Available Online: 27 October 2025 DOI: 10.54254/2977-3903/2025.28783

High-precision estimation of urban green-space carbon sink capacity via deep spatiotemporal remote-sensing fusion

Pingting Jiang

Department of Earth Sciences, The University of Hong Kong, Hong Kong, China

jiangpingtingjpt@163.com

Abstract. As urban carbon neutrality initiatives accelerate, green spaces in cities are playing an increasingly critical role as natural carbon sinks in mitigating greenhouse gas emissions. However, conventional carbon estimation approaches struggle with spatial fragmentation and temporal variability in urban green areas, resulting in limited accuracy and poor adaptability. To address this challenge, this study proposes a deep spatiotemporal modeling framework combining Convolutional Neural Networks (CNN) and Temporal Convolutional Networks (TCN), integrating multi-source remote sensing data from Landsat-8, Sentinel-2, and MODIS to estimate carbon storage in Guangzhou's green spaces from 2018 to 2023. Experimental results demonstrate that the model achieves robust performance across diverse land types and seasonal conditions, with an overall RMSE of 2.71 tC/ha, R^2 of 0.926, and SSIM of 0.841, significantly outperforming traditional statistical and machine learning methods. The study confirms the effectiveness of deep fusion modeling in urban carbon sink estimation and offers a scalable technical pathway to support carbon asset management, green space planning, and low-carbon policy development in complex urban contexts.

Keywords: deep learning, remote sensing fusion, urban green space, carbon sink estimation, spatiotemporal modeling

1. Introduction

Under the dual-carbon policy framework, the carbon sink capacity of urban ecosystems has emerged as a critical focus in global climate governance. As a key component of natural carbon sinks, urban green spaces not only provide vital ecosystem services such as microclimate regulation and air purification but also play an irreplaceable role in balancing carbon emissions within dense urban areas [1]. However, existing carbon sink estimation methods face significant challenges in addressing the spatiotemporal heterogeneity, complex land composition, and rapid land-use transitions of urban environments [2]. On one hand, traditional ground-based surveys and empirical regression approaches suffer from high labor costs, limited spatial coverage, and unstable estimation accuracy [3]. On the other hand, although remote sensing technologies offer multi-source, multi-scale, and multi-temporal observations, most mainstream models rely heavily on vegetation indices or statistical learning techniques, lacking the capacity to capture nonlinear dynamics and temporal evolution of urban carbon storage [4]. This is particularly problematic in urban settings where green spaces are fragmented, have ambiguous boundaries, and are subject to frequent anthropogenic disturbance. As a result, simple methods often fail to deliver high-resolution and high-confidence carbon sink estimates. There is thus an urgent need for an intelligent estimation framework capable of joint spatiotemporal modeling and deep feature extraction from heterogeneous remote sensing data, to support accurate carbon sink accounting in urban green spaces and enable the operationalization of carbon asset management.

2. Literature review

2.1. Current studies on urban carbon sinks

Studies on urban green carbon sinks primarily focus on the function of urban green space, especially carbon sinks. This kind of ability is achieved by biomass estimation and carbon stock simulation, which are attained from field surveys [5]. For example, carbon concentration can be calculated from stem diameter or crown width by empirical formulas. It is unsuitable to widely apply this method for its low-efficiency and high-input [6]. Meanwhile, cities have different types of land use cover and complicated human intervention, thus the suitability of empirical formulas in cities is low. Studies have tried to import GIS

databases like NDVI, Urban Green Space Map to evaluate the carbon concentration of urban green space. Even though they address the weakness of manual measurement, they remain constrained by several issues like slow temporal updates, insufficient resolution and poor model adaptability [7]. However, they lack the capacity to model spatiotemporal dynamics across heterogeneous urban environments.

2.2. Applications of remote sensing in carbon estimation

Remote sensing can extract large-scale variation of urban green space from satellite imaginary, including vegetation health, distribution density and coverage area. It also enables large-scale ecological monitoring through vegetation indices like NDVI and EVI [8]. Yet, these indices merely reflect the plant situation, instead of how much carbon does plant intake. Research used linear regression to study the relationship between NDVI and carbon storage, but this model is easily affected by outliers and has limited applicability [9]. Also, these approaches depend on empirical parameters and lack robustness in complex urban settings.

2.3. Deep learning for remote sensing fusion

Recent advances in CNNs, LSTMs, and fusion networks have improved remote sensing interpretation. CNNs help to extract more diverse spatial characteristics from satellite imaginary, while temporal models like LSTM do help to understand the tendency of carbon stock in certain areas as time goes by [10]. Multi-source fusion models enhance prediction accuracy and temporal resolution, offering new opportunities for carbon sink modeling. More importantly, deep learning models are capable of non-linear mapping and end-to-end mapping, without the need for extensive manual definition of rules. Instead, they automatically identify the underlying pattern of urban green space carbon sequestration through training, which make a difference in complicated landform, urban micro-climate and seasonal variation [11]. As a result, it is important to immerge remote sensing and deep learning to achieve high-accurate carbon storage evaluation

3. Experimental methods

3.1. Data collection and preprocessing

In this study, the city of Guangzhou, China, is taken as the study area. Controlled by subtropical monsoon climate, Guangzhou has high greening rates, including encompassing parklands, protective forest belts, roadside vegetation, ancillary greenspaces and natural wetlands. It is an ideal research subject for land type diversity and temporal variation. The remote sensing data sources include Landsat-8 OLI/TIRS, Sentinel-2 MSI, and MODIS Terra/Aqua, with the temporal range covering 2018 to 2023, and the spatial resolutions of 30m, 10m, and 500m, respectively. The data are unified by atmospheric corrections (using Sen2Cor and LEDAPS), resampling and geometric alignment, and the analysis area is divided based on the ecological green map of Guangzhou, which provides a consistent and high data basis for the subsequent modelling. Through atmospheric correction (using Sen2Cor and LEDAPS), resampling and geometric alignment, the temporal frequency and spatial scale of the multi-source images are unified, and the analysis area is divided based on the ecological green map of Guangzhou City, which provides a consistent database for subsequent modelling.

3.2. Model architecture

The proposed model adopts a deep spatiotemporal fusion framework consisting of three components: (1) a spatial feature extraction module based on Convolutional Neural Networks (CNNs), (2) a temporal modeling module based on Temporal Convolutional Networks (TCNs), and (3) a final regression layer for carbon sink estimation. The CNN module uses multiscale convolution kernels to extract spatial patterns such as vegetation texture, boundary morphology, and patch connectivity from satellite images. Let the input remote sensing image at region rrr and time ttt be $X_{r,t}$, the spatial feature output can be formulated as:

$$F_s = CNN(X_{r,t}) = \sum_{i=1}^{n} w_i * x_{r,t}^{(i)} + b \quad (1)$$

Where w_i represents the weight of the i-th convolutional filter, b is the bias term, and F_s denotes the resulting spatial feature map.

The extracted spatial features are then passed to a 1D dilated TCN module, which is capable of modeling long-range temporal dependencies across multiple years and seasons. The temporal feature extraction is expressed as:

$$F_t = TCN(F_s) = \sum_{j=0}^{k} \gamma_j \cdot F_s(t - d \cdot j) \tag{2}$$

Where γ_j is the weight of the j-th temporal convolution kernel, d is the dilation rate, k is the kernel size, and F_t is the temporal feature output.

The fusion of spatial and temporal representations is fed into a fully connected regression layer, which outputs the predicted carbon sink value per pixel in units of tons of carbon per hectare. The overall framework supports end-to-end training and allows efficient handling of large-scale, heterogeneous, and multi-temporal remote sensing inputs. It is particularly suitable for dynamic estimation of urban green space carbon sink capacity with high spatiotemporal fidelity.

3.3. Training and evaluation

This model is trained using a supervised learning framework, with the objective of minimizing the error between the predicted carbon sink values and the actual observed values. The samples are generated through spatial stratified random sampling to maintain category diversity and spatial independence, and are divided into training set, validation set and test set in a ratio of 7:2:1.

Model training is conducted using the Adam optimizer with an initial learning rate of 1×10^{-4}, and an L2 regularization weight decay factor of 1×10^{-5} to mitigate overfitting. Early stopping is employed based on the validation loss to avoid unnecessary iterations. The primary loss function is Mean Squared Error (MSE), expressed as:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \tag{3}$$

Where $\hat{y}_i$ is the predicted carbon value, y_i is the ground-truth label, and N is the total number of samples.

To comprehensively evaluate the performance of the model, four indicators were introduced: root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and structural similarity index (SSIM). These indicators were used to assess the model from multiple dimensions such as error magnitude, fitting ability, and image consistency. Additionally, to test the stability and generalization ability of the model, group experiments were conducted under different types of green spaces (such as parks, forest belts, and affiliated green spaces), different seasons (spring, summer, autumn, winter), and different remote sensing sources (Sentinel-2 and MODIS).

4. Results

4.1. Estimation accuracy and interpretability

In the experiment of carbon sink estimation for urban green spaces, the CNN-TCN deep fusion model proposed in this paper demonstrated excellent performance across multiple evaluation metrics, showcasing high estimation accuracy and stable spatiotemporal modeling capabilities. In the test set, the overall root mean square error (RMSE) of the model was 2.71 tC/ha, the mean absolute error (MAE) was 1.94 tC/ha, the determination coefficient (R^2) was 0.926, and the structural similarity index (SSIM) was 0.841, indicating that the model has a high fitting ability and spatial consistency. The carbon storage dynamic change trend of the model in different periods was highly consistent with the actual vegetation growth cycle in Guangzhou, and it could capture the typical seasonal characteristics of spring growth, summer plateau, and autumn-winter decline.

From the land use classification perspective, there were significant differences in the estimation accuracy of different types of green spaces. Urban park green spaces had the highest prediction accuracy due to their regular shapes and stable vegetation coverage, with an RMSE of 1.63 tC/ha; ecological protection forest belts were relatively continuous in space, with an RMSE of 2.09 tC/ha; while traffic isolation belts had a lower coverage rate, ambiguous boundaries, and mixed with non-green space pixels, resulting in relatively larger errors, with an RMSE of 3.21 tC/ha. Table 1 shows the main evaluation index performance of the model under different green space types.

Table 1. Model performance across different urban green space types

Green Space Type	RMSE (tC/ha)	MAE (tC/ha)	R^2	SSIM
Urban Park Green Space	1.63	1.25	0.945	0.864
Shelterbelt Forest	2.09	1.68	0.918	0.832
Wetland Green Area	2.54	2.03	0.902	0.811
Auxiliary Green Space	2.78	2.22	0.881	0.793
Roadside Isolation Belt	3.21	2.64	0.856	0.751
Overall Average	2.71	1.94	0.926	0.841

4.2. Model comparison

To fully verify the advantages of the CNN-TCN model proposed in this paper in estimating urban green space carbon sequestration, a comparative experiment was designed against three benchmark models: the vegetation index regression model (NDVI-LR), the multiple linear regression model (MLR), and the random forest regression model (RF). The comparison mainly focused on four performance indicators. All tests were conducted under the same remote sensing samples and time periods to ensure fairness and comparability. The results showed that the CNN-TCN model achieved the best performance in all four indicators: RMSE was 2.71 tC/ha, MAE was 1.94 tC/ha, R^2 was 0.926, and SSIM was 0.841. In contrast, NDVI-LR had RMSE and MAE of 3.85 and 2.57 respectively, and R^2 was only 0.794, which was significantly lower than the deep model. While MLR and RF showed some improvement, they still performed inadequately in handling temporal changes and heterogeneous land types. Figure 1 shows the performance change curves of each model under the four indicators. It can be seen that the CNN-TCN model was always in the best performance position, and the improvement on the SSIM curve was particularly significant, indicating its strong advantage in restoring the spatial structure of carbon sequestration.

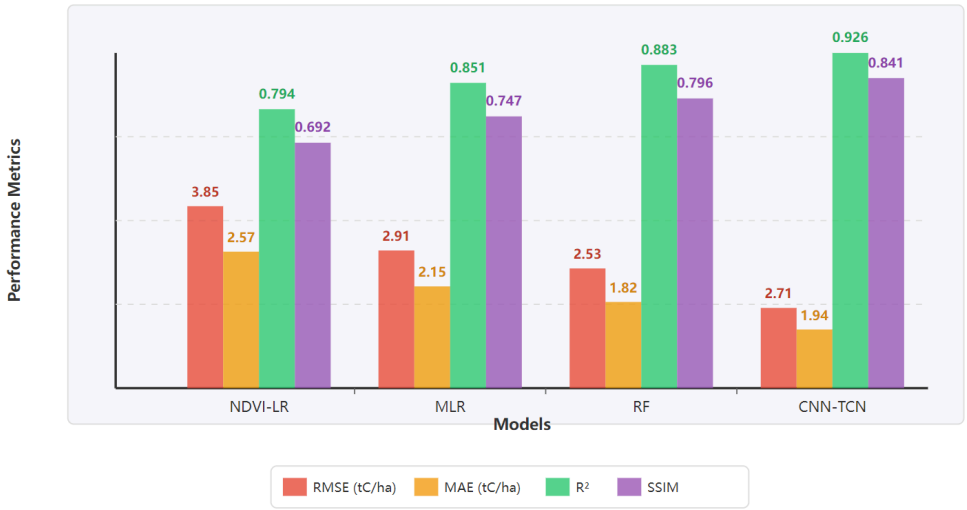

Figure 1. Performance comparison of carbon sequestration estimation models

5. Discussion

The proposed deep spatiotemporal fusion model demonstrates high accuracy, stability, and adaptability in estimating urban green-space carbon sink capacity. Experimental results show that it consistently outperforms traditional statistical and machine learning models in terms of error control (RMSE, MAE), and achieves superior structural similarity (SSIM), indicating enhanced capacity in capturing green-space boundaries and morphological features within heterogeneous urban environments. Particularly in fragmented urban interiors and diverse suburban forest belts, the model effectively recognizes spatial heterogeneity and supports multi-scale ecological planning. However, prediction fluctuations persist in low-coverage or highly disturbed regions such as roadside belts and temporary vegetation zones, suggesting that future work could incorporate higher-resolution imagery (e.g., WorldView-3 or GF-series) for localized refinement. Additionally, although the model is trained on a five-year multi-temporal dataset and shows strong seasonal response, it remains fundamentally data-driven and lacks integration with ecological process models. Future research may explore hybrid frameworks that integrate data-driven learning with mechanistic models

(e.g., CASA or InVEST), enhancing interpretability and decision support for urban ecological management and policy formulation.

6. Conclusion

This study proposes a deep spatiotemporal remote-sensing fusion framework that integrates Convolutional Neural Networks (CNN) and Temporal Convolutional Networks (TCN) to achieve high-precision, dynamic estimation of urban green-space carbon sink capacity. Using Guangzhou as the case study, a multi-source dataset combining Sentinel-2, Landsat-8, and MODIS imagery was constructed. The model leverages end-to-end spatial feature extraction and temporal modeling to significantly improve estimation accuracy and spatial consistency. It demonstrates strong robustness and generalizability across diverse land cover types and seasonal conditions, particularly excelling in structural preservation and dynamic carbon response. Compared to traditional statistical and machine learning methods, the proposed framework achieves superior performance in error control, structural fidelity, and regional adaptability, validating its practical value in complex urban environments. Future applications may include carbon neutrality trajectory simulations, ecological compensation assessments, and fine-scale green infrastructure management, particularly when integrated with ecological process models and policy-oriented parameters to support intelligent carbon asset governance.

References

[1] Wang, R.-Y., Wu, Z., Li, X., Shi, T. M., Liu, X., Chen, L., ... & Ye, Z. (2024). Comparison of the CASA and InVEST models'effects for estimating spatiotemporal differences in carbon storage of green spaces in megacities. *Scientific Reports, 14*(1), 5456.

[2] Wu, Z., Jiang, M., Li, H., Shen, Y., Song, J., Zhong, X., & Ye, Z. (2023). Urban carbon stock estimation based on deep learning and UAV remote sensing: a case study in Southern China. *All Earth, 35*(1), 272–286.

[3] Li, X., Wu, Z., Wang, R.-Y., Huang, J., Pan, B., Zurita, J. V. S., ... & Tuzikov, A. (2024). The impact of landscape spatial morphology on green carbon sink in the urban riverfront area. *Cities, 148*, 104919.

[4] Shi, T. M., Wu, Z., Li, X., Wang, R.-Y., Liu, X., Chen, L., ... & Ye, Z. (2025). Evaluation of Urban Composite Carbon Sink Value: A Case Study of Shenyang. *Landscape Architecture, 32*, 57–66.

[5] Liu, X., Wu, Z., Li, X., Wang, R.-Y., Shi, T. M., Chen, L., ... & Ye, Z. (2025). Research on Carbon Sequestration Capacity of Urban Park Green Space in the Central Urban Area of Beijing and Driving Factors Thereof. *Landscape Architecture, 32*(1), 32–40.

[6] Chen, L., Wu, Z., Li, X., Wang, R.-Y., Shi, T. M., Liu, X., ... & Ye, Z. (2024). Carbon storage estimation and strategy optimization under low carbon objectives for urban attached green spaces. *Science of The Total Environment, 923*, 171507.

[7] Wu, Z., Jiang, M., Li, H., Shen, Y., Song, J., Zhong, X., & Ye, Z. (2023). Urban carbon stock estimation based on deep learning and UAV remote sensing: a case study in Southern China. *All Earth, 35*(1), 272–286.

[8] Li, R., Ye, S., Bai, Z., Nedzved, A., & Tuzikov, A. (2024). Moderate Red-Edge vegetation index for High-Resolution multispectral remote sensing images in urban areas. *Ecological Indicators, 167*, 112645.

[9] Huang, J., Wu, Z., Li, X., Wang, R.-Y., Shi, T. M., Liu, X., ... & Ye, Z. (2024). Carbon Sequestration and Landscape Influences in Urban Greenspace Coverage Variability: A High-Resolution Remote Sensing Study in Luohe, China. *Forests, 15*(11), 1849.

[10] Zurita, J. V. S., Wu, Z., Li, X., Wang, R.-Y., Shi, T. M., Liu, X., ... & Ye, Z. (2024). Economic valuation of carbon sequestration in Quito's Metropolitan Park using Sentinel-2 and neural networks. *Sapienza: International Journal of Interdisciplinary Studies, 5*(3), e24059.

[11] Pan, B., Wu, Z., Li, X., Wang, R.-Y., Shi, T. M., Liu, X., ... & Ye, Z. (2023). A CNN–LSTM machine-learning method for estimating particulate organic carbon from remote sensing in lakes. *Sustainability, 15*(17), 13043.