


Article

Spatiotemporal Modeling of Carbon Fluxes over Complex Underlying Surfaces along the North Shore of Hangzhou Bay

Kaidi Zhang ^{1,2,3,*} , Min Zhao ^{1,4}, Zhenyu Zhao ^{1,4}, Xucheng Shen ^{1,4}, Yanyu Lu ^{2,3} and Jun Gao ^{1,4,*}

¹ School of Environmental and Geographical Sciences, Shanghai Normal University, Shanghai 200234, China; zhaomin@shnu.edu.cn (M.Z.); 1000495335@smail.shnu.edu.cn (Z.Z.); 1000557247@smail.shnu.edu.cn (X.S.)

² Anhui Province Key Laboratory of Atmospheric Science and Satellite Remote Sensing, Anhui Institute of Meteorological Sciences, Hefei 230031, China

³ Shouxian National Climatology Observatory, Huaihe River Basin Typical Farm Eco-Meteorological Experiment Field of CMA, Shouxian 232200, China

⁴ Yangtze River Delta Urban Wetland Ecosystem National Field Scientific Observation and Research Station, Shanghai 200234, China

* Correspondence: zhangkaidi@ahmi.org.cn (K.Z.); gaojun@shnu.edu.cn (J.G.)

Abstract: Urban areas contribute to over 80% of carbon dioxide emissions, and considerable efforts are being undertaken to characterize spatiotemporal variations of CO₂ (carbon dioxide) at a city, regional, and national level, aiming at providing pipelines for carbon mission reduction. The complex underlying surface composition of urban areas makes process-based and physiology-based models inadequate for simulating carbon flux in this context. In this study, long short-term memory (LSTM), support vector machine (SVM), random forest (RF), and artificial neural network (ANN) were employed to develop and investigate their viability in estimating carbon flux at the ecosystem level. All the data used in our study were derived from the long-term chronosequence observations collected from the flux towers within urban complex underlying surface, along with meteorological reanalysis datasets. To assess the generalization ability of these models, the following statistical metrics were utilized: coefficient of determination (R²), root mean square error (RMSE), and mean absolute error (MAE). Our analysis revealed that the RF model performed the best in simulating carbon flux over long time series, with the highest R² values reaching up to 0.852, and exhibiting the smallest RMSE and MAE values at 0.293 μmol·m⁻²·s⁻¹ and 0.157 μmol·m⁻²·s⁻¹. As a result, the RF model was chosen for simulating carbon flux at spatial scale and assessing the impact of urban impervious surfaces in the simulation. The results showed that the RF model performs well in simulating carbon flux at the spatial scale. The input of impervious surface area index can improve the performance of the RF model in simulating carbon flux, with R² values of 84.46% (with the impervious surface area index in) and 83.74% (without the impervious surface area index in). Furthermore, the carbon flux in Fengxian District, Shanghai, exhibited significant spatial heterogeneity: the CO₂ flux in the western part of Fengxian District was less than in the eastern part, and the CO₂ flux gradually increased from the west to the east. In addition, we creatively introduced the diurnal impervious surface area index based on the Kljun model, and clarified the influence of impervious surface on the spatiotemporal simulation of CO₂ flux over the complex urban underlying surface. Based on these findings, we conclude that the RF models can be effectively applied for estimating carbon flux on the complex underlying urban surface. The results of our study reduce the uncertainty in modeling carbon cycling in terrestrial ecosystems, and make the variety of models for the carbon cycling of terrestrial ecosystems more diverse.



Citation: Zhang, K.; Zhao, M.; Zhao, Z.; Shen, X.; Lu, Y.; Gao, J.

Spatiotemporal Modeling of Carbon Fluxes over Complex Underlying Surfaces along the North Shore of Hangzhou Bay. *Atmosphere* **2024**, *15*, 727. <https://doi.org/10.3390/atmos15060727>

Academic Editor: Dmitry Belikov

Received: 20 May 2024

Revised: 7 June 2024

Accepted: 15 June 2024

Published: 17 June 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/>).

Keywords: urban; complex underlying surface; carbon and water fluxes; machine learning; simulation

1. Introduction

Human transportation-related activities are responsible for 80% of the carbon dioxide (CO₂) emission in densely populated urban areas [1], making them a significant source

of atmospheric carbon. Changes in urban environments resulting from human activities primarily revolve around several aspects, such as the combustion of fossil fuels and the rapid increase in impervious surfaces [2,3]. The rapid increase in urban impervious surface area has altered the biogeochemical processes associated with the soil–atmosphere gas exchange within the urban ecosystems and their response to global climate change. An urban impervious surface not only weakens surface transpiration and evaporation but also interferes with the exchange of water and heat between the surface and the atmosphere [4].

In the continual evolution of cities and the escalating challenges posed by climate change, the establishment of eddy covariance (EC) observation systems within urban areas has emerged as a crucial tool for comprehending urban environmental transformations. The research on carbon, water, and energy exchange between complex underlying urban surfaces and the atmosphere by the eddy covariance observation system mainly has focused on dynamic characteristics, influencing factors, and flux footprint determination. Lietzke et al. [5] presented a comprehensive overview of recent studies on urban CO₂ emissions, and the longest time series analysis reported in the literature spans six years [6,7]. Furthermore, Schmutz et al. [8] conducted a decade-long examination of the long-term sequence of urban ecosystems, offering insights into the CO₂ flux and CO₂ molar concentration of this flux through observation at the University of Basel in Switzerland.

Since the 20th century, there has been increasing recognition of the predictive power of machine learning algorithms to solve a variety of problems, such as determining the relative importance of environmental factors affecting the mechanisms of carbon exchange [9], elucidating the nonlinear procession of carbon interaction between surface and atmosphere [10], providing a useful tool for carbon flux upscaling [11], and combining different process-based models to reduce predictive uncertainty [12], interpolated missing data on carbon flux, energy flux, and climate variables based on observations from the Global Flux Towers [13]. Existing studies mainly have focused on the simulation of carbon flux in natural ecosystems. The simulation and prediction of carbon flux in ecosystems remain challenging, especially for highly heterogeneous urban underlying surfaces. Tramontana et al. [14] conducted a comprehensive cross-validation analysis on the spatial prediction of various carbon and energy fluxes, highlighting the particular difficulty in predicting net ecosystem exchange (NEE). Meyer [15] proposed a sequence feature selection algorithm based on spatial cross-validation to remove spatial autocorrelation predictors. However, challenges persist in expanding the scale, with data quality of EC data being a significant obstacle. Wang et al. [16] utilized the random forest (RF) model to assess the relationship between land use and carbon emissions. Their results showed that carbon emissions from retail and residential land categories accounted for a large proportion, and carbon emissions from terrace houses were higher than emissions from other residential building categories. Reitz et al. [17] used the RF model to predict the daily CO₂ flux at 250 m spatial resolution in the Ruhr catchment area of western Germany from 2010 to 2018; the results proved that although the model results underestimated the variance of CO₂ flux, they could accurately reflect the average value. Thus, spatial prediction was more difficult than time series prediction.

Previous studies on the observation and simulation of carbon flux mostly focused on natural ecosystems with relatively simple underlying surface types, such as forest ecosystems, grassland ecosystems, and farmland ecosystems [18–20]. However, due to the high heterogeneity of urban complex underlying surface and intensive human activities, it is difficult for physiological models to be used to estimate the variations and spatial distribution of carbon flux in urban. At the same time, activities in urban areas also increase the uncertainty of carbon flux simulation. Therefore, a data-driven-based model is an effective way to solve the simulation of carbon flux over urban complex underlying surfaces. Although there were researchers using machine learning to study the interpolation and scale inference of carbon flux in urban, the results were not very good [16,17]. In most of these studies, only meteorological and vegetation factors were considered in the model, and there were also studies that believed the impact of land use type on the spatial distribution

of carbon flux, but only considered different ecosystem types, and did not study the spatial distribution of carbon flux of different land uses in the complex urban underlying surface.

In conclusion, previous research could not develop a suitable model to study the spatiotemporal distribution of CO₂ flux in urban areas. Hence, our objectives are the following: (1) Using four machine learning models to simulate the long time series of carbon flux over the complex urban underlying surfaces, and evaluating the model performance; (2) Utilizing the best-fitting model to scale up the carbon flux to Fengxian District, Shanghai.

2. Materials and Methods

2.1. Site Description

The study area is located in Fengxian District, Shanghai, along the north bank of Hangzhou Bay in the middle and lower reaches of the Yangtze River in China (Figure 1). The study area is flat and located in the mid-temperate zone geographically, with a subtropical monsoon climate. The prevailing wind direction is southeast, and the climate is humid and mild [21]. The annual average temperature stands at around 16.1 °C, with an annual rainfall of approximately 1191.5 mm. Moreover, the annual frost-free period lasts for about 225 days [22]. The underlying surface in the study is complex and fragmented, including forests, grasslands, farmlands, water bodies, and buildings. Each land use type collectively influences the carbon cycle within the study area [22].

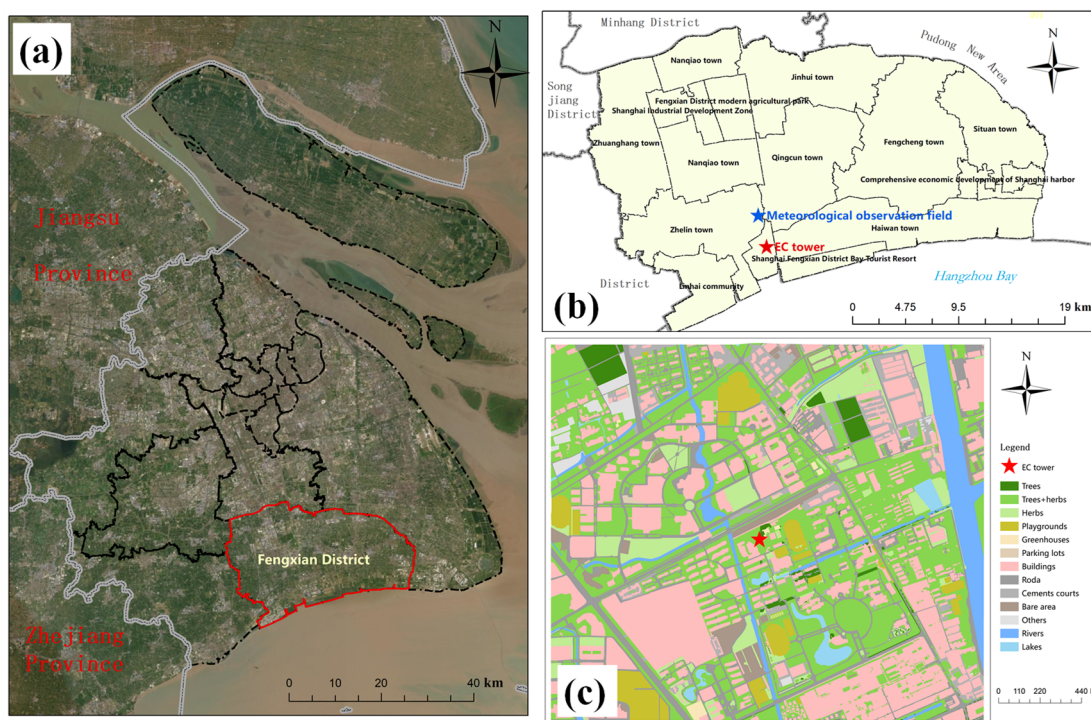


Figure 1. (a) An image of Shanghai; (b) a distribution map of streets and towns in Fengxian District of Shanghai; and (c) a land use map of the underlying surface for the vorticity tower area in 2019.

2.2. Measurements and Data Processing

The eddy covariance flux observation tower (EC tower, 121°30′38.96″ E, 30°50′32.26″ N) is located in Fengxian Bay University City, Shanghai. CO₂ flux was measured using an open-path eddy covariance (OPEC) system, which consisted of an open-path and fast-response infrared gas analyzer (Model LI-7500, Li-Cor Inc., Lincoln, NE, USA) to monitor the densities of CO₂ and H₂O, and a 3D sonic anemometer (Model WindMaster Pro, Gill Instruments Ltd., Lymington, UK) to measure the fluctuations of three-dimensional wind speed and virtual temperature [21]. The height of EC monitoring was 20 m. The raw data were recorded and saved to a data logger (Model CR 3000, Campbell Scientific Inc., North Logan, UT, USA) (Figure 1).

Micrometeorological measurement systems monitored multiple parameters, including relative humidity and air temperature (HMP-45C, Vaisala Inc., Helsinki, Finland), wind speed (AR-100, Vector Instruments, Weymouth, UK), photosynthetically active radiation (PAR) (Li-190SB, Li-Cor Inc., Lincoln, NE, USA) at 15 m above the ground surface, and soil temperature (Ts) (CS615-L, Campbell Scientific, Logan, UT, USA) at 5 cm below the soil surface. The meteorological data were recorded at 30 min intervals using a datalogger (Model CR 3000, Campbell Scientific).

In this study, we selected five years of flux and meteorological data (2011, 2012, 2017, 2018, and 2019). We used EddyPro 7.1.1 software (Li-COR, Lincoln, NE, USA) to calculate CO₂ flux at a time interval of 30 min. EddyPro 7.1.1 was also employed for various data treatments, including axis rotation coordination, frequency response correction, Webb Pearman Leuning (WPL) density correction, wild point removal, and flux data quality indicator establishment (where 0 represents the best quality, 1 represents the middle quality, and 2 represents the worst quality) [23–25]. Then we deleted the flux generated within 1 h before and after the rain and when the friction velocity (u^* , u^*) was less than 0.15 m/s, and we also deleted the flux data quality marked as 2. When the CO₂ flux was negative, it represented that the whole ecosystem was in a state of absorbing CO₂, and when the flux was positive, it represented that the whole ecosystem was in a state of releasing CO₂. After data quality control, we gap-filled CO₂ flux using the REddyProc package in R.

2.3. Daily Scale Urban Impervious Surface Area Data

The Kljun flux footprint model is a novel algorithm based on scale (dimension) analysis developed by Kljun et al. [26] that can be used to calculate the crosswind integral function of a flux source area. According to the Kljun flux source calculation model, we calculated the daily scale flux source area and obtained the two-dimensional coordinates and a two-dimensional plane diagram of the flux source area. Finally, the daily scale land use/land cover information within the flux source area was obtained. The computation of the footprints was conducted by MatLab 2015b.

2.4. Grid Data

The environmental factors affecting carbon flux at a single site and regional carbon flux have some differences. In this study, combined with the research results of Papale [11], Chen [27], Li [28], and Liu [29], we used the six factors of land use, atmospheric temperature, soil moisture, air relative humidity, precipitation, and photosynthetically active radiation to study the scale increase in regional carbon flux. Explanatory variables were compiled from various sources and were of different spatial and temporal resolutions, as shown in Table 1. The time span of all data is 2011–2019.

Table 1. Predictor variables used for model simulation.

| Nr | Name | Source | Temporal Resolution | Spatial Resolution | Unit |
|----|---------------------------------|-----------------------|---------------------|--------------------|---|
| 1 | Soil moisture | European Space Agency | 1 day | 0.25° | / |
| 2 | Precipitation | ERA-Interim | 1 day | 1° | mm |
| 3 | Temperature | ERA-Interim | 1 day | 0.1° | °C |
| 4 | Relative humidity | ERA-Interim | 1 month | 0.25° | % |
| 5 | Photosynthetic active radiation | ERA-Interim | 1 month | 0.125° | $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ |
| 6 | Land use | European Space Agency | 1 year | 300 m | / |

2.5. Statistical Analysis

The Lindeman–Merenda–Gold (LMG) method [30] was used to quantify the relative contributions of each factor to the daily changes of carbon flux over complex underlying surfaces. We performed the algorithm using the “relaimpo” package in R, which contains variance decomposition methods for multiple linear regression models. The LMG method

estimated the relative importance of each variable by decomposing the square sum into non-negative contributions shared by each variable, and obtained the LMG value by averaging the sum of squares (R^2) of all possible orders, as follows:

$$LMG(x_k) = \frac{1}{k!} \sum_{q(k)} seqR^2\{x_k|S_k(q)\},$$

where $q(k)$ is the order in which the independent variable x_k enters the multiple linear regression model; $S_k(q)$ is the set of independent variables before x_k enters the model in q permutations; $seqR^2\{x_k|S_k(q)\}$ is the R^2 increment when x_k enters the model when the model contains the independent variable set $S_k(q)$; and $LMG(x_k)$ is the average increase in R^2 caused by the independent variable x_k .

2.6. Machine Learning and Model Evaluation

Long short-term memory (LSTM) is a fully interconnected neural network which has positive and negative feedback connections between neurons [31]. It initially considers the long-term dependence of the learning phase and can overcome the shortcomings of recurrent neural networks (RNNs) [32]. We used the backpropagation algorithm with gradient descent to calculate the weights and bias terms in the training phase to minimize the objective function across time. Each LSTM unit included four parts: input gate; output gate; forgetting gate; and memory unit. LSTM used these memory units to control the impact of historical information on current information, which ultimately enabled the model to persist and transmit information.

The artificial neural network (ANN) referred to a feedforward neural network based on Pytorch [33]. The architecture of the neural network model of the algorithm comprises three primary parts: the input layer, the hidden layer (middle layer), and the output layer. Typically, only one input layer and one output layer are considered. However, the number of hidden layers varies in different studies. In this study, all operations used a hidden layer. The parameter value played a decisive role in the performance of the ANN model during the training, verification, and generalization phases. We employed the trial-and-error method to determine the number of nodes in the hidden layer, based on the squared error between the output value and the observed value of the training network. This trained model was then applied to the model validation and testing (prediction) stages. In this study, the number of trained nodes in the hidden layer was 128.

The RF model is a tree-based ensemble method utilized to manage high-dimensional regression simulations in which forest development is based on multiple interconnected trees [34,35]. In regression problems, the basic units of RF are regression trees. Each regression tree is constructed using random initial data sampling, where in a random subset of “ m ” attributes is used in each data sample to choose attributes with the most significant information. RF generates a ranking of the most important attributes in forest development based on the cumulative importance of the node partition in each tree. The regression trees are independent of each other. Each node in a regression tree randomly selects a subset of characteristic variables, then picks the optimal subset of variables from these subsets to split the branches. The final estimation is the average of estimates from all regression trees. RF is neutral to outliers and can avoid overfitting when dealing with high-dimensional features [36].

The support vector machine (SVM) method is known for its strong generalization ability [37,38]. When developing an SVM model, it is particularly important to select the appropriate kernel function. In our study, we compared and evaluated several kernel functions based on the dataset to ensure the accuracy of the predictions. The radial basis function (RBF) kernel yielded the best performance for our SVM model. Additionally, after determining the kernel function, we considered other parameters that influence the SVM model’s simulation ability, such as the insensitive loss factor, error penalty factor, and kernel function parameters. In this study, the insensitive loss coefficient was set to 0.01 by default. We utilized grid search to determine the kernel function parameters.

The model accuracy was assessed using root mean square error (RMSE), mean absolute error (MAE), and R-squared value (R^2). R^2 reflects the overall simulation performance of the model, RMSE indicates the general quality of the simulation, and MAE measures the average deviation of the simulation results [39]. The model was considered highly accurate when R^2 approached 1, and both RMSE and MAE approached 0.

The operating system of this experiment was Windows 10, using Python 3.8.5 as the development language, and using Jupyter Notebook 6.4 and Spyder 3.5 as the development platform.

3. Results

3.1. Important Driving Factors of Carbon Flux in a Long Time Series

To identify the key environmental variables influencing the long-term carbon fluxes at the site, we utilized the LMG model to analyze the contribution of urban impervious surface area (IMS), as well as atmospheric temperature (T_a), soil temperature 10 cm from the surface ($T_s_{10\text{ cm}}$), air relative humidity (RH), and net radiation (Rn) on the daily average carbon flux changes. The results are given in Table 2. Within the observation scope of the flux tower, each variable exhibited distinct effects on the fluctuations of daily average carbon flux within the ecosystem. Among these factors, air temperature emerged as the primary driver behind the variations in daily average carbon fluxes, explaining 61.22% of the observed trend of daily average CO_2 flux. Additionally, $T_s_{10\text{ cm}}$ and Rn contributed 13.12% and 13.30%, respectively, to the changes in daily average CO_2 flux. The impervious surface area had a relatively moderate influence, accounting for 8.23% of the variability in daily average CO_2 flux.

Table 2. Contribution of influence factors to daily average carbon flux.

| Items | Abbreviation of Factors | Contribution (%) |
|----------------------|-------------------------|------------------|
| CO ₂ flux | T_a | 61.222 |
| | $T_s_{10\text{ cm}}$ | 13.12 |
| | Rn | 13.304 |
| | RH | 4.136 |
| | IMS | 8.219 |

T_a , RH, $T_s_{10\text{ cm}}$, Rn, and IMS factors had a significant influence on the change in carbon flux. Therefore, we considered these five factors to be input factors to simulate the carbon flux in a long time series.

3.2. Evaluation of Model Performance for Long-Term CO₂ Flux

In the assessment of daily average CO_2 flux prediction for urban complex underlying surfaces using SVM, LSTM, ANN, and RF models, notable differences were observed in the agreement between predicted and observed values. Despite this, most predictions clustered closely around the 1:1 line, suggesting that these machine learning models could effectively forecast daily average CO_2 flux for complex urban environments (Figure 2). The performance metrics varied significantly among the models. The RF model exhibited the lowest RMSE of $0.293\ \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ and the highest R^2 of 0.852, indicating superior performance compared to the other models. The SVM, LSTM, and ANN models had RMSE values of $0.413\ \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, $0.461\ \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, and $0.438\ \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, respectively, all higher than the RF model. Under the condition of similar RMSE values, the LSTM model demonstrated a higher R^2 value (0.830) than both the SVM model (0.702) and the ANN model (0.688). This suggested that the LSTM model's predictive performance surpassed that of the SVM and ANN models. The MAE values of the RF model ($0.157\ \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$), SVM model ($0.233\ \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$), and the LSTM model ($0.212\ \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) were lower than that of the ANN model ($0.260\ \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$). The ANN model recorded the highest RMSE and MAE values, accompanied by the lowest R^2 , indicating its comparatively weaker predictive performance among the models.

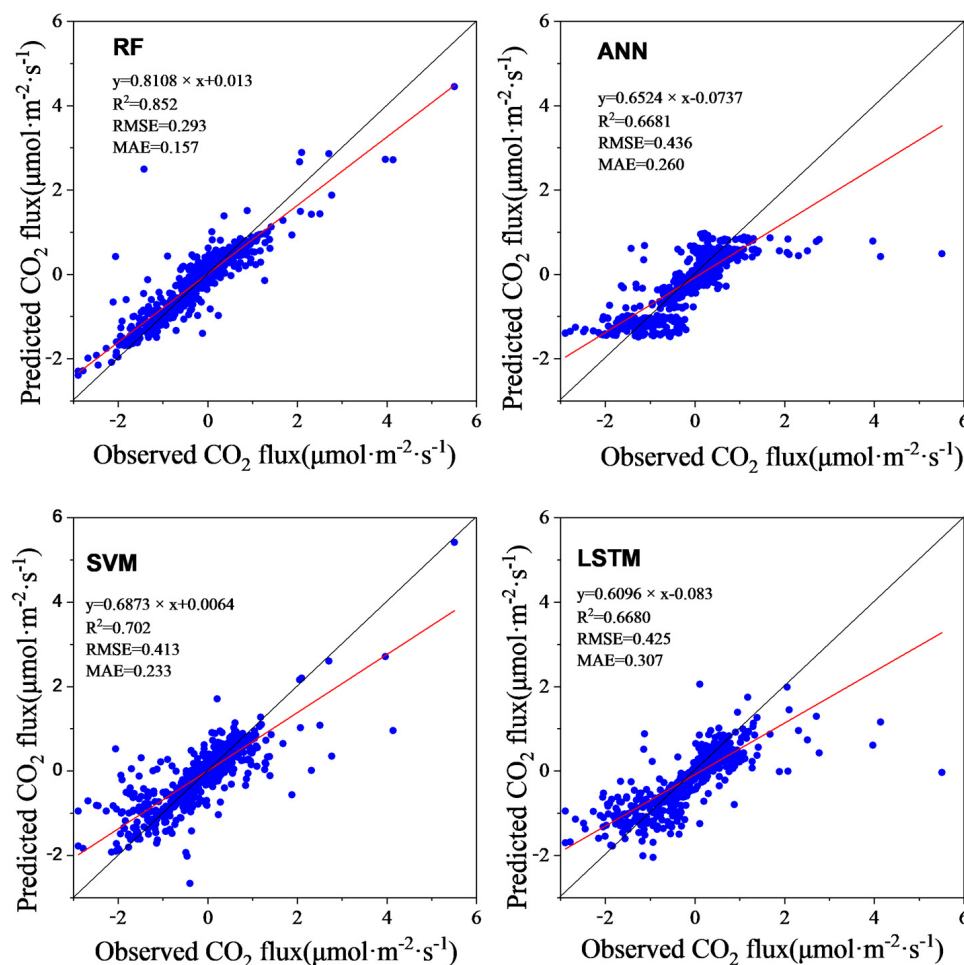


Figure 2. Comparison of daily mean CO₂ flux predicted by different machine learning models and observed daily mean CO₂ flux: SVM: support vector machine; LSTM: long short-term memory network; ANN: artificial neural network; RF: random forest. The red line was the fitting line, the black one was 1:1 line.

3.3. Simulation of Regional Carbon Flux Distribution Based on the RF Model

Based on the findings from the preceding section, we utilized the RF model to simulate the annual CO₂ flux at a spatial resolution of 500 m for Fengxian District in the years 2011, 2012, 2017, 2018, and 2019 (Figure 3). Our results demonstrated that the RF model exhibited strong simulation capabilities, with an R² value of 0.8446, an RMSE of 0.2808 μmol·m⁻²·s⁻¹, and an MAE of 0.1462 μmol·m⁻²·s⁻¹. Fengxian District acted as a net CO₂ source on average between 2011 and 2019; the average net CO₂ exchange was 1.02 g·m⁻²·d⁻¹. The analysis of the annual average CO₂ flux in Fengxian District revealed that the western region exhibited lower values (mean CO₂ flux of 0.79 g·m⁻²·d⁻¹) compared to the eastern region (mean CO₂ flux of 1.25 g·m⁻²·d⁻¹), while the northern region had lower values (mean CO₂ flux of 0.87 g·m⁻²·d⁻¹) than the southern region (mean CO₂ flux of 1.22 g·m⁻²·d⁻¹), with a gradual increase in CO₂ flux values from west to east. In the western region, the CO₂ flux values demonstrated a circular pattern, with Nanqiao Town and Fengxian District Modern Agricultural Park serving as the central points, and the CO₂ flux values increasing outward in a concentric manner. Furthermore, the distribution characteristics and range of CO₂ flux varied significantly from year to year.

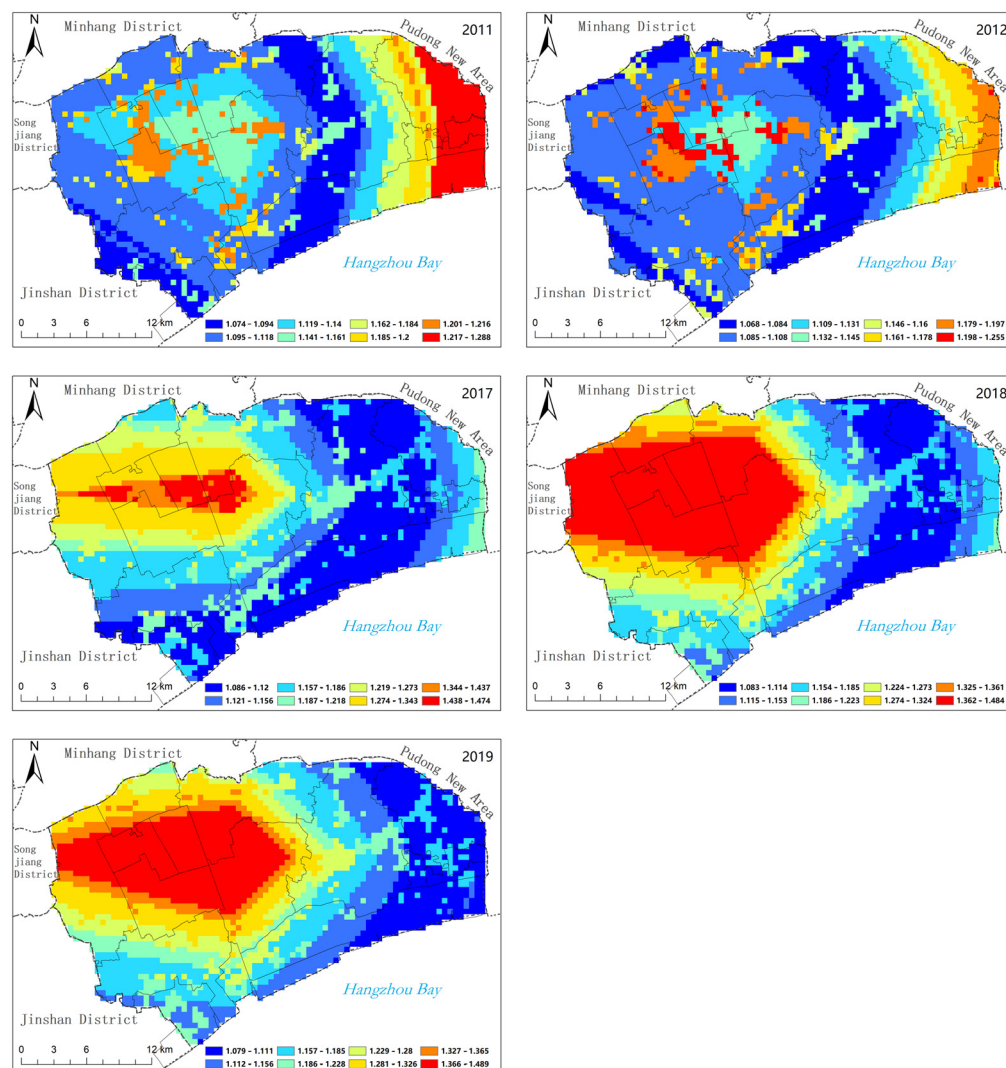


Figure 3. Spatial distribution of annual average CO₂ flux in Fengxian District of Shanghai along the north bank of Hangzhou Bay in 2011, 2012, 2017, 2018, and 2019.

3.4. Effects of Impervious Surface Area in Simulating CO₂ Flux

In this study, we introduced the factor of land use change, specifically, the incorporation of impervious surface area, and assessed the disparity in simulation performance of the carbon flux model with and without impervious surface area factor (Figure 4). It was observed that upon incorporating impervious surface area into the training model (Figure 4), the RF model demonstrated an ability to account for 84.46% of the daily average CO₂ flux variation over the complex underlying urban surface. The resulting RMSE was $0.2808 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, with an MAE of $0.1462 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, and the predicted value and observed value generally were distributed near the 1:1 line. In contrast, the RF model accounted for 83.74% of the daily average CO₂ flux variation over the complex underlying urban surface without the inclusion of impervious surface area in its training (Figure 4). The RMSE was $0.2872 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, and the MAE was $0.147 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Consequently, based on this comparison, we designated the incorporation of impervious surface area as a contributing factor, which effectively enhanced the model's performance and consequently refined the accuracy of simulation outcomes.

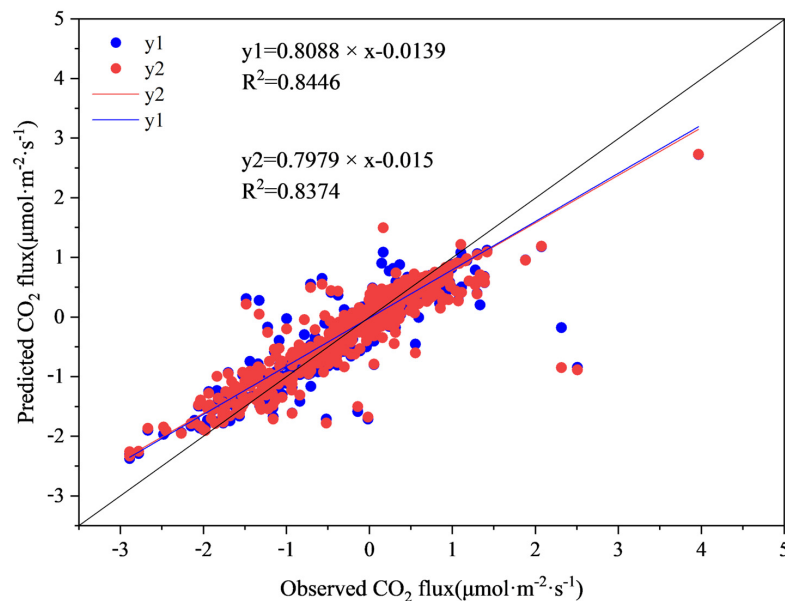


Figure 4. Comparison of the predicted daily average carbon flux with the observed daily average carbon flux: y1 is CO₂ flux predicted by random forest with impervious surface input; y2 is CO₂ flux values predicted by RF without impervious surface input.

3.5. Interannual Variation of CO₂ Flux Spatial Distribution

This study utilized an RF model to simulate the spatial distribution patterns of carbon fluxes in Fengxian District of Shanghai for the years 2011, 2012, 2017, 2018, and 2019. Additionally, a simple linear regression method was employed to quantify the interannual variations in the spatial patterns of carbon fluxes in Fengxian District (Figure 5). Over the span of five years, the interannual variability of annual CO₂ fluxes in Fengxian District, Shanghai, exhibited very distinct characteristics: more than 90% of the region experienced a decrease in annual CO₂ flux values, with the maximum reduction rate reaching 17.83%. These reductions were primarily concentrated in the eastern Shanghai Harbor Industrial Zone, western Nanqiao Town, and Zhuanghang Town of Fengxian District. Conversely, there were fewer areas where the annual CO₂ flux values increased, and these were more scattered, primarily located in the eastern part of Haiwan Town, with a growth rate of up to 6.74%. Combined with the land use changes in Fengxian District of Shanghai, it was found that the areas with decreased CO₂ flux values generally corresponded to regions with denser vegetation, while areas with increased CO₂ flux values were associated with increased land for construction purposes [40].

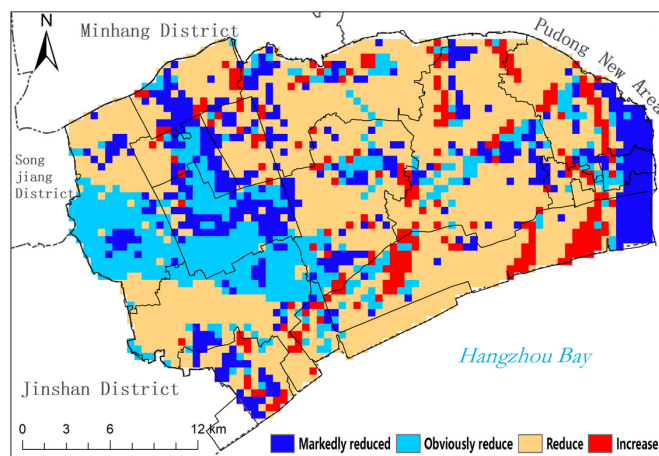


Figure 5. Interannual spatial variation pattern of annual average carbon flux in Fengxian District of Shanghai in 2011, 2012, 2017, 2018, and 2019.

4. Discussion

In this study, we adopted three traditional machine learning models (ANN, SVR, RF) and one deep learning model (LSTM). Our modeled results are shown to be within a reasonable range and highlight the importance of impervious surface area in simulating CO₂ flux over urban complex underlying surfaces. Machine learning models are automatically able to learn complex nonlinear relationships from input data [41,42]. In view of our restricted comprehension of the complicated physical, chemical, and biological interactions in the carbon flux cycle within urban ecosystems, there exists significant uncertainty in the simulation of carbon flux [14]. Leveraging machine learning models enables us to more precisely quantify the spatial and temporal dynamics of carbon flux on regional and global scales. Relatively few studies have examined the time series of urban CO₂ flux [43,44]. Schmidt et al. [43] employed ANN and RBF to observe and simulate CO₂ flux. In a six-week observational simulation conducted 65 m above an urban community in Münster, Germany, they achieved an R² of 0.67. Järvi et al. [44] predicted CO₂ flux based on the ANN model algorithm using five years of data from Helsinki, Finland, achieving an R² up to 0.4; the findings of Menzer et al. [13] showed that the ANN algorithm successfully provided similar performance in urban environments (with R² values ranging from 0.60 to 0.86). The performance of various machine learning models in estimating NEE also varied. The R² of ANN in estimating CO₂ flux in this study was 0.67, consistent with the results in Münster, while the RF model attained an R² value of 0.85 in Münster. Zeng et al. [45] used the RF model to estimate the carbon flux of global terrestrial ecosystems based on eddy covariance data, achieving favorable outcomes (with R² values of 0.97 for gross primary production (GPP), 0.96 for ecosystem respiration (RECO), and 0.94 for NEE when combining all training data).

By comparing the results of previous studies and the results in our study, we found that when ANN simulated CO₂ flux over urban complex surfaces, there was a difference in the simulation performance, which might be related to the input factors or the length of time [46,47]. The RF model has shown excellent simulation performance in many studies, which was consistent with the results of this study.

In the simulation and prediction of carbon fluxes in the study area, quantifying the actual contributions of land use and land cover changes to carbon fluxes [48,49] shows significant uncertainty. One of the reasons is the lack of core data that disaggregate these fluxes into individual grids [49]. Furthermore, limited data or the exclusion of certain processes (such as tree felling and conversion of land for cultivation) may lead to the underestimation of carbon dioxide emissions or transfers resulting from land use changes [49–52]. Most studies establish regional-scale simulation models for carbon fluxes based on data from individual sites to construct machine learning models applied to those sites [53–55]. However, utilizing data from multiple flux sites to establish a generalized simulation model proves effective in addressing this issue. Such a generalized model can also be used to infer carbon fluxes at meteorological stations and provide additional observational datasets for studies on flux changes across multiple regions. The RF model is considered a reasonable and suitable method for simulating CO₂ fluxes from site to regional scales. Firstly, as a machine learning algorithm, the RF model selects the optimal output from multiple regression trees to capture the features of the data, effectively enhancing the accuracy of flux data [56,57]. Secondly, by extracting the multivariate functional relationships between observed data and explanatory variables, the RF model can integrate data from different sources and simplify complex processes, addressing nonlinear issues in ecosystems [58].

This study proposed a novel model for simulating the spatiotemporal scale of carbon fluxes in urban complex underlying surfaces, leveraging four machine learning algorithms for the first time. Such a model demonstrated commendable performance in handling flux data while effectively circumventing the computation of intricate parameters and ecological processes. By utilizing footprint models to quantify the area of impervious surfaces at a daily scale, the model achieved enhanced accuracy in simulating the spatiotemporal variability of carbon flux in Fengxian District, Shanghai, thereby reducing the uncertainty

associated with spatiotemporal scale simulation of carbon flux in the region. Nevertheless, this study acknowledges several uncertainties and limitations. Firstly, it did not quantify the errors generated during the process of data scale transformation. Secondly, due to the opaque nature of machine learning models, it was difficult to fully understand the impact of the models on the prediction results during the model training process. Additionally, the variable resampling process only employed linear interpolation, neglecting other resampling methods. Hence, future research endeavors should focus on exploring appropriate methods to mitigate the scale effects caused by scale conversion.

5. Conclusions

In this study, we employed machine learning algorithms to proficiently simulate the spatiotemporal fluctuations of carbon flux on the complex underlying surfaces of Fengxian District, located on the northern bank of Hangzhou Bay, by integrating diverse sources of observation data. The primary findings obtained in this study are summarized concisely as follows:

- (1) Our study demonstrated that the four machine learning models used in our study can accurately simulate the long-term carbon flux over the complex underlying surfaces, with the RF model exhibiting the highest simulation performance.
- (2) The RF model can accurately portray the spatiotemporal distribution characteristics of carbon flux in Fengxian District, Shanghai.
- (3) Spatial heterogeneity in carbon flux was evident in Fengxian District on the north bank of Hangzhou Bay: the carbon flux value in the western region was lower compared to this in the eastern region, with a gradual increase observed from west to east within Fengxian District.
- (4) When simulating the spatiotemporal carbon flux of complex underlying surfaces using machine algorithms, the incorporation of the impervious surface area index marginally improved the accuracy of long-term carbon flux simulations. At a spatial scale, regions with larger impervious surface areas exhibit higher carbon flux values, indicating a strong correlation between carbon flux distribution and land use patterns. Consequently, the incorporation of the impervious surface area index serves as a relatively significant indicator for simulating spatial-scale carbon flux.

This study concludes that the RF model can accurately simulate the carbon flux of complex urban underlying surfaces and confirms the significant role of impervious surface area in precisely predicting the spatiotemporal scale of carbon flux on such surfaces. The innovation of this paper is the obtainment of the impervious area of the daily scale based on the Kljun model, and the results in our study demonstrate that the incorporation of the daily impervious surface area index improves the accuracy of long-term carbon flux simulations over the urban complex underlying surface. This not only reduces the uncertainty in modeling carbon cycling in terrestrial ecosystems but also broadens the variety of models for the carbon cycling of terrestrial ecosystems.

Author Contributions: Conceptualization, K.Z., M.Z. and J.G.; methodology, K.Z. and Z.Z.; software, K.Z. and X.S.; validation, Z.Z., X.S. and Y.L.; formal analysis, K.Z. and Z.Z.; investigation, K.Z. and X.S.; resources, K.Z. and M.Z.; data curation, K.Z., Z.Z. and X.S.; writing—original draft preparation, K.Z.; writing—review and editing, K.Z.; visualization, K.Z. and Z.Z.; supervision, J.G., Y.L. and M.Z.; project administration, K.Z. and J.G.; funding acquisition, J.G. and K.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Joint Research Project for Meteorological Capacity Improvement [grant number 22NLTSQ011] and the Key Program of National Natural Science Foundation of China [grant no. 41730642]. And the APC was funded by [the Key Program of National Natural Science Foundation of China].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is unavailable due to privacy or ethical restrictions.

Acknowledgments: We greatly appreciate the suggestions for illuminating the manuscript given by the anonymous reviewers and give thanks also to the editorial staff.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Grimmond, C.S.B.; King, T.S.; Cropley, F.D.; Nowak, D.J.; Souch, C. Local-scale fluxes of carbon dioxide in urban environments: Methodological challenges and results from Chicago. *Environ. Pollut.* **2002**, *116*, 243–254. [[CrossRef](#)] [[PubMed](#)]
- Schmid, H.P. Experimental design for flux measurements: Matching scales of observations and fluxes. *Agric. For. Meteorol.* **1997**, *87*, 179–200. [[CrossRef](#)]
- Papale, D.; Valentini, A. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Glob. Change Biol.* **2003**, *9*, 525–535. [[CrossRef](#)]
- Chen, H.; Huang, J.J.; Liang, H.; Wang, W.; Li, H.; Wei, Y.; Jiang, A.Z.; Zhang, P. Can evaporation from urban impervious surfaces be ignored? *J. Hydrol.* **2023**, *616*, 128582. [[CrossRef](#)]
- Lietzke, B.; Vogt, R.; Feigenwinter, C.; Parlow, E. On the controlling factors for the variability of carbon dioxide flux in a heterogeneous urban environment. *Int. J. Climatol.* **2015**, *13*, 3921–3941. [[CrossRef](#)]
- Crawford, B.; Grimmond, C.S.B.; Christen, A. Five years of carbon dioxide fluxes measurements in a highly vegetated suburban area. *Atmos. Environ.* **2011**, *45*, 896–905. [[CrossRef](#)]
- Gioli, B.; Toscano, P.; Lugato, E.; Matese, A.; Miglietta, F.; Zaldei, A.; Vaccari, F.P. Methane and carbon dioxide fluxes and source partitioning in urban areas: The case study of Florence, Italy. *Environ. Pollut.* **2012**, *164*, 125–131. [[CrossRef](#)]
- Schmutz, M.; Vogt, R.; Feigenwinter, C.; Parlow, E. Ten years of eddy covariance measurements in Basel, Switzerland: Seasonal and interannual variabilities of urban CO₂ mole fraction and flux. *J. Geophys. Res. Atmos.* **2016**, *121*, 8649–8667. [[CrossRef](#)]
- Gevrey, M.; Dimopoulos, I.; Lek, S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Model.* **2003**, *160*, 249–264. [[CrossRef](#)]
- Evrendilek, F. Quantifying biosphere-atmosphere exchange of CO₂ using eddy covariance, wavelet denoising, neural networks, and multiple regression models. *Agric. For. Meteorol.* **2013**, *171–172*, 1–8. [[CrossRef](#)]
- Papale, D.; Black, T.A.; Carvalhais, N.; Cescatti, A.; Chen, J.; Jung, M.; Kiely, G.; Lasslop, G.; Mahecha, M.D.; Margolis, H.; et al. Effect of spatial sampling from European flux towers for estimating carbon and water fluxes with artificial neural networks. *J. Geophys. Res. Biogeosci.* **2015**, *120*, 1941–1957. [[CrossRef](#)]
- Wang, T.; Brender, P.; Ciais, P.; Piao, S.; Mahecha, M.D.; Chevallier, F.; Reichstein, M.; Ottlé, C.; Maignan, F.; Arain, A.; et al. State-dependent errors in a land surface model across biomes inferred from eddy covariance observations on multiple timescales. *Ecol. Model.* **2012**, *246*, 11–25. [[CrossRef](#)]
- Menzer, O.; Meiring, W.; Kyriakidis, P.C.; McFadden, J.P. Annual sums of carbon dioxide exchange over a heterogeneous urban landscape through machine learning based gap-filling. *Atmos. Environ.* **2015**, *101*, 312–327. [[CrossRef](#)]
- Tramontana, G.; Jung, M.; Schwalm, C.R.; Ichii, K.; Camps-Valls, G.; Ráduly, B.; Reichstein, M.; Arain, M.A.; Cescatti, A.; Kiely, G.; et al. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences* **2016**, *13*, 4291–4313. [[CrossRef](#)]
- Meyer, H.; Reudenbach, C.; Hengl, T.; Katurji, M.; Nauss, T. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environ. Model. Softw.* **2018**, *101*, 1–9. [[CrossRef](#)]
- Wang, G.; Han, Q.; Vries, B.D. Assessment of the relation between land use and carbon emission in Eindhoven, the Netherlands. *J. Environ. Manag.* **2019**, *247*, 413–424. [[CrossRef](#)] [[PubMed](#)]
- Reitz, O.; Graf, A.; Schmidt, M.; Ketzler, G.; Leuchner, M. Upscaling net ecosystem exchange over heterogeneous landscapes with machine learning. *J. Geophys. Res. Biogeosci.* **2021**, *126*, JG005814. [[CrossRef](#)]
- Dou, X.; Chen, B.; Black, T.A.; Jassal, R.S.; Che, M. Impact of nitrogen fertilization on forest carbon sequestration water loss in a chronosequence of three douglasfir stands in the pacific northwest. *Forests* **2015**, *6*, 1897–1921. [[CrossRef](#)]
- Dou, X.M.; Yang, Y.G. Comprehensive Evaluation of Machine Learning Techniques for Estimating the Responses of Carbon Fluxes to Climatic Forces in Different Terrestrial Ecosystems. *Atmosphere* **2018**, *9*, 83. [[CrossRef](#)]
- Qin, Z.; Su, G.-L.; Zhang, J.-E.; Ouyang, Y.; Yu, Q.; Li, J. Identification of important factors for water vapor flux and CO₂ exchange in a cropland. *Ecol. Model.* **2010**, *221*, 575–581. [[CrossRef](#)]
- Zhang, K.; Gong, Y.; Fa, H.; Zhao, M. CO₂ flux characteristics of different plant communities in a subtropical urban ecosystem. *Sustainability* **2019**, *11*, 4879. [[CrossRef](#)]
- Zhang, K.; Gong, Y.; Escobedo, F.J.; Bracho, R.; Zhang, X.; Zhao, M. Measuring multi-scale urban forest carbon flux dynamics using an integrated eddy covariance technique. *Sustainability* **2019**, *11*, 4335. [[CrossRef](#)]
- Starr, G.; Staudhammer, C.L.; Loescher, H.W.; Mitchell, R.; Whelan, A.; Hiers, J.K.; O'Brien, J.J. Time series analysis of forest carbon dynamics: Recovery of *Pinus palustris* physiology following a prescribed fire. *New For.* **2015**, *46*, 63–90. [[CrossRef](#)]
- Starr, G.; Staudhammer, C.L.; Wiesner, S.; Kunwor, S.; Loescher, H.W.; Baron, A.F.; Whelan, A.; Mitchell, R.J.; Boring, L. Carbon dynamics of *Pinus palustris* ecosystems following drought. *Forests* **2016**, *7*, 98. [[CrossRef](#)]

25. Whelan, A.; Starr, G.; Staudhammer, C.L.; Loescher, H.W.; Mitchell, R.J. Effects of drought and prescribed fire on energy exchange in longleaf pine ecosystems. *Ecosphere* **2015**, *6*, 1–22. [[CrossRef](#)]
26. Kljun, N.; Calanca, P.; Rotach, M.W.; Schmid, H.P. A simple parameterisation for flux footprint predictions. *Bound.-Layer Meteorol.* **2004**, *112*, 503–523. [[CrossRef](#)]
27. Chen, C.; Cleverly, J.; Zhang, L.; Yu, Q.; Eamus, D. Modelling seasonal and interannual variations in carbon and water fluxes in an aridzone acacia savanna woodland, 1981–2012. *Ecosystems* **2016**, *19*, 625–644. [[CrossRef](#)]
28. Li, Z.; Zhou, G.S.; Liu, S.H.; Sui, X. Seasonal contribution and interannual variation of evapotranspiration over a reed marsh (*Phragmites australis*) in Northeast China from 3-year eddy covariance data. *Hydrol. Process.* **2010**, *24*, 1039–1047.
29. Liu, B.; Zhao, W.; Wen, Z.; Zhang, Z. Response of water and energy exchange to the environmental variable in a desert–oasis wetland of Northwest China. *Hydrol. Process.* **2014**, *28*, 6098–6112. [[CrossRef](#)]
30. Gromping, U. Relative importance for linear regression in R: The package relaimpo. *J. Stat. Softw.* **2006**, *17*, 27. [[CrossRef](#)]
31. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
32. Fischer, T.; Krauss, C. Deep learning with long short-term memory networks for financial market predictions. *Eur. J. Oper. Res.* **2018**, *270*, 654–669. [[CrossRef](#)]
33. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
34. Chen, D.; Gong, L.; Xu, C.Y.; Halldin, S. A High resolution, gridded dataset for monthly temperature normals (1971–2000) in Sweden. *Geogr. Ann.* **2007**, *89*, 249–261. [[CrossRef](#)]
35. Karimi, S.; Shiri, J.; Kisi, O.; Xu, T. Forecasting daily streamflow values: Assessing heuristic models. *Nord. Hydrol.* **2018**, *49*, 658–669. [[CrossRef](#)]
36. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
37. Raghavendra, N.S.; Deka, P.C. Support vector machine applications in the field of hydrology: A review. *Appl. Soft Comput.* **2014**, *19*, 372–386. [[CrossRef](#)]
38. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [[CrossRef](#)]
39. Yang, Q.C.; Zhang, X.S. Improving SWAT for simulating water and carbon fluxes of forest ecosystems. *Sci. Total Environ.* **2016**, *569–570*, 1478–1488. [[CrossRef](#)]
40. Hong, S.O.; Kim, J.; Byun, Y.H.; Hong, J.; Hong, J.W.; Lee, K.; Park, Y.S.; Lee, S.S. Intra-urban Variations of the CO₂ Fluxes at the Surface-Atmosphere Interface in the Seoul Metropolitan Area. *Asia-Pac. J. Atmos. Sci.* **2023**, *59*, 417–431. [[CrossRef](#)]
41. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science.* **2015**, *349*, 255–260. [[CrossRef](#)] [[PubMed](#)]
42. Liu, Z.; Peng, C.; Work, T.; Candau, J.N.; DesRochers, A.; Kneeshaw, D. Application of machine learning methods in forest ecology: Recent progress and future challenges. *Environ. Rev.* **2018**, *26*, 339–350. [[CrossRef](#)]
43. Schmidt, A.; Wrzesinsky, T.; Klemm, O. Gapfilling and quality assessment of CO₂ and water vapour fluxes above an urban area with radial basis function neural networks. *Bound.-Layer Meteorol.* **2008**, *126*, 389–413. [[CrossRef](#)]
44. Järvi, L.; Nordbo, A.; Junninen, H.; Riikonen, A.; Moilanen, J.; Nikinmaa, E.; Vesala, T. Seasonal and annual variation of carbon dioxide surface fluxes in Helsinki, Finland, in 2006–2010. *Atmos. Chem. Phys.* **2012**, *12*, 8475–8489. [[CrossRef](#)]
45. Zeng, J.; Matsunaga, T.; Tan, Z.-H.; Saigusa, N.; Shirai, T.; Tang, Y.; Peng, S.; Fukuda, Y. Global terrestrial carbon fluxes of 1999–2019 estimated by upscaling eddy covariance data with a random forest. *Sci. Data* **2020**, *7*, 313. [[CrossRef](#)] [[PubMed](#)]
46. Chen, Z.; Lin, T.; Tang, N.; Xia, X. A parallel genetic algorithm based feature selection and parameter optimization for support vector machine. *Sci. Program.* **2016**, *2016*, 2739621. [[CrossRef](#)]
47. Horemans, J.A.; Janssens, I.A.; Gielen, B.; Roland, M.; Deckmyn, G.; Verstraeten, A.; Neirynek, J.; Ceulemans, R. Weather, pollution and biotic factors drive net forest-atmosphere exchange of CO₂ at different temporal scales in a temperate-zone mixed forest. *Agric. For. Meteorol.* **2020**, *291*, 108059. [[CrossRef](#)]
48. Sleeter, B.M.; Liu, J.; Daniel, C.; Rayfield, B.; Sherba, J.; Hawbaker, T.J.; Zhu, Z.; Selmants, P.C.; Loveland, T.R. Effects of contemporary landuse and landcover change on the carbon balance of terrestrial ecosystems in the United States. *Environ. Res. Lett.* **2018**, *13*, 045006. [[CrossRef](#)]
49. Arneeth, A.; Sitch, S.; Pongratz, J.; Stocker, B.D.; Ciais, P.; Poulter, B.; Bayer, A.D.; Bondeau, A.; Calle, L.; Chini, L.P.; et al. Historical carbon dioxide emissions caused by land-use changes are possibly larger than assumed. *Nat. Geosci.* **2017**, *10*, 79–84. [[CrossRef](#)]
50. Caspersen, J.P.; Pacala, S.W.; Jenkins, J.C.; Hurtt, G.C.; Moorcroft, P.R.; Birdsey, R.A. Contributions of landuse history to carbon accumulation in U.S. Forests Science. *Science* **2000**, *290*, 1148–1151. [[CrossRef](#)]
51. Yue, C.; Ciais, P.; Houghton, R.A.; Nassikas, A.A. Contribution of land use to the interannual variability of the land carbon cycle. *Nat. Commun.* **2020**, *11*, 3170. [[CrossRef](#)]
52. Houghton, R.A.; House, J.I.; Pongratz, J.; Van Der Werf, G.R.; DeFries, R.S.; Hansen, M.C.; Le Quééré, C.; Ramankutty, N. Carbon emissions from land use and land–cover change. *Biogeosciences* **2012**, *9*, 5125–5142. [[CrossRef](#)]
53. Cui, X.; Goff, T.; Cui, S.; Menefee, D.; Wu, Q.; Rajan, N.; Nair, S.; Phillips, N.; Walker, F. Predicting carbon and water vapor fluxes using machine learning and novel feature ranking algorithms. *Sci. Total Environ.* **2021**, *775*, 145130. [[CrossRef](#)]
54. Dou, X.M.; Yang, Y.G. Estimating forest carbon fluxes using four different data-driven techniques based on long-term eddy covariance measurements: Model comparison and evaluation. *Sci. Total Environ.* **2018**, *627*, 78–94. [[CrossRef](#)]

55. Hang, I.H.; Hsieh, C.I. Gap-filling of surface fluxes using machine learning algorithms in various ecosystems. *Water* **2020**, *12*, 3415. [[CrossRef](#)]
56. Zhu, X.J.; Yu, G.R.; Chen, Z.; Zhang, W.K.; Han, L.; Wang, Q.F.; Chen, S.P.; Liu, S.M.; Wang, H.M.; Yan, J.H.; et al. Mapping Chinese annual gross primary productivity with eddy covariance measurements and machine learning. *Sci. Total Environ.* **2022**, *857*, 159390. [[CrossRef](#)]
57. Zhang, W.; Luo, G.; Yuan, X.; Li, C.; Xie, M.; Wang, Y.; Ma, X.; Shi, H.; Hamdi, R.; Hellwich, O.; et al. New data-driven method for estimation of net ecosystem carbon exchange at meteorological stations effectively increases the global carbon flux data. *Methods Ecol. Evol.* **2023**, *14*, 2449–2463. [[CrossRef](#)]
58. Zheng, J.; Zhang, Y.; Wang, X.; Zhu, J.; Zhao, G.; Zheng, Z.; Tao, J.; Zhang, Y.; Li, J. Estimation of Net Ecosystem Productivity on the Tibetan Plateau Grassland from 1982 to 2018 Based on Random Forest Model. *Remote Sens.* **2023**, *15*, 2375. [[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.