

GAN-Based Sample Augmentation for Improving Above-Ground Biomass Inversion of Salt Marsh Vegetation Using Multi-Source Remote Sensing Data

Xiangtao Zhao

National Ocean Technology Center

Jianhua Zhu

Besmile@263.net

National Ocean Technology Center

Zhen Tian

National Ocean Technology Center

Zhiyong Xue

National Ocean Technology Center

Tingting Han

Liaoning Natural Resources Service Center

Jifang Wei

State Oceanic Administration

Research Article

Keywords: wetland monitoring, remote sensing inversion, adversarial neural networks, vegetation biomass

Posted Date: October 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9709398/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

GAN-Based Sample Augmentation for Improving Above-Ground Biomass Inversion of Salt Marsh Vegetation Using Multi-Source Remote Sensing Data

Xiangtao Zhao,^{1,2} Jianhua Zhu,^{1,2,*} Zhen Tian,¹ Zhiyong Xue,¹ Tingting Han,³ Jifang Wei⁴

¹National Ocean Technology Center, Tianjin, China

²Key Laboratory of Ocean Observation Technology, Tianjin, China

³Liaoning Natural Resources Service Center, Liaoning, China;

⁴North China Sea Environmental Monitoring Center, State Oceanic Administration, Qingdao, China

Corresponding author Jianhua Zhu, E-mail: Besmile@263.net

Abstract

Coastal wetlands are important components of blue carbon ecosystems and play a key role in climate change mitigation. Accurate above-ground biomass (AGB) estimation is essential for evaluating their carbon sink capacity. In this study, multi-year field AGB data were integrated with Sentinel-2 multispectral imagery to perform regional-scale biomass inversion in the Liaohe wetland. To address the limitation of insufficient field samples, a generative adversarial network (GAN) was employed for sample augmentation, and a neural network (NN) model was developed for AGB estimation. The results indicate that the proposed approach achieved high accuracy, with R^2 values of 0.79 and 0.75 for *Phragmites australis* and *Suaeda glauca*, respectively, along with relatively low RMSE and MAE values. Compared with models trained on the original dataset, the augmented dataset significantly improved model performance. Application to September 2024 imagery further revealed that the mean AGB of *Phragmites australis* and *Suaeda glauca* was 514.22 g/m² and 230.07 g/m², respectively, with clear spatial heterogeneity across the study area. Overall, the proposed method improves the accuracy and robustness of AGB inversion and provides reliable support for blue carbon assessment and wetland ecosystem management.

Keywords: wetland monitoring, remote sensing inversion, adversarial neural networks, vegetation biomass.

1 Introduction

In response to the challenges of global climate change, the Chinese government announced its goal of achieving carbon neutrality by 2060 during the general debate of the 75th session of the United Nations General Assembly on September 22, 2020 (Zhao et al. 2022). Carbon neutrality refers to offsetting greenhouse gas emissions through measures such as afforestation, energy conservation, and emission reduction, and represents a key pathway for global climate

governance. As an important component of coastal “blue carbon” ecosystems, coastal wetlands exhibit higher carbon sequestration efficiency and stability than most terrestrial ecosystems, thereby playing a crucial role in climate change mitigation (Lovelock and Duarte 2019). It is estimated that more than 50% of the carbon sink capacity of global marine ecosystems originates from blue carbon wetlands, including salt marshes, mangroves, and seagrass meadows (Wang et al. 2021; Amani et al. 2018; Davranche et al. 2010; Ghosh et al. 2016; McLeod et al. 2011). Their carbon sequestration capacity is reflected not only in carbon fixation within plant tissues (e.g., leaves, stems, and roots), but also in the long-term storage of carbon in sediments and non-living organic matter, such as litter and necromass. In China, salt marshes are the most widely distributed type of blue carbon ecosystem, mainly occurring in temperate coastal regions such as Bohai Bay, the Jiangsu coast, and the Yangtze River Delta (Wang et al. 2020). Representative vegetation types include *Phragmites australis*, halophytic vegetation, and the introduced species *Spartina alterniflora* (Wang et al. 2021).

Aboveground biomass (AGB) is a key indicator of wetland ecosystem primary productivity and carbon sink capacity (Byrd et al. 2014; Castillo et al. 2017; Yang et al. 1999). However, conventional biomass surveys rely on manual field sampling, which is labor-intensive, inefficient, and spatially limited (Goetz and Dubayah. 2011; Gallant. 2015). In recent years, advances in remote sensing technology have provided new opportunities for large-scale AGB monitoring. Statistical regression models based on multispectral satellite data and field measurements have become widely used in biomass estimation (Zhao et al. 2021; Berni et al. 2009; Chen et al. 2015; Li et al. 2015). For instance, Doughty et al. established a linear relationship between NDVI and AGB in a Californian salt marsh, achieving an R^2 of only 0.36 (Doughty and Cavanaugh et al. 2019), whereas Zhao et al. developed a multiple regression model in the Momoge wetland with an R^2 of 0.75, highlighting the critical influence of index selection and modeling strategies on accuracy (Zhao et al. 2022). Nevertheless, the number and quality of field samples remain the key factors constraining AGB inversion accuracy (Mutanga and Skidmore. 2004; Ling et al. 2014; Liu et al. 2019).

To improve model robustness and generalization, researchers have increasingly incorporated multi-source remote sensing data—such as synthetic aperture radar (SAR), LiDAR, optical multispectral imagery, and UAV-based high-resolution images—into multi-feature modeling frameworks. These data sources provide complementary spectral, structural, and textural information that enhances the estimation of aboveground biomass as well as other vegetation functional trait parameters (e.g., leaf area index, chlorophyll content, canopy structure) across different wetland types (Mutanga et al. 2012; Yang et al. 2018; Zhu et al. 2020; Wang et al. 2023; Fu et al. 2025). In addition, advancements in deep learning—especially convolutional neural networks, transformer-based architectures, and hybrid deep feature extraction approaches—have shown superior capability in capturing complex nonlinear relationships between multi-source inputs and vegetation functional traits compared with traditional machine learning models (Fu. et al. 2024). Although these multi-source and deep learning-enhanced methods significantly improve estimation accuracy, they still depend on high-quality ground-truth data for model training and validation. In coastal wetlands, however, complex microtopography, heterogeneous vegetation distribution, and dynamic tidal inundation pose significant challenges for field sampling, limiting the acquisition of sufficient and spatially representative field data within short timeframes, which constrains model performance and applicability (Melgani and Bruzzone. 2004; Soja et al. 2021).

In this context, the present study takes the Liaohe Estuary coastal wetland as a representative case to address the problem of AGB inversion under limited sample availability. Generative Adversarial Networks (GAN), originally proposed as a data generation framework based on adversarial learning, have demonstrated strong capability in modeling complex data distributions and generating realistic synthetic samples. In recent years, GAN have been widely applied in remote sensing tasks such as data augmentation, image super-resolution, missing data reconstruction, and ecological parameter estimation, effectively alleviating the limitations caused by insufficient training samples and improving model generalization performance (Hemati et al. 2025; Sha et al. 2025; Wu et al. 2024;). Building upon these advancements, we introduce a GAN-based strategy for sample augmentation to generate diverse and reliable virtual samples from a small set of field observations. Subsequently, a neural network inversion model was constructed to estimate the aboveground biomass of *Phragmites australis* and salt marsh vegetation. The findings provide critical technical support for carbon stock assessment, blue carbon ecosystem monitoring, and the protection and restoration of the Liaohe wetlands, while also offering a methodological reference for carbon sink function evaluation in similar coastal ecosystems.

2 Dataset

2.1 Study area

The Liaohe Wetland, located in the northern part of Liaodong Bay at the junction of Panjin and Yingkou in Liaoning Province, represents a typical coastal wetland ecosystem in China. Geographically, it extends from 121° 30' to 122° 00' E and 40° 45' to 41° 10' N, covering approximately 128,000 hectares. This region constitutes the core coastal wetland area of the Liaohe River Delta. The wetland lies within the north temperate continental monsoon climate zone, with a mean annual temperature of 9.8 °C and average annual precipitation ranging between 650 and 800 mm. The geomorphology is dominated by low-lying alluvial plains formed by marine regression and riverine deposition, with surface elevations generally between 2.0 and 4.0 m above sea level. The soils are characterized by alternating marine and fluvial sediments, high salinity, and poor aeration, which limit plant diversity and inhibit the growth of non-halophytic vegetation.

Vegetation in the region is relatively homogeneous and dominated by salt-tolerant species, particularly *Phragmites australis* and halophytic herbs, forming a representative salt marsh ecosystem. Fig. 1 illustrates the spatial extent of the study area and the distribution of field sampling sites within the Liaohe salt marsh wetland.

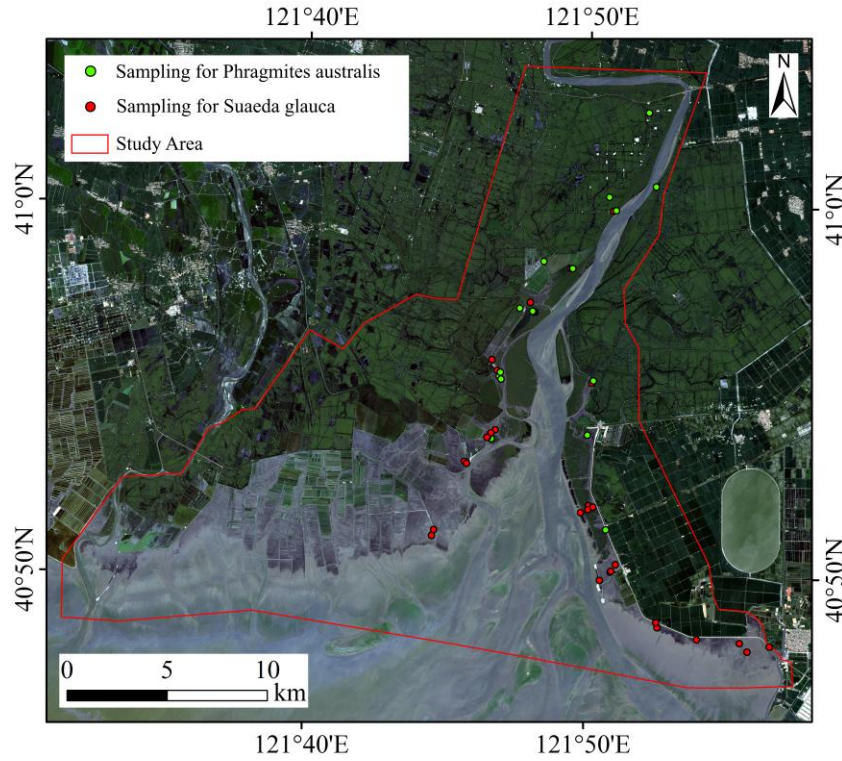


Fig. 1 Study area of salt marsh wetlands in the Liaohe estuary.

2.2 Research data

The data used in this study include field-measured AGB and Sentinel-2 multispectral imagery acquired during three field campaigns conducted between 2021 and 2024 (2021, 2023, and 2024). Field sampling covered representative salt marsh vegetation types, namely *Phragmites australis* and halophytic species. At each sampling site, geographic coordinates, vegetation type, and corresponding AGB values were systematically recorded, providing a reliable basis for model construction and validation. Sentinel-2 imagery was selected to match the field measurement periods as closely as possible. The data include visible (blue, green, red) and near-infrared bands, which effectively capture the spectral characteristics of wetland vegetation. The high spatial resolution and multi-band configuration of Sentinel-2 provide essential support for vegetation classification and biomass inversion.

2.2.1 Measured AGB data

AGB data were collected from the eastern and western banks of the Liaohe Wetland during three field campaigns conducted in July 2021, November 2023, and September 2024. A total of 53 valid sampling sites were obtained, including 21 sites dominated by *Phragmites australis* and 32 sites dominated by *Suaeda glauca*. Random sampling was adopted in each survey area to capture the spatial heterogeneity of vegetation distribution. At each site, one quadrat (0.5 m × 0.5 m) was established for field investigation. Within each quadrat, vegetation species, height, and other community parameters were recorded. Subsequently, all aboveground plant material was clipped at ground level, placed into labeled plastic bags, and transported to the laboratory. Samples were cleaned of impurities, transferred to

paper bags, and oven-dried at 80 °C until constant weight. The dry biomass was then weighed, and the AGB per unit area (g/m^2) was calculated based on the dry weight and quadrat area. For model development and validation, the measured samples were randomly divided into two subsets: 50% were used for sample augmentation and model training, while the remaining 50% were reserved exclusively for accuracy validation. The validation set consisted only of real measured samples to ensure an unbiased evaluation of model performance.

2.2.2 Sentinel-2 multispectral image data

The remote sensing data used in this study consist of Level-1C multispectral imagery acquired by the Sentinel-2A satellite on June 20, 2021, October 28, 2023, and September 2, 2024. All data were obtained from the European Space Agency (ESA) Copernicus Open Access Hub (currently the Copernicus Data Space Ecosystem; <https://dataspace.copernicus.eu/>), which provides free and open access to Sentinel mission products. The selected scenes fully cover the Liaohe salt marsh wetland and were chosen to correspond closely with field measurement periods to ensure temporal consistency. The Sentinel-2 constellation, comprising Sentinel-2A and Sentinel-2B, operates at an orbital altitude of approximately 786 km and provides a revisit time of 10 days for a single satellite (5 days combined). Each satellite is equipped with the MultiSpectral Instrument (MSI), which acquires data in 13 spectral bands spanning the visible, near-infrared (NIR), and shortwave infrared (SWIR) regions. The imagery offers multiple spatial resolutions, including 10 m resolution in the visible and NIR bands, 20 m in red-edge and SWIR bands, and 60 m for atmospheric correction bands, with a swath width of approximately 290 km. These characteristics make Sentinel-2 data highly suitable for large-scale and high-precision vegetation monitoring. Prior to analysis, the Level-1C imagery underwent standard preprocessing procedures, including radiometric calibration, atmospheric correction, and geometric correction (resampling). These steps were performed to minimize radiometric distortions and geometric inconsistencies, thereby restoring surface reflectance characteristics and ensuring the reliability of subsequent vegetation classification and biomass inversion analyses.

3 Methodology

This section describes the methodological framework adopted in this study, including Sentinel-2 data preprocessing, spectral feature extraction, saltmarsh vegetation classification using the Random Forest algorithm, GAN-assisted AGB sample expansion, and biomass inversion based on a neural network model. The overall workflow of the proposed method is illustrated in Fig. 2.

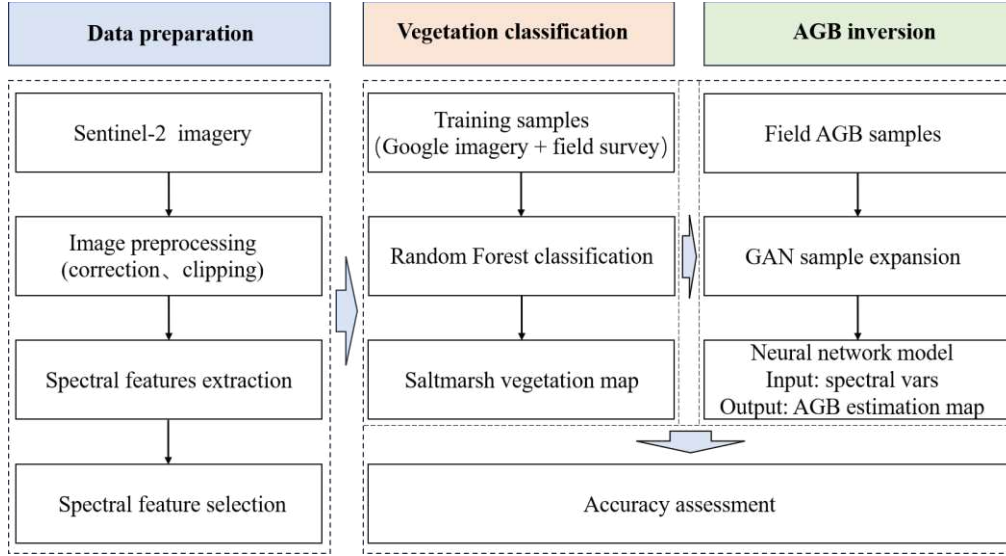


Fig. 2 Workflow of the proposed framework for saltmarsh vegetation classification and AGB inversion.

3.1 Spectral feature extraction and selection

Satellite imagery provides valuable spectral information for characterizing vegetation growth conditions at the canopy scale. Previous studies have extensively employed optical remote sensing data for vegetation classification and biomass inversion. The spectral reflectance of different bands can be used to infer vegetation characteristics (Forkuor et al. 2020), while linear or nonlinear combinations of these bands can generate spectral variables that are sensitive to specific vegetation properties (Huete et al. 2002; McFeeters et al. 1996).

To comprehensively characterize the spectral properties of salt marsh vegetation, both original spectral bands and derived spectral indices were used as input features for AGB inversion. The original spectral features were obtained from Sentinel-2 imagery, including blue (Band 2), green (Band 3), red (Band 4), red-edge (Band 5), near-infrared (Band 8), and shortwave infrared bands (Band 11 and Band 12), which are closely related to vegetation structure, chlorophyll content, and moisture conditions. Based on these spectral bands, multiple vegetation indices were calculated to enhance the representation of vegetation and environmental characteristics, including NDVI, EVI, DVI, NDWI, NDVIre1, GNDVI, SAVI, mNDWI, NDBI, and NDMI. The definitions of these indices are summarized in Table 1.

Table 1 Definition of spectral indices used in this study.

Feature	Formula	Description
NDVI	$\frac{B8 - B4}{B8 + B4}$	Vegetation vigor
EVI	$\frac{B8 - B4}{1 + B8 + 6 * B4 - B3}$	Enhanced vegetation index
DVI	$B8 - B4$	Vegetation density

NDWI	$\frac{B3 - B8}{B3 + B8}$	Water content
NDVire1	$\frac{B8 - B5}{B8 + B5}$	Red-edge vegetation index
GNDVI	$\frac{B8 - B3}{B8 + B3}$	Chlorophyll-related index
SAVI	$\frac{B8 - B4}{B8 + B4 + 0.5}$	Soil-adjusted vegetation index
mNDWI	$\frac{B3 - B11}{B3 + B11}$	Moisture indicator
NDBI	$\frac{B11 - B8}{B11 + B8}$	Built-up/soil indicator
NDMI	$\frac{B8 - B11}{B8 + B11}$	Vegetation moisture

Note: B3, B4, B5, B8, and B11 denote Sentinel-2 spectral bands.

In total, 17 spectral features were constructed, including 7 original bands and 10 derived indices. To evaluate the relationships between spectral features and AGB, both Pearson and Spearman correlation coefficients were calculated. Pearson correlation was used to assess linear relationships, while Spearman correlation was applied to capture potential nonlinear monotonic relationships. The correlation analysis results are presented in Fig. 3.

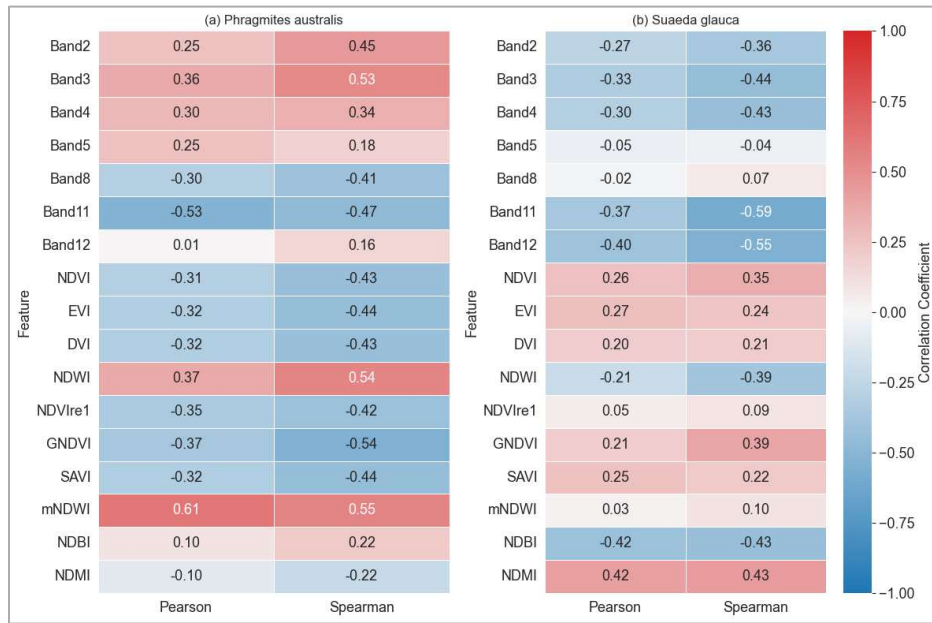


Fig. 3 Correlation coefficients between spectral features and AGB for different vegetation types.

The correlation analysis indicates that the relationships between individual spectral features and AGB are generally moderate, which is expected given the complex and nonlinear nature of biomass distribution in salt marsh wetlands, where vegetation structure, soil background, and moisture conditions jointly influence spectral responses. Notably, Spearman correlation coefficients are generally higher than Pearson coefficients, suggesting that nonlinear relationships play an important role in AGB estimation. Considering the differences in spectral responses between

vegetation types, feature selection was performed separately for *Phragmites australis* and *Suaeda glauca*. For each vegetation type, the top 8 features with the highest correlation coefficients were selected as input variables for subsequent AGB inversion modeling (Table 2). Specifically, moisture-related indices (e.g., mNDWI and NDWI) show relatively strong correlations with AGB for *Phragmites australis*, whereas SWIR bands and moisture-related indices (e.g., NDMI) exhibit higher relevance for *Suaeda glauca*, reflecting their distinct ecological and spectral characteristics. Although some features exhibit relatively low linear correlations, they were retained due to their potential contributions under nonlinear modeling frameworks, and thus machine learning models capable of capturing nonlinear relationships were employed for AGB inversion.

It should be noted that different feature sets were adopted for different tasks in this study. All 17 spectral features were used for vegetation classification to fully exploit spectral information and improve class separability. In contrast, for GAN-based sample expansion and subsequent AGB inversion modeling, feature selection was performed separately for each vegetation type, and only the top 8 features with the highest correlation coefficients were used as model inputs. This strategy helps reduce feature redundancy and noise while improving model efficiency and generalization performance.

3.2 Wetland classification based on Random Forest

Wetland classification is a systematic framework established based on wetland genesis, geomorphological characteristics, hydrological conditions, and ecological functions, providing a fundamental basis for wetland resource investigation, monitoring, and management. In this study, the classification scheme was developed in accordance with the Chinese national standard Current Land Use Classification (GB/T 21010–2017), which ensures the consistency and comparability of land cover identification. Combined with field investigations conducted in the Liaohe Estuary, land cover types within the study area were categorized into six classes: arable land, aquaculture ponds, rivers and seawater, mudflats, *Phragmites australis*, and *Suaeda glauca*. The classification system was further refined according to dominant vegetation types and ecological characteristics to better represent the coastal wetland environment.

For model construction, multi-source features extracted from remote sensing imagery were used as input variables. All 17 spectral features were employed for vegetation classification to fully exploit spectral information and improve class separability. A total of 60 representative samples for each land cover type were selected based on high-resolution imagery from Google Maps and validated using field survey data to ensure their accuracy and representativeness. The sample selection followed standard remote sensing interpretation principles, including spatial homogeneity, clear boundary definition, and temporal consistency with satellite acquisition.

The Random Forest algorithm was employed to construct the classification model due to its strong capability in handling nonlinear relationships and high-dimensional feature spaces. In this study, the number of decision trees was set to 300, and the number of variables considered at each split was determined automatically based on the square root of the total number of input features. These parameter settings improve model stability and help reduce the risk of overfitting. To evaluate the classification performance, a confusion matrix was constructed based on independent validation samples, and overall accuracy (OA) and the Kappa coefficient were calculated. These metrics provide a

quantitative assessment of classification accuracy and agreement, ensuring the reliability of the classification results. The classification outputs provide a robust basis for subsequent AGB inversion analysis.

3.3 GAN-Based AGB sample augmentation model

A GAN-based sample augmentation method was developed in this study to improve the accuracy of large-scale AGB inversion in salt marsh wetlands. The model was trained using multi-year field-measured AGB data and multi-temporal Sentinel-2 spectral features to generate high-quality synthetic samples. This approach alleviates the limitation of scarce field observations, enhances the generalization ability of the inversion model, and improves the reliability of large-scale biomass estimation.

GAN is a deep learning framework consisting of two competing neural networks: a generator and a discriminator (Lv et al. 2018).³⁸ In this study, the generator is designed to produce synthetic AGB samples that follow the distribution of real field measurements, while the discriminator aims to distinguish between real and generated samples. During adversarial training, the generator and discriminator iteratively optimize their respective performances, resulting in synthetic samples that closely resemble the real data in terms of statistical characteristics (Fig. 4). This approach improves the diversity and balance of training samples, mitigates overfitting due to limited field data, and provides an extended dataset for subsequent AGB inversion modeling.

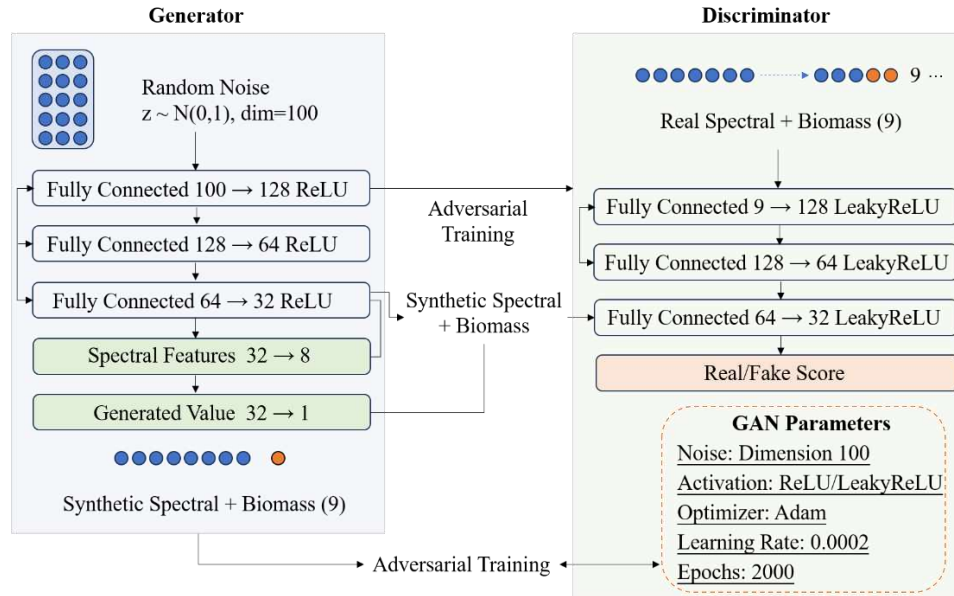


Fig. 4 Framework of GAN-based AGB sample augmentation used in this study.

In this study, the generator consists of four fully connected layers with 128, 64, and 32 hidden neurons, respectively. The ReLU activation function was used in the hidden layers, and the output layer produced nine variables, including eight spectral features and one biomass value. The discriminator also consists of four fully connected layers with 128, 64, and 32 neurons. LeakyReLU activation functions were applied to improve training stability, and the final output layer used a sigmoid function to estimate the probability that an input sample is real. The model was trained using the

Adam optimizer with a learning rate of 0.0002. The batch size was set to 16 and the number of training epochs was set to 2000. The dimension of the random noise vector input to the generator was set to 100.

Based on measured AGB samples and their corresponding multi-temporal Sentinel-2 spectral features, the GAN model was trained to generate additional synthetic samples. Separate extended datasets were constructed for *Phragmites australis* and *Suaeda glauca*. The generated samples were then combined with the original field measurements to form an expanded training dataset. Subsequently, this dataset was used together with multi-source remote sensing features to develop AGB inversion models, enabling spatial estimation of biomass for both vegetation types across the study area. By increasing the number and diversity of training samples, the proposed approach alleviates the limitation of insufficient field observations and enhances the robustness and predictive performance of the biomass estimation models.

3.4 Neural network AGB inversion model

The neural network AGB inversion models of *Phragmites australis* and saline *Suaeda glauca* flora were constructed by combining the extended AGB sample data with the multi-source feature variables extracted from remote sensing images to realize the accurate inversion of aboveground biomass in the Liaohe Estuary region, respectively. The neural network algorithm constructs a data-driven inversion model by simulating the nonlinear mapping relationship between input variables and target variables (AGB). Its core lies in continuously optimizing the model parameters by using the error back propagation mechanism, i.e., updating the weights and thresholds of each layer of the network by minimizing the mean-square error between the predicted outputs and the true values, so as to improve the prediction accuracy of the model (Lee et al. 2011; Montesano et al. 2011).

In this study, the model adopts an improved feed-forward neural network structure with 8 spectral eigen factors as input variables and measured AGB values as output variables, and the ratio of the training set to the validation set is set at 7:3. The number of nodes in the hidden layer is set at 18 in the network structure, and the output layer is 1 node. In order to ensure the nonlinear fitting ability and training stability of the model, the activation function of the input layer is selected as a Sigmoid function, and the hidden layer and the output layer are both linear functions. The maximum number of iterations is set to 20000 and the termination error threshold is 0.01 to ensure that the network reaches the convergence condition and constructs the optimal inverse model. After training, the validation dataset was input into the neural network to perform spatial inversion of AGB for *Phragmites australis* and *Suaeda glauca* in the Liaohe River salt marsh. This process generated spatially continuous AGB distribution maps for the entire study area.

To evaluate the effectiveness of the proposed neural network (NN) model, two widely used machine learning algorithms, Random Forest (RF) and Support Vector Regression (SVR), were selected as baseline models for comparison. All models were trained using the same input features and expanded sample dataset. A consistent data partition strategy was adopted, with identical training and testing sets applied across all models to ensure a fair comparison. The expanded samples were used exclusively in the training phase.

Three evaluation indexes, namely, the coefficient of determination R^2 , the root mean square error (RMSE), and the mean absolute error (MAE), were used to evaluate the accuracy and reliability of the AGB estimation model. The calculation formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}, \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|, \quad (3)$$

where, y_i denotes the measured value, x_i denotes the predicted value, $\bar{y}$ denotes the average of the measured values, and n denotes the number of samples.

4 Results

4.1 Wetland classification results

In order to accurately extract the *Phragmites australis* and *Suaeda glauca* vegetation, the Random Forest algorithm was used to carry out the feature classification on the remote sensing image of salt marsh wetland acquired in September 2024, and the training samples and validation samples were selected on Google Maps. Training samples and validation samples were selected on Google Maps for feature category identification and accuracy assessment. Following classification, the distributions of two vegetation types, *Phragmites australis* and *Suaeda glauca*, were extracted, and their respective areas were quantified. The classification results are presented in Fig. 5.

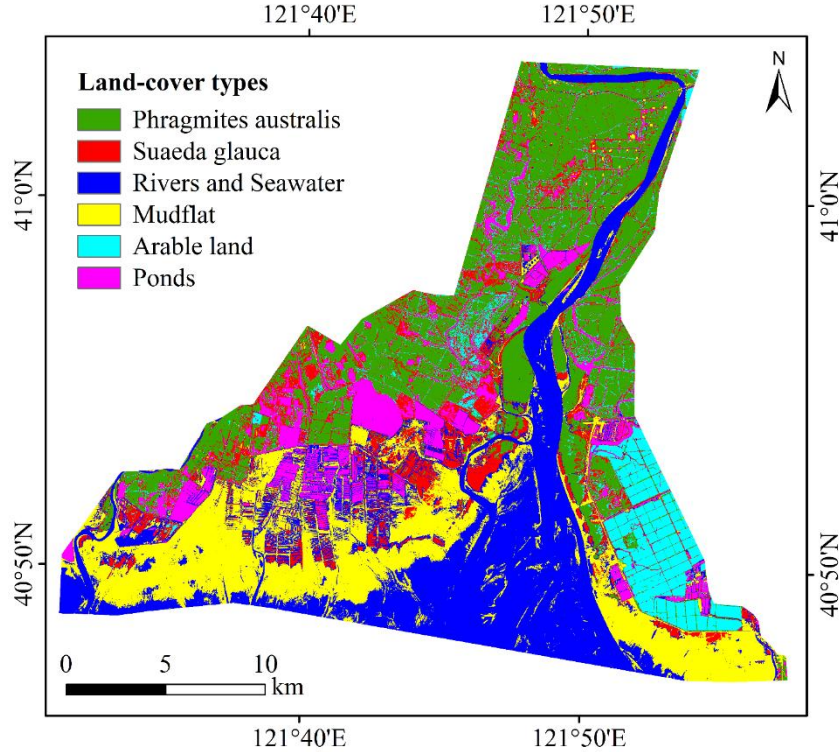


Fig. 5 Liaohe Estuary Feature Classification Results for September 2024.

The accuracy of the classification results was verified using the Confusion Matrix method, and the evaluation indexes included Overall Accuracy and Kappa Coefficient. In order to reduce the interference of human subjective factors and improve the reliability of accuracy evaluation, the “Confusion Matrix” tool in ENVI software was used for verification. The results showed that the overall classification accuracy was 92.39% and the Kappa coefficient was 0.91, which indicated that the Random Forest algorithm has high accuracy and stability in salt marsh wetland feature classification, and is able to effectively extract the target vegetation types.

According to the statistics of the feature classification results, the distribution area of *Phragmites australis* in the study area was 131.69 km², and the distribution area of *Suaeda glauca* was 66.05 km². In general, the coverage of *Phragmites australis* in the Liaohe estuary is significantly larger than that of salt flats, and the *Phragmites australis* are mainly distributed in the upper reaches of the Liaohe River, while the salt flats are mostly distributed in the lower reaches of the river and the areas near the coast.

4.2 Evaluation of AGB Sample Augmentation

In order to improve the stability and generalization ability of the AGB inversion model, this study used GAN to extend the in situ measured AGB samples. By this method, the number of AGB samples of *Phragmites australis* and *Suaeda glauca* was extended to 300 respectively to alleviate the modeling bias caused by insufficient samples. The expanded samples enhanced the diversity and representativeness of the dataset while maintaining the distributional characteristics of the original data. The statistical characteristics of the expanded samples are shown in Table 2.

Table 2 Adversarial neural network AGB sample extension results.

Vegetation type	Sample size	AGB range (g/m ²)	AGB mean (g/m ²)
Phragmites australis	300	1.69-1741.87	761.84
Suaeda glauca	300	0.82-235.46	82.21

Table 2 shows that the number of AGB samples increased substantially after expansion, while their value ranges and mean values remained consistent with those of the original samples, indicating that the expansion process preserved the statistical characteristics of the original dataset. Among them, the AGB of *Phragmites australis* ranges from 1.69 to 1741.87 g/m², with a mean value of 761.84 g/m²; the AGB of *Suaeda glauca* ranges from 0.82 to 235.46 g/m², with a mean value of 82.21 g/m², which is in line with its ecological growth characteristics. The reasonableness of the sample distribution indicates that the adversarial neural network can effectively simulate the numerical characteristics of the original samples, which provides a reliable sample basis for the subsequent construction of a robust AGB inversion model.

For visualization, 50 samples were randomly selected from the augmented dataset to construct the combined histogram–KDE–scatter (three-panel) plots, ensuring a clear and representative comparison.

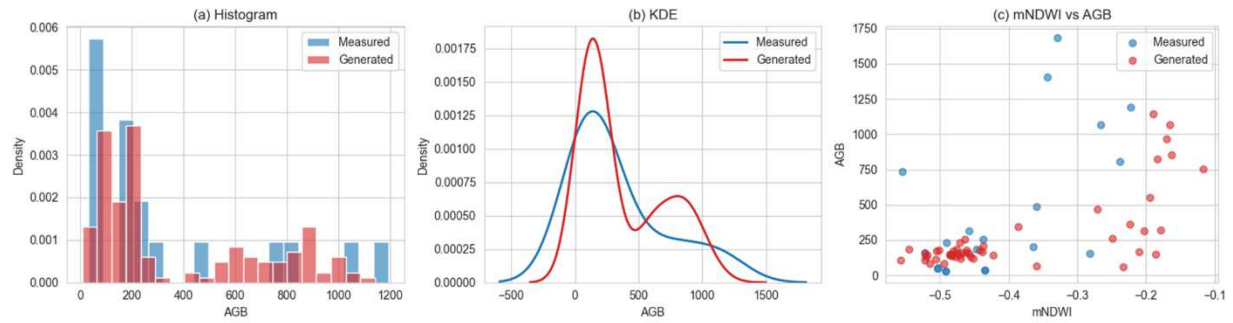


Fig. 6 Comparison of measured and GAN-generated AGB samples for *Phragmites australis*, including (a) histogram, (b) kernel density estimation (KDE), and (c) relationships between mNDWI and AGB.

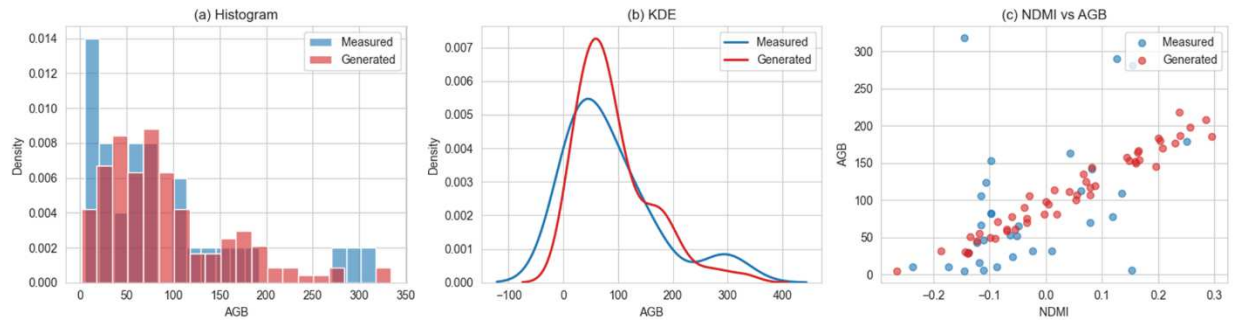


Fig. 7 Comparison of measured and GAN-generated AGB samples for *Suaeda glauca*, including (a) histogram, (b) kernel density estimation (KDE), and (c) relationships between NDMI and AGB.

As shown in Fig. 6 and 7, the generated samples exhibit distributions highly consistent with those of the measured data, indicating that the GAN model effectively captures the statistical characteristics of the original dataset. Both the histograms and KDE curves show strong agreement, with no significant distribution shift or abnormal outliers

observed. In addition, the scatter plots demonstrate that the relationships between spectral features and AGB are well preserved in the generated samples. These results confirm that the GAN-based sample expansion method can reliably reproduce both the statistical distributions and feature AGB relationships of the original data, providing effective supplementary samples for subsequent AGB inversion modeling.

4.3 Spatial Distribution and Validation of AGB Inversion

Following the extraction of vegetation distribution in the Liaohe wetland, RF, SVR, and NN were used for AGB inversion based on the expanded sample dataset. The models were trained separately for *Phragmites australis* and *Suaeda glauca*. A consistent data partition strategy was adopted, with identical training and validation datasets applied across all models to ensure a fair comparison. The performance of the models is summarized in Table 3.

Table 3 Accuracy of AGB inversion modelling

Vegetation type	Model	R ²	RMSE (g/m ²)	MAE (g/m ²)
<i>Phragmites australis</i>	SVR	0.68	257.31	189.52
	RF	0.73	246.88	178.45
	NN	0.83	230.93	160.52
<i>Suaeda glauca</i>	SVR	0.58	35.42	26.38
	RF	0.65	29.66	25.46
	NN	0.79	26.55	21.70

As shown in Table 3, the neural network model consistently outperformed RF and SVR for both vegetation types, achieving higher R² and lower RMSE and MAE values. This indicates that the neural network is more capable of capturing the nonlinear relationship between spectral features and AGB, especially when trained with the expanded dataset. Although RF and SVR achieved relatively stable performance, their predictive capability was limited compared to the neural network model, likely due to their weaker ability to model complex nonlinear interactions in heterogeneous salt marsh environments. Specifically, for *Phragmites australis*, the neural network model achieved a high coefficient of determination (R² = 0.83), although the prediction error was relatively large (RMSE = 230.93 g/m²). This can be attributed to the wide range and high variability of biomass in this species. In contrast, the inversion results for *Suaeda glauca* showed lower error (RMSE = 26.55 g/m²) and more stable performance (R² = 0.79), indicating that the model performs more consistently for vegetation types with relatively homogeneous biomass distribution.

Based on these results, the neural network model was selected for large-scale AGB inversion in the study area. The trained model was subsequently applied to the entire wetland region to generate spatially continuous AGB maps for both vegetation types. Furthermore, independent field-measured samples were used to validate the inversion results, confirming that the proposed approach provides reliable and accurate estimation of AGB at the regional scale. Overall, the results demonstrate that the neural network model, combined with sample expansion, can effectively improve the accuracy and robustness of AGB inversion in salt marsh wetlands.

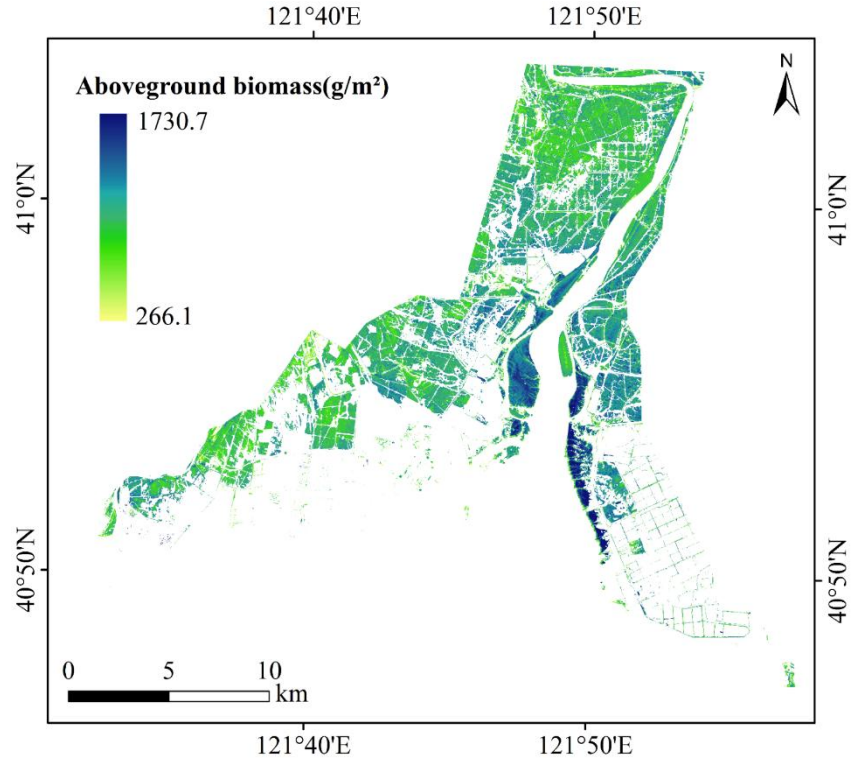


Fig. 8 AGB inversion results on *Phragmites australis* ground in September 2024.

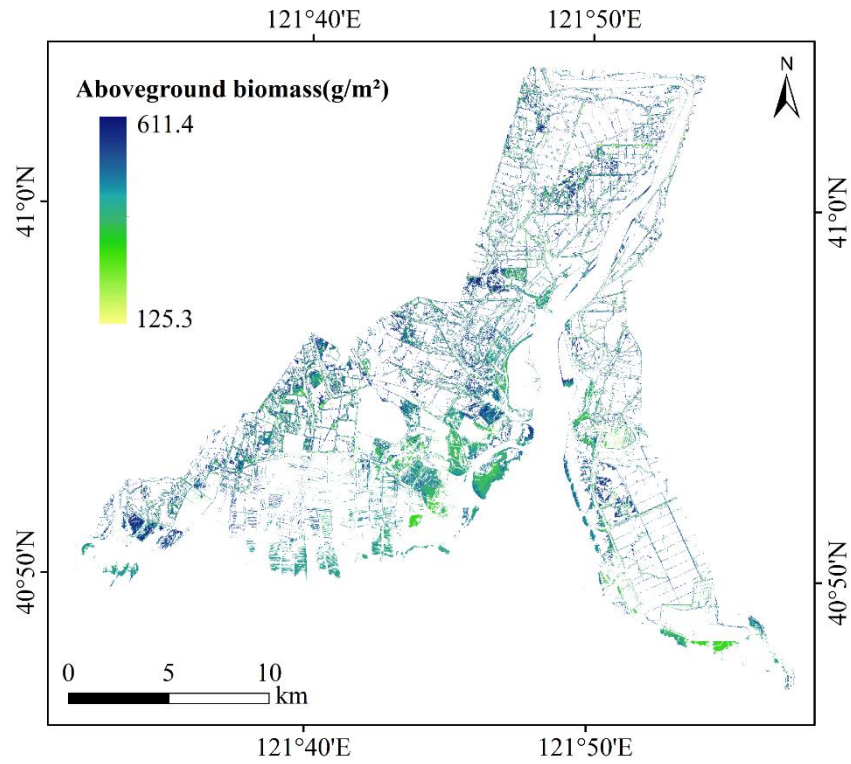


Fig. 9 AGB inversion results on *Suaeda glauca* ground in September 2024.

According to the statistics of the inversion results, the AGB of *Phragmites australis* in the Liaohu estuary region ranged 266.04 g/m² to 1500.37 g/m², with an average AGB of 514.22 g/m² and a total aboveground biomass of about

6.72×10¹⁰ g. It was mainly located in the area around the estuary, with a high level of biomass accumulation. The AGB of *Suaeda glauca* ranged from 125.06 g/m² to 554.04 g/m², with a mean value of 230.07 g/m² and a total biomass of 1.35×10¹⁰ g. It was mainly distributed in the relatively open areas of the mudflat, and the overall aboveground biomass was slightly lower than that of the *Phragmites australis* area.

To verify the accuracy of the AGB inversion results, model performance was evaluated using independent measured samples. The R²、RMSE and MAE were calculated. Notably, only real measured samples were used for validation, while the expanded samples were used exclusively for model training, ensuring an unbiased evaluation. In addition, to evaluate the effectiveness of the sample expansion strategy, model performance before and after augmentation was compared for both vegetation types. The results are presented in Table 4.

Table 4. AGB inversion accuracy evaluation table

	Vegetation type	R ²	RMSE (g/m ²)	MAE (g/m ²)
Before sample	<i>Phragmites australis</i>	0.60	328.80	292.10
expansion	<i>Suaeda glauca</i>	0.62	56.36	51.63
After sample	<i>Phragmites australis</i>	0.79	236.27	192.57
expansion	<i>Suaeda glauca</i>	0.75	46.04	34.27

Compared with the results before sample expansion, the model achieved improved predictive performance for both salt marsh vegetation types after augmentation. Among them, the R² of *Phragmites australis* improved by 0.19, the RMSE decreased by 92.53 g/m², and the MAE decreased by 99.53 g/m², indicating that the model's fitting ability was significantly enhanced on high-variability samples. The R² of *Suaeda glauca* was improved by 0.13, and the RMSE and MAE were reduced by 10.32 g/m² and 17.36 g/m², respectively, indicating that the predictive effect of the model was also further optimized in the low and medium biomass zones. The sample extension strategy has a significant role in improving the model generalization ability and prediction accuracy, especially in the AGB inversion of *Phragmites australis* vegetation, which verifies the effectiveness of enhancing the original sample set in data-driven modeling.

In order to further visualize the accuracy and prediction effect of the model construction, scatter plots of the measured and predicted values are plotted in this paper. Through the scatter distribution and its fitting relationship with the 1:1 reference line, the fitting ability and error characteristics of the model at different stages can be clearly observed, so as to more accurately evaluate its prediction performance. Fig. 10 and 11 show the scatter plots of the application accuracy evaluation of the model before and after the sample expansion, respectively, and the comparative analysis helps to verify the actual effect of the sample expansion on the model performance enhancement.

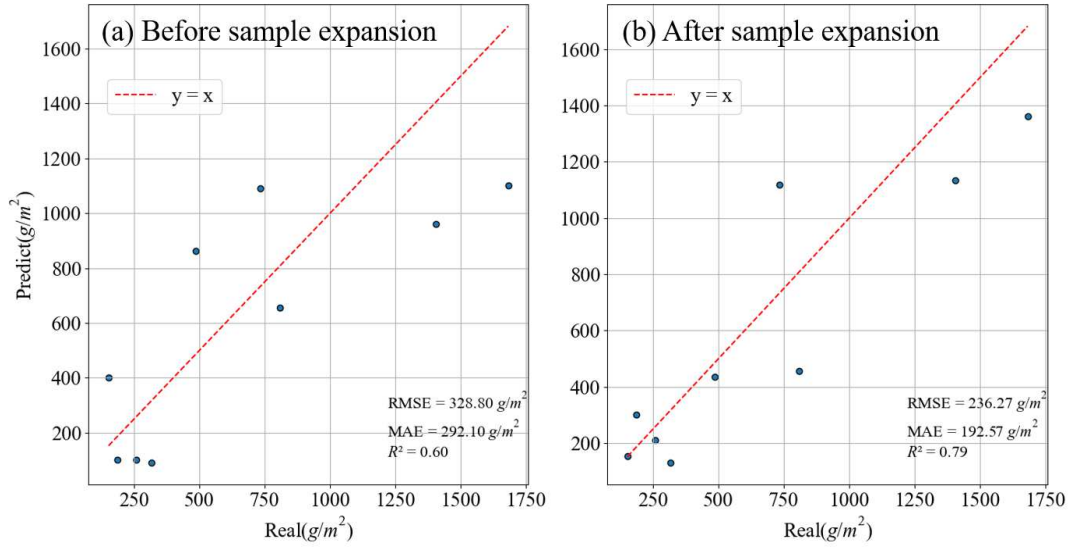


Fig. 10 Evaluation of inversion accuracy of *Phragmites australis*.

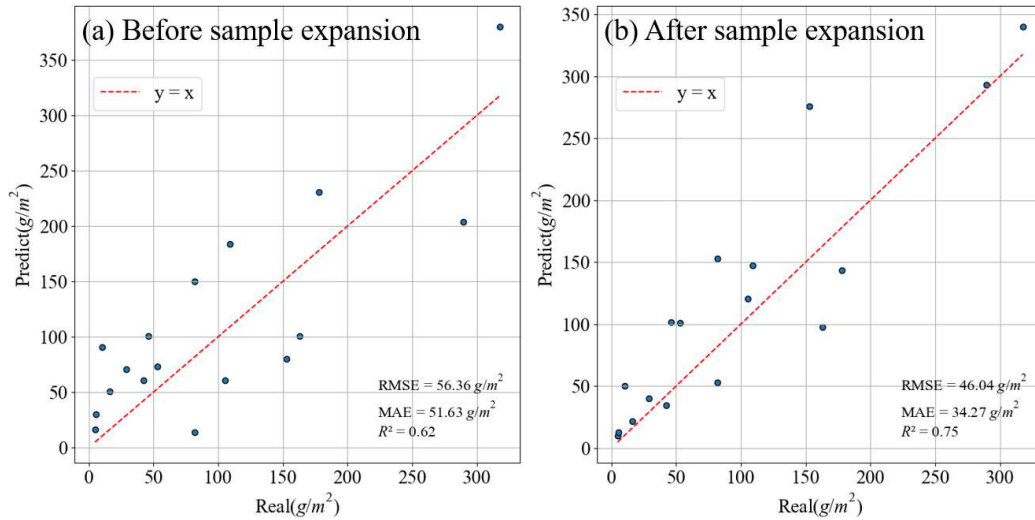


Fig. 11 Evaluation of inversion accuracy of *Suaeda glauca*.

The scatter plot indicates that, compared with the results before sample expansion, the predicted values after expansion are more closely aligned with the 1:1 line, reflecting improved agreement with measured AGB. Although some deviations are present, they may result from discrepancies between field measurements and image-derived spectral information. Overall, the model performance is enhanced after sample expansion.

5 Discussion

This study presents a remote sensing-based framework for estimating the AGB of salt marsh vegetation in the Liaohe Estuary by combining Sentinel-2 multispectral imagery, limited field observations, and a GAN-based sample extension strategy. The proposed approach addresses a key constraint in coastal wetland biomass inversion, namely the scarcity of in situ measurements caused by complex hydrological conditions and limited accessibility.

By augmenting the original AGB dataset with GAN-generated samples that preserve the statistical properties of measured data, the robustness and generalization ability of the neural network inversion models were substantially improved. The results show clear accuracy gains for both *Phragmites australis* and *Suaeda glauca*, with notable increases in the coefficient of determination and corresponding reductions in RMSE and MAE after sample expansion. The improvement is particularly evident for *Phragmites australis* vegetation, which exhibits large spatial variability in biomass and is more sensitive to sample imbalance in data-driven modeling.

The spatial inversion results derived from Sentinel-2 imagery in September 2024 reveal distinct biomass distribution patterns between vegetation types. *Phragmites australis* show higher AGB values and total biomass, mainly distributed in the upper estuarine areas, whereas *Suaeda glauca* is characterized by lower biomass levels and a more fragmented distribution across tidal flats. These patterns are consistent with field observations and known ecological characteristics, supporting the reliability of the proposed inversion framework.

Overall, this study demonstrates that GAN-assisted sample augmentation is an effective solution for improving biomass estimation accuracy in data-scarce coastal wetlands. The methodology relies on freely available multispectral satellite data and can be readily transferred to other salt marsh systems with similar environmental conditions, providing a practical reference for large-scale wetland biomass monitoring and carbon stock assessment using remote sensing.

Acknowledgements We would like to thank the anonymous reviewers for their comments and their insightful suggestions significantly improved this paper. We thank the Natural Resources Ministry of Harbin Eco-logical Center for providing data.

Author contributions Conceptualization, methodology, formal analysis, and writing—original draft preparation, Zhao Xiangtao; software, validation, data curation, and visualization, Zhao Xiangtao and Xue Zhiyong; investigation, Zhao Xiangtao and Han Tingting; supervision, project administration, funding acquisition, and writing—review and editing, Zhu Jianhua and Wei Jifang. All authors have read and agreed to the published version of the manuscript.

Funding This work was supported by the directional Fund Project of Key Laboratory of Ocean Observation Technology (2023klootB03).

Data availability The Sentinel-2 imagery used in this study is publicly available from the European Space Agency (ESA) Copernicus Open Access Hub. The processed data and related materials supporting the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

Clinical trial number Not applicable.

Open Access

References

- Amani M, Salehi B, Mahdavi S, Brisco B (2018) Spectral analysis of wetlands using multi-source optical satellite imagery. *ISPRS J Photogramm Remote Sens* 144:119–136. <https://doi.org/10.1016/j.isprsjprs.2018.07.005>
- Berni JAJ, Zarco-Tejada PJ, Suarez L, Fereres E (2009) Thermal and multispectral UAV remote sensing for vegetation monitoring. *IEEE Trans Geosci Remote Sens* 47:722–738. <https://doi.org/10.1109/TGRS.2008.2010457>
- Byrd KB, O’Connell JL, Di Tommaso S, Kelly M (2014) Evaluation of sensor types and environmental controls on mapping biomass. *Remote Sens Environ* 149:166–180. <https://doi.org/10.1016/j.rse.2014.04.003>
- Castillo JAA, Apan AA, Maraseni TN, Salmo SG (2017) Estimation of mangrove biomass using Sentinel imagery. *ISPRS J Photogramm Remote Sens* 134:70–85. <https://doi.org/10.1016/j.isprsjprs.2017.10.016>
- Chen Y, Zhao X, Jia X (2015) Spectral–spatial classification of hyperspectral data based on DBN. *IEEE J Sel Top Appl Earth Obs Remote Sens* 8(6):2381–2392. <https://doi.org/10.1109/JSTARS.2015.2388577>
- Davranche A, Lefebvre G, Poulin B (2010) Wetland monitoring using classification trees and SPOT-5 seasonal time series. *Remote Sens Environ* 114(3):552–562
- Doughty CL, Cavanaugh KC (2019) Mapping coastal wetland biomass from UAV imagery. *Remote Sens* 11(5):540. <https://doi.org/10.3390/rs11050540>
- Forkuor G, Zoungrana JBB, Dimobe K et al (2020) Biomass mapping in dryland forest using Sentinel data. *Remote Sens Environ* 236:111496. <https://doi.org/10.1016/j.rse.2019.111496>
- Fu BL, Jiang LH, Yao H et al (2024) Mangrove height retrieval using ML and UAV-LiDAR. *Int J Digit Earth* 17(1):2392851. <https://doi.org/10.1080/17538947.2024.2392851>
- Fu BL, Wei YY, Jiang LH et al (2025) Mangrove biomass estimation using UAV-LiDAR and SAR. *Ecol Inform* 88:103160. <https://doi.org/10.1016/j.ecoinf.2025.103160>
- Gallant AL (2015) The challenges of remote monitoring of wetlands. *Remote Sens* 7(8):10938–10950. <https://doi.org/10.3390/rs70810938>
- Ghosh S, Mishra DR, Gitelson AA (2016) Long-term monitoring of tidal wetlands. *Remote Sens Environ* 173:39–58. <https://doi.org/10.1016/j.rse.2015.11.015>
- Goetz S, Dubayah R (2011) Advances in remote sensing for forest carbon. *Carbon Manag* 2(3):231–244. <https://doi.org/10.4155/cmt.11.18>
- Hemati M, Mahdianpari M, Shiri H et al (2025) Wetland biomass mapping using PlanetScope. *Remote Sens Appl Soc Environ* 38:101604. <https://doi.org/10.1016/j.rsase.2025.101604>
- Huete A, Didan K, Miura T et al (2002) MODIS vegetation indices performance. *Remote Sens Environ* 83(1–2):195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Lee K, Mizutani N, Fujii T (2011) Prediction of wave breaking using ANN. *J Coast Res* 27(2):318–328. <https://doi.org/10.2112/JCOASTRES-D-10-00094.1>
- Li L, Chen Y, Xu T et al (2015) Super-resolution mapping of wetland inundation based on ANN and GA. *Remote Sens Environ* 164:142–154. <https://doi.org/10.1016/j.rse.2015.04.009>
- Ling C, Sun H, Zhang H et al (2014) Above-ground biomass estimation using WorldView-2 data. In: *Proc Int Workshop Earth Obs Remote Sens Appl*, 428–432

- Liu Y, Gong W, Xing Y et al (2019) Forest biomass estimation using Sentinel-1/2 and DEM. *ISPRS J Photogramm Remote Sens* 151:277–289. <https://doi.org/10.1016/j.isprsjprs.2019.03.016>
- Lovelock CE, Duarte CM (2019) Dimensions of blue carbon and emerging perspectives. *Biol Lett* 15(3):20180781. <https://doi.org/10.1098/rsbl.2018.0781>
- Lv Y, Yuan R, Wang T et al (2018) Fault diagnosis based on CEEMDAN. *Materials* 11(6):1009. <https://doi.org/10.3390/ma11061009>
- McFeeters SK (1996) NDWI for water delineation. *Int J Remote Sens* 17(7):1425–1432. <https://doi.org/10.1080/01431169608948714>
- McLeod E, Chmura GL, Bouillon S et al (2011) A blueprint for blue carbon. *Front Ecol Environ* 9(10):552–560. <https://doi.org/10.1890/110004>
- Melgani F, Bruzzone L (2004) Classification of hyperspectral images with SVM. *IEEE Trans Geosci Remote Sens* 42:1778–1790. <https://doi.org/10.1109/TGRS.2004.831865>
- Montesano P, Ranson KJ, Zhang Z et al (2011) Forest biomass mapping from lidar and radar. *Remote Sens Environ* 115(11):2906–2916. <https://doi.org/10.1016/j.rse.2011.03.021>
- Mutanga O, Adam E, Cho MA (2012) High density biomass estimation using WorldView-2 and RF. *Int J Appl Earth Obs Geoinf* 18:399–406. <https://doi.org/10.1016/j.jag.2012.03.012>
- Mutanga O, Skidmore AK (2004) Narrow band vegetation indices overcome saturation in biomass estimation. *Int J Remote Sens* 25(19):3999–4014. <https://doi.org/10.1080/01431160310001654923>
- Sha JH, Zhuo ZJ, Zhou QQ et al (2025) Saltmarsh vegetation fractional cover estimation. *Diversity* 17(1):3. <https://doi.org/10.3390/d17010003>
- Soja MJ, Quegan S, d'Alessandro MM et al (2021) Biomass mapping with P-band SAR. *Remote Sens Environ* 253:112153. <https://doi.org/10.1016/j.rse.2020.112153>
- Wang F, Sanders CJ, Santos IR et al (2021) Global blue carbon accumulation in tidal wetlands increases with climate change. *Natl Sci Rev* 8(9):nwaa296. <https://doi.org/10.1093/nsr/nwaa296>
- Wang F, Tang J, Ye S, Liu J (2021) Blue carbon sink function of Chinese coastal wetlands. *Bull Chin Acad Sci* 36(3):241–251
- Wang Q, Lu X, Zhang H et al (2023) Comparison of ML methods for LAI and biomass estimation. *Forests* 14(8):1688. <https://doi.org/10.3390/f14081688>
- Wang X, Xiao X, Zou Z et al (2020) Mapping coastal wetlands of China using Landsat time series. *ISPRS J Photogramm Remote Sens* 163:312–326. <https://doi.org/10.1016/j.isprsjprs.2020.03.014>
- Wu N, Zhang C, Zhuo W et al (2024) Impact of vegetation types on biomass inversion. *Remote Sens* 16(24):4762. <https://doi.org/10.3390/rs16244762>
- Yang S, Feng Q, Liang T et al (2018) Modeling grassland biomass based on ANN and remote sensing. *Remote Sens Environ* 204:448–455. <https://doi.org/10.1016/j.rse.2017.10.011>
- Yang Z, Wang XK, Miao H (1999) Chinese terrestrial ecosystem services. *Acta Ecol Sin* 19(5):607–613
- Zhao X, Ma X, Chen B et al (2022) Challenges toward carbon neutrality in China. *Resour Conserv Recycl* 176:105959. <https://doi.org/10.1016/j.resconrec.2021.105959>

496 Zhao YH, Hou P, Jiang JB et al (2021) Progress in quantitative inversion of vegetation parameters. Natl Remote Sens
 497 Bull 25(11):2173–2197. <https://doi.org/10.11834/jrs.20210402>
 498 Zhao YX, Mao DH, Zhang DY et al (2022) Mapping *Phragmites australis* aboveground biomass in the Momoge
 499 Wetland Ramsar Site based on Sentinel-1/2 images. Remote Sens 14(3):694. <https://doi.org/10.3390/rs14030694>
 500 Zhu Y, Liu KW, Myint S et al (2020) Integration of GF2, GF3 and UAV data for biomass estimation. Remote Sens
 501 12(12):2039. <https://doi.org/10.3390/rs12122039>