

Beyond Boundaries: Identifying Spatial Scales of Enclave–Habitat Associations Under Alternative National Park Management Scenarios

Section S1. Data sources and preprocessing

The data used included Landsat remote sensing imagery, digital elevation model (DEM) data, mean annual temperature, mean annual precipitation, road network data, water data, settlement data, and enclave boundary data for Wuyishan National Park. Raster datasets were preprocessed in ArcGIS 10.8 through mosaicking, projection, resampling, and clipping to ensure a spatial resolution of 30 m for all raster layers and a common coordinate system (WGS 1984 UTM Zone 50N). Land-use data for 2005, 2020, and 2025 were obtained through visual interpretation of Landsat remote-sensing imagery and Google Earth satellite imagery. The 2005 and 2020 land-use datasets were used to simulate future land-use changes within Wuyishan National Park and its enclaves under multiple scenarios. The 2025 dataset was used to characterize current conditions in the national park and its enclaves, including land-use composition, landscape patterns, and habitat quality. Land-use data from 2005 and 2025 were also used to assess the spatial associations between changes in enclave landscape patterns and changes in habitat quality in the national park.

Table S1. Data sources in the study.

Data	Source	Data type	Spatial resolution
Landsat remote sensing imagery	Google Earth Engine platform (https://code.earthengine.google.com/)	–	30 m
Digital elevation model (DEM) data	ALOS PALSAR dataset (https://search.asf.alaska.edu)	Raster	12.5 m
Mean annual temperature data	National Earth System Science Data Center (https://www.geodata.cn)	Raster	1000 m
Mean annual precipitation data	National Earth System Science Data Center (https://www.geodata.cn)	Raster	1000 m
Road network data	OpenStreetMap dataset (https://www.openstreetmap.org)	Vector	–
Water data	OpenStreetMap dataset (https://www.openstreetmap.org)	Vector	–
Settlement data	OpenStreetMap dataset (https://www.openstreetmap.org)	Vector	–
Enclave boundary data for Wuyishan National Park	Wuyishan National Park Administration	Vector	–

Section S2. Rationale for driving factor selection in the PLUS Model

Table S2. Drivers of natural and social factors.

Data type	Driving factor	Description	Unit
Natural factor	DEM[1]	Calculated from the DEM data of Wuyishan National Park enclaves	m
Natural factor	Mean annual temperature[2]	Derived from spatial interpolation of meteorological data	°C
Natural factor	Mean annual precipitation[2]	Derived from spatial interpolation of meteorological data	mm
Natural factor	Slope[3]	Calculated from the DEM data of Wuyishan National Park enclaves	°
Natural factor	Aspect[3]	Calculated from the DEM data of Wuyishan National Park enclaves	°
Socioeconomic factor	Distance to roads[1]	Calculated using Euclidean distance in ArcGIS	m
Socioeconomic factor	Distance to water[2]	Calculated using Euclidean distance in ArcGIS	m
Socioeconomic factor	Distance to settlements[3]	Calculated using Euclidean distance in ArcGIS	m

Section S3. Rationale for the landscape pattern index selection

Table S3. Selection and ecological significance of landscape pattern indices.

Landscape pattern index	Unit of measure	Typology	Ecological significance
Patch density (PD)	Number per 100 hectares	Density and variation metric	Indicates the number of patches per unit area, reflecting the spatial configuration of the landscape and the degree of landscape fragmentation.
Largest patch index (LPI)	%	Area metric	Measures the proportion of the largest patch area relative to the total landscape area, reflecting the dominance of the largest patch in the landscape structure.
Aggregation index (AI)	%	Aggregation–dispersion metric	Measures the spatial connectivity and aggregation of patches of the same type, reflecting landscape integrity and its capacity to support ecological processes.
Contagion index (CONTAG)	%	Aggregation–dispersion metric	Measures the distribution and aggregation of different patch types, reflecting landscape continuity and fragmentation. Higher values indicate stronger continuity and lower fragmentation.
Shannon’s diversity index (SHDI)	/	Aggregation–dispersion metric	Measures the diversity of patch types in the landscape, reflecting the richness of species or patch types. Higher values indicate greater landscape diversity and ecological complexity.
Shannon’s evenness index (SHEI)	/	Aggregation–dispersion metric	Measures the evenness of patch-type distribution. Values closer to 1 indicate a more even distribution of patch types and a more balanced ecosystem.

Section S4. Habitat quality model formulation and parameters

The Habitat Quality module of the InVEST model was used to estimate habitat quality within the enclaves and Wuyishan National Park. Model parameters were specified with reference to previous studies and the InVEST User Guide, taking field observations into account. The inputs included land-use maps, threat raster layers, a threat parameter table, a habitat suitability and sensitivity table, and the half-saturation constant.

Farmland, tea plantations, and impervious surfaces were selected as habitat threats. Separate raster layers for each threat were extracted from the corresponding land-use maps for 2025 and for each scenario in 2035 and 2050. Within the valid analysis area, cells containing the relevant threat were assigned a value of 1, and other cells were assigned a value of 0. The maximum influence distance, relative weight, and distance-decay type assigned to each threat are reported in Table S4. Habitat suitability scores were assigned to each land-use type on a scale from 0 to 1, where 0 represents land classified as non-habitat and 1 represents the highest suitability within the assessment. The habitat suitability scores and the sensitivity of each land-use type to the three threats are reported in Table S5.

Table S4. Threat factors, maximum influence distances, relative weights, and distance-decay types.

Threat factor	Maximum influence distance (km)	Relative weight	Distance-decay type
Farmland	2.0	0.7	Linear
Impervious surface	6.0	1.0	Exponential
Tea plantation	4.0	0.5	Linear

Table S5. Habitat suitability and sensitivity of land-use types to each threat.

Land-use type	Habitat suitability	Sensitivity to farm-land	Sensitivity to tea plantation	Sensitivity to im-pervious surface
Farmland	0.5	0	0.6	0.7
Tea plantation	0.5	0.6	0	0.8
Forest land	1	0.7	0.8	0.9
Water	0.9	0.4	0.5	0.7
Impervious surface	0	0	0	0
Grassland	0.8	0.7	0.8	0.9

Habitat quality and threat distance decay were calculated using Equations (1) and (2), respectively:

$$Q_{xj} = H_j \left(1 - \frac{D_{xj}^z}{D_{xj}^z + k^z} \right) \quad (1)$$

$$i_{rxy} = \begin{cases} 1 - \frac{d_{xy}}{d_{rmax}}, & \text{linear decay} \\ \exp\left(\frac{-2.99d_{xy}}{d_{rmax}}\right), & \text{exponential decay} \end{cases} \quad (2)$$

Where Q_{xj} represents the habitat quality of land-use type j in grid cell x ; H_j is the habitat suitability of land-use type j ; D_{xj} is the habitat degradation degree of land-use type j in grid cell x ; k is the half-saturation constant, set to 0.5; z is a scaling parameter, set to 2.5; i_{rxy} represents the influence of threat r in source grid cell y on grid cell x ; d_{xy} is the distance between cells x and y ; and d_{rmax} is the maximum effective distance of threat r .

Section S5. K-means classification of enclave types

Enclaves were classified using the proportions of six land-use classes within each enclave. Seven enclaves without valid land-use values were excluded before classification, leaving 145 enclaves. The land-use values were converted to proportions and transformed using the centred log-ratio (CLR) transformation. K-means clustering was performed on the CLR-transformed variables using the Hartigan–Wong algorithm, 50 random initialisations, and a fixed random seed

of 1. Candidate solutions with $K = 4 - 8$ were evaluated using the mean silhouette coefficient calculated from Euclidean distances in CLR space. A minimum cluster size of 15 enclaves was required to provide sufficient observations for subsequent type-specific spatial association analyses. The $K = 4$ solution was therefore selected. The final $K = 4$ model was refitted using 100 random initialisations and the same random seed. The CLR cluster centres are reported in Table S6.

The initial k-means classification produced four enclave types based on their mean land-use composition. Cluster 1 was identified as the tea plantation-dominated type and contained 41 enclaves. Cluster 2 was identified as the built-up disturbance type, characterised by the combined presence of tea plantations and impervious surfaces, and contained 31 enclaves. Cluster 3 was identified as the tea-forest mosaic type, with tea plantation and forest land as the dominant land-use classes, and contained 48 enclaves. Cluster 4 was identified as the agricultural-use type because it had the highest mean farmland proportion among the four clusters and contained 25 enclaves. Within Cluster 3, enclaves with forest land $\geq 60\%$, tea plantations $\leq 40\%$, and impervious surfaces $\leq 10\%$ were further classified as forest-dominated. Thirteen enclaves met these criteria. The final classification therefore comprised 41 tea plantation-dominated enclaves, 31 built-up disturbance enclaves, 35 tea-forest mosaic enclaves, 25 agricultural-use enclaves, and 13 forest-dominated enclaves. The forest-dominated enclaves were excluded from the subsequent spatial association analyses. Their land-use composition was similar to that of the surrounding forest matrix, providing limited contrast in the land-use conditions examined. The subsequent spatial association analyses therefore included 132 enclaves representing the other four types.

Table S6. CLR cluster centres and sample sizes for the selected $K = 4$ solution

Cluster	Farmland CLR	Tea plantation CLR	Forest land CLR	Water CLR	Impervious surface CLR	Grassland CLR	Number of enclaves
1	−2.065960	11.463300	−2.349340	−2.349340	−2.349340	−2.349340	41
2	−5.250740	5.682768	1.809728	−2.873550	6.221882	−5.590090	31
3	−4.010590	7.692834	8.219941	−4.138290	−3.847050	−3.916850	48
4	3.971858	5.430906	0.187812	−6.175900	3.504792	−6.919470	25

Section S6. ZOI exposure construction and spatial model evaluation

For enclave i , landscape metric m , and scenario s , source strength was calculated as the area-weighted signed change in the landscape metric:

$$S_{i,m,s} = A_i \Delta M_{i,m,s} \quad (1)$$

where A_i is the area of enclave i , and $\Delta M_{i,m,s}$ is the signed change in landscape metric m between the corresponding simulation periods. The shortest distance between sampling point p and enclave i was denoted by d_{pi} .

For a candidate radius r , the linear distance-decay function was defined as:

$$w(d_{pi}, r) = \begin{cases} 1 - \frac{d_{pi}}{r}, & d_{pi} \leq r \\ 0, & d_{pi} > r \end{cases} \quad (2)$$

The distance-weighted contribution of enclave i to sampling point p was then calculated as:

$$C_{pi,m,s}(r) = S_{i,m,s} w(d_{pi}, r) \quad (3)$$

Under the dominant-source assumption, only the contribution from the enclave with the largest absolute contribution was retained for each sampling point:

$$E_{p,m,s}(r) = C_{pi^*,m,s}(r), i^* = \arg \max_i | C_{pi,m,s}(r) | \quad (4)$$

where $E_{p,m,s}(r)$ is the *ZOI* exposure value used in the regression model.

The resulting $E_{p,m,s}(r)$ was entered into the following model:

$$\Delta HQ_{p,s} = \beta_0 + \beta_1 E_{p,m,s}(r) + \beta_2 Slope_p + \beta_3 Road_p + \beta_4 Home_p + \beta_5 DEM_p + \varepsilon_p \quad (5)$$

For each scenario, 30,000 valid raster cells within the national park and outside the enclave masks were sampled. Cells with missing values for habitat quality change or any control variable were excluded before sampling. The sampled cells were assigned to five spatial cross-validation folds using 10 km spatial blocks, and the same fold structure was retained across candidate radii and landscape metrics. For fold k , the model was fitted using the training blocks and evaluated using the held-out blocks.

The fold-level coefficient of determination was calculated using the mean response in the training data to calculate the total sum of squares for the test data:

$$CV - R^2 = \frac{1}{K} \sum_{k=1}^K \left[1 - \frac{\sum_{p \in T_k} (y_p - \hat{y}_{p,-k})^2}{\sum_{p \in T_k} (y_p - \bar{y}_{-k})^2} \right] \quad (6)$$

where T_k denotes the test cells in fold k , $\hat{y}_{p,-k}$ is the prediction generated by the model trained without fold k , and $\bar{y}_{-k}$ is the mean response of the corresponding training data.

As an additional spatial diagnostic, residual spatial autocorrelation was assessed using full-sample fitted residuals from the control-only models and the *ZOI*-augmented models at their selected predictive spatial thresholds. A symmetrised eight-nearest-neighbour spatial-weights matrix with row standardisation was used to calculate Moran's I . Moran's I was positive and statistically significant for all models ($p < 0.001$). Values for the control-only models ranged from 0.6837 to 0.7330, whereas those for the overall, type-specific, and area-specific *ZOI*-augmented models ranged from 0.6809 to 0.7314, 0.6770 to 0.7327, and 0.6759 to 0.7327, respectively. The inclusion of the *ZOI* exposure variable produced small reductions in residual spatial autocorrelation, but substantial residual spatial dependence remained.

Section S7. Rationale for selecting the maximum candidate radius and predictive gain criterion

To evaluate the sensitivity of the predictive spatial thresholds to the search range and predictive gain criterion, 12 parameter combinations were compared, comprising maximum candidate radii of 30, 40, 50, and 60 km and predictive gain criteria of 90%, 95%, and 99%. Candidate radii for the enclave *ZOI* started at 0.5 km and increased in 0.5 km intervals. For each search range, the maximum incremental cross-validated predictive gain was first identified, and the minimum candidate radius reaching at least the specified proportion of that maximum was then identified. Thus, the threshold for each search range was calculated relative to the maximum predictive gain within that range. Sampling points, control variables, and spatial cross-validation folds were kept identical across all comparisons. The following summaries are based on five-fold spatial block cross-validation under the four future scenarios and cover 24 combinations in the overall analysis, 96 in the type-specific analysis, and 72 in the area-specific analysis.

S7.1 Sensitivity to the maximum candidate radius

Changes in the maximum candidate radius affected the three analytical levels differently. In the overall analysis, the number of combinations with detected thresholds increased from 14 at a maximum radius of 30 km to 15 at 40 km and remained unchanged thereafter. In the type-specific analysis, 77 thresholds were detected for each of the four search ranges. In the area-specific analysis, the number of detected thresholds increased from 47 at 30 km to 52 at 50 km and remained at 52 when the maximum candidate radius was extended to 60 km (Table S7).

Table S7. Number of detected thresholds under different maximum candidate radii and stability between adjacent search ranges

Analysis level	Total number of combinations	Detected at 30 km	Detected at 40 km	Detected at 50 km	Detected at 60 km
Overall	24	14	15	15	15
Type-specific	96	77	77	77	77
Area-specific	72	47	50	52	52

note: The analytical combinations were scenario × landscape metric for the overall analysis, enclave type × scenario × landscape metric for the type-specific analysis, and enclave area group × scenario × landscape metric for the area-specific analysis.

All thresholds in the overall analysis remained unchanged between the 50 and 60 km search ranges, and neither the maximum predictive gain nor the radius at which it occurred changed. However, the PD threshold under ECMS was 30 km when the maximum candidate radius was set to either 30 or 40 km, but increased to 48.5 km when the search range was extended to 50 km. In the type-specific analysis, the main change occurred for PD in tea plantation-dominated enclaves under ECM, for which the threshold increased from 30 to 35 km when the maximum candidate radius increased from 30 to 40 km and remained unchanged as the search range was extended to 50 and 60 km. Across the four future scenarios, no type-specific threshold reached the upper limit of the search range when the maximum candidate radius was 40 km or greater.

The area-specific analysis showed greater sensitivity to the search range. Thresholds were detected for some combinations only after the search range was extended, whereas thresholds for other combinations shifted to larger radii. Representative changes are shown in Table S8.

Table S8. Representative combinations sensitive to the maximum candidate radius and their predictive spatial thresholds

Analysis level	Group	Scenario	Metric	30 km maximum	40 km maximum	50 km maximum	60 km maximum
Overall	/	ECMS	LPI	ND	39.0	39.0	39.0
		ECMS	PD	30.0	30.0	48.5	48.5
Type-specific	Tea plantation-dominated	ECM	PD	30.0	35.0	35.0	35.0
		Small	BAU	AI	6.0	34.0	34.0
		Small	BAU	LPI	ND	39.5	48.5
		Small	BAU	SHDI	ND	39.0	41.0
Area-specific	Small	ECMS	LPI	30.0	38.0	42.5	42.5
	Small	ECMS	PD	7.0	7.0	7.0	58.5
	Medium	BCM	CONTAG	ND	ND	50.0	50.0
	Medium	BCM	PD	ND	ND	50.0	58.5
	Large	ECMS	PD	30.0	30.0	48.5	48.5

note: Thresholds are reported in km. ND indicates that no positive incremental predictive gain was detected within the corresponding candidate range under the model validation procedure. Only representative combinations are shown; the table does not include all combinations that changed across search ranges.

S7.2 Sensitivity to the predictive gain criterion

With the maximum candidate radius fixed at 60 km, the median difference between the thresholds obtained using the 90% and 99% criteria was 0.5 km in the overall, type-specific, and area-specific analyses. In the overall analysis, all 15 detected thresholds differed by no more than 1 km between these two criteria, while all 77 detected thresholds in the

type-specific analysis differed by no more than 2 km. In the area-specific analysis, 42 of the 52 detected thresholds differed by no more than 1 km and 46 differed by no more than 2 km, although a small number of combinations showed larger differences (Table S9). The most pronounced exception in the area-specific analysis was PD for small enclaves under ECMS, for which the 90% threshold was 7 km, whereas both the 95% and 99% thresholds were 58.5 km (Table S10). The maximum incremental predictive gain for this combination within the 60 km candidate range was approximately 0.000262, indicating that a small difference in predictive gain could correspond to a large shift in the threshold radius. This threshold estimate was therefore sensitive to the choice of criterion and should not be interpreted as a precise management distance.

Table S9. Differences between thresholds derived using the 90% and 99% predictive gain criteria at a maximum candidate radius of 60 km

Analysis level	Number of combinations with detected thresholds	Median difference (km)	Maximum difference (km)	Difference ≤ 1 km	Difference ≤ 2 km
Overall	15	0.5	1.0	15/15	15/15
Type-specific	77	0.5	2.0	71/77	77/77
Area-specific	52	0.5	51.5	42/52	46/52

note: The difference was calculated as the 99% threshold minus the 90% threshold for the same analytical combination. Only combinations with detected thresholds were included.

Table S10. Area-specific combinations with threshold differences greater than 2 km between the 90% and 99% predictive gain criteria

Area group	Scenario	Metric	90% threshold (km)	95% threshold (km)	99% threshold (km)
Small	BAU	LPI	50.0	51.5	54.0
Small	BAU	SHDI	39.0	41.0	44.0
Small	ECM	PD	19.0	20.0	22.0
Small	ECMS	LPI	39.5	42.5	44.5
Small	ECMS	PD	7.0	58.5	58.5
Medium	BCM	PD	57.0	58.5	60.0

Based on these comparisons, 60 km was adopted as the common maximum candidate radius. First, maximum candidate radii of 30 and 40 km did not capture predictive gains that emerged at larger radii for some combinations, which could result in thresholds occurring at the search boundary, shifts in threshold radius, or failure to detect thresholds identified at larger spatial scales. Second, although the results of the overall and type-specific analyses under the four future scenarios were stable between 50 and 60 km, results for some area-specific combinations continued to change beyond 50 km. Using 60 km as the common maximum candidate radius therefore ensured that these observed changes were included in the analysis. The 95% criterion was retained as the main predictive gain criterion. It identifies the minimum candidate radius at which at least 95% of the maximum incremental predictive gain within the candidate range is reached, allowing a 5% tolerance relative to the observed maximum. Compared with the 90% criterion, it retains a larger proportion of the maximum predictive gain, while compared with the 99% criterion, it allows a larger departure from the maximum. The 95% criterion was therefore used as a common operational compromise. The sensitivity results showed only small threshold shifts between the 90% and 99% criteria for most combinations, supporting its use in the main analysis.

Section S8. Results of habitat quality assessment within enclaves under different scenarios

Habitat quality in Wuyishan National Park enclaves showed a generally consistent trend, with improvements projected for both 2035 and 2050 under all scenarios. In terms of habitat quality classes, the areas of low- and medium-low-quality habitats generally decreased, whereas medium-high- and high-quality habitats expanded, although the magnitude of improvement varied across scenarios. Habitat quality improvement was most pronounced under BCM, with a substantial expansion of high-quality habitats and a marked reduction in low-quality habitats. BAU and ECMS showed moderate improvement, with steady expansion of high-quality habitats and a continued decline in medium-low-quality habitats. ECM showed relatively weaker improvement. Spatially, enclave habitat quality was characterized by contiguous high-quality areas and scattered low-quality patches. Habitat quality remained high in the contiguous forested areas in the northern and central parts of the national park, whereas low-quality habitats were mainly concentrated around southeastern villages and park margins (Figure S1–S5).

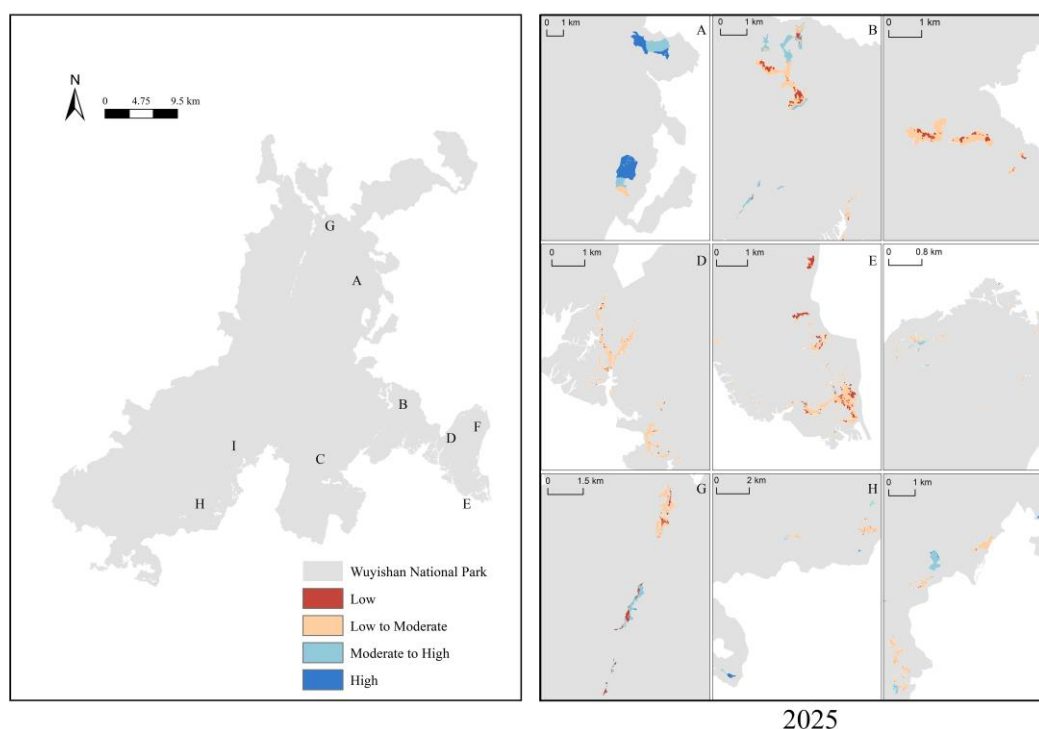


Figure S1. Habitat quality distribution in Wuyishan National Park enclaves in 2025.

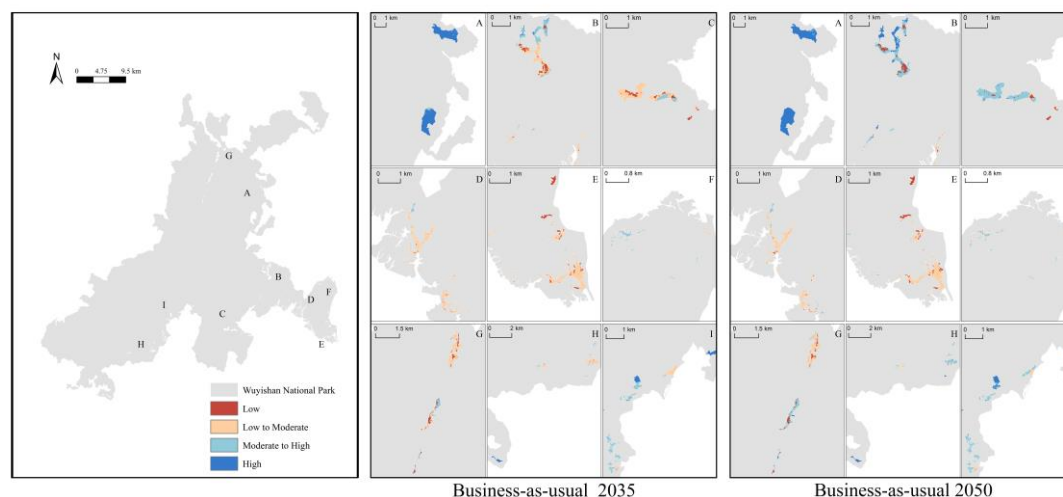


Figure S2. Habitat quality distribution of Wuyishan National Park enclaves under BAU in 2035 and 2050.

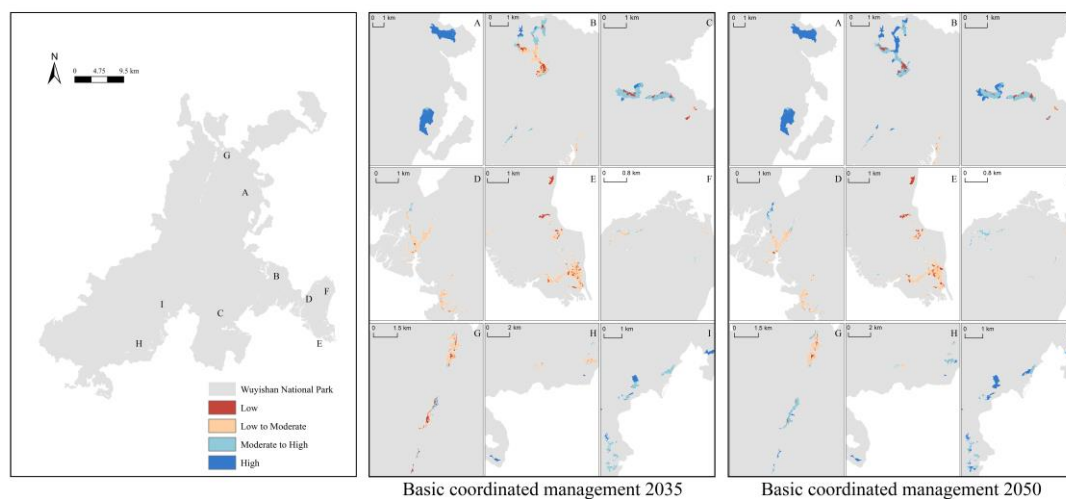


Figure S3. Habitat quality distribution of Wuyishan National Park enclaves under BCM in 2035 and 2050.

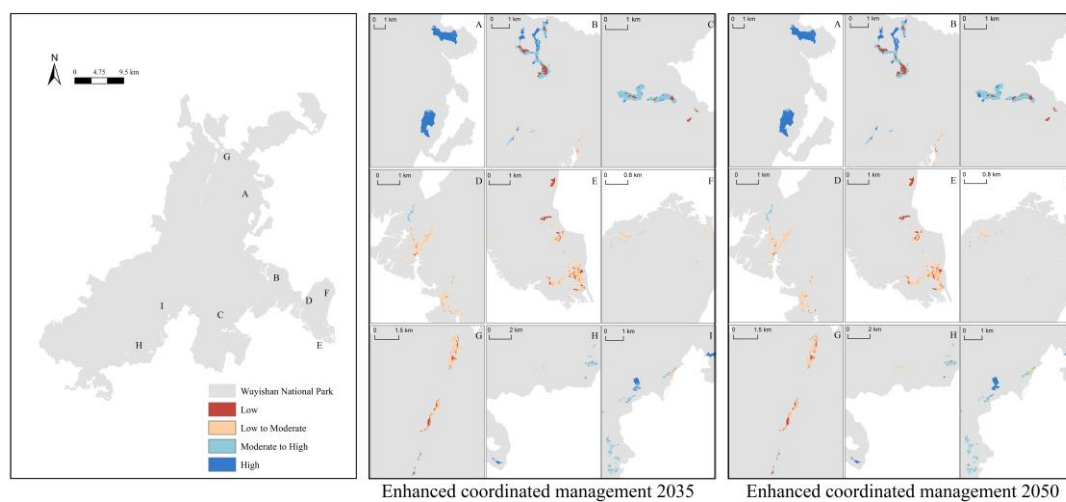


Figure S4. Habitat quality distribution of Wuyishan National Park enclaves under ECM in 2035 and 2050.

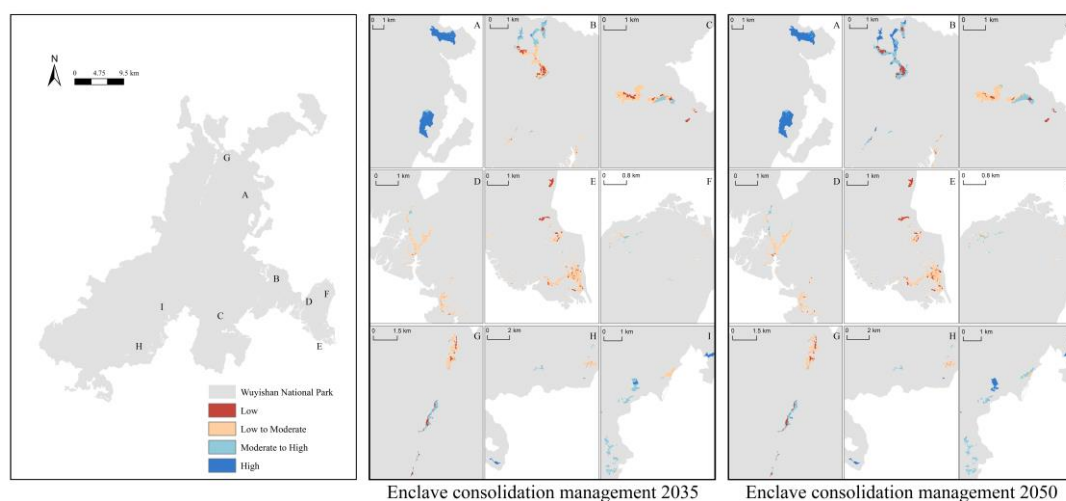


Figure S5. Habitat quality distribution of Wuyishan National Park enclaves under ECMS in 2035 and 2050.

References

1. Huang, C.; Zhou, Y.; Wu, T.; Zhang, M.; Qiu, Y. A cellular automata model coupled with partitioning CNN-LSTM and PLUS models for urban land change simulation. *Journal of Environmental Management* **2024**, *351*, 119828.
2. Liu, J.; Kong, X.; Zhu, Y.; Zhang, B. A study on land use change simulation based on PLUS model and the U-Net structure: a case study of Jilin Province. *Ecological Indicators* **2025**, *176*, 113619.
3. Tan, G.; Zhu, J.; Chen, Z. Evolution and dynamic prediction of urban-rural settlement patterns in a Chinese county-level city under new-type urbanization. *Habitat International* **2025**, *165*, 103540.