emission and consumption datasets, including both simulation and observational data. Experimental results show that CHAM-net consistently outperforms all baselines in all datasets.

## 2 Problem Formulation

The task of global methane flux prediction can be formulated as a site-level time-series regression problem. For each site $i \in \{1, \ldots, N\}$, we are given a sequence of environmental drivers (e.g., soil properties and temperature) over a time period, denoted as $X_i = \{x_1, x_2, \ldots, x_T\}$, where each $x_t \in \mathbb{R}^D$ is a feature vector of dimension $D$ at time $t$ (e.g., a specific date), $D$ is the total number of input drivers, and $T$ is the length of each sequence. Following prior works in methane prediction and other environmental monitoring tasks [Liu *et al.*, 2024a; Sun *et al.*, 2025], we cut the data into yearly sequences (i.e., $T$=365) to facilitate modeling seasonal patterns. The goal is to predict the methane flux for the corresponding year $Y_i = \{y_1, y_2, \ldots, y_T\}$, where $y_t \in \mathbb{R}$ is the target label, e.g., methane emission or consumption. In our proposed method, we also leverage multi-year historical records. For each site $i$, we use $X_i^{(k)}$ and $Y_i^{(k)}$ to represent the environmental drivers and methane data from a historical year $k \in \{1, \ldots, K\}$.

The datasets used in this paper can be categorized into simulation and observational datasets, each with distinct characteristics as follows:

- **Simulation Dataset** We use process-based simulation datasets, which incorporate biogeochemical processes to simulate methane fluxes by solving differential equations.

Given globally available input drivers, they enable estimation of methane fluxes at the global scale. However, due to the complexity of model computations and uncertainties in multiple input drivers, the highest spatial resolution is limited to 0.5 degree, corresponding to approximately 3,000 $\text{km}^2$ per grid.

- **Observational Dataset** In observational datasets, both input drivers and methane fluxes are directly measured at each site using eddy covariance techniques, providing real-world ground truth observations. However, these sites are spatially sparse and geographically discontinuous, with footprints on the order of hundreds of square meters. Moreover, observational datasets cover substantially shorter time spans than simulation datasets. The details of the datasets used in this paper are provided in Section 4.1.

# 3 Design and Methodology

In this section, we introduce the design of CHAM-net (Contrastive Hierarchical Adaptive Meta-network) architecture.

## 3.1 Model Overview

The proposed CHAM-net model explicitly incorporates historical information to learn site-specific characteristics and long-term dynamics. By employing a hybrid hierarchical meta-learning architecture, the model extracts the most informative historical trends and scales for each site. These learned site-specific representations are then injected into the decoder and combined with current-year inputs to improve the final prediction.

Figure 2 illustrates the overall architecture of the CHAM-net model. For each site, a configurable length of historical years' data is used as the **support set**, while the current year data forms the **query set**. When the historical year length exceeds one year, a cross-attention module is first applied to compute each year's relevance to the current-year inputs. The attention weights are then used to extract context-aware representations capturing local characteristics that influence current-year dynamics. These representations are then propagated to the decoder stage to guide the final prediction.

The workflow consists of two hierarchical phases:

- **Context Encoder.** The context encoder is designed to extract representations for capturing site-specific characteristics that affect current-year dynamics. Although many of these characteristics are not directly observable, they can be inferred from the dynamic responses of methane fluxes $Y$ to environmental drivers $X$, i.e., $Y = f(X; \theta)$. The key idea of the context encoder is to *inversely* infer these characteristics from historical methane data, by embedding them in a site-specific representation that serves as the parameters $\theta$ for the function $f$. However, directly inferring the $\theta$ as the full parameterization of the function $f$ may not precisely capture the site-specific characteristics as the mapping is also influenced by many confounding factors. Hence, we reformulate the model $f$ as $Y = f(g(X); \theta)$, where $g$ serves as a global feature extractor (via a bidirectional GRU) that captures shared underlying processes while $\theta$ parameterizes the site-specific information. The extractor $g$ is learned end-to-end under supervised training, which helps better define and stabilize the inverse problem for the target task. A contrastive learning objective is further introduced to amplify inter-site differences, which ensures the discrimination of site-specific characteristics.
- **Adaptive Decoder.** The resulting site-specific representations are then transformed and injected into the hidden state of a Long Short-Term Memory (LSTM) [Hochreiter and Schmidhuber, 1997]-based decoder. The decoder shares parameters across all sites, but conditions its temporal dynamics on the injected site context. The final output of methane fluxes is produced from this context-enhanced hidden state.

The model is trained through a bi-level optimization process. The inner-loop optimization aims to extract site-specific representations from historical data observations. In the outer loop, the model utilizes these representations to condition and enhance the prediction, while updating model parameters to optimize predictive performance and to shape the inner-loop objective.

In the following, we provide details of the model components as well as the training process.

## 3.2 Context Encoder

The goal of the encoder is to convert the historical support set $\mathcal{S}^{(k)} = \{(X_i^{t,(k)}, Y_i^{t,(k)})\}_{t=1}^{T}$ into a compact context-aware representation $\mathbf{W}_i$ for each site $i$, and then use it to enhance the decoder through site-specific conditioning.

**BiGRU Module**. We first encode data from each historical year $k$ using a bidirectional GRU:

$$\mathbf{G}_i^{(k)} = \text{BiGRU}(\mathbf{X}_i^{(k)}) \in \mathbb{R}^{T \times 2H}, \quad (1)$$

where $H$ is the size of the hidden dimension.

**Site-specific MLP**. To capture the site-specific characteristics, we assign each site $i$ a dedicated MLP projection head parameterized by $\mathbf{W}_i^k$, which can be represented by:

$$\hat{\mathbf{Y}}_i^{(k)} = \mathbf{G}_i^{(k)} \mathbf{W}_i^k + b_i, \quad (2)$$

where $\mathbf{W}_i^k \in \mathbb{R}^{2H \times 1}$ and $b_i \in \mathbb{R}$. $\hat{\mathbf{Y}}_i^{(k)}$ represents the predicted labels of year $k$, which are later used to compute the inner-loop loss.

In our proposed method, we use the learned $\mathbf{W}_i^k$ as the representation of site-specific behavior. The MLP weights effectively distill rich historical information from both the drivers and methane fluxes into a compact representation. The dimensionality of $\mathbf{W}_i^k$ can be adjusted based on the output dimension of the previous GRU layer.

**Year-wise Attention**. When incorporating multiple historical years (i.e., $K > 1$), we weight each year based on its relevance to the current-year prediction. We therefore compute attention weights that measure the year-wise importance $\alpha_i^{(k)}$ of each historical year, as follows:

$$\alpha_i^{(k)} = \text{Attn}\big(\mathbf{X}_i^{(0)}, \mathbf{X}_i^{(k)}\big), \quad \text{s.t.,} \sum_{k=1}^{K} \alpha_i^{(k)} = 1, \quad (3)$$

where $\mathbf{X}_i^{(0)}$ is the current-year inputs. We employ the standard multiple-head attention mechanism to calculate the

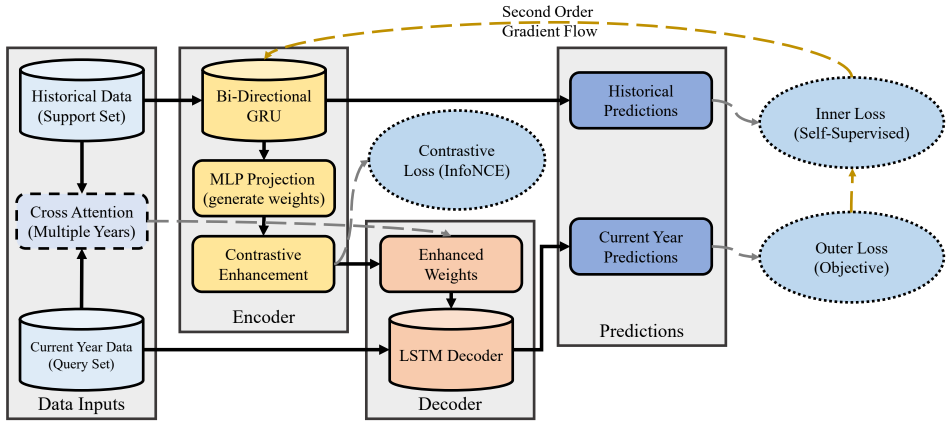


Figure 2: CHAM-net structure overview.

similarity between the current year and the historical years, and then normalize the attention weights across $K$ historical years via softmax. The attention weights are used in the outer-loop for aggregating $\mathbf{W}_i^k$ from historical years, which will be discussed in Section 3.3.

**Contrastive Learning**. To better ensure site discrimination and encode persistent site characteristics (e.g., long-term scaling and trend patterns) into $\mathbf{W}_i^k$, we introduce a contrastive objective directly defined over the site-specific representation $\mathbf{W}_i^k$. We treat the representation from the same site but different years as positive pairs $(\mathbf{W}_i^k, \mathbf{W}_i^+)$ and other sites in the batch as negatives. The contrastive loss is Info Noise-Contrastive Estimation (InfoNCE) loss, which is defined as:

$$\mathcal{L}_{cont} = -\log \frac{\exp(\text{sim}(\mathbf{W}_i^k, \mathbf{W}_i^+)/\tau)}{\sum_{j=1}^{B} \exp(\text{sim}(\mathbf{W}_i^k, \mathbf{W}_j)/\tau)}, \tag{4}$$

where $B$ is the batch size, $\text{sim}(\cdot)$ is the cosine similarity, and $\tau$ is the configurable temperature parameter.

To improve the robustness of site-specific representations, we adopt a stochastic perturbation method to augment the generated $\mathbf{W}_i^k$ in contrastive learning:

$$\tilde{\mathbf{W}}_i^k = \mathbf{W}_i^k + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{5}$$

This augmentation encourages the embeddings to remain stable under small variations of site-specific parameters.

### 3.3 Adaptive Decoder

The decoder serves as the base-learner that leverages the outcome of the encoder to condition the actual flux prediction on the Query Set $\mathcal{Q}_i = \{X_{curr}^t\}_{t=1}^T$. In our implementation, the decoder is an LSTM-based model. We adopt LSTM because LSTM-based architectures have consistently demonstrated robust performance in prior methane flux prediction studies [Luo *et al.*, 2023; Chen *et al.*, 2024; Sun *et al.*, 2025].

We first aggregate $\tilde{\mathbf{W}}_i^k$ from multiple historical years using attention weights $\alpha_i^{(k)}$, and then project the site-specific embedding into the hidden state space of the decoder:

$$\tilde{\mathbf{z}}_i = \phi(\sum_{k=1}^{K} \alpha_i^{(k)} \tilde{\mathbf{W}}_i^k), \tag{6}$$

where $\phi(\cdot)$ is a fully connected projection.

The $\tilde{\mathbf{z}}_i$ will be further normalized with the original hidden state of the LSTM decoder with a learnable weight $\beta_i$:

$$\tilde{\mathbf{h}}_i = \frac{\tilde{\mathbf{h}}_i + \beta_i \tilde{\mathbf{z}}_i}{1 + |\beta_i|}, \tag{7}$$

Then the final prediction is obtained by putting the current-year inputs and the enhanced hidden state into the LSTM decoder:

$$\hat{\mathbf{Y}}_i^{(0)} = \text{LSTM}_\psi(\mathbf{X}_i^{(0)}, \tilde{\mathbf{h}}_i), \tag{8}$$

where $\psi$ denotes the other parameters of the LSTM, and $\hat{\mathbf{Y}}_i^{(0)}$ is the predicted current-year labels.

### 3.4 Optimization and Loss Functions

The training of CHAM-net is formulated as a bilevel optimization problem, involving nested inner- and outer-loop loss computations and the corresponding bilevel backpropagation process. The second-order backpropagation gradient flow is illustrated in Figure 2.

**Inner-loop Reconstruction Loss**. The inner-loop objective is designed to extract site representations using historical information. It consists of a historical reconstruction loss and a contrastive enhancement loss:

$$\mathcal{L}_{\text{inner}}(\theta) = \mathcal{L}_{\text{hist}}(\theta) + \lambda_{\text{con}} \mathcal{L}_{\text{con}}(\theta), \tag{9}$$

where $\theta = \{\mathbf{W}_i^k\}_{k=1}^K$ is the parameters in the inner-loop, $\mathcal{L}_{\text{con}}(\theta)$ is the contrastive loss defined in Equation 4, and $\lambda_{\text{con}}$ is a configurable hyperparameter that controls the contribution of the contrastive loss. The historical reconstruction loss $\mathcal{L}_{\text{hist}}(\theta)$ is defined as:

$$\mathcal{L}_{\text{hist}}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \alpha_i^{(k)} \left\| \hat{\mathbf{Y}}_i^{(k)} - \mathbf{Y}_i^{(k)} \right\|_2^2, \tag{10}$$

where $N$ is the total number of sites in training.

The encoder parameters are updated through a differentiable inner optimization step:

$$\theta' = \theta - \gamma \nabla_\theta \mathcal{L}_{\text{inner}}(\theta), \tag{11}$$

where $\gamma$ is the inner learning rate.

**Outer-loop Objective**. The outer-loop takes the learned representation and performs current-year prediction, and the loss is calculated by the mean squared error across all the training sites:

$$\mathcal{L}_{\text{outer}}(\psi, \theta^*) = \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\mathbf{Y}}_i^{(0)}(\theta', \psi) - \mathbf{Y}_i^{(0)} \right\|_2^2 \qquad (12)$$

The outer optimization updates the parameters $\psi$ for both the encoder and decoder parameters, i.e., the BiGRU parameters in the encoder and the LSTM parameters in the decoder. $\theta^*$ represents the updated site-specific representations obtained from the inner-loop (via Eq. 11). Given that $\theta^*$ also contributes to the final prediction, the outer-loop errors are backpropagated to the representations $\theta$, which forms a second-order gradient flow during the training process.

The complete training objective is formulated as follows:

$$\begin{aligned} \min_{\psi} \quad & \mathcal{L}_{\text{outer}}(\psi, \theta^*) \\ \text{s.t.} \quad & \theta^* = \arg\min \mathcal{L}_{\text{inner}}(\theta) \end{aligned} \qquad (13)$$

where $\psi$ is the parameters of the outer LSTM, $\theta$ is the parameters of inner-loop. Note that the outer-loop parameters $\psi$ are shared across all sites while the inner-loop parameters $\theta$ are specific to each site.

***Why meta-learning*** Traditional methane flux prediction models either rely on globally shared parameters or fail to fully exploit historical information. In our setting, historical context must be explicitly encoded into the prediction process. However, since input drivers and labels vary dynamically across training and evaluation, efficiently integrating this information poses a challenge. Meta-learning provides a natural solution. Each site can be viewed as a distinct and evolving task, characterized by unique long-term dynamics and temporal responses. Through the inner-loop, the model learns site-specific historical information and dynamically adapts it to the outer-loop objectives, offering an effective mechanism for improving the prediction accuracy.

# 4 Evaluation

## 4.1 Experimental Setup

**Baselines.** We compare CHAM-net against nine competitive long-term time-series forecasting models. These include Transformer-based methods, such as the original Transformer [Vaswani *et al.*, 2017], iTransformer [Liu *et al.*, 2024b], Pyraformer [Liu *et al.*, 2022], DUET [Qiu *et al.*, 2025], and PatchTST [Nie *et al.*, 2022], as well as two MLP-based architectures, TSMixer [Chen *et al.*, 2023] and TimeMixer [Wang *et al.*, 2024]. We further include two RNN-based approaches, namely LSTM [Hochreiter and Schmidhuber, 1997] and P-sLSTM [Kong *et al.*, 2025]. Notably, CHAM-net can also be regarded as an RNN-based model, as it relies on BiGRU and LSTM components to extract historical information and condition the prediction.

**Datasets.** We evaluate all models using both simulation and observational datasets. An overview of the datasets is given below, with more details provided in the extended version and the open-source repository[2].

- *Emission Datasets*. We leverage the methane emission datasets introduced in [Sun *et al.*, 2025], which include a global simulation dataset at 0.5-degree spatial resolution with daily granularity, TEM [Zhuang *et al.*, 2004] (**TEM-E**), and an observational dataset, FLUXNET-CH4 [Delwiche *et al.*, 2021] (**FLUXNET-E**).
- *Consumption Datasets*. We use two consumption simulation datasets, **TEM-C** [Zhuang *et al.*, 2013] and **MeMo** [Murguia-Flores *et al.*, 2018; Murguia-Flores *et al.*, 2021]. We additionally include an observational dataset, **FLUXNET-C** derived from [Delwiche *et al.*, 2021], which consists of methane sinks from 28 upland sites.

**Data Splitting and Evaluation Task.** We focus on *temporal extrapolation* for both methane emission and consumption tasks. Following [Sun *et al.*, 2025], simulation datasets are split chronologically into two equal halves, with the earlier period used for training and the later period used for testing. For the observational datasets FLUXNET-E and FLUXNET-C, data are split on a per-site basis, where six-sevenths of the temporal records are used for training and the remaining one-seventh for testing. This 6/7–1/7 split ensures sufficient training data for effective model learning while mitigating overfitting risks.

**Evaluation Metrics.** We use normalized root of mean squared error (nRMSE) and the coefficient of determination ($R^2$) to evaluate model performance. nRMSE measures the magnitude of prediction errors normalized by the scale of the observations, enabling fair comparison across sites with substantially different magnitudes. Lower nRMSE indicates better performance. $R^2$ quantifies the proportion of variance in the observations explained by the model, reflecting its ability to capture the temporal patterns. Higher $R^2$ values correspond to a better model fit.

## 4.2 Results

### Predictive performance

Table 1 reports the predictive performance on all simulation and observational datasets. For each column, models are trained and evaluated solely on the corresponding dataset, without pretraining or fine-tuning across datasets. We observe that CHAM-net consistently outperforms all baselines in terms of both nRMSE and $R^2$. Specifically, on the emission datasets TEM-E and FLUXNET-E, CHAM-net achieves nRMSE values of 0.43 and 0.88, respectively. The nRMSEs are at least 0.1 lower than those of competing methods, indicating its superior ability to capture the magnitude of methane emissions over time. CHAM-net also attains the highest $R^2$ value on FLUXNET-E, showing more accurate modeling of the emission patterns. The combination of lower nRMSE and higher $R^2$ demonstrates that CHAM-net effectively captures both emission scales and temporal dynamics, which is critical for reliable future methane budget estimation. Moreover, the improvement reflects that the learned site-specific representations provide valuable insights into localized emission behavior, which can help inform targeted methane mitigation strategies.

For the consumption datasets, CHAM-net demonstrates substantial improvements on TEM-C, reducing nRMSE by more than 0.4 and increasing $R^2$ by over 0.2, indicating

[2]Dataset and code link: https://zenodo.org/records/20450697.

a significant performance gain. For the MeMo dataset, which is generated using relatively simpler biogeochemical equations and has more regular patterns, nearly all models achieve lower nRMSE and higher $R^2$ compared to the TEM-C dataset. Results on both TEM-C and MeMo demonstrate that explicitly leveraging historical information substantially enhances predictive performance for current-year consumption. For the observational dataset FLUXNET-C, model performance is generally limited due to the small magnitude of consumption fluxes and high levels of environmental noise. Nevertheless, CHAM-net still achieves the highest $R^2$ value of 0.31 and the lowest nRMSE of 1.27 among all methods.

## Case Analysis

Figure 3 presents two representative site examples from the FLUXNET-E dataset. We compare CHAM-net with the two best baseline models, Pyraformer and TSMixer. As shown in Figure 3 (a), CHAM-net achieves a closer match to both the scale and temporal evolution of methane emissions at site FI-Lom, particularly after day 200. While all three models capture the overall temporal pattern, CHAM-net more accurately estimates the peak value. In contrast, Pyraformer overestimates the peak value, whereas TSMixer underestimates it. Figure 3 (b) illustrates results for site DE-SfN [Schmid and Klatt, 2014], which represents a particularly challenging case due to the presence of negative flux values in the ground truth. Even under this difficult setting, CHAM-net more effectively captures the site-specific emission scale than the baseline models, resulting in higher $R^2$, lower nRMSE, and a total emission estimate that is closer to the observed values.

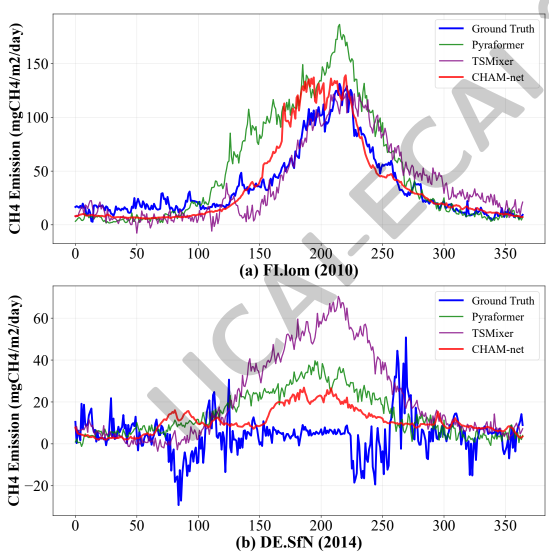


Figure 3: Predictions on representative FLUXNET-E sites.

## Model Analysis

*Historical Year Length.* We first investigate the optimal length of historical data used in the inner-loop of CHAM-net. As shown in Figure 4, we report performance using the $R^2$ value to compare different historical year window lengths. We evaluate historical year lengths of 2, 4, 6, and 8 years to determine the optimal setting for the simulation datasets TEM-E, TEM-C, and MeMo. Due to the sparsity and varying temporal spans of the observational datasets (e.g., some sites contain only three years of data), we use a single historical year length for FLUXNET-E and FLUXNET-C datasets in our experiments. This setting is sufficient to yield explicit performance improvements, as demonstrated in Table 1. For the simulation datasets, Figure 4 shows that performance varies marginally across different lengths of historical window. Nevertheless, a four-year historical window consistently achieves the best performance.

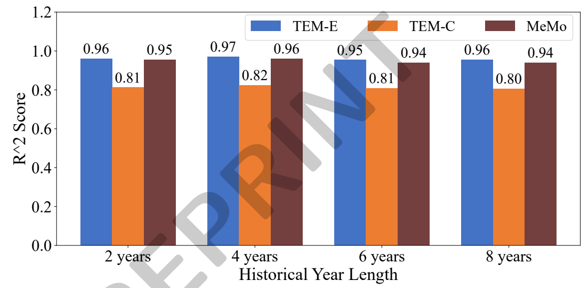


Figure 4: Sensitivity study of CHAM-net with respect to historical year length.

*Learned Representations Analysis.* To interpret the learned representations and identify the most influential inputs, we analyze the correlation between the learned representations and the input variables. Taking FLUXNET-E as an example, we extract the site-specific representations for all sites after training and apply principal component analysis (PCA) to identify the top three dominant components. We then compute the Pearson product–moment correlation coefficients between these components and the site-wise input features.

As shown in the heatmap in Figure 5, the learned representations capture meaningful physical relationships when encoding historical information. In the figure, the x-axis represents the selected 12 input features, while the y-axis corresponds to the three most important components identified by PCA, namely W1 to W3. In the heatmap, positions in read are indicative of strong associations. The results show that BD (`topsoil bulk density`), vegetation types (`cltveg` and `vegetation`), `climate`, solar radiation (`SOLR`), and air temperature (`TAIR`) are among the most influential features. These findings are consistent with the established understanding of methane processes in natural ecosystems and suggest that the learned representations can provide useful insights for improved methane forecasting and mitigation strategies.

## Ablation Studies

We also conduct the ablation experiments to examine the contribution of each component in our framework. Using the emission datasets as an example, we evaluate three variants of CHAM-net: (i) a mean-pooling-based CHAM-net, (ii) CHAM-net with cross-attention, and (iii) CHAM-net

| Methods | TEM-E | | FLUXNET-E | | TEM-C | | MeMo | | FLUXNET-C | |
|---|---|---|---|---|---|---|---|---|---|---|
| | nRMSE | $R^2$ | nRMSE | $R^2$ | nRMSE | $R^2$ | nRMSE | $R^2$ | nRMSE | $R^2$ |
| LSTM | 0.57 | 0.94 | 1.08 | 0.53 | 1.56 | 0.56 | 0.22 | 0.89 | 1.31 | 0.27 |
| P-sLSTM | 1.34 | 0.71 | 1.06 | 0.54 | 1.82 | 0.41 | 0.30 | 0.82 | 1.41 | 0.16 |
| Transformer | 0.52 | 0.95 | 1.09 | 0.52 | 1.45 | 0.62 | 0.20 | 0.91 | 1.40 | 0.16 |
| iTransformer | 1.24 | 0.75 | 1.29 | 0.33 | 1.64 | 0.52 | 0.29 | 0.83 | 1.36 | 0.21 |
| Pyraformer | 0.92 | 0.87 | 1.05 | 0.56 | 1.49 | 0.60 | 0.22 | 0.89 | 1.29 | 0.29 |
| TSMixer | 0.55 | 0.95 | 0.99 | 0.60 | 1.47 | 0.61 | 0.24 | 0.88 | 1.35 | 0.23 |
| TimeMixer | 1.85 | 0.45 | 1.56 | 0.02 | 1.79 | 0.42 | 0.30 | 0.82 | 1.48 | 0.07 |
| PatchTST | 1.81 | 0.48 | 1.26 | 0.36 | 1.93 | 0.33 | 0.43 | 0.62 | 1.38 | 0.20 |
| DUET | 1.62 | 0.58 | 1.18 | 0.44 | 1.81 | 0.41 | 0.29 | 0.83 | 1.38 | 0.19 |
| **CHAM-net** | **0.43** | **0.97** | **0.88** | **0.68** | **0.99** | **0.82** | **0.15** | **0.95** | **1.27** | **0.31** |

Table 1: Main results across all datasets and models. Lower nRMSE and higher $R^2$ indicate better performance.

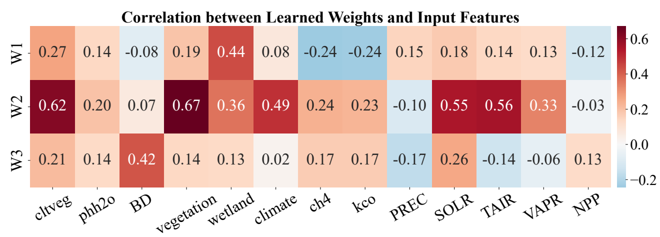


Figure 5: Correlation between learned weights and input features.

with both cross-attention and contrastive learning. For reference, we also include the LSTM as baseline.

As shown in the Figure 6, the incorporation of historical information accounts for the largest performance improvement. Contrastive learning provides the second-largest gain, as it enhances the model's ability to distinguish site-specific characteristics and extract informative features from historical data, leading to more expressive embeddings. Cross-attention yields a relatively smaller improvement in this setting. This is because input features evolve slowly over time, resulting in high similarity across historical years. Thus, it offers limited benefit over mean pooling. Nevertheless, cross-attention remains a configurable component and may offer greater benefits when applied to other domains where inputs diversity across years is more pronounced.

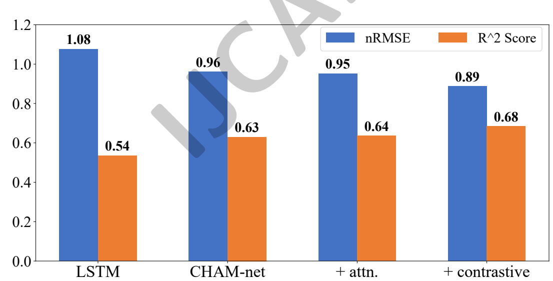


Figure 6: Ablation study of different model variants (in $R^2$).

## 5 Related Work

**Methane Prediction.** Previous methane flux prediction studies [Kim *et al.*, 2020; Irvin *et al.*, 2021; Luo *et al.*, 2023; Chen *et al.*, 2024; Saha *et al.*, 2025] primarily rely on Random Forest, decision tree, extreme gradient boosting (XGB), and Artificial Neural Network (ANN) approaches trained on relatively small observational datasets, which often focused on specific regions and coarse spatial resolutions. The work of [Sun *et al.*, 2025] introduced the first global wetland methane emission dataset that integrates both simulation and observational data, enabling large-scale machine learning studies. To the best of our knowledge, this work is the first to jointly model methane emission and consumption in natural ecosystems using machine learning, achieving state-of-the-art performance and providing new insights for global methane budget estimation and analysis.

**Knowledge-guided machine learning.** Knowledge-guided machine learning (KGML) [Willard *et al.*, 2022; Karpatne *et al.*, 2024; Yu *et al.*, 2025] has been successfully applied in various environmental studies, including carbon dioxide modeling [Liu *et al.*, 2024a] and lake temperature profiling [Jia *et al.*, 2021; Yu *et al.*, 2024], by embedding physical knowledge into the loss functions of ML models to enhance performance. While directly incorporating biogeochemical equations into methane prediction models is beyond the scope of this work, it represents a promising direction for future research.

## 6 Conclusion

In this paper, we propose a contrastive hierarchical adaptive meta-learning framework that explicitly leverages site-specific historical information to capture both spatial heterogeneity and temporal dynamics in methane prediction. By learning site-aware representations, our model improves prediction accuracy for both methane emission and consumption across simulation and observational datasets. These improvements support more reliable estimation of the global methane budget and provide insights that may inform effective strategies for natural methane mitigation. Experimental results demonstrate that our approach consistently outperforms all other baselines, achieving the lowest nRMSE and the highest $R^2$ on both methane emission and consumption datasets.

## Acknowledgements

Rongchao Dong and Xiaowei Jia were partially supported by the National Science Foundation (NSF) grants 2203581, 2239175, 2316305, 2147195, 2425845, and 2530609; the USGS award G22AC00266; and the NASA grants 80NSSC24K1061 and 80NSSC25K0013. Licheng Liu and Youmi Oh are supported by the Department of Energy (DOE) grant DE-SC0024360, NSF ESIIL 2153040. Yiqun Xie is supported in part by the NSF under Grant No. 2126474, 2147195, 2425844, and 2530610; NASA under grant 80NSSC25K0013 and 80NSSC25K7221; Google's AI for Social Good Impact Scholars program. We also sincerely thank all reviewers for their thoughtful comments and feedback, and all our collaborators for their insightful contributions.

## References

[Bohrer and Morin, 2015] Gil Bohrer and Timothy Morin. FLUXNET-CH4 US-ORv Olentangy River Wetland Research Park, 2015. Dataset. Time range: 2011-2015.

[Chen *et al.*, 2023] Si-An Chen, Chun-Liang Li, Nate Yoder, Sercan Ö. Arik, and Tomas Pfister. Tsmixer: An all-mlp architecture for time series forecasting. *CoRR*, abs/2303.06053, 2023.

[Chen *et al.*, 2024] Shuo Chen, Licheng Liu, Yuchi Ma, Qianlai Zhuang, and Narasinha J Shurpali. Quantifying global wetland methane emissions with in situ methane flux data and machine learning approaches. *Earth's Future*, 12(11):e2023EF004330, 2024.

[Delwiche *et al.*, 2021] Kyle B Delwiche, Sara Helen Knox, Avni Malhotra, Etienne Fluet-Chouinard, Gavin McNicol, Sarah Feron, Zutao Ouyang, Dario Papale, Carlo Trotta, Eleonora Canfora, et al. Fluxnet-ch4: A global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands. *Earth System Science Data Discussions*, 2021:1–111, 2021.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[Hospedales *et al.*, 2021] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):5149–5169, 2021.

[Irvin *et al.*, 2021] Jeremy Irvin, Sharon Zhou, Gavin McNicol, Fred Lu, Vincent Liu, Etienne Fluet-Chouinard, Zutao Ouyang, Sara Helen Knox, Antje Lucas-Moffat, Carlo Trotta, et al. Gap-filling eddy covariance methane fluxes: Comparison of machine learning model predictions and uncertainties at fluxnet-ch4 wetlands. *Agricultural and Forest Meteorology*, 308:108528, 2021.

[Jia *et al.*, 2021] Xiaowei Jia, Jared Willard, Anuj Karpatne, Jordan S Read, Jacob A Zwart, Michael Steinbach, and Vipin Kumar. Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles. *ACM/IMS Transactions on Data Science*, 2(3):1–26, 2021.

[Karpatne *et al.*, 2024] Anuj Karpatne, Xiaowei Jia, and Vipin Kumar. Knowledge-guided machine learning: Current trends and future prospects. *arXiv preprint arXiv:2403.15989*, 2024.

[Kim *et al.*, 2020] Yeonuk Kim, Mark S Johnson, Sara H Knox, T Andrew Black, Higo J Dalmagro, Minseok Kang, Joon Kim, and Dennis Baldocchi. Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis. *Global Change Biology*, 26(3):1499–1518, 2020.

[Koebsch and Jurasinski, 2018] Franziska Koebsch and Gerald Jurasinski. FLUXNET-CH4 DE-Hte Huetelmoor, 2018. Dataset. Time range: 2011-2018.

[Kong *et al.*, 2025] Yaxuan Kong, Zepu Wang, Yuqi Nie, Tian Zhou, Stefan Zohren, Yuxuan Liang, Peng Sun, and Qingsong Wen. Unlocking the power of lstm for long term time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 11968–11976, 2025.

[Liu *et al.*, 2022] Shizhan Liu, Hang Yu, Cong Liao, Jianguo Li, Weiyao Lin, Alex X. Liu, and Schahram Dustdar. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

[Liu *et al.*, 2024a] Licheng Liu, Wang Zhou, Kaiyu Guan, Bin Peng, Shaoming Xu, Jinyun Tang, Qing Zhu, Jessica Till, Xiaowei Jia, Chongya Jiang, et al. Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems. *Nature communications*, 15(1):357, 2024.

[Liu *et al.*, 2024b] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024.

[Lohila *et al.*, 2010] Annalea Lohila, Mika Aurela, Juha-Pekka Tuovinen, Tuomas Laurila, Juha Hatakka, Juuso Rainne, and Timo Mäkelä. FLUXNET-CH4 FI-Lom Lompolojankka, 2010. Dataset. Time range: 2006-2010.

[Luo *et al.*, 2023] Ran Luo, Jingyi Wang, and Ian Gates. Machine learning for accurate methane concentration predictions: short-term training, long-term results. *Environmental Research Communications*, 5(8):081003, 2023.

[Malley *et al.*, 2023] Christopher S Malley, Nathan Borgford-Parnell, Seraphine Haeussling, Ioli C Howard, Elsa N Lefèvre, and Johan CI Kuylenstierna. A roadmap to achieve the global methane pledge. *Environmental Research: Climate*, 2(1):011003, 2023.

[Mar *et al.*, 2022] Kathleen A Mar, Charlotte Unger, Ludmila Walderdorff, and Tim Butler. Beyond co2 equivalence: The impacts of methane on climate, ecosystems, and health. *Environmental science & policy*, 134:127–136, 2022.

[Matthes *et al.*, 2018] Jaclyn Matthes, Cove Sturtevant, Patty Oikawa, Samuel Chamberlain, Daphne Szutu, Ariane Ortiz, Joseph Verfaillie, and Dennis Baldocchi. FLUXNET-CH4 US-Myb Mayberry Wetland, 2018. Dataset. Time range: 2010-2018.

[Murguia-Flores *et al.*, 2018] Fabiola Murguia-Flores, Sandra Arndt, Anita L Ganesan, Guillermo Murray-Tortarolo, and Edward RC Hornibrook. 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, 2018.

[Murguia-Flores *et al.*, 2021] Fabiola Murguia-Flores, Anita L Ganesan, Sandra Arndt, and Edward RC Hornibrook. Global uptake of atmospheric methane by soil from 1900 to 2100. *Global Biogeochemical Cycles*, 35(7):e2020GB006774, 2021.

[Nie *et al.*, 2022] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *CoRR*, abs/2211.14730, 2022.

[Ocko *et al.*, 2021] Ilissa B Ocko, Tianyi Sun, Drew Shindell, Michael Oppenheimer, Alexander N Hristov, Stephen W Pacala, Denise L Mauzerall, Yangyang Xu, and Steven P Hamburg. Acting rapidly to deploy readily available methane mitigation measures by sector can immediately slow global warming. *Environmental Research Letters*, 16(5):054042, 2021.

[Qiu *et al.*, 2025] Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. Duet: Dual clustering enhanced multivariate time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pages 1185–1196, 2025.

[Saha *et al.*, 2025] Esha Saha, Oscar Wang, Amit K Chakraborty, Pablo Venegas Garcia, Russell Milne, and Hao Wang. Dispersion based recurrent neural network model for methane monitoring in albertan tailings ponds. *Journal of Environmental Management*, 395:127748, 2025.

[Saunois *et al.*, 2025] Marielle Saunois, Adrien Martinez, Benjamin Poulter, Zhen Zhang, Peter A Raymond, Pierre Regnier, Josep G Canadell, Robert B Jackson, Prabir K Patra, Philippe Bousquet, et al. Global methane budget 2000–2020. *Earth System Science Data*, 17(5):1873–1958, 2025.

[Schmid and Klatt, 2014] Hans Schmid and Janina Klatt. FLUXNET-CH4 DE-SfN Schechenfilz Nord, 2014. Dataset. Time range: 2012-2014.

[Stocker, 2014] Thomas Stocker. *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge university press, 2014.

[Sun *et al.*, 2025] Yiming Sun, Shuo Chen, Shengyu Chen, Chonghao Qiu, Licheng Liu, Youmi Oh, Sparkle L Malone, Gavin McNicol, Qianlai Zhuang, Chris Smith, et al. X-methanewet: A cross-scale global wetland methane emission benchmark dataset for advancing science discovery with ai. *arXiv preprint arXiv:2505.18355*, 2025.

[Tian *et al.*, 2015] Hanqin Tian, Guangsheng Chen, Chaoqun Lu, Xiaofeng Xu, Wei Ren, Bowen Zhang, Kamaljit Banger, Bo Tao, Shufen Pan, Mingliang Liu, et al. Global methane and nitrous oxide emissions from terrestrial ecosystems due to multiple environmental changes. *Ecosystem Health and Sustainability*, 1(1):1–20, 2015.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[Wang *et al.*, 2024] Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024.

[Willard *et al.*, 2022] Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, 55(4):1–37, 2022.

[Yu *et al.*, 2024] Runlong Yu, Chonghao Qiu, Robert Ladwig, Paul C Hanson, Yiqun Xie, Yanhua Li, and Xiaowei Jia. Adaptive process-guided learning: An application in predicting lake do concentrations. In *2024 IEEE International Conference on Data Mining (ICDM)*, pages 580–589. IEEE, 2024.

[Yu *et al.*, 2025] Runlong Yu, Chonghao Qiu, Robert Ladwig, Paul Hanson, Yiqun Xie, and Xiaowei Jia. Physics-guided foundation model for scientific discovery: An application to aquatic science. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 28548–28556, 2025.

[Zhang *et al.*, 2017] Zhen Zhang, Niklaus E Zimmermann, Andrea Stenke, Xin Li, Elke L Hodson, Gaofeng Zhu, Chunlin Huang, and Benjamin Poulter. Emerging role of wetland methane emissions in driving 21st century climate change. *Proceedings of the National Academy of Sciences*, 114(36):9647–9652, 2017.

[Zhuang *et al.*, 2004] Qianlai Zhuang, Jerry M Melillo, David W Kicklighter, Ronald G Prinn, A David McGuire, Paul A Steudler, Benjamin S Felzer, and Shaomin Hu. 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), 2004.

[Zhuang *et al.*, 2013] Qianlai Zhuang, Min Chen, Kai Xu, Jinyun Tang, Eri Saikawa, Yanyu Lu, Jerry M Melillo, Ronald G Prinn, and A David McGuire. Response of global soil consumption of atmospheric methane to changes in atmospheric climate and nitrogen deposition. *Global Biogeochemical Cycles*, 27(3):650–663, 2013.