Supplementary File 1

Integrating ecological networks and multi-scenario optimization: a novel framework for constructing ecological security pattern

S1. Calculation of ecosystem services

S1.1 Habitat quality

This study evaluated habitat quality based on the InVEST model. The habitat quality model, utilizing data on human disturbance, land cover, and expert knowledge, generates reliable indicators of habitat response to threat sources[1] (**Supplementary Table S7**). The InVEST habitat quality model integrates data on land cover and biodiversity threat factors to produce habitat quality maps. Habitat quality is defined as the capacity of ecosystems to offer conditions conducive to individual and population survival, contingent upon the availability of subsistence resources, reproductive opportunities, and organism presence[2]. We determined the threat data for different land use types in the study area according to the InVEST User's Guide. The calculation formula of habitat quality is as follows:

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

$$D_{xj} = \sum_{r=1}^{R}\sum_{y=1}^{Y_r}\left(\frac{w_r}{\sum_{r=1}^{R} w_r}\right) \times r_y \times i_{rxy} \times \beta_x \times S_{jr}, \tag{2}$$

where: Q_{xj} is the habitat quality of raster x of land use type j; H_j is the habitat suitability of land use type j; D_{xj} is the threat level of raster x of land use type j; z denotes the normalization constant; k is the scaling constant, R is the number of threat factors; r is the threat factor, Y_r is the number of rasters occupied by the threat factor r; w_r is the threat weight of the threat factor r; y is the number of rasters of the threat factor r; i_{rxy} is the effect of the threat factor r on each raster of the habitat; β_x is the accessibility class of raster x; and S_{jr} is the sensitivity index of land use type j to the threat factor r.

The spatial attenuation of threat factors is divided into two categories (**Supplementary Table S8**): linear attenuation and exponential attenuation. The respective characterization functions are as follows:

Linear attenuation

$$i_{rxy} = 1 - \left(\frac{d_{xy}}{d_{r\max}} \right), \quad (3)$$

Exponential attenuation

$$i_{rxy} = \exp\left(-\left(\frac{2.99}{d_{r\max}} \right) \times d_{xy} \right), \quad (4)$$

where: i_{rxy} is the stress degree of the threat factor on grid x; d_{xy} is the distance between grid x and grid y; and $d_{r\max}$ is the maximum influence distance of the threat factor.

Supplementary Table S7 Sensitivity of land use type to habitat threat factor

Land use type	Habitat suitability	Threat factor			
		Farmland	Railroad	Urban land	Road
Urban	0	0	0	0	0
Forest	0.91	0.8	0.65	0.65	3
Cropland	0.5	0.3	0.5	0.5	0.6
Water	0.9	0.6	0.6	0.75	0.6
Grassland	0.7	0.55	0.4	0.6	0.4
Shrub	1	0.5	0.6	0.6	0.4
Bareland	0	0	0	0.2	0

Supplementary Table S8 Threat factors and their stress intensity

Threat factor	Maximum threat distance	Weight	Spatial recession type
Farmland	8	0.7	linear
Urban	10	1	exponential
Road	2	0.6	exponential
Railroad	5	0.6	exponential

S1.2 Water yield

This study estimated water yield using the InVEST model. The water yield model estimates the relative contribution of water from various landscape sections, offering insights into how alterations in land use patterns influence annual surface water yields and hydropower generation. Within the InVEST model, annual water production at each pixel scale is computed based on the Budyko curve and average annual precipitation[3] (**Supplementary Table S9**). The specific equations are as follows:

$$\frac{\mathrm{AET}_x}{P_x}=1+\frac{\mathrm{PET}_x}{P_x}-\left[1+\left(\frac{\mathrm{PET}_x}{P_x}\right)^{\omega_x}\right]^{\frac{1}{\omega_x}}, \tag{5}$$

$$\mathrm{PET}_x=\mathrm{Kc}\left(l_x\right)\times \mathrm{ET}_{0x}, \tag{6}$$

$$\mathrm{ET}_{0x}=0.0013\times 0.408\times \mathrm{RA}\times\left(T_{\mathrm{av}}+17\right)\times\left(T_{\mathrm{D}}-0.0123P_x\right)^{0.76}, \tag{7}$$

$$\omega_x=Z\times\frac{\mathrm{AWC}_x}{P_x}+1.25, \tag{8}$$

$$\begin{aligned}\mathrm{AWC}_x=&\,54.509-0.132\times P_{\mathrm{SAN}}-0.003\times P_{\mathrm{SAN}}^2-0.055\times P_{\mathrm{SIL}}-0.006\times P_{\mathrm{SIL}}^2\\&-0.738\times P_{\mathrm{CLA}}+0.007\times P_{\mathrm{CLA}}^2-2.688\times P_{\mathrm{C}}+0.501\times P_{\mathrm{C}}^2\end{aligned}, \tag{9}$$

where P_x is the annual precipitation (mm) of grid x, and AET_x is the actual evapotranspiration (mm) of grid x. In general, it is not easy to obtain data on actual evapotranspiration directly; therefore, this study uses the approximation of the Budyko curve derived from the ratio of actual evapotranspiration to the precipitation instead[4]. Here, PET_x is the potential evapotranspiration (mm) of grid x; ω_x is a non-physical parameter representing the natural climate-soil attribute, which represents the effective precipitation and annual water availability of the soil; $\mathrm{Kc}(l_x)$ is the evapotranspiration coefficient of the vegetation of the landscape type l in the grid x; ET_{0x} is the average annual evapotranspiration (mm) of the grid; RA is the solar zenith radiation

(MJ/mm); T_{av} and T_D are the mean and difference of the mean annual maximum and mean annual minimum temperatures (°C), respectively; Z is the total water constant stored and released by plants in the soil; AWC_x is the amount of water efficiently utilized by vegetation; P_{SAN}, P_{SIL}, P_{CLA}, and P_C are the contents of sand, silt, clay, and organic carbon in the soil, respectively.

Supplementary Table S9 Biophysical attributes table

Land use type	Root_depth	Kc	LULC_veg
Urban	500	0.3	0
Forest	7000	1	1
Cropland	2100	0.75	1
Water	500	1	0
Grassland	2000	0.65	1
Shrub	5200	0.95	1
Bareland	10	0.2	0

S1.3 Soil conservation

The soil conservation module of the InVEST model was developed based on the Generalized Soil Loss Equation calculation method at the metric scale. The generalised soil loss equation is the most commonly used method for calculating soil retention. It works by subtracting the actual soil erosion from the potential soil erosion for each grid to obtain the soil retention of each grid. The InVEST model is an improvement of the generalised soil loss equation for calculating regional soil retention, i.e., the InVEST model fully considers sediment deposition, and the calculation equation is as follows:

$$\mathrm{SR} = \mathrm{RKLS} - \mathrm{ULSE} + \mathrm{SD}, \tag{10}$$

$$\mathrm{RKLS} = R \times K \times L \times S, \tag{11}$$

$$\mathrm{ULSE} = R \times K \times L \times S \times C \times P, \tag{12}$$

$$\mathrm{SD} = \mathrm{SE} - \left(1 - \mathrm{SDR}\right), \tag{13}$$

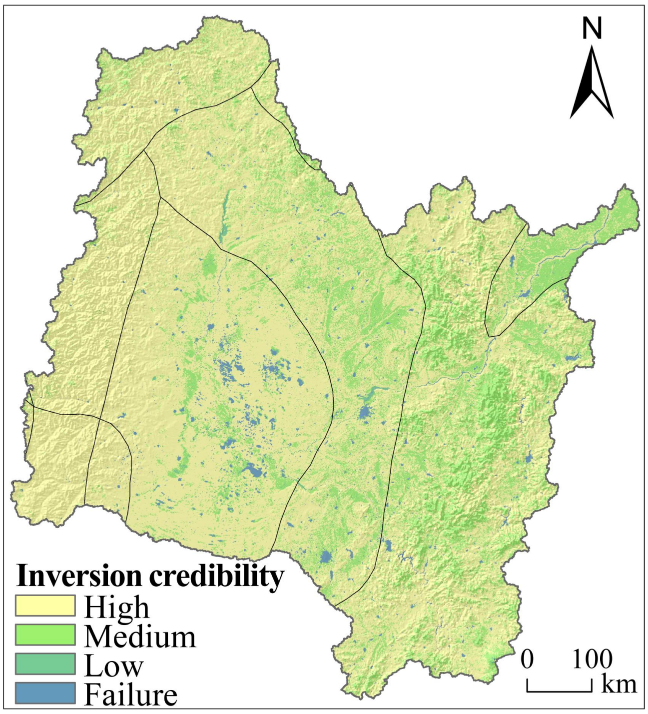
where, SR is soil retention rate; RKLS is soil erosion rate of bare land; ULSE is soil erosion rate of land type with management measures or vegetation cover; SD is sediment deposition rate; SE is sediment movement and migration rate; SDR is sediment transport rate; R is rainfall erosion rate; K is the soil erosion factor, which is measured using data from Williams et al.[5]; L and S are slope length and slope coefficient, C is vegetation cover and management factor; and P is soil and vegetation cover (**Supplementary Table S10**).

Supplementary Table S10 C and P factors for different land use types

Land use type	*C*	*P*
Urban	0	0.1
Forest	0.005	1
Cropland	0.228	0.352
Water	0	0
Grassland	0.112	1
Shrub	0.005	1
Bareland	0.01	1

S1.4 Carbon storage

Carbon storage was replaced by Net primary productivity (NPP) in this study. NPP data were obtained from the MODIS MOD17A3HGF dataset released by National Aeronautics and Space Administration (NASA), which provides year-by-year NPP data at 500 m resolution, synthesized from 8-day net photosynthesis products (MOD17A2H), with a raw spatial resolution of 500 m and a temporal resolution of 1 year. The dataset is corrected for the effects of clouds and aerosols[6], and contains year-by-year quality control data (referred to as NPP_QC) to ensure the quality reliability of NPP products[7]. Based on the statistical analysis of the NPP_QC data in the Songhua River Basin in 2020, the confidence level of the corresponding NPP data was classified into four grades: high, medium, low and inversion failure. The results showed that the cumulative percentage of high and medium confidence levels for vegetation NPP data quality reached more than 97% in total, indicating that this dataset has a very high accuracy in the SRB (**Supplementary Fig. S8**).



Supplementary Fig. S8 Spatial distribution of inversion credibility for NPP of SRB in 2020.

S2. Parameter setting of MPSA

The parameter configurations for Morphological Spatial Pattern Analysis (MSPA) were defined as follows: Foreground Connectivity was assigned a value of "8". Given the limited urban habitat area, an edge width of "1" pixel was selected during the analysis, which equates to an actual distance of 1 km. Transition and Intext features were both activated. Subsequently, based on the characteristics of landscape categories, the cores were delineated. Specific parameter settings and effects are shown in **Supplementary Table S11.**

Supplementary Table S11 The parameter settings in MSPA

Parameter	Option	Setting	Effect
Foreground Connectivity	8/4	8	Fewer bridges
Edge Width	1, 2, 3, 4, …	1	More cores
Intext-set	On/Off	On	Landscape type showing inside and outside
Transition-set	On/Off	On	More bridges

S3. Weighting methodology

S3.1 Entropy-weight method

The entropy weighting method is based on Shannon entropy, which was originally proposed by

Shannon and Weaver[8]. Shannon entropy measures information uncertainty using probability theory. Entropy is well-suited to measure the relative strength of contrasting criteria in representing the average intrinsic information conveyed by a decision. Shannon developed measure H that satisfies the following properties for all p_i in an estimated joint probability distribution P[9]. The weights are determined using the entropy-weight method as follows:

$$P_{ij} = \frac{z_{ij}}{\sum_{i=1}^{m} z_{ij}}, \tag{14}$$

$$H_j = -k \sum_{i=1}^{m} P_{ij} \ln\left(P_{ij}\right), \tag{15}$$

$$k = \left(\ln\left(m\right)\right)^{-1}, \tag{16}$$

$$div_j = 1 - H_j, \tag{17}$$

$$w_j = \frac{div_j}{\sum_{j=1}^{n} div_j}, \tag{18}$$

where, P_{ij} is the normalisation index; z_{ij} is the normalisation matrix; i=1, 2, …, m; j=1, 2, …, n; H_j is the entropy value; k is a constant; div_j is the scatter of information inherent in each criterion; and w_j is the entropy weight of each criterion.

S3.2 GeoDetector

GeoDetector is a set of statistical methods for detecting spatial dissimilarities and revealing the driving forces behind them[10]. In this paper, the optimal parameter GeoDetector was applied to use five classification methods, including equal interval classification, natural breakpoint classification, quantile classification, geometric classification and standard deviation classification[11]. The classification levels were set from 3 to 7, and the q-values under different classification methods and classification levels were calculated respectively to filter the classification method and classification level with the largest q-value of the influencing factors as the optimal parameter for GeoDetector analysis, and use it as the one of the resistance surface weights. The GeoDetector calculation is as follows:

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \times \sigma_h^2}{N \times \sigma^2}, \tag{19}$$

where, q is the divergence factor; L is the number of variable categories; N_h and N are the number of cells in category h and the number of cells in the whole region; h is a specific type, the number of cells in category h and the number of cells in the whole region; and σ_h and σ are the variance in category h and the variance in the whole region.

S4. Identification of ecological corridors, pinch points, and barrier points

S4.1 Identification of ecological corridors

Ecological corridors, as a channel for species flow and energy transfer between source sites, should be the flow path with the least resistance. In this study, the Linkage Mapper tool was used to extract ecological corridors and potential ecological corridors respectively.

S4.2 Identification of pinch points

Ecological pinch points refer to the areas with the highest electric current density, i.e., the areas with the highest density of biological migration between the source sites, and they are the key nodes in ecological protection. The ecological nodes in the corridor were identified using the "All to one" model in the "Pinchpoint Mapper". The current density based on the connected corridors can be obtained by linking a certain source ground to the ground, injecting current into the remaining source ground, and interacting among all the source grounds. In this study, the electric current density is divided into six levels and the highest sixth level is taken as the most ecological pinch points area.

S4.3 Identification of barrier points

Barrier points refer to areas in the corridor that impede the flow of species and information; their reduction can improve connectivity between sources. Using the Barrier Mapper module with 1,000 m as the Minimum Detection Radius, 10,000 m as the Maximum Detection Radius, and 1,000 m as the Radius Step

Value, the "Maximum" mode and "Calculate percent improvement scores relative to corridor LCD" were selected to obtain the "Barrier Mapper". The resulting values were classified into six levels using the natural breaks method, with the highest sixth category identified as barrier points.

S4. Ecological network topology calculation

The Comprehensive Importance (CI) of each ecological patch was calculated using the following formulas:

$$CI_i = \alpha DC_i + \beta CC_i + \gamma BC_i, \tag{20}$$

$$DC_i = \frac{k_i}{n-1}, \tag{21}$$

$$CC_i = \frac{n-1}{\sum_{j=1}^{n} d_{ij}}, \tag{22}$$

$$BC_i = \sum_{i \geq j} \frac{d_{ij}^n}{d_{ij}}, \tag{23}$$

where, CI_i is the comprehensive importance of node; DC_i is the degree centrality; CC_i is the closeness centrality; BC_i is the betweenness centrality; α, β, and γ denote the weights; n represents the total number of nodes; k_i refers to the degree of node I; d_{ij} is the distance between nodes i and j; and d_{ij}^n represents the number of shortest paths between node i and node j passing through node n.

To evaluate changes in stability before and after adding PEC, we analyzed network robustness using the following **Eq. (24)**[12]:

$$R = \frac{C}{N - N_r}, \tag{24}$$

where, R represents network stability, C denotes the number of nodes in the largest connected subgraph after removing certain ecological nodes, indicates the total number of nodes in the network, and N_r is the number of nodes removed.

S5. Configuration of genetic algorithm parameters

Genetic algorithms (GAs) are global optimization techniques that mimic natural selection and genetic mechanisms, enabling efficient exploration of complex solution spaces[13]. In this study, the parameters of the genetic algorithm were configured as follows: real-number encoding was used, the population size was set to 200, the maximum number of iterations was 200,000, and the mutation probability was 0.5. These parameter settings were chosen to maintain algorithmic diversity in the search space while enhancing its global optimization performance. During the optimization process, a comprehensive objective function was developed to balance average risk, total cost, and variation coefficient of width incorporated adaptive weighting and high-risk area penalty mechanisms to ensure that the optimization outcomes align more closely with practical requirements. To enhance the computational efficiency of the algorithm, this study incorporated an early stopping mechanism that halts the optimization process if the optimal fitness value remains constant for 100 consecutive generations. This approach effectively prevents the algorithm from becoming trapped in local optima or engaging in unproductive computations within the solution space, thereby significantly improving computational efficiency. Leveraging the global optimization capabilities of genetic algorithms, this study successfully optimized the design of corridor width and elucidated the relationship between width and regional risk, providing a robust scientific foundation and technical support for ecological landscape planning and design.

To balance the optimization objectives for corridors of varying lengths, we categorized each corridor based on its length (L) into four distinct length categories (LC): short, medium, long, and extra-long. Different adaptability weights (w_a) were assigned to each category to ensure balanced optimization objectives across corridors of different lengths. The classification criteria are as follows:

$$LC = \begin{cases} \text{Short, } L \leq 10\text{ km} \\ \text{Medium, } 10\text{ km} < L < 50\text{ km} \\ \text{Long, } 50\text{ km} < L < 100\text{ km} \\ \text{Extra long, } L \geq 100\text{ km} \end{cases}, \tag{25}$$

$$w_a = \begin{cases} 0.7, \text{ Short} \\ 1.0, \text{ Medium} \\ 0.5, \text{ Long} \\ 0.3, \text{ Extra long} \end{cases}, \tag{26}$$

This study formulates an objective function that incorporates average RV, Total Cost (TC), and Variation Coefficient of Width (CW) to balance these considerations, and optimizes the corridor width to minimize the value of this objective function. The related formulas are as follows:

$$AR = \frac{\sum_{i=1}^{N} L_i \cdot RA_i}{\sum_{i=1}^{N} L_i}, \tag{27}$$

$$RA_i = \frac{AR_i \cdot RP_i}{W_i + 1} \cdot w_a, \tag{28}$$

$$TC = \sum_{i=1}^{N} L_i \cdot W_i, \tag{29}$$

$$CW = \frac{\sigma(W)}{\mu(W)}, \tag{30}$$

where, L_i represents the length of corridor i, RA_i denotes the adjusted risk value of corridor i, and RP_i signifies the risk penalty multiplier for the ith corridor, which is determined based on the quantile of the risk value. Specifically, if the average risk value of a corridor exceeds the 75th percentile of the overall average risk values, the penalty multiplier is set to 2; otherwise, it remains at 1. $\sigma(W)$ is the standard deviation of the corridor width; $\mu(W)$ is the mean of the corridor width.

To ensure the consistency of the order of magnitude across all components, **Eqs (27)**, **(29)**, and **(30)** were processed individually, as presented in **Eqs (31)**, **(32)**, and **(33)**.

$$AR_{\text{norm}} = AR \times 10^4, \tag{31}$$

$$TC_{\text{norm}} = TC \times 10^{-10}, \tag{32}$$

$$CW_{\text{norm}} = CW, \tag{33}$$

References

[1] Terrado M, Sabater S, Chaplin-Kramer B, Mandle L, Ziv G, Acuna V. 2016. Model development for the

assessment of terrestrial and aquatic habitat quality in conservation planning. *Science of the Total Environment* 540:63-70 https://doi.org/10.1016/j.scitotenv.2015.03.064

[2] Hall LS, Krausman PR, Morrison ML. 1997. The habitat concept and a plea for standard terminology. *Wildlife society bulletin* 25(1):173-182

[3] Budyko MI, Miller DH. 1974. Climate and life. *Academic press*

[4] Zhang L, Dawes W, Walker G. 2001. Response of mean annual evapotranspiration to vegetation changes at catchment scale. *Water resources research* 37:701-708 https://doi.org/10.1029/2000WR900325

[5] Williams J, Renard K, Dyke P. 1983. EPIC: A new method for assessing erosion's effect on soil productivity. *Journal of Soil and water Conservation* 38:381-383

[6] Zhao MS, Running SW. 2010. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science* 329(5994):940-943 https://doi.org/10.1126/science.1192666

[7] Running S, Zhao MS. 2015. User's Guide. Daily GPP and annual NPP (MOD17A2/A3) products NASA earth observing system MODIS land algorithm. Version 3.0 for collection 6. https://lpdaac.usgs.gov/sites/default/files/public/product_docu-mentation/mod17_user_guide.pdf

[8] Shannon CE. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27(3):379-423

[9] Shemshadi A, Shirazi H, Toreihi M, Tarokh MJ. 2011. A fuzzy VIKOR method for supplier selection based on entropy measure for objective weighting. *Expert Systems with Applications* 38(10):12160-12167 https://doi.org/10.1016/j.eswa.2011.03.027

[10] Wang JF, Xu CD. 2017. Geodetector: Principle and prospective. *Acta Geographica Sinica* 72(1):116-134 (in Chinese)

[11] Song YZ, Wang JF, Ge Y, Xu CD. 2020. An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: cases with different types of spatial data. *Giscience & Remote Sensing* 57(5):593-610 https://doi.org/10.1080/15481603.2020.1760434

[12] Lu Y, Liu Y, Xing L, Liu Y. 2023. Robustness test of multiple protection strategies for ecological networks from the perspective of complex networks: Evidence from Wuhan Metropolitan Area, China. *Land Degradation & Development* 34(1):52-71

[13] Xu AK, Hu MJ, Shi J, Bai QZ, Li XH. 2024. Construction and optimization of ecological network in inland river basin based on circuit theory, complex network and ecological sensitivity: A case study of Gansu section of Heihe River Basin. *Ecological Modelling* 488:110578 https://doi.org/10.1016/j.ecolmodel.2023.110578