Journal
Environmental Research Letters, 19(11)

ISSN
1748-9318

Authors
Zhu, Qing
Yuan, Kunxiaojia
Li, Fa
et al.

Publication Date
2024-11-01

DOI
10.1088/1748-9326/ad8019

ENVIRONMENTAL RESEARCH
LETTERS

LETTER • OPEN ACCESS

Critical needs to close monitoring gaps in pan-tropical wetland CH_4 emissions

To cite this article: Qing Zhu *et al* 2024 *Environ. Res. Lett.* **19** 114046

IOP Publishing   Environ. Res. Lett. 19 (2024) 114046   https://doi.org/10.1088/1748-9326/ad8019

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
20 May 2024

REVISED
5 September 2024

ACCEPTED FOR PUBLICATION
26 September 2024

PUBLISHED
7 October 2024

LETTER

Critical needs to close monitoring gaps in pan-tropical wetland CH_4 emissions

Qing Zhu[1,*], Kunxiaojia Yuan[1], Fa Li[2,4], William J Riley[1], Alison Hoyt[2], Robert Jackson[2], Gavin McNicol[3], Min Chen[4], Sara H Knox[5,6], Otto Briner[3], David Beerling[7], Nicola Gedney[8], Peter O Hopcroft[9], Akihito Ito[10], Atul K Jain[11], Katherine Jensen[12], Thomas Kleinen[13], Tingting Li[14], Xiangyu Liu[15], Kyle C McDonald[12], Joe R Melton[16], Paul A Miller[17], Jurek Müller[18], Changhui Peng[19], Benjamin Poulter[20], Zhangcai Qin[21], Shushi Peng[22], Hanqin Tian[23], Xiaoming Xu[11], Yuanzhi Yao[23], Yi Xi[24], Zhen Zhang[25], Wenxin Zhang[17], Qiuan Zhu[26] and Qianlai Zhuang[15]

1 Climate and Ecosystem Sciences Division, Climate Sciences Department, Lawrence Berkeley National Laboratory, Berkeley, CA, United States of America
2 Department of Earth System Science, Stanford University, Stanford, CA, United States of America
3 Department of Earth and Environmental Sciences, University of Illinois Chicago, Chicago, IL, United States of America
4 Department of Forest and Wildlife Ecology, University of Wisconsin Madison, Madison, WI, United States of America
5 Department of Geography, The University of British Columbia, Vancouver, British Columbia, Canada
6 Department of Geography, McGill University, Montreal, Quebec, Canada
7 School of Biosciences, University of Sheffield, Sheffield, United Kingdom
8 MetOffice Hadley Centre, Joint Centre for Hydrometeorological Research, Wallingford, United Kingdom
9 School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham, United Kingdom
10 Department of Forest Science, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan
11 Department of Atmospheric Sciences, University of Illinois, Urbana, IL 61821, United States of America
12 Department of Earth and Atmospheric Sciences, City College of New York, City University of New York, New York City, NY, United States of America
13 Max Planck Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany
14 LAPC, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, People's Republic of China
15 Department of Earth, Atmospheric, Planetary Sciences, Purdue University, West Lafayette, IN, United States of America
16 Climate Research Division, Environment and Climate Change Canada, Victoria, BC, Canada
17 Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, 223 62 Lund, Sweden
18 Climate and Environmental Physics, Physics Institute and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
19 Department of Biology Sciences, University of Quebec at Montreal, C.P. 8888, Succ. Centre-Ville, Montreal, QC H3C 3P8, Canada
20 NASA Goddard Space Flight Center, Biospheric Science Laboratory, Greenbelt, MD 20771, United States of America
21 School of Atmospheric Sciences, Sun-Yat-Sen University, and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519000, People's Republic of China
22 Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100871, People's Republic of China
23 Schiller Institute for Integrated Science and Society, Department of Earth and Environmental Sciences, Boston College, Chestnut Hill, MA 02467, United States of America
24 Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL (CEA-CNRS-UVSQ), Université Paris-Saclay, 91191 Gif-sur-Yvette, France
25 National Tibetan Plateau Data Center (TPDC), State Key Laboratory of Tibetan Plateau Earth System, Environment and Resource (TPESER), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, People's Republic of China
26 College of Geography and Remote Sensing, Hohai University, Nanjing 210098, People's Republic of China
* Author to whom any correspondence should be addressed.

E-mail: qzhu@lbl.gov

Keywords: spatial representativeness, global freshwater wetlands, CH4 emissions

Supplementary material for this article is available online

Abstract

Global wetlands are the largest and most uncertain natural source of atmospheric methane (CH_4). The FLUXNET-CH_4 synthesis initiative has established a global network of flux tower infrastructure, offering valuable data products and fostering a dedicated community for the measurement and analysis of methane flux data. Existing studies using the FLUXNET-CH_4 Community Product v1.0 have provided invaluable insights into the drivers of ecosystem-to-regional spatial patterns and daily-to-decadal temporal dynamics in temperate, boreal, and Arctic climate regions. However, as the wetland CH_4 monitoring network grows, there is a critical knowledge gap about where new monitoring infrastructure ought to be located to improve understanding of the global wetland CH_4 budget. Here we address this gap with a spatial representativeness analysis at existing and hypothetical observation sites, using 16 process-based wetland biogeochemistry models and machine learning. We find that, in addition to eddy covariance monitoring sites, existing chamber sites are important complements, especially over high latitudes and the tropics. Furthermore, expanding the current monitoring network for wetland CH_4 emissions should prioritize, first, tropical and second, sub-tropical semi-arid wetland regions. Considering those new hypothetical wetland sites from tropical and semi-arid climate zones could significantly improve global estimates of wetland CH_4 emissions and reduce bias by 79% (from 76 to 16 $TgCH_4\ y^{-1}$), compared with using solely existing monitoring networks. Our study thus demonstrates an approach for long-term strategic expansion of flux observations.

1. Introduction

Methane (CH_4) is a highly potent greenhouse gas with a global warming potential for sustained ecosystem emissions 45 times greater than that of carbon dioxide (CO_2) over a 100 year time span (Neubauer and Megonigal 2015). Among natural ecosystems, freshwater inland wetlands function as hotspots of sustained CH_4 emissions due to strong microbial activity operating under seasonal or permanent anaerobic soil conditions (Bridgham *et al* 2013). At a global scale, wetlands contribute approximately 20%–30% of total CH_4 emissions. However, uncertainties remain large within and between modeling approaches. Using bottom-up models, where parameterization is informed by process-level understanding, annual wetland CH_4 emissions to the atmosphere have been estimated at 102–182 $TgCH_4\ y^{-1}$. Meanwhile, using top-down inversion models constrained by observed atmospheric CH_4 concentrations, yearly emissions have been estimated at 159–200 $TgCH_4\ y^{-1}$ (Saunois *et al* 2020). More precise quantification of wetland CH_4 emissions is much needed, yet it presents a formidable challenge to current state-of-the-art modeling approaches, evident in these wide model ensemble ranges.

Surface monitoring networks designed to directly measure wetland CH_4 emissions can reduce modeled global CH_4 emission uncertainties by (1) informing bottom-up model parameterization (Yuan *et al* 2022b, Chinta *et al* 2024), (2) providing ecosystem scale observational benchmarks (Yuan *et al* 2022a, Chang *et al* 2023, Ito *et al* 2023), and (3) providing data constraints for top-down inversions (Saunois *et al* 2020). The FLUXNET-CH_4 Synthesis Activity, launched in 2018, is already facilitating global wetland CH_4 model improvements through the recent compilation and publication of the FLUXNET-CH_4 Community Product v1.0; a global, standardized dataset of eddy covariance (EC) flux measurements from 81 sites worldwide (Knox *et al* 2019, 2021, Delwiche *et al* 2021). The EC method allows for the continuous acquisition of high-frequency time series data of wetland CH_4 emissions at the ecosystem scale and can be paired with concurrent biometeorological and biogeochemical drivers of CH_4 emission, e.g., air and soil temperature, vegetation activity, latent heat flux, and soil moisture (Keenan *et al* 2019). Meanwhile, surface chamber measurements are also an important measurement strategy across wetland ecosystems (Turetsky *et al* 2014, Bao *et al* 2021). Chamber-based observations provide CH_4 emission estimates for specific wetland patches, including microtopographic features and plant functional types, and thus provide an important spatially explicit complement to EC towers. Combining both flux tower and chamber measurements will lead to better spatial coverage and will improve understanding of the spatiotemporal dynamics of wetland CH_4 emissions across ecosystem-to-global scales (Yuan *et al* 2024).

The FLUXNET-CH_4 monitoring network is relatively new and still underdeveloped in contrast to the ecosystem CO_2 exchange monitoring systems (e.g. AmeriFlux, FLUXNET) (Baldocchi 2003), especially when considering spatial coverage (Delwiche *et al* 2021). Consequently, global upscaling efforts using the current network of *in situ* measurements of wetland CH_4 emissions to reconstruct or forecast the global CH_4 budget introduces large uncertainties because the spatiotemporal coverage of tower and chamber flux data is limited. In particular, there are only a handful of flux observation sites in the regions ranging from 60° S to 30° N, which account for >60% of global wetland CH_4 emissions and >70% of emissions uncertainties (McNicol *et al* 2023). The ongoing expansion of the *in situ* tower and chamber CH_4 flux network is therefore vital for a global CH_4 emission monitoring system (Griffis *et al* 2020, Soosaar *et al* 2022). Yet, a comprehensive understanding of how this network will ultimately contribute to process-level understanding and predictability of global wetland CH_4 emissions remains largely unexplored.

Our study aims to outline potential advantages for wetland CH_4 emission monitoring at representative locations. We first evaluate the global representativeness of the existing observational network. Then we assess the relative benefit of adding new sites, particularly in currently under-sampled locations hypothesized to be important for the global CH_4 emission budget. It is important to note that we focus on the spatial representativeness of the wetland CH_4 monitoring network, defined by how well CH_4 flux measurements obtained from a network depict the CH_4 flux conditions across a larger regional or global domain (Hargrove *et al* 2003, Chu *et al* 2021). Specifically, we attempt to address two fundamentally important questions: *(1) How much uncertainty in global wetland CH_4 emission estimates results from insufficient monitoring of diverse wetlands? (2) How should we prioritize future wetland CH_4 monitoring sites to best improve global emission estimates?* To address these questions, we developed a machine-learning approach based on an expanded observational dataset and the latest Global Carbon Project CH_4 (GCP-CH_4) bottom-up model ensembles (Ito *et al* 2023, Zhang *et al* 2023).

2. Methodology and data

Overall, our approach combines multi-model ensemble estimates of global wetland CH_4 emissions (section 2.1), site locations from current EC and chamber measurements (section 2.2), and spatial representative analysis with machine learning approach (section 2.3) to evaluate the representativeness of the existing monitoring network. In section 2.4, we further quantify the potential benefit of hypothetical monitoring sites from different climate zones. Lastly, we examine the uncertainties associated with randomness of new site locations within each climate zone and the selection of GCP model ensembles (section 2.5).

2.1. GCP-CH_4 global wetland CH_4 emission

The global monthly estimates of wetland CH_4 emissions came from the latest GCP-CH_4 products based on a cohort of biogeochemical models. This product extended the previous 2000–2017 estimate (Saunois *et al* 2020) to 2000–2020 (Ito *et al* 2023) and all models applied the same Wetland Area and Dynamics for Methane Modeling (WAD2M) (dynamic) global wetland extent estimates and climate forcing (Zhang *et al* 2021). Wetland CH_4 emissions were all gridded to $0.5° \times 0.5°$ spatial resolution.

The GCP-CH_4 multi-model product provided 16 bottom-up model estimates of global wetland CH_4 emissions. Although these models calculate CH_4 production, oxidation, and transport based on broadly similar wetland hydrological and biogeochemical processes, the global estimates differed by up to a factor of two. Therefore, we calculated the multi-model ensemble mean for all 16 models and used it as our benchmark (figure 1(a)) to minimize potential bias due to individual models.

2.2. Wetland CH_4 monitoring networks

The existing network for monitoring wetland CH_4 emissions has two types of measurements: high-frequency measurements using the EC technique (Knox *et al* 2019) and chamber measurements (Turetsky *et al* 2014). In this study, we considered 84 EC sites (Deshmukh *et al* 2020, Delwiche *et al* 2021, Roman *et al* 2021) and 96 chamber sites (table S1), for a total of 180 globally-distributed sites. The existing monitoring network (figure 1(b)) provides valuable insights into the diverse locations that possess the capacity for monitoring, although chamber measurements are not continuous in a high-frequency manner (e.g. 30 min data). We used the location information of the monitoring network (section 2.3) to sample corresponding gridcells (2000–2020 monthly time series) from the GCP-CH_4 models (figure 1 blue arrow).

This work represents an idealized experiment aimed at forward-looking assessment to determine the optimal locations for EC tower sites. We acknowledge that the existing EC and chamber datasets lack continuous temporal coverage from 2000 to 2020. We focus on the spatial representativeness of the monitoring network rather than the temporal variability (seasonal, inter-annual) of CH_4 emissions. We assessed the accuracy of a global wetland CH_4 emission reconstruction, utilizing the machine learning model described in section 2.3. This evaluation was conducted with samples from corresponding gridcells at CH_4 monitoring sites (figure 1(b)).

2.3. Experiment design for representativeness analysis of existing sites

Here, we define the 'representativeness' of the wetland CH_4 monitoring sites by how well CH_4 emission measurements at those sites could represent the spatial variability of global wetland CH_4 emissions (Hargrove *et al* 2003, Chu *et al* 2021). Based on this definition, we designed a *sample-and-reconstruct* approach to evaluate the representativeness of the sampled sites.

First, we assembled a dataset from the GCP-CH_4 model ensemble mean at existing monitoring sites, including air temperature, rainfall, pressure, wind speed, solar radiation, relative humidity, inundation fraction, longitude, latitude, and CH_4 emissions. Then, we trained (using 80% randomly sampled data) and tested (with the remaining 20%) an XGBoost (eXtreme Gradient Boosting) model (Chen *et al* 2015) to capture the relationship between environmental predictors and wetland CH_4 emissions. The XGBoost model is a widely used ensemble learning

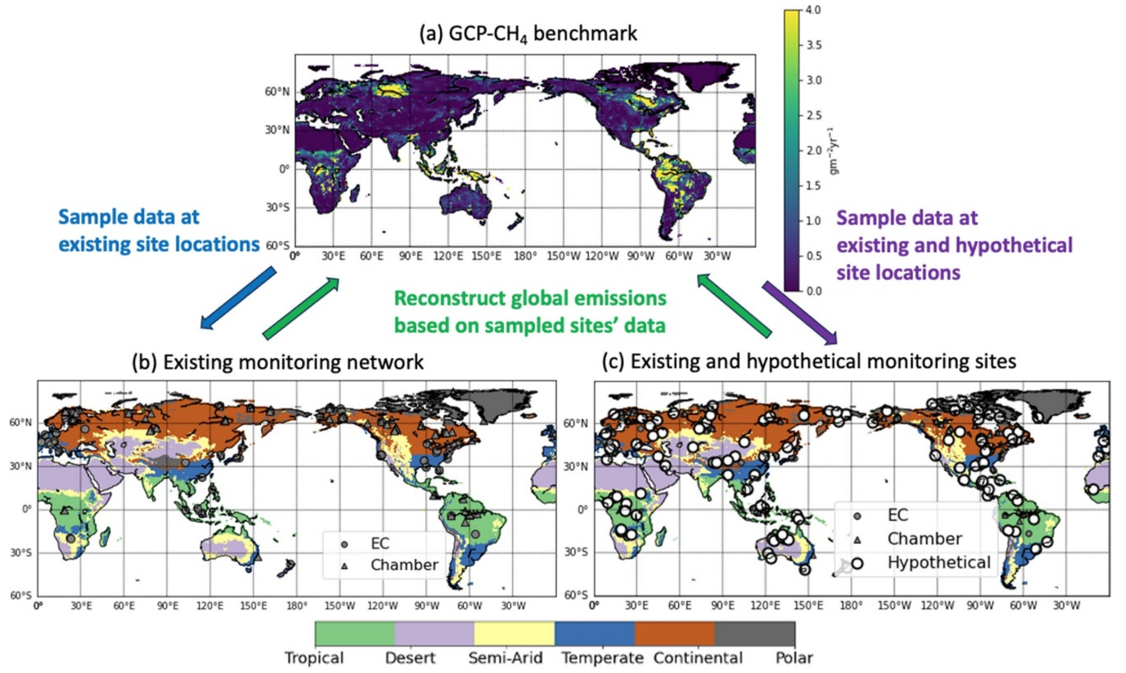

Figure 1. Overview of our sample-and-reconstruct approach to evaluate the representativeness of the current and expanded monitoring networks. Data were sampled from the GCP-CH_4 benchmark (a) based on the location of the current monitoring network in (b) or the combination of the existing and 20 new hypothetical sites (c). Then the data were used to train the machine learning model and compared to the original GCP-CH_4 model benchmark.

algorithm. This algorithm employs a strategic optimization process to iteratively refine decision trees, thereby achieving good predictive precision and generalization across a diverse range of complex datasets (Osman *et al* 2021, Guo *et al* 2023).

Then, we extrapolated the trained model to all wetland gridcells in the GCP-CH_4 benchmark and calculated the global pattern of wetland CH_4 emissions. The reconstructed global CH_4 emission was compared with the GCP-CH_4 model benchmark (figure 1 green arrow). With this approach, site representativeness could be confirmed if the sampled sites were sufficient to reconstruct the GCP-CH_4 model benchmark (Shirley *et al* 2023).

2.4. Experiment design for strategic network expansion

Besides the existing monitoring network sites (section 2.2), we are also interested in seeking under-represented locations for future network expansion. We evaluated the benefit of new monitoring sites by quantifying how adding CH_4 emission measurements would improve the reconstruction global CH_4 emission estimate (figure 1(c)). Twenty randomly selected gridcells with a mean annual inundation fraction larger than 1% at each of 5 Köppen climate zones (tropical, semi-arid, temperate, continental, and polar zones (Kalvová *et al* 2003)) were selected, and additional samples were drawn from GCP-CH_4 models at those hypothetical new sites (figure 1 purple arrow).

We conducted three sets of experiments with this dataset: (1) only at FLUXNET-CH_4 EC network sites (hereafter referred to as 'network1'); (2) at FLUXNET- CH_4 EC and chamber sites (hereafter referred to as 'network2'); and (3) at FLUXNET-CH_4 EC sites, chamber sites, and 20 hypothetical new sites. The value of adding a set of new sites could be quantified by the relative reduction in CH_4 emission bias, compared to the GCP-CH_4 model ensemble benchmark.

2.5. Uncertainty associated with hypothetical sites locations and model selection

Each climate zone covers many wetland gridcells, which may significantly differ in CH_4 emission magnitude and seasonality. To quantify the sampling uncertainty for the 20 hypothetical new sites within each climate zone, we randomly repeated the site selection 10 times and reported the associated uncertainty using the standard deviation of the reconstruction errors.

The GCP-CH_4 includes 16 bottom-up models, each of which exhibits different spatial and temporal patterns of wetland CH_4 emissions. In the main analysis (section 2.1), we used the model ensemble mean to avoid the uncertainty associated with any individual model. However, each GCP models' performance vary dramatically when compared with FLUXNET-CH_4 site observations (Chang *et al* 2023). To determine whether our main conclusions depend on the choice of the multi-model ensemble mean, we also conducted the same analysis using the top three models selected by the FLUXNET-CH_4 dataset and the International Land Model Benchmarking package (Chang *et al* 2023).

3. Results and discussion

3.1. Is the current monitoring network sufficient for global wetland CH_4 emission estimates?

Our model showed high performance when trained and tested at current EC sites (figure S1(a)) and chamber sites (figure S1(b)), with R^2 values for training and testing both >0.95, and mean absolute percentage errors of only 1%–6%. These results suggest that the trained XGBoost ML models robustly captured the relationships between environmental drivers (e.g. temperature, inundation) and wetland CH_4 emissions, and were thus reliable tools for interpolation across regions that experience similar environmental conditions.

Extrapolating the XGBoost ML model to global wetlands, reconstructed CH_4 emissions with network1's sites generally overestimated wetland CH_4 emissions across all climate zones (figure 2 blue line), with the highest biases at tropical and high-latitude wetland ecosystems. The reconstruction with network1's sites estimated annual wetland CH_4 emissions to be 282 $TgCH_4$ $year^{-1}$, which is 70% higher than the GCP-CH_4 model benchmark of 165 $TgCH_4$ $year^{-1}$. Although the reconstruction using network2′ sites still overestimated global wetland CH_4 emissions (242 $TgCH_4$ $year^{-1}$), significant improvements were achieved over tropical and high-latitude regions (figure 2 orange line). For example, the regional reconstruction error was reduced from 16.9 $TgCH_4$ $year^{-1}$ to 4.2 $TgCH_4$ $year^{-1}$ over regions north of 60 N by considering chamber observational sites.

Although both models (trained with network1 vs. network2 sites) achieved similar high accuracy during training, the global extrapolations were substantially different between the two trained models (figure 2). This result indicates that, on the one hand, the model can accurately capture relationships between environmental predictors and CH_4 emissions, but on the other hand, extrapolation was sensitive to the number and locations of monitoring sites. After considering current chamber sites as EC tower sites, however, relatively high errors still exist over tropical and tropical regions. In summary, even if long-term continuous data were available across the existing EC and chamber monitoring network, large errors would persist in global (e.g. 70% over-estimation of global emissions) and regional (e.g. 51% over-estimation in the tropics) estimates. Therefore, more sites are needed to reduce regional and global biases.

3.2. Where to prioritize new monitoring sites of wetland CH_4 emissions?

The existing EC and chamber monitoring network has most of its sites in continental and polar Köppen climate zones (figure 1(b)). This incomplete spatial coverage is consistent with our results that a large portion of reconstruction errors occurred over tropical regions (30° S–30° N) (figure 2). Specifically, the reconstruction with EC and chamber sites (network2) estimated 184 $TgCH_4$ $year^{-1}$ emissions from tropical wetland, which 51% higher than the GCP-CH_4 model benchmark of 122 $TgCH_4$ $year^{-1}$ over this region). To quantify the benefit of adding new tropical sites, we randomly selected 10 sets of 20 monitoring sites over the tropical zone and trained and tested 10 new XGBoost models. The new models consistently had high accuracy over training and testing datasets with R^2 both >0.95 (figure S2). The newly reconstructed global wetland CH_4 emissions were substantially improved over the tropics (figure 3 green line), and overestimated emissions only by 18% (figure 4 green bar). The reconstruction was not sensitive to the random selection of tropical sites (figure 3 green shaded area), partly because the hypothetical site locations were limited to GCP-CH_4 gridcells with >1% surface inundation. The reconstruction errors were most prominent between 30° S and 10° S. In summary, sampling new sites from tropical wetlands could reduce the reconstruction error from 76 to 30 $TgCH_4$ $year^{-1}$ (absolute error).

Similarly, global reconstructions with 10 sets of 20 new sites from semi-arid, temperate, continental, and polar Köppen climate zones indicate that new sites over semi-arid regions reduced the reconstruction error from 10 to 5 $TgCH_4$ $year^{-1}$ over southern hemisphere sub-tropical semi-arid regions (figure 3 magenta line), while new sites in temperate, continental, and polar Köppen climate zones provided much smaller benefits. Overall, by comparing global reconstruction accuracy, we conclude that expanding the monitoring network for wetland CH_4 emissions should prioritize, first, tropical and second, subtropical semi-arid regions

Our sample-and-reconstruct approach might be sensitive to the spatial variation of wetland CH_4 emissions in our GCP-CH_4 model benchmark (i.e. the multi-model ensemble mean of the 16 GCP-CH_4 models). We note that GCP-CH_4 estimates are not real observations, and individually show deviations from observed emissions (Chang *et al* 2023). To investigate the robustness of our conclusions, we conducted an additional analysis that used only the 3 highest scoring GCP-CH_4 models based on an independent global benchmark analysis (Chang *et al* 2023). The top 3 models (TRIPLEX, LPJ-MPI, and DLEM) were used to derive the global mean spatial pattern of CH_4 emissions by sampling the

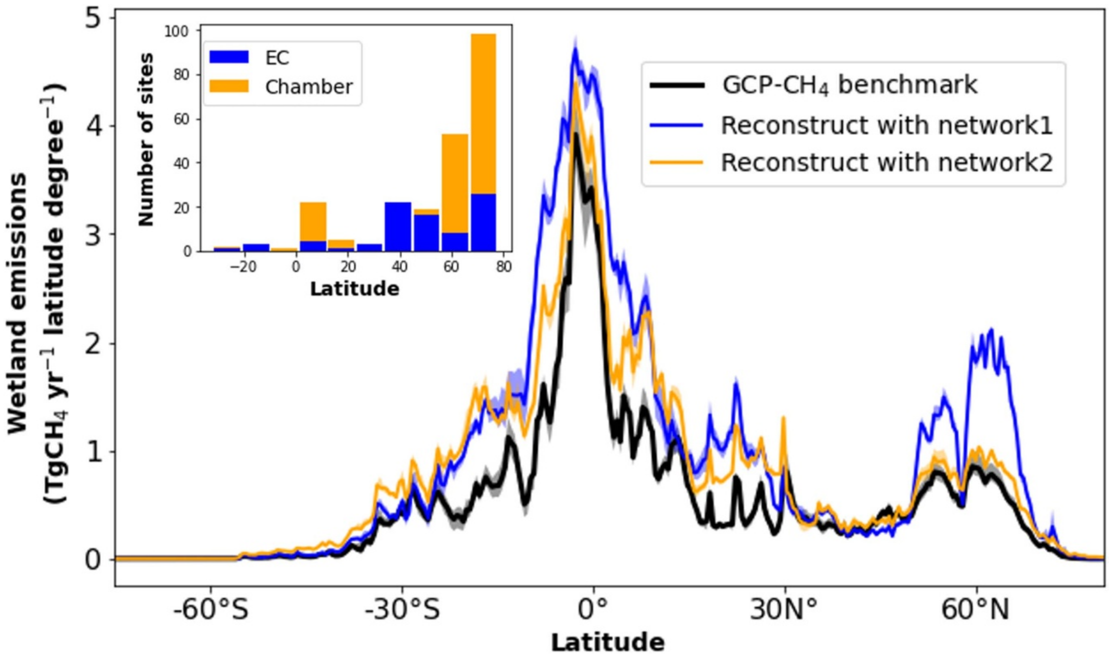

Figure 2. Reconstructed global wetland CH_4 emissions using samples from network1 (eddy covariance sites only, blue line) or from network2 sites (eddy covariance and chamber sites, orange line). The shaded area around the line is the uncertainty range. The inserted panel shows the number of EC and chamber sites along latitudes.

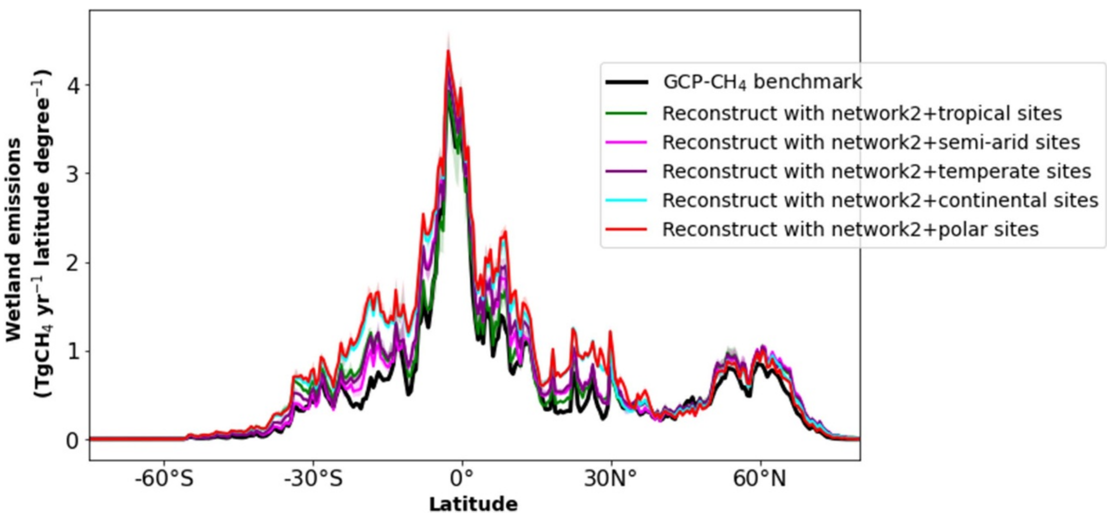

Figure 3. Reconstructed global wetland CH_4 emissions using samples from network2 (eddy covariance and chamber sites locations) and 20 randomly sampled hypothetical sites, separately added for tropical, semi-arid, temperate, continental, and polar Köppen climate zones (see distribution in figure 1(c)).

models at locations of existing and hypothetical monitoring networks (same as the previous analyses). Although the benchmark has changed, our sample-and-reconstruct approach suggested similar results: (1) observations from the chamber site locations are an important supplement to the EC monitoring network; (2) tropical wetlands are not well represented in the existing monitoring network, thus future development of tropical sites is critical to reducing biases in global wetland CH_4 emissions budgets; and (3) more observational sites are needed across sub-tropical semi-arid Köppen climate zones (30° S–10° S) where wetlands are sparsely distributed (figures S3 and S4).

3.3. Sensitivity to the number new monitoring sites

Given our conclusion that prioritizing monitoring sites over tropical regions can substantially reduce CH_4 emission uncertainties, we next evaluated how much the global reconstruction bias could be reduced by increasing the number of hypothetical sites. We therefore conducted additional model experiments that randomly sampled 50, 100, 150, and 200

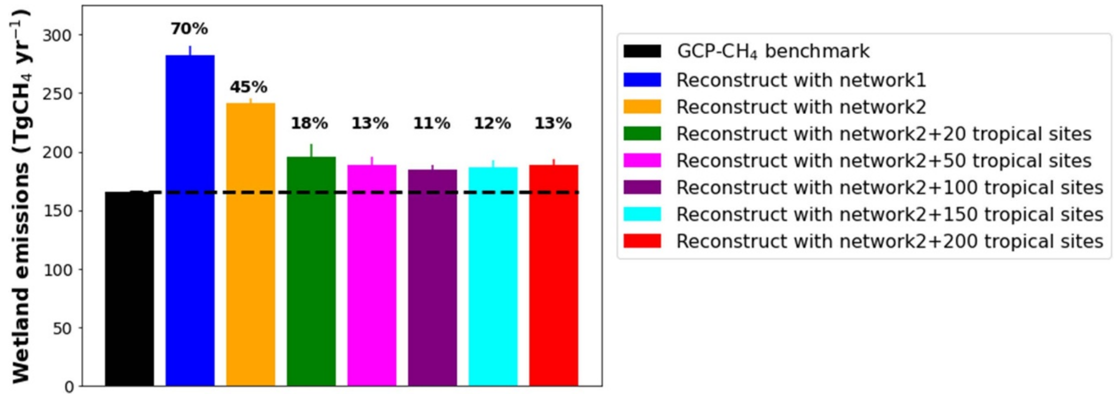

Figure 4. Global wetland CH_4 emissions from the GCP-CH_4 benchmark and the reconstructed emissions using several sets of hypothetical sites over tropical regions. Network1 represents the eddy covariance site locations and network2 includes both eddy covariance and chamber site locations. Numbers reported on top of each bar represent the percentage bias of the reconstruction, compared with the GCP-CH_4 benchmark. The vertical line at the top of each error bar represents the uncertainty of the ML model reconstruction.

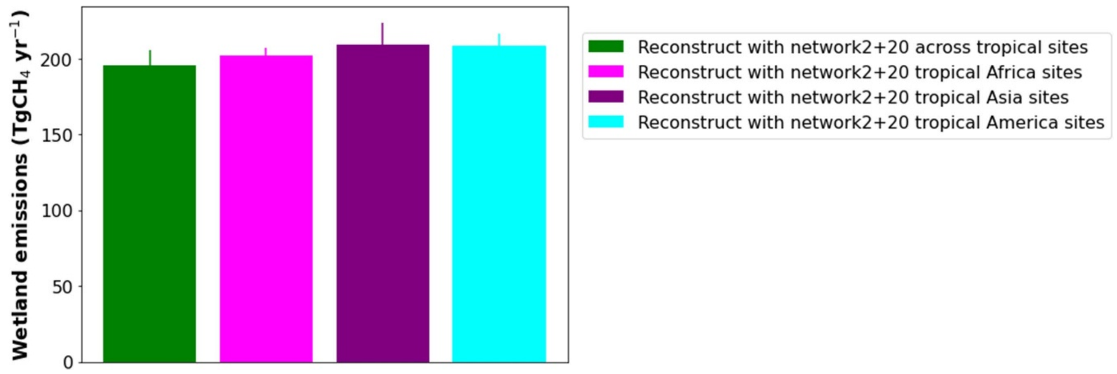

Figure 5. Sensitivity of the reconstructed global wetland CH_4 emissions to the locations of hypothetical tropical monitoring sites.

sites across the tropics. Each random sampling was repeated 10 times to evaluate the impacts of random selection. As the number of monitoring sites increased from 20 to 200, the global reconstruction error declined from 30 to 23 $TgCH_4$ year^{-1} (figure 4).

Furthermore, sampling sites from tropical Africa, America, and Asia individually did not significantly change the reconstruction. The results suggested that sampling sites across the entire tropical region is to be preferred (figure 5 green bar). Sampling sites from tropical Africa are better than sampling sites from tropical America or tropical Asia, because of the relatively lower reconstruction error (figure 5 magenta bar versus purple and cyan bars).

3.4. Towards a representative CH_4 monitoring network

We acknowledge that the establishment of new monitoring sites is a complex problem, *e.g.* should a network preferentially monitor high emission (hotspots) sites (Gloor *et al* 2021, France *et al* 2022, Rößger *et al* 2022, Shaw *et al* 2022, Soosaar *et al* 2022) or better cover wetlands that span the emissions range. For large-scale analysis, spatial representativeness is an important metric (Yang *et al* 2008). For analyses of EC CO_2 emissions, it has been reported that the historical placement of EC towers in highly productive landscapes may bias the global upscaling effort (Ran *et al* 2016). Previous analysis using the FLUXNET-CH_4 dataset has highlighted the lack of spatial coverage over tropical wetlands (Delwiche *et al* 2021). Such gaps in representativeness have been identified in both geographic space (based on longitude and latitude) and bioclimatic space (based on temperature, latent heat, and vegetation index). In addition to the existing tropical sites in the FLUXNET-CH_4 monitoring network, we considered chamber site locations, thus greatly enhancing the spatial coverage over tropical wetlands. However, even with these additional sites, our analysis still showed significant biases (figure 2), indicating the need for more EC sites at tropical and semi-arid wetland sites. Our best reconstruction used existing sites plus 20 hypothetical tropical and semi-arid wetland sites. The resultant global reconstruction of the GCP-CH_4 benchmark, has an absolute bias of only 16 $TgCH_4$ y^{-1}, while the global reconstruction using solely existing sites has an absolute bias of 76 TgCH_4 y^{-1}.

Although a previous study (Delwiche *et al* 2021) using dissimilarity metrics at FLUXNET-CH_4 sites suggested relatively high representativeness over arctic wetlands, our analysis quantitatively showed the high value of monitoring the wetland-dense regions (i.e. Siberian lowland and Hudson Bay lowland), that are currently represented in the chamber network, but not the FLUXNET-CH_4 network.

Our sample-and-reconstruct approach with hypothetical new sample sites was able to address the question of where to prioritize new sites, from the perspective of global CH_4 emissions estimates, modeling, and analysis. This work represents an idealized experiment aimed at determining the optimal locations for placing EC tower sites. We found that future investments in monitoring wetlands over tropical and sub-tropical semi-arid Köppen climate zones are recommended (figure 3). We also recommend enhanced monitoring of the fine-scale wetland hotspots across temperate, continental, and polar zones. We acknowledge that our analysis is subject to the spatial resolution and accuracy of the GCP-CH_4 model benchmark ($0.5° \times 0.5°$, $\sim$50 Km), and thus insensitive to fine-scale heterogeneity of real-world wetlands. Previous efforts with a similar sample-and-reconstruct approach for CO_2 emissions have shown the importance of finer-scale heterogeneity in the representativeness analysis and choice of new monitoring sites (Shirley *et al* 2023). Further, our analysis did not consider seasonal and inter-annual variability in CH_4 emissions and its drivers, but the methods we describe here are well-suited to explore those issues in future work.

This study used $0.5° \times 0.5°$ resolution gridded products for wetland CH_4 emissions, which means that fine-scale processes and interactions influencing CH_4 emissions were not considered. We acknowledge that sub-grid heterogeneity is an important factor in representativeness analysis, though accounting for such heterogeneity is challenging at a global scale. This challenge arises because: (1) high-resolution gridded products for wetland CH_4 emissions do not exist; and (2) within the current monitoring network, only a limited number of $0.5° \times 0.5°$ gridcells are covered by multiple monitoring sites, making it difficult to robustly estimate sub-grid heterogeneity.

We conservatively selected two gridcells with the largest number of monitoring sites for the EC and chamber networks, respectively (figure S5). The gridcell at Sacramento–San Joaquin Delta, USA, is covered by 9 EC sites, while the gridcell at Stordalen, Sweden, is covered by 17 chamber sites. The derived sub-grid heterogeneity in these gridcells is relatively similar, with coefficients of variation (CV) of 72% and 68%, respectively (CV = standard deviation of annual CH_4 emission divided by mean annual CH_4 emission). However, all other $0.5° \times 0.5°$ gridcells contain fewer than 5 EC sites or 10 chamber sites, making it difficult to generalize this sub-grid heterogeneity to global wetlands.

Although applying sub-grid heterogeneity to global estimates is challenging, we argue that our representativeness analysis framework (figure 1) could be seamlessly adapted to a finer resolution when high-resolution gridded products become available in the future. Ideally, using a product at the hundred-meter scale for wetland CH_4 emissions, combined with EC site data, would significantly improve the analysis. Therefore, a high-resolution global product of wetland CH_4 emissions is urgently needed to improve representativeness analysis and guide the strategic deployment of new wetland monitoring sites.

4. Conclusions

In this study, we assessed the representativeness of the global wetland CH_4 emissions monitoring network using a machine learning-based sample-and-reconstruct approach. We utilized a combination of the GCP-CH_4 model simulations, which provide estimates of wetland CH_4 emissions from 2000 to 2020, and existing wetland CH_4 monitoring locations including both EC and chamber sites. Our results revealed that, while our machine learning model effectively captured predictive relationships at existing sites, extrapolation to global wetlands often led to a large overestimation of CH_4 emissions, particularly in tropical and arctic regions. We conclude that the existing monitoring network lacks the necessary spatial representativeness to reliably construct global wetland CH_4 emission estimates. Considering new monitoring sites, particularly over tropical and semi-arid climate zones, could substantially improve the network's representativeness and reduce global wetland CH_4 emission by 79% (from 76 to 16 TgCH_4 y^{-1})

Data availability statement

All data that support the findings of this study are included within the article and the supplementary file: GCP-CH_4 bottom-up estimates of wetland CH_4 emission are archived in the International Carbon Observation System: https://doi.org/10.18160/gcp-ch4-2019.

FLUXNET-CH4 data are available at https://fluxnet.org/data/fluxnet-ch4-community-product/, https://doi.org/10.5281/zenodo.4672601. Chamber site locations are documented in supplementary file.

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This research is mainly supported by the NASA Carbon Monitoring System grant (#NNH20ZDA001N) and the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Scientific Focus Area Project sponsored by the Earth and Environmental Systems Modeling (EESM) Program under the Office of Biological and Environmental Research of the US Department of Energy Office of Science. K Y is also supported by the Early Career Development Grants (ECDG), which was sponsored by the Earth and Environmental Sciences Area of Lawrence Berkeley National Laboratory. A I is supported by the Environmental Research and Technological Development Fund (JPMEERF24S12200) by the Ministry of the Environment, Japan. Z Qin is supported by the National Key Research and Development Program of China (2023YFF0805403). A H, F L, and R J are supported by the Moore Foundation grant Advancing the Understanding of Methane Emissions from Tropical Wetlands. N G was funded by the Met Office Climate Science for Service Partnership (CSSP) Brazil project which is supported by the Department for Science, Innovation & Technology (DSIT), UK.

ORCID iDs

Qing Zhu https://orcid.org/0000-0003-2441-944X
Kunxiaojia Yuan https://orcid.org/0000-0002-1336-5768
William J Riley https://orcid.org/0000-0002-4615-2304
Robert Jackson https://orcid.org/0000-0001-8846-7147
Nicola Gedney https://orcid.org/0000-0002-2165-5239
Peter O Hopcroft https://orcid.org/0000-0003-3694-9181
Akihito Ito https://orcid.org/0000-0001-5265-0791
Atul K Jain https://orcid.org/0000-0002-4051-3228
Thomas Kleinen https://orcid.org/0000-0001-9550-5164
Benjamin Poulter https://orcid.org/0000-0002-9493-8600
Zhangcai Qin https://orcid.org/0000-0001-9414-4854
Xiaoming Xu https://orcid.org/0000-0003-1405-8089
Yi Xi https://orcid.org/0000-0003-2654-4615
Zhen Zhang https://orcid.org/0000-0003-0899-1139

References

Baldocchi D D 2003 Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future *Glob. Change Biol.* **9** 479–92

Bao T, Jia G and Xu X 2021 Wetland heterogeneity determines methane emissions: a pan-arctic synthesis *Environ. Sci. Technol.* **55** 10152–63

Bridgham S D, Cadillo-Quiroz H, Keller J K and Zhuang Q 2013 Methane emissions from wetlands: biogeochemical, microbial, and modeling perspectives from local to global scales *Glob. Change Biol.* **19** 1325–46

Chang K Y *et al* 2023 Observational constraints reduce model spread but not uncertainty in global wetland methane emission estimates *Glob. Change Biol.* **29** 4298–312

Chen T *et al* 2015 Xgboost: extreme gradient boosting *R package version 0.4–2* vol 1 pp 1–4

Chinta S, Gao X and Zhu. Q 2024 Machine learning driven sensitivity analysis of E3SM land model parameters for wetland methane emissions *J. Adv. Model. Earth Syst.* **16** e2023MS004115

Chu H *et al* 2021 Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites *Agric. For. Meteorol.* **301** 108350

Delwiche K B *et al* 2021 FLUXNET-CH4: a global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands *Earth Syst. Sci. Data* **13** 3607–89

Deshmukh C S *et al* 2020 Impact of forest plantation on methane emissions from tropical peatland *Glob. Change Biol.* **26** 2477–95

France J L *et al* 2022 Very large fluxes of methane measured above Bolivian seasonal wetlands *Proc. Natl Acad. Sci.* **119** e2206345119

Gloor M *et al* 2021 Large methane emissions from the pantanal during rising water-levels revealed by regularly measured lower troposphere CH_4 profiles *Glob. Biogeochem. Cycles* **35** e2021GB006964

Griffis T J *et al* 2020 Hydrometeorological sensitivities of net ecosystem carbon dioxide and methane exchange of an Amazonian palm swamp peatland *Agric. For. Meteorol.* **295** 108167

Guo X, Gui X, Xiong H, Hu X, Li Y, Cui H, Qiu Y and Ma C 2023 Critical role of climate factors for groundwater potential mapping in arid regions: insights from random forest, XGBoost, and LightGBM algorithms *J. Hydrol.* **621** 129599

Hargrove W W, Hoffman F M and Law B E 2003 New analysis reveals representativeness of the AmeriFlux network *EOS Trans. Am. Geophys. Union* **84** 529–35

Ito A *et al* 2023 Cold-season methane fluxes simulated by GCP-CH_4 models *Geophys. Res. Lett.* **50** e2023GL103037

Kalvová J, Halenka T, Bezpalcová K and Nemešová I 2003 Köppen climate types in observed and simulated climates *Stud. Geophys. Geod.* **44** 185–202

Keenan T F, Moore D J P and Desai A 2019 Growth and opportunities in networked synthesis through AmeriFlux *New Phytol.* **222** 1685–7

Knox S H *et al* 2019 FLUXNET-CH_4 synthesis activity: objectives, observations, and future directions *Bull. Am. Meteorol. Soc.* **100** 2607–32

Knox S H *et al* 2021 Identifying dominant environmental predictors of freshwater wetland methane fluxes across diurnal to seasonal time scales *Glob. Change Biol.* **27** 3582–604

McNicol G *et al* 2023 Upscaling wetland methane emissions from the FLUXNET-CH4 Eddy covariance network (UpCH_4 v1. 0): model development, network assessment, and budget comparison *AGU Adv.* **4** e2023AV000956

Neubauer S C and Megonigal J P 2015 Moving beyond global warming potentials to quantify the climatic role of ecosystems *Ecosystems* **18** 1000–13

Osman A I A, Ahmed A N, Chow M F, Huang Y F and El-Shafie A 2021 Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia *Ain Shams Eng. J.* **12** 1545–56

Ran Y, Li X, Sun R, Kljun N, Zhang L, Wang X and Zhu G 2016 Spatial representativeness and uncertainty of eddy covariance carbon flux measurements for upscaling net ecosystem productivity to the grid scale *Agric. For. Meteorol.* **230** 114–27

Roman T *et al* 2021 AmeriFlux FLUXNET-1F PE-QFR Quistococha Forest Reserve, Ver. 3–5 AmeriFlux AMP (https://doi.org/10.17190/AMF/1832157)

Rößger N, Sachs T, Wille C, Boike J and Kutzbach L 2022 Seasonal increase of methane emissions linked to warming in Siberian tundra *Nat. Clim. Change* **12** 1031–6

Saunois M *et al* 2020 The global methane budget 2000–2017 *Earth Syst. Sci. Data* **12** 1561–623

Shaw J T *et al* 2022 Large methane emission fluxes observed from tropical Wetlands in Zambia *Glob. Biogeochem. Cycles* **36** e2021GB007261

Shirley I A, Mekonnen Z A, Grant R F, Dafflon B and Riley W J 2023 Machine learning models inaccurately predict current and future high-latitude C balances *Environ. Res. Lett.* **18** 014026

Soosaar K *et al* 2022 High methane emission from palm stems and nitrous oxide emission from the soil in a peruvian Amazon peat swamp forest *Front. For. Glob. Change* **5** 849186

Turetsky M R *et al* 2014 A synthesis of methane emissions from 71 northern, temperate, and subtropical wetlands *Glob. Change Biol.* **20** 2183–97

Yang F, Zhu A-X, Ichii K, White M A, Hashimoto H and Nemani R R 2008 Assessing the representativeness of the AmeriFlux network using MODIS and GOES data *J. Geophys. Res.* G **113** G4

Yuan K *et al* 2022a Causality guided machine learning model on wetland CH_4 emissions across global wetlands *Agric. For. Meteorol.* **324** 109115

Yuan K, Li F, McNicol G, Chen M, Hoyt A, Knox S, Riley W J, Jackson R and Zhu Q 2024 Boreal–Arctic wetland methane emissions modulated by warming and vegetation activity *Nat. Clim. Change* **14** 282–8

Yuan K, Zhu Q, Riley W J, Li F and Wu H 2022b Understanding and reducing the uncertainties of land surface energy flux partitioning within CMIP6 land models *Agric. For. Meteorol.* **319** 108920

Zhang Z *et al* 2023 Characterizing performance of freshwater wetland methane models across time scales at FLUXNET-CH_4 sites using wavelet analyses *J. Geophys. Res.* G **128** e2022JG007259

Zhang Z, Fluet-Chouinard E, Jensen K, McDonald K, Hugelius G, Gumbricht T, Carroll M, Prigent C, Bartsch A and Poulter B 2021 Development of the global dataset of wetland area and dynamics for methane modeling (WAD2M) *Earth Syst. Sci. Data* **13** 2001–23