water

MDPI

Article

Quantifying Drivers of Methane Hydrobiogeochemistry in a Tidal River Floodplain System

Z. Jason Hou [1,*], Nicholas D. Ward [1,2], Allison N. Myers-Pigg [1], Xinming Lin [1], Scott R. Waichler [1], Cora Wiese Moore [1,3], Matthew J. Norwood [1], Peter Regier [1] and Steven B. Yabusaki [1]

[1] Pacific Northwest National Laboratory, Richland, WA 99354, USA; nicholas.ward@pnnl.gov (N.D.W.); allison.myers-pigg@pnnl.gov (A.N.M.-P.); xinming.lin@pnnl.gov (X.L.); scott.waichler@pnnl.gov (S.R.W.); cora.wiesemoore@pnnl.gov (C.W.M.); matthew.norwood@pnnl.gov (M.J.N.); peter.regier@pnnl.gov (P.R.); yabusaki@pnnl.gov (S.B.Y.)
[2] School of Oceanography, University of Washington, Seattle, WA 98195, USA
[3] Merrick & Company, San Antonio, TX 78258, USA
* Correspondence: zhangshuan.hou@pnnl.gov

Abstract: The influence of coastal ecosystems on global greenhouse gas (GHG) budgets and their response to increasing inundation and salinization remains poorly constrained. In this study, we have integrated an uncertainty quantification (UQ) and ensemble machine learning (ML) framework to identify and rank the most influential processes, properties, and conditions controlling methane behavior in a freshwater floodplain responding to recently restored seawater inundation. Our unique multivariate, multiyear, and multi-site dataset comprises tidal creek and floodplain porewater observations encompassing water level, salinity, pH, temperature, dissolved oxygen (DO), dissolved organic carbon (DOC), total dissolved nitrogen (TDN), partial pressure of carbon dioxide (pCO_2), nitrous oxide (pN_2O), methane (pCH_4), and the stable isotopic composition of methane ($\delta^{13}CH_4$). Additionally, we incorporated topographical data, soil porosity, hydraulic conductivity, and water retention parameters for UQ analysis using a previously developed 3D variably saturated flow and transport floodplain model for a physical mechanistic understanding of factors influencing groundwater levels and salinity and, therefore, CH_4. Principal component analysis revealed that groundwater level and salinity are the most significant predictors of overall biogeochemical variability. The ensemble ML models and UQ analyses identified DO, water level, salinity, and temperature as the most influential factors for porewater methane levels and indicated that approximately 80% of the total variability in hourly water levels and around 60% of the total variability in hourly salinity can be explained by permeability, creek water level, and two van Genuchten water retention function parameters: the air-entry suction parameter α and the pore size distribution parameter m. These findings provide insights on the physicochemical factors in methane behavior in coastal ecosystems and their representation in local- to global-scale Earth system models.

Keywords: methane; hydrobiogeochemistry; drivers; floodplain; machine learning; ModEx

check for updates

Citation: Hou, Z.J.; Ward, N.D.; Myers-Pigg, A.N.; Lin, X.; Waichler, S.R.; Wiese Moore, C.; Norwood, M.J.; Regier, P.; Yabusaki, S.B. Quantifying Drivers of Methane Hydrobiogeochemistry in a Tidal River Floodplain System. *Water* **2024**, *16*, 171. https://doi.org/10.3390/w16010171

Academic Editor: Douglas A. Burns

Received: 18 November 2023
Revised: 28 December 2023
Accepted: 29 December 2023
Published: 2 January 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

A large uncertainty for global greenhouse gas (GHG) budgets is how the increasing inundation and salinization of coastal freshwater ecosystems will affect terrestrial GHG sources and sinks [1]. Methane, being a potent GHG with a higher radiative trapping efficiency than carbon dioxide, plays a pivotal role in these budgets. The need for a systematic understanding of methane cycling dynamics and how they balance the efficient burial of carbon in coastal ecosystems is particularly acute considering the variable nature of methane cycling and the intricate interplay of factors influencing these processes [2]. Global GHG budgets do not currently include a robust quantification of the methane emissions from brackish coastal wetlands [3]. While brackish wetlands generally emit less methane and more CO_2 than their freshwater counterparts [4], the emissions originating

from coastal floodplains might partially counteract the carbon burial rates observed in coastal sediments [5] because methane has considerably greater global warming potential than CO_2 [6].

Furthermore, the dynamics of methane in coastal freshwater ecosystems under the influence of rising inundation and salinization remain a subject of uncertainty and ongoing investigation [7]. To address these uncertainties and enhance our understanding, it is imperative to undertake a systematic and predictive study that establishes well-defined associations between observed methane dynamics and on-site field conditions. This approach draws inspiration from analogous studies conducted in inland terrestrial ecosystems or freshwater wetlands, where such linkages have contributed significantly to our comprehension of methane's behavior [8,9].

Developing predictive models of methane's behavior can benefit from the identification of the most consequential processes, properties, and conditions affecting subsurface methane concentrations and, ultimately, fluxes [10]. High-resolution reactive transport models can represent a complex network of coupled biogeochemical reactions under dynamic hydrological conditions [11]. However, to feasibly represent coastal ecosystems in global-scale Earth system models, we must understand how the behavior of methane and other climatically relevant elements co-vary with basic physicochemical parameters [12]. For example, water table depth and temperature can largely explain methane's behavior in inland terrestrial ecosystems [13,14], while salinity is likely another major parameter needed to explain coastal methane dynamics.

To gain a comprehensive mechanistic understanding of methane dynamics, a diverse dataset encompassing field observations and numerical model outputs can be combined synergistically, with ensemble machine learning (ML) models serving as a valuable tool to enhance accuracy and reliability, ultimately forming an ML-enabled model–experiment (ModEx) framework. Prior studies have employed ML-ModEx to analyze coastal and inland floodplains, extracting patterns and quantitative relationships from observed variables to model and understand their behavior. For example, principal component analysis (PCA) [15], artificial neural network (ANN) [16–18], and tree-based methods [19,20], as well as linear regression models [21], have been used to build prediction models, access relationships among environmental variables, or reduce input dimensions. ANN models can simulate and predict water quality variables such as the chlorophyll concentration, salinity, and dissolved oxygen in estuaries and coastal systems [22,23]. Random Forest (RF) [24], a tree-based ensemble ML technique, was implemented to model wetland inundation patterns across a large semiarid floodplain [25] and spatiotemporal oxygen dynamics across coastal terrestrial–aquatic interfaces [20] as well as to predict methane emissions and identify environmental drivers of variation in methane of different ecosystems [26,27]. Boosted regression trees (BRTs) [28] were used to identify controls on sedimentation and nutrient transformations in a study of restored urban stream–floodplain systems [29].

Despite this progress, there is still a big knowledge gap in the quantitative mechanistic understanding of the driving factors of methane cycling in coastal floodplain ecosystems, and few studies have quantified parameter uncertainty and identified model sensitivity to hydrologic conditions in the representation of methane processes. We hypothesize that a subset of physical, spatial, and temporal parameters and factors play critical roles in methane hydrobiogeochemistry, which can be tested with comprehensive ensemble ML approaches, for example, PCA, ANN, RF, generalized linear model (GLM), and gradient boosting machine (GBM). These ML models identify and rank the most influential environmental conditions on observed methane concentrations, develop predictive methane concentration relationships, and quantify the uncertainty and sensitivity of methane distributions to simulation parameters and conditions that could be used to prioritize characterization and the sampling/monitoring strategy. Specifically, we use observations in freshwater floodplains along a tidal stream responding to recently restored seawater inundation [30] to identify statistically significant relationships between methane behavior and site conditions.

Site characterization included the topography and soil properties. Physical and chemical data were routinely collected across multiple transects spanning transverse gradients from the tidal creek to the floodplain hillslope.

Our study aims to establish a foundational understanding of key environmental factors influencing methane's behavior in floodplains, providing a basis for future investigations into methane fluxes. By integrating a numerical model, we linked methane dynamics with subsurface material properties, identifying sensitive model parameters. We found that specific physical parameters, such as permeability and water retention parameters related to air entry and pore size distribution, significantly influence methane behavior, as evidenced by their correlations with the water level and salinity.

2. Site and Data

2.1. Site Description

Beaver Creek is a first-order watershed in western Washington State, USA. The creek and floodplain are tidally influenced across much of its extent, although prior to 2014, a barrier prevented tidal access, and the floodplain supported freshwater vegetation and soil microbial communities [31,32]. The removal of the barrier to tidal access in 2014 restored saltwater access to the stream, which is a 2.5 km tributary that drains into the Johns River and Grays Harbor estuary (Washington State Department of Fish and Wildlife, 2019; [30]). Four field sampling sites in order of decreasing salinity exposure are referenced herein: the marsh, the downstream floodplain, the upstream floodplain, and the wetland (Figure 1). The study focuses on three transects (marsh, downstream floodplain, upstream floodplain) with designated data acquisition locations, all of which are subject to varying water-level fluctuations.

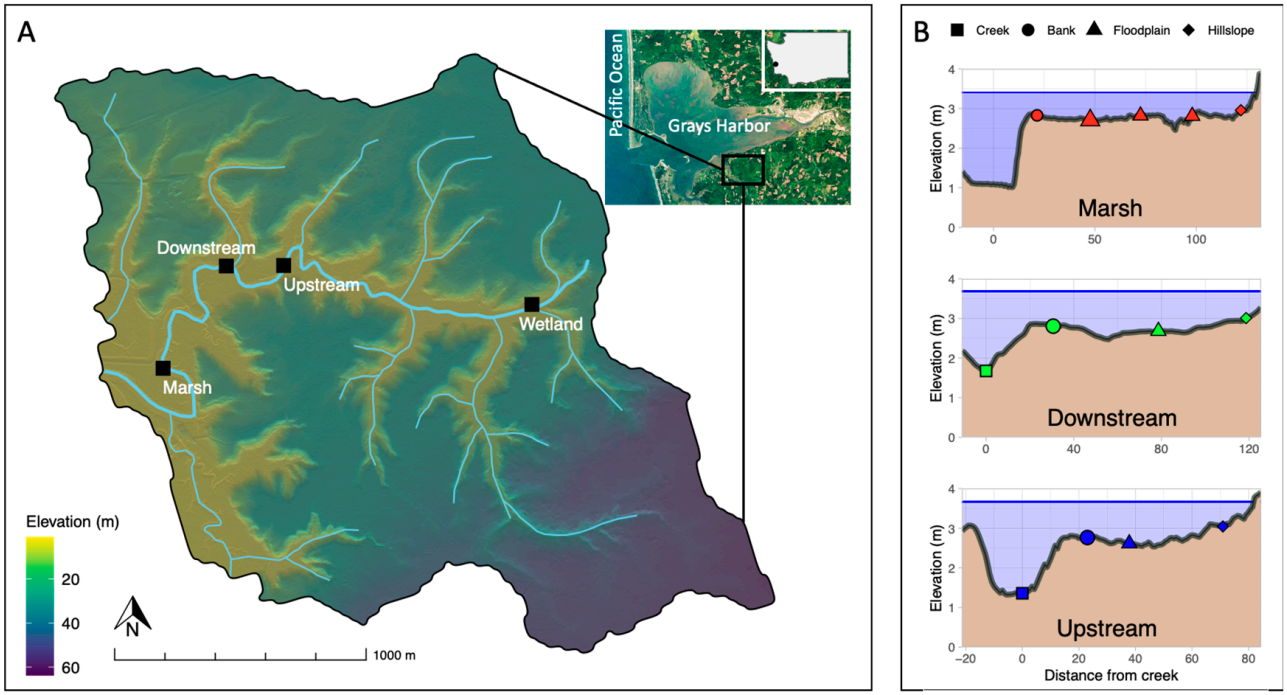

Figure 1. The Beaver Creek watershed and the locations of four floodplain field sites (**Panel A**); three transects of focus with data acquisition locations (**Panel B**).

2.2. Data Collection and Generation

We generated a novel multivariate hydrobiogeochemical dataset across the Beaver Creek watershed from January 2018 to January 2020. Shallow groundwater wells, extending one meter deep, were strategically placed along three hillslope-to-stream transects within the tidally influenced reaches (Figure 1B), along with one location in the non-tidal freshwa-

ter wetland at the watershed's headwaters (Figure 1A). Multi-attribute in situ sensors were installed in the wells and creek to monitor water temperature, salinity, pH, and water level at 30-minute intervals [30,33].

Porewaters were collected using sipper rods deployed at 30, 60, and 90 cm (M.H.E. Products PushPoint Research Samplers) [34], while surface waters were collected at the surface of the stream in 1 L acid-washed brown HDPE grab bottles. Overall, there were 157, 230, 69, and 22 point samples for marsh, downstream, upstream, and wetland, respectively. Point samples were processed for salinity, temperature, pH, dissolved O_2 (DO), oxidation–reduction potential (ORP), dissolved organic carbon (DOC), total dissolved N (TDN), pCO_2 (partial pressure of CO_2), pCH_4 (partial pressure of CH_4), $\delta^{13}CH_4$ (the stable isotopic composition of methane), pN_2O (partial pressure of N_2O), water level, and temperature, using a combination of field and laboratory analyses. Flow-through measurements were taken for pH, ORP, dissolved oxygen, salinity, and temperature using either an Ekton handheld sensor for samplings prior to August 2018 or a YSI Pro handheld multiparameter sonde for later samplings, when possible. When it was not possible to collect flow-through measurements, linear interpolations for in situ sensor reporting of temperature, salinity, and pH were used to fill in data at the reported collection date and times, within one minute of the actual collection time using the R package 'zoo' function 'na.approx' [35].

Filtered samples (0.2 µm) were analyzed for DOC and TDN using a Shimadzu TOC-L system equipped with a total nitrogen module at the PNNL Marine and Coastal Research Laboratory [33] and in the Bianchi Lab at the University of Florida [36]. Headspace gas measurements and analyses followed the procedure in [34]. Briefly, 20 mL of porewater and 40 mL of pure N_2 gas were mixed in the sampling syringe, which was shaken for approximately 1 min to equilibrate the gases in the water sample. Porewater gas concentrations were corrected from ppm for the dissolution of gas in the water by converting them to µmol/L with Henry's law using temperature- and salinity-dependent coefficients described by [37]. Linear interpolations within one minute of collection time were used to fill in any missing data at the reported collection date and times using in situ sensor data for temperature and salinity, using the R package 'zoo' function 'na.approx' [35]. In sites where porewaters were unable to be extracted, a 60 mL gas sample was taken through the soil probe. pCO_2, pCH_4, pN_2O, were analyzed as partial pressures by direct injection on a Picarro G2508 Cavity Ring-Down Spectrometer with a flow limiter installed to reduce gas flow rates. Samples were diluted with pure nitrogen when GHG levels were above the instrument's threshold. All values were corrected for dilutions based on the common gas law. $\delta^{13}CH_4$ was analyzed using a Los Gatos Research (LGR) Methane Carbon Isotope Analyzer (MCIA) providing methane measurements up to 1 ppm and corrected as detailed above for dilution.

Complementing the field observational data, numerical flow and transport simulations were employed to augment our datasets. These simulations were essential for developing a more systematic and physics-based understanding of the floodplain system. The mathematical equations that depict the flow and transport processes, along with the associated relative permeability and water retention functions, were taken from [38]. To solve the equations numerically, we utilized the PFLOTRAN saturated–unsaturated subsurface simulator [39]. PFLOTRAN solves a system of generally nonlinear partial differential equations describing multiphase, multicomponent and multiscale reactive flow and transport in porous materials, and it was used in this study to model 3-D variably saturated water flow and salinity transport in the upstream floodplain site. We summarize the model specification below; for a more detailed description, please see [30]. Briefly, the rectilinear model grid was based on a LiDAR digital elevation map (DEM) and had 3 m cells in the horizontal direction and 0.1 m cell thickness in the vertical direction. The grid extent was 243 m east–west by 153 m north–south by 7.2 m in the vertical direction. Soil was specified as a homogeneous silty clay. Hydraulic properties included permeability, porosity, and van Genuchten water retention function parameters with the Mualem relative permeability function. Flow and salinity in the floodplain were driven by time-varying water level

and salinity observations from the Creek and Hillslope sensors. The bottom boundary at elevation of 0 m and two short sections joining the Hillslope and Creek boundaries were designated as no-flow. A seepage-face boundary condition was specified on the floodplain ground surface, where inundation was simulated according to the stream water level. The model used one year of observations which began on 24 April 2018 as input. The freshwater initial condition of the site prior to the restoration of seawater access to Beaver Creek in 2014 was created by repeating the boundary condition water levels with zero salinity for two model years. This "spin-up" generated a dynamic steady-state of a repeating annual cycle of water levels and porewater saturation in the model domain. The resulting hydrologic initial condition was used to start the four-year model run with salinity to reconstruct the post-restoration salinization of the floodplain by quasi-monthly inundation from the creek. Model output from the final year compared favorably with the data collected during April 2018–April 2019 [30].

3. Machine Learning of Field and Simulation Data

In coastal ecosystems, ML can support data collection, integrate diverse data into models, identify key processes, and offer data-driven process representations. Principal component analysis (PCA) is a statistical method that transforms correlated initial variables into uncorrelated principal components (PCs) [40], assessing the genetic distance and relatedness between the original variables. In our study, we applied PCA to point-sampling data from the Beaver Creek system transects to evaluate methane concentrations' relationships with environmental conditions (e.g., pCO_2, pH, DOC, and temperature). We used standardized environmental conditions (e.g., pCO_2, pH, DOC, and temperature) as inputs for PCA, centered to have a mean of zero and unit standard deviation. PCA was limited to porewater data from the marsh, downstream floodplain, and upstream floodplain due to data insufficiency in the surface waters.

Random Forest (RF) is an ensemble machine learning method that creates multiple decision trees with bootstrapped data samples and random subsets of predictor variables, leading to a strong predictive model [41,42]. It also calculates variable importance, typically using Mean Decrease Impurity (MDI). Gradient Boosting Machine (GBM) is another tree-based ensemble ML method, differing from RF in its constructive ensemble strategy. GBM sequentially adds new models to the ensemble, training them to minimize the error of the previous ensemble [43]. In this work, the ensemble tree-based models (e.g., RF and GBM) were built to identify and rank the most influential environmental conditions controlling the methane behavior in a first-order tidal stream floodplain. The response variable of the RF and GBM models is the observed pCH_4, and the predictor variables include water level, salinity, temperature, landscape, pH, DO, DOC, location, and sampling depth. Then, we developed GLM, RF, GBM, and ANN models to evaluate the floodplain modeling parameters' potential influence on the simulated water levels and salinity under diverse conditions. The predictor variables of these models are the floodplain modeling parameters, which include (1) permeability, with an estimated range of $[1 \times 10^{-14}, 1 \times 10^{-12}]$ m^2; (2) water-level offset in the creek, with a range of $[-0.1, 0.1]$ m; (3) porosity, with a range of [0.35, 0.65]; (4) residual saturation, with a range of [0.0756, 0.222]; (5) the van Genuchten model parameter α (i.e., inverse of air entry suction) in a range of $[5.1 \times 10^{-5}, 3.67 \times 10^{-4}]$ Pa^{-1}; and (6) the van Genuchten model parameter m (i.e., pore size distribution) with a range of [0.0826, 0.359]. The response variables are the simulated model misfits (e.g., mean absolute error, MAE) of the water levels and salinity collected at different sites (e.g., marsh, downstream, upstream) and at different time scales (e.g., half-hourly, semi-diurnal, diurnal, and monthly) in order to understand the scale-dependent behaviors of the system dynamics [44]. Multi-fold cross-validation was done by splitting the dataset into 80% training and 20% testing subsets, and the finalized models were then applied to the full dataset and compared for their goodness of fit (R-squared) as well as the relative parameter importance of the parameters for each of the response variables.

Principal steps to this process include: (1) Identify the most significant features correlated with methane concentrations which may include landscape position and depth, salinity and water level, oxygen, CO_2, pH, $\delta^{13}CH_4$, and temperature; (2) Assess floodplain modeling parameters in an existing variably saturated flow and salinity transport model of the floodplain for influence on the simulated water levels and salinity under various conditions; (3) Link the parameters in this model to the observed methane behavior. This is achieved by combining the sensitivity of modeled hydrology and salinity to hydraulic parameters with the observed correlations/significance of methane behavior in relation to the floodplain's water table elevation and salinity.

4. Results

4.1. Machine Learning Results

PCA of porewaters indicates that water level and salinity along the first principal component are the most significant in explaining the overall variability in the porewater dataset and are strongly correlated (Figure 2). Porewater pCH_4 and temperature along the second principal component are also important contributors to the system variability. Porewater pCH_4 is positively correlated with DOC and pCO_2 and highly negatively correlated with $\delta^{13}CH_4$. Porewater measurements in the downstream transect are the least clustered and exhibit the largest spatial variability as a result of tide–topography interplay (Figure 2). These correlations are beneficial in comprehending methane dynamics because $\delta^{13}CH_4$ can distinguish between biogenic and thermogenic methane origins, as biogenic methane tends to have more negative $\delta^{13}CH_4$ values, while thermogenic methane has less negative or even positive values. Simultaneously, the levels of porewater pCH_4 can serve as indicators of microbial methane production rates and the associated biogeochemical processes. The negative correlation between $\delta^{13}CH_4$ and pCH_4, DO, and pCO_2 indicates that hydrogenotrophic methanogenesis is a dominant process in methane production in the study area, as this pathway typically results in more negative $\delta^{13}C$ values, and that oxygen and carbon dioxide levels influence the availability of substrates for methane-producing microorganisms. The raw data can be found in the figshare repository at https://doi.org/10.6084/m9.figshare.c.6759357.v1 (accessed on 28 December 2023).

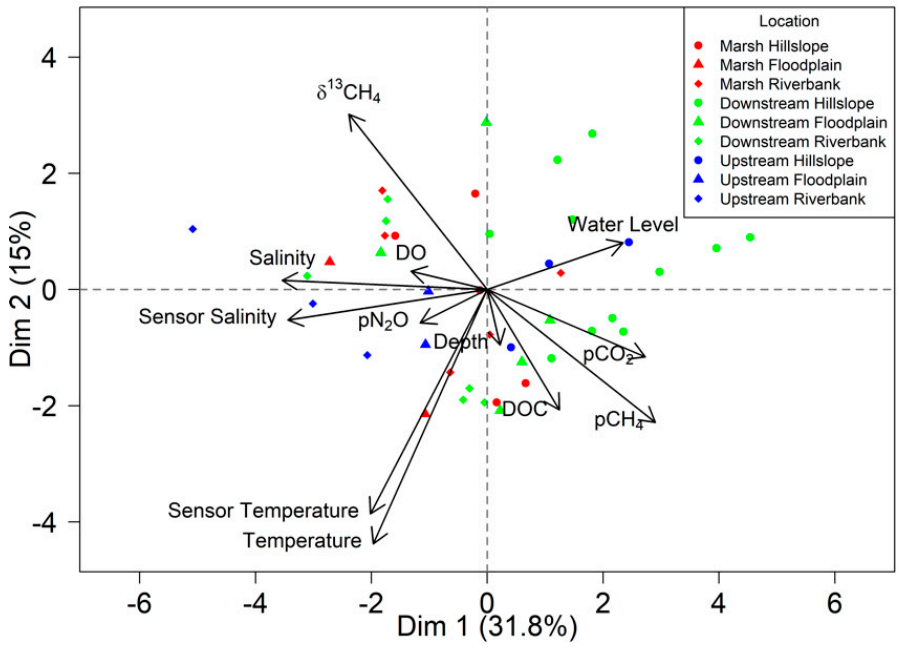

Figure 2. The principal component analysis (PCA) biplot for all the measured attributes in the porewater across landscapes.

RF and GBM models were built to identify and rank the most influential environmental conditions controlling methane behavior in a first-order tidal stream floodplain system. More than 60% of the variation in pCH_4 can be explained by the two predictive models. Feature importance rankings from both RF and GBM show that DO and water level contribute the most to pCH_4 variations (Figure 3). Even though the rankings from the RF and GBM models are a bit different, both show that the salinity, pH, temperature, and site locations play non-negligible but secondary roles. Specifically, the methane concentrations were found to have significant cross-dependence with water table elevation (positive), temperature (positive), and salinity (negative), as shown in Figure 2. Considering the large variation in the dynamic water levels and salinity across the floodplain, these cross-dependencies can be useful for estimating their impacts on methane distributions, especially when and where data were not collected but modeled outputs are available. In this case, the parameter uncertainty and sensitivity of a 3D variably saturated flow and transport floodplain model were used to further identify the extended influence of modeling inputs on methane behavior using the identified cross-dependence in the context of highly resolved spatiotemporal modeled water levels and salinity.

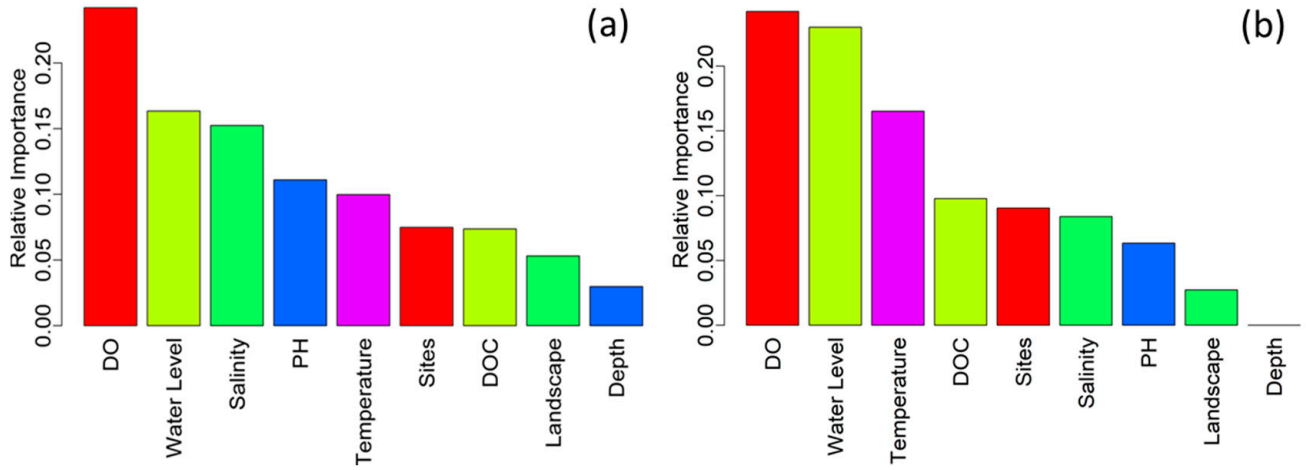

Figure 3. Feature importance ranking from the (**a**) RF and (**b**) GBM models with porewater pCH_4 as the target variable. DO and water level contribute the most to pCH_4 variations.

4.2. Uncertainty Quantification Results

Ensemble numerical simulation outputs were used to understand the drivers and their impacts on salinity and water level variations using data from four different time scales (i.e., half-hourly, semi-diurnal, monthly). The goodness of fit analyses were conducted using four different models (i.e., GBM, GLM, RF, and ANN). As shown in Figure 4, all four models agree that log permeability (logK) is the most important variable for explaining both salinity variability and water-level variability. For example, logK alone can explain ~90% of the total explained variance of water levels and ~70% of the total variance of salinity in the GLM model. All four models are better at explaining the salinity variability than water level changes, especially at the upstream floodplain, where more than 80% of the total variability in salinity is explained for data of all time scales. Out of the six floodplain modeling parameters, logK is the dominant factor followed by the van Genuchten α parameter (α) as shown in Figure 4.

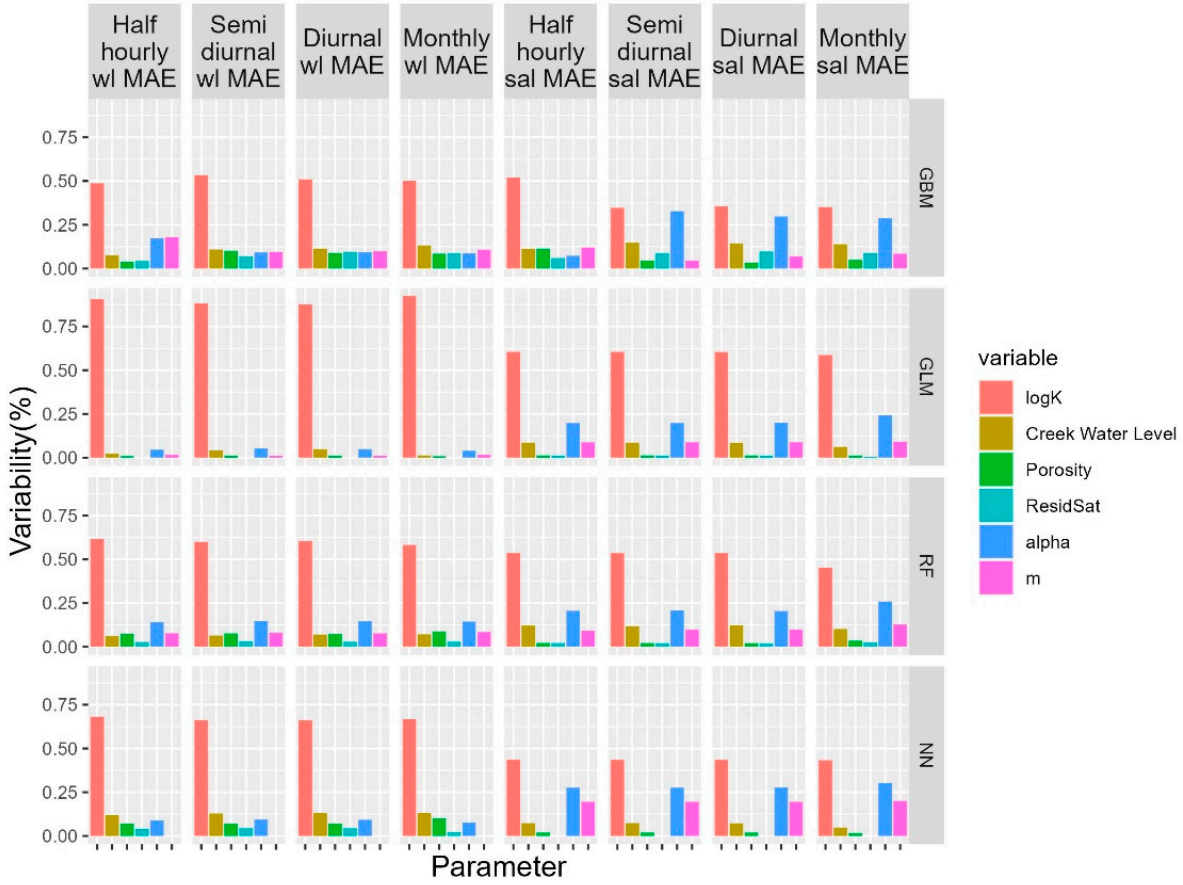

Figure 4. Floodplain model parameters' relative contributions ranked by GBM, GLM, RF, and ANN. Among the eight response variables, wl MAE denotes water level mean absolute error, while sal MAE represents salinity mean absolute error at half-hourly, semi-diurnal, diurnal, and monthly scales, respectively.

5. Discussion

Our research underscores the pivotal roles played by water level and salinity in shaping the intricate dynamics of methane within coastal wetlands and ecosystems undergoing transition, which is consistent with the findings of some previous studies [4,7]. Using ML techniques, we have not only reaffirmed their central role in environmental variability and methane concentrations within this ecosystem but also uncovered other crucial factors. Specifically, we have identified that the source of methane (i.e., hydrogenotrophic, acetoclastic, or methylotrophic methanogenesis) and the extent of oxidation, as indicated by its ^{13}C isotopic signature, significantly influence floodplain methane abundance, as determined by the ensemble ML models.

Constraining methane sources and the oxidation extent is particularly important to consider in coastal systems under a changing climate, as these transition zones may undergo dynamic shifts in terminal electron acceptors [45], microbial activity [31], and subsurface transport [46]. Additionally, factors such as other greenhouse gas dynamics and oxygen content play important roles in shaping methane concentrations within the floodplain. These findings highlight the importance of considering a multitude of environmental variables for understanding methane dynamics, whether in inland or coastal ecosystems [47]. It is also important to note that feedbacks between important features like water level and oxygen, and the spatial and temporal variability of these relationships [20], add additional complexity to our predictive understanding of methane. The negative correlation of methane concentrations with salinity observed here may arise from various underlying processes, some of which we are unable to fully disentangle within the scope

of this study. These processes include dilution and displacement effects [48], methane oxidation [12,49,50], competition among terminal electron acceptors [51], and the influence of salinity on heterotrophs [52,53]. For example, one of the immediate effects of salinity pulses in previously freshwater systems could be the potential for short-term increases in methane concentrations. The introduction of saline water can alter the redox conditions in the subsurface, creating more favorable conditions for methanogenesis, although the presence of sulfate and sulfate-reducing bacteria in saline environments can lead to the suppression of methanogenesis and the promotion of anaerobic methane oxidation coupled to sulfate reduction. While the short-term response may be driven by the displacement of freshwater and the introduction of organic matter from adjacent saline waters; long-term, chronic saltwater intrusion could have lasting impacts on methane concentrations. In salt marshes and coastal environments, chronic exposure to elevated salinity levels can result in a shift in the microbial communities with uncertain implications for the promotion of methane-producing microorganisms and methane production.

The pre-existing floodplain flow and transport model demonstrated a remarkable ability to replicate observed dynamic water levels and salinities across the floodplain. This validation was performed through comparisons with data collected from three strategically positioned wells spanning the floodplain, extending from its proximity to the hillslope to its proximity to the streambank of Beaver Creek [30]. These three landscape positions represent distinctly different regimes of exposure to conditions in the stream. As shown in Figure 1, the near-bank floodplain is closest to the stream and strongly influenced by the stream water level and salinity dynamics. The central floodplain has the lowest ground surface elevation, which corresponds to the highest inundation magnitude and duration for stream water reaching this location. Conversely, the near-hillslope floodplain has the highest elevations in the floodplain and is only exposed to stream waters with the highest water levels of the year [30,33]. The simulated water table elevations and salinities in the floodplain are sensitive to hydraulic parameters, shedding light on how these factors interplay within the freshwater floodplain system.

Our findings highlight the robust statistical correlations that bridge porewater pCH_4 levels to water levels and salinity, as well as the evident influence of hydraulic parameters on modeled water levels and salinity. These insights expand our understanding of methane behaviors beyond mere hydraulic considerations. While it may seem obvious that permeability is a sensitive factor affecting simulated water levels and salinity, its correlation with methane allows for a systematic link between permeability and spatiotemporal methane behavior. For example, lower permeability leads to longer residence time and higher floodplain salinization, which are associated with lower porewater methane concentrations. The van Genuchten α parameter, which represents the inverse of the air-entry pressure, was also a sensitive parameter for the modeled salinity. This parameter controls the water content as saturation is approached, which is critical to the capacitance of the subsurface to accept infiltration and accompanying salinity during an inundation event. This alteration of salinization adversely affects methane concentrations. Unlike the air entry parameter, porosity was not found to be a sensitive parameter for salinity, despite its potential role in controlling transport velocity and the capacitance of the unsaturated zone. This phenomenon can be partly attributed to Beaver Creek's relatively low permeability, resulting in lateral transport across the floodplain occurring on decadal time scales, somewhat diminishing the influence of porosity on transport velocity. Furthermore, the sensitivity analysis indicates that the degree of saturation has more of an impact on capacitance than the porosity itself. The positive correlation of methane observations with water table elevation appears to be consistent with the common observation of more reducing conditions below the water table in organic-rich environments. Although inundation events lead to the infiltration of salinity-bearing surface water, they also contribute to higher water tables that can raise methane-bearing groundwater (of lower redox potential) into previously unsaturated zones.

6. Conclusions

This study developed UQ and ML models for learning the field monitoring data of major processes, properties, and conditions controlling methane's behavior, as well as for evaluating ensemble numerical simulation outputs. By combining a unique multivariate hydrobiogeochemical dataset with a physics-based floodplain hydrology and salinity model, along with ensemble ML techniques, we have advanced our understanding of the major drivers and influencing factors shaping methane dynamics in this dynamic ecosystem.

Our analysis has shed light on the key factors influencing pCH_4 variations, confirming and highlighting the dominant roles of dissolved oxygen and the water level. Furthermore, our findings underscore the noteworthy influences of salinity, pH, temperature, and sampling location. Notably, methane concentrations exhibit intriguing patterns, displaying a positive correlation with water table elevation and temperature and a negative correlation with salinity. It is noteworthy that soil texture should be regarded as equally or even more crucial than factors like DO, DOC, temperature, pH, nitrate, and so on in influencing methane behavior. In the upstream floodplain, six modeling parameters explained over 80% of the salinity variability, with permeability (logk) being the most influential, followed by the van Genuchten α parameter, which underscores the soil texture's importance in coastal floodplains with increased seawater exposure, suggesting the potential for parameter reduction or surrogate models. Our future study will involve an evaluation of the spatial transferability of the developed understanding, where additional data will be needed to cover more study sites, similar to the effort in [54], or integrate other potentially important parameters such as soil temperature [55].

These quantitative relationships derived from our integrated field and numerical simulation data also provide a basis for developing a conceptual model for such a typical floodplain, where soils are poised for methanogenesis, controlled by the presence of oxidants, and where near-surface floodplain soils are generally methane-oxidizing in the presence of atmospheric O_2 and sulfate-bearing salinity. In contrast, the saturated zone is largely anaerobic [33], conducive to methane production and persistence in the absence of salinity (i.e., sulfate). The inundation and progressive salinization of the lower floodplain elevations is leading to lower porewater methane levels, consistent with local increases in dilution, oxidation, and competition for substrate.

While our study primarily focused on methane behavior, we recognize that the interplay between oxygen-consuming and oxygen-producing organisms can have a significant impact on the overall ecosystem dynamics and plan to further investigate and refine our understanding in future research endeavors.

Author Contributions: Conceptualization, Z.J.H. and S.B.Y.; data acquisition, N.D.W., A.N.M.-P., C.W.M., M.J.N., and P.R.; machine learning, X.L.; simulation, S.R.W.; writing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been funded by the PREMIS Initiative conducted under the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory (PNNL), by the DOE COMPASS-FME (https://compass.pnnl.gov/), and by the River Corridor Science Focus Area (SFA) supported by the DOE Office of Biological and Environmental Research (BER) Environmental System Science (ESS) Program.

Data Availability Statement: Data used in this manuscript have been compiled and made publicly available at https://doi.org/10.6084/m9.figshare.c.6759357.v1 (accessed on 28 December 2023).

Acknowledgments: Modeling was performed using resources available through Research Computing at PNNL. PNNL is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830. We thank Don Lentz, the Hancock Timber Company, and Washington Department of Fish & Wildlife for access to the Beaver Creek field site.

Conflicts of Interest: Author Cora Wiese Moore was employed by the Merrick & Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Al-Haj, A.N.; Fulweiler, R.W. A synthesis of methane emissions from shallow vegetated coastal ecosystems. *Glob. Change Biol.* **2020**, *26*, 2988–3005. [CrossRef] [PubMed]
2. Liu, L.; Wang, D.; Chen, S.; Yu, Z.; Xu, Y.; Li, Y.; Ge, Z.; Chen, Z. Methane emissions from estuarine coastal wetlands: Implications for global change effect. *Soil Sci. Soc. Am. J.* **2019**, *83*, 1368–1377. [CrossRef]
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. The global methane budget 2000–2017. *Earth Syst. Sci. Data* **2020**, *12*, 1561–1623. [CrossRef]
4. Poffenbarger, H.J.; Needelman, B.A.; Megonigal, J.P. Salinity influence on methane emissions from tidal marshes. *Wetlands* **2011**, *31*, 831–842. [CrossRef]
5. Kauffman, J.B.; Giovanonni, L.; Kelly, J.; Dunstan, N.; Borde, A.; Diefenderfer, H.; Cornu, C.; Janousek, C.; Apple, J.; Brophy, L. Total ecosystem carbon stocks at the marine-terrestrial interface: Blue carbon of the Pacific Northwest Coast, United States. *Glob. Change Biol.* **2020**, *26*, 5679–5692. [CrossRef] [PubMed]
6. Neubauer, S.C.; Megonigal, J.P. Moving beyond global warming potentials to quantify the climatic role of ecosystems. *Ecosystems* **2015**, *18*, 1000–1013. [CrossRef]
7. Chamberlain, S.D.; Hemes, K.S.; Eichelmann, E.; Szutu, D.J.; Verfaillie, J.G.; Baldocchi, D.D. Effect of drought-induced salinization on wetland methane emissions, gross ecosystem productivity, and their interactions. *Ecosystems* **2020**, *23*, 675–688. [CrossRef]
8. Melton, J.; Wania, R.; Hodson, E.; Poulter, B.; Ringeval, B.; Spahni, R.; Bohn, T.; Avis, C.; Beerling, D.; Chen, G. Present state of global wetland extent and wetland methane modelling: Conclusions from a model inter-comparison project (WETCHIMP). *Biogeosciences* **2013**, *10*, 753–788. [CrossRef]
9. Knox, S.H.; Bansal, S.; McNicol, G.; Schafer, K.; Sturtevant, C.; Ueyama, M.; Valach, A.C.; Baldocchi, D.; Delwiche, K.; Desai, A.R. Identifying dominant environmental predictors of freshwater wetland methane fluxes across diurnal to seasonal time scales. *Glob. Change Biol.* **2021**, *27*, 3582–3604. [CrossRef]
10. Weston, N.B.; Neubauer, S.C.; Velinsky, D.J.; Vile, M.A. Net ecosystem carbon exchange and the greenhouse gas balance of tidal marshes along an estuarine salinity gradient. *Biogeochemistry* **2014**, *120*, 163–189. [CrossRef]
11. Danczak, R.E.; Yabusaki, S.B.; Williams, K.H.; Fang, Y.; Hobson, C.; Wilkins, M.J. Snowmelt induced hydrologic perturbations drive dynamic microbiological and geochemical behaviors across a shallow riparian aquifer. *Front. Earth Sci.* **2016**, *4*, 57. [CrossRef]
12. Ward, N.D.; Bianchi, T.S.; Martin, J.B.; Quintero, C.J.; Sawakuchi, H.O.; Cohen, M.J. Pathways for methane emissions and oxidation that influence the net carbon balance of a subtropical cypress swamp. *Front. Earth Sci.* **2020**, *8*, 573357. [CrossRef]
13. Xu, X.; Yuan, F.; Hanson, P.J.; Wullschleger, S.D.; Thornton, P.E.; Riley, W.J.; Song, X.; Graham, D.E.; Song, C.; Tian, H. Reviews and syntheses: Four decades of modeling methane cycling in terrestrial ecosystems. *Biogeosciences* **2016**, *13*, 3735–3755. [CrossRef]
14. Calabrese, S.; Garcia, A.; Wilmoth, J.L.; Zhang, X.; Porporato, A. Critical inundation level for methane emissions from wetlands. *Environ. Res. Lett.* **2021**, *16*, 044038. [CrossRef]
15. Alcázar, J.; Palau, A. Establishing environmental flow regimes in a Mediterranean watershed based on a regional classification. *J. Hydrol.* **2010**, *388*, 41–51. [CrossRef]
16. Fathian, F.; Mehdizadeh, S.; Sales, A.K.; Safari, M.J.S. Hybrid models to improve the monthly river flow prediction: Integrating artificial intelligence and non-linear time series models. *J. Hydrol.* **2019**, *575*, 1200–1213. [CrossRef]
17. Maier, H.R.; Jain, A.; Dandy, G.C.; Sudheer, K.P. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environ. Model. Softw.* **2010**, *25*, 891–909. [CrossRef]
18. Wang, W.-C.; Chau, K.-W.; Cheng, C.-T.; Qiu, L. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J. Hydrol.* **2009**, *374*, 294–306. [CrossRef]
19. De'ath, G.; Fabricius, K.E. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* **2000**, *81*, 3178–3192. [CrossRef]
20. Regier, P.J.; Ward, N.D.; Myers-Pigg, A.N.; Grate, J.; Freeman, M.J.; Ghosh, R.N. Seasonal drivers of dissolved oxygen across a tidal creek–marsh interface revealed by machine learning. *Limnol. Oceanogr.* **2023**, *68*, 2359–2374. [CrossRef]
21. Heuvelmans, G.; Muys, B.; Feyen, J. Regionalisation of the parameters of a hydrological model: Comparison of linear regression models with artificial neural nets. *J. Hydrol.* **2006**, *319*, 245–265. [CrossRef]
22. Alizadeh, M.J.; Kavianpour, M.R. Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Mar. Pollut. Bull.* **2015**, *98*, 171–178. [CrossRef] [PubMed]
23. Shamshirband, S.; Jafari Nodoushan, E.; Adolf, J.E.; Abdul Manaf, A.; Mosavi, A.; Chau, K.-w. Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters. *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 91–101. [CrossRef]
24. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
25. Shaeri Karimi, S.; Saintilan, N.; Wen, L.; Valavi, R. Application of machine learning to model wetland inundation patterns across a large semiarid floodplain. *Water Resour. Res.* **2019**, *55*, 8765–8778. [CrossRef]
26. Mosher, J.J.; Fortner, A.M.; Phillips, J.R.; Bevelhimer, M.S.; Stewart, A.J.; Troia, M.J. Spatial and temporal correlates of greenhouse gas diffusion from a hydropower reservoir in the southern United States. *Water* **2015**, *7*, 5910–5927. [CrossRef]

27. Peltola, O.; Vesala, T.; Gao, Y.; Räty, O.; Alekseychik, P.; Aurela, M.; Chojnicki, B.; Desai, A.R.; Dolman, A.J.; Euskirchen, E.S. Monthly gridded data product of northern wetland methane emissions based on upscaling eddy covariance observations. *Earth Syst. Sci. Data* **2019**, *11*, 1263–1289. [CrossRef]
28. De'Ath, G. Boosted trees for ecological modeling and prediction. *Ecology* **2007**, *88*, 243–251. [CrossRef]
29. McMillan, S.K.; Noe, G.B. Increasing floodplain connectivity through urban stream restoration increases nutrient and sediment retention. *Ecol. Eng.* **2017**, *108*, 284–295. [CrossRef]
30. Yabusaki, S.B.; Myers-Pigg, A.N.; Ward, N.D.; Waichler, S.R.; Sengupta, A.; Hou, Z.; Chen, X.; Fang, Y.; Duan, Z.; Serkowski, J.A. Floodplain inundation and salinization from a recently restored first-order tidal stream. *Water Resour. Res.* **2020**, *56*, e2019WR026850. [CrossRef]
31. Sengupta, A.; Indivero, J.; Gunn, C.; Tfaily, M.M.; Chu, R.K.; Toyoda, J.; Bailey, V.L.; Ward, N.D.; Stegen, J.C. Spatial gradients in soil-carbon character of a coastal forested floodplain are associated 1 with abiotic features, but not microbial communities 2. *Biogeosciences* **2019**, *16*, 3911–3928. [CrossRef]
32. Wang, W.; McDowell, N.G.; Ward, N.D.; Indivero, J.; Gunn, C.; Bailey, V.L. Constrained tree growth and gas exchange of seawater-exposed forests in the Pacific Northwest, USA. *J. Ecol.* **2019**, *107*, 2541–2552. [CrossRef]
33. Regier, P.; Ward, N.D.; Indivero, J.; Wiese Moore, C.; Norwood, M.; Myers-Pigg, A. Biogeochemical control points of connectivity between a tidal creek and its floodplain. *Limnol. Oceanogr. Lett.* **2021**, *6*, 134–142. [CrossRef]
34. Ward, N.; Indivero, J.; Gunn, C.; Wang, W.; Bailey, V.; McDowell, N. Longitudinal gradients in tree stem greenhouse gas concentrations across six Pacific Northwest coastal forests. *J. Geophys. Res. Biogeosci.* **2019**, *124*, 1401–1412. [CrossRef]
35. Zeileis, A.; Grothendieck, G. Zoo: S3 infrastructure for regular and irregular time series. *arXiv* **2005**, arXiv:math/0505527.
36. Arellano, A.; Bianchi, T.S.; Osburn, C.; D'Sa, E.; Ward, N.D.; Oviedo-Vargas, D.; Joshi, I.; Ko, D.; Shields, M.; Kurian, G. Mechanisms of organic matter export in estuaries with contrasting carbon sources. *J. Geophys. Res. Biogeosci.* **2019**, *124*, 3168–3188. [CrossRef]
37. Wanninkhof, R. Relationship between wind speed and gas exchange over the ocean revisited. *Limnol. Oceanogr. Methods* **2014**, *12*, 351–362. [CrossRef]
38. Nielsen, D.; Van Genuchten, M.T.; Biggar, J. Water flow and solute transport processes in the unsaturated zone. *Water Resour. Res.* **1986**, *22*, 89S–108S. [CrossRef]
39. Hammond, G.E.; Lichtner, P.C.; Mills, R. Evaluating the performance of parallel subsurface simulators: An illustrative example with PFLOTRAN. *Water Resour. Res.* **2014**, *50*, 208–228. [CrossRef]
40. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **2008**, *26*, 303–304. [CrossRef]
41. Boulesteix, A.L.; Janitza, S.; Kruppa, J.; König, I.R. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2012**, *2*, 493–507. [CrossRef]
42. Regier, P.; Duggan, M.; Myers-Pigg, A.; Ward, N. Effects of random forest modeling decisions on biogeochemical time series predictions. *Limnol. Oceanogr. Methods* **2023**, *21*, 40–52. [CrossRef]
43. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorobot.* **2013**, *7*, 21. [CrossRef] [PubMed]
44. 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*, 188–204. [CrossRef]
45. RoyChowdhury, T.; Bramer, L.; Hoyt, D.W.; Kim, Y.-M.; Metz, T.O.; McCue, L.A.; Diefenderfer, H.L.; Jansson, J.K.; Bailey, V. Temporal dynamics of CO_2 and CH_4 loss potentials in response to rapid hydrological shifts in tidal freshwater wetland soils. *Ecol. Eng.* **2018**, *114*, 104–114. [CrossRef]
46. Trifunovic, B.; Vázquez-Lule, A.; Capooci, M.; Seyfferth, A.L.; Moffat, C.; Vargas, R. Carbon dioxide and methane emissions from a temperate salt marsh tidal creek. *J. Geophys. Res. Biogeosci.* **2020**, *125*, e2019JG005558. [CrossRef]
47. Covey, K.; Soper, F.; Pangala, S.; Bernardino, A.; Pagliaro, Z.; Basso, L.; Cassol, H.; Fearnside, P.; Navarrete, D.; Novoa, S. Carbon and beyond: The biogeochemistry of climate in a rapidly changing Amazon. *Front. For. Glob. Change* **2021**, *4*, 11. [CrossRef]
48. McNicol, G.; Silver, W.L. Separate effects of flooding and anaerobiosis on soil greenhouse gas emissions and redox sensitive biogeochemistry. *J. Geophys. Res. Biogeosci.* **2014**, *119*, 557–566. [CrossRef]
49. Sawakuchi, H.O.; Neu, V.; Ward, N.D.; Barros, M.d.L.C.; Valerio, A.M.; Gagne-Maynard, W.; Cunha, A.C.; Less, D.F.; Diniz, J.E.; Brito, D.C. Carbon dioxide emissions along the lower Amazon River. *Front. Mar. Sci.* **2017**, *4*, 76. [CrossRef]
50. Malyan, S.K.; Bhatia, A.; Kumar, A.; Gupta, D.K.; Singh, R.; Kumar, S.S.; Tomer, R.; Kumar, O.; Jain, N. Methane production, oxidation and mitigation: A mechanistic understanding and comprehensive evaluation of influencing factors. *Sci. Total Environ.* **2016**, *572*, 874–896. [CrossRef]
51. Cord-Ruwisch, R.; Seitz, H.-J.; Conrad, R. The capacity of hydrogenotrophic anaerobic bacteria to compete for traces of hydrogen depends on the redox potential of the terminal electron acceptor. *Arch. Microbiol.* **1988**, *149*, 350–357. [CrossRef]
52. Cunha, M.; Almeida, M.; Alcântara, F. Patterns of ectoenzymatic and heterotrophic bacterial activities along a salinity gradient in a shallow tidal estuary. *Mar. Ecol. Prog. Ser.* **2000**, *204*, 1–12. [CrossRef]
53. He, Z.; Geng, S.; Shen, L.; Lou, L.; Zheng, P.; Xu, X.; Hu, B. The short-and long-term effects of environmental conditions on anaerobic methane oxidation coupled to nitrite reduction. *Water Res.* **2015**, *68*, 554–562. [CrossRef] [PubMed]

54. Knox, S.H.; Jackson, R.B.; Poulter, B.; McNicol, G.; Fluet-Chouinard, E.; Zhang, Z.; Hugelius, G.; Bousquet, P.; Canadell, J.G.; Saunois, M. FLUXNET-CH 4 synthesis activity: Objectives, observations, and future directions. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 2607–2632.
55. Liu, J.; Zhou, Y.; Valach, A.; Shortt, R.; Kasak, K.; Rey-Sanchez, C.; Hemes, K.S.; Baldocchi, D.; Lai, D.Y. Methane emissions reduce the radiative cooling effect of a subtropical estuarine mangrove wetland by half. *Glob. Change Biol.* **2020**, *26*, 4998–5016. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.