Environmental Science & Technology

pubs.acs.org/est

Article

Unraveling Spatially Diverse and Interactive Regulatory Mechanisms of Wetland Methane Fluxes to Improve Emission Estimation

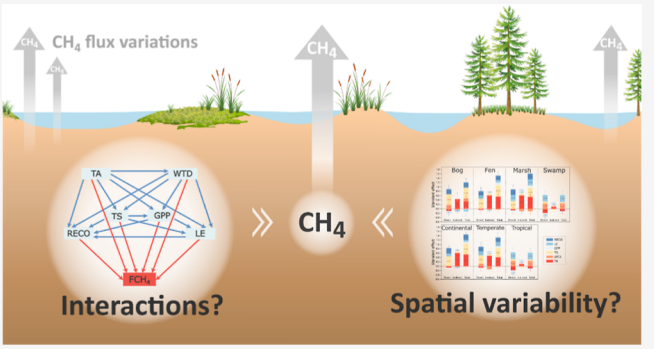
Haonan Guo, Shihao Cui, Claudia Kalla Nielsen, Johannes Wilhelmus Maria Pullens, Chunjing Qiu, and Shubiao Wu*

Cite This: *Environ. Sci. Technol.* 2024, 58, 15052−15065

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Methane fluxes (FCH_4) vary significantly across wetland ecosystems due to complex mechanisms, challenging accurate estimations. The interactions among environmental drivers, while crucial in regulating FCH_4, have not been well understood. Here, the interactive effects of six environmental drivers on FCH_4 were first analyzed using 396,322 half-hourly measurements from 22 sites across various wetland types and climate zones. Results reveal that soil temperature, latent heat turbulent flux, and ecosystem respiration primarily exerted direct effects on FCH_4, while air temperature and gross primary productivity mainly exerted indirect effects by interacting with other drivers. Significant spatial variability in FCH_4 regulatory mechanisms was highlighted, with different drivers demonstrated varying direct, indirect, and total effects among sites. This spatial variability was then linked to site-specific annual-average air temperature (17.7%) and water table (9.0%) conditions, allowing the categorization of CH_4 sources into four groups with identified critical drivers. An improved estimation approach using a random forest model with three critical drivers was consequently proposed, offering accurate FCH_4 predictions with fewer input requirements. By explicitly accounting for environmental interactions and interpreting spatial variability, this study enhances our understanding of the mechanisms regulating CH_4 emissions, contributing to more efficient modeling and estimation of wetland FCH_4.



KEYWORDS: *methane emissions, freshwater wetland, environmental drivers, interactions, spatial variability, structural equation modeling*

1. INTRODUCTION

Methane (CH_4) is the second-largest contributor to global radiative forcing from greenhouse gases after carbon dioxide (CO_2). The 100-year global warming potential of CH_4 is 27−30 times higher than that of CO_2.[1] In 2022, CH_4 accounted for 19% (0.650 W m^{-2}) of the total radiative forcing, according to the NOAA Annual Greenhouse Gas Index (https://gml.noaa.gov/aggi/aggi.html). Wetlands are the largest natural source of CH_4, responsible for about 20−30% of the global total CH_4 emissions.[2,3] Currently, there is a significant uncertainty (23−54%, based on top-down or bottom-up model ensembles) in estimating CH_4 contributions from natural wetlands,[3,4] preventing the closure of CH_4 budget at regional-to-global scales. Improving the estimation of wetland CH_4 fluxes (FCH_4) will help more accurately attribute CH_4 sources and effectively mitigate emissions, ultimately contributing to the goal of limiting global warming. Nevertheless, there exist severe challenges that hinder the accurate estimation of global wetland FCH_4, with one of the most prominent being our unclear understanding of the complex mechanisms regulating CH_4 emissions at ecosystem scale.[5]

Wetland CH_4 emissions are the net result of CH_4 production, consumption, and transportation, all of which are regulated by multiple environmental factors.[2] CH_4 production in wetlands is primarily an anaerobic process performed by methanogens, which convert organic matter into CH_4.[6] The production rate is controlled by drivers including temperature, substrate availability, redox potential, and soil pH, which affect methanogen activity by regulating environmental conditions and food sources.[7−10] Once produced, CH_4 can be transported from the soil to the atmosphere via diffusion, ebullition, or plant aerenchyma transport.[11,12] These pathways are influenced by factors such as temperature, water depth, atmospheric pressure, turbulent conditions, and the type and density of vegetation, which determine the primary transport pathway and its rate.[13−16] During transportation, CH_4 can be

Received: June 17, 2024
Revised: July 18, 2024
Accepted: July 19, 2024
Published: August 12, 2024

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065

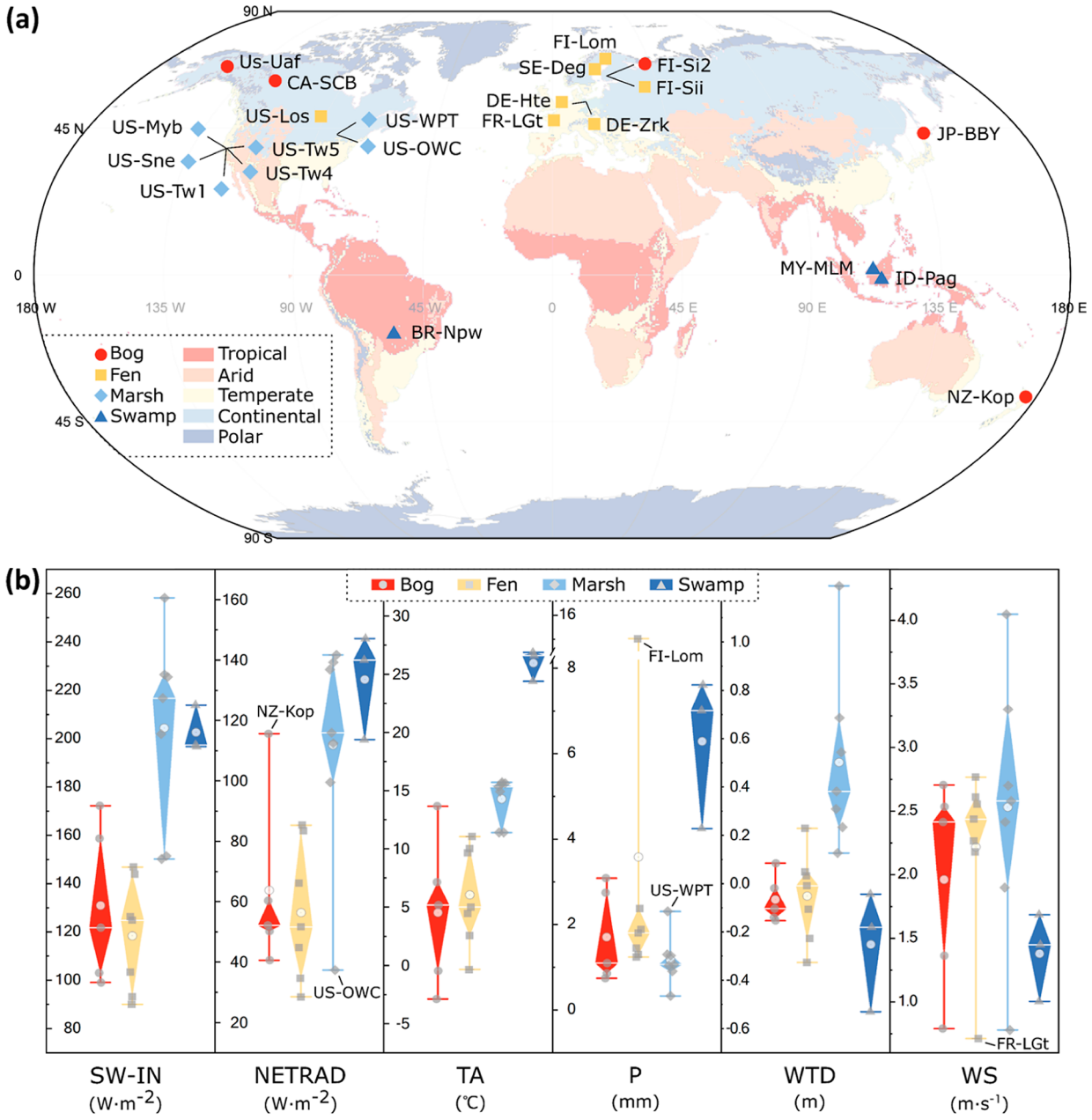


Figure 1. (a) Locations and wetland types of the 22 EC sites across different Köppen climate zones. (b) Annual-average meteorological and water table conditions of different types of wetlands. The values for each site were calculated by averaging daily measurements over one or more complete years. The white central marks, white dots, and the bottom and top end points of diamond boxes indicate the median, average, and the 25th and 75th percentiles, respectively. The whiskers extend up and down to the data points with maximum and minimum values. Outliers exceeding 1.5 times the interquartile range are labeled. SW-IN, incoming shortwave radiation; NETRAD, net radiation; TA, air temperature; P, precipitation; WTD, water table depth; WS, wind speed.

consumed through both aerobic and anaerobic oxidation.[17,18] Critical drivers for these processes include oxygen availability, soil moisture, electron acceptors, microbial communities, temperature, and pH.[2,7,19,20] These factors collectively influence the activity of aerobic and anaerobic methanotrophs, thereby determining the efficiency and extent of CH_4 oxidation before it reaches the atmosphere.

The complexity of CH_4 regulatory mechanisms is evident not only in the multitude of potential environmental drivers,[21,22] but more importantly, in their complicated interactions. For example, air temperature (TA) could have a confounding effect when assessing the influence of soil temperature (TS), soil moisture, and plant productivity, as TA affects all these variables.[23−25] Meanwhile, soil moisture can influence both TS[23] and plant productivity.[26] These interactions suggest the possibility of indirect effects of environmental factors on CH_4 emissions, which may play a comparable or even more significant role than direct effects. Previous studies, however, have not adequately explored these interactions, largely due to the inherent limitation of statistical methods such as correlation analysis, regression analysis, generalized additive modeling, and mutual information. These methods often fall short in explicitly accounting for the interdependencies and interactions among the driv-

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065

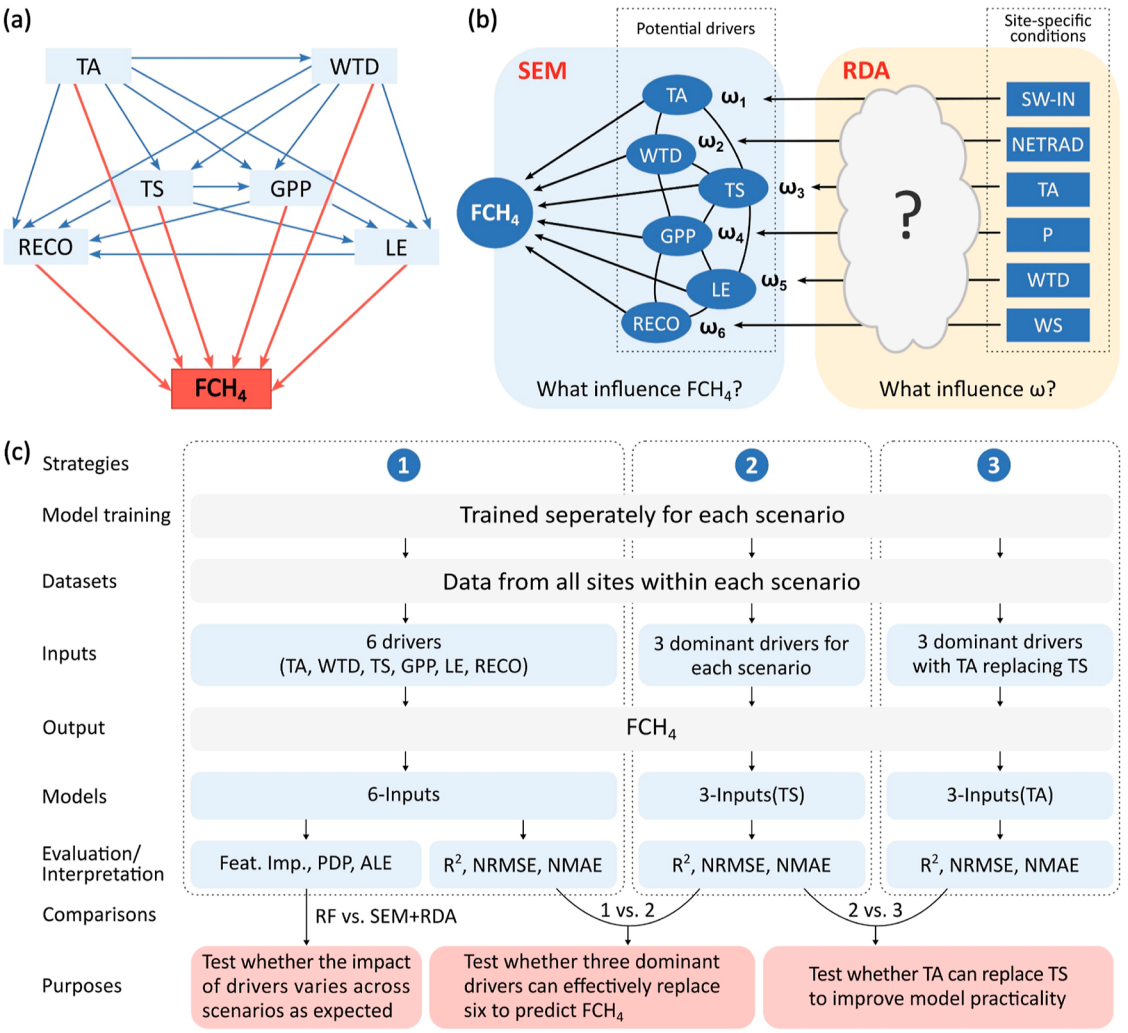


Figure 2. (a) SEM model built based on the interactions between environmental drivers and methane fluxes. The red arrows indicate the direct effect of environmental drivers on methane flux, and the blue arrows indicate the interactions among environmental drivers. (b) Schematic diagram showing the purposes of SEM and RDA analyses. (c) Schematic diagram showing the differences between various RF modeling strategies and their respective purposes. TA, air temperature; WTD, water table depth; TS, soil temperature; GPP, gross primary productivity; LE, latent heat turbulent flux; RECO, ecosystem respiration; FCH_4, methane turbulent flux; SW-IN, incoming shortwave radiation; NETRAD, net radiation; P, precipitation; WS, wind speed.

ers.[22,27−29] Therefore, a study addressing this limitation is critically needed to reveal deeper connections between potential drivers and CH_4 emissions, leading to new insights that enhance the estimation of global wetland FCH_4.

The complexity of CH_4 regulatory mechanisms is also reflected in its notable spatial variability, which has been clearly observed in many previous studies based on global synthesis of eddy covariance (EC) measurements. For instance, Knox et al.[14,27] found that TA and TS were the strongest predictors of FCH_4 globally. However, water table depth (WTD) became more important when focusing on wetland sites with nonconsistent inundation or small temperature variations. Additionally, Chang et al.[30] demonstrated that the relationship between FCH_4 and TA varies significantly among sites even within the same wetland type. Similarly, Yuan et al.[31] and Delwiche et al.[32] revealed that CH_4 emissions were more sensitive to temperature in cooler regions compared to warmer ones. These findings highlight the challenge of developing universally applicable CH_4 models without a clear understanding of the spatial variability of CH_4 regulatory mechanisms. A deeper investigation into this aspect could thus offer valuable insights essential for refining CH_4 models and improving global estimates. Regrettably, current research lacks statistical evidence needed to explain which factors primarily lead to this spatial variability and how much they contributed.

In this study, we incorporated half-hourly FCH_4 measurements from 22 EC sites (Figure 1), covering four freshwater wetland types (i.e., bog, fen, marsh, and swamp) and three Köppen climate zones (i.e., continental, temperate, and tropical), to construct comprehensive data sets. We utilized structural equation modeling (SEM) to quantitatively assess the interactive effects of six primary environmental drivers (i.e., TA, WTD, TS, gross primary productivity (GPP), latent heat turbulent flux (LE), and ecosystem respiration (RECO)) on FCH_4. SEM is a powerful method for elucidating interdependencies among environmental drivers and has been widely adopted in previous studies on gas flux dynamics in natural ecosystems.[33−35] Despite potential limitations in analyzing autocorrelated time-series data, the application of

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065

SEM in our study would be instrumental in uncovering the deeper mechanisms driving FCH_4. Additionally, we used redundancy analysis (RDA) to statistically interpret the spatial variations in the dominant drivers of FCH_4. Building on the RDA findings, we introduced a novel approach integrating principal component analysis (PCA) to categorize CH_4 sources and using three dominant drivers to accurately estimate FCH_4. Finally, we employed random forest (RF) models to test the effectiveness of our categorization approach and the reliability of the identified dominant drivers.

We hypothesize that the dominant drivers of FCH_4 vary across sites with different meteorological and water table conditions, and accounting for such spatial variability when building CH_4 models can enhance our ability to estimate and predict wetland FCH_4. Overall, our study aims to (i) investigate how environmental drivers interactively affect FCH_4, (ii) explain why and how the dominant drivers of FCH_4 vary across sites, and (iii) test whether the dominant drivers identified for each site category can effectively replace six drivers in predicting wetland FCH_4.

2. MATERIALS AND METHODS

2.1. Data Set and Site Description. The data set used in this study is from the FLUXNET-CH_4, a synthesis activity led by the Global Carbon Project, which compiles and standardizes global EC FCH_4 measurements via the regional networks of AmeriFlux, EuroFlux, AsiaFlux, and OzFlux.[14,32] The EC technique is a promising method for providing high-quality data stream, demonstrating advantages of ecosystem-scale capabilities, noninvasive monitoring, quasi-continuous data collection, and high-frequency sampling.[13]

We focused on four major types of natural freshwater wetlands (bog, fen, marsh, and swamp), and selected 22 sites (Figure 1a), each with data for all six environmental drivers studied and at least one full year of FCH_4 measurements. The sites were classified based on the site-specific literature[32] (Table S1). Half-hourly FCH_4 and six environmental drivers were considered in this study: TA, WTD, TS, GPP, LE, and RECO. The negative and positive values of WTD indicate water levels below and above the soil surface, respectively. When more than one TS observations are available, we used the average of all observations measured within the upper 20 cm of the soil (see Delwiche et al.[32] for detailed TS probe depth), which has been typically demonstrated as the active layer for CH_4 production.[36−38] The raw half-hourly data were used since they preserve the original dynamic patterns without any artificial gap-filling and have the finest temporal resolution. The six environmental drivers were selected due to their relatively strong control over FCH_4 as analyzed in previous studies.[27,31] Data rows with missing values were deleted. The annual-average meteorological and water table conditions of each site, calculated as the average of daily values across one or more full years, are presented in Figure 1b and Table S1.

2.2. Structural Equation Modeling. SEM was employed to explicitly quantify the interactive effects of six environmental drivers on FCH_4 at 22 wetland sites (Figure 2b). SEM is a multivariate approach capable of representing hypotheses about complex networks and multiple causality.[39,40] Based on biogeochemical knowledge, an SEM model showing the relational network (pathways) of selected variables was built (Figure 2a). The biogeochemical basis and corresponding literature supporting each pathway are provided in Table S2. The model happened to be saturated, rendering fit indices irrelevant, with only the path coefficients mattering. In the model, direct effect was defined as those influences unmediated by any other variable, indirect effect was defined as those influences mediated by at least one intervening variable, and total effect was defined as the sum of direct and indirect effects.[41] The same model was applied separately to each of the 22 sites using IBM SPSS AMOS software (v24.0.0), and corresponding standardized direct, indirect, and total effects were obtained (Figures 3a, S1 and Table S3). To further attribute the indirect effect of TA to different pathways, PE_{IVi} (proportional distribution of the indirect effect of TA in the IV_i pathway) was calculated using the following equation

$$PE_{IV_i}(\%) = \frac{|E_{di\_TA\_IV_i} \times E_{tot\_IV_i\_FCH_4}|}{\sum_{i=1}^{5} |E_{di\_TA\_IV_i} \times E_{tot\_IV_i\_FCH_4}|} \times 100\% \quad (1)$$

where $E_{di\_TA\_IV_i}$ is the direct effect of TA on intervening variable IV_i, and $E_{tot\_IV_i\_FCH_4}$ is the total effect of IV_i on FCH_4. i ranges from 1 to 5, and IV_1, IV_2, IV_3, IV_4, and IV_5 indicate WTD, TS, GPP, LE, and RECO, respectively. The results are presented in Figure 3c and Table S4.

2.3. One-Way Analysis of Variance. One-way analysis of variance (one-way ANOVA) was employed to test differences among the direct, indirect, and total effects of the same environmental driver, as well as differences among environmental drivers within the same type of effect, and differences among portioned effects in different pathways. Standard $P < 0.05$ criterion and Dunnett's T3 post hoc analysis were employed.

2.4. Redundancy Analysis. RDA was used to elucidate the variations in mechanisms regulating CH_4 emissions across different sites (Figure 2b). RDA, a multivariate and constrained ordination technique, provides valuable insights into the extent to which changes in the independent variables account for the variations in the dependent variables.[42] The standardized effects listed in Table S3, along with the annual-average conditions provided in Table S1, were respectively chosen as the dependent and independent variables for performing RDA using Canoco 5 software. The standardized effects were transformed into absolute values before analysis to capture the strength of the effects rather than their direction. The results are presented as an ordination plot and a summarized conditional effects table (Figure 5 and Table S5). Conditional effects in this context denote the contributions of individual independent variables in explaining the observed variation in the dependent variables, while considering the influence of other independent.[43]

2.5. Principal Component Analysis. PCA was employed for the categorization of all wetland sites based on their annual-average air temperature (aa-TA) and annual-average water table depth (aa-WTD). PCA is a statistical technique that reduces data dimensionality by capturing the primary patterns as uncorrelated variables known as principal components. Typically, the first two principal components are used for two-dimensional data visualization, allowing the visual identification of clusters of data points exhibiting similar characteristics.[44] Detailed reasons for selecting PCA as site categorization method are provided in the Supporting Information. Here, PCA was conducted on the aa-TA and aa-WTD data of 22 sites using Canoco 5 software. According to the areas divided by the arrows representing aa-TA and aa-WTD (Figure 6a), the 22 sites were then categorized into four scenarios: scenario A (sites with high aa-TA and high aa-WTD), scenario

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065

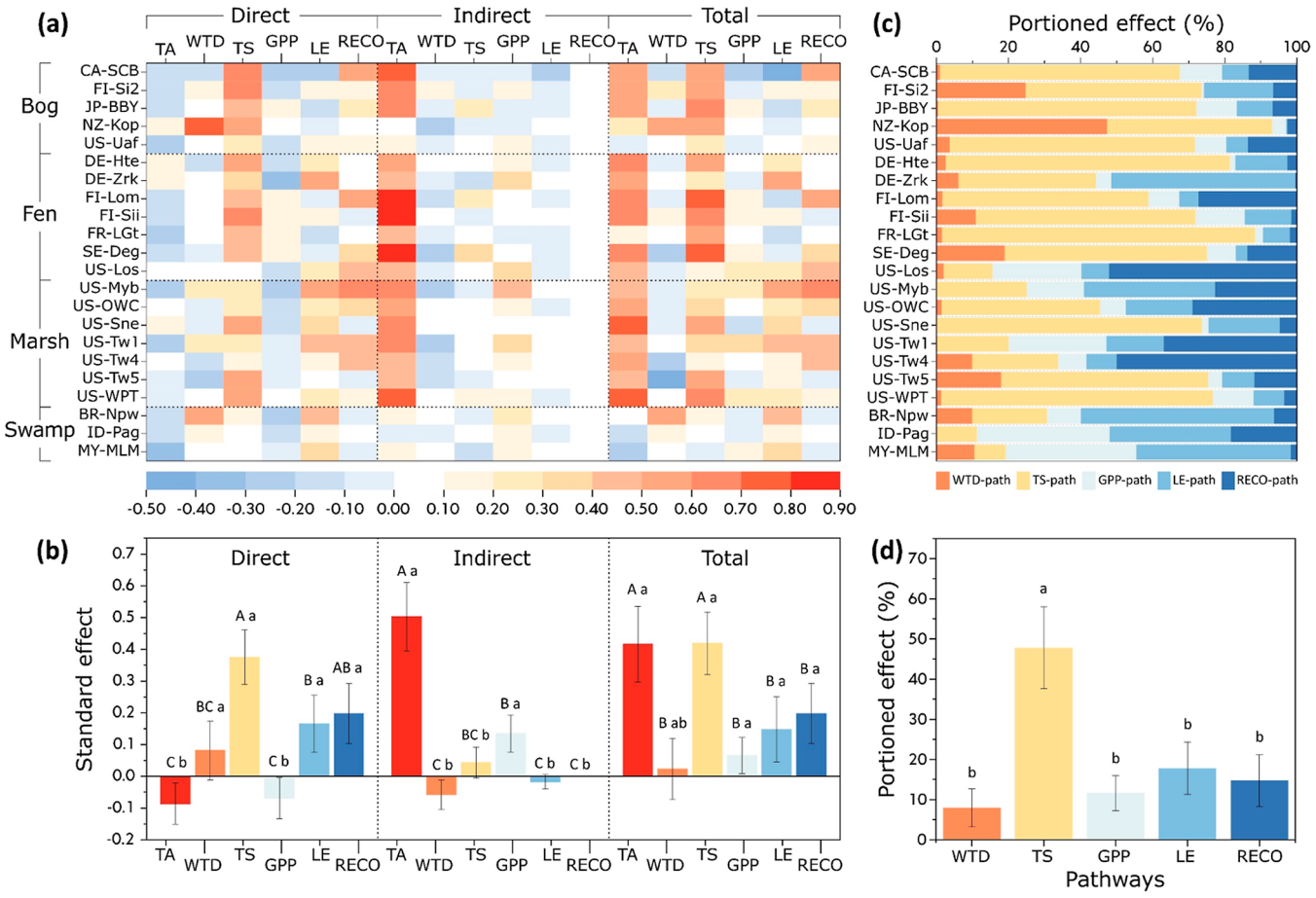


Figure 3. (a) Heatmap showing the standardized direct, indirect, and total effects of environmental drivers on methane fluxes across 22 EC sites. Specific values are provided in Table S3. (b) Statistical analysis of the standardized effects (n = 22). Error bars represent mean ±95% confidence interval of standardized effects. Different capital letters indicate significant differences (P < 0.05) among environmental drivers under the same type of effect. Different lowercase letters indicate significant differences (P < 0.05) among the direct, indirect, and total effects of the same environmental driver. (c) Proportional distribution (%) of the indirect effect of air temperature on methane fluxes in different pathways across 22 EC sites (Materials and Methods). Specific values are provided in Table S4. (d) Statistical analysis of the portioned effects (n = 22). Error bars represent mean ±95% confidence interval of portioned effects. Different lowercase letters indicate significant differences (P < 0.05) among pathways. TA, air temperature; WTD, water table depth; TS, soil temperature; GPP, gross primary productivity; LE, latent heat turbulent flux; RECO, ecosystem respiration.

B (sites with high aa-TA and low aa-WTD), scenario C (sites with low aa-TA and high aa-WTD), and scenario D (sites with low aa-TA and low aa-WTD). Additionally, the range of each region was defined based on the extreme values of aa-TA and aa-WTD within the corresponding scenario (Figure 6b), hypothesizing that sites falling within the same region exhibit similar underlying mechanisms regulating CH_4 emissions.

2.6. Random Forest. RF was used to test the reliability of the dominant drivers identified for each site category and whether they could effectively substitute for the six drivers in predicting FCH_4. RF is an ensemble machine learning model that combines multiple decision trees to make predictions, showing advantages of robustness, versatility, and interpretability.[45] In this study, FCH_4 was modeled separately for each scenario through three distinct input strategies (Figure 2c): employing all six environmental drivers, focusing on three dominant drivers (as detailed in Figure 6b), or utilizing three dominant drivers with TA substituting TS. RF modeling was carried out using Python (v3.7.6). Prior to model training, the original data set was shuffled using the "shuffle" function from the "sklearn.utils" module to introduce greater randomness into the subsequent data set division, thus improving the generalization performance of the model. The data set was then randomly divided into a training set (80%) and a test set (20%). RF regressor was trained with the training set, and the model hyperparameters were fine-tuned using "GridSearchCV" with 5-fold cross-validation. The hyperparameter "n_estimators" was tuned within the range of [1, 200], while "min_samples_leaf" was tuned within [1, 10]. Additionally, "max_depth" was set to 50, "random_state" was set to 0, and default values were maintained for other hyperparameters. Description of each hyperparameter in RF modeling and the fine-tuned hyperparameters are provided respectively in Tables S6 and S7. The "RandomForestRegressor" was imported from the "sklearn.ensemble" package, while the "GridSearchCV" functionality was imported from the "sklearn.model_selection" package. Following training and validation, the generalization ability of the RF model was assessed using the test set by computing the R^2, RMSE, NRMSE, MAE, and NMAE scores (Figure 7a and Table S7). In the case of a data set comprising n samples, the R^2, RMSE, NRMSE, MAE, and NMAE scores were determined as follows

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065

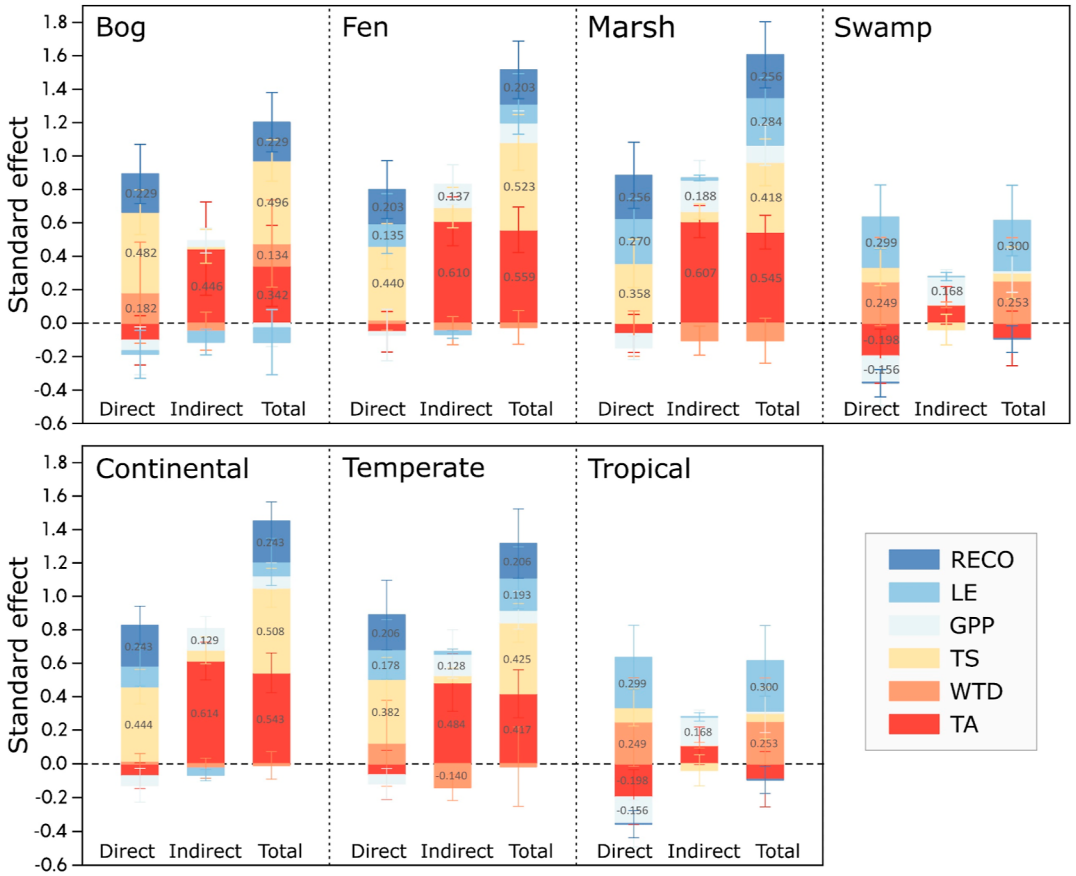


Figure 4. Comparison of the standardized effects of environmental drivers on methane fluxes among different wetland types and Köppen climate zones. The number of sites (n) included in each category is as follows: bog = 5, fen = 7, marsh = 7, swamp = 3; continental = 12, temperate = 7, tropical = 3. Error bars represent mean ±95% confidence interval of standardized effects. Specific values are only shown for the columns with a height ≥10% of the total height of stacked columns. TA, air temperature; WTD, water table depth; TS, soil temperature; GPP, gross primary productivity; LE, latent heat turbulent flux; RECO, ecosystem respiration.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (3)$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMAE}}{\bar{y}} \quad (4)$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \quad (5)$$

$$\mathrm{NMAE} = \frac{\mathrm{MAE}}{\bar{y}} \quad (6)$$

where y_i is the measured value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of measured values.

The importance of different inputs in the 6-input models was analyzed using the "feature_importances_" attribute, and corresponding outcomes are presented in Figure 7b. Additionally, the partial dependence plots (PDP) and accumulated local effects (ALE) were generated for the 6-input models using the "partial_dependence" function from the "sklearn.inspection" package and the "ALE" function from the "alibi.explainers" package. These plots are presented in Figures S3 and S4, respectively.

3. RESULTS

3.1. General Regulatory Mechanisms of Wetland CH_4 Emissions. For each of the 22 wetland sites, we systematically assessed the direct effects (i.e., those unmediated by any other variable), indirect effects (i.e., those mediated by at least one intervening variable), and total effects (i.e., the sum of direct and indirect effects) of environmental drivers on FCH_4. This analysis was conducted using a SEM model (Figure 2a), which was constructed based on established biogeochemical knowledge (Materials and Methods). Detailed results for each site can be found in Figure S1. The SEM model offered the best explanation for marshes, achieving an average R^2 of 0.52. This was followed by fens, bogs, and swamps, with average R^2 of 0.49, 0.36, and 0.18, respectively. The SEM results are visually represented in a heatmap (Figure 3a) and subjected to statistical analysis (Figure 3b).

Overall, our results reveal a generally positive total effect of various environmental drivers on FCH_4, suggesting the potential for increased levels of TA, TS, WTD, GPP, LE, and RECO to enhance FCH_4. However, we also observed negative direct effects for TA and GPP, along with a negative indirect effect for WTD. When comparing the different types of effect of the same environmental driver over all sites (indicated by the lowercase letters in Figure 3b), TS, LE, and RECO were found to have a greater direct effect, whereas TA and GPP exerted a more indirect influence. Additionally, when

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065

examining the effects of different environmental drivers within the same type of effect (indicated by the capital letters in Figure 3b), TS emerged as the dominant factor in terms of direct effect, whereas TA played a dominant role in terms of indirect effect. Remarkably, among all environmental drivers, both TA and TS demonstrated significantly higher total effects, nearing 0.42, while the effect of WTD ranked the bottom.

In order to elucidate the underlying source of the indirect effect of TA on FCH_4, an effect distribution analysis was conducted for each wetland site individually (Materials and Methods). The corresponding results are depicted in Figure 3c and subjected to statistical analysis, as shown in Figure 3d. Notably, a clear trend emerged, indicating that a significant portion of the indirect effect was attributed to the TS pathway (47.84%), surpassing the proportions assigned to the WTD (7.97%), GPP (11.65%), LE (17.78%), and RECO (14.76%) pathways. This suggests that the indirect effect of TA primarily originates from its influence on TS, thereby providing further evidence for the crucial role played by TS in regulating FCH_4 from wetlands. Overall, our findings highlight the significant potential of employing TS instead of WTD as the primary factor in modeling or estimating wetland CH_4 emissions.

3.2. Variations in Regulatory Mechanisms Across Sites. Although the mechanisms regulating CH_4 emissions at the 22 wetland sites exhibited clear general patterns, variations were still observed among the sites, as indicated by the error bars from the statistical analysis in Figure 3b,d. To account for these variations, we categorized the sites based on wetland type and Köppen climate zone, enabling us to compare the differences in the mechanisms regulating CH_4 emissions among the respective groups (Figure 4). Among the investigated wetland types, bogs, fens, and marshes showed a similar pattern, which demonstrated TA and TS as the dominant drivers, with total effects ranging from 0.342 to 0.559. Meanwhile, RECO played a secondary role, with total effects ranging from 0.203 to 0.256. The only difference observed was a slightly greater significance of LE in marshes compared to bogs and fens. In contrast, swamps displayed a distinct pattern, where WTD and LE exerted stronger total effects of 0.253 and 0.300, respectively, while the effects of TA and TS were less prominent. Furthermore, the effect of RECO in swamps was notably weaker (<0.01) compared to the other three types of wetlands.

When comparing different Köppen climate zones, clearer trends were observed. The effects of TA, TS, and RECO displayed a consistent decrease from continental to temperate and then tropical zones, while the effects of WTD and LE followed the opposite trend (Figure 4). Specifically, the standardized total effect (absolute value) of TA and TS declined from >0.50 in the continental zone to <0.10 in the tropical zone. Similarly, the effect of RECO decreased from approximately 0.24 in the continental zone to <0.10 in the tropical zone. In contrast, the effects of WTD and LE increased from <0.10 in the continental zone to >0.25 in the tropical zone. Based on the observed variations, it can be inferred that the mechanism regulating CH_4 emissions is influenced by the natural conditions, such as mean annual temperature and precipitation, of the specific wetland site.

3.3. Site-Specific Temperature and Water Table Dependence of CH_4 Regulatory Mechanisms. RDA was employed to identify the specific natural conditions influencing the regulatory mechanisms of CH_4 emissions. The standardized effects of environmental drivers (refer to Figure 3a and Table S3) were correlated with the annual-average meteorological and water table conditions of specific sites (refer to Figure 1b and Table S1) (Materials and Methods). These conditions include incoming shortwave radiation (SW-IN), net radiation (NETRAD), TA, precipitation (P), WTD, and wind speed (WS). Based on the conditional explaining quantity presented in Figure 5, it became evident that the aa-TA (17.7%, $P < 0.01$) and aa-WTD (9.0%, $P < 0.10$) played crucial roles in influencing the relative significance of environmental drivers in regulating CH_4 emissions, which outweighed the influence of SW-IN (4.3%), NETRAD (1.4%), P (2.0%), and WS (6.6%). When considered together, aa-TA and aa-WTD explained 26.7% of the variations in regulatory mechanisms across the sites. Furthermore, by examining the arrow direction in Figure 5, we observed that the aa-TA was positively correlated with the direct and total effects of WTD and LE, while negatively correlated with the indirect and total effects of TA and GPP, as well as the direct and total effects of TS and RECO. This aligns with the variation trends observed across different climate zones (Figure 4). In contrast, aa-WTD exhibited positive correlations with the indirect and total effects of TA and GPP, as well as the direct and total effects of TS and RECO, while demonstrated negative correlations with the direct and total effects of WTD and LE. This suggests that in a specific wetland ecosystem, when one of the critical conditions, either temperature or water table, is fully satisfied to provide the necessary energy or create an anoxic environment for methanogenic activities, it becomes less important in regulating CH_4 emissions, while the remaining conditions become more important.

Based on these findings, we propose four scenarios to explore potential changes in the importance of various environmental drivers in regulating CH_4 emissions. These scenarios are (1) sites with higher aa-TA and aa-WTD; (2) sites with higher aa-TA but lower aa-WTD; (3) sites with lower aa-TA but higher aa-WTD; (4) sites with lower aa-TA and aa-WTD. In scenarios 2 and 3, it is straightforward to anticipate the changes, as the alterations caused by aa-TA and aa-WTD occur in the same direction. Specifically, in scenario 2, WTD and LE would emerge as more influential drivers, while the effects of TA, TS, RECO, and GPP would gain greater significance in scenario 3. However, scenario 1 presents a more complex situation, as the changes caused by aa-TA and aa-WTD are in different directions. Still, inferences can be drawn from the approximate correlations indicated by the projected length in Figure 5. In scenario 1, WTD and RECO, showing the strongest positive correlation with the aa-TA and aa-WTD, respectively, would become more significant. Similarly, in scenario 4, WTD and RECO would play a greater role in regulating CH_4 emissions. These inferences will be applied and tested in the subsequent section of our study.

3.4. Site Categorization and Random Forest Modeling. The CH_4 sources were categorized using PCA based on their aa-TA and aa-WTD conditions. As depicted in Figure 6a, the 22 wetland sites were divided into four groups, corresponding to the four scenarios described in the previous section, by the arrows representing the aa-TA and aa-WTD. These scenarios disrupt the original classification based on wetland type (i.e., bogs, fens, marshes, and swamps). In order to enhance the readability of the PCA results, the 22 sites were then plotted on a Cartesian coordinate system, considering their aa-TA and aa-WTD conditions. Subsequently, the boundaries of each region were determined based on the

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065

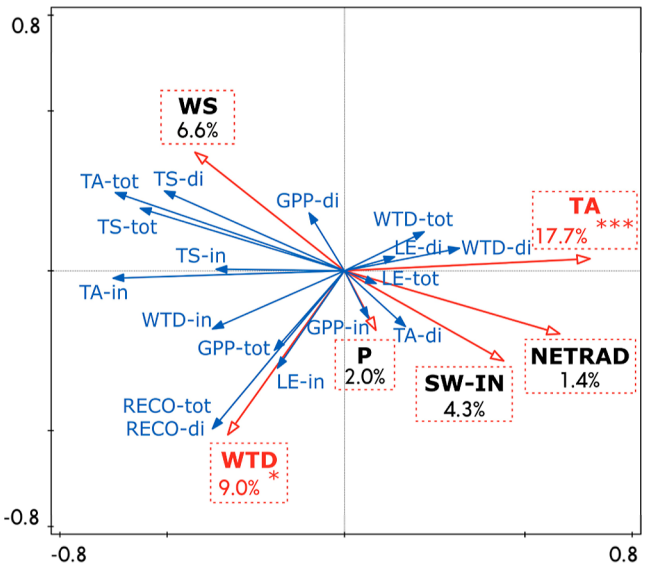


Figure 5. RDA of the effects of environmental drivers on methane fluxes and annual-average meteorological and water table conditions. Each arrow points to the direction of the maximum increase of the corresponding variable. Acute and obtuse angles between arrows indicate positive and negative correlations between variables, respectively. The approximate correlation between two variables can be read by projecting the length of one arrow onto the direction line of the other arrow. The number below explanatory variables (TA, WTD, WS, SW-IN, P, and NETRAD) represents their conditional explaining quantity (%) for the total variation in response variables (written in blue font). The asterisks next to the number indicate significance level (*P < 0.10, ***P < 0.01). GPP, gross primary productivity; LE, latent heat turbulent flux; NETRAD, net radiation; P, precipitation; RECO, ecosystem respiration; SW-IN, incoming shortwave radiation; TA, air temperature; TS, soil temperature; WS, wind speed; WTD, water table depth; -di, direct effect; -in, indirect effect; -tot, total effect.

conditions of the most marginal sites within each scenario. The resulting visualization in Figure 6b exhibits the four regions distinguished by different colors. We hypothesize that sites falling into the same region share similar dominant drivers for FCH_4, as indicated in Figure 6b. These drivers were inferred from Figure 3, which demonstrated the critical role of TS in regulating FCH_4, and Figure 5, which demonstrated the potential changes in the importance of various drivers in different scenarios. It is worth noting that despite TA showing a comparable total effect to TS, it was not considered as a dominant driver. This decision was based on the understanding that TA primarily influenced FCH_4 indirectly by regulating TS. Additionally, there was a strong correlation observed between TA and TS, as illustrated in Figure S2. Therefore, if TS has already been chosen as a representative driver, incorporating TA does not benefit the model performance. However, from an alternative perspective, a model incorporating TA offers broader practicality than one reliant on TS, considering the practical ease of obtaining TA data compared to TS data.

To test the above hypothesis, RF was employed to model the FCH_4 data for each scenario (Materials and Methods). These models were developed using three different input configurations: the full set of six environmental drivers, the three dominant drivers (Figure 6b), and the three dominant drivers with TA substituting TS. The comparison between the measured and predicted FCH_4 values in Figure 7a reveals that, when utilizing all six drivers as inputs, the models developed for the two scenarios with high aa-WTD (i.e., scenario HTA−HWTD and LTA−HWTD) demonstrated a good testing performance ($R^2 \approx 0.85$). Following this, the LTA−LWTD scenario (i.e., low aa-TA and aa-WTD) achieved a respectable R^2 of 0.77, while the HTA−LWTD scenario (i.e., high aa-TA and low aa-WTD) presented the lowest R^2 of 0.68. Remarkably, the feature importance for each scenario (Figure 7b) aligns really well with our theoretically inferred dominant

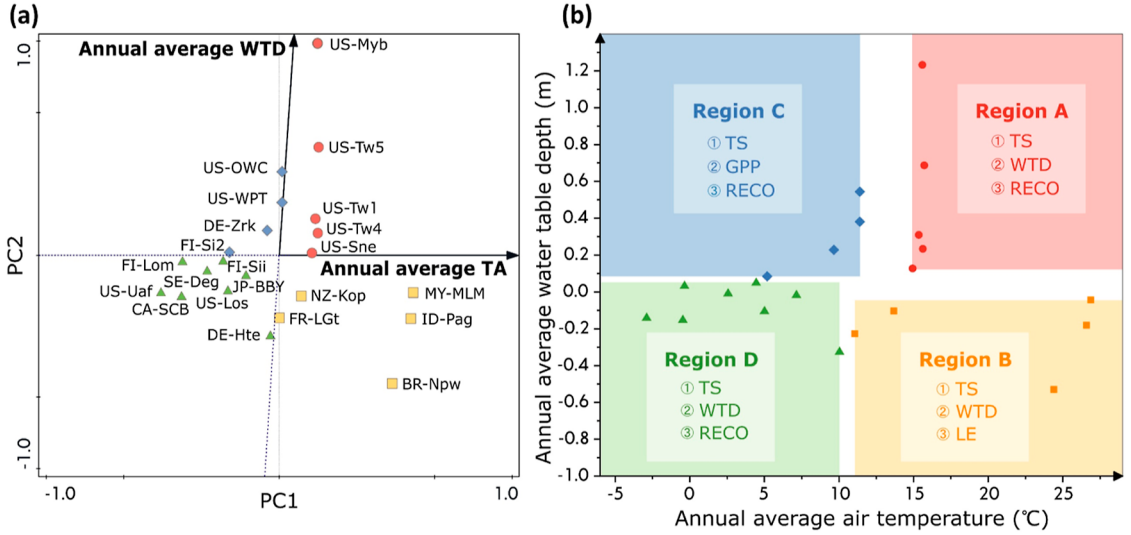


Figure 6. Categorization of all EC sites based on principal components analysis of aa-TA and aa-WTD. (a) PCA results of site-level aa-TA and aa-WTD. The 22 sites were categorized into four scenarios (indicated by points of different colors and shapes) according to the areas divided by the arrows representing aa-TA and aa-WTD. (b) The range of regions A, B, C, and D defined by the categorization results in (a) (see details in Materials and Methods). Region A: high aa-TA and high aa-WTD; region B: high aa-TA and low aa-WTD; region C: low aa-TA and high aa-WTD; region D: low aa-TA and low aa-WTD. The environmental drivers listed in each region are the key drivers of methane fluxes from sites that fall into that region. TA, air temperature; WTD, water table depth; TS, soil temperature; GPP, gross primary productivity; LE, latent heat turbulent flux; RECO, ecosystem respiration.

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065

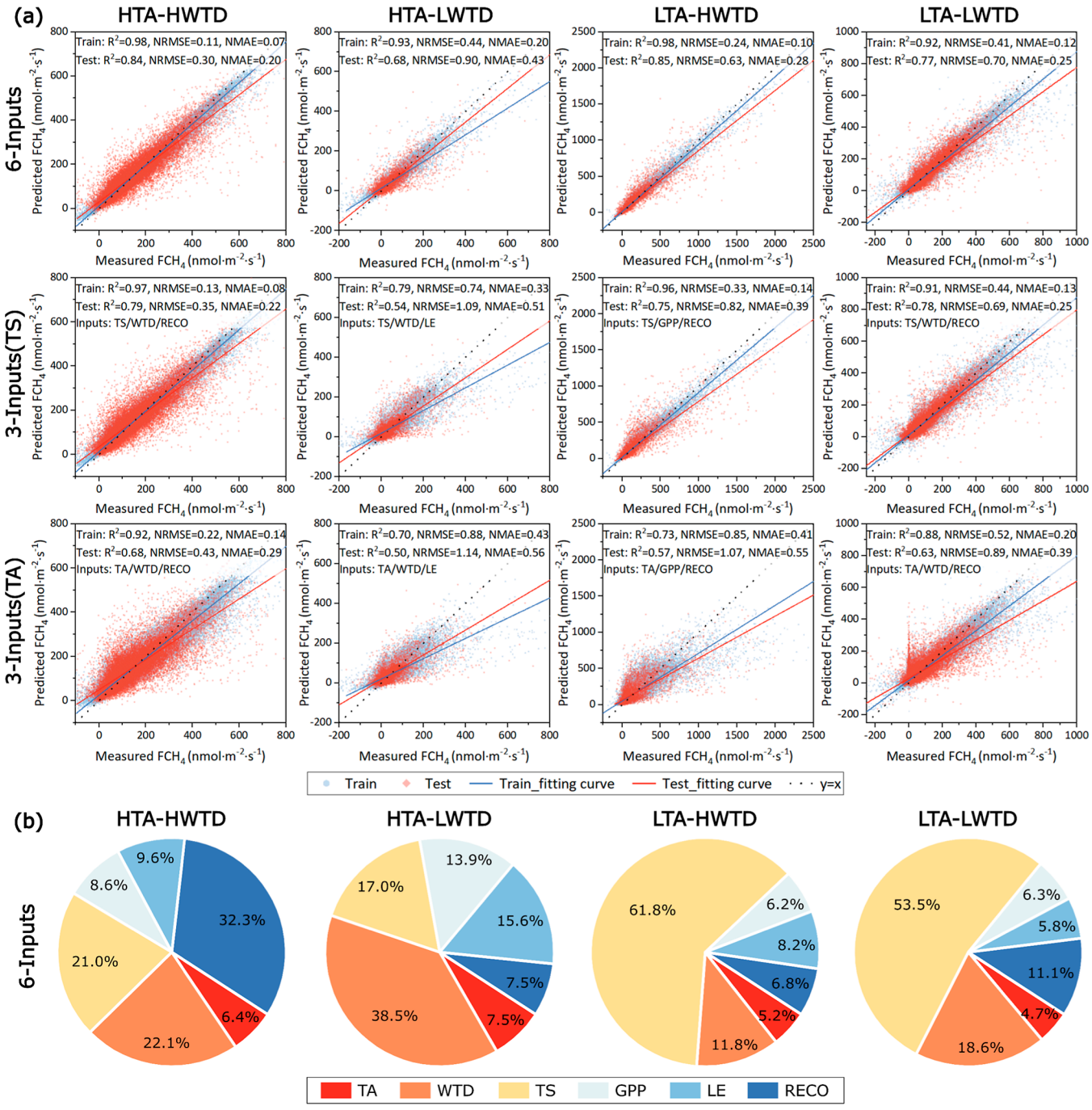


Figure 7. (a) Comparison of RF modeling performance of methane fluxes using different input variables. The data size (n) of each scenario is as follows: HTA−HWTD (high annual-average air temperature and water table depth), n = 139,365; HTA−LWTD (high annual-average air temperature and low annual-average water table depth), n = 48,426; LTA−HWTD (low annual-average air temperature and high annual-average water table depth), n = 38,411; LTA−LWTD (low annual-average air temperature and water table depth), n = 170,120. (b) Feature importance of the 6-input RF models. TA, air temperature; WTD, water table depth; TS, soil temperature; GPP, gross primary productivity; LE, latent heat turbulent flux; RECO, ecosystem respiration.

drivers. For example, the three dominant drivers identified for scenarios HTA−HWTD, HTA−LWTD and LTA−HWTD were also determined by the RF models as the three most crucial inputs for predicting FCH_4. Additionally, the importance of TS in scenarios with low aa-TA and high aa-WTD was significantly higher than in those with high aa-TA and low aa-WTD. Similarly, the importance of WTD in scenarios with low aa-WTD and high aa-TA was markedly higher than in those with high aa-WTD and low aa-TA. These patterns align perfectly with the correlations observed in Figure 5, strongly supporting our hypothesis regarding the potential shifts in CH_4 drivers under different scenarios.

When reducing inputs to three dominant drivers, the HTA−HWTD and LTA−HWTD scenarios displayed a modest decrease of less than 0.10 in R^2. This indicates that the selected inputs accounted for a significant portion of the variations in FCH_4. Interestingly, the LTA−LWTD scenario demonstrated a marginal enhancement in model performance, with R^2

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065

increasing from 0.77 to 0.78. This indicates that the model built with three dominant drivers not only retained FCH_4 variation information but also displayed slightly improved generalization ability. However, for the HTA−LWTD scenario, the R^2 of the 3-inputs (TS) model decreased by 0.14 compared to the 6-inputs model, indicating some loss of information although the 3-inputs (TS) model still captured the majority of FCH_4 variations.

Upon substituting TS with TA, the R^2 of the HTA−LWTD scenario exhibited a marginal decrease of 0.04, while the remaining three scenarios experienced reductions in R^2 ranging from 0.1 to 0.2. This suggests that TA could serve as an ideal surrogate predictor for wetland sites characterized by high aa-TA and low aa-WTD, primarily due to the stronger dependence of CH_4 emissions on the WTD than on TS in this scenario. For the remaining scenarios, prioritizing TS as input is preferable, while acknowledging that TA-based models remain valid alternatives ($R^2 > 0.55$) in cases of limited TS data availability.

4. DISCUSSION

The net emission of CH_4 from wetlands into the atmosphere depends on the rates of production, consumption, and transportation.[7] Our SEM results reveal a generally positive total effect of various environmental drivers on FCH_4 (Figure 3a,b). These effects could be attributed to several mechanisms, including creating favorable environmental conditions (TA, TS, and WTD)[46,47] and supplying labile carbon substrates (GPP and RECO)[27,48] for methanogenic activities, as well as facilitating CH_4 transport through diffusion, ebullition, and aerenchyma in vascular plants (TA, TS, WTD, LE, and GPP).[7,11,13] However, we also observed negative direct effects for TA and GPP, and a negative indirect effect for WTD. The negative direct effect of TA can be attributed to its promotion of microbial consumption of CH_4 in the upper soil layer,[7] where the impact of TA can more readily reach compared to the deeper layers. The negative direct effect of GPP may be attributed to rhizosphere oxygenation, which favors the oxidation of CH_4 in this otherwise anoxic zone.[49] Furthermore, the negative indirect effect of WTD can be primarily explained by its negative correlation with TS and RECO, as supported by previous studies[50−52] and also evidenced by the detailed SEM results presented in Figure S1.

Our SEM analysis identified TS as the most critical driver for wetland FCH_4 (Figure 3b,d), aligning with many related studies.[4,5,13,14,27,31,32,46] However, the effect of WTD was not significant, ranking lowest among the six drivers. This agrees with previous studies based on the FLUXNET-CH_4 database,[14,27,31] but contradicts several studies based on microcosm experiments,[38,53,54] chamber measurements,[55,56] and observations at lower latitude sites.[57] One reason could be the data bias toward higher latitude sites, which typically exhibit lower aa-TA, resulting in stronger effects of TA and TS but weaker effect of WTD, as well discussed in previous studies.[4,27,32] Another reason might be that CH_4 regulatory mechanisms differ at various spatial scales.[58] The interplay of factors such as temperature, vegetation, and soil conditions at the larger ecosystem scale could mask or moderate the direct effect of water level changes observed in smaller-scale chamber studies and controlled experiments.

Among different wetland types, bogs, fens, and marshes exhibited a similar pattern, with TA and TS being the primary drivers and RECO as a secondary driver (Figure 4), consistent with Yuan et al.[31] The slightly higher significance of LE in marshes could be attributed to the prevalence of aerenchymatous vegetation at all marsh sites (Table S1). LE characterizes the influence of vapor pressure deficit or humidity gradients on pressurized ventilation in aerenchymatous vegetation, serving as a useful proxy for CH_4 transport through vascular plants.[11,13] The greater direct and total effects of LE in marsh sites indicate suggest that the produced CH_4 could be primarily transported through vascular plants. In contrast, swamps were dominated by WTD and LE, while the effects of TA and TS were less significant (Figure 4). This may be due to the higher aa-TA but lower aa-WTD at swamp sites (Figure 1b), ensuring sufficient temperature for methanogenesis while limiting factors become anoxic condition and transport rate. This aligns with Knox et al.,[27] who found WTD to be crucial in wetlands with smaller temperature variations, such as seasonal tropical or subtropical wetlands. Additionally, the effect of RECO in swamps (<0.01) was notably weaker than in other wetland types. This can be explained by the dominance of trees in swamp sites (Table S1). Tree-dominated swamps tend to have more labile carbon in the soil than *Sphagnum* moss-dominated ecosystems, primarily due to the faster decomposition of deciduous plant litter compared to bryophytic litter, and the additional dissolved organic carbon from rainfall passing through the forest canopy.[59] This results in less reliance on breaking down complex carbon compounds through respiration for methanogenesis in these environments.

Generally, our SEM results highlight the dominant role of TS in driving wetland FCH_4, and the existence of diverse regulatory mechanisms that vary across sites. Building on previous research, our work delves deeper by explicitly considering interactions among environmental drivers, thereby providing new and deeper insights into the mechanisms regulating CH_4 emissions. Nevertheless, we also acknowledge the limitations of our analysis. Our SEM relied on long-term, high-temporal-resolution data from widely distributed EC stations. While this data set excels in spatial and temporal coverage, it regrettably lacks soil-related indicators such as dissolved organic carbon,[60] pH,[61] and various microbial indices,[62] which are more fundamental in driving FCH_4 compared to GPP, LE, and RECO. The absence of these important soil-related variables may partly explain our model's relatively poor performance at certain sites, particularly swamps. However, we recognize that monitoring these variables on a seasonal basis could pose substantial financial and logistical hurdles, thus challenging their inclusion in the data set. Additionally, standard SEM is typically designed to analyze static and linear relationships, while our data set consists of time-series data that might exhibit autocorrelations, asynchronous effects, and nonlinear relationships between variables such as WTD, GPP, and FCH_4.[14,32] These inherent data characteristics may lead to biased parameter estimates. Despite these challenges, we believe our SEM still provides reliable results for several reasons. First, according to Knox et al.,[27] various analytical approaches tend to produce consistent results in identifying dominant drivers for FCH_4, regardless of their suitability for handling asynchronous and nonlinear relationships, probably due to the large data set size, which enhances the stability of the estimates. Additionally, our SEM results align well with previous studies using methods such as transfer entropy[31] and relative mutual information,[27] which are capable of addressing nonlinearity and asynchrony. Moreover, the dominant drivers identified for different scenarios in our

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065

study, as inferred theoretically from SEM and RDA results, show good consistency with the feature importance determined by RF, further supporting the reliability of our SEM findings. Nevertheless, it is highly recommended that future research explores advanced approaches to incorporate temporal and nonlinear relationships into SEM, or test other methods such as conditional variable importance for various machine learning models,[63,64] and conditional or partial mutual information,[65−67] thus further enhancing our understanding of wetland CH_4 emissions.

In this study, we also propose a novel approach for better estimation of wetland CH_4 emissions. This involves: first, gaining an in-depth understanding of FCH_4 regulatory mechanisms; second, categorizing sites based on their unique natural conditions and individually prioritizing dominant drivers for each category; and last, modeling and testing through machine learning. Our approach provides a possible solution for making accurate predictions with a few critical predictors when working with limited data, potentially easing the modeling process for large-scale CH_4 estimation. Furthermore, our approach places greater emphasis on accounting for the spatial variability of the dominant drivers when building CH_4 models. It highlights the importance of tailoring predictors to suit the unique annual average temperature and water table conditions of different wetland sites, rather than using a one-size-fits-all set of predictors across all locations. Still, there is room for improvement in our proposed approach. For example, the spatial variability of FCH_4 regulatory mechanisms could be better interpreted by incorporating vegetation-related indicators, such as dominant vegetation coverage,[16] into RDA. The undefined areas in the CH_4 source categorization plot (Figure 6b) could be more precisely delineated through including additional EC data from diverse wetland sites, especially those in the tropical regions. A more effective substitution of TS by TA may be achieved through establishing a stronger TA-TS connection or leveraging targeted temperature models, thus enhancing model practicality without compromising prediction accuracy. These considerations also suggest potential directions for follow-up studies.

Moreover, our study has broader implications. The extensive statistical analysis conducted prior to ML modeling establishes a solid theoretical foundation. The strong alignment between the theoretically inferred dominant drivers and the feature importance determined by the algorithm allows us to shed light on the inner workings of RF models. This greatly enhances our ability to explain these models and, consequently, makes them more trustworthy.[68] Additionally, our results offer a deeper understanding of wetland CH_4 emissions at the ecosystem scale. This could aid in gap-filling EC CH_4 measurements, thus contributing to the creation of a higher-quality local or global database of wetland FCH_4.[69] More importantly, our findings and proposed approach show strong potential in refining natural wetland FCH_4 estimations. Given the current urgency of global warming, this will aid in the more accurate attribution of CH_4 sources and the formulation of targeted mitigation strategies,[4] ultimately contributing to combating climate change.

■ ASSOCIATED CONTENT

Data Availability Statement

The data that support the findings of this study are accessible through the FLUXNET-CH_4 Community Product, available at https://fluxnet.org/data/fluxnet-ch4-community-product/. DOIs for individual site data are provided in Table S1.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.4c06057.

Description of the studied EC sites; supporting literature for the influence pathways within SEM; description of hyperparameters in RF modeling; detailed results of SEM, RDA, and RF modeling; correlation between TA and TS; PDP and ALE plots of the 6-input RF models; performance of RF modeling trained on the entire data set; reasons for selecting PCA for site categorization (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Shubiao Wu − *Department of Agroecology, Aarhus University, Tjele 8830, Denmark;* orcid.org/0000-0003-1203-0680; Email: wushubiao@agro.au.dk

Authors

Haonan Guo − *Department of Agroecology, Aarhus University, Tjele 8830, Denmark;* orcid.org/0009-0003-3054-122X

Shihao Cui − *Department of Agroecology, Aarhus University, Tjele 8830, Denmark*

Claudia Kalla Nielsen − *Department of Agroecology, Aarhus University, Tjele 8830, Denmark*

Johannes Wilhelmus Maria Pullens − *Department of Agroecology, Aarhus University, Tjele 8830, Denmark*

Chunjing Qiu − *Research Center for Global Change and Complex Ecosystems, School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China; Institute of Eco-Chongming, East China Normal University, Shanghai 200241, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.4c06057

Author Contributions

H.G. and S.W. conceived and designed the study. H.G. processed data analysis and produced figures. H.G., S.C., C.K.N and J.W.M.P. contributed to data interpretation. H.G. drafted the manuscript. All authors contributed to the paper revision.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by the European Union's Horizon Europe program (WET HORIZONS, GA number 101056848). H.G. acknowledges support from the China Scholarship Council for the PhD scholarship. We express our gratitude to all authors associated with the EC sites for their diligent efforts and the provision of openly accessible data (a detailed list is provided in Supporting Information).

■ REFERENCES

(1) Forster, P.; Storelvmo, T.; Armour, K.; Collins, W.; Dufresne, J. L.; Frame, D.; Lunt, D. J.; Mauritsen, T.; Palmer, M. D.; Watanabe, M.; Wild, M.; Zhang, H. The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth*

Assessment Report of the Intergovernmental Panel on Climate Change; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B., Eds.; Cambridge University Press, 2021; pp 923−1054.

(2) Bridgham, S. D.; Cadillo-Quiroz, H.; Keller, J. K.; Zhuang, Q. Methane emissions from wetlands: Biogeochemical, microbial, and modeling perspectives from local to global scales. *Global Change Biol.* **2013**, *19* (5), 1325−1346.

(3) Saunois, M.; Stavert, A. R.; Poulter, B.; Bousquet, P.; Canadell, J. G.; Jackson, R. B.; Raymond, P. A.; Dlugokencky, E. J.; Houweling, S.; Patra, P. K.; Ciais, P.; Arora, V. K.; Bastviken, D.; Bergamaschi, P.; Blake, D. R.; Brailsford, G.; Bruhwiler, L.; Carlson, K. M.; Carrol, M.; Castaldi, S.; et al. The global methane budget 2000−2017. *Earth Syst. Sci. Data* **2020**, *12* (3), 1561−1623.

(4) McNicol, G.; Fluet-Chouinard, E.; Ouyang, Z.; Knox, S.; Zhang, Z.; Aalto, T.; Bansal, S.; Chang, K. Y.; Chen, M.; Delwiche, K.; Feron, S.; Goeckede, M.; Liu, J.; Malhotra, A.; Melton, J. R.; Riley, W.; Vargas, R.; Yuan, K.; Ying, Q.; Zhu, Q.; et al. Upscaling wetland methane emissions from the FLUXNET-CH_4 eddy covariance network (UpCH4 v1.0): Model development, network assessment, and budget comparison. *AGU Adv.* **2023**, *4* (5), No. e2023AV000956.

(5) Peltola, O.; Vesala, T.; Gao, Y.; Räty, O.; Alekseychik, P.; Aurela, M.; Chojnicki, B.; Desai, A. R.; Dolman, A. J.; Euskirchen, E. S.; Friborg, T.; Göckede, M.; Helbig, M.; Humphreys, E.; Jackson, R. B.; Jocher, G.; Joos, F.; Klatt, J.; Knox, S. H.; Kowalska, N.; et al. Monthly gridded data product of northern wetland methane emissions based on upscaling eddy covariance observations. *Earth Syst. Sci. Data* **2019**, *11* (3), 1263−1289.

(6) Peter Mayer, H.; Conrad, R. Factors influencing the population of methanogenic bacteria and the initiation of methane production upon flooding of paddy soil. *FEMS Microbiol. Ecol.* **1990**, *73*, 103−111.

(7) Whalen, S. C. Biogeochemistry of methane exchange between natural wetlands and the atmosphere. *Environ. Eng. Sci.* **2005**, *22* (1), 73−94.

(8) Bergman, I.; Klarqvist, M.; Nilsson, M. Seasonal variation in rates of methane production from peat of various botanical origins: effects of temperature and substrate quality. *FEMS Microbiol. Ecol.* **2000**, *33*, 181−189.

(9) Wang, W.; Zeng, C.; Sardans, J.; Wang, C.; Tong, C.; Peñuelas, J. Soil methane production, anaerobic and aerobic oxidation in porewater of wetland soils of the Minjiang River Estuarine, China. *Wetlands* **2018**, *38*, 627−640.

(10) Laanbroek, H. J. Methane emission from natural wetlands: interplay between emergent macrophytes and soil microbial processes. A mini-review. *Ann. Bot.* **2010**, *105*, 141−153.

(11) Villa, J. A.; Ju, Y.; Stephen, T.; Rey-Sanchez, C.; Wrighton, K. C.; Bohrer, G. Plant-mediated methane transport in emergent and floating-leaved species of a temperate freshwater mineral-soil wetland. *Limnol. Oceanogr.* **2020**, *65* (7), 1635−1650.

(12) Morin, T. H.; Bohrer, G.; Frasson, R. P. d. M.; Naor-Azreli, L.; Mesi, S.; Stefanik, K. C.; Schäfer, K. V. R. Environmental drivers of methane fluxes from an urban temperate wetland park. *J. Geophys. Res.: Biogeosci.* **2014**, *119*, 2188−2208.

(13) Morin, T. H. Advances in the eddy covariance approach to CH_4 monitoring over two and a half decades. *J. Geophys. Res.: Biogeosci.* **2019**, *124* (3), 453−460.

(14) Knox, S. H.; Jackson, R. B.; Poulter, B.; McNicol, G.; Fluet-Chouinard, E.; Zhang, Z.; Hugelius, G.; Bousquet, P.; Canadell, J. G.; Saunois, M.; Papale, D.; Chu, H.; Keenan, T. F.; Baldocchi, D.; Torn, M. S.; Mammarella, I.; Trotta, C.; Aurela, M.; Bohrer, G.; Zona, D. FLUXNET-CH_4 synthesis activity: Objectives, observations, and future directions. *Bull. Am. Meteorol. Soc.* **2019**, *100* (12), 2607−2632.

(15) Nadeau, D. F.; Rousseau, A. N.; Coursolle, C.; Margolis, H. A.; Parlange, M. B. Summer methane fluxes from a boreal bog in northern Quebec, Canada, using eddy covariance measurements. *Atmos. Environ.* **2013**, *81*, 464−474.

(16) Bhullar, G. S.; Iravani, M.; Edwards, P. J.; Olde Venterink, H. Methane transport and emissions from soil as affected by water table and vascular plants. *BMC Ecol.* **2013**, *13*, 32.

(17) Wahlen, M. The global methane cycle. *Annu. Rev. Earth Planet. Sci.* **1993**, *21*, 407−426.

(18) Fan, L.; Schneider, D.; Dippold, M. A.; Poehlein, A.; Wu, W.; Gui, H.; Ge, T.; Wu, J.; Thiel, V.; Kuzyakov, Y.; Dorodnikov, M. Active metabolic pathways of anaerobic methane oxidation in paddy soils. *Soil Biol. Biochem.* **2021**, *156*, 108215.

(19) Chowdhury, T. R.; Dick, R. P. Ecology of aerobic methanotrophs in controlling methane fluxes from wetlands. *Appl. Soil Ecol.* **2013**, *65*, 8−22.

(20) Valenzuela, E. I.; Avendaño, K. A.; Balagurusamy, N.; Arriaga, S.; Nieto-Delgado, C.; Thalasso, F.; Cervantes, F. J. Electron shuttling mediated by humic substances fuels anaerobic methane oxidation and carbon burial in wetland sediments. *Sci. Total Environ.* **2019**, *650*, 2674−2684.

(21) Bao, T.; Jia, G.; Xu, X. Wetland heterogeneity determines methane emissions: A pan-Arctic synthesis. *Environ. Sci. Technol.* **2021**, *55* (14), 10152−10163.

(22) Rinne, J.; Tuittila, E.-S.; Peltola, O.; Li, X.; Raivonen, M.; Alekseychik, P.; Haapanala, S.; Pihlatie, M.; Aurela, M.; Mammarella, I.; Vesala, T. Temporal variation of ecosystem scale methane emission from a boreal fen in relation to temperature, water table position, and carbon dioxide fluxes. *Global Biogeochem. Cycles* **2018**, *32* (7), 1087−1106.

(23) Park, K.; Kim, Y.; Lee, K.; Kim, D. Development of a shallow-depth soil temperature estimation model based on air temperatures and soil water contents in a permafrost area. *Appl. Sci.* **2020**, *10* (3), 1058.

(24) Huang, Y.; Ciais, P.; Luo, Y.; Zhu, D.; Wang, Y.; Qiu, C.; Goll, D. S.; Guenet, B.; Makowski, D.; De Graaf, I.; Leifeld, J.; Kwon, M. J.; Hu, J.; Qu, L. Tradeoff of CO_2 and CH_4 emissions from global peatlands under water-table drawdown. *Nat. Clim. Change* **2021**, *11* (7), 618−622.

(25) Verma, S.; Nema, M. K. Development of an empirical model for sub-surface soil moisture estimation and variability assessment in a lesser Himalayan watershed. *Model. Earth Syst. Environ.* **2022**, *8* (3), 3487−3505.

(26) Deng, L.; Wang, K.; Li, J.; Zhao, G.; Shangguan, Z. Effect of soil moisture and atmospheric humidity on both plant productivity and diversity of native grasslands across the Loess Plateau, China. *Ecol. Eng.* **2016**, *94*, 525−531.

(27) Knox, S. H.; Bansal, S.; McNicol, G.; Schafer, K.; Sturtevant, C.; Ueyama, M.; Valach, A. C.; Baldocchi, D.; Delwiche, K.; Desai, A. R.; Euskirchen, E.; Liu, J.; Lohila, A.; Malhotra, A.; Melling, L.; Riley, W.; Runkle, B. R. K.; Turner, J.; Vargas, R.; Zhu, Q.; et al. Identifying dominant environmental predictors of freshwater wetland methane fluxes across diurnal to seasonal time scales. *Global Change Biol.* **2021**, *27* (15), 3582−3604.

(28) Sturtevant, C.; Ruddell, B. L.; Knox, S. H.; Verfaillie, J.; Matthes, J. H.; Oikawa, P. Y.; Baldocchi, D. Identifying scale-emergent, nonlinear, asynchronous processes of wetland methane exchange. *J. Geophys. Res.: Biogeosci.* **2016**, *121* (1), 188−204.

(29) White, J. R.; Shannon, R. D.; Weltzin, J. F.; Pastor, J.; Bridgham, S. D. Effects of soil warming and drying on methane cycling in a northern peatland mesocosm study. *J. Geophys. Res.* **2008**, *113*, G00A06.

(30) Chang, K. Y.; Riley, W. J.; Knox, S. H.; Jackson, R. B.; McNicol, G.; Poulter, B.; Aurela, M.; Baldocchi, D.; Bansal, S.; Bohrer, G.; Campbell, D. I.; Cescatti, A.; Chu, H.; Delwiche, K. B.; Desai, A. R.; Euskirchen, E.; Friborg, T.; Goeckede, M.; Helbig, M.; Hemes, K. S.; et al. Substantial hysteresis in emergent temperature sensitivity of global wetland CH_4 emissions. *Nat. Commun.* **2021**, *12* (1), 2266.

(31) Yuan, K.; Zhu, Q.; Li, F.; Riley, W. J.; Torn, M.; Chu, H.; McNicol, G.; Chen, M.; Knox, S.; Delwiche, K.; Wu, H.; Baldocchi, D.; Ma, H.; Desai, A. R.; Chen, J.; Sachs, T.; Ueyama, M.; Sonnentag, O.; Helbig, M.; Tuittila, E. S.; et al. Causality guided machine learning

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065

model on wetland CH_4 emissions across global wetlands. *Agric. For. Meteorol.* **2022**, *324*, 109115.
(32) Delwiche, K. B.; Knox, S. H.; Malhotra, A.; Fluet-Chouinard, E.; McNicol, G.; Feron, S.; Ouyang, Z.; Papale, D.; Trotta, C.; Canfora, E.; Cheah, Y.-W.; Christianson, D.; Alberto, M. C. R.; Alekseychik, P.; Aurela, M.; Baldocchi, D.; Bansal, S.; Billesbach, D. P.; Bohrer, G.; Bracho, R.; et al. FLUXNET-CH_4: a global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands. *Earth Syst. Sci. Data* **2021**, *13* (7), 3607−3689.
(33) Koebsch, F.; Sonnentag, O.; Jarveoja, J.; Peltoniemi, M.; Alekseychik, P.; Aurela, M.; Arslan, A. N.; Dinsmore, K.; Gianelle, D.; Helfter, C.; Jackowicz-Korczynski, M.; Korrensalo, A.; Leith, F.; Linkosalmi, M.; Lohila, A.; Lund, M.; Maddison, M.; Mammarella, I.; Mander, U.; Minkkinen, K.; et al. Refining the role of phenology in regulating gross ecosystem productivity across European peatlands. *Global Change Biol.* **2020**, *26* (2), 876−887.
(34) Chen, Q.; Long, C.; Chen, J.; Cheng, X. Differential response of soil CO_2, CH_4, and N_2O emissions to edaphic properties and microbial attributes following afforestation in central China. *Global Change Biol.* **2021**, *27* (21), 5657−5669.
(35) Wang, F.; Chen, Y.; Li, T.; Wang, C.; Wang, D.; Fu, B.; Lv, Y.; Wu, X. Grazing reduces the soil-atmosphere exchange of greenhouse gases during freeze-thaw cycles in meadow steppes in Inner Mongolia. *Front. Ecol. Evol.* **2021**, *9*, 795203.
(36) Joyce, J.; Jewell, P. W. Physical controls on methane ebullition from reservoirs and lakes. *Environ. Eng. Geosci.* **2003**, *9* (2), 167−178.
(37) Angle, J. C.; Morin, T. H.; Solden, L. M.; Narrowe, A. B.; Smith, G. J.; Borton, M. A.; Rey-Sanchez, C.; Daly, R. A.; Mirfenderesgi, G.; Hoyt, D. W.; Riley, W. J.; Miller, C. S.; Bohrer, G.; Wrighton, K. C. Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nat. Commun.* **2017**, *8* (1), 1567.
(38) Zhao, M.; Han, G.; Li, J.; Song, W.; Qu, W.; Eller, F.; Wang, J.; Jiang, C. Responses of soil CO_2 and CH_4 emissions to changing water table level in a coastal wetland. *J. Clean. Prod.* **2020**, *269*, 122316.
(39) Grace, J. B.; Anderson, T. M.; Olff, H.; Scheiner, S. M. On the specification of structural equation models for ecological systems. *Ecol. Monogr.* **2010**, *80* (1), 67−87.
(40) Harrison, L.; Stephan, K.; Friston, K. CHAPTER 38-Effective Connectivity. In *Statistical Parametric Mapping*; Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W., Eds.; Academic Press, 2007; pp 508−521.
(41) Bollen, K. A. Total, direct, and indirect effects in structural equation models. *Socio. Methodol.* **1987**, *17*, 37−69.
(42) McArdle, B. H.; Anderson, M. J. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology* **2001**, *82* (1), 290−297.
(43) Majdi, N.; Traunspurger, W.; Boyer, S.; Mialet, B.; Tackx, M.; Fernandez, R.; Gehner, S.; Ten-Hage, L.; Buffan-Dubau, E. Response of biofilm-dwelling nematodes to habitat changes in the Garonne River, France: Influence of hydrodynamics and microalgal availability. *Hydrobiologia* **2011**, *673* (1), 229−244.
(44) Greenacre, M.; Groenen, P. J. F.; Hastie, T.; D'Enza, A. I.; Markos, A.; Tuzhilina, E. Principal component analysis. *Nat. Rev. Methods Primers* **2022**, *2* (1), 100.
(45) Schonlau, M.; Zou, R. Y. The random forest algorithm for statistical learning. *STATA J.* **2020**, *20* (1), 3−29.
(46) Turetsky, M. R.; Kotowska, A.; Bubier, J.; Dise, N. B.; Crill, P.; Hornibrook, E. R.; Minkkinen, K.; Moore, T. R.; Myers-Smith, I. H.; Nykanen, H.; Olefeldt, D.; Rinne, J.; Saarnio, S.; Shurpali, N.; Tuittila, E. S.; Waddington, J. M.; White, J. R.; Wickland, K. P.; Wilmking, M. A synthesis of methane emissions from 71 northern, temperate, and subtropical wetlands. *Global Change Biol.* **2014**, *20* (7), 2183−2197.
(47) Westermann, P. Temperature regulation of methanogenesis in wetlands. *Chemosphere* **1993**, *26* (1−4), 321−328.
(48) Whiting, G. J.; Chanton, J. P. Primary production control of methane emission from wetlands. *Nature* **1993**, *364* (6440), 794−795.
(49) Colmer, T. D. Long-distance transport of gases in plants: a perspective on internal aeration and radial oxygen loss from roots. *Plant Cell Environ.* **2003**, *26* (1), 17−36.
(50) Helfter, C.; Campbell, C.; Dinsmore, K. J.; Drewer, J.; Coyle, M.; Anderson, M.; Skiba, U.; Nemitz, E.; Billett, M. F.; Sutton, M. A. Drivers of long-term variability in CO_2 net ecosystem exchange in a temperate peatland. *Biogeosciences* **2015**, *12* (6), 1799−1811.
(51) Chen, X.; Hu, Q. Groundwater influences on soil moisture and surface evaporation. *J. Hydrol.* **2004**, *297* (1−4), 285−300.
(52) Zhang, Z.; Chen, X.; Pan, Z.; Zhao, P.; Zhang, J.; Jiang, K.; Wang, J.; Han, G.; Song, Y.; Huang, N.; Ma, S.; Zhang, J.; Yin, W.; Zhang, Z.; Men, J. Quantitative estimation of the effects of soil moisture on temperature using a soil water and heat coupling model. *Agriculture* **2022**, *12* (9), 1371.
(53) Yang, J.; Liu, J.; Hu, X.; Li, X.; Wang, Y.; Li, H. Effect of water table level on CO_2, CH_4 and N_2O emissions in a freshwater marsh of Northeast China. *Soil Biol. Biochem.* **2013**, *61*, 52−60.
(54) Zhu, X.; Song, C.; Chen, W.; Zhang, X.; Tao, B. Effects of water regimes on methane emissions in peatland and gley marsh. *Vadose Zone J.* **2018**, *17* (1), 1−7.
(55) Calabrese, S.; Garcia, A.; Wilmoth, J. L.; Zhang, X.; Porporato, A. Critical inundation level for methane emissions from wetlands. *Environ. Res. Lett.* **2021**, *16* (4), 044038.
(56) Evans, C. D.; Peacock, M.; Baird, A. J.; Artz, R. R. E.; Burden, A.; Callaghan, N.; Chapman, P. J.; Cooper, H. M.; Coyle, M.; Craig, E.; Cumming, A.; Dixon, S.; Gauci, V.; Grayson, R. P.; Helfter, C.; Heppell, C. M.; Holden, J.; Jones, D. L.; Kaduk, J.; Levy, P.; et al. Overriding water table control on managed peatland greenhouse gas emissions. *Nature* **2021**, *593* (7860), 548−552.
(57) Goodrich, J. P.; Campbell, D. I.; Roulet, N. T.; Clearwater, M. J.; Schipper, L. A. Overriding control of methane flux temporal variability by water table dynamics in a Southern Hemisphere, raised bog. *J. Geophys. Res.: Biogeosci.* **2015**, *120* (5), 819−831.
(58) Stiehl-Braun, P. A.; Hartmann, A. A.; Kandeler, E.; Buchmann, N.; Niklaus, P. A. Interactive effects of drought and N fertilization on the spatial distribution of methane assimilation in grassland soils. *Global Change Biol.* **2011**, *17* (8), 2629−2639.
(59) Avagyan, A.; Runkle, B. R. K.; Hartmann, J.; Kutzbach, L. Spatial variations in pore-water biogeochemistry greatly exceed temporal changes during baseflow conditions in a boreal river valley mire complex, Northwest Russia. *Wetlands* **2014**, *34* (6), 1171−1182.
(60) Osterloh, K.; Tauchnitz, N.; Spott, O.; Hepp, J.; Bernsdorf, S.; Meissner, R. Changes of methane and nitrous oxide emissions in a transition bog in central Germany (German National Park Harz Mountains) after rewetting. *Wetl. Ecol. Manag.* **2018**, *26* (1), 87−102.
(61) Abdalla, M.; Hastings, A.; Truu, J.; Espenberg, M.; Mander, U.; Smith, P. Emissions of methane from northern peatlands: a review of management impacts and implications for future management options. *Ecol. Evol.* **2016**, *6* (19), 7080−7102.
(62) Chen, X.; Ma, H.; Zheng, Y.; Liu, J.; Liang, X.; He, C. Changes in methane emission and methanogenic and methanotrophic communities in restored wetland with introduction of Alnus trabeculosa. *J. Soils Sediments* **2017**, *17* (1), 181−189.
(63) Molnar, C.; König, G.; Bischl, B.; Casalicchio, G. Model-agnostic feature importance and effects with dependent features: a conditional subgroup approach. *Data Min. Knowl. Discov.* **2023**, 1.
(64) Strobl, C.; Boulesteix, A. L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinf.* **2008**, *9*, 307.
(65) Frenzel, S.; Pompe, B. Partial mutual information for coupling analysis of multivariate time series. *Phys. Rev. Lett.* **2007**, *99*, 204101.
(66) Zhao, J.; Zhou, Y.; Zhang, X.; Chen, L. Part mutual information for quantifying direct associations in networks. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 5130−5135.
(67) Fleuret, F. Fast binary feature selection with conditional mutual information. *J. Mach. Learn. Res.* **2004**, *5*, 1531−1555.
(68) Maxwell, A. E.; Sharma, M.; Donaldson, K. A. Explainable boosting machines for slope failure spatial predictive modeling. *Rem. Sens.* **2021**, *13* (24), 4991.

(69) Kim, Y.; Johnson, M. S.; Knox, S. H.; Black, T. A.; Dalmagro, H. J.; Kang, M.; Kim, J.; Baldocchi, D. 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 Biol.* **2020**, *26*, 1499−1518.

https://doi.org/10.1021/acs.est.4c06057
Environ. Sci. Technol. 2024, 58, 15052−15065