



The role of reservoir size in driving methane emissions in China

Zilin Wang^a, Meili Feng^{a,*}, Matthew F. Johnson^{b,**}, Aldo Lipani^c, Faith Chan^{a,d,**}

^a School of Geographical Sciences, University of Nottingham Ningbo China, Ningbo 315100, China

^b School of Geography, Sir Clive Granger Building, University of Nottingham, University Park, Nottingham NG7 2R, United Kingdom

^c University College London (UCL), London, UK

^d Water@Leeds Research Institute and School of Geography, University of Leeds, Leeds LS2 9JT, UK

ARTICLE INFO

Keywords:

Methane
Reservoirs
China
Machine Learning
Greenhouse gas

ABSTRACT

Reservoirs play a crucial role as sources of methane (CH₄) emissions, with emission rates and quantities varying widely depending on reservoir size due to factors such as surface area, water depth, usage, operational methods, and spatial distribution. Gaining insights into emission characteristics across different reservoir sizes can aid in designing and managing reservoirs to mitigate CH₄ emissions effectively. In this study, machine learning models were applied to estimate both diffusive and ebullitive CH₄ emissions across 97,435 reservoirs in China, spanning five categories of storage capacity. This comprehensive assessment covers nearly all reservoirs within the country, revealing total CH₄ emissions of approximately 5,414 Gg. Reservoirs > 0.01 km³ are responsible for about 90 % of these emissions, primarily due to high diffusive flux rates and extensive surface areas. Elevated CH₄ diffusion in reservoirs > 0.01 km³ is largely influenced by their thermal stratification and capacity for organic matter accumulation. Furthermore, these reservoirs are particularly vulnerable to climate warming, which could accelerate CH₄ emission rates more rapidly in larger reservoirs than in smaller ones (below 0.01 km³). Consequently, prioritising CH₄ management in reservoirs > 0.01 km³ is imperative. Nevertheless, the high ebullitive flux of CH₄ in reservoirs < 0.01 km³, linked to their shallow depth, highlighting the potential for significant CH₄ ebullition from smaller aquatic systems. Given large and small-ranged reservoirs' distinct spatial distribution patterns, targeted management strategies are recommended: project-level management for large reservoirs and basin-level approaches for smaller reservoirs.

1. Introduction

In the past two decades, extensive studies have unequivocally identified that reservoirs are substantial sources of greenhouse gas (GHG) emissions (Barros et al., 2011; Harrison et al., 2021). Research during this period has endeavoured not only to quantify GHG fluxes from reservoirs but also to investigate the critical factors that drive these emissions. Such factors include eutrophication, temperature, water level, sedimentation, duration since impoundment, wind speed, and stream flow, each playing a crucial role in the production and release of carbon dioxide (CO₂) and methane (CH₄) (Deemer et al., 2016).

The varying storage capacities of reservoirs give rise to differences in surface area, water depth, organic matter decomposition rates, usage, operational practices, and distribution patterns. These differences subsequently lead to the formation of distinct emissions profiles. Several studies have investigated CO₂ and CH₄ emissions in reservoir areas,

reporting a decrease in emissions as the reservoir area increases (Grinham et al., 2018; Hølgerson and Raymond, 2016; DelSontro et al., 2018). However, these studies predominantly classify reservoirs based on surface area, without adequately considering how variations in storage capacity influence emission patterns. Understanding the impact of storage capacity variations on emission profiles is crucial for optimising reservoir design to reduce GHG emissions while meeting water storage needs. Additionally, this knowledge supports the development of more focused and effective policies to mitigate the environmental impacts of reservoir emissions.

CH₄ is a significant carbon source from reservoirs, responsible for approximately 44 % of anthropogenic CH₄ emissions from freshwater systems (Saunio et al., 2024). Its release occurs mainly via three pathways: diffusion, ebullition, and degassing (Wang et al., 2024). CH₄ diffusion follows the general gas diffusing mechanism, where CH₄ is emitted through the water surface due to concentration gradients. In the

* Corresponding author.

** Corresponding authors.

E-mail addresses: meili.feng@nottingham.edu.cn (M. Feng), M.Johnson@nottingham.ac.uk (M.F. Johnson), Faith.Chan@nottingham.edu.cn (F. Chan).

<https://doi.org/10.1016/j.watres.2025.123441>

Received 7 December 2024; Received in revised form 23 February 2025; Accepted 4 March 2025

Available online 5 March 2025

0043-1354/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

case of CH₄ ebullition, CH₄ production typically takes place within anoxic sediments (Lessmann et al., 2023). Due to the low porosity of deeper sediments, CH₄ does not easily dissolve in water but tends to accumulate in bubbles that are saturated with methane (Sobek et al., 2012). In shallow waters, where hydrostatic pressure is insufficient to contain these bubbles, they ascend through the water column, break at the surface, and release CH₄ into the atmosphere (Bastviken et al., 2004). In deeper water columns, however, CH₄ bubbles can dissolve gradually, with some emitted through diffusion, while the remainder is oxidised into CO₂. Additionally, CH₄ may accumulate in the bottom water layer (epilimnion), especially in anoxic conditions. When water from the epilimnion is released for hydropower, CH₄ is discharged into the atmosphere via turbines and flood control spillways, a process known as CH₄ degassing (Jager et al., 2022).

By 2022, China accounted for nearly half of the world's dams, encompassing 95,926 reservoirs, including 814 large (>0.1 km³), 4192 medium (0.01–0.1 km³), and 90,290 small (<0.01 km³) reservoirs (Ministry of Water Resources, 2022). Gaining insight into carbon emissions from China's reservoirs is crucial to the global effort to shape effective climate change mitigation strategies. Therefore, this study focuses on analysing carbon emissions from Chinese reservoirs to explore viable carbon reduction strategies based on this extensive case study.

Increasingly, machine learning models are being used to estimate carbon emissions from inland water bodies. These models are powerful tools, capable of identifying linear and non-linear relationships within datasets and detecting complex patterns without necessitating a detailed understanding of variable interactions (Zhong et al., 2021). For instance, various Artificial Neural Network (ANN) algorithms have been successfully used to predict CO₂ emissions from reservoirs, outperforming traditional multiple linear regression models (Chen et al., 2018). Rocher-Ros et al. (2023) also employed a Random Forest (RF) machine learning model to estimate global CH₄ levels from rivers and streams, achieving R² values between 0.45 and 0.68. However, limited data on CH₄ emissions from reservoirs has restricted the development of machine learning applications in this area as such models typically rely on relatively large datasets for training and validation.

This study aims to use a newly compiled dataset on CH₄ emissions to apply machine learning models for estimating CH₄ emissions from documented reservoirs across China. The specific objectives are as follows:

1. Examine variations in CH₄ emissions across reservoirs of different sizes and purposes.
2. Calculate cumulative CH₄ emissions from reservoirs of varying scales, identifying the primary contributors to national emissions.
3. Analyze the variation of key variables across different reservoir sizes and propose potential carbon reduction strategies, providing insights for optimizing reservoir management practices to mitigate greenhouse gas emissions.

2. Methods

2.1. Literature review data collection

The database for CH₄ diffusive and ebullition emissions was constructed primarily from the G-res Dataset (Prairie et al., 2021) and compilations by Li et al. (2018) and Zheng et al. (2022), which included sampling data up to 2015. We expanded this dataset by incorporating more recent measurements from 2015 to 2024 (see Supplementary Data). Data were systematically sourced from Scopus, and Google Scholar, focusing on studies with keywords like 'carbon', 'emissions', and 'reservoirs' in the title. After an initial selection, titles, abstracts, and full texts were manually reviewed to retain only studies relevant to CH₄ emissions from reservoirs. Ultimately, this curation resulted in 528 diffusive CH₄ emissions records from 284 reservoirs and 106 ebullitive CH₄ emissions records from 70 reservoirs.

2.2. Predicting variables for CH₄ emissions

To predict CH₄ diffusive and ebullition emissions, spatially explicit datasets were used, including reservoir area (km²) and volume (million m³) from the Chinese Reservoir Database (CRD) (Song et al., 2022a). Climate-relevant data such as temperature (°C), precipitation (kg/m²/month), wind speed (m/s), potential evapotranspiration (PET) (kg/m²/month), vapour pressure deficit (VPD) (Pa), humidity (%), and net primary productivity (NPP) (gC/m²/yr) were extracted from the Climatologies High Resolution for the Earth's Land Surface Areas (CHELSA) dataset version 2.1, providing monthly averages from 1981 to 2010 at spatial resolutions of 0.25°–1° (Karger et al., 2017; Brun et al., 2022). We also included average monthly discharge (m³/s) from 1976 to 2005 (Bosmans et al., 2022) and global horizontal irradiance (GHI) (kWh/m²/month) (NASA, 2008). The specific monthly values corresponding to CH₄ observations were assigned, while average values were used for multi-month or yearly observations. The potential relationship between the input variables and CH₄ emissions is illustrated in Table 1.

Table 1

The potential relationship may exist between the input variables and CH₄ emissions.

| Variables | Relationships with CH ₄ emissions |
|---|--|
| Reservoir area (km²) | The size of reservoirs determines the availability of organic carbon and sediment deposition zones, which influence methanogenesis (Grinham et al., 2018). |
| Volume (million m³) | Reservoir volume affects the distribution of anaerobic zones, organic matter deposition, and methane production potential (Wang et al., 2024). |
| Temperature (°C) | The higher the temperature, the lower the solubility of CH ₄ and the faster the production of CH ₄ (Keller and Stallard, 1994; West et al., 2016). |
| Precipitation (kg/m²/month) | Increased precipitation disturbs the water surface, promoting gas exchange and potentially flushing organic material into reservoirs, fuelling emissions (Deemer et al., 2016; Li et al., 2022). |
| Wind speed (m/s) | Higher wind speeds disturb the water surface, increasing turbulence and facilitating methane release into the atmosphere (Keller and Stallard, 1994). |
| Potential Evapotranspiration (kg/m²/month) | Higher evapotranspiration rates may increase water loss, which may concentrate nutrients and organic matter, which may increase CH ₄ concentration gradients and accelerate diffusion (Lee et al., 2018). |
| Vapour Pressure Deficit (Pa) | High VPD is often linked to warmer temperatures that enhance methanogenesis. Associated with drought-induced vegetation mortality, which contributes organic matter to reservoirs (Xu et al., 2023). |
| Humidity (%) | High humidity often co-occurs with accelerated decomposition of organic carbon (Waksman and Gerretsen, 1931) which may promote CH ₄ production. |
| Net Primary Productivity (gC/m²/yr) | Higher NPP supplies organic matter to sediments, enhancing methanogenesis. Oxygen production from photosynthesis may also regulate methane oxidation (Jager et al., 2022). |
| Discharge (m³/s) | Decreased discharge promotes sediment deposition, creating more drawdown areas that serve as hotspots of CH ₄ ebullition (Harrison et al., 2017). |
| Global Horizontal Irradiance (kWh/m²/month) | Solar radiation heats the water surface, stimulating microbial activity and CH ₄ production. It can also influence reservoir stratification (Harrison et al., 2021). |

2.3. Machine learning modelling

2.3.1. Data processing

The CH₄ emissions comprising 528 observations about CH₄ diffusion and 106 observations related to ebullition, were uniformly split into a 90 % training set and a 10 % testing set. To standardise feature scales and prevent any single variable from disproportionately influencing the model, we applied the Min-Max Scaler to the training dataset, scaling data between 0 and 1 using the minimum and maximum values of the observations. Subsequently, the resulting scaler from the training set was applied to the testing set. Given the logarithmic distribution of CH₄ emissions, a Log₁₀ transformation was used on each observation. The cross-validation technique was employed on the training set to assist in the hyperparameter tuning of the models. This was achieved through the implementation of a five-fold setup, which served to mitigate the issue of overfitting. Each fold serves as a validation set once across five rounds, with the average R² for each iteration providing an accurate estimate of model performance.

2.3.2. Hyperparameter tuning

To predict CH₄ diffusive and ebullitive emissions, five machine learning models were evaluated: Backpropagation Neural Networks (BPNNs), General Regression Neural Networks (GRNNs), K-Nearest Neighbors Regressor (KNN), Support Vector Regression (SVR), and Random Forest (RF) (Mitchell, 1997), all implemented using Scikit-learn (version 1.2.2) (Pedregosa et al., 2011). Model structures were optimised through hyperparameter tuning (detailed in Supplementary Material) to maximise predictive accuracy (Pedregosa et al., 2011).

2.3.3. Exhaustive feature selection

The datasets contained irrelevant or redundant features that did not significantly contribute to target variable prediction. These features can potentially introduce noise and degrade model performance. Furthermore, different models may perform variably depending on feature combinations. While linear regression often uses Pearson correlation to remove features with low correlation to target variables, this approach may overlook non-linear relationships between features and dependent variables. To address this limitation, an exhaustive wrapper feature selection method was employed, which is regarded as a brute force technique for feature selection.

Drawing on conclusions from prior training, this method generates all possible subsets, constructs a learning model for each, and chooses the one that yields the optimal performance. This study identified a total of 11 features, which are reservoir area, volume, temperature, precipitation, PET, VPD, NPP, humidity, wind speed, discharge, and GHI. The 11 features can generate a total of 2047 subset feature combinations. Each subset was tested to identify the optimal subset features for five machine learning algorithms with their best hyperparameter configurations based on the full feature set. The subset that had the highest R² was selected.

2.3.4. Feature importance

The significance of features in the selected model was estimated using Shapley values through the SHAP package (version 0.42.1) in Python (version 3.11.7). The SHAP library, developed by (Lundberg and Lee, 2017), calculates Shapley values to quantify each feature's contribution by measuring expected changes in the model's prediction when including or excluding the feature across all possible feature subsets (Molnar, 2022). Each input variable is assigned an individual Shapley value.

The KernelExplainer was used to calculate feature importance, with the training set first used to initialise the KernelExplainer to understand how the trained model functions and typical feature distributions. SHAP values were subsequently computed for the testing set, offering an accurate assessment of each feature's significance in unseen data. For each dataset in the testing set, the mean Shapley value was calculated by

averaging the SHAP values and ranking features by their mean Shapley values to determine their importance in the model.

2.3.5. Upscale study area and modelling dataset

The machine learning model was subsequently applied to upscale CH₄ diffusive and ebullitive emissions across reservoirs in China, where the total CH₄ flux rate is defined as the sum of the CH₄ diffusive and ebullitive emission rates. This analysis included 97,435 reservoirs from the CRD database (Song et al., 2022b), with data on reservoir area (km²), storage capacity (km³), and geographical locations. China's water systems are highly developed, encompassing numerous rivers. The Ministry of Water Resources and the National Bureau of Statistics have divided China into nine river basins based on the distribution of its major rivers, with river basin boundaries obtained from the Resource and Environment Science and Data Centre (<https://www.resdc.cn/>). Reservoirs are assigned to these river basins according to their geographic locations (Fig. 2).

2.4. Scale classification and data collection on reservoir usage

To analyse carbon emissions across reservoirs with varying storage capacities, reservoirs were grouped by size according to national classifications. While size categories differ internationally, this study applied the classification standards from the Ministry of Water Resources of China. Based on these guidelines, reservoirs were classified into five groups by storage capacity: <0.001 km³, 0.001–0.01 km³, 0.01–0.1 km³, 0.1–1 km³, and >1 km³. These categories were then consolidated into two main groups: reservoirs with a volume <0.01 km³ and reservoirs with a volume >0.01 km³.

To further examine emissions relative to reservoir purposes, usage data were sourced from the GRanD database and Global Dam Watch database (GDW), covering 671 reservoirs (Lehner et al., 2011; Lehner et al., 2024). In China, reservoir water flows are allocated with 10 % designed for ecological purposes, though only primary usage was considered here.

2.5. Data analysis

2.5.1. Statistical analysis

To assess the statistical significance of CH₄ emissions with reservoir size and usage, we used the *scipy.stats* and *scikit-posthoc* packages in Python (version 3.7). Levene's test was applied for homogeneity and the Shapiro-Wilks' test for normality. Given that the continuous variables did not meet the assumptions of normal distribution and homogeneity, the nonparametric Kruskal-Wallis (KW) test was employed to assess statistical significance across multiple data groups. Furthermore, the Dunn's test was employed for post-hoc multi-group comparisons.

2.5.2. Regression analysis

Regression analysis was applied to assess how reservoir size affects the dynamics of independent variables using the optimal machine learning model. Five representative volumes and areas were defined for each storage capacity category: 10 km³ and an area of 100 km² for reservoirs > 1 km³, 1 km³ and 10 km² for reservoirs between 1 and 0.1 km³, 0.1 km³ and 1 km² for size between 0.1 and 0.01 km³, 0.01 km³ and 0.1 km² for reservoirs between 0.01 and 0.001 km³, and 0.001 km³ and 0.01 km² for reservoirs < 0.001 km³. A total of 1000 samples were generated using the distribution of variables in the CRD dataset, replicating each sample set five times. Each group was assigned consistent values for either volume or reservoir area, and these variables were then put into the trained model to predict CH₄ flux rates.

3. Results

3.1. Model comparison

For CH₄ diffusive emissions, the SVR model achieved the highest R² of 0.66, with an RMSE of 0.425 (log₁₀ mg CH₄/m²/h) (Fig. 1). The optimal variable set for SVR includes volume, temperature, VPD, NPP, and GHI. Regarding CH₄ ebullitive emissions, both the BPNN and RF models demonstrate superior performance compared to the remaining three models, reaching an R² of 0.79 and 0.78, and both have RMSE of 0.33 (log₁₀ mg CH₄/m²/h) (cross-validation results see Figures S1-S10). However, the RF model requires only 3 input variables (i.e. reservoir

area, temperature, and discharge), which reduces the risk of overfitting and is more computationally efficient compared to BPNN, which requires 7 input features. Thus, in this study, the RF model was employed to estimate CH₄ ebullition (variable sets see Table S1).

3.2. Spatial distribution patterns of CH₄ emissions

Fig. 2 shows the estimated total emission rates (CH₄ diffusive and CH₄ ebullitive pathways) of 97,435 reservoirs in China, with CH₄ diffusive emissions estimated using the SVR model and ebullitive emissions estimated using the RF model. Reservoirs larger than 0.01 km³ are primarily located in the YRB, PRB and the NERB (full name see

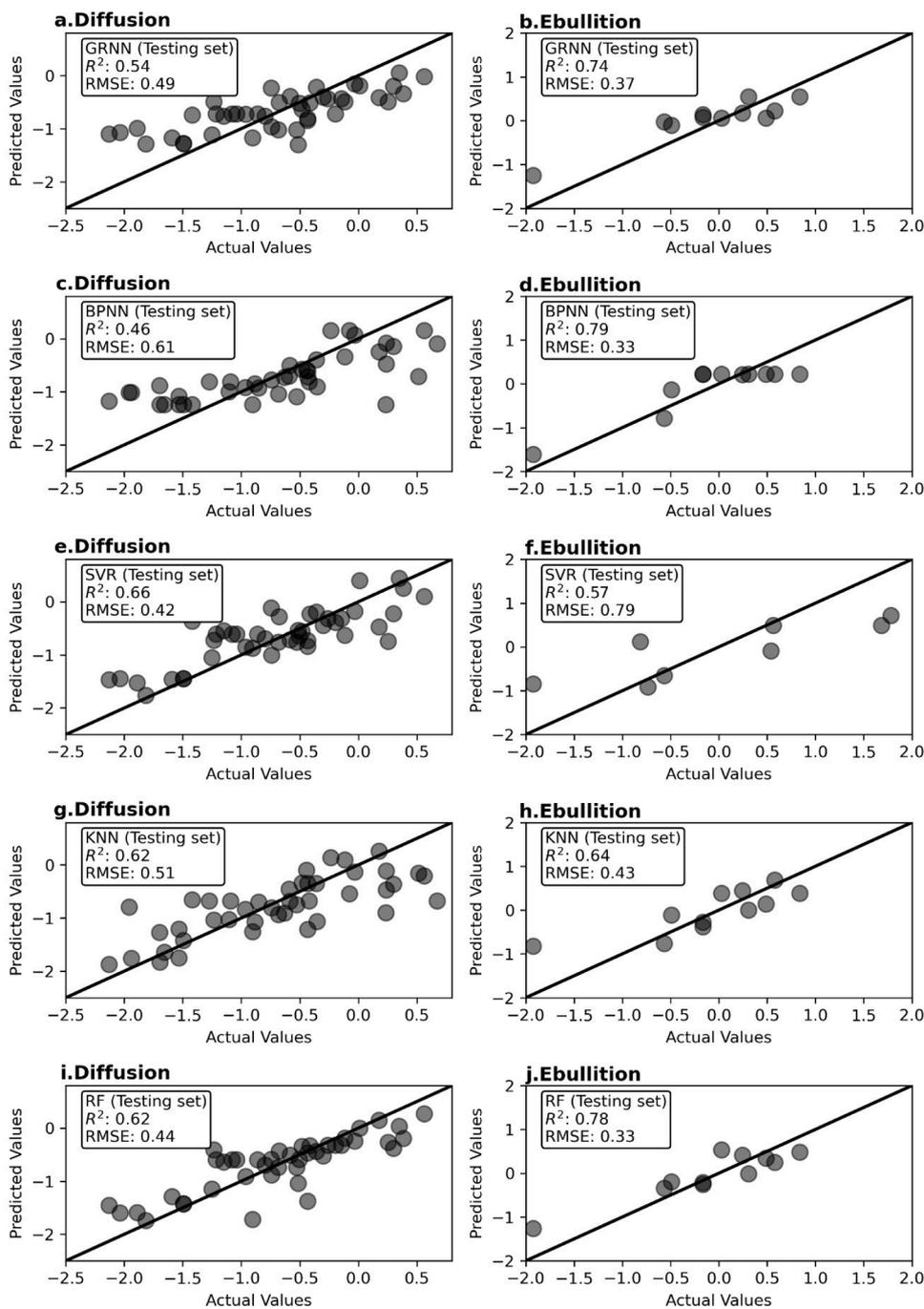


Fig. 1. Predictions of machine learning models versus observed testing data for CH₄ emissions: the left panel shows the diffusive CH₄ emission rate prediction, while the right panel shows the ebullitive CH₄ emission rate. Models in the panel a) and b) are GRNN, c) and d) are BPNN, e) and f) are SVR, g) and h) are KNN, and i) and j) are RF.

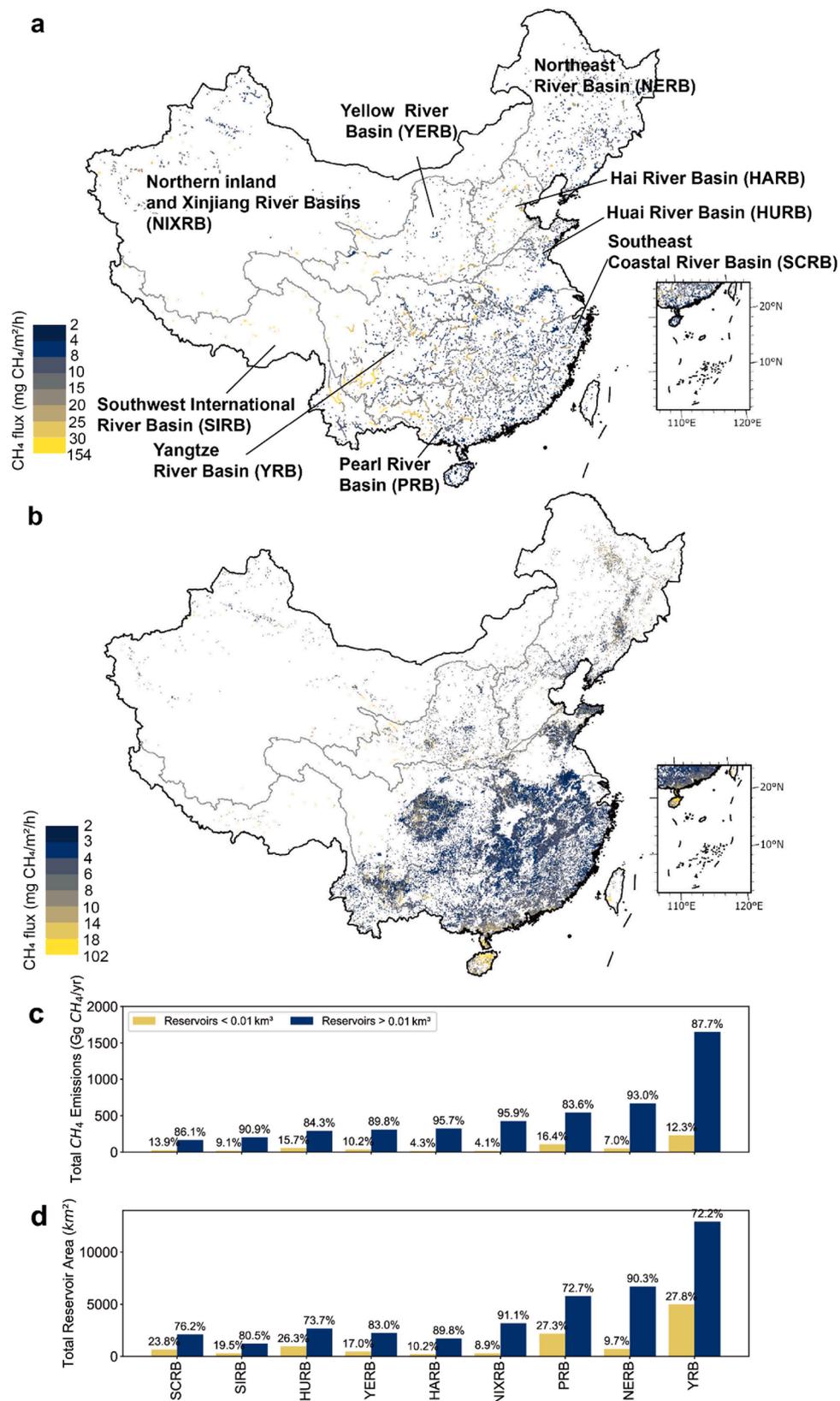


Fig. 2. Interpolated flux rates for (a) reservoirs > 0.01 km³, and (b) reservoirs < 0.01 km³. (c) Basin-level total CH₄ emissions and (d) total reservoir areas across nine major river basins in China, with annotations indicating each size category's contribution to the basin.

Fig. 2), regions rich in hydrological resources for hydropower development. The highest mean CH₄ flux rates for reservoirs > 0.01 km³ are observed in the SIRB, YRB, and HARB (Fig. 2). For reservoirs < 0.01 km³, they are clustered in the middle and downstream regions of YRB, PRB and SCRB, where they serve as critical infrastructure for irrigation water supply. The CH₄ flux rates of reservoirs < 0.01 km³ are more evenly distributed across basins, with a maximum mean rate observed in the NIXRB, and the minimum mean rate in the SCRB. But from the map, a high-rate cluster can be observed in the Hainan island located at the southernmost part of the PRB.

The maximum flux rate in reservoirs < 0.01 km³ is lower than in those > 0.01 km³, with an overall tendency towards lower emission rates. In terms of total CH₄ emissions, reservoirs > 0.01 km³ collectively contribute 4587 Gg CH₄ per year, while reservoirs < 0.01 km³ add 557 Gg CH₄ annually. Fig. 2c illustrates the basin-level total emissions from reservoirs of all sizes, with reservoirs > 0.01 km³ size accounting for an average of 89 % of total CH₄ emissions. In HARB and NIXRB, their contribution exceeds 95 % of total emissions, driven largely by their expansive reservoir areas, which constitute nearly 90 % of the total reservoir area (Fig. 2d).

3.3. Reservoir volume and CH₄ emission rates

Boxplots of the five volume-based categories reveal that reservoirs > 0.01 km³ tend to have higher estimated fluxes than smaller ones. This pattern is particularly evident in CH₄ diffusive emissions, where the interquartile range (IQR) progressively increases from the smallest category to the largest category (Fig. 3).

Among reservoir categories, reservoirs > 1 km³ had the highest mean CH₄ diffusive flux rate, followed by reservoirs size 0.1 to 1 km³. Reservoirs 0.001 to 0.01 km³ and < 0.001 km³ exhibited lower mean diffusive flux rates. The Dunn post hoc test confirms significant differences ($p < 0.01$) between reservoirs < 0.01 km³ and reservoirs > 0.01 km³, while no significant difference was found between reservoirs 0.1 to 1 km³ and reservoirs > 1 km³ ($p = 0.33$) (Fig. 3a).

CH₄ ebullition flux rates exhibit a less pronounced increasing trend across reservoir sizes (Fig. 3b). Reservoirs < 0.01 km³ showed higher IQR

than reservoirs 0.01 to 1 km³, and reservoirs > 1 km³ had the highest flux overall. Post hoc analysis revealed significant variations between reservoirs < 0.01 km³ and reservoirs > 0.01 km³ ($p < 0.01$) but no significant differences between reservoirs 0.01 to 0.1 km³ and reservoirs > 1 km³ ($p = 0.19$) (Fig. 3b).

3.4. Reservoir usage and CH₄ emission rates

Boxplots of CH₄ diffusive and ebullitive emission rates across four primary reservoir usage categories are shown in Fig. 4. Hydropower reservoirs generally show the highest CH₄ emissions, with water supply reservoirs having the lowest rates. Emission rates for both irrigation and flood control reservoirs are comparable across both diffusion and ebullition pathways. Statistical analysis indicates no significant difference in diffusive emissions rates between flood control and hydroelectric reservoirs, while other pairs showed significant differences ($p < 0.01$; see Fig. 4c). In contrast, for ebullitive emissions, statistical tests revealed notable differences solely between irrigation and water supply reservoirs, with other categories showing no significant relationships (Fig. 4f).

Fig. 4 illustrates the spatial distribution of reservoirs by scale and usage. Hydropower reservoirs, predominantly in southern China, exhibit the highest emission flux rates. Irrigation reservoirs are dispersed in clustered patterns, mainly along the middle and lower reaches of the Yangtze River, the lower sections of the Yellow River, and the Pearl River Basin. Flood control reservoirs are concentrated in the east, reflecting China's terrain, which slopes from high western plateaus and mountains to the lower plains and hills in the east and south. In contrast, water supply reservoirs, the least numerous of four types, form noticeable clusters around metropolitan areas such as Beijing-Tianjin in the north, Shanghai along the eastern coast, and Kunming in the southwestern. Additionally, clusters of water supply reservoirs are found near other cities.

The substantial variability in CH₄ diffusive emissions across reservoir types is related to their storage characteristics. Reservoirs designated for irrigation and water supply purposes are often constrained by geographical and environmental limitations, necessitating integration

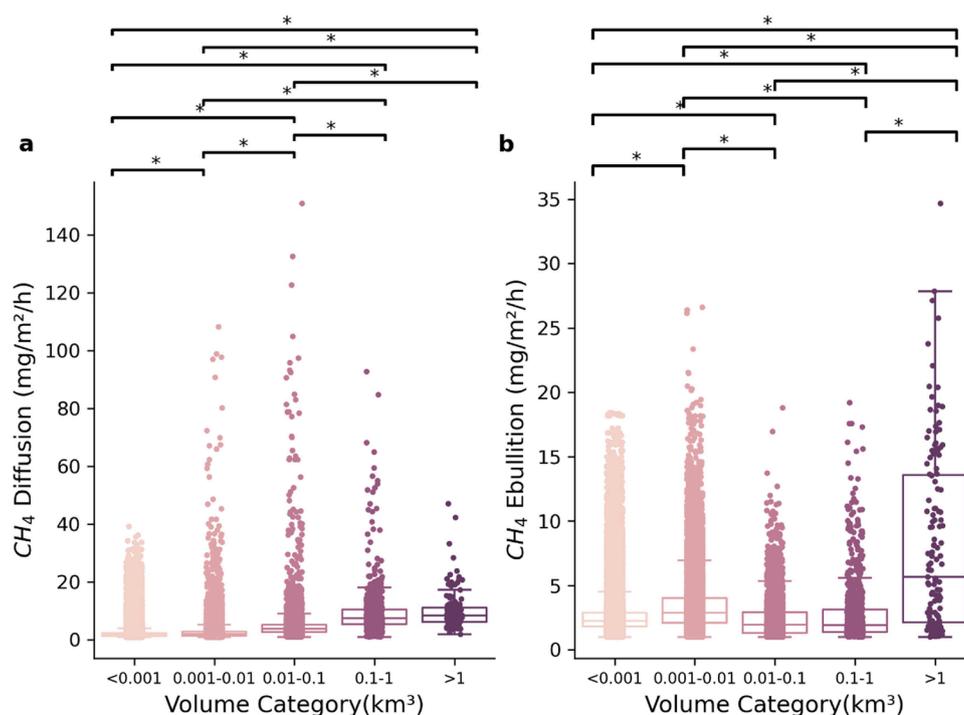


Fig. 3. Boxplots showing CH₄ diffusion (a) and ebullition (b) by reservoir size categories of estimated reservoirs in China.

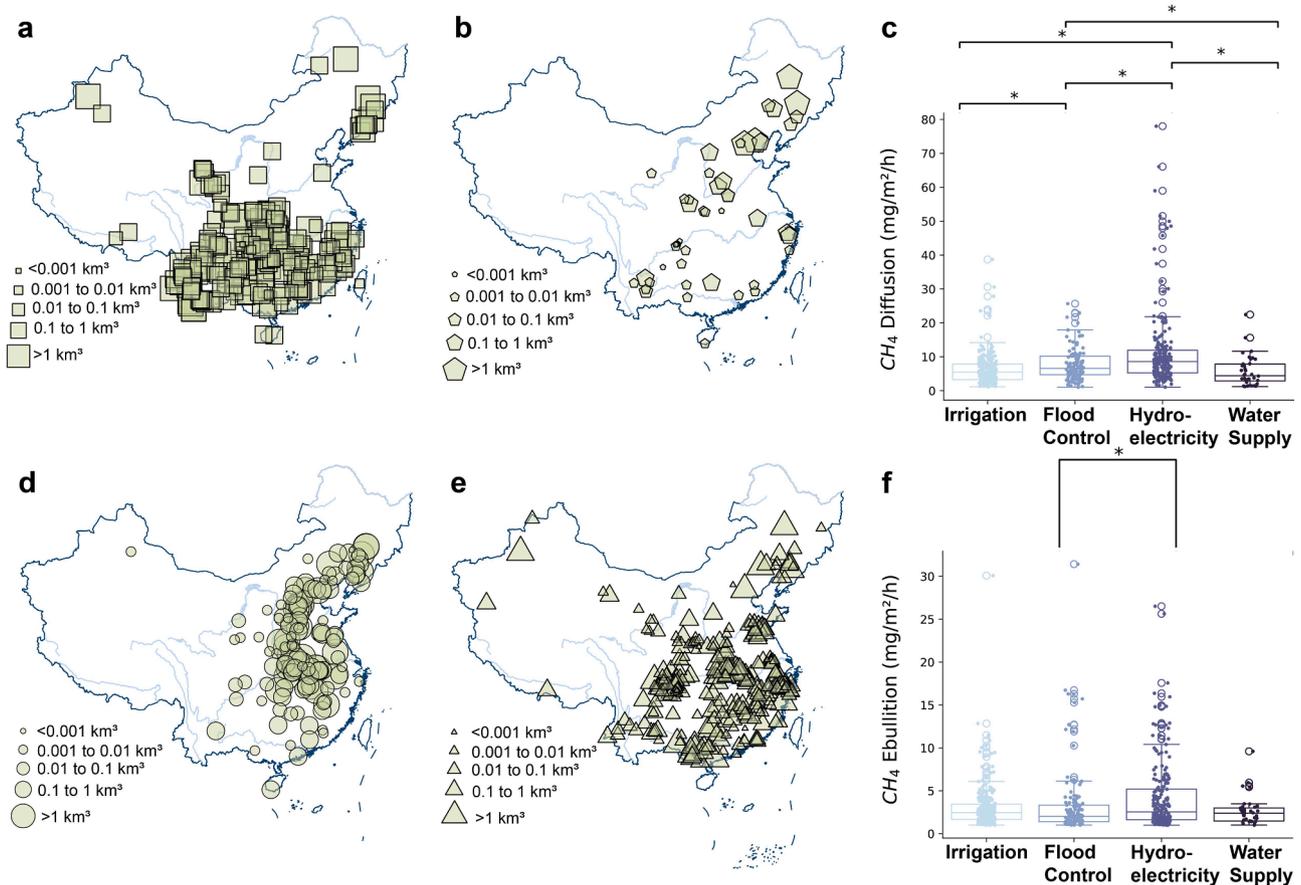


Fig. 4. Reservoirs of various sizes are categorised by primary use: a) hydropower; b) water supply; d) flood control; and e) irrigation. Boxplots c) and f) showing CH_4 diffusion (c) and ebullition (f) by reservoir usage type of estimated reservoirs in China.

into diverse landscapes with minimum ecological disruptions (Jurik et al., 2019). Smaller reservoirs are generally sufficient for irrigation and water supply, as these needs are seasonal and temporary, and also less expensive to construct and maintain (Wisser et al., 2010). Consequently, smaller reservoirs are commonly found in areas with extensive agriculture and human settlements (see Fig. 4), highlighting their importance in supporting urban and agricultural development (Grinham et al., 2018). Large reservoirs, however, are primarily situated in mountainous areas, such as the upper reaches of the Yangtze, Lancang, and Pearl Rivers, where significant elevation change facilitates their development

(Song et al., 2022a).

Flood control reservoirs, characterised by large storage capacities for managing excess water, are typically located in low-lying areas (see Fig. 4) with abundant vegetation, which decomposes underwater and contributes to organic matter accumulation. This could explain why CH_4 diffusive emissions from flood control reservoirs, as depicted in the boxplot, are comparable to those of hydropower reservoirs. However, ebullitive emission rates are notably lower in flood control reservoirs, possibly due to reduced water column disturbance by lower discharge, which inhibits CH_4 bubble release.

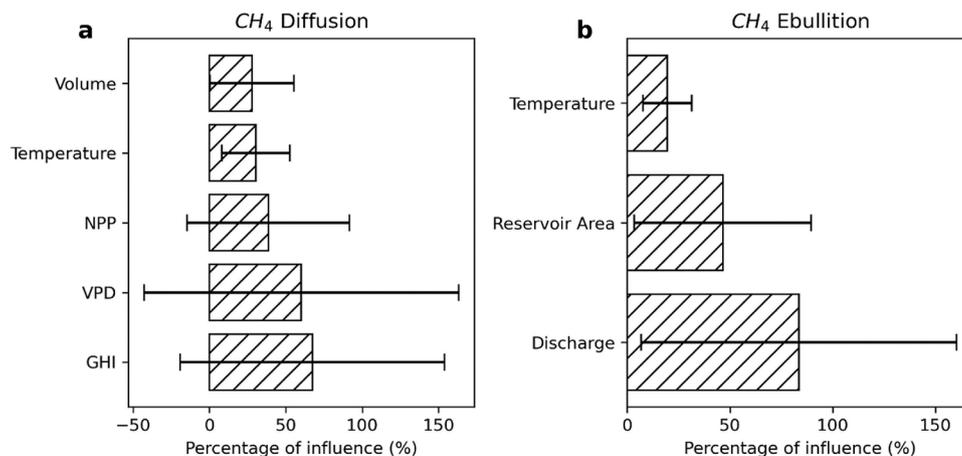


Fig. 5. Feature importance for the SVR model on the testing dataset for diffusive CH_4 emissions (plot a) and the RF model on the testing dataset for ebullitive CH_4 emissions (plot b).

3.5. Variables and carbon flux dynamics

This study examines a range of variables impacting carbon flux, including key climatic factors such as temperature and GHI, as well as reservoir area and volume. The feature importance for each prediction model is shown in Fig. 5, with Shapley values expressed as percentages, representing each feature’s contribution to the deviations from the mean prediction.

For CH₄ diffusive emissions, GHI is the most significant factor, with a mean Shapley value of 0.22, followed by VPD at 0.19 (see Figure S11). This indicates that GHI shifts predictions by approximately 0.22 units

from the dataset average. Given the log₁₀ transformation of CH₄ values for model compatibility, GHI affects the original prediction by around 1.659 times, or roughly 66%, while VPD contributes about 54%. Temperature and NPP rank third and fourth, each with Shapley values of approximately 0.14, indicating an influence of 36%. Volume has the lowest impact, influencing predictions by about 26%. For CH₄ ebullitive emissions, discharge has the highest influence at around 83%, followed by reservoir area at 46 % and temperature at 20%.

In addition to the commonly discussed predictor variables such as temperature, GHI, and reservoir area, this study introduces new variables. For example, VPD, the gap between saturated and actual vapour

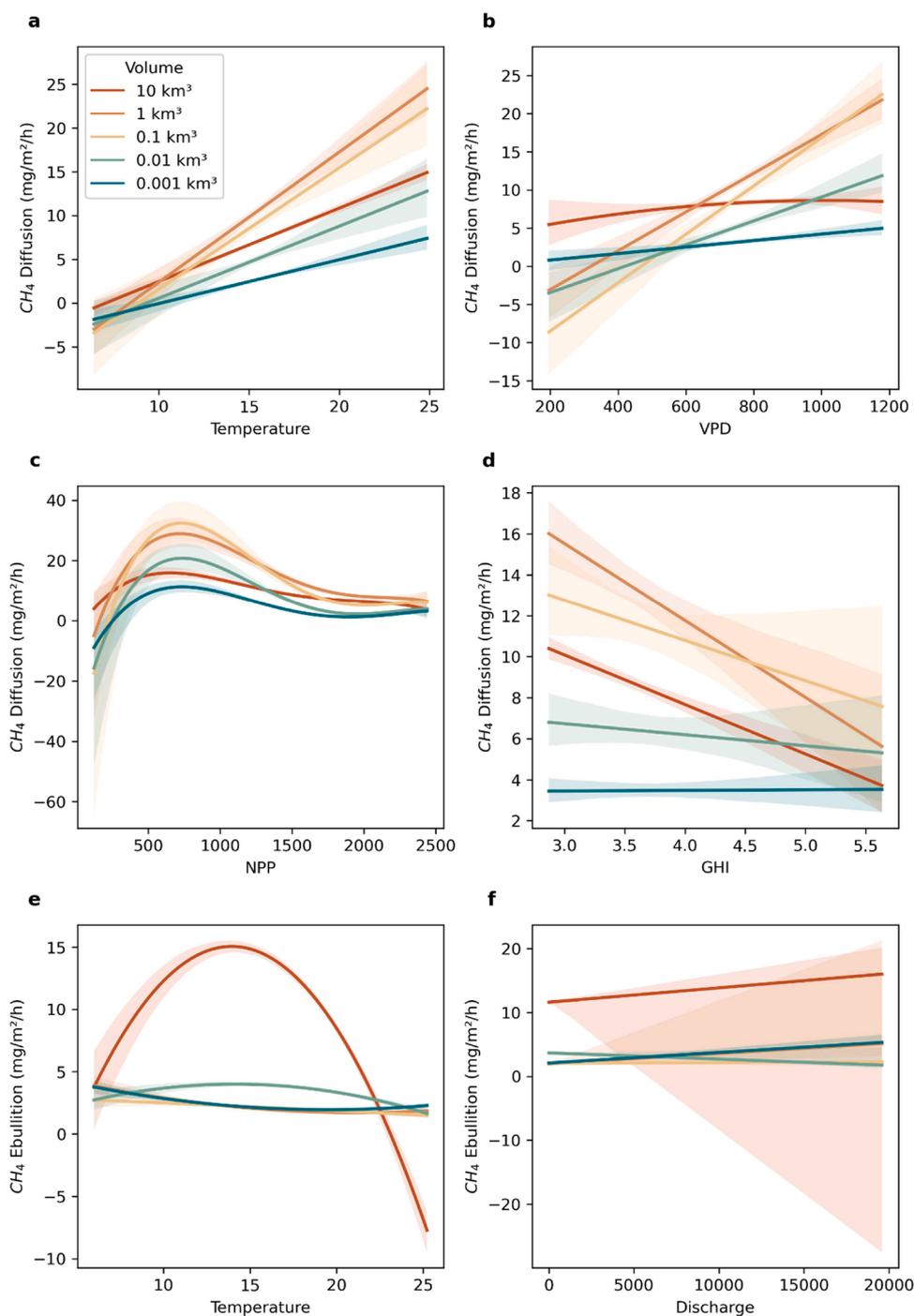


Fig. 6. Regression lines depicting relationships between independent variables and CH₄ fluxes for different reservoir sizes based on the generated 1000 samples based on the statistical distribution of variables in the CRD dataset. Figures (a-d) present diffusive CH₄ emissions, while (e-f) show ebullitive CH₄ emissions, with variables for (a) temperature, (b) VPD, (c) NPP, (d) GHI, (e) temperature, and (f) discharge.

pressures, determines the atmosphere's moisture absorption capacity (Zhong et al., 2023). A higher VPD correlates with increased CH₄ flux rates, as depicted in Fig. 6b, potentially due to warmer temperatures that often co-occur with high VPD (Amitrano et al., 2019). Elevated temperatures can enhance the metabolic rates of methanogenic bacteria in sediments, thereby increasing CH₄ diffusion into the atmosphere (Yvon-Durocher et al., 2014).

High VPD also intensifies evaporation, thinning the water layer and heightening the CH₄ concentration gradient, accelerating diffusion rates (Lee et al., 2018). High VPD levels are associated with drought-induced forest mortality (Yuan et al., 2019), where increased VPD prompts stomatal closure to mitigate water loss under high atmospheric water demand. This response negatively impacts carbon balance by depleting carbohydrate reserves, desiccating plant tissue, and causing plant death (Xu et al., 2023). As a result, elevated VPD can increase the mortality rate of plants and algae, contributing organic material to water bodies, which serves as a rich source for methanogenesis (Yuan et al., 2019).

Reservoir productivity, represented by NPP, is another key factor influencing carbon flux (Hertwich, 2013; Deemer et al., 2016; Chen et al., 2018). NPP representing the increase in dry organic matter in an ecosystem through photosynthesis CO₂ uptake, influences CH₄ emissions by producing organic matter that serves as a substrate for CH₄ production, and generating O₂ for CH₄ oxidation (Venkiteswaran et al., 2013).

Fig. 6c illustrates a non-linear relationship between NPP and CH₄ diffusion, with CH₄ flux rates rising at NPP values up to 1000 gC/m²/yr and subsequently declining. At lower NPP, organic matter supports CH₄ production (Deemer et al., 2016; Hertwich, 2013), though limited oxygen from low NPP constraints CH₄ oxidation. However, once NPP surpasses a threshold, increased oxygen availability enhances oxidation, reducing CH₄ emissions (Pacheco et al., 2014). This inverse trend of CH₄ emissions at high productivity levels has also been observed in other studies, where increased primary production in lentic systems can turn them from CO₂ sources to sinks.

Fig. 6 also illustrates the varying regression relationships between predictor variables and CH₄ flux rates across different reservoir sizes (Regression Coefficient in Tables S2 and S3). Reservoirs > 0.01 km³ exhibit a larger regression coefficient compared to reservoirs < 0.01 km³, suggesting that CH₄ emissions rates from larger reservoirs are more sensitive to changes in climatic variables. For example, in the case of temperature, in CH₄ diffusion, rising temperatures lead to an increased flux across all reservoir sizes, with reservoirs > 0.01 km³ experiencing a more pronounced increase compared to smaller ones. The sensitivity to temperature variation is also evident in CH₄ ebullition, where reservoirs at 10 km³ size exhibit greater fluctuations. Below 15 °C, rising temperatures are associated with an increase in emissions rates, whereas beyond 15 °C, further temperature increases are linked to a decrease in emissions rates.

4. Discussion

4.1. Mechanisms behind high CH₄ emissions rate of larger reservoirs

Two primary physical processes drive high CH₄ diffusive emission rates in larger reservoirs: sediment and nutrient trapping, and thermal stratification (Winton et al., 2019). The first mechanism involves dams effectively trap sediments (Winton et al., 2019), which contributes significantly to CH₄ emissions. Organic matter in sediment provides substrates and anoxic conditions ideal for methanogenesis (Maeck et al., 2013). Larger reservoirs, with greater storage capacities, tend to trap more sediment, creating substantial submerged organic carbon that promotes CH₄ production under anaerobic conditions.

The second mechanism, thermal stratification, is more pronounced in larger reservoirs due to their greater capacity for heat retention, resulting in longer stratification seasons compared to smaller reservoirs (Nhiwatiwa and Marshall, 2006). This prolonged stratification supports extended hypoxic conditions in the hypolimnion, favouring CH₄

production (Kalf, 2002). In contrast, smaller reservoirs often experience vertical mixing at night, oxygenating the water and reducing CH₄ production (Yang et al., 2018).

The greater slope in the regression line between climatic variables and CH₄ diffusive emissions rate in larger reservoirs, particularly for temperature-related variables may be attributed to the higher susceptibility of larger reservoirs to thermal stratification as temperatures rise. In contrast, for shallower reservoirs, the increase in temperature inhibits stratification, leading to uniform warming across the vertical profile (Lee et al., 2018). Prolonged thermal stratification in larger reservoirs creates conditions for CH₄ generation.

The elevated CH₄ ebullition observed in reservoirs < 0.01 km³ (compared to reservoirs 0.01 to 1 km³) may be attributed to their extended shallow areas or littoral zones. Larger reservoirs typically have deeper central areas, reducing the extent of littoral zones, while smaller reservoirs are shallower, leading to more extensive littoral areas. Seekell et al. (2021) analysed 54 Swedish lakes and demonstrated that lakes with lower mean-to-maximum depth ratios have relatively larger littoral areas than those with high ratios. High water depth also blocks the penetration of light, influencing the growth of vegetation and inhibiting the expansion of the littoral zone (Seekell et al., 2021).

The widespread shallower area promotes CH₄ ebullition due to lower hydrostatic pressure, allowing CH₄ bubbles to ascend in the water column and burst into the atmosphere (Beaulieu et al., 2018). In deeper water, a significant proportion of CH₄ in bubbles can dissolve before reaching the surface (McGinnis et al., 2006), contributing to higher CH₄ ebullition rates in smaller reservoirs.

However, the susceptibility to thermal stratification may also lead to intense fluctuations in CH₄ ebullition rates in larger reservoirs, as seen in Fig. 6c for reservoirs at 10 km³. Below 15 °C, rising temperatures promote methanogenesis. More CH₄ is produced which can easily escape into the atmosphere. However, when temperatures exceed 15 °C, CH₄ ebullition decreases as stratification inhibits the passage of CH₄ bubbles through the oxycline, preventing their release into the atmosphere. This constraint is further intensified at higher temperatures. Nevertheless, the accumulated CH₄ that failed to pass through the oxycline can still be released into the air through a spike of emissions during fall turnover when water mixing allows gas exchange from the hypolimnion to the epilimnion, or through degassing pathways (Guérin et al., 2016; Harrison et al., 2021). Although in other reservoir sizes temperature variations minimally impact CH₄ ebullition, Aben et al. (2017) indicate that increased water temperatures can enhance CH₄ ebullition rates in shallow freshwater systems—such as ponds, rivers, and lakes—where stratification is limited.

Looking forward, both reservoir sizes are expected to experience increased CH₄ emissions under a warming climate, albeit through different processes. In smaller reservoirs, intensified biological activity raises emissions, while in medium to large reservoirs, prolonged and more extensive thermal stratification facilitates methanogenesis and results in increased emissions.

4.2. Scale-specific emissions management

The debate over whether smaller or larger reservoirs offer superior solutions for water management remains unresolved. Larger dams are generally perceived to have greater environmental impacts due to their substantial carbon footprint, extensive inundation areas, deforestation, and sedimentation (Abbasi and Abbasi, 2011; Liang et al., 2021). Moreover, it was revealed that larger dams such as Three Gorges Dam (TGD) resulted in a 79 % decrease in CO₂ export and a 50 % reduction in CH₄ export to the sea, highlighting the 'large-dam effects' on both the upstream and downstream regions of the YRB (Ni et al., 2022). These dams intercept organic matter from flowing downstream, trapping it behind the reservoirs, which may promote CH₄ emissions from the reservoirs. Small dams are often accused of their cumulative effects on hydrology, biochemistry, and ecology (Habets et al., 2018; Donchyts

et al., 2022). Examining the benefits of smaller versus larger dams from different perspectives yields varying conclusions.

In terms of CH₄ emissions, larger dams are the dominant contributors across China, with higher CH₄ diffusion rates and greater national emissions due to the extensive surface areas of these reservoirs. Moreover, stratification in large reservoirs might intensify CH₄ production, increasing emissions rates. Although it was initially assumed that China's numerous reservoirs < 0.01 km³ would contribute higher total CH₄ emissions, their combined surface area is still smaller than that of reservoirs > 0.01 km³, resulting in a relatively lower contribution to total CH₄ emissions. Consequently, CH₄ management efforts should prioritise reservoirs > 0.01 km³.

Nevertheless, the boxplot (Fig. 3b) indicated that reservoirs < 0.01 km³ had higher ebullition rates than reservoirs 0.01 to 1 km³. This is likely attributed to their expansive littoral area and relatively lower depth. Moreover, the CH₄ ebullition rate data collected in our study may often be underestimated due to the challenges of capturing bubble fluxes over time (Maeck et al., 2014). CH₄ ebullition is often measured over short periods lasting only hours or days, whereas long-term continuous sampling would provide more representative estimates (Maeck et al., 2014). Thus, in certain regions, the potential for significant CH₄ ebullition from reservoirs < 0.01 km³ may be higher than previously expected.

Addressing reservoir carbon emissions by focusing solely on larger dams is insufficient. A sustainable water management strategy requires attention to both small and large reservoirs, with targeted measures based on their spatial distribution patterns. In Figs. 3 and 5, reservoirs < 0.01 km³ tend to cluster near farmland and urban areas, where high population densities drive water demand. Conversely, reservoirs > 0.01 km³ are more dispersed and located predominantly in mountainous regions. Such diverse distribution patterns necessitate different carbon reduction strategies: reservoirs > 0.01 km³ are best managed at the project level, while reservoirs < 0.01 km³ benefit from basin-scale management.

Larger reservoirs are typically managed by government agencies or large corporations (Donchyts et al., 2022), with the resources for carbon reduction measures, such as water quality control and methane recovery systems. CH₄ recovery involves installing CH₄ collection mechanisms at the reservoir's hypolimnion to capture CH₄ for power generation. These systems are more economically feasible for larger reservoirs given the substantial capital investment required (Wood et al., 2023). In contrast, for smaller reservoirs, the cost-benefit ratio is less favourable (Wood et al., 2023).

Similar to water quality control, larger reservoirs also benefit from consistent monitoring, which is less feasible for smaller reservoirs due to their large numbers and local ownership, making individualized monitoring logistically and financially challenging. Therefore, while project-scale management is suitable for larger reservoirs, basin-scale strategies are more suitable for particularly managing smaller reservoirs.

Basin-scale management could involve regulatory measures to restrict nutrient discharge from wastewater, thereby preventing water quality deterioration at the source (Preisner et al., 2020). Collaborating with local farmers to implement best management practices at appropriate times and locations, supported by policies and incentives, is another effective approach (Bijay and Craswell, 2021). Additionally, public education campaigns aimed to raise awareness about environmentally friendly practices could improve water resource use in agriculture (Zhang et al., 2023).

4.3. Limitations and opportunities from machine learning models

In this study, machine learning was applied to estimate CH₄ emissions from Chinese reservoirs. Unlike commonly used regression models, which provide interpretable relationships between variables and emissions rates, the black-box nature of machine learning models limits their interpretability. To investigate how the trained machine learning model

understands the relationship between input variables and emissions rates, two techniques were employed: Shapley values and regression lines to visualize the trained relationships. However, these techniques are limited as they can only provide a vague demonstration of the potential statistical relationships, rather than offering specific mathematical equations. Furthermore, the learned relationship between input variables and emissions rates in machine learning models may involve high dimensionality, which current two- and three-dimensional techniques cannot fully capture. Therefore, to enhance the interpretability of machine learning models, more innovative expression techniques must be explored to better leverage the potential of machine learning in understanding the underlying mechanisms in environmental research.

5. Conclusion

- Reservoirs > 0.01 km³ contribute nearly 90 % of total CH₄ emissions due to their significant CH₄ diffusive flux rates, which are attributed to their thermal stratification regime high capacity for receiving organic matter, and expansive surface areas.
- Reservoirs < 0.01 km³ exhibited higher CH₄ ebullitive flux compared to reservoirs between 0.01–1 km³, owing to their shallow depth, emphasizing the importance of addressing CH₄ reduction in smaller aquatic systems.
- Hydroelectric and flood control reservoirs tend to have higher CH₄ diffusive emission rates, which is related to reservoir size and usage: reservoirs < 0.01 km³ primarily serve irrigation and water supply, while reservoirs > 0.01 km³ are designated for hydropower and flood control. Reservoirs < 0.01 km³ are often clustered near agricultural or urban regions, whereas reservoirs > 0.01 km³ are more dispersed and located in mountainous or low-lying areas.
- Given the spatial variability of reservoir sizes, carbon reduction efforts require tailored approaches: project-level management for reservoirs > 0.01 km³ and basin-scale management for reservoirs < 0.01 km³.

CRedit authorship contribution statement

Zilin Wang: Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Meili Feng:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Matthew F. Johnson:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Aldo Lipani:** Writing – review & editing, Validation, Software, Methodology. **Faith Chan:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study is funded by the National of Science Foundation of China (W2432029; 32161143025; 51909126; 41850410497); The Zhejiang Natural Science Foundation of China, The Zhejiang Basic and Commonweal Programme (grant number ZJWY23E0900005; LY22E080018); The Ningbo Natural Science Foundation (2023J193); The Mongolian Foundation for Science and Technology (NSFC_202201, CHN2022/276); The National Key R&D Program of China (2022YFE0119200); The Key Project of Innovation LREIS (KPI006); The National University of Mongolia (P2023–4429, P2022–4256); The Hong Kong Environmental Council Fund (ECF: 44/2020).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2025.123441.

Data availability

Data will be made available on request. Code script at https://github.com/zy22159-tech/machine_learning_GHGemissions_China_reservoirs.

References

- Abbasi, T., Abbasi, S.A., 2011. Small hydro and the environmental implications of its extensive utilization. *Renew. Sustain. Energy Rev.* 15 (4), 2134–2143.
- Aben, R.C.H., Barros, N., van Donk, E., Frenken, T., Hilt, S., Kazanjian, G., Lamers, L.P. M., Peeters, E.T.H.M., Roelofs, J.G.M., de Senerpont Domis, L.N., Stephan, S., Velthuis, M., Van de Waal, D.B., Wik, M., Thornton, B.F., Wilkinson, J., DelSontro, T., Kosten, S., 2017. Cross continental increase in methane ebullition under climate change. *Nat. Commun.* 8 (1), 1682.
- Amitrano, C., Arena, C., Roupael, Y., De Pascale, S., De Micco, V., 2019. Vapour pressure deficit: the hidden driver behind plant morphofunctional traits in controlled environments. *Ann. Appl. Biol.* 175 (3), 313–325.
- Barros, N., Cole, J.J., Tranvik, L.J., Prairie, Y.T., Bastviken, D., Huszar, V.L.M., del Giorgio, P., Roland, F., 2011. Carbon emission from hydroelectric reservoirs linked to reservoir age and latitude. *Nat. Geosci.* 4 (9), 593–596.
- Bastviken, D., Cole, J.J., Pace, M., Tranvik, L., 2004. Methane emissions from lakes: dependence of lake characteristics, two regional assessments, and a global estimate. *Glo. Biogeochem. Cycles* 18 (4).
- Beaulieu, J.J., Balz, D.A., Birchfield, M.K., Harrison, J.A., Nietch, C.T., Platz, M.C., Squier, W.C., Waldo, S., Walker, J.T., White, K.M., Young, J.L., 2018. Effects of an experimental water-level drawdown on methane emissions from a eutrophic reservoir. *Ecosystems* 21 (4), 657–674.
- Bijay, S., Craswell, E., 2021. Fertilizers and nitrate pollution of surface and ground water: an increasingly pervasive global problem. *SN. Appl. Sci.* 3 (4), 518.
- Bosmans, J., Wanders, N., Bierkens, M.F.P., Huijbregts, M.A.J., Schipper, A.M., Barbarossa, V., 2022. FutureStreams, a global dataset of future streamflow and water temperature. *Sci. Data* 9 (1), 307.
- Brun, P., Zimmermann, N.E., Hari, C., Pellissier, L. and Karger, D.N. (2022) 'CHELSA-BIOCLIM+ A novel set of global climate-related predictors at kilometre-resolution'. Available at: https://www.envdat.com/dataset/bioclim_plus (Accessed).
- Chen, Z., Ye, X., Huang, P., 2018. Estimating Carbon Dioxide (CO₂) emissions from reservoirs using artificial neural networks. *Water* 10 (1), 26.
- Deemer, B.R., Harrison, J.A., Li, S., Beaulieu, J.J., DelSontro, T., Barros, N., Bezerra-Neto, J.F., Powers, S.M., Dos Santos, M.A., Vonk, J.A., 2016. Greenhouse gas emissions from reservoir water surfaces: a new global synthesis. *Bioscience* 66 (11), 949–964.
- DelSontro, T., Beaulieu, J.J., Downing, J.A., 2018. Greenhouse gas emissions from lakes and impoundments: upscaling in the face of global change. *Limnol. Oceanogr. Lett.* 3 (3), 64–75.
- Donchys, G., Winsemius, H., Baart, F., Dahm, R., Schellekens, J., Gorelick, N., Iceland, C., Schmeier, S., 2022. High-resolution surface water dynamics in Earth's small and medium-sized reservoirs. *Sci. Rep.* 12 (1), 13776.
- Grinham, A., Albert, S., Deering, N., Dunbabin, M., Bastviken, D., Sherman, B., Lovelock, C.E., Evans, C.D., 2018. The importance of small artificial water bodies as sources of methane emissions in Queensland, Australia. *Hydro. Earth Syst. Sci.* 22 (10), 5281–5298.
- Guérin, F., Deshmukh, C., Labat, D., Pighini, S., Vongkhamsoo, A., Guédant, P., Rode, W., Godon, A., Chanudet, V., Descloux, S., Serça, D., 2016. Effect of sporadic destratification, seasonal overturn, and artificial mixing on CH₄ emissions from a subtropical hydroelectric reservoir. *Biogeochemistry* 13 (12), 3647–3663.
- Habets, F., Molénat, J., Carluet, N., Douez, O., Leenhardt, D., 2018. The cumulative impacts of small reservoirs on hydrology: a review. *Sci. Total Environ.* 643, 850–867.
- Harrison, J.A., Deemer, B.R., Birchfield, M.K., O'Malley, M.T., 2017. Reservoir water-level drawdowns accelerate and amplify methane emission. *Environ. Sci. Technol.* 51 (3), 1267–1277.
- Harrison, J.A., Prairie, Y.T., Mercier-Blais, S., Soued, C., 2021. Year-2020 global distribution and pathways of reservoir methane and carbon dioxide emissions according to the greenhouse gas from reservoirs (G-res) model. *Glo. Biogeochem. Cycles* 35 (6), e2020GB006888.
- Hertwich, E.G., 2013. Addressing biogenic greenhouse gas emissions from hydropower in LCA. *Environ. Sci. Technol.* 47 (17), 9604–9611.
- Holgerson, M.A., Raymond, P.A., 2016. Large contribution to inland water CO₂ and CH₄ emissions from very small ponds. *Nat. Geosci.* 9 (3), 222–226.
- Jager, H.I., Griffiths, N.A., Hansen, C.H., King, A.W., Matson, P.G., Singh, D., Pilla, R.M., 2022. Getting lost tracking the carbon footprint of hydropower. *Renew. Sustain. Energy Rev.* 162, 112408.
- Jurik, L., Zelenáková, M., Kaletová, T., Arifjanov, A., 2019. Small water reservoirs: sources of water for irrigation. In: Negm, A.M., Zelenáková, M. (Eds.), *Water Resources in Slovakia: Part I: Assessment and Development*. Springer International Publishing, Cham, pp. 115–131.
- Kalff, J., 2002. *Limnology: inland water ecosystems*. Prentice Hall New Jersey.
- Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* 4 (1), 170122.
- Keller, M., Stallard, R.F., 1994. Methane emission by bubbling from Gatun Lake, Panama. *J. Geophys. Res.: Atmos.* 99 (D4), 8307–8319.
- Lee, R.M., Biggs, T.W., Fang, X., 2018. Thermal and hydrodynamic changes under a warmer climate in a variably stratified hypereutrophic reservoir. *Water*.
- Lehner, B., Beames, P., Mulligan, M., Zarfl, C., De Felice, L., van Soesbergen, A., Thieme, M., Garcia de Leaniz, C., Anand, M., Belletti, B., Brauman, K.A., Januchowski-Hartley, S.R., Lyon, K., Mandl, L., Mazany-Wright, N., Messenger, M.L., Pavelsky, T., Pekel, J.-F., Wang, J., Wen, Q., Wishart, M., Xing, T., Yang, X., Higgins, J., 2024. The global dam watch database of river barrier and reservoir information for large-scale applications. *Sci. Data* 11 (1), 1069.
- Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J.C., Rödel, R., Sindorf, N., Wisser, D., 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Front. Ecol. Environ.* 9 (9), 494–502.
- Lessmann, O., Encinas Fernández, J., Martínez-Cruz, K., Peeters, F., 2023. Methane emissions due to reservoir flushing: a significant emission pathway? *EGU sphere* 2023, 1–18.
- Li, S., Bush, R.T., Santos, I.R., Zhang, Q., Song, K., Mao, R., Wen, Z., Lu, X.X., 2018. Large greenhouse gases emissions from China's lakes and reservoirs. *Water. Res.* 147, 13–24.
- Li, Y., Zhou, Y., Zhou, L., Zhang, Y., Xu, H., Jang, K.-S., Kothawala, D.N., Spencer, R.G. M., Jeppesen, E., Brookes, J.D., Davidson, T.A., Wu, F., 2022. Changes in water chemistry associated with rainstorm events increase carbon emissions from the inflowing river mouth of a major drinking water reservoir. *Environ. Sci. Technol.* 56 (22), 16494–16505.
- Liang, X., Ye, Z., Feng, J., Li, H., Wang, Y., 2021. Sediment retention effect of reservoirs and microscale surface depressions. *Water Conserv. Sci. Eng.* 6 (3), 153–162.
- Lundberg, S.M. and Lee, S.-I. 'A unified approach to interpreting model predictions'. *Neural Inf. Process. Syst.*
- Maack, A., DelSontro, T., McGinnis, D.F., Fischer, H., Flury, S., Schmidt, M., Fietzek, P., Lorke, A., 2013. Sediment Trapping by dams creates methane emission hot spots. *Environ. Sci. Technol.* 47 (15), 8130–8137.
- Maack, A., Hofmann, H., Lorke, A., 2014. Pumping methane out of aquatic sediments - ebullition forcing mechanisms in an impounded river. *Biogeosciences* 11, 2925–2938.
- McGinnis, D.F., Greinert, J., Artemov, Y., Beaubien, S.E., Wüest, A., 2006. Fate of rising methane bubbles in stratified waters: How much methane reaches the atmosphere? *J. Geophys. Res.: Oceans* 111 (C9).
- Ministry of Water Resources, 2022. *Statistic Bulletin on China Water Activities*. Ministry of Water Resources of the People's Republic of China. Available at: http://www.mwr.gov.cn/sj/tjgb/slfztjgb/202312/t20231221_1698710.html (Accessed: 1/31 2024).
- Mitchell, T.M., 1997. *Machine Learning*. McGraw-Hill International Editions, McGraw-Hill.
- Molnar, C., 2022. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed.
- NASA, 2008. *Solar: Average Monthly and Annual Global Horizontal Irradiance Data, One-Degree Resolution of the World from NASA/SSE, 1983-2005*. Available at: <https://earthworks.stanford.edu/catalog/stanford-xx487wn6207>.
- Nhiwatiwa, T., Marshall, B.E., 2006. Seasonal and diurnal stratification in two small Zimbabwean reservoirs. *Afr. J. Aquat. Sci.* 31 (2), 185–196.
- Ni, J., Wang, H., Ma, T., Huang, R., Ciais, P., Li, Z., Yue, Y., Chen, J., Li, B., Wang, Y., Zheng, M., Wang, T., Borthwick, A.G.L., 2022. Three Gorges Dam: friend or foe of riverine greenhouse gases? *Natl. Sci. Rev.* 9 (6), nwac013.
- Pacheco, F.S., Roland, F., Downing, J.A., 2014. 'Eutrophication reverses whole-lake carbon budgets. *Inland. Waters.* 4 (1), 41–48.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Prairie, Y.T., Mercier-Blais, S., Harrison, J.A., Soued, C., Giorgio, P.d., Harby, A., Alm, J., Chanudet, V., Nahas, R., 2021. A new modelling framework to assess biogenic GHG emissions from reservoirs: the G-res tool. *Environ. Model. Softw.* 143, 105117.
- Preisner, M., Neverova-Dziopak, E., Kowalewski, Z., 2020. An analytical review of different approaches to wastewater discharge standards with particular emphasis on nutrients. *Environ. Manage* 66 (4), 694–708.
- Rocher-Ros, G., Stanley, E.H., Loken, L.C., Gasson, N.J., Raymond, P.A., Liu, S., Amatulli, G., Sponseller, R.A., 2023. Global methane emissions from rivers and streams. *Nature* 621 (7979), 530–535.
- Saunio, M., Martinez, A., Poulter, B., Zhang, Z., Raymond, P., Regnier, P., Canadell, J.G., Jackson, R.B., Patra, P.K., Bousquet, P., Ciais, P., Dlugokencky, E.J., Lan, X., Allen, G.H., Bastviken, D., Beerling, D.J., Belikov, D.A., Blake, D.R., Castaldi, S., Crippa, M., Deemer, B.R., Dennison, F., Etiope, G., Gedney, N., Höglund-Isaksson, L., Holgerson, M.A., Hopcroft, P.O., Huguelius, G., Ito, A., Jain, A.K., Janardan, R., Johnson, M.S., Kleinen, T., Krummel, P., Lauerwald, R., Li, T., Liu, X., McDonald, K. C., Melton, J.R., Mühle, J., Müller, J., Murguía-Flores, F., Niwa, Y., Noce, S., Pan, S., Parker, R.J., Peng, C., Ramonet, M., Riley, W.J., Rocher-Ros, G., Rosentretter, J.A., Sasakawa, M., Segers, A., Smith, S.J., Stanley, E.H., Thanwerdas, J., Tian, H., Tsuruta, A., Tubiello, F.N., Weber, T.S., van der Werf, G., Worthy, D.E., Xi, Y., Yoshida, Y., Zhang, W., Zheng, B., Zhu, Q., Zhu, Q., Zhuang, Q., 2024. Global methane budget 2000-2020. *Earth Syst. Sci. Data Discuss.* 2024, 1–147.
- Seekell, D., Cael, B., Norman, S., Byström, P., 2021. Patterns and variation of littoral habitat size among lakes. *Geophys. Res. Lett.* 48 (20), e2021GL095046.
- Sobek, S., DelSontro, T., Wongfun, N., Wehrli, B., 2012. Extreme organic carbon burial fuels intense methane bubbling in a temperate reservoir. *Geophys. Res. Lett.* 39 (1).

- Song, C., Fan, C., Zhu, J., Wang, J., Sheng, Y., Liu, K., Chen, T., Zhan, P., Luo, S., Yuan, C., Ke, L., 2022a. A comprehensive geospatial database of nearly 100 000 reservoirs in China. *Earth Syst. Sci. Data* 14 (9), 4017–4034.
- Song, C., Fan, C., Zhu, J., Wang, J., Sheng, Y., Liu, K., Chen, T., Zhan, P., Luo, S., Yuan, C., Ke, L., 2022b. A comprehensive geospatial database of nearly 100000 reservoirs in China. *Earth Syst. Sci. Data* 14 (9), 4017–4034.
- Venkiteswaran, J.J., Schiff, S.L., Louis, St., V. L., Matthews, C.J.D., Boudreau, N.M., Joyce, E.M., Beaty, K.G., Bodaly, R.A., 2013. Processes affecting greenhouse gas production in experimental boreal reservoirs. *Glob. Biogeochem. Cycles* 27 (2), 567–577.
- Waksman, S.A., Gerretsen, F.C., 1931. Influence of temperature and moisture upon the nature and extent of decomposition of plant residues by microorganisms. *Ecology* 12 (1), 33–60.
- Wang, Z., Chan, F.K.S., Feng, M., Johnson, M.F., 2024. Greenhouse gas emissions from hydropower reservoirs: emission processes and management approaches. *Environ. Res. Lett.* 19 (7), 073002.
- West, W.E., Creamer, K.P., Jones, S.E., 2016. Productivity and depth regulate lake contributions to atmospheric methane. *Limnol. Oceanogr.* 61 (S1), S51–S61.
- Winton, R.S., Calamita, E., Wehrli, B., 2019. Reviews and syntheses: dams, water quality and tropical reservoir stratification. *Biogeosciences* 16 (8), 1657–1671.
- Wisser, D., Frolking, S., Douglas, E.M., Fekete, B.M., Schumann, A.H., Vörösmarty, C.J., 2010. The significance of local water resources captured in small reservoirs for crop production – a global-scale analysis'. *J. Hydrol. (Amst)* 384 (3), 264–275.
- Wood, T.K., Gurgan, I., Howley, E.T., Riedel-Kruse, I.H., 2023. Converting methane into electricity and higher-value chemicals at scale via anaerobic microbial fuel cells. *Renew. Sustain. Energy Rev.s* 188, 113749.
- Xu, S., Gentine, P., Li, L., Wang, L., Yu, Z., Dong, N., Ju, Q., Zhang, Y., 2023. Response of ecosystem productivity to high vapor pressure deficit and low soil moisture: lessons learned from the global eddy-covariance observations. *Earths Future* 11 (8), e2022EF003252.
- Yang, Y., Wang, Y., Zhang, Z., Wang, W., Ren, X., Gao, Y., Liu, S., Lee, X., 2018. Diurnal and seasonal variations of thermal stratification and vertical mixing in a shallow fresh water lake. *J. Meteorol. Res.* 32 (2), 219–232.
- Yuan, W., Zheng, Y., Piao, S., Ciaia, P., Lombardozzi, D., Wang, Y., Ryu, Y., Chen, G., Dong, W., Hu, Z., Jain, A.K., Jiang, C., Kato, E., Li, S., Lienert, S., Liu, S., Nabel, J.E. M.S., Qin, Z., Quine, T., Sitch, S., Smith, W.K., Wang, F., Wu, C., Xiao, Z., Yang, S., 2019. Increased atmospheric vapor pressure deficit reduces global vegetation growth. *Sci. Adv.* 5 (8), eaax1396.
- Yvon-Durocher, G., Allen, A.P., Bastviken, D., Conrad, R., Gudas, C., St-Pierre, A., Thanh-Duc, N., del Giorgio, P.A., 2014. Methane fluxes show consistent temperature dependence across microbial to ecosystem scales. *Nature* 507 (7493), 488–491.
- Zhang, M., Qin, J., Tan, H., Mao, H., Tu, X. and Jian, J. (2023) 'Education level of farmers, market-oriented reforms, and the utilization efficiency of agricultural water resources in China', pp. 1–21.
- Zheng, Y., Wu, S., Xiao, S., Yu, K., Fang, X., Xia, L., Wang, J., Liu, S., Freeman, C., Zou, J., 2022. Global methane and nitrous oxide emissions from inland waters and estuaries. *Glob. Chang. Biol.* 28 (15), 4713–4725.
- Zhong, S., Zhang, K., Bagheri, M., Burken, J.G., Gu, A., Li, B., Ma, X., Marrone, B.L., Ren, Z.J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B.M., Xiao, X., Yu, X., Zhu, J.-J., Zhang, H., 2021. Machine learning: new ideas and tools in environmental science and engineering. *Environ. Sci. Technol.* 55 (19), 12741–12754.
- Zhong, Z., He, B., Wang, Y.-P., Chen, H.W., Chen, D., Fu, Y.H., Chen, Y., Guo, L., Deng, Y., Huang, L., Yuan, W., Hao, X., Tang, R., Liu, H., Sun, L., Xie, X., Zhang, Y., 2023. Disentangling the effects of vapor pressure deficit on northern terrestrial vegetation productivity. *Sci. Adv.* 9 (32), eadf3166.