Front. Earth Sci. 2025, 19(1): 54–67
https://doi.org/10.1007/s11707-024-1127-9

RESEARCH ARTICLE

Evaluation of carbon storage in a metropolitan area based on long-term time series of Land Use/Cover Change dynamics

Luyue TU[1,2], Jiayi PAN (✉)[1,3]

1 School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China
2 Key Laboratory of Poyang Lake Wetland and Watershed Research (Ministry of Education), Nanchang 330022, China
3 Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong 999077, China

Abstract This research focuses on analyzing land use and cover changes in Metropolitan Area of Wuhan between 1988 and 2023, utilizing a comprehensive data set from Landsat remote sensing and machine learning techniques to understand their implications for carbon storage. It finds that the Random Forest (RF) algorithm outperforms others like Support Vector Machine (SVM), Gradient Boosting Trees (GBT), and Classification and Regression Trees in identifying land use types, achieving high accuracy and a Kappa coefficient exceeding 0.98. Significant changes in Wuhan's landscape have been noted, especially the marked decrease in arable land and increase in urban construction, reflecting the pressures of economic development and urban expansion on natural resources and their impact on the ecosystem. The study uses the InVEST model to assess how these land use transformations affect carbon storage, revealing a significant decrease in carbon storage from 1988 to 2023, with a total reduction of approximately 428.59 × 10^4 t from 1988 to 2023, largely attributed to the conversion of key carbon sequestering lands such as arable lands and forests into urban areas. This transition, particularly from arable land to urban construction land, underscores the challenges faced in managing land use changes without compromising environmental sustainability and carbon storage capacities.

Keywords carbon storage, machine learning methods, InVEST model, land cover/use

1 Introduction

Amidst intensifying global climate change, the continuing rise in atmospheric carbon dioxide concentrations has triggered significant global warming, profoundly impacting ecosystem stability and biodiversity conservation. In response to this global challenge, there is an increasing recognition of the imperative to enhance the carbon storage capacity of Earth's ecosystems and reduce carbon emissions as key strategies to combat climate change. Terrestrial ecosystems, as critical components of the global carbon cycle, mitigate the pace of climate change effectively by absorbing atmospheric carbon dioxide (Tao et al., 2001). However, land use and cover changes, as direct manifestations of terrestrial ecosystem alterations, have a direct impact on major carbon storage carriers such as aboveground biomass, belowground biomass, soil, and dead organic matter (Sharp et al., 2018).

Numerous studies have revealed significant differences in the carbon sequestration capacities of different land use types. These variations directly influence the distribution of vegetation and soil, thereby affecting the carbon storage of terrestrial ecosystems (Li et al., 2020). For example, between 1980 and 2010, land use changes led to a loss of approximately 279 × 10^6 t of carbon in China's terrestrial ecosystems (Zhang et al., 2015). Internationally, reports by the Intergovernmental Panel on Climate Change (IPCC) highlight that land use changes contribute about 1.5 × 10^9 t of carbon emissions to the atmosphere annually, profoundly impacting the global terrestrial ecosystem carbon cycle (Shukla et al., 2019).

As we navigate a critical transition in global carbon emission management, the urgency for coordinated and effective strategies and technologies to achieve 'carbon peak and carbon neutrality' goals intensifies. Understanding carbon reduction and sequestration across various ecosystems is crucial. Urban ecosystems, in particular, pose a unique challenge due to limited studies and a lack of comprehensive analysis regarding spatial heterogeneity

Received February 8, 2024; accepted April 13, 2024

E-mail: panj@cuhk.edu.hk

and interactions within these areas. Consequently, accurately assessing the impact of land use changes on carbon storage and examining the spatiotemporal dynamics of carbon storage in urban ecosystems are essential. Such analysis will serve as crucial reference points for precisely calculating ecosystem carbon storage, aiding in the development of informed management strategies.

In recent years, advances in machine learning and deep learning technologies have significantly enhanced their application in the field of land use classification, particularly in processing extensive data sets and intricate scenarios, where these technologies have shown remarkable efficacy. Methods such as Support Vector Machines (SVM), Classification and Regression Trees (CART), Gradient Boosting Regression Trees (GBT), and Random Forest (RF) have been commonly employed. Nguyen et al., (2020) conducted an evaluation involving a parametric classifier (logistic regression) alongside three non-parametric machine learning classifiers—namely, an enhanced *k*-nearest neighbors, random forest, and support vector machine—for the purpose of classifying multi-temporal Sentinel 2 satellite imagery within Dak Nong Province, Vietnam, into Land Use and Land Cover (LULC) categories. The study reported an Overall Accuracy (OA) ranging from 63.9% to 80.3%, with SVM achieving the highest precision. Furthermore, Singh and Pandey (2021) utilized algorithms such as maximum likelihood, random forest, and support vector machine to investigate the LULC conditions within the Alaknanda River basin, located in the north-western Himalayan region of India. Their findings suggested that, in mountainous basins characterized by substantial elevation variances and steep terrains, classification algorithms based on machine learning are superior to traditional statistical classification methods.

The estimation of carbon storage serves as a pivotal methodology for quantifying the carbon sequestration and storage capacities within ecosystems, playing a significant role in evaluating carbon cycles and devising climate change mitigation strategies. Traditional methodologies, such as in situ sampling and carbon flux observation, are predominantly confined to small-scale investigations and are characterized by high costs and labor intensiveness. The InVEST model, through the synthesis of existing Geographic Information System (GIS) data and streamlined methods for assessing ecosystem services, offers advantages including the ease of parameter acquisition and the visualization of results. This model facilitates rapid and cost-effective estimations of carbon storage across expansive spatial scales. Kohestani et al. (2023) have integrated the InVEST model with Markov chain analysis to map and evaluate the implications of landscape transformations within the mountainous river basins of northern Iran on carbon storage capacities. Furthermore, Zhao et al. (2018) employed the InVEST model, utilizing MCD12Q1 land classification data, to model the carbon storage and its driving factors within the Tibetan Plateau region. Piyathilake et al. (2021) adopted the InVEST model for ecosystem services and trade-offs to model and estimate carbon storage in Sri Lanka's Uva Province. Their findings indicate that the designated area, covering 850000 ha, presently encompasses an estimated 96.63 billion tons of carbon, with discernible disparities across various land use categories.

Contemporary studies have predominantly concentrated on ecological conservation regions such as watersheds, wetlands, and coastal areas, with urban territories receiving comparatively less scrutiny. Nonetheless, given their role as catalysts for national economic advancement and as pivotal areas for new urbanization initiatives, the significance of urban regions in ecosystem preservation is indisputable. As a key node of the Yangtze River Economic Belt, Wuhan's role in the central rise strategy of China is pivotal. Thoroughly analyzing the impact of regional land use under urban development in Wuhan not only provides a basis for local ecological restoration and ecological safety pattern optimization but also contributes significantly to the green development of the Middle Yangtze River Economic Belt. Zhang et al. (2022) employed the FLUS coupled model to project land use change scenarios within the urban agglomeration under the "Three Lines" policy framework and applied the InVEST model to quantitatively assess the impacts of land use alterations in Wuhan City on carbon storage across various scenarios. The findings indicated a consistent decline in the total carbon storage within the Wuhan urban agglomeration from 2000 to 2015, primarily attributed to the transition from farmland to built-up areas. Lu (2023) leveraged land use data spanning five periods to conduct a dynamic evaluation of wetland utilization changes in the Wuhan urban agglomeration between 2000 and 2020, utilizing the InVEST model to investigate shifts in wetland ecosystem services. Moreover, Zhang et al. (2023) utilized the Markov-FLUS coupled model to forecast land use transformations in Huangpi District of Wuhan by 2035 under scenarios of natural development, economic prioritization, and comprehensive development, employing the InVEST model to quantitatively explore the implications of land use variations on carbon storage under these scenarios. Present research primarily explores the ecosystem service values within the Wuhan wetland ecosystem or projects variations in carbon storage within the Wuhan urban agglomeration under specific future scenarios, often failing to elucidate the dynamic interplay between land use changes and the evolution of carbon storage over extended temporal sequences.

Focusing on Wuhan metropolitan area, Therefore, this study selects eight phases of Landsat series remote sensing imagery data from 1988 to 2023 as the basis, applies the GEE platform, and combines various machine

learning algorithms such as Random Forest, Support Vector Machine, Gradient Boosting Tree, and Decision Tree to interpret the remote sensing imagery, obtaining long-term land use classification data for the Wuhan area. By appropriately selecting landscape indices to analyze land use changes in Wuhan metropolitan area and using the carbon storage module of the InVEST model combined with the GIS platform, this study assesses the spatiotemporal differentiation characteristics of regional carbon storage. The purpose of this study is to explore the impact of land use changes in Wuhan on ecosystem carbon storage, providing a scientific basis for enhancing land use efficiency, promoting the coordinated development of the ecological economy, and improving carbon storage levels in the metropolitan area.

2 Study area and data

2.1 Study area

Wuhan metropolitan area, situated in the heartland of central China, occupies a crucial geographical position in the eastern part of the Jianghan Plain and serves as the political, economic, and cultural hub of Hubei Province of China (Geographic coordinates: 29°58′N to 31°22′N, 113°41′E to 115°05′E, Fig. 1). The area is divided by the confluence of the Chang Jiang (Yangtze R.) and Han Jiang, which imparts a unique character to its urban ecological environment and land use patterns. Wuhan governs 13 districts, covering a total land area of 8569.15 km^2, and it has a permanent population of approximately 13.739 million. In 2022, the city's Gross Domestic Product (GDP) reached 1886.643 billion yuan, reflecting its central role in regional economic development (Zhang et al., 2017).

The land use characteristics of Wuhan metropolitan area predominantly feature urbanized areas and extensive agricultural lands, surrounded by various forests and wetlands (Huang et al., 2017). These topographic and land cover types play a pivotal role in the region's carbon storage capacity and carbon cycling processes. Wuhan experiences a subtropical monsoon climate, with an average annual temperature ranging between 15.8°C and 17.5°C and annual precipitation between 1150 and 1450 mm (Yan et al., 2017).

2.2 Research data

This study uses Landsat TM/ETM + series remote sensing imagery data from eight autumn and winter seasons, specifically for the years 1988, 1993, 1998, 2003, 2008, 2013, 2018, and 2023, to extract land use types in Wuhan from 1988 to 2023, as shown in Table 1.

The satellite images are sourced from Google Earth Engine (GEE) platform. Landsat 8 OLI is outfitted with a suite of nine bands. This array comprises eight spectral bands, each possessing a spatial resolution of 30 m, alongside a singular panchromatic band, which boasts a finer spatial resolution of 15 m. The instrument captures a swath width of 185 by 185 km. Conversely, the Landsat 5 Thematic Mapper (TM) features a total of seven bands,

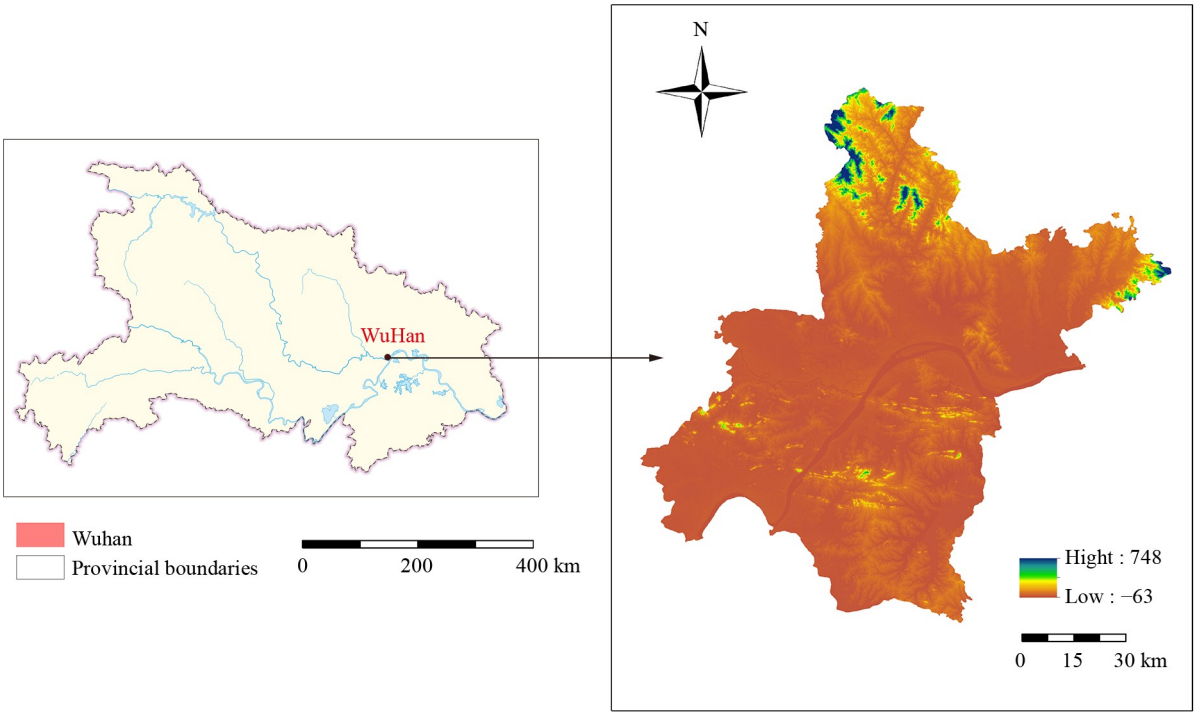

Fig. 1 The geographical location of Wuhan.

Table 1 Landsat images used in this study

Year	Sensor	Resolution/m
1988	Landsat 5 TM	30
1993	Landsat 5 TM	30
1998	Landsat 5 TM	30
2003	Landsat 5 TM	30
2008	Landsat 5 TM	30
2013	Landsat 8 OLI	30
2018	Landsat 8 OLI	30
2023	Landsat 8 OLI	30

encompassing six bands with a spatial resolution of 30 m and one thermal infrared band with a spatial resolution of 120 m, as listed in Table 1.

3 Research method

3.1 Feature construction

3.1.1 Spectral features

Index Features involve selecting appropriate combinations of spectral bands for mathematical operations to create indices that provide a more stable and pronounced representation of land cover than single-band data. To better represent the characteristics of various land covers such as farmland, forest land, grassland, water bodies, urban land, and unused land, the following indices have been selected: Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Modified Normalized Difference Water Index (MNDWI), Normalized Difference Built-up Index (NDBI), and Enhanced Vegetation Index (EVI). These indices are effective in identifying the fundamental types of land surface.

NDVI is a critical spectral index in remote sensing science, dedicated to the quantitative assessment of vegetation biomass and growth conditions on the Earth's surface. The NDVI calculation employs a method that involves measuring and analyzing the difference and ratio between the surface reflectance in Near-Infrared (NIR) and Red light wavelengths (Red) (Tian and Min, 1998), given by

$$\mathrm{NDVI}=\frac{\mathrm{NIR}-\mathrm{Red}}{\mathrm{NIR}+\mathrm{Red}}. \tag{1}$$

The index values, which span from −1 to +1, signify the health and density of green vegetation when approaching +1. In contrast, values nearing 0 or negative values are indicative of barren lands or aquatic environments.

MNDWI is utilized for identifying water bodies in satellite imagery. This index capitalizes on the spectral reflectance differences between the green and mid-infrared bands to maximize the contrast between water bodies and other land cover types (Xu, 2005):

$$\mathrm{MNDWI}=\frac{(\mathrm{P(Green)}-\mathrm{P(SWIR)})}{(\mathrm{P(Green)}+\mathrm{P(SWIR)})}, \tag{2}$$

where Green pertains to the specific green wavelength values (such as the B1 band for Landsat 5 and B2 band for Landsat 8), while SWIR denotes the Short-Wave Infrared wavelength values (like SWIR B5 for Landsat 5 and SWIR B6 for Landsat 8), with P representing a constant.

NDBI serves as an essential tool within remote sensing for the delineation and surveillance of urban infrastructure and built-up areas. The NDBI's computation is primarily based on the disparity in reflectance observed between the Short-Wave Infrared (SWIR) and Near-Infrared (NIR) wavelengths (Zha et al., 2003), as detailed in the formula,

$$\mathrm{NDBI}=\frac{\mathrm{SWIR}-\mathrm{NIR}}{\mathrm{SWIR}+\mathrm{NIR}}. \tag{3}$$

Reflectance characteristics of the SWIR wavelength are notably responsive to construction materials such as concrete and brick, whereas NIR wavelength reflectance is more sensitive to vegetative surfaces, resulting in positive NDBI values for urban areas and negative values for vegetated zones.

EVI constitutes an advanced development in spectral index technology, engineered to refine vegetation signal extraction and bolster the precision and dependability of vegetation monitoring efforts (Huete et al., 2002). As an improvement over the NDVI, the EVI leverages data from the NIR, Red, and Blue wavelengths, incorporating atmospheric correction and soil background suppression strategies to mitigate the effects of atmospheric dispersion and soil background noise, given by

$$\mathrm{EVI}=2.5\times\frac{(\mathrm{NIR}-\mathrm{Red})}{(\mathrm{NIR}+6\times\mathrm{Red}-7.5\times\mathrm{Blue}+1)}. \tag{4}$$

3.1.2 Texture features

The utilization of texture features, stemming from the unique texture characteristics of different land cover types, can significantly enhance the accuracy of identification and classification (Haralick et al., 1973). The Gray-Level Co-occurrence Matrix (GLCM), a classic and efficient method for texture extraction, has been widely applied in land cover classification studies. The computation of GLCM is executed through specialized coding. In this study, a range of commonly used GLCM texture features were selected to assist in the classification process. These include the Angular Second Moment of NDVI, contrast, correlation, variance, Inverse Difference Moment, average gray level, and entropy.

3.1.3 Terrain characteristics and sample points

Incorporating terrain features significantly enhances the precision of classification in remote sensing image analysis. This study utilizes the SRTMGL1_003 Digital Elevation Data product to compute key terrain parameters, including aspect, hillshade, slope, and elevation. These parameters, treated as independent features, are integrated into the remote sensing image classification analysis, thereby improving the sensitivity and accuracy of the model in detecting surface changes.

This study focuses on selecting representative and pure pixels as sampling points, uniformly chosen from land satellite images accessed through the GEE platform. To facilitate effective model training and validation, these sample points are randomly allocated, following a 7:3 ratio for dividing into training and testing sets. Subsequently, classification analysis is conducted using Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Tree (GTB), and Classification and Regression Tree (CART) methods. This process generates confusion matrices, evaluates overall accuracy, and calculates the Kappa coefficient, providing key assessment metrics for the study.

3.2 Machine learning methods

In this study, four advanced machine learning algorithms, namely Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Trees (GBT), and CART—were employed for the detailed classification of land use types in Wuhan city.

3.2.1 Random Forest

The Random Forest, an ensemble machine learning technique, achieves accurate classification decisions by amalgamating numerous weak classifiers, such as classification and regression trees, into a robust model (Breiman, 2001). Its essence lies in randomly selecting subsets of samples and features from the entire data set to construct each weak classifier. This approach not only reduces noise but also ensures effective model fitting. During the construction of a Random Forest, each weak classifier flexibly determines its node splits based on randomly selected features, thereby enhancing the model's generalization capability across diverse data sets.

A distinctive feature of Random Forest is the integration of independent classifiers, which adeptly handles nonlinear and multidimensional features, making it particularly suitable for solving complex classification problems. When building a Random Forest, two key parameters must be defined: the number of classification trees and the number of predictor variables for node splitting. While having multiple classification trees may increase computational complexity, it also significantly improves classification performance. In the implementation on the GEE platform for this study, the number of classification trees was set to 20, with a bagging fraction of 0.8, and other parameters were left at their default settings.

3.2.2 Support Vector Machines

Support Vector Machine (SVM) is an innovative machine learning method founded on statistical learning theory (Vapnik, 1995). It is particularly suited for handling nonlinear problems with small sample sizes. The application of SVM is extensive, encompassing a variety of problems in both classification and regression. Its core principle involves mapping data that is originally nonlinearly separable into a higher dimensional space and then effectively dividing it using a hyperplane. This process optimizes the model's generalization ability by seeking to minimize structured risk, thereby enabling the extraction of reliable statistical patterns, even in situations with limited sample sizes.

3.2.3 Gradient Boosting Trees

Gradient Boosting Trees (GBT) is an iterative ensemble learning method based on constructing multiple decision trees (Friedman et al., 2000). This approach combines the predictions of all participating decision trees by approximating the residuals of the current model based on the negative gradient of the loss function, thus yielding the final prediction. GBT algorithm demonstrates flexibility in handling various types of data, whether they are continuous or categorical. Compared to Support Vector Machines (SVM), GBT typically achieves higher prediction accuracy with fewer parameter adjustments. Moreover, when specific loss functions like Huber and Quantile are introduced, the algorithm exhibits greater robustness in handling outliers. To achieve optimal classification performance, we set up a foundation of 20 classification trees as the basis for the GBT model in our research.

3.2.4 Classification and Regression Trees

Classification and Regression Trees (CART) is a classic algorithm designed for both classification and regression tasks, particularly tailored for handling continuous data. This method encompasses the processes of feature selection, tree construction, and pruning using binary split techniques (Breiman et al., 1984). It divides the data set into two subsets, ensuring that each intermediate node ultimately differentiates into two branches. CART method handles various types of features, whether they are categorical or continuous, through binary partitioning. In this process, if a categorical feature with three attributes can be split into two attributes, the third

attribute is assigned to the same category. However, it is worth noting that the CART algorithm tends to allow trees to grow naturally during decision tree construction, which may lead to overfitting. Therefore, adopting appropriate pruning strategies is an effective means to mitigate overfitting.

3.3 Accuracy evaluation methods

To accurately assess the performance of the classification models in this study, we have selected the Kappa coefficient and Overall Accuracy (OA) as the primary evaluation metrics (Yi and Zhang, 2012). The Kappa coefficient is a crucial statistical measure that quantifies the level of consistency in the classification results, taking into account the chance agreement produced by random classification, given by

$$K = \frac{p_o - p_e}{1 - p_e}, \tag{5}$$

where p_o is the observed agreement among raters (i.e., the overall accuracy), and p_e is the expected agreement by chance.

On the other hand, Overall Accuracy (OA), serving as a quantifiable standard for evaluating the precision of remote sensing classification, directly reflects the accuracy of land cover classification, as detailed in the formula, given by

$$\mathrm{OA} = \frac{\sum_{i=1}^{n} x_{ii}}{N}, \tag{6}$$

where X_{ii} represents the number of correctly classified instances for each category i, n is the total number of categories, and N is the total number of instances in the data set.

Together, these two evaluation methods offer a comprehensive perspective for assessing the model's performance across various categories and its overall classification accuracy.

3.4 InVEST Model's carbon storage and sequestration module

The InVEST model was developed by Stanford University with the Nature Conservancy (TNC) and the World Wildlife Fund (WWF) (Yang et al., 2012). This model is specifically designed for quantifying and assessing various ecosystem service functions. In particular, the Carbon Storage and Sequestration module of the InVEST model utilizes land-use data to estimate carbon storage in specific landscapes or carbon sequestration over specific time periods. It achieves this by considering various factors, including aboveground biomass, belowground biomass, soil carbon stocks, and dead organic matter.

The carbon storage in the InVEST is calculated by

$$C_i = C_{i-\mathrm{above}} + C_{i-\mathrm{below}} + C_{ci-\mathrm{dead}} + C_{i-\mathrm{soil}}, \tag{7}$$

and

$$C_{i-\mathrm{total}} = C_i \times A_i, \tag{8}$$

where C_i represents the different carbon density parameters for land class i, including total carbon density, aboveground carbon density, belowground carbon density, dead organic matter carbon density, soil organic matter carbon density, and total carbon storage. Additionally, A_i represents the corresponding area for land class i. The determination of carbon density for each land class (Table 2) is detailed in the reference cited. Here, i ranges from 1 to 6, corresponding to land classes such as cropland, water bodies, forests, grasslands, and unused land types. The InVEST model's carbon storage component necessitates data comprising: extant land use information, formatted as TIFF files where each land use category is designated by a unique numerical code; and carbon density information, formatted as CSV files. The model designates the minimum grid of each land use category as the assessment unit, facilitating the calculation of carbon storage for each carbon pool through the input of respective carbon density data. Subsequently, the model ascertains the total carbon storage for each map unit and the aggregate study area by aggregating the carbon storage across all carbon pools.

4 Results and analysis

4.1 Comparison of machine learning methods

Utilizing the Google Earth Engine (GEE) platform, synthesized Landsat OLI imagery for the autumn and winter of 2023 in Wuhan metropolitan area was subjected to interpretative classification, with the resulting overall accuracy and Kappa coefficient detailed in Table 3. Upon comparison of four machine learning methodologies, it is

Table 2 Carbon density of each LULC type in Wuhan (C_a—aboveground carbon density, C_b—belowground carbon density, C_s—soil organic carbon density, and C_d—dead organic carbon density)

LULC Types	Carbon Density/(t·km^{-2})			
	C_a	C_b	C_s	C_d
Cropland	402	75	211	9813
Forest	2262	103	278	12675
Grassland	360	1170	728	9043
Water	159	0	398	6403
Built-up land	83	8	0	4371
Other	59	64	96	2842

Table 3 Overall accuracies and Kappa coefficients of the four machine learning algorithms for the Landsat image of Oct 2023

Algorithms	Overall accuracy	Kappa coefficient
RF	0.99	0.99
SVM	0.97	0.96
GTB	0.95	0.93
CART	0.94	0.92

determined that the precision of each remote sensing classification algorithm in discerning and categorizing land cover types within Wuhan metropolitan area was comparatively high, with all methodologies achieving an overall classification accuracy above 94% and Kappa coefficients above 92%. Among them, the Random Forest algorithm was identified as the most effective, achieving both classification accuracy and Kappa coefficient scores of 99%.

The confusion matrices reflecting the classification outcomes of the four machine learning methodologies are depicted in Table 4. It is demonstrated that the Random Forest (RF) method achieved the smallest overall classification error. In comparison with the other three machine learning approaches, the RF method was characterized by a reduced incidence of pixel mixing among different land categories, exhibiting superior performance in the classification of cropland, where instances of mixed pixels were not detected.

4.2 Long-term land use change in Wuhan metropolitan area

After conducting an analysis of accuracy comparisons among four machine learning methods and scrutinizing confusion matrices, the Random Forest approach was ultimately selected for interpreting Landsat imagery of the Wuhan metropolitan area spanning from 1988 to 2023. Table 4 demonstrates that both the overall accuracy and Kappa coefficient of the Random Forest method exceeded 0.99, justifying its utilization for further exploration.

Through the analysis of long-term land use imagery of Wuhan metropolitan area from 1988 to 2023 (Fig. 2), the transformation of urban land use patterns over a span of 35 years can be vividly observed. The expansion of urban construction land, originating from the city center and

Table 4 Confusion matrices of four machine learning algorithms for the Landsat image of Oct 2023

Algorithms	Land type	Cropland	Forest	Grassland	Water	Build–up land	Unused land
RF	Cropland	67	0	0	0	0	0
	Forest	0	679	0	2	0	2
	Grassland	0	1	94	0	0	0
	Water	0	2	0	442	0	0
	Build-up land	0	0	0	0	2503	2
	Unused land	1	2	0	0	0	326
SVM	Cropland	63	0	0	0	0	4
	Forest	1	674	4	4	0	0
	Grassland	0	0	90	5	0	0
	Water	0	0	0	444	0	0
	Build-up land	1	0	0	0	2503	1
	Unused land	0	0	0	0	1	328
GTB	Cropland	62	0	1	0	0	4
	Forest	0	679	1	2	0	1
	Grassland	0	6	89	0	0	0
	Water	0	2	0	442	0	0
	Build-up land	0	1	0	0	2503	1
	Unused land	3	2	0	0	0	324
CART	Cropland	63	0	0	0	0	4
	Forest	3	668	0	8	0	4
	Grassland	0	1	94	0	0	0
	Water	0	4	0	440	0	0
	Build-up land	0	0	0	0	2504	1
	Unused land	4	0	0	0	2	323

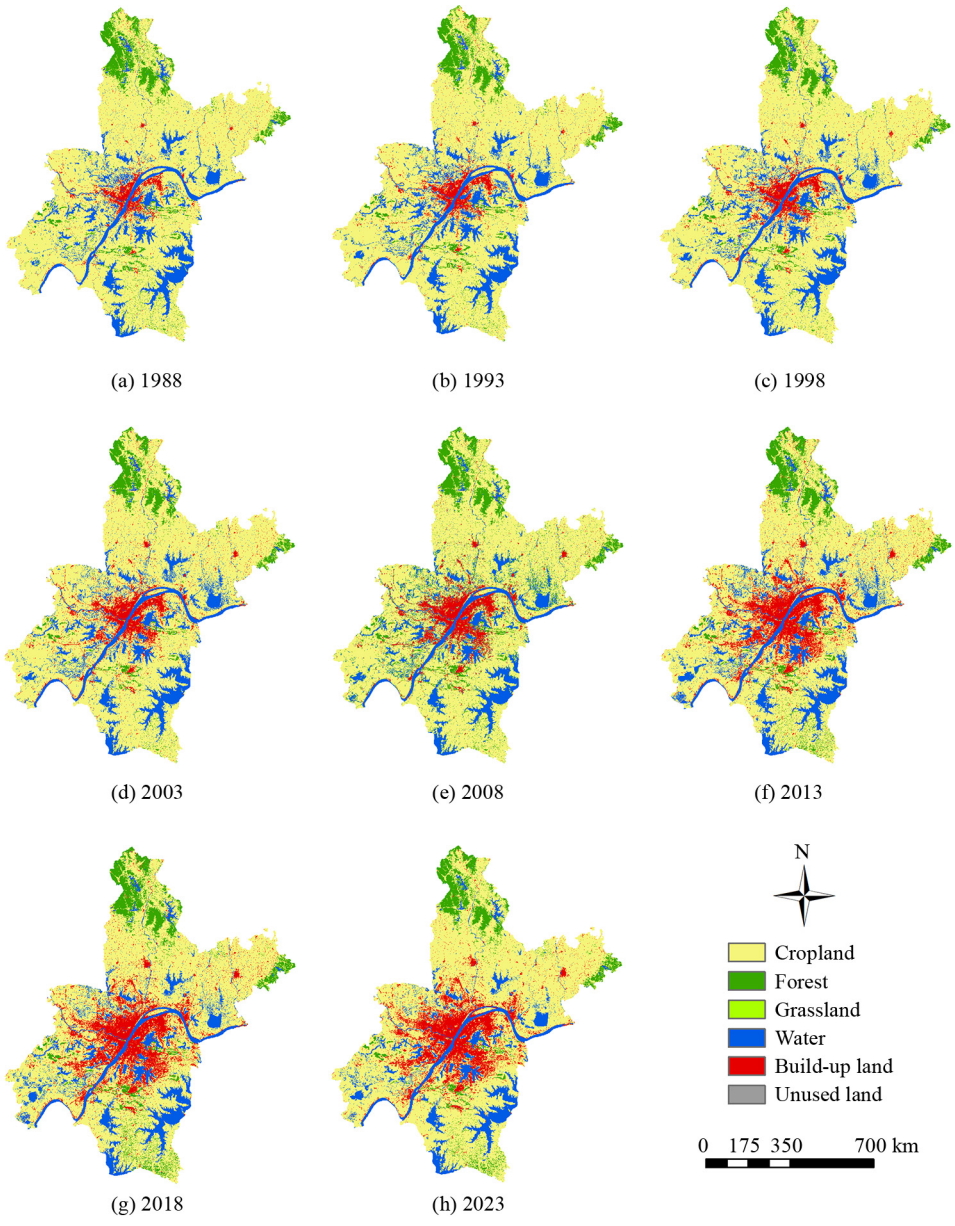

Fig. 2 Wuhan metropolitan area land types obtained from Landsat images with the random forest method.

radiating outward along the major transportation routes and the extensive Yangtze River system, exhibited a pronounced radial growth.

Croplands, serving as the primary land type within the city's boundaries, were initially prevalent around the outskirts of the city center but have been progressively encroached upon by the expansion of new built-up areas as urbanization advanced. Wuhan, celebrated as the "City of a Hundred Lakes", features rivers and lakes that create a network linking urban development zones with agricultural lands. Meanwhile, forest resources are predominantly found in the northern and south-eastern regions of the metropolitan area, with urban parks and the vicinity of lakes interspersed with patches of greenery, adding ecological and natural hues to the urban landscape.

As depicted in Fig. 3, significant changes in the proportions of different land use types were observed in Wuhan City from 1988 to 2023. Table 3 meticulously documents the dynamics in the area of each land type throughout these 35 years. In this timeframe, net increases were exhibited by urban construction lands and forest lands. Notably, the area designated for construction witnessed a significant enlargement by 811.57 km^2, with an average annual expansion of about 23.19 km^2, causing its percentage to skyrocket from 3.37% to 13.17%.

Despite the fluctuating growth of forest lands, an overall ascending trend was evident, with their area expanding from 592.32 km^2 to 678.47 km^2 and their proportion enhancing from 6.9% to 7.9%. Meanwhile, the areas allocated to cropland, grassland, water bodies, and unused lands were characterized by an overall reduction.

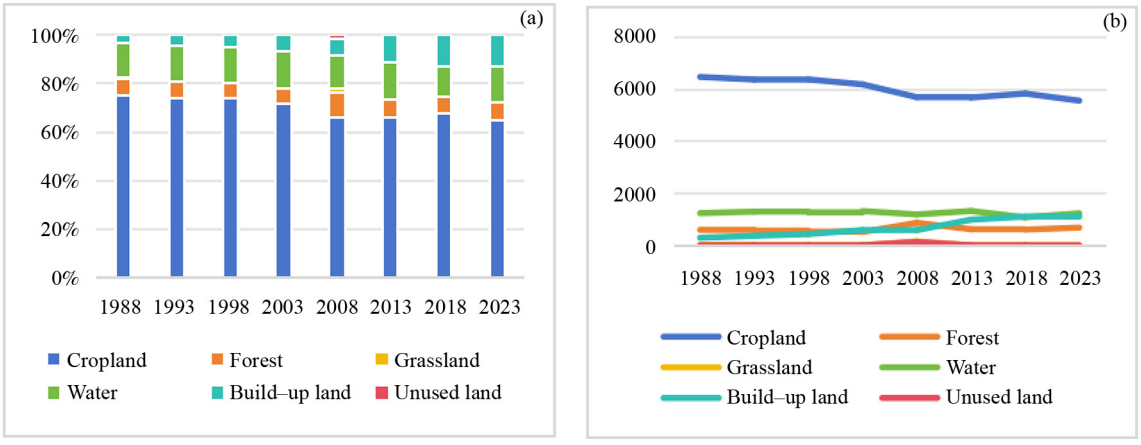

Fig. 3 Percentage (a) and changes (b) of Wuhan LULC types from 1988 to 2023.

The decrease in cropland was notably pronounced, dwindling from 6450.97 km^2 to 5815.21 km^2. Despite the yearly reduction, cropland maintained its status as the predominant land use category in Wuhan metropolitan area, holding an average proportion of 69%.

Table 5 illustrates the matrix of conversions among land cover types within Wuhan metropolitan area spanning from 1988 to 2023, revealing that the primary changes in land cover involved the conversion from cropland to both construction land and water bodies, in addition to the conversion from forest to cropland. Throughout this 35-year period, the conversion from cropland to construction land was observed to be the most extensive, reaching a total of 266.81 km^2, followed by cropland converting to water bodies, with a cumulative area of approximately 261.96 km^2. Meanwhile, the transformation of cropland into forest land was comparatively less significant, amounting to a total of 44.1747 km^2. It is noteworthy that portions of water bodies and forests were also transformed into cropland, corresponding to approximately 160.07 km^2 and 125.52 km^2, respectively. These transitions signify profound alterations in the distribution of land resources in Wuhan metropolitan area, with the significant conversion of cropland to construction land and water bodies highlighting the trends toward urban expansion and aquatic resource management.

Over the last 30-five years, significant changes in land use have been witnessed in Wuhan metropolitan area, marked by a substantial reduction in the areas of cropland and grassland, and a considerable expansion of urban construction land, with an increase in forest land as well. These transformations reflect the swift urbanization process and the growing demand for urban construction land in Wuhan. Serving as the economic nexus of central China, Wuhan metro has been undergoing sustained urbanization and expansion of its construction land. The expansion of land designated for residential, commercial, and industrial purposes has been propelled by urban development strategies. Particularly noteworthy is the period following 2008, during which Wuhan's economic shift from traditional sectors to high-tech sectors spurred the construction of high-tech industrial parks, thereby driving the demand for construction land. This increase in construction land is intimately associated with the realignment of Wuhan metro's economic framework, illustrating its strategic shift toward the cultivation of high-tech industries.

4.3 Long-term temporal changes in carbon storage in Wuhan metropolitan area

Estimations conducted using the InVEST model have indicated that carbon storage in Wuhan City was characterized by an overall downward trend from 1988 to 2023. It was observed that carbon storage diminished

Table 5 Land type transfer matrix from 1988 to 2023

Area conversion/km^2	Cropland	Forest	Grassland	Water	Build–up land	Unused land
Cropland	5876.90	44.17	1.11	261.96	266.81	0.01
Forest	125.52	462.06	0.29	1.13	3.32	0.00
Grassland	2.86	5.73	1.51	0.71	1.44	0.12
Water	160.07	0.50	0.05	1044.11	28.18	0.24
Build-up land	0.43	0.00	0.00	3.67	284.90	0.00
Unused land	0.13	0.00	0.31	0.26	0.35	0.54
Total	6165.91	512.47	3.27	1311.84	585.00	0.91

from 8783.87 × 10^4 t in 1988 to 8355.27 × 10^4 t by 2023, culminating in a net decrease of 428.6 × 10^4 t. This period saw carbon storage undergoing two distinct phases: a sustained reduction was recorded from 1988 to 2003, with a total decrease of 258.9 × 10^4 t, equating to a drop of approximately 2.9%. Conversely, from 2003 to 2008, a brief uptick in carbon storage was noted, with an increase of roughly 195.49 × 10^4 t, translating to a growth of 2.3% (Table 6).

Regarding different types of land use, it has been observed that carbon storage within various land categories in Wuhan metropolitan area underwent changes throughout the study period. It is generally observed that the carbon storage in croplands has demonstrated a trend of overall decline, with a decrease from 6774.16 × 10^4 t in 1988 to 5748.96 × 10^4 t in 2023, with the period from 1998 to 2008 experiencing the most rapid rate of decline. Excluding minor increases during 1993−1998 and 2013−2018, a net reduction in carbon storage was recorded in cropland for the remaining years. Forest land's carbon storage, conversely, exhibited an overall increase, marked by a substantial rise and subsequent decrease between 2003 and 2013, alongside relatively minor variations in other intervals, resulting in a net increment of 61.83 × 10^4 t from 1988 to 2003. Carbon storage in grasslands and unused lands showed a declining trend, except for a slight increment in 2008, with net reductions of 12.88 × 10^4 t and 0.3 × 10^4 t, respectively. Built-up land's carbon storage was observed to increase, from 128.95 × 10^4 t in 1988 to 606.66 × 10^4 t in 2023, averaging an annual growth of 13.65 × 10^4 t. The carbon storage in water bodies displayed relatively stable inter-annual variations, with a total increase of 70.26 × 10^4 t.

The spatial distribution of carbon storage in the Wuhan region is characterized by notable heterogeneity, as depicted in Fig. 4, with a distribution pattern that is intricately linked to the land use types within the area. Through an examination of detailed annual distribution maps, it is observed that areas under forest and cultivation, especially the forested regions in the north and the wetland areas along the river, are associated with higher levels of carbon storage. Owing to the rich vegetation and water resources, the ecosystems in these regions demonstrate enhanced capabilities for carbon sequestration. In contrast, areas utilized for urban development, grasslands, and those left unused in the central urban district are marked by lower carbon storage capacities. This pattern underscores the conversion of natural ecological spaces as a result of urban sprawl, thereby highlighting the critical role of preserving green spaces in the context of urban planning.

In this research, changes in carbon storage in Wuhan City from 1988 to 2023 were categorized into significant reductions, no significant change, and significant increases (Fig. 5). Over the course of this period, the overall fluctuation in carbon storage within Wuhan metropolitan area is observed to be relatively steady, with the predominance of areas showing no significant change, comprising 82.38%. Although the regions experiencing an increase in carbon storage represented a mere 4.15%, these were predominantly situated in suburban forests and newly developed areas, reflecting the beneficial effects of vegetation recovery and urban greening initiatives. In contrast, areas experiencing significant reductions were primarily located in the urban core, making up 13.47% of the study area, where the most notable shifts in carbon storage occurred. This was largely due to the area being a hub of economic activity, where extensive construction activities led to a significant loss of original vegetation, subsequently reducing its carbon storage potential. This changing pattern underscores the distinctive role of urban central regions in the carbon cycle and emphasizes the importance of ecological and environmental considerations in future urban planning strategies.

4.4 Impact of land use changes in Wuhan on carbon storage

The areas and carbon densities associated with various land use types in Wuhan metropolitan area have been identified as significant factors influencing regional carbon storage. Between 1988 and 2023, land use changes in Wuhan led to a reduction of 907.95 km^2 in

Table 6 Statistics of carbon storage during 1988–2023 in Wuhan

LULC	Cropland	Forest	Grassland	Water	Build-up land	Unused land	Total
1988	6774.16	1008.02	13.98	858.27	128.95	0.49	8783.87
1993	6669.44	959.60	8.47	898.92	162.28	0.47	8699.19
1998	6673.63	889.94	4.63	880.31	192.36	0.37	8641.23
2003	6474.82	872.11	3.69	913.04	261.03	0.28	8524.98
2008	5961.72	1467.01	169.34	824.08	259.19	39.13	8720.46
2013	5946.48	1050.29	2.01	915.45	438.25	0.06	8352.54
2018	6004.71	986.80	0.50	767.47	574.24	0.17	8333.90
2023	5748.96	1069.85	1.10	928.53	606.66	0.18	8355.27

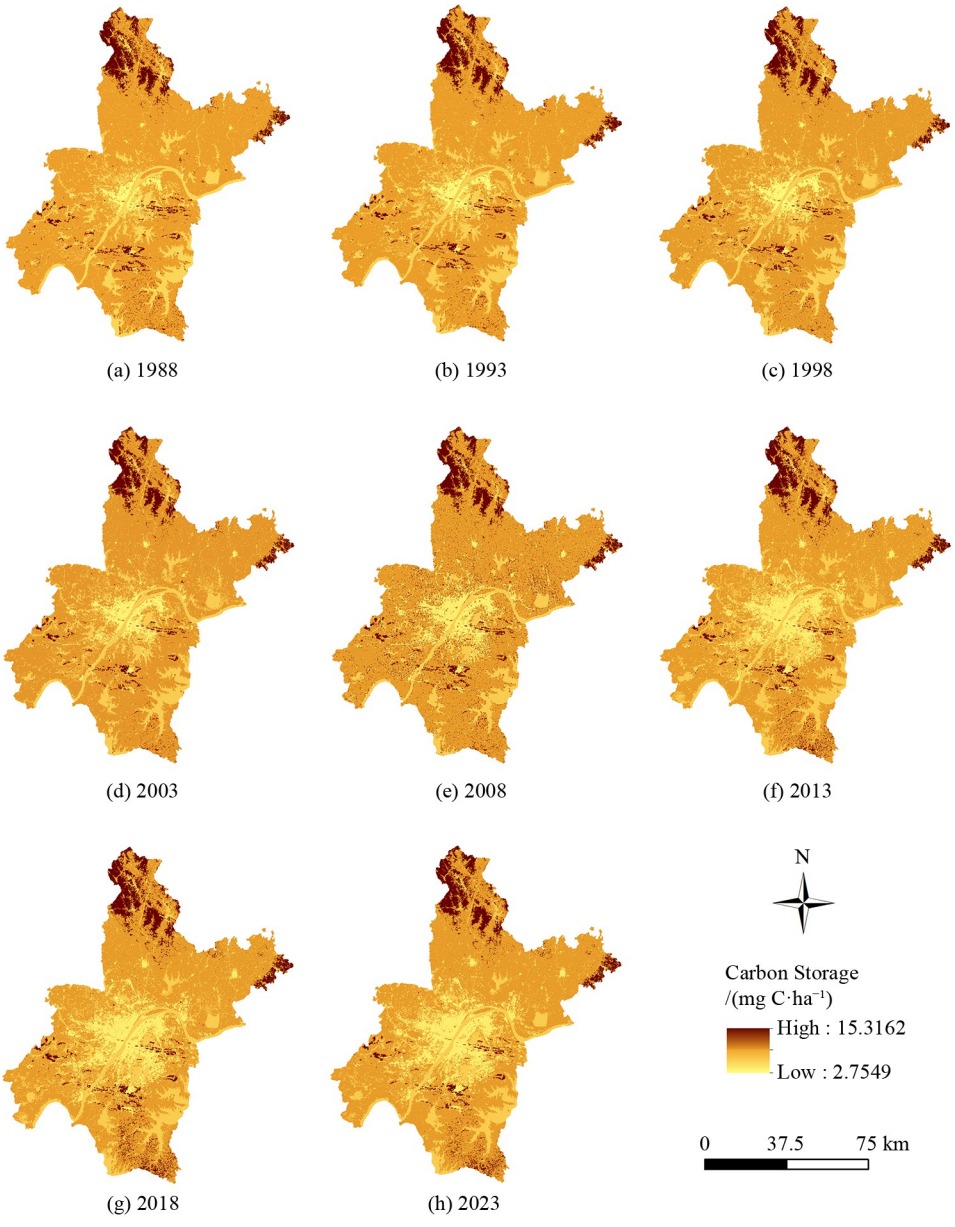

Fig. 4 Spatial variation of carbon storage in Wuhan from 1988 to 2023.

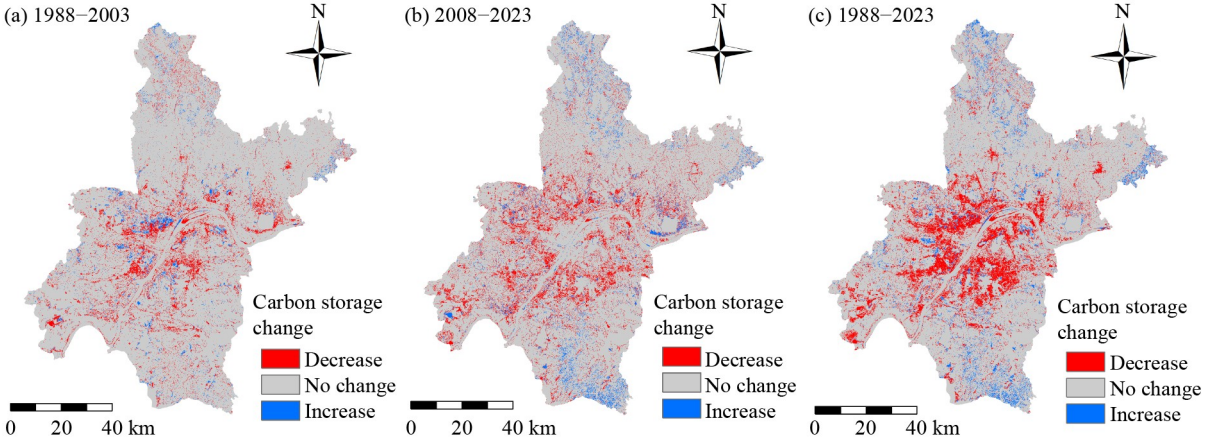

Fig. 5 Carbon storage changes over different periods in Wuhan City.

cropland, an increase of 86.15 km^2 in forested areas, a decrease of 11.19 km^2 in grassland, a decrease of 6.69 km^2 in water bodies, and a decrease of 1.22 km^2 in unused land. Throughout this 35-year timeframe, a downward trend in carbon storage was observed, culminating in a total reduction of 428.59 × 10^4 t. The transitions among different types of land use have resulted in changes to the regional carbon storage and its spatial distribution, with cropland, forest land, and built-up areas being the primary land categories impacting the variations in Wuhan's carbon storage.

From 1988 to 2023, it was observed that the conversion of cropland to other land types in Wuhan City led to a carbon storage loss quantified at 215.77 × 10^4 t (Table 7). The most significant conversion was from cropland to built-up areas, which encompassed an area of 266.81 km^2 and resulted in a carbon loss of 154.57 × 10^4 t. Losses from the conversion of forest land to other uses were tallied at 84.6 × 10^4 t, with the conversion from forest to cropland, covering an area of 125.52 km^2, accounting for a loss of 79.35 × 10^4 t. The most extensive area of conversion from water bodies to cropland was recorded at 160.07 km^2, contributed to an aggregate increase in carbon storage of 54.49 × 10^4 t.

5 Discussions

The Carbon module of the InVEST model, utilizing land use data, has been recognized for providing an efficient and intuitive approach for the estimation of carbon storage. This approach has found widespread application in the assessment of carbon storage across numerous global regions. Presently, extensive efforts are being made by researchers to conduct quantitative studies on the relationship between changes in land use and carbon storage within ecosystems such as watersheds, coastal and inland wetlands, and forests, including the projection of future scenarios. While considerable progress has been realized in carbon storage evaluation and landscape pattern analysis, studies focusing on the spatiotemporal dynamics of regional carbon storage have been comparatively scarce.

In the case of Wuhan metropolitan area, this research has employed Landsat remote sensing imagery data spanning eight periods from 1988 to 2023. Various machine learning algorithms, including Random Forest, Support Vector Machine, Gradient Boosting Trees, and Decision Trees, were compared through the Google Earth Engine (GEE) platform to derive long-term land use classification data. The changes in land use within Wuhan City were analyzed, and the InVEST model's Carbon module was utilized in conjunction with the GIS platform to evaluate the spatiotemporal variations in regional carbon storage.

Upon comprehensive consideration, the Random Forest (RF) algorithm exhibits superior performance, as indicated by its exceptional overall accuracy and Kappa coefficients (both exceeding 0.98). This superiority can be attributed to its effective handling of high-dimensional data, nonlinear modeling capabilities, diversity of models and error correction mechanisms, as well as its superior adaptability and internal validation mechanisms. These characteristics render the RF particularly advantageous in the classification of land cover from remote sensing data, especially within complex and variable real-world application scenarios, demonstrating a level of performance that is difficult for other models to match.

The findings of the study reveal that the total carbon storage in Wuhan City has exhibited a downward trend over 35 years, with an aggregate reduction of approximately 2.589 billion tons. This decline is mainly due to the outcomes of transitions among various land use types. The conversion from forests and croplands, which act as the primary carbon sinks in the area, has been identified as having a critical impact on the decrease in carbon storage.

These natural ecosystems were significant carbon sinks, capable of absorbing substantial amounts of carbon dioxide through photosynthesis. When they are converted into urban land, not only is the capacity for carbon storage lost, but the energy consumption involved in the construction of buildings and roads also leads to an increase in carbon emissions. Moreover, the ground is hardened, reducing the storage of organic carbon in the soil. Hardened surfaces impede water infiltration and plant growth, thereby diminishing soil carbon storage and the carbon cycling process. As the green spaces in urban areas decrease, the city's ability to absorb carbon is reduced. Trees and vegetation not only absorb carbon

Table 7 Land type transfer matrix from 1988 to 2023

Area conversion/km^2	Cropland	Forest	Grassland	Water	Build–up land	Unused land
Cropland	6148.95	27.92	0.05	−89.17	−154.57	−0.01
Forest	−79.35	775.53	−0.17	−1.10	−4.02	0.00
Grassland	−0.12	3.39	1.64	−0.27	−0.90	−0.09
Water	54.49	0.48	0.02	737.04	−6.73	−0.08
Build-up land	0.45	0.00	0.00	0.88	133.04	0.00
Unused land	0.09	0.00	0.22	0.09	0.04	0.20

dioxide but also help regulate the urban microclimate by providing shade and increasing evapotranspiration, reducing the urban heat island effect. The reduction of green spaces limits these natural processes, negatively impacting the urban carbon cycle. The urban heat island effect results in higher temperatures in urban areas compared to surrounding rural areas, further increasing the use of air conditioning and cooling systems, thereby increasing energy consumption and carbon emissions. Additionally, higher temperatures can accelerate the decomposition of organic matter in the urban environment, leading to more carbon emissions.

However, the application of the InVEST model for estimating carbon storage in Wuhan City is subject to certain uncertainties. First, the model's approach to simplifying the carbon cycle process, estimating through static carbon densities associated with different land use types, could diminish the accuracy of these estimates. Secondly, although carbon density data from the literature were selected and consolidated for this study, the absence of direct measurement data on the carbon density for various land use types in Wuhan City across different periods, along with difficulties in spatial verification of the chosen carbon density data, might influence the outcomes. Furthermore, the resolution of the land use type maps employed in this research is 30 m, with the time period selected being September to October, potentially impacting the model's precision in capturing land use changes.

Despite the inherent uncertainties in the InVEST model, the outcomes of the study are consistent with the actual conditions observed in Wuhan metropolitan area, indicating the model's considerable applicability.

Consequently, the investigation into the spatiotemporal distribution of changes in carbon storage within Wuhan City over the preceding 35 years is capable of distinctly illustrating the effects of land use alterations on carbon storage, offering insights for the metro's economic and ecological harmonious advancement. In forthcoming studies, an emphasis should be placed on augmenting field monitoring and surveys of carbon densities pertinent to diverse land use categories, aiming to capture long-term dynamic data, which would enhance the precision of model-based estimations. Moreover, against the backdrop of Wuhan metropolitan area's advancement in modern manufacturing and China's objective to reach its carbon emission peak by 2030, intensified efforts in researching urban expansion simulations and quantitative evaluations of additional ecosystem services are imperative. These measures will furnish scientific evidence for assessing the repercussions of urban development on ecosystem services, thereby facilitating Wuhan City's pursuit of sustainable progress in its economic and ecological realms.

6 Conclusions

1) In this research, sophisticated machine learning methodologies, namely Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Trees (GBT), and Classification and Regression Trees (CART), were utilized for the intricate classification of Wuhan metropolitan area's land use employing Landsat satellite imagery. Random Forest (RF), chosen for its exemplary performance in terms of overall accuracy and Kappa coefficient (both surpassing 0.98), was demonstrated to be reliably effective in distinguishing the unique characteristics of diverse land use types.

2) An analysis of land use changes disclosed considerable transformations in land coverage within Wuhan metropolitan area over the nearly 35-year span. A marked reduction in areas of cropland and water alongside a significant escalation in built-up land, particularly the drastic decrease of cropland (14.07%) and the extensive enlargement of built-up land (290.97%), mirrored the pressures of land resource utilization instigated by economic advancement. The conversions within urban central locales specifically accentuated the dichotomy between land conservation and agricultural land preservation.

3) Longitudinal evaluations of carbon storage indicated a general declining trajectory in Wuhan metropolitan area's total carbon storage from 1988 to 2023, with a total diminution of around 428.59×10^4 t predominantly caused by shifts in land use types, notably from cropland and forest land, which significantly affected carbon storage capacities.

4) Moreover, estimations of carbon storage highlighted a reduction of approximately 428.59×10^4 t in Wuhan metropolitan area from 1988 to 2023 due to land use transformations, chiefly from cropland and forest land. The diminishment from water body conversions, though beneficial to carbon storage, was considerably outweighed by the adverse effects from cropland and forest land conversions.

5) Variations in carbon storage within the ecosystem of Wuhan metropolitan area illustrated the direct consequences of land use alterations, stressing the necessity for synchronizing economic growth with ecological conservation. Future endeavors should prioritize mitigating the decline of forests, grasslands, and croplands, curtailing the repurposing of cropland for non-agricultural intents, and bolstering ecological engineering efforts to augment carbon storage capabilities within the regional ecosystem.

Acknowledgments This study was supported by National Key R&D Program of China (No. 2021YFB3900400).

Competing interests The authors declare that they have no competing interests.

References

Breiman L (2001). Random forests. Mach Learn, 45(1): 5–32

Breiman L, Friedman J H, Olshen R A, Stone C J (1984). Classification and regression trees. Biometrics, 40(3): 874

Friedman J, Hastie T, Tibshirani R (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann Stat, 28(2): 337–407

Haralick R M, Shanmugam K, Dinstein I (1973). Textural features for image classification. IEEE Trans Syst Man Cybern, SMC-3(6): 610–621

Huang D, Li R, Qiu J, Shi Y, Liu J (2017). Analysis on spatio-temporal variation of land use and its policy-driven factors in Wuhan metropolitan area. J Geo-Inform, 19(1): 80–90 (in Chinese)

Huete J, Didan K, Miura T, Rodriguez E P, Gao X, Ferreira L G (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens Environ, 83(1−2): 195–213

Kohestani N, Rastgar S, Heydari G, Jouibary S S, Amirnejad H (2023). Spatiotemporal modeling of the value of carbon sequestration under changing land use/land cover using InVEST model: a case study of Nour-rud Watershed, Northern Iran. Environ Dev Sustain, 26(6): 14477–14505

Li L, Song Y, Wei X, Dong J (2020). Exploring the impacts of urban growth on carbon storage under integrated spatial regulation: a case study of Wuhan, China. Ecol Indic, 111: 106064

Lu Y F (2023). Changes and Scenario Simulation of Wetland Ecosystem Services in Wuhan Urban Agglomeration. Dissertation for Master's Degree. Wuhan: Huazhong Agricultural University

Nguyen H T T, Doan T M, Tomppo E, McRoberts R E (2020). Land use/land cover mapping using multitemporal Sentinel-2 imagery and four classification methods—a case study from Dak Nong, Vietnam. Remote Sens (Basel), 12(9): 1367

Piyathilake I D U H, Sumudumali R G I, Udayakumara E P N, Ranaweera L V, Jayawardana J M C K, Gunatilake S K (2021). Modeling predictive assessment of soil erosion related hazards at the Uva province in Sri Lanka. Model Earth Syst Environ, 7(3): 1947–1962

Sharp R, Chaplin-Kramer R, Wood S A, Guerry A (2018). InVEST User's Guide

Shukla P R, Skea J, Calvo Buendia E, Masson-Delmotte V, Portner H O, Roberts D C, Zhai P, Slade R, Connors S, van Diemen R, Ferrat M, Haughey E, Luz S, Neogi S, Pathak M, Petzold J, Portugal Pereira J, Vyas P, Huntley E, Kissick K, Belkacemi M, Malley J (2019). Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems. IPCC: 2019

Singh G, Pandey A (2021). Evaluation of classification algorithms for land use land cover mapping in the snow-fed Alaknanda River Basin of the Northwest Himalayan Region. Appl Geomat, 13(4): 863–875

Tao B, Ge Q S, Li K R, Zhang Y (2001). Research progress in terrestrial ecosystem carbon cycle. Geogr Res, 12(5): 565–574

Tian Q J, Min X J (1998). Progress in vegetation index research. Adv Earth Sci, 13: 327–333

Vapnik V N (1995). The Nature of Statistical Learning Theory. New York: Springer

Xu H Q (2005). A study on information extraction of water body with the modified normalized difference water index (MNDWI). J Remote Sens, 9(5): 589–595 (in Chinese)

Yan J, Liu H, Ge Q, Zheng J, Hao Z, Wang Y (2017). Reconstruction and analysis of annual mean temperature of Wuhan for the 1906–2015 Period. Prog Geogr Sci, 36(9): 1176–1183 (in Chinese)

Yang Y Y, Dai E F, Fu H (2012). A framework for assessing the value of ecosystem services functions based on the InVEST model. J Capital Normal Univ Nat Sci Ed, 33(3): 7 (in Chinese)

Yi L, Zhang G (2012). Object-oriented remote sensing imagery classification accuracy assessment based on confusion matrix. In: Proceedings of the 20th International Conference on Geoinformatics, Hong Kong, China, 1–8

Zha Y, Gao J, Ni S (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. Int J Remote Sens, 24(3): 583–594

Zhang J, Zhang C, Dong H, Zhang L, He S (2023). Spatial–temporal change analysis and multi-scenario simulation prediction of land-use carbon emissions in the Wuhan urban agglomeration, China. Sustainability (Basel), 15(14): 11021

Zhang M, Huang X, Chuai X, Yang H, Lai L, Tan J (2015). Impact of land use type conversion on carbon storage in terrestrial ecosystems of China: a spatial-temporal perspective. Sci Rep, 5(1): 10233

Zhang P, Liu S, Du J (2017). A map spectrum-based spatiotemporal clustering method for GDP variation pattern analysis using nighttime light images of the Wuhan urban agglomeration. ISPRS Int J Geoinf, 6(6): 160

Zhang X, Ren W, Peng H (2022). Urban land use change simulation and spatial responses of ecosystem service value under multiple scenarios: a case study of Wuhan, China. Ecol Indic, 144: 109526

Zhao Z, Liu G, Mou N, Xie Y, Xu Z, Li Y (2018). Assessment of carbon storage and its influencing factors in Qinghai-Tibet Plateau. Sustainability (Basel), 10(6): 1864