Supplementary Materials for

Revealing coastal vegetation structural diversity through LiDAR-derived relative entropy

Guanpu Qi[1]†, Jiawei Wei[2]†, Zaizhou Xin[3], Longcheng Yao[1], Xiong Chen[2*], Jun Ma[1*]

[1] State Key Laboratory of Wetland Conservation and Restoration, National Observations and Research Station for Wetland Ecosystems of the Yangtze Estuary, Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, and Institute of Eco-Chongming, School of Life Sciences, Fudan University, # 2005 Songhu Road 10, Shanghai, 200438, China.

[2] School of Information Science and Engineering, Fudan University, # 2005 Songhu Road 10, Shanghai, 200438, China.

[3] Key Laboratory of Computational Physical Sciences (Ministry of Education), Institute of Computational Physical Sciences, State Key Laboratory of Surface Physics, and Department of Physics, Fudan University, Shanghai 200433, China.

†These authors contributed to the work equally.

*Corresponding author

E-mail:

Jun Ma: ma_jun@fudan.edu.cn

Xiong Chen: **chenxiong@fudan.edu.cn**

Tel. Jun Ma: +86-21-31246690.

Tel. Xiong Chen: +86-21-65643231.

This file includes:

Figures

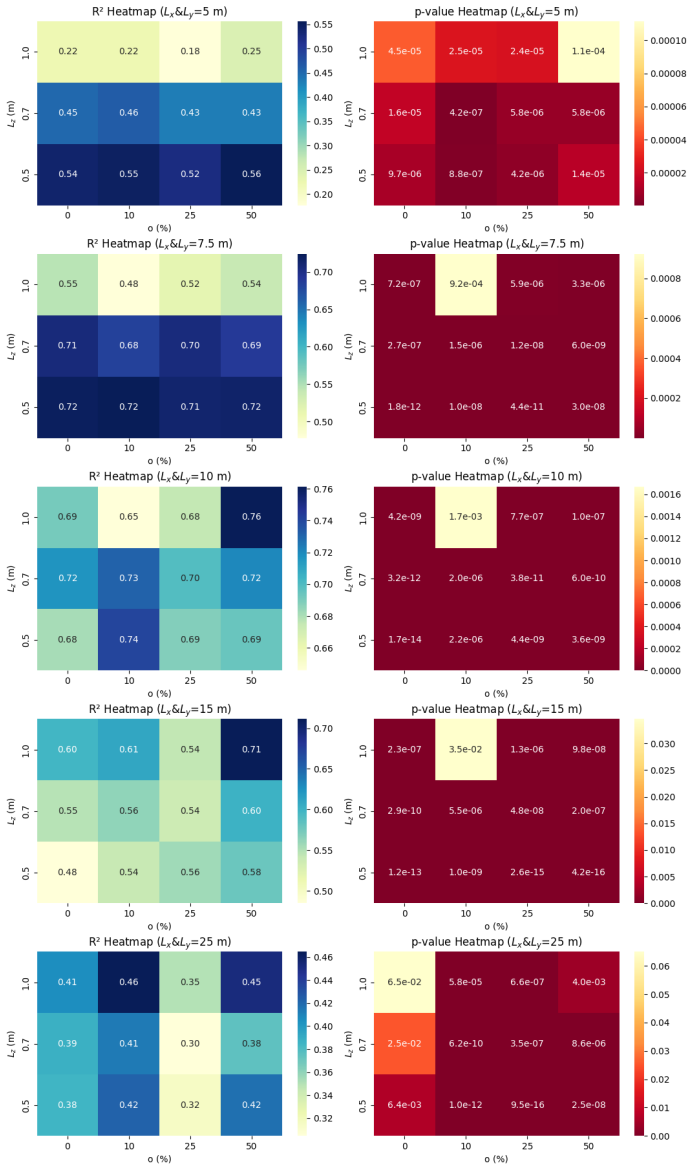


Fig. S1 | Heatmap of R^2 and p-values for different hyperparameter combinations. We conducted a grid search over the hyperparameters of the proposed method to find the optimal combination for the application. In the tests, we performed regression analysis with the Shannon index on a 60 m × 50 m plot and ANOVA analysis on a 25 m × 25 m plot. In these experiments, the R^2 value and p-value for each parameter combination were selected as evaluation metrics. Specifically, we explored window side lengths (L_x & L_y) of 5 m, 7.5 m, 10 m, 15 m, and 25 m; window heights (L_z) of 0.5 m, 0.7 m, and 1 m; and overlap ratios of 0, 0.1, 0.25, and 0.5 (o). **Figure S1** shows heatmaps of the two metrics under different parameter combinations.

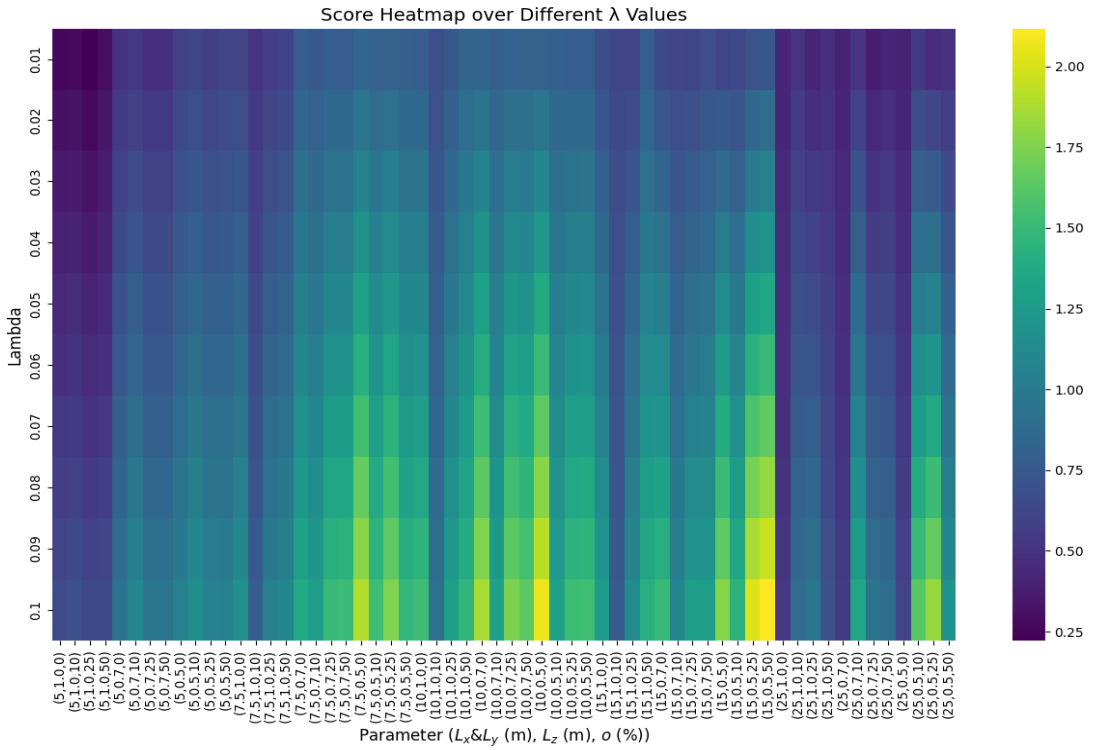


Fig. S2 | Heatmap of scores for different hyperparameter combinations under varying trade-off factors.

We calculated the final score for each parameter combination according to the following formula:

$$\text{Score} = R^2 - \lambda \log_{10}^{\text{p-value}} \quad \text{(S1)}$$

where λ is a weighting factor used to balance the two evaluation metrics, we varied λ from 0.01 to 0.1 and computed the scores for each case, as shown in **Figure S2**. Among all parameter combinations, those with a window height of L_z=0.5 m achieved higher scores, which may be related to the average height of low-lying vegetation. In particular, combinations of small windows without overlap (10, 0.5, 0) and large windows with overlap (15, 0.5, 25) and (15, 0.5, 50) achieved the best scores across all tested settings. Considering the additional terrain normalization required during processing and the pre-trimming of water body regions in the nationwide field dataset, we selected the optimal combination (10, 0.5, 0) under λ = 0.02 to ensure computational efficiency.

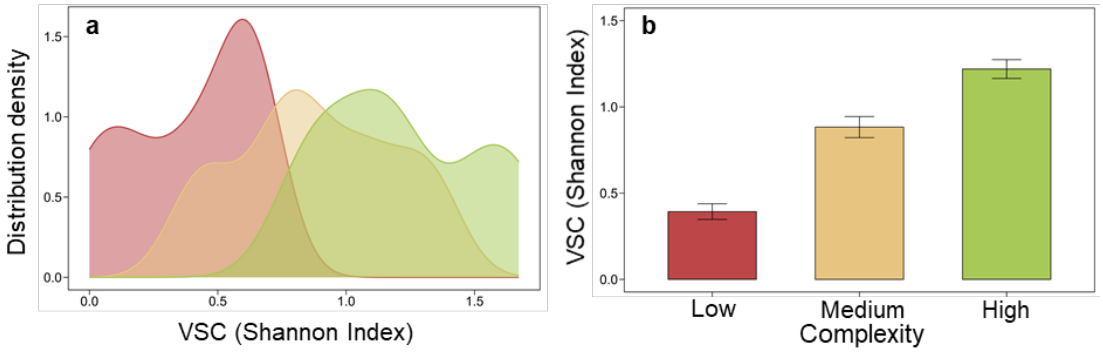
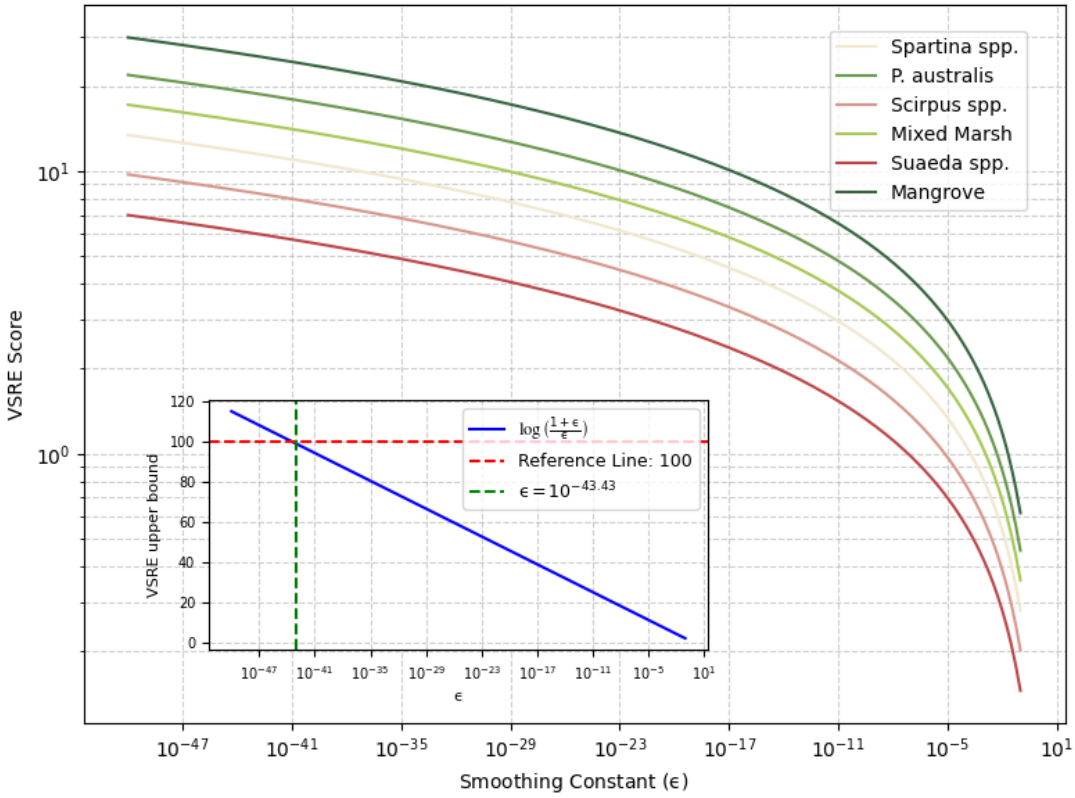


Fig. S3 | VSC index distribution characteristics constructed based on the Shannon index. a, the VSC distribution curve was fitted using kernel density estimation. **b,** Mean values of VSC in groups with different human-defined complexity. Error bars indicate standard deviation.



58

59 **Fig. S4** | The variation trend of average VSRE scores for different vegetation
60 communities in field data concerning the smoothing constant and the theoretical upper
61 bound of VSRE as a function of the smoothing constant.

62 To ensure numerical stability in the computation of RE between voxel-based 3D
63 histograms, we adopt a modified formulation:

64 $$RE(\Phi||\Psi) = \sum_{i,j,k} \Phi_{ijk} log \frac{\Phi_{ijk} + \epsilon}{\Psi_{ijk} + \epsilon} \quad \text{(S2)}$$

65 Where, Φ and Ψ are the normalized point cloud distributions in the two windows. ϵ is
66 a small positive constant introduced to prevent divergence caused by zero entries in
67 either distribution. This adjustment stabilizes the computation and implicitly constrains
68 the upper bound of the VSER. Specifically, when considering the extremal case where
69 a single bin in Φ concentrates the entire probability mass (i.e., $\Phi_a = 1$ for one index
70 and zero elsewhere), and Ψ is uniform or zero in that bin, the RE divergence simplifies
71 to:

$$RE_{max} = log\left(\frac{1+\epsilon}{\epsilon}\right) \quad \text{(S3)}$$

is expression shows that the maximum divergence is directly determined by the choice of ϵ. The smaller the value of ϵ, the higher the potential upper bound for the RE divergence. In our experiments with field data, we set:

$$RE_{max} = log\left(\frac{1+10^{-43.43}}{10^{-43.43}}\right) \approx log(10^{43.43}) = 100 \quad \text{(S4)}$$

Figure S4 presents the distributions of the VSRE scores across various plant species from field data in China, under different smoothing constants ϵ. The figure also shows the theoretical upper bound of VSRE as a function of ϵ. As illustrated, the choice of ϵ exerts a significant influence on the numerical stability and the sensitivity of the VSRE metric. When ϵ is set too large (e.g., $\epsilon > 10^{-2}$), the added smoothing overwhelms the original differences between distributions, compressing the RE values into a narrow range. Consequently, the VSRE scores fail to distinguish structural complexity across vegetation types, resulting in a substantial loss of discriminative power. Conversely, when ϵ is excessively small (e.g., $\epsilon < 10^{-50}$), although the sensitivity to structural differences increases, the computation becomes highly susceptible to numerical instability due to finite-precision floating-point arithmetic. Minor numerical errors may be amplified in such cases, resulting in unreliable VSRE estimates.

By setting $\epsilon = 10^{-43.43}$, the theoretical upper bound of the VSRE score is approximately 100, placing the indicator within a practical, interpretable [0, 100] range. This choice ensures that the VSRE metric maintains sufficient sensitivity to capture structural variations in the point cloud while preserving numerical stability across different datasets. Moreover, normalizing VSRE within [0, 100] facilitates intuitive interpretation and straightforward comparison across studies and ecosystems.

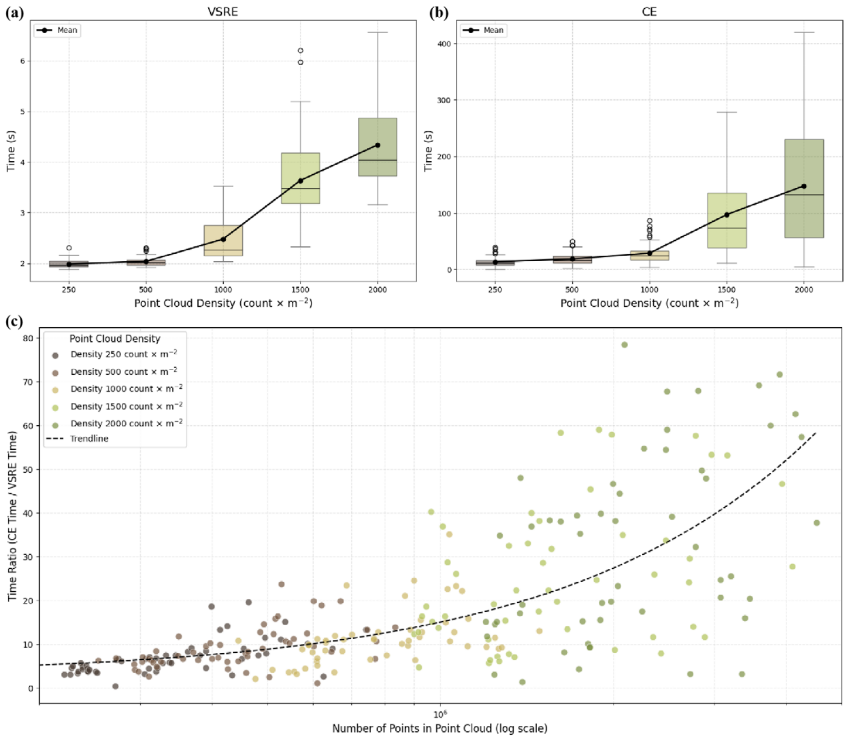


Fig. S5 | Comparison of Computational Efficiency Between VSRE and CE Methods. a-b, Variation trends of computation time for VSRE and CE concerning point cloud density. **c**, The time ratio (CE Time/VSRE Time) as a function of point cloud size. All statistics were derived from 25m × 25m samples in fixed plots.

In **Figure S5**, we compare the computational efficiency between the VSRE and CE methods using 25m×25m plot data. Panels (a) and (b) display box plots of processing time for both methods across varying point cloud densities. As density increases, the computational time rises for both approaches due to the inherent processing demands of larger point clouds. Nevertheless, our VSRE method consistently requires less than one-tenth the time of CE. Panel (c) presents the time ratio (CE/VSRE) as a function of point count, revealing that VSRE's efficiency advantage becomes more pronounced with increasing point numbers, reaching a 70-fold improvement at approximately 5×10^6 points. This significant gain stems from VSRE's bin-wise voxel distribution statistics, which directly characterize local point cloud distributions rather than CE's computationally intensive point-wise resampling and projection operations. The bin-wise approach maintains robustness against variations in point count while substantially reducing computational overhead during distribution construction. It can be expected that as the point cloud grows larger, this efficiency gap will continue to widen.

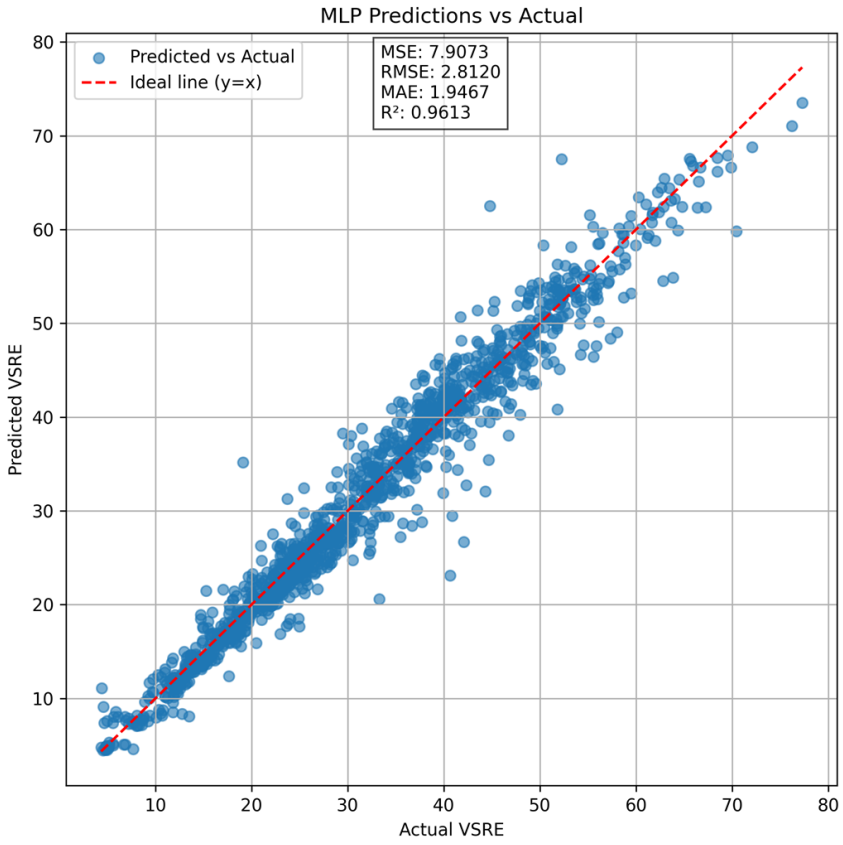


Fig. S6 | Accuracy of VSRE spatial mapping. After initial screening, the best-performing MLP model was selected for VSRE prediction. A grid search was conducted to optimize the model's hyperparameters. The optimal configuration included a network architecture with hidden layers (256, 128, 64), a learning rate of 0.001, a dropout rate of 0.1, a batch size of 64, and batch normalization enabled (bn=True). **Fig. S6** shows the performance of the fine-tuned model on the entire dataset.

Tables

Table S1 | Coordinates and VSRE of the forest LiDAR sample

ID	Longitude	Latitude	Height Min	Height Max	VSRE
1	119.5685625	28.1306505	183.24	202.88	51.65
2	119.5684155	28.13112325	187.47	207.67	42.50
3	119.5715543	28.12946875	186.32	209.44	41.50
4	119.5709303	28.12837475	210.87	230.49	57.84
5	119.5717565	28.128487	188.57	210.84	44.95
6	119.572456	28.13168525	216.79	244.13	35.64
7	119.5608293	28.12190925	193.30	225.95	35.48
8	119.5612788	28.12004175	187.43	216.56	36.28
9	119.5847853	28.22370175	664.73	690.83	22.01
10	119.5853543	28.22377425	686.36	703.42	43.18
11	119.5738523	28.1344965	239.03	263.80	63.46
12	119.582558	28.22422925	664.00	693.83	33.45
13	119.5827285	28.22397425	642.43	669.91	39.11
14	119.5851405	28.22349825	664.55	690.21	27.85
15	119.5854633	28.22349775	666.43	692.79	21.71
16	119.5796698	28.08116525	210.64	246.07	27.44
17	119.57957	28.08140575	213.75	244.79	36.95
18	119.5799563	28.08170375	242.39	271.01	32.08
19	119.5189663	28.0811495	230.55	270.53	14.22
20	119.5828998	28.0871795	255.99	276.02	47.33
21	119.5826563	28.08743325	235.27	262.78	61.26
22	119.5823155	28.08752925	216.27	245.22	50.07
23	138.7030558	25.45691525	205.01	234.99	47.97
24	119.5822858	28.08710225	232.16	256.44	43.09
25	119.5814625	28.08639575	212.10	235.55	42.61
26	119.5811858	28.086243	209.90	240.85	36.33
27	119.5807738	28.08612875	197.18	231.28	34.12
28	119.5807183	28.08586275	207.02	232.85	43.42
29	119.580713	28.08558425	208.65	232.27	31.59
30	119.5807443	28.08524175	208.14	230.94	28.19

125 **Table S2 |** VSRE of different vegetation communities in different provinces

Province	*Spartina* spp.	*P. australis*	*Scirpus* spp.	Mixed Marsh	*Suaeda* spp.	Mangrove
Liaoning	12.80	14.05	/	/	4.38	/
Shandong	10.75	15.43	/	/	5.42	/
Jiangsu	7.15	7.73	9.49	/	12.12	/
Shanghai	/	21.84	6.04	14.49	/	/
Zhejiang	13.79	11.53	/	13.94	/	/
Fujian	12.87	20.12	9.04	15.86	6.28	55.39
Guangdong	23.31	/	/	22.66	/	28.52
Guangxi	15.96	17.65	/	/	/	20.92

126

Table S3 | Average accuracy of vegetation classification in different provinces.

Province	Type	Samples	Mean_UA	Mean_PA	Mean_F1-Scores	Kappa	OA
Guangxi	3	423	0.97	0.96	0.96	0.95	0.97
Guangdong	4	321	0.96	0.95	0.96	0.93	0.96
Fujian	5	524	0.85	0.87	0.86	0.84	0.89
Zhejiang	5	510	0.90	0.91	0.90	0.87	0.90
Shanghai	5	460	0.83	0.86	0.85	0.86	0.90
Jiangsu	6	1335	0.84	0.90	0.85	0.88	0.91
Shandong	5	870	0.92	0.89	0.91	0.89	0.92
Hebei	4	203	0.90	0.90	0.90	0.87	0.90
Liaoning	4	746	0.86	0.89	0.87	0.87	0.91
Tianjin	4	280	0.87	0.87	0.87	0.85	0.90

Indices such as UA (User's Accuracy, 1), PA (Producer's Accuracy, 2), OA (Overall Accuracy, 3), and Kappa (5) are commonly used as statistical metrics of consistency between classified data and observed data. The range generally varies from 0.20, indicating slight consistency, 0.20~0.40 for fair consistency, 0.40~0.60 for moderate consistency, 0.60~0.80 for substantial consistency, and 0.80~1 for almost perfect consistency. The F1-score (4) is a metric for measuring the performance of a classification model, particularly suited to situations with imbalanced class distributions. The F-score ranges from 0 to 1, where values closer to 1 indicate better precision and recall, while values closer to 0 indicate poorer performance.

$$UA = \frac{TP}{TP+FP} \quad (1)$$

$$PA = \frac{TP}{TP+FN} \quad (2)$$

$$OA = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$F1-Socres = \frac{2PA*UA}{PA+UA} \quad (4)$$

$$Kappa = \frac{N^2*OA-\sum_{i=1}^{n} a_i*b_i}{N^2-\sum_{i=1}^{n} a_i*b_i} \quad (5)$$

TP (True Positive, correctly classified positive samples by the model), TN (True Negative, correctly classified negative samples by the model), FP (False Positive, incorrectly classified positive samples by the model), FN (False Negative, incorrectly classified negative samples by the model), a_i (the total actual observations for a sample type), b_i (the total model classifications for a sample type).

Table S4 | Saltmarsh area in different provinces (km^2).

Province	*Mudflat*	*Spartina* spp.	*P.australis*	*Scirpus* spp.	*Suaeda* spp.	*Mixed marsh*	*Mangrove*	Vegetated	Total
Fujian	1048.7	68.6	13.6	93.0	/	/	13.1	175.2	1223.9
Guangdong	712.5	105.8	/	/	/	1.0	67.5	106.8	819.3
Guangxi	573.3	127.7	/	/	/	/	61.6	127.7	701
Hebei	453.7	1.2	1.2	/	1.0	/	/	3.4	457.1
Jiangsu	3196.0	141.6	77.1	46.2	39.1	0.7	/	304.7	3500.7
Liaoning	1539.1	/	66.4	/	75.6	20.8	/	162.8	1701.9
Shandong	1752.8	97.3	65.5	/	63.1	13.0	/	238.9	1991.7
Shanghai	422.9	151.9	98.0	151.5	/	14.9	/	416.3	839.2
Tianjin	92.1	1.6	1.4	/	0.7	/	/	3.7	95.8
Zhejiang	894.6	112.2	98.5	52.7	/	/	0.0	263.4	1158

Table S5| Model Comparison Based on Cross-Validation MAE

Category	Method	CV MAE Mean
Statistical	Ridge	6.6151
Machine Learning	Random Forest	4.6750
Machine Learning	Gradient Boosting	5.0676
Machine Learning	SVR	7.8506
Machine Learning	XGBoost	4.5278
Machine Learning	KNN	4.2419
Machine Learning	Decision Tree	5.8195
Deep Learning	MLP	3.8087
Deep Learning	Residual	3.8815
Deep Learning	Auto Encoder	3.9343
Deep Learning	Transformer	4.2711

MAE (mean absolute error) is the average magnitude of prediction errors and serves as an indicator of overall model accuracy. All metrics in this table were calculated as averages from tenfold cross-validation on the dataset. To select the most appropriate model for the task, a set of machine learning and deep learning algorithms was compared using their default parameter settings.

Terrain simulation

To focus on structural diversity, the input point cloud data must first undergo terrain undulation removal. A Gaussian mixture model in the horizontal plane efficiently fits rugged terrain due to its multi-modal nature. Given a set of LiDAR point cloud data $\mathcal{P} = \{(\mathrm{x_I}, \mathrm{y_I}, \mathrm{z_i})\}_{\mathrm{i=1}}^{\mathrm{N}}$, where x_i, y_i, and z_i denote the 3D coordinates of the i^{th} point, the fitted surface $\hat{z} = f(x, y)$ can be expressed as(Wolfe, 1970):

$$\hat{z}(x, y) = \sum_{k=1}^{K} \omega_k \cdot G_k(x, y) \tag{S5}$$

Where $G(x, y) \sim \mathcal{N}(\vec{\mu}, \Sigma)$ is a bivariate Gaussian distribution. For the k^{th} distribution:

$$G_k(x, y) = A_k exp\left(-\frac{1}{2}\begin{bmatrix} x - \mu_k^{(x)} \\ y - \mu_k^{(y)} \end{bmatrix}^T \Sigma_k^{-1} \begin{bmatrix} x - \mu_k^{(x)} \\ y - \mu_k^{(y)} \end{bmatrix}\right) \tag{S6}$$

adopt learnable A_k as the amplitude, $\left(x - \mu_k^{(x)}, y - \mu_k^{(y)}\right)$ as the distribution center. Σ_k is the covariance matrix defining the distribution's orientation and spread:

$$\Sigma_k = \begin{bmatrix} {\sigma_k^{(x)}}^2 & \rho_k \sigma_k^{(x)} \sigma_k^{(y)} \\ \rho_k \sigma_k^{(x)} \sigma_k^{(y)} & {\sigma_k^{(y)}}^2 \end{bmatrix} \tag{S7}$$

Where ρ_k is the correlation coefficient and $\sigma_k^{(\cdot)}$ denotes the standard deviation in a given direction. To effectively fit real terrain, independent distributions with weights ω_k are linearly combined to form peaks ($\omega_k > 1$) or valleys ($\omega_k < 1$). The point cloud $\mathcal{P}$ is divided into grids (P × Q) along the x and y directions, extracting the minimum z-coordinate points in each grid as the ground morphology data S:

$$\mathrm{S} = \left\{\arg\min_{(x,y,z)\in\mathcal{P}} \mathrm{z} \mid (\mathrm{x}, \mathrm{y}) \in (\mathrm{x_p}, \mathrm{x_{p+1}}] \times (\mathrm{y_p}, \mathrm{y_{p+1}}]\right\}_{\mathrm{p=1,q=1}}^{\mathrm{P,Q}} \tag{S8}$$

Residuals are defined as:

$$r_i = z_i - \hat{z}(x_i, y_i) \forall z_i \in S \tag{S9}$$

To further enhance the robustness of outlier fitting, we adopt the Tukey loss function(Beaton and Tukey, 1974; Belagiannis et al., 2015), which suppresses the

influence of mild outliers without compromising overall fitting performance. This improves the robustness of terrain normalization:

$$\mathcal{L}(r_i) = \begin{cases} \frac{\delta^2}{6}\left(1 - \left(1 - \frac{{r_i}^2}{\delta^2}\right)^3\right), |r_i| \le \delta \\ \frac{\delta^2}{6}, |r_i| > \delta \end{cases} \quad \text{(S10)}$$

To prevent degeneration of the covariance matrix and improve model stability, a regularization term is added to the final loss(Grave et al., 2011). Consequently, the optimization objective for the fitting stage is summarized as:

$$\arg\min_{\Theta} \sum_{i=1}^{P\times Q} \mathcal{L}(r_i) + \lambda \sum_{k=1}^{K} \mathrm{Tr}(\Sigma_k^{-1}) \quad \text{(S11)}$$

Where the model parameters are defined as:

$$\Theta = \{(\omega_k, A_k, \vec{\mu}_k, \Sigma_k)\}_{k=1}^{K} \quad \text{(S12)}$$

Reference

Beaton, A. E. and Tukey, J. W.: The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data, Technometrics, 16, 147-185, 1974.

Belagiannis, V., Rupprecht, C., Carneiro, G., and Navab, N.: Robust optimization for deep regression, Proceedings of the IEEE international conference on computer vision, 2830-2838,

Grave, E., Obozinski, G. R., and Bach, F.: Trace lasso: a trace norm regularization for correlated designs, Advances in neural information processing systems, 24, 2011.

Wolfe, J. H.: Pattern clustering by multivariate mixture analysis, Multivar. Behav. Res., 5, 329-350, 1970.