https://doi.org/10.15302/J-LAF-1-020073

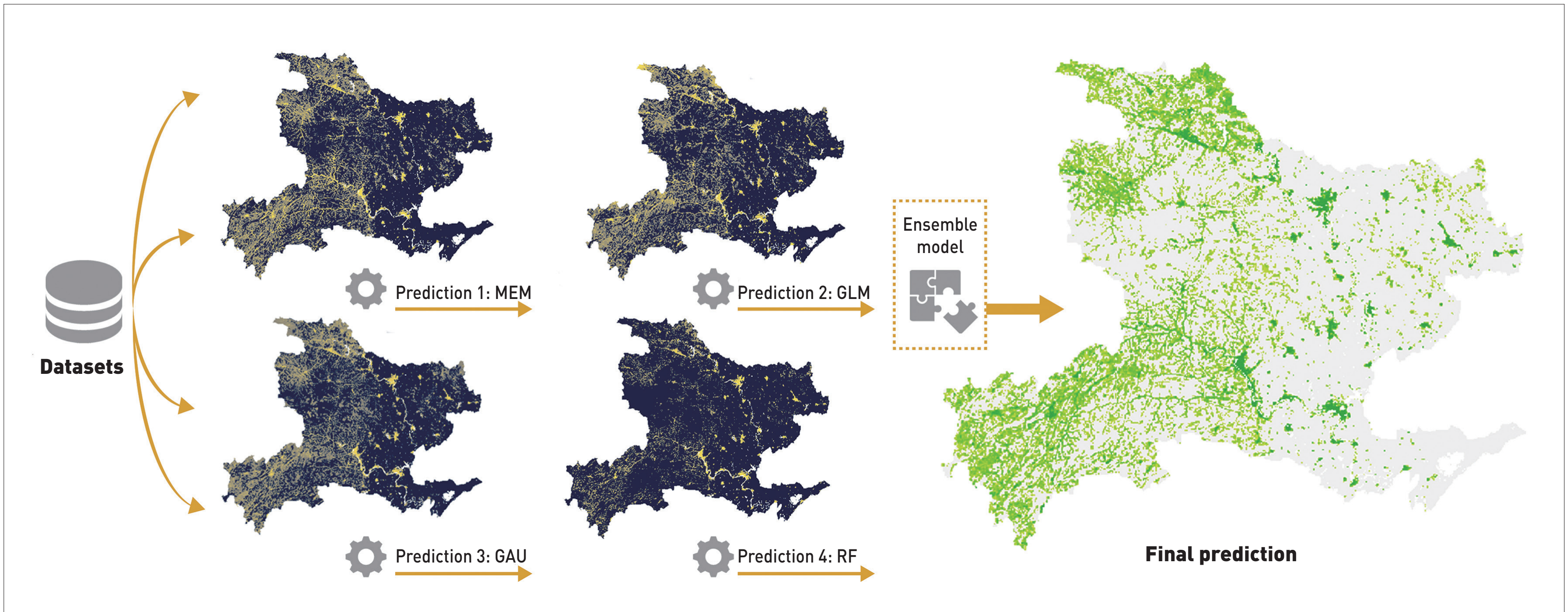
Spatial Potential of Recreational Services in Western Hubei Region in Light of the "All-for-One Tourism" Development —A Machine Learning Approach Based on Ensemble Model

WEN Chen[1,2], CHA Jing[1,2], XU Liquan[1,2,3,*], XU Haiyun[4]

1 School of Architecture and Urban Planning, Huazhong University of Science and Technology, Wuhan 430074, China
2 Key Laboratory of City Simulation, Ministry of Natural Resources, Wuhan 430074, China
3 Built Heritage Research Center, Huazhong University of Science and Technology, Wuhan 430074, China
4 School of Architecture and Urban Planning, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

*CORRESPONDING AUTHOR
Address: No. 1037, Luoyu Road, Hongshan District, Wuhan 430074, Hubei Province, China
Email: xlq2019hust@hust.edu.cn

GRAPHICAL ABSTRACT



HIGHLIGHTS

- Addresses the key technical difficulty of spatially locating of continuous areas in all-for-one tourism development
- Builds an SDM model of ensemble machine learning with multi-sourced social-ecological data
- Quantifies the recreation potential of continuous areas at the regional scale by taking the western Hubei region as the case study

KEYWORDS

All-for-One Tourism;
Recreational Services;
Cultural Ecosystem Services;
Spatial Potential Prediction;
Machine Learning;
Western Hubei

RECEIVED DATE

2022-12-13

ABSTRACT

Under the call for "all-for-one tourism" development, the focus of regional-scale recreational services is shifting from the construction of individual scenic spots and tourist areas towards the comprehensive planning of tourist destinations, so as to propel China's rural revitalization and regional coordinated development. In research and practice, however, it is still challenging to identity and evaluate spatial locations for developing tourism according their cultural and environmental resource and characteristics and prioritizing the high-potential ones. Employing the whole western Hubei region as a case study, this paper proposes a method of assessing recreation potential within the research framework on cultural ecosystem services, and uses multi-sourced social-ecological data to develop an SDM model via ensemble machine learning. Through analyses of the environmental features of 336 recreational hotspots in the study area, the model predicts the areas with high recreation potential in continuous areas. This study intends to establish a technique path to examine the regional-scale pattern of recreational spaces via numerical analysis of environmental features, and to provide a reference for relevant spatial development strategies of all-for-one tourism and rural revitalization.

EDITED BY Tina TIAN

1 Background

In recent years, both the development of China's ecological civilization and its spatial management system have stepped into a new stage. Local governments are making efforts to explore regional coordinated development and urban-rural integration under the "all-for-one tourism" strategy—considering the entire region a holistic tourist area requires promoting the integration of tourism and all industries, leveraging existing tourism resources, using famous scenic spots, rural tourism hotspots, and other tourism elements to drive the development of regional tourism as a whole, and finally enhancing regional recreation potential[1]. The strategy for developing regional tourism will move from managing individual scenic and tourist areas to creating a series of tourist destinations that are connected and integrated with each other[2].

The development of all-for-one tourism is not uniform spatially, and the allocation of resources to important development areas requires informed decision-making. In practice, it means to identify the areas with high potential of recreational services based on an overall assessment of local natural resources, cultural and humanity resources, infrastructure, etc., and to prioritize their development by evaluating related efficiency factors. Hence, regional-scale analyses of ecological-cultural tourism resources have become an emerging research topic[3]~[5]. Some research has offered supportive spatial strategies, such as development modes based on the evaluation of tourism competitiveness via analyses on touristic resources, population and economic impulse, and infrastructure distribution[6]~[9]. However, existing studies have quantified and mapped local ecological-cultural tourism resources less from a spatial perspective, particularly identifying and prioritizing the areas with high recreation potential within a certain region based on its indigenous ecological-cultural tourism resources. The absence would limit the understanding and development of urban-rural integration strategies.

This study aims to create a machine-learning-based framework for spatial analysis under cultural ecosystem services (CES) theories by adopting the methods for the spatial assessment of CES recreation potential and using multi-sourced data to predict the recreation potential of the given region. Adopting the concepts in CES research, this research defines the "recreation potential" as the possibility of natural environments providing recreational activities or experiences[10][11]. This paper aims to achieve two associated research objectives: 1) forecasting the recreation potential of continuous areas in a region using locally well-known recreational hotspots; and 2) exploring the region's spatial pattern of recreational services based on numerical analysis of environmental features.

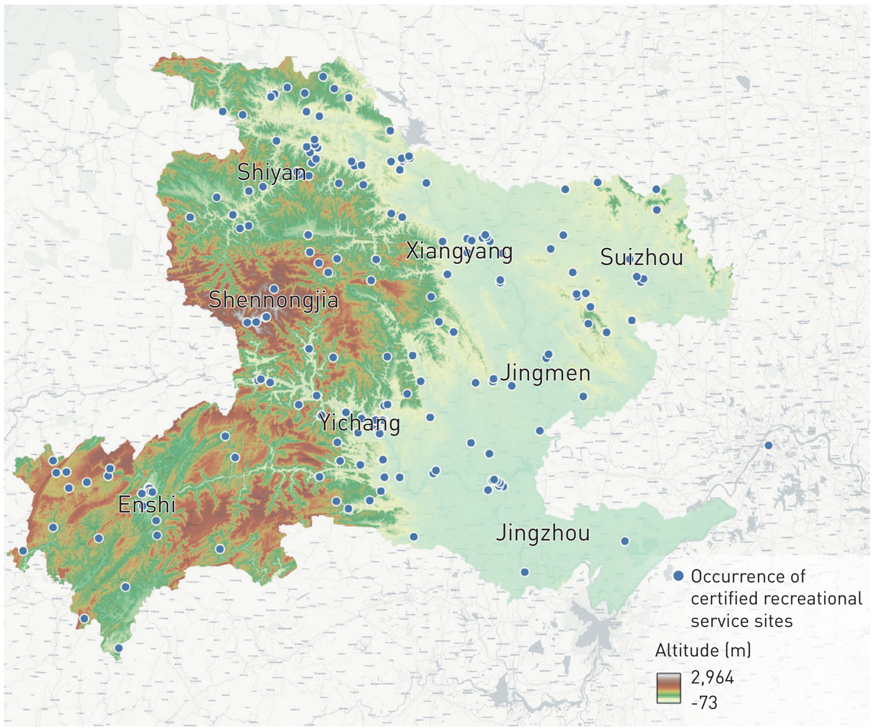
2 Literature Review

CES refers to the non-material well-being gained by humans through interaction with nature, including recreational, aesthetic, spiritual, religious, instructional, therapeutic, and artistic experience[12]. Recreational services are subjective and difficult to quantify since they depend not only on the physical landscape, including the type, quantity, and quality of the environmental setting and the associated infrastructure[12], but also on the public demand and social culture. The research methodology adopted from CES theories can be used for the quantitative evaluation of recreation potential, examining the connections between human needs and landscape environmental factors and better getting integrated with planning frameworks[13][14].

Existing approaches used in CES research on the spatial assessment of recreation potential include spatial overlay (multi-criteria decision analysis), network analysis method, spatial regression method, modeling with multiple indicators, and spatial visualization with the social media-based method[15]~[17]. Several studies, for instance, utilized AHP (analytical network process) to identify environmental factors that impact recreation potential, and then computed recreation potential using GIS overlay methods[18][19]. Other studies examined the recreation potential by looking at the spatial centralities between different villages and what kinds of tourism services they have. Some new technologies, such as overlay analysis integrated with citizen science, PPGIS, and methods for visualizing big data from social medias, have also turned out to be effective in relevant studies[16][20]~[25] (Table 1).

Table 1: Recent studies on spatially assessing the regional recreation potential under CES theories

Source	Study area	Criteria of recreational services	Computing method	Notes
Ref. [22]	Wusheng in Sichuan Province, China	Natural resources, cultural landmarks, landscape aesthetics, accessibility, and supporting industry development	Quantifying each criterion of recreational potential through multi-sourced spatial data, and then integrating them through AHP method and multi-criteria decision analysis based on expert scoring	Assessment of recreation potential in continuous areas, considering both natural and cultural factors
Ref. [23]	Three Parallel Rivers of Yunnan Protected Areas, covering 16 counties and cities in the northwest of Yunnan Province, China	Landscape diversity, naturalness, river and lake elements, the rated level of scenic spots	After quantifying and normalizing each indicator into a five-levels rating system, different types of indicators are then combined with an equal weight	Exploring the potential of natural recreation, and the equal weighting method assuming that the impact of each criterion is equal
Ref. [24]	Taizhou in Zhejiang Province, China	Land use, slope degree, distance to urban areas, distance to water bodies, historical and cultural resources	Quantifying each criterion of recreation potential through spatial data analysis, and normalizing them into several levels with standard scores accordingly; the final result is obtained by adding them up	Factors of recreation potential were scored directly by an expert group, but the scoring process was not described in detail
Ref. [16]	Southeastern region in Lithuania	Natural categories (water bodies, naturalness, protected area status, topographic diversity, etc.); cultural categories (tourist paths, bike lanes, world cultural heritage, viewpoints, etc.)	The indicators of each factor were standardized by z-value for overlay analysis; the results were validated by a questionnaire survey	The different factors were still treated as equal weights, but public surveys were introduced to demonstrate the validity of the research
Ref. [20]	Jeollanam-do, southwest the Republic of Korea	Assessment of rural tourism potential from the perspective of transportation network	Accessibility was assessed by using multiple centrality indicators of graph theory for evaluating rural recreation potential	The study focused on the villages' development perspective, especially on accessibility; the recreation features of the villages were not discussed
Ref. [25]	Basque Country, northern Spain	Naturalness, status of nature reserves, water bodies, topography, geographical points of interest	This study introduced the concept of "viewshed" as a statistical unit and calculated the factors of recreation potential within each unit	A photo-comparison based questionnaire was introduced as a validation



1 © Wen Chen, Cha Jing, Xu Liquan, Xu Haiyun

Recent research has utilized species distribution modeling (SDM) to forecast the spatial occurrences of socio-cultural activities/behaviors at a regional scale and reported valuable findings. SDM, developed from ecological research, primarily builds a relationship between biological observation points and environmental features. It is capable of fitting models using machine learning techniques and producing predicted results for continuous areas[26][27]. Among socio-cultural topics, SDM has been used to predict geographic distribution of human emotion with physical environmental factors[28] and public preferences of national parks on natural environmental characteristics[29]. Although geographical environmental factors are not the sole ones that directly impact socio-cultural activities, SDM can still reflect humanity factors via environmental variables as proxy indicators[28]. Moreover, since humanity factors are usually difficult to spatially quantify and visualize, their close relationship with environmental features can be used as proxies to study social topics[30].

However, existing regional-scale research sees the following limitations. First, overlay method and multi-criteria analysis method are subjective when defining the weights of different influencing factors, and they often lack replication and validation tests; besides, the expert scoring method often introduces ambiguity. Second, local well-known recreational sites are less covered in current modeling, which leads to the ignorance of regional uniqueness. Lastly, places with sparse data representation, such as wilderness and remote villages, have been neglected in existing research, even in SDM modeling.

3 Research Data and Methods

3.1 Study Area

The research selects the whole western Hubei region, China as its study area (Fig. 1), covering eight municipal units, namely Xiangyang, Yichang, Shiyan, Jingzhou, Jingmen, Suizhou, Enshi, and Shennongjia. Most of the 27.3 million people[31] in the region live in mountainous areas. The western Hubei region is not only culturally diverse and rich (Fig. 2), but also has a pressing mission to shake off poverty by propelling its tourism development. In the past few years, the Hubei government has attached importance to the promotion of all-for-one tourism in the study area. It thus is critical and urgent to investigate the spatial pattern of the region's recreation potential in order to inform the priority of its tourism development.

2 © Rural Construction Center of Hubei Province (RCC-HB)

1. Locations of the verified recreational sites in the study area
2. Examples of the studied sample points of recreational sites

3.2 Building Machine Learning Datasets

3.2.1 Occurrence Dataset

This study employed a machine learning modeling, using certified recreational sites for occurrence observations and multi-sourced social-ecological data as environmental features.

The occurrence dataset collects the records of objective events. In this study, it refers to the places where recreational services have already been verified, recognized, or acknowledged in the region. Based on literature review and preliminary research findings, the study collected data of three types of recreational hotspots to construct the occurrence dataset: 1) national A-level tourist destinations, scenic spots, and the sites on the officially recommended Hubei Tourism Routes; 2) national-level ancient villages; and 3) popular travel destinations ranked by Weibo, a mainstream social media in China. These recreational hotspots are all recognized by the government or the market.

The steps of spatial data processing were as follows. The research team retrieved the spatial coordinates of the aforementioned recreation occurrences via Amap API geocoding, and transformed them into UTM projection coordinates. Through the preprocessing, it was found that the Weibo check-in data were highly related with the population distribution in the region, leading to problems of duplicated information and absence of some occurrences. Thus, check-in data within the major urban areas of all the municipal units were omitted, and only those in towns and villages were kept. The dataset contained 336 sample points in total.

3.2.2 Dataset of Environmental Features

The research team reviewed existing CES literature on modeling recreational potential (Table 2) before building the dataset of environmental features[19][30][32]~[37]. We gathered and selected 16 indicators, ranging from land cover, terrain, landscape composition, climate to transportation. Following the "spatially explicit" modeling paradigm[38], the study converted each feature into raster data and ensured the entire region covered, to achieve consistency and reproducibility in the modeling process.

In this study, data of both land use and landscape indices require specialized processing. Because the machine learning algorithms employed in the integrated model require numerical rather than categorical feature data, the study transformed the land use classification of a given area into the specific proportion of four land cover types, including forest, built-up, agricultural, and water. This processing method can also well depict the land cover types at each pixel that are often mixed. The study used the moving window method to construct the global landscape indices within the neighborhood of each pixel to provide spatial results, including JOINENT (joint entropy at the landscape level) and other indicators. Due to the huge amount of computing load that went into this analysis, the study selected indicators of SHDI (Shannon's diversity index) and JOINENT to measure the diversity and complexity of the features. In this study, the spatial resolution of the environmental feature dataset was unified to 1,000-meter[39]. With bilinear resampling, all variables were processed into the same resolution. After constructing the environmental feature dataset, all environmental feature values of the 336 sample points in the occurrence dataset will be extracted.

3.3 Modeling SDM with Integrated Machine Learning Frameworks

3.3.1 Preprocessing

The SDM model is used to forecast the probability of events occurring in a spatial area. Regional-scale research often shows records of a few locations, lacking precise information about the occurrence of recreational sites (dependent variables) in certain continuous space. SDM model can learn the environmental factors (independent variables) of known recreational hotspots and then apply the predicted probability to the whole study area. Also, SDM makes it possible to learn the relationship between occurrence data and environmental features data by using different algorithms and parameters, supporting a flexible modeling process, the prediction results often fluctuate though. In this research, the data processing was performed via the programming languages R and Python, and the results were visualized via QGIS.

The predictive performance of the SDM model is sensitive to the spatial extent to be predicted and the co-linearity between the environmental feature variables. During the preprocessing, the whole western Hubei region was determined as the spatial extent, and the environmental filtering method was used to avoid the impact of close data points on modeling results. Following the input of required data, the SDM model randomly generated a pseudo-absences dataset to solve the potential co-linearity problem of the environmental feature dataset. Before using machine learning modeling, the study conducted principal component analysis for dimensionality reduction in environmental feature variables to reduce co-linearity.

The study employed the approach of block partition[40] to split the data of the 336 sample points into training set and testing set. Specifically, it divided the dataset into four blocks (sometimes known as "folds"), with three blocks serving as the training set and

Table 2: Dataset of environmental features in this study

Environmental feature	Code	Description	Data source	Reference(s)
Land cover ratio–forest land	Fractions_Forest	Calculating the percentage of forest land in the analysis unit	EU Copernicus Global Land Cover Data (Global CGLS-LC100)	Ref. [33]
Land cover ratio–built up area	Fractions_Built	Calculating the percentage of built-up area in the analysis unit	EU Copernicus Global Land Cover Data (Global CGLS-LC100)	Ref. [34]
Land cover ratio–agriculture	Fractions_ Cropland	Calculating the percentage of farmland in the analysis unit	EU Copernicus Global Land Cover Data (Global CGLS-LC100)	Ref. [32]
Land cover ratio–water	Fractions_Water	Calculating the percentage of water in the analysis unit	EU Copernicus Global Land Cover Data (Global CGLS-LC100)	Ref. [35]
Mean temperature of the warmest month	LST	Measuring the surface temperatures in different sections of the site, for the purpose of examining important tourist capacity such as summer cooling	Based on Landsat 8 satellite image, calculated for the time period from July 1, 2021 to July 1, 2022	Ref. [36]
Annual mean NDVI	NDVI	Quantifying the greenness and ecological conditions	Based on Landsat 8 satellite image, calculated for the year 2021	Ref. [33]
Annual mean EVI	EVI	Responding to plant biophysical parameters; this indicator corrects atmospheric soil noise and is more sensitive to dense forest areas	Based on Landsat 8 satellite image, calculated for the year 2021	Ref. [37]
Shannon diversity index	SHDI	Landscape index for diversity of land covers; in this study, a moving window algorithm is used with a neighborhood radius of 1 km	Based on the land cover data GlobeLand30	Refs. [33][35]
Landscape joint entropy	JOINENT	Landscape index for complexity; in this study, a moving window algorithm is used with a neighborhood radius of 1 km	Based on the land cover data GlobeLand30	Ref. [33]
Patch richness	PR	Landscape index for the patches; in this study, a moving window algorithm is used with a neighborhood radius of 1 km	Based on the land cover data GlobeLand30	Ref. [33]
Elevation	ELEV	Measurement of terrain height	ALOS-2 elevation data	Refs. [30][35]
Slope	SLOPE	Measurement of terrain steepness	Based on the elevation data	Refs. [30][35]
Terrain ruggedness index	TRI	Measurement of terrain variability	Based on the elevation data	Ref. [30]
Terrain position index	TPI	Measuring the topographic relief of data points relative to their immediate neighbors	Based on the elevation data	Ref. [30]
Distance to the nearest town	DIST_TOWN	Measurement of transportation convenience	Calculated with national town-level settlement location data	Ref. [32]
Distance to the nearest waterbody	DIST_WATER	Measurement of proximity to waterbody	Based on the land cover data GlobeLand30	Refs. [19][35]

Table 3: Three sets of environmental features for modeling

No.	Description	Environmental variables
Model 1	Basic model: site composition and topography	Fractions_Forest, Fractions_Built, Fractions_Cropland, Fractions_Water, ELEV, SLOPE, TRI, TPI
Model 2	Advanced model: site composition, topography, and proximity to water and towns	Fractions_Forest, Fractions_Built, Fractions_Cropland, Fractions_Water, ELEV, SLOPE, TRI, TPI
Model 3	Complete model: site composition, topography, proximity to water and towns, landscape index, and spectral index	Fractions_Forest, Fractions_Built, Fractions_Cropland, Fractions_Water, ELEV, SLOPE, TRI, TPI, LST, NDVI, EVI, SHDI, JOINENT, PR

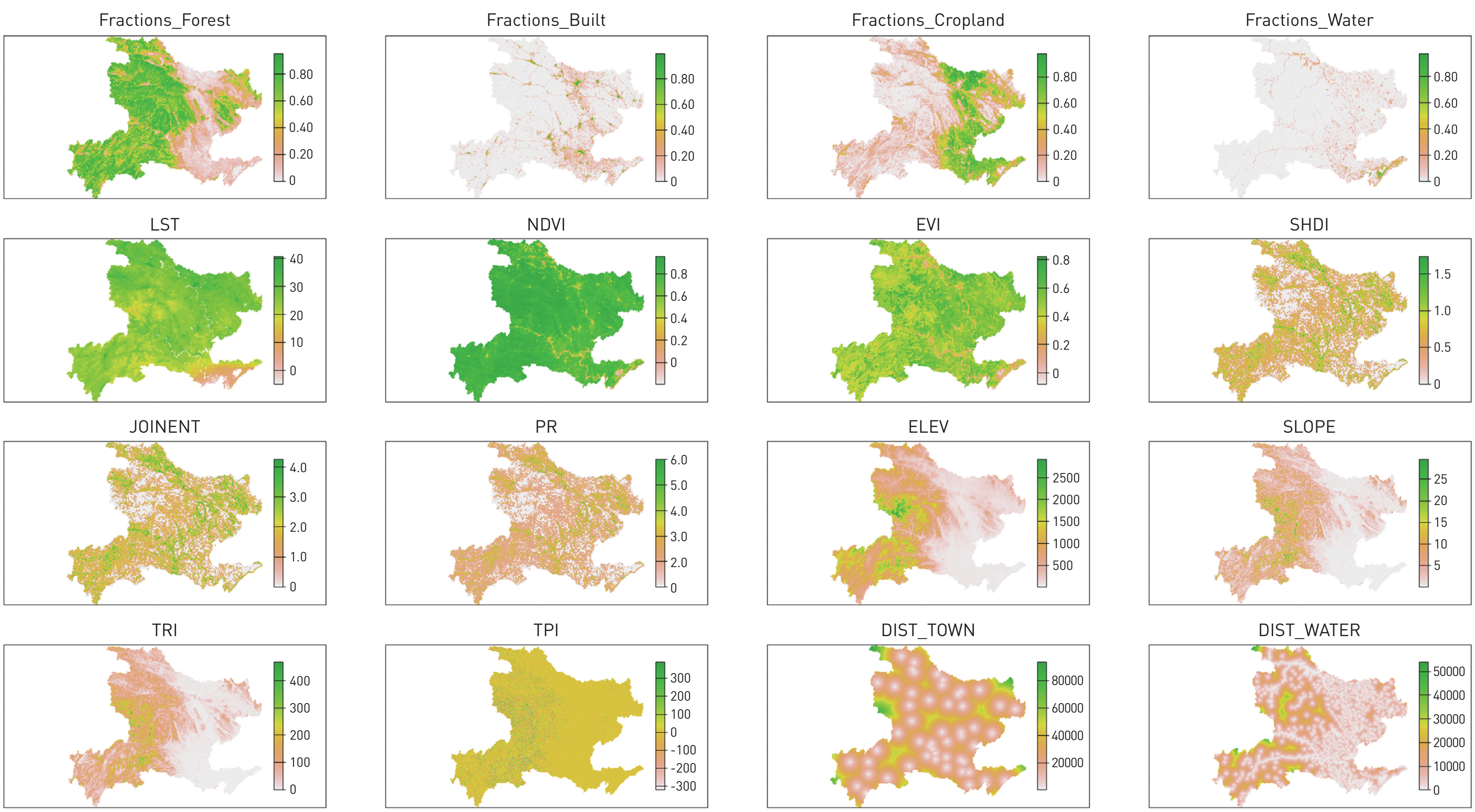
the fourth serving as the testing set. Then, each block was used alternatively and iteratively as the testing set, and the remaining blocks as the training set. This method was used to evaluate the performance of the model and reduce the bias produced by the uneven partitioning of the dataset.

This study made three sets of models with different variables (Table 3) so that the prediction performance of different environmental variables could be measured. Each set of models followed the same analytic procedure, and the significance of each environmental feature was determined by comparing the prediction performance of the three.

3.3.2 Data Fitting and Verification With the Ensemble Model

The data were initially modeled by using four types of machine learning algorithms: Maximum Entropy Models, Gaussian Process Models, Generalized Linear Models, and Random Forest Models. These four algorithms are commonly used in SDM modeling research, with different assumptions, data processing

3. The spatial modeling results of environmental features



methodologies, and strengths/weaknesses.

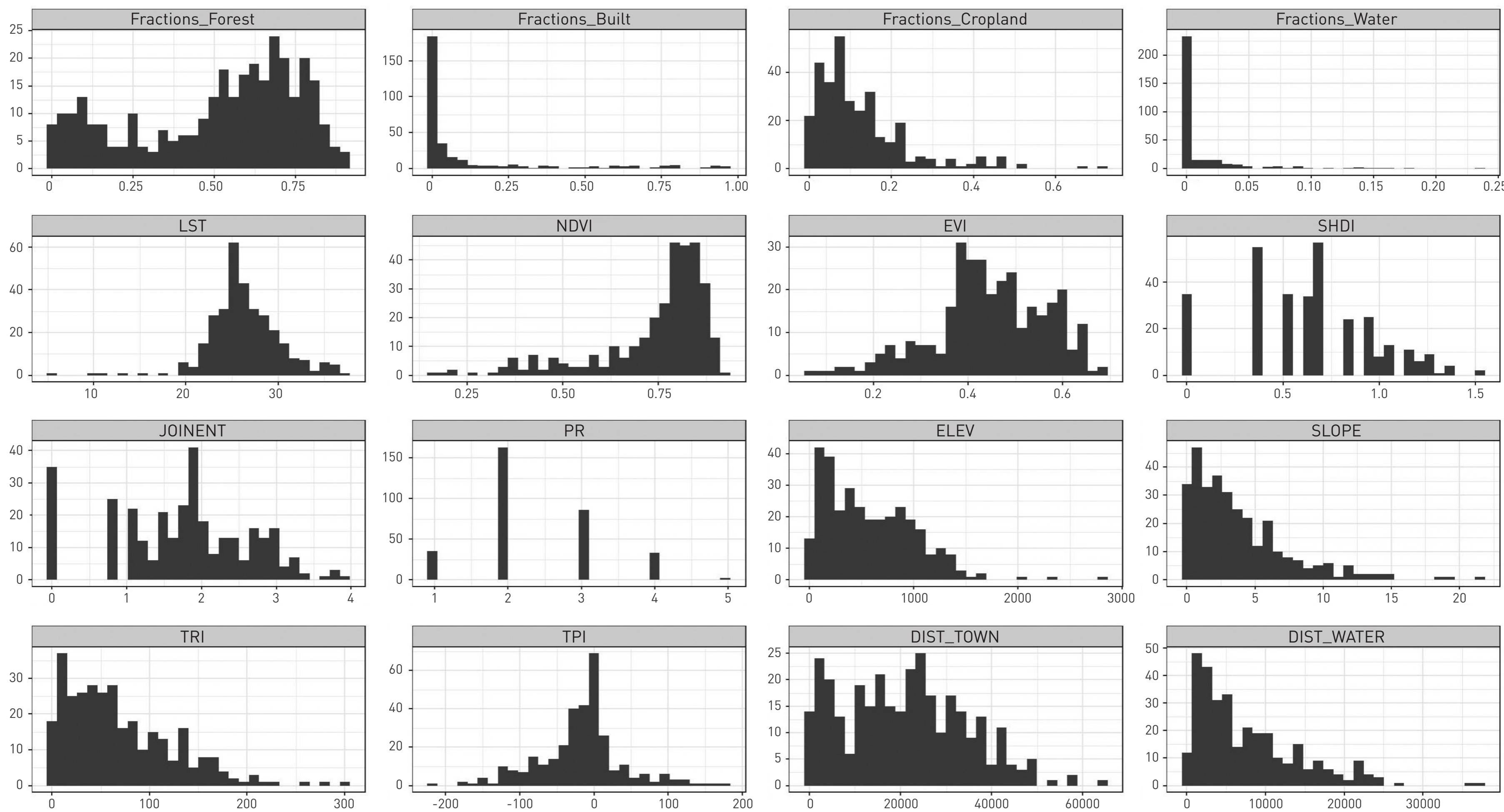
This study employed the ensemble machine learning approach, which integrates multiple machine learning algorithms for an improved performance, and also to avoid arbitrarily selecting a single algorithm or weighting parameter. Using performance evaluation indicators for each model, a weighted average of the prediction outcome of several separate algorithms can be calculated. For model validation, TSS (True Skill Statistics), AUC (Area under the ROC Curve), and JACCARD (Jaccard similarity coefficient), were used as evaluation indicators. Although MSE (mean square error) and MAE (mean absolute error) are frequently used for evaluating the effectiveness of prediction models, they are less employed in SDM modeling. This is because MSE and MAE do not consider true negatives, which refers to the absence of occurrences in this study. In SDM, predicting the absence of an event is equally as significant as predicting its occurrence. Therefore, indicators such as AUC and TSS are more suitable for evaluating the performance of SDM model. The tuned model was used to estimate the probability of the spatial occurrences of recreation potential throughout the study area. The prediction results were then visualized as recreational potential maps.

4 Research Results

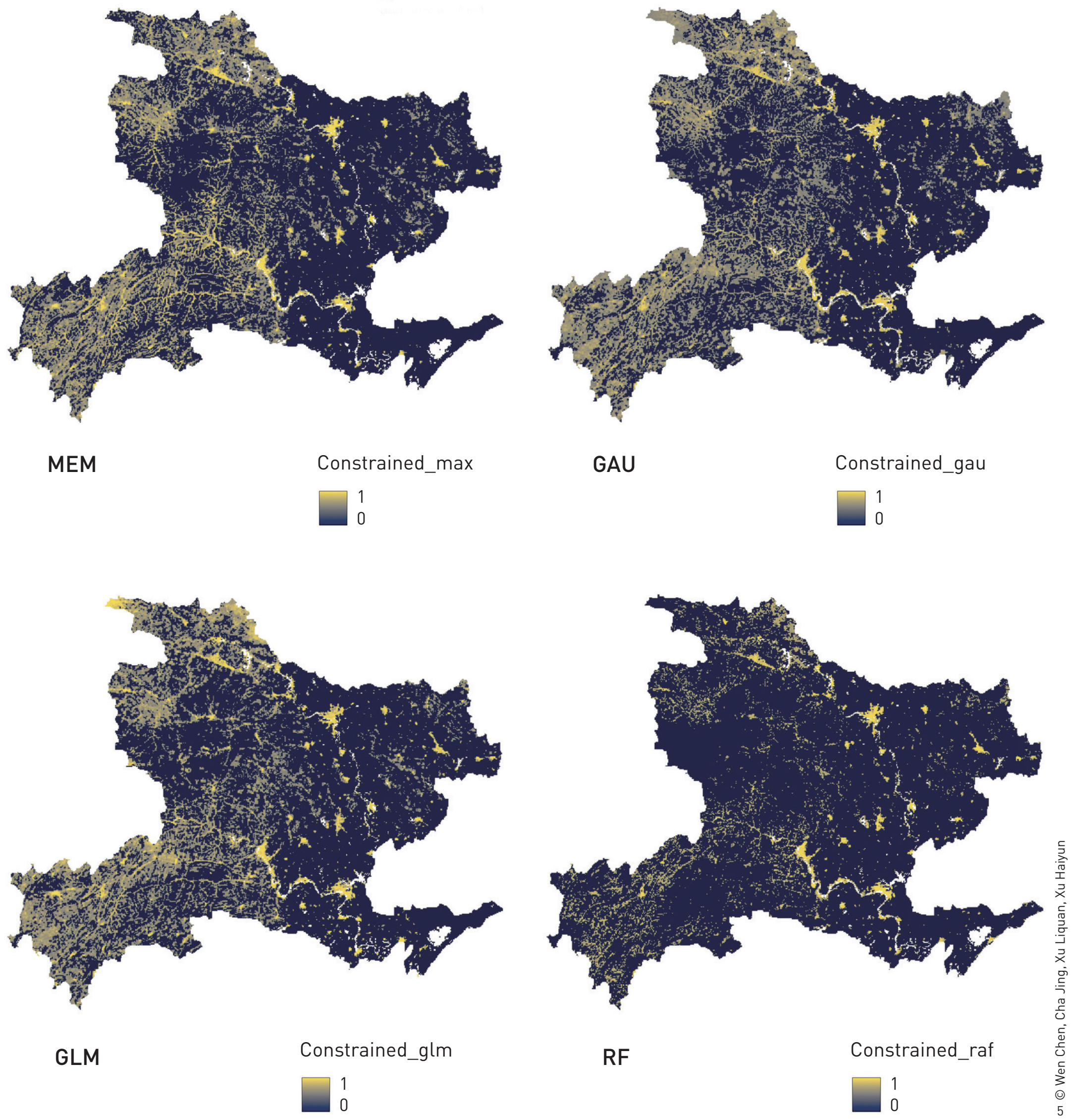
4.1 Analysis Results of the Occurrence Data and Environmental Feature Data

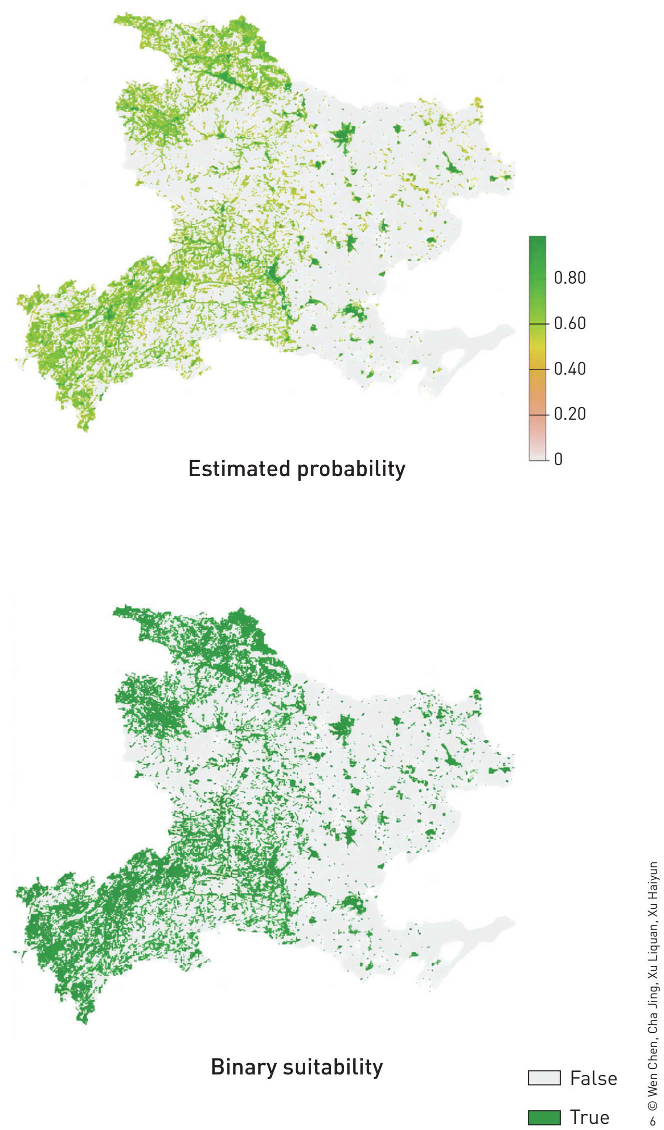
The 336 sample points in the occurrence dataset are mostly located in Shiyan, Enshi, Xiangyang, and Yichang. Analysis showed that the environmental features of sample points are not distributed uniformly and demonstrate obvious clustering effects (Figs. 3, 4). The average elevation of the sample points is 591 m, and the average slope is 3.85 degrees. For the kilometer grid to which these occurrences belong, forests account for more than 50 percent of the total area, while built-up areas and agricultural land account for 7.8% and 17.8%, respectively. The closest town is 22 kilometers away, while the closest lake and river are in a

4. Numerical distributions of environmental features



4 © Wen Chen, Cha Jing, Xu Liquan, Xu Haiyun





5. Predicted recreation potentials via different algorithms

6. Estimated probability via the ensemble model and the binary suitability

distance of 7.6 kilometers. The high annual average NDVI and EVI indices for the entire study area suggest that the ecological status of the western Hubei region is favorable; Suizhou, Jingmen, and Jingzhou areas have lower elevations and relatively flat terrains, with a greater proportion of farmland and connected river network. Through analysis of landscape indices, the study found several belt-shaped areas with a higher degree of landscape diversity and complexity.

Because the dataset of environmental features consists of variables under multiple categories and different ranges of values, principal component analysis was used to avoid possible collinearity and parameter estimation distortions by reducing the dimensionality of data. This study selected 10 principal components from the raw data of environmental features, covering 95% of the cumulative explained variation (Table 3), as independent variables in the modeling. Model 3, which incorporates all environmental variables, had the best performance and was used for spatial visualization. Figure 5 displays the estimated results of four individual algorithms. Figure 6 displays estimated probability of recreation potential map and the binary suitability map of the study area.

4.2 Prediction Results and Performance of the Models

Among the predication results of different algorithms, RF predicted fewer high-potential areas, and most of them are located in two chains from northwest to southeast: one goes through

Jingzhou, Zhijiang, Yidu, Yichang, Zigui, and Badong; and the other links up Xiangyang, Gucheng, Laohekou, Danjiangkou, and Shiyan, with scattered points in their surroundings. MEM, GAU, and GLM all agree that the bank areas along the Yangtze River and a large number of mountainous areas are of high potential, in addition to the two chains predicted above. Most of the differences among these three algorithms come from the thresholds for identifying high-potential locations and the corresponding sizes.

Giving the same dataset of environmental features, different algorithms produced somewhat varied prediction outcomes, reflecting the uncertainty of the models. Considering AUC and JACCARD, the index values of each by the three algorithms were roughly above 0.7 and 0.5 with slight fluctuations, showing that the models perform well. Nonetheless, their performance differs considerably regarding the TSS evaluation. The value by each algorithm varies between 0.27 and 0.42, showing the randomness exists among the models.

The ensemble model combined multiple individual algorithms by setting a threshold, weighting the average prediction results and therefore enhancing the model's overall performance in precision (Table 4). The ensemble model forecasted that the areas with the highest recreation potential are scattered across the study area; on the whole, the west side of the region having a relatively greater recreation potential, with several belt-shaped continuous areas (Fig. 6). Also, compared to the results by individual algorithms, the ensemble model's estimated probability results suggested a refined prediction outcome that further identified the high-potential locations at the area where the Yangtze River meets the mountains, partly covering the mountainous areas of Xiangyang and Yichang.

Table 4: PCA extraction of the environmental features

Principle component	Cumulative percentage of explained variance
1	0.325
2	0.504
3	0.604
4	0.679
5	0.744
6	0.797
7	0.848
8	0.895
9	0.927
10	0.956
11	0.974
12	0.987
13	0.992
14	0.997
15	0.999
16	1

5 Discussion

5.1 Reflection on Theory and Methodology

Recreational services are complex social-ecological phenomena. Predicting recreation potential in a spatially explicit manner has become an important means in regional planning for developing strategies. This study uses an SDM-based machine learning ensemble model to predict regional recreation potential under CES theories. The following summarizes the innovativeness of the study.

Theories of CES take into account both public preferences and objective environmental factors, and it has been challenging to assess and map recreation potential in continuous areas[19][30]. This study presents an alternative method to evaluate regional-scale recreation potential by using the environmental features of known recreational hotspots as modeling parameters for machine learning. This method not only provides a numerical understanding of the patterns of recreation potential, but it also overcome the uncertainty and subjectivity inherent in the commonly used expert scoring method, thus providing an appropriate option for regional-scale recreation potential assessment.

It should be mentioned that different research methods have their own advantages and disadvantages because they are appropriate for different case scenarios and would complement each other. The commonly used overlay analysis quantifies recreation potential by weighting up different criteria leveraging the varied expertise and experience of experts, and has the advantages of robust interpretability and flexibility and of representing the interests of various stakeholders. The uncertainty of this method can be effectively reduced by increasing the number of experts. Machine learning approaches, on the other hand, learn spatial patterns from multiple data sources using various algorithms and witness advantages of high reproducibility, efficiency, and scalability. As sample points for empirical research, this case chose national A-level tourist destinations, scenic spots, the sites on the officially recommended Hubei Tourism Routes, national-level ancient villages, and popular travel destinations ranked by Weibo. Such recreational sites can represent the popular tourist destinations in a region as public preferences, making it cover a broader service population that was previously difficult to address in spatial studies related to CES[17].

Applying SDM modeling at present to investigate the occurrence rules of social events is still in its early stages[28][21]. This study used several environmental features to examine the recreation potential on a regional scale, by using three prediction models that were gradually set with different categories of variables to probe into how well each environmental feature works on the predicting. In addition to land use and topography variables, the introducing of variables like water and town proximity, landscape index, spectral index, and climate improved the model's prediction performance in the study area (Table 5). This also validates the previous studies on the factors influencing recreation potential[41][42]. Before creating a dataset, it is vital to determine what is representative of the recreational hotspots in the given research site. For different research sites and goals, different datasets should be established.

Also, well-defined evaluation metrics can help measure the performance of how well the model predicts within the framework of machine learning[43]. The technical approach suggested in this study allows for continuous predictions at the regional scale without dividing specific statistical units (e.g., administrative districts, scenic spots, etc.) This method can also be used to study rural and wilderness areas that do not have clear boundaries, offering a suitable tool for the research on all-for-one tourism development.

5.2 Interpretation of Important Research Findings

This study predicts the recreation potential of the western Hubei region. Some of the spatial patterns observed in this study confirm previous research findings, illustrating the validity of the research. For example, one research project on the eco-tourism potential in the Wuling Mountain area used the multi-criteria decision analysis method combined with the AHP method for expert scoring. The results showed that the tourism potential in the Wuling Mountain area of Hubei Province was high in the west and low in the east[44]. Our study corroborates this finding and advances it with a predicted range of continuous areas from the Wuling Mountain area to the plain, which also respond to the agenda of all-for-one tourism. Additionally, there are national-level quantitative studies on CES as a reference. For example, the study Spatial Distribution Dataset of Ecosystem Service Values in China[45] visualized the aesthetics and landscape values, according to which almost all areas in Shiyan, Shennongjia, and Enshi. Our study shares a similar scope with it while with more precise estimated results of the internal heterogeneity regarding recreation potential.

5.3 Application of the Model

Conventional techniques for predicting recreation potential frequently rely on market-oriented reports and economic index. For spatial modeling in this study, we used environmental feature variables, which have two-folding benefits to associated planning practices. First, GIS-based modeling and evaluation can provide spatially explicit predictions in continuous areas, and it can overcome the drawbacks of data imbalance widely confronted by traditional methods, such like few samples could be collated from rural or wildness areas[46][42]. Second, the modeling results can be used in planning practice to help build the regional continuous spatial system for the all-for-one tourism. Regional tourism brand development can be also promoted by highlighting the prioritized zones of high recreation potential and integrating the areas with outstanding social and ecological resources for tourism. In addition, spatial modeling can help figure out the challenges and barriers to regional development by inspecting the spatial patterns between high-potential places. Typically, such challenges are found in places with poor infrastructure (e.g., roads, supporting facilities), which need to be responded in future development.

Still, there are constrains to how modeling results can be used in planning practice. For instance, the accuracy of prediction results of the model is determined by the categories, time points, and quality of the input data. Thus it is essential to construct datasets based on empirical research and literature review. It is also important to consider as many factors as possible—environmental, cultural, political, etc.—before the modeling, and turn them into measurable variables in a high-resolution form. Second, when attempting to apply regional-scale modeling results onto smaller-scale sites—for example, from regional to local—practitioners should be aware of the uncertainty across the scales. At a smaller scale, the determination of recreation potential levels may be heavily influenced by social and cultural factors, rather than natural or environmental ones. Hence,

Table 5: Model performance

Algorithm	Threshold	AUC	TSS	JACCARD
MEM	max_sens_spec	0.760852879	0.42029041	0.551155987
GAU	equal_sens_spec	0.722440861	0.314135732	0.493431703
GAU	max_sens_spec	0.722440861	0.367243352	0.532686921
GAU	max_sorensen	0.722440861	0.310617867	0.582119991
GLM	equal_sens_spec	0.733507474	0.33074915	0.504009131
GLM	max_sens_spec	0.733507474	0.389194449	0.54421299
GLM	max_sorensen	0.733507474	0.284418985	0.582530324
RF	equal_sens_spec	0.725745829	0.313885039	0.492374892
RF	max_sens_spec	0.725745829	0.380363152	0.521307786
RF	max_sorensen	0.725745829	0.271024969	0.568871933
MeanW	equal_sens_spec	0.745306141	0.374121363	0.527618818
MeanW	max_sens_spec	0.745306141	0.41126667	0.529763553
MeanW	max_sorensen	0.745306141	0.361814927	0.595895734

NOTES

1. MEM: Maximum Entropy Model; GAU: Gaussian Process Model; GLM:Generalized Linear Model; RF: Random Forest Model; MeanW: Weighted Average Ensemble Model based on the performance of the aforementioned models.
2. Threshold in the table is used to obtain binary prediction values (potential or not): max_sens_spec stands for maximizing the sum of sensitivity and specificity of the model; equal_sens_spec stands for making sensitivity and specificity equal; and max_sorensen stands for maximizing the Sorensen evaluation metric.

a modeling study cannot substitute for field research. In planning, the results of modeling should only be used as a guide for making spatial recommendations as an incorporated part of workflows.

5.4 Limitations of the Research

This study has two main limitations. First, this study did not take into account many humanity factors. Humanity variables are well-acknowledged in tourism research, but they are also challenging to incorporate into continuous spatial modeling. Due to data and method limitations, it is often impossible to directly measure many humanity factors at the regional scale. Based on relevant research, this paper has included a number of humanity-related environmental features, such as built-up areas, farmland, and the distance to towns, partly addressing the problem. The authors attempted to include more humanity and cultural factors (such as socio-economic development level and ethnic minority distribution) into the research, but saw a failure due the lack of data and the inability to spatially represent them precise enough for effective modeling. Moreover, although the study used Weibo check-in data within the study area, the amount of valid samples was limited. Second, when dealing with the occurrence dataset, although this study included various types of known recreational hotspots, each category was treated equally according to the data requirements of SDM, without distinguishing the qualitative differences between types. Future research needs to concentrate on how to assess the occurrences of multi-category social events and aggregate the results within the framework of SDM modeling.

6 Conclusions

Using machine-learning-based SDM ensemble models, this paper proposes a research pathway to predict regional recreation potential for developing all-for-one tourism. The results show that, under CES theories, this approach has a sound performance in estimating the recreation potential by acquiring knowledge about the recognized recreational hotspots in the region and relevant multi-sourced social-ecological data, which exhibits its capacity of supporting the establishment of spatially continuous areas in the all-for-one tourism. The final ensemble model performed well in predicting and had a high accuracy by the testing dataset. When comparing several commonly used algorithms and the ensemble model, the research found that different algorithms generate varied prediction results with the same data, suggesting that the uncertainty must be taken into account in its application. The ensemble model offers a way of predicting that avoids "placing all eggs in a one basket" and has a significance as practical guidance.

The research results of the western Hubei region indicate that a predictive model of recreation potential at the regional scale can be built on environmental features such as land use types, landscape composition, climate factors, and transportation. However, its evaluation system should consider cultural and recreational hotspots that have been recognized or certified by the market or the government. By "learning" the environmental features of the well-known recreational hotspots, the study can help support all-for-one tourism through the identification of high-potential areas, so as to promote the regional tourism development.

REFERENCES

[1] Yang, Z., (2016). The connotation of all-for-one tourism and its development stages. *Tourism Tribune*, 31(12), 1–3.

[2] Li, X., Ma, L., & Chen, L. (2016). Development of all-for-one tourism: Logic and focus. *Tourism Tribune*, 31(9), 22–24.

[3] Hegetschweiler, K. T., de Vries, S., Arnberger, A., Bell, S., Brennan, M., Siter, N., Olafsson, A. S., Voigt, A., & Hunziker, M. (2017). Linking demand and supply factors in identifying cultural ecosystem services of urban green infrastructures: A review of European studies. *Urban Forestry & Urban Greening*, (21), 48–59.

[4] He, Y., & Xie, H. (2019). Exploring the spatiotemporal changes of ecological carrying capacity for regional sustainable development based on GIS: A case study of Nanchang City. *Technological Forecasting and Social Change*, (148), 119720.

[5] Liu, L. (2017). Universe tourism in the strategy of rural revitalization: An analytical framework. *Reform*, (12), 80-92.

[6] Huang, Q. (2016). A research on the spatial structure of tourism resource of ecological and cultural tourism circle in western Hubei. *Central China Normal University Journal of Postgraduates*, 23(2), 152–156.

[7] Fang, S., & Lan, Y. (2017). Evaluation on the urban tourism competitiveness in western Hubei eco-culture circle based on SPSS. *Journal of Hubei University of Arts and Science*, 38(11), 50–55.

[8] Zhang, Y. (2017). *Research on Comprehensive Evaluation and Promotion of Tourism Competitiveness of Western Hubei Eco-Cultural Tourism Circle* [Master's thesis]. Central China Normal University.

[9] Liao, F. (2015). *A Fractal Study on Spatial Structure of Tourist Scenic Spots in Western Hubei Eco-Cultural Circle* [Master's thesis]. Central China Normal University.

[10] Cortinovis, C., & Geneletti, D. (2018). Mapping and assessing ecosystem services to support urban planning: A case study on brownfield regeneration in Trento, Italy. *One Ecosystem*, (3), e25477.

[11] Wen, C., Albert, C., & von Haaren, C. (2022). Nature-based recreation for the elderly in urban areas: Assessing opportunities and demand as planning support. *Ecological Processes*, (11), 44.

[12] Fish, R., Church, A., & Winter, M. (2016). Conceptualising cultural ecosystem services: A novel framework for research and critical engagement. *Ecosystem Services*, (21), 208–217.

[13] Plieninger, T., Bieling, C., Fagerholm, N., Byg, A., Hartel, T., Hurley, P., López-Santiago, C. A., Nagabhatla, N., Oteros-Rozas, E., Raymond, C. M., van der Horst, D., & Huntsinger, L. (2015). The role of cultural ecosystem services in landscape management and planning. *Current Opinion in Environmental Sustainability*, (14), 28–33.

[14] Tratalos, J. A., Haines-Young, R., Potschin, M., Fish, R., & Church, A. (2016). Cultural ecosystem services in the UK: Lessons on designing indicators to inform management and policy. *Ecological Indicators*, (61), 63–73.

[15] Zhao, Q., Li, J., Liu, J., Cuan, Y., & Zhang, C. (2019). Integrating supply and demand in cultural ecosystem services assessment: A case study of Cuihua Mountain (China). *Environmental Science and Pollution Research*, 26, 6065–6076.

[16] Kalinauskas, M., Bogdzevič, K., Gomes, E., Inácio, M., Barcelo, D., Zhao, W., & Pereira, P. (2023). Mapping and assessment of recreational cultural ecosystem services supply and demand in Vilnius (Lithuania). *Science of the Total Environment*, (855), 158590.

[17] Cheng, X., Van Damme, S., Li, L., & Uyttenhove, P. (2019). Evaluation of cultural ecosystem services: A review of methods. *Ecosystem services*, (37), 100925.

[18] Ghorbanzadeh, O., Pourmoradian, S., Blaschke, T., & Feizizadeh, B. (2019). Mapping potential nature-based tourism areas by applying GIS-decision making systems in East Azerbaijan Province, Iran. *Journal of Ecotourism*, 18(3), 261–283.

[19] Paracchini, M. L., Zulian, G., Kopperoinen, L., Maes, J., Schägner, J. P., Termansen, M., Zandersen, M., Perez-Soba, M., Scholefield, P. A., & Bidoglio, G. (2014). Mapping cultural ecosystem services: A framework to assess the potential for outdoor recreation across the EU. *Ecological Indicators*, (45), 371–385.

[20] Lee, S. H., Choi, J. Y., Yoo, S. H., & Oh, Y. G. (2013). Evaluating spatial centrality for integrated tourism management in rural areas using GIS and network analysis. *Tourism Management*, (34), 14–24.

[21] Park, D., Lee, G., Kim, W. G., & Kim, T. T. (2019). Social network analysis as a valuable tool for understanding tourists' multi-attraction travel behavioral intention to revisit and recommend. *Sustainability*, 11(9), 2497.

[22] Wang, Z., Wang, S., & Xu, M. (2020). Assessment of recreation services at the county-level: A case study based on Wusheng City, Sichuan Province. *South Architecture*, (3), 27–33.

[23] Guo, Y., Yang, F., Wang, J., & Wu, R. (2020). Assessment of the tourism and recreation cultural ecosystem services in Three Parallel Rivers Region. *Acta Ecologica Sinca*, 40(13), 4351–4361.

[24] Li, W., & Wang, Q. (2020). Recreation suitability analysis for regional landscape: With the aim of serving for decision-making in multi-functional greenway network planning and further implementation. *Journal of Xi'an University of Architecture & Technology (Natural Science Edition)*, 52(1), 121–128.

[25] Peña, L., Casado-Arzuaga, I., & Onaindia, M. (2015). Mapping recreation supply and demand using an ecological and a social evaluation approach. *Ecosystem services*, (13), 108–118.

[26] Pérez-Valladares, C. X., Moreno-Calles, A. I., Mas, J. F., & Velazquez, A. (2022). Species distribution modeling as an approach to studying the processes of landscape domestication in central southern Mexico. *Landscape Ecology*, (37), 461–476.

[27] Zhang, J., & Li, S. (2017, December). A review of machine learning based species' distribution modelling. In: *2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII)* (pp. 199–206). IEEE.

[28] Li, Y., Fei, T., Huang, Y., Li, J., Li, X., Zhang, F., Kang, Y., & Wu, G. (2021). Emotional habitat: Mapping the global geographic distribution of human emotion with physical environmental factors using a species distribution model. *International Journal of Geographical Information Science*, 35(2), 227–249.

[29] Clemente, P., Calvache, M., Antunes, P., Santos, R., Cerdeira, J. O., & Martins, M. J. (2019). Combining social media photographs and species distribution models to map cultural ecosystem services: The case of a natural park in Portugal. *Ecological Indicators*, (96), 59–68.

[30] Hermes, J., Albert, C., & von Haaren, C. (2018). Assessing the aesthetic quality of landscapes in Germany. *Ecosystem Services*, (31), 296–307.

[31] People's Government of Hubei Province. (2016, May 31). *Notice of Provincial People's Government on Issuance of the 13th Five-Year Plan for Tourism Development in Hubei Province*.

[32] Jin, C., Fan, L., & Lu, Y. (2010). Spatial distribution of agricultural tourism based on accessibility in case of Jiangsu Province. *Journal of Natural Resources*, 25(9), 1506–1518.

[33] Weyland, F., & Laterra, P. (2014). Recreation potential assessment at large spatial scales: A method based in the ecosystem services approach and landscape metrics. *Ecological Indicators*, (39), 34–43.

[34] Gosal, A. S., Giannichi, M. L., Beckmann, M., Comber, A., Massenberg, J. R., Palliwoda, J., Roddis, P., Schägner, J. P., Wilson, J., & Ziv, G. (2021). Do drivers of nature visitation vary spatially? The importance of context for understanding visitation of nature areas in Europe and North America. *Science of the Total Environment*, (776), 145190.

[35] Chhetri, P., & Arrowsmith, C. (2008). GIS-based modelling of recreational potential of nature-based tourist destinations. *Tourism Geographies*, 10(2), 233–257.

[36] Chan, N. W., & Wichman, C. J. (2020). Climate change and recreation: Evidence from North American cycling. *Environmental and Resource Economics*, (76), 119–151.

[37] Rugel, E. J., Henderson, S. B., Carpiano, R. M., & Brauer, M. (2017). Beyond the Normalized Difference Vegetation Index (NDVI): Developing a natural space index for population-level health research. *Environmental Research*, (159), 474–483.

[38] Hernández-Morcillo, M., Plieninger, T., & Bieling, C. (2013). An empirical review of cultural ecosystem service indicators. *Ecological Indicators*, (29), 434–444.

[39] Tao, Y., Tao, Q., Sun, X., Qiu, J., Pueppke, S. G., Ou, W., Guo, J., & Qi, J. (2022). Mapping ecosystem service supply and demand dynamics under rapid urban expansion: A case study in the Yangtze River Delta of China. *Ecosystem Services*, (56), 101448.

[40] Velazco, S. J. E., Rose, M. B., de Andrade, A. F. A., Minoli, I., & Franklin, J. (2022). FLEXSDM: An R package for supporting a comprehensive and flexible species distribution modelling workflow. *Methods in Ecology and Evolution*, 13(8), 1661-1669.

[41] Li, F., Li, F., Li, S., & Long, Y. (2020). Deciphering the recreational use of urban parks: Experiments using multi-source big data for all Chinese cities. *Science of the Total Environment*, (701), 134896.

[42] Hu, B., & Zhao, J. (2022). Factors promoting nature-based outdoor recreation during the daytime and evening. *Journal of Outdoor Recreation and Tourism*, (40), 100572.

[43] Gobeyn, S., Mouton, A. M., Cord, A. F., Kaim, A., Volk, M., & Goethals, P. L. M. (2019). Evolutionary algorithms for species distribution modelling: A review in the context of machine learning. *Ecological Modelling*, (392), 179–195.

[44] Liu, Q., & Wang, Z. (2021). The potential and spatial heterogeneity of ecotourism in the Wuling Mountain Area, China. *Mountain Research*, 39(4), 563–575.

[45] Xu, X. (2018, January 1). *Spatial Distribution of Terrestrial Ecosystem Service Values in China*. Resource and Environment Science Data Registration and Publication System.

[46] Martin, C. L., Momtaz, S., Gaston, T., & Moltschaniwskyj, N. A. (2016). A systematic quantitative review of coastal and marine cultural ecosystem services: Current status and future research. *Marine Policy*, (74), 25–32.

https://doi.org/10.15302/J-LAF-1-020073 中图分类号 | TU984, TP181 文献标识码 | A

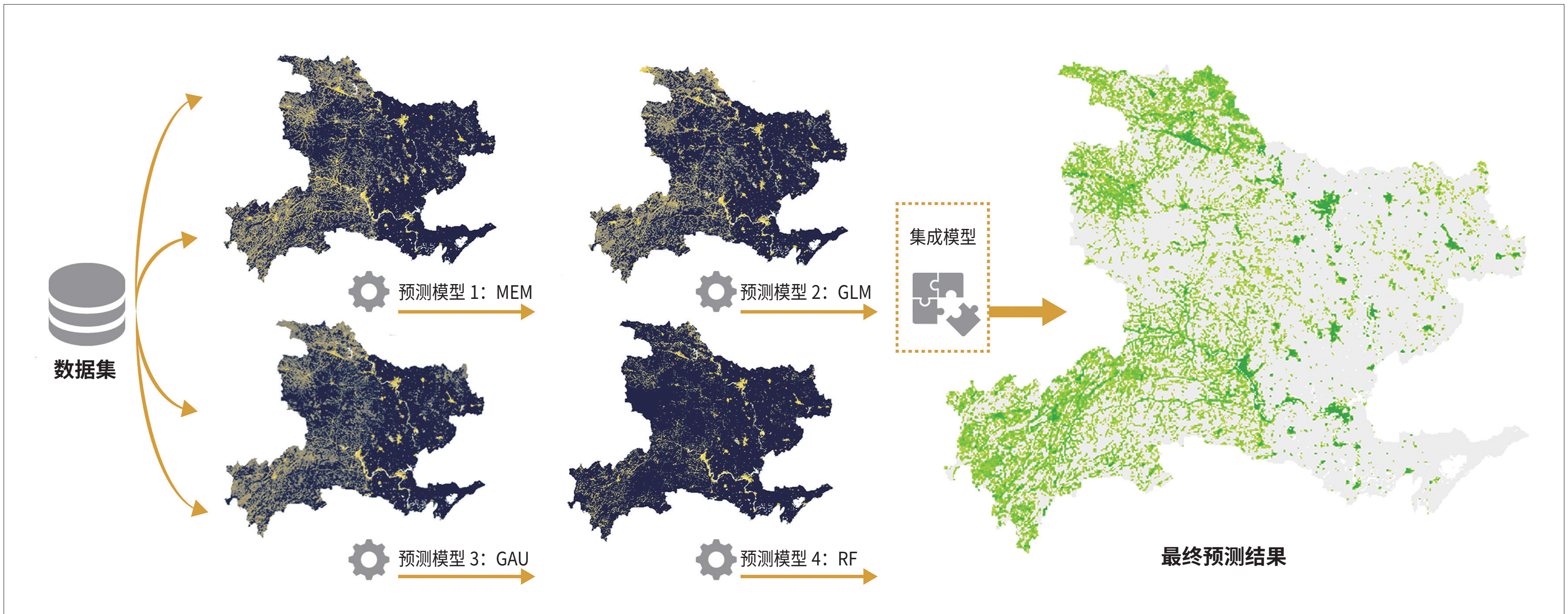
“全域旅游”导向下鄂西游憩服务的空间潜力预测研究——基于集成模型的机器学习方法

文晨[1,2]，茶静[1,2]，徐利权[1,2,3,*]，徐海韵[4]

1 华中科技大学建筑与城市规划学院，武汉 430074
2 自然资源部城市仿真重点实验室，武汉 430074
3 华中科技大学建成遗产研究中心，武汉 430074
4 北京建筑大学建筑与城市规划学院，北京 100044

*通讯作者
地址： 湖北省武汉市洪山区珞喻路1037号
邮编： 430074
邮箱： xlq2019hust@hust.edu.cn

图文摘要



文章亮点

- 尝试解决全域旅游发展中“连点成片”空间定位这一关键技术难点
- 以社会－生态多源数据构建了一个结合集成机器学习的SDM模型
- 以鄂西地区为实证对象，量化分析了区域尺度下连续空间的游憩潜力

关键词

全域旅游；
游憩服务；
生态系统文化服务；
空间潜力预测；
机器学习；
鄂西

摘要

在建设“全域旅游”的背景下，区域尺度的游憩服务发展将从单一景点景区的建设转向旅游目的地的综合统筹，助力乡村振兴和区域协调发展。然而，在全域旅游“连点成片”的过程中，如何根据本地独特的环境禀

赋识别出具有较高游憩潜力的区域并据此评估发展的优先程度，仍是研究与实践的热点和难点。基于此，本研究以鄂西地区为例，引入生态系统文化服务理论中潜力评估的研究方法，运用社会-生态多源数据构建了一个结合集成机器学习的SDM模型。该模型对研究区域内336个已知游憩服务热点的环境特征进行了剖析，并预测了连续空间中高游憩潜力区域的概率分布。本研究提供了一条从环境特征变量数值关系角度理解区域尺度游憩空间规律的技术路径，旨在为全域旅游和乡村振兴的空间发展策略提供参考。

收稿时间

2022-12-13

编辑 田乐

1 背景概述

近年来，随着中国的生态文明建设和空间治理体系进入新的发展阶段，全国各地积极探索区域协调发展，强化城乡统筹融合，并深化了“全域旅游”的概念。全域旅游是指把整个区域作为旅游区考虑，推动旅游和全产业融合，依托现有旅游资源，以知名景点景区、乡村旅游热点带动相关旅游片区发展，全面激活游憩服务潜力[1]。在此背景下，区域尺度的旅游发展模式正在经历从单一景点景区建设管理到旅游目的地综合统筹的转变[2]。

全域旅游发展并不是均匀发力，重点发展区的选位和资源投入需要科学决策。实践中需要根据当地的自然资源、人文资源、基础设施等禀赋，识别出具有较高游憩服务潜力的区域进行重点发展，并借此考虑不同区位的开发成效，制定优先发展策略。因此，区域尺度的城市及乡村生态—文化旅游资源潜力评估也成为了学界的新兴关注点[3]~[5]。一些研究从空间角度提出了支持性策略，如梳理景区旅游资源、人口、经济、基础设施的分布和集中度等，并在此基础上评价旅游竞争力和探讨开发模式[6]~[9]。然而，现有研究较少从空间上对当地独特的生态—文化旅游资源进行量化和可视化，尤其是以全域视角前瞻性地识别出高游憩潜力区域，并从中筛选出适合优先发展的区域。这些研究的缺失阻碍了精细化城乡融合发展策略的制定。

在此背景下，本研究旨在通过引入生态系统文化服务（cultural ecosystem services，CES）理论，尤其是运用CES游憩潜力空间测定的方法论，搭建一套基于机器学习的空间分析框架，运用多源数据进行区域尺度游憩潜力预测。研究沿用CES研究的语境，将“游憩服务潜力”（recreation potential）界定为自然环境提供游憩活动或体验的可能性[10][11]。本文旨在达到两个相互关联的研究目标：1）根据本地一些受认可的游憩服务热点预测全域连续空间的游憩服务潜力；2）从环境特征数值关系的角度理解区域尺度下游憩服务的空间规律。

2 文献综述

CES是指人类通过与自然接触所能得到的非物质类类福祉，包括游憩、审美体验、精神感知、宗教、教育、治愈、启发创造等[12]。游憩服务具有抽象、主观、难以量化等特点，因为其不仅取决于景观资源的特征——包括生态环境类型、数量、多样性（组合）、开发程度（基础设施）等[12]——也取决于公众需求和社会文化。沿用CES理论来研究游憩服务潜力，可以借力于生态系统服务成熟的研究框架来辅助量化评价，将人的需求和景观的环境特征建立联系，从而更好地融入规划框架[13][14]。

现有CES研究用于空间测定游憩服务潜力的方法主要有空间叠加法（多准则决策分析）、网络分析法、空间回归法、指标建模法、基于社交网络的空间可视化法等[15]~[17]。例如，多个研究结合层次网络分析（AHP）或其他权重赋值方法，确认了影响游憩服务潜力的不同环境变量，然后通过GIS的叠加法计算出空间的游憩服务潜力[18][19]。另有研究通过分析不同乡村的空间中心性、相互间的联系强度及服务设施，综合测算了它们的旅游潜力[20]。一些新的方法（如融入市民科学的叠加分析、PPGIS、基于社交网络大数据的可视化方法）在相关研究中也展现了良好的适用性[16][20]~[25]（表1）。

近年来，有研究开始使用物种分布模型（Species Distribution Modeling，SDM）进行区域尺度社会文化现象的空间预测，取得了良好成效。SDM源自生态学研究，主要是将生物的观测记录点信息和环境特征建立联系，可通过机器学习方法拟合模型并生成连续空间下的预测结果[26][27]。在社会文化话题中，SDM被用于通过建成环境预测人类情感表达[28]，或者通过自然环境特征预测国家公园的社会偏好[29]。在相关研究中，尽管地理环境因素并不是直接促成社会文化现象发生的唯一因素，但使用SDM进行预测的理论依据是地理环境因素仍可以作为代理变量从不同角度反映人文因素[28]。同时，由于人文要素往往难以在空间中精确量化和可视化，借助其与地理环境因素之间相互渗透和影响的关系成为

表 1：CES 框架下游憩服务潜力空间测定的近期研究

研究文献	研究区域	所研究的游憩服务准则	计算方法	评论
参考文献 [22]	四川省武胜县	自然资源、文化景点、景观美感、可达性、配套产业建设	通过多源空间数据量化各项评价准则，再通过基于专家打分法的 AHP 和多准则决策分析综合叠加	连续空间下的游憩服务潜力评估，考虑了自然和人文要素
参考文献 [23]	“三江并流”区域，云南省西北部的 16 个市县	景观多样性、自然度、河流湖泊元素、景点等级	将各个指标量化后转化为 5 个等级，再把各类指标等权重叠加	侧重挖掘自然游憩服务潜力，等权重法认为各准则的影响力一致
参考文献 [24]	浙江省台州市	土地利用、坡度、与城镇的距离关系，与水系与湿地的距离关系、历史文化资源	将每个准则使用空间数据进行量化后，分为若干等级，并分别赋予标准分值，通过叠加得到最终结果	通过专家小组对因子直接赋分，但打分过程未详细介绍
参考文献 [16]	立陶宛东南部	自然类别（水体、自然度、保护地状态、地形多样性等）；文化类别（游览路径、自行车道、世界文化遗产、观景点）	各个类别的指标进行 z 值标准化后，进行叠加分析，叠加结果经过了问卷调查的验证	不同因子仍按等权重处理，但引入了市民验证的程序，证明了有效性
参考文献 [20]	韩国西南部全罗南道	交通路网视角下乡村旅游潜力评估	使用图论的多个中心性指标评价可达性，以此作为乡村游憩服务潜力	以村落作为研究对象，侧重开发角度，主要关注交通便捷，未讨论乡村本身的特征
参考文献 [25]	西班牙北部巴斯克地区	自然度、自然保护地状态、水体、地形、地理兴趣点等	引入“视域”概念作为统计单元，计算各个单元内涵盖的游憩服务准则因子	引入基于照片比选的问卷作为验证

了一种研究的替代手段[30]。

然而，现有区域尺度的研究仍存在以下不足。首先，常用的叠加法和多准则分析法在设定不同影响因素的权重时较为主观，缺乏复现和验证研究，而目前所采用的专家打分法也通常会引入不确定性因素。其次，现有建模研究在制定指标时较少挖掘本地已得到认证或认可的游憩服务点，导致地域性难以体现。最后，除城镇、村落、景点外，数据样本较为稀缺的荒野等连续空间较少被纳入研究，包括SDM建模。

3 研究资料与方法

3.1 研究区域

研究场地位于湖北西部——鄂西区域全境（图1），包含襄阳、宜昌、十堰、荆州、荆门、随州、恩施、神农架等八市（州、林区），覆盖人口超2 730万[31]，主要为山地环境。作为一个区域概念，鄂西既是生态系统文化服务多样且丰富的地区（图2），也是发展旅游扶贫的重点区域。近年来，湖北省政府重视鄂西全域旅游发展，强调根据区域禀赋构建湖北旅游新格局。因此，针对如何前瞻性预测和规划全域旅游服务潜力空间格局的研究具备典型性和迫切性。

3.2 机器学习数据集构建

3.2.1 发生数据集

研究采用机器学习的建模框架，将现有受认证的游憩服务离散点位作为发生观测，将鄂西地区内的社会—生态多源数据作为环境特征。

其中，发生数据集是对客观发生事件的记录观察，在本研究中指研

究场地中已经产生游憩服务的点位。根据文献和预调研结果，研究选择了三类游憩服务热点数据构建区域内的发生数聚集：1）国家A级旅游景点景区和湖北省精品旅游线路中所涉景点；2）国家级传统村落；3）社交网络微博热门打卡地点，即用户曾到访该热点。这些游憩服务热点代表了受官方认证或市场认可的游憩服务。

本研究采用了如下空间信息处理方式。首先，研究团队通过高德API地理编码获取上述游憩服务热点的空间坐标，并将之转换为UTM投影坐标系；而后在预处理过程中，研究发现所获得的微博签到数据与区域人口分布高度相关，存在仅代表居住点而非游憩打卡点的问题，所以研究剔除了所有主城区范围内的打卡数据，仅合并保留了乡镇、乡村的热门签到点位。最终得到的数据集包括336个空间点位。

3.2.2 环境特征数据集

在构建环境特征数据集时，研究团队参考了现有CES文献中有关游憩潜力建模的研究[19][30][32]~[37]（表2），将土地覆盖、地形地貌、景观构成、气候、交通便利等多类别共计16个变量作为本研究的环境特征类别。为保证建模过程的标准化和可复现性，研究遵循“空间明确”原则[38]，将每一个特征处理为栅格数据，并覆盖场地内所有区域。

在本研究中，土地利用和景观指数这两类数据的处理方式较为特殊。由于集成模型的机器学习方法要求特征数据不可为定类数据，仅允许数值，因此研究将土地利用类型转化为了四类用地（林地、建成区、农田、水体）的构成比例值，这种处理方法也可以较为准确地反映场地内土地覆盖的混合状况。计算景观指数时，研究使用了“移动视窗”的方法计算了每一个像元邻域内的全局景观指数，包括JOINENT（景观联合熵值）等指标。这种分析计算负荷较重，研究仅选取了JOINENT和SHDI（香农多样性指数）两个指标来量化多样性和复杂性特征。根据同类区域尺度的研究[39]，研究将环境特征数据集的空间分辨率统一为1 000m，非此分辨率的变量采用重抽样统一处理为1 000m精度。环境特征数据集构建完毕后，将由发生数聚集提取336个已知游憩服务发生点所在位置的所有环境特征数值。

表 2：本研究构建的环境特征数据集

环境特征名称	编码	计算方法或描述	数据来源	文献来源
土地覆盖占比－林地	Fractions_Forest	分析单元中林地占比，本研究采用重抽样处理	欧盟“哥白尼全球土地覆盖”数据（Global CGLS-LC100）	参考文献 [33]
土地覆盖占比－建成区	Fractions_Built	分析单元中建成区占比，本研究采用重抽样处理	欧盟“哥白尼全球土地覆盖”数据（Global CGLS-LC100）	参考文献 [34]
土地覆盖占比－农田	Fractions_Cropland	分析单元中农田占比，本研究采用重抽样处理	欧盟“哥白尼全球土地覆盖”数据（Global CGLS-LC100）	参考文献 [32]
土地覆盖占比－水体	Fractions_Water	分析单元中水体占比，本研究采用重抽样处理	欧盟“哥白尼全球土地覆盖”数据（Global CGLS-LC100）	参考文献 [35]
最热月平均地表温度	LST	分析不同区位的地表温度，目的为分析“避暑”等重要的旅游吸引力	根据 Landsat8 卫星影像反演。计算时间段为 2021-07-01 至 2022-07-01	参考文献 [36]
年平均标准化植被指数	NDVI	量化绿色植被相对丰度	根据 Landsat8 卫星影像计算，计算时间段为 2021 年	参考文献 [33]
年平均植被增强指数	EVI	反映植被生物物理参数，矫正了大气土壤噪音，并对茂密的林区更敏感	根据 Landsat8 卫星影像计算，计算时间段为 2021 年	参考文献 [37]

续表见下页

表 2：本研究构建的环境特征数据集

环境特征名称	编码	计算方法或描述	数据来源	文献来源
香农多样性指数	SHDI	景观指数，描述植物群落的多样性，本研究采用移动视窗算法，邻域半径为 1km	根据土地覆盖数据 Globeland30 计算	参考文献 [33][35]
景观联合熵值	JOINENT	景观指数，描述格局的复杂程度，本研究采用移动视窗算法，邻域半径为 1km	根据土地覆盖数据 Globeland30 计算	参考文献 [33]
斑块丰度	PR	景观指数，描述景观斑块的丰富程度，本研究采用移动视窗算法，邻域半径为 1km	根据土地覆盖数据 Globeland30 计算	参考文献 [33]
高程	ELEV	场地海拔高度	ALOS-2 高程数据	参考文献 [30][35]
坡度	SLOPE	场地坡度	根据高程数据计算	参考文献 [30][35]
地形粗糙指数	TRI	反映地表起伏和侵蚀程度	根据高程数据计算	参考文献 [30]
地形位置指数	TPI	反映地形地貌类型	根据高程数据计算	参考文献 [30]
到达最近的城镇距离	DIST_TOWN	衡量交通便利状况	根据全国镇一级居民点位置数据计算	参考文献 [32]
到达最近的水体距离	DIST_WATER	衡量与水体的远近关系	根据土地利用数据 Globeland30 计算	参考文献 [19][35]

3.3 融合集合机器学习框架的SDM建模

3.3.1 建模前预处理

SDM模型被用于预测某一事件在空间中发生的概率。区域尺度下，研究通常缺乏游憩服务（因变量）在连续空间内的详细发生状况，仅有若干点位的观测记录。所以SDM模型“以点探面”方法能够通过学习已知游憩服务热点的环境特征（自变量），将预测概率推广至全域。由于SDM模型允许基于不同的算法和参数来学习发生数据和环境特征数据的关系，因此建模过程灵活，但预测结果往往存在差异。本研究数据分析过程使用编程语言R和Python实现；计算结果借助QGIS进行地图可视化。

SDM模型对预测范围和环境特征的共线性较为敏感。在建模预处理阶段，研究确认预测范围为鄂西全域，并使用环境过滤法来预筛选预测范围内的发生数据集，避免数据点因空间距离过近而对建模结果产生影响。按照模型要求输入数据后，算法针对环境特征数据集的潜在共线性问题，随机生成了伪缺失数据集，并使用了主成分分析的方法对数据进行降维。

研究使用了空间块划分法[40]对336个游憩服务点位进行训练集和验证集的拆分。具体来说，研究将数据集分为四个块（也称为“折叠”），其中三个块用作训练集，另一个块用作验证集；而后循环遍历每个块，将之作为验证集，并使用其余的块作为训练集来训练模型。这种技术可以帮助评估模型的性能，并减少因数据集划分不均匀而引起的偏差。

为了分析不同的环境变量对模型预测性能的贡献，研究构建了纳入了不同环境特征变量的三组模型（表3），每组模型历经了同样的分析流程，通过比较最终的模型预测性能来分析不同环境特征对于预测游憩服务潜力的显著程度。

3.3.2 使用集合模型对数据进行拟合与验证

研究使用最大熵模型（MEM）、高斯过程模型（GAU）、广义线性模型（GLM）、随机森林（RF）这四类机器学习算法进行建模。这些算

法是SDM模型研究中常用的算法，具有不同的假设、数据处理过程和优劣势。

为了避免武断选择某一算法和权重参数，研究采用基于集成模型的机器学习方法，将若干机器学习算法合并，达到全局更优的预测效果。研究首先对不同单独算法的预测结果进行加权平均，权重的依据是各模型的性能评价指标，包括TSS（真实技巧统计值）、AUC（ROC曲线下面积）和JACCARD（杰卡德相似系数）。尽管MSE（均方误差）和MAE（平均绝对误差）是常用于评估统计模型性能的指标，但它们在SDM类型的建模中较少使用。这是因为MSE和MAE不考虑真负数，在本研究语境中即发生数据集之外的"缺席"区域；而在SDM模型中，预测事件的缺席与预测其存在一样重要。因此，TSS等指标更适合于评估SDM模型的预测性能。调试完毕的模型将用于预测鄂西游憩服务潜力的空间概率，并完成相应的游憩潜力制图。

4 研究结果

4.1 发生数据和环境特征数据分析结果

发生数据集包含336个游憩服务发生点，主要位于十堰、恩施、襄阳、宜昌。通过样本点环境特征分析可知，游憩服务发生点的各个环境特征变量并非均匀分布，在多个环境特征中具有明显的聚集效应（图3，4）。游憩服务发生点所处位置的平均海拔为591m，平均坡度为3.85°；所属公里网格的林地占比超过50%，建成区占比7.8%，农田占比17.7%；距离最近的城镇约22km，距离最近的湖泊和河流约7.6km。整体而言，鄂西区域的生态状态较好，体现为NDVI和EVI指数在全域范围均较高；其中海拔较低和相对平坦的随州、荆门、荆州等市域范围农田占比高，水系网络发达；通过景观指数，研究识别出若干带状区域具有较高的景观多样性和复杂度。

由于环境特征数据集包含多个类别且不同取值范围的变量，为了避免可能存在的共线性和参数估计失真，研究使用10个主成分提取了原始数据中累计达95%的可解释变异（表3），这10个主成分作为自变量参与模型构建，以实现数据降维。在三组模型中，纳入了所有环境特征变量的模型三性能表现最好，将以此作为空间可视化的对象。图5展示了4种单独的算法的预测结果，图6展示了基于模型三的全域游憩潜力的预测概率地图和二元适应性地图。

4.2 预测结果和模型表现

不同算法下的预测结果显示，随机森林模型预测的高游憩服务潜力区域相对较少，主要体现为以西北－东南向的两个带状区域，包括荆州－枝江－宜都－宜昌－秭归－巴东带状区域，以及襄阳－谷城－老河口－丹江口－十堰带状区域及周边的星点区域。其他三个模型——最大熵模型、高斯过程模型、广义线性模型——也识别出了长江沿岸和这两个带状区域，但又额外识别出大量山地地区作为高游憩服务潜力区域。这三个模型之间的差异主要体现在识别高游憩服务潜力区域的阈值及判定的面积。

在运用相同环境特征数据集的情况下，三个模型的预测结果略有差异，体现出了模型的不确定性。以评价指标AUC和JACCARD来看，三个模型的指标数值大致为0.7和0.5以上，且差异较小，显示了较好的模型表现；但是以评价指标TSS来看，各模型表现差异较大，指标数值在0.27~0.42间波动，说明模型仍有随机性。

集成模型通过预设的阈值能够融合多个单独的算法，加权平均预测

表 3：用于构建模型的三组环境变量

编号	说明	环境变量
模型一	基础模型：用地构成、地形	Fractions_Forest、Fractions_Built、Fractions_Cropland、Fractions_Water、ELEV、SLOPE、TRI、TPI
模型二	进阶模型：用地构成、地形、水域和城镇的邻近度	Fractions_Forest、Fractions_Built、Fractions_Cropland、Fractions_Water、ELEV、SLOPE、TRI、TPI
模型三	完整模型：用地构成、地形、水域和城镇的邻近度、景观指数、光谱指数	Fractions_Forest、Fractions_Built、Fractions_Cropland、Fractions_Water、ELEV、SLOPE、TRI、TPI、LST、NDVI、EVI、SHDI、JOINENT、PR

结果，从而能够提高整体的模型表现，提升预测准确率（表4）。预测结果显示，最高游憩服务潜力地区较为分散，整体研究区域中都有分布，但是西部地区的整体游憩服务潜力较高，而且一些明显条带状区域具备连片条件（图6）。另外，相较于单独的算法，集成模型提供的概率值预测结果进一步细化了预测结果，精细识别出了位于长江水系与荆山等山脉交汇处的高游憩服务潜力区域，例如在襄阳和宜昌市域范围内的部分区域。

5 讨论

5.1 理论和方法论意义

游憩服务是复杂的社会—生态耦合现象。在规划实践中，预测区域空间的游憩服务潜力已经成为制定发展策略的重要参考依据。本研究沿用CES中游憩服务潜力测定中的相关理论，使用SDM框架下的机器学习集成模型进行预测。研究具有以下创新意义。

在CES理论中，由于同时涉及群体偏好和客观环境因素，游憩服务潜力在连续空间中的评价一直是相关研究与实践的难点[19][30]。本研究通过现有已知的游憩服务热点的环境特征作为机器学习的模型参数，既能够从数值角度理解游憩服务潜力的规律，又能够克服常用的专家打分带来的不确定性及主观性，提供了一种区域尺度游憩服务潜力评估的新路径。

需要指出的是，不同的方法有各自的优劣势，因为各自的适用场景有所不同，并且能够互补。常用的基于专家打分的叠加法通过运用受访者的专业知识和经验对评估的影响因子进行加权，具有可解释性强、操作灵活、能代表不同利益相关方等特点。通过增加受访专家的数量，该方法还能够有效降低不确定性。而机器学习方法则是通过不同的算法从多源数据中学习空间模式，具有可复现、效率高、延展性强等特点。在实证研究中，本案例选择了国家A级旅游景点景区、湖北省精品旅游线路中所涉景点、国家级传统村落、社交网络微博热门打卡地点作为样本点。这些类别代表了区域尺度下的常见旅游目的地，体现了复合的群体偏好，回应了CES在相应空间研究中关于偏好的难点[17]。

在技术方法论上，目前运用SDM模型探索社会事件的发生规律仍在起步阶段[28]，本研究则以多类环境特征变量在区域尺度推进了对游憩服务潜力的探索。其中，研究使用了三组模型进行建模，分步骤地纳入了不同类别的环境特征变量，以剖析不同变量的预测能力。结果表明，在用地构成、地形等环境特征变量之上，加入水域和城镇的邻近度、景观指数、光谱指数、气候等变量能提升模型在研究场地的预测表现（表5）。这也印证了现有文献中对于游憩服务潜力影响因素的研究[41][42]。鉴于CES，特别是游憩服务是复杂的社会—生态现象，在构建数据集时应预先调查研究场地的现有游憩服务热点具有何种吸引力特征，从而针对不同的研究场地和目的构建不同的数据集。

表 4：环境特征的主成分提取

主成分	解释方差的累计百分数
1	0.325
2	0.504
3	0.604
4	0.679
5	0.744
6	0.797
7	0.848
8	0.895
9	0.927
10	0.956
11	0.974
12	0.987
13	0.992
14	0.997
15	0.999
16	1

其次，在机器学习框架下，模型的预测能力能够借助完善的评价指标进行分析[43]。本研究提出的技术路线能够在区域尺度产生连续的预测结果，不依赖特定统计单元（如行政区或特定景区）的划分。因此，该方法同样适用于边界难以确定的乡村和荒野地区，适合全域旅游的研究。

5.2 实证研究重要结果解释

本研究提供了鄂西地区连续空间中游憩服务潜力的预测结果。其

表 5：模型的性能评价

模型	阈值标准	AUC	TSS	JACCARD
MEM	max_sens_spec	0.760852879	0.42029041	0.551155987
GAU	equal_sens_spec	0.722440861	0.314135732	0.493431703
GAU	max_sens_spec	0.722440861	0.367243352	0.532686921
GAU	max_sorensen	0.722440861	0.310617867	0.582119991
GLM	equal_sens_spec	0.733507474	0.33074915	0.504009131
GLM	max_sens_spec	0.733507474	0.389194449	0.54421299
GLM	max_sorensen	0.733507474	0.284418985	0.582530324
RF	equal_sens_spec	0.725745829	0.313885039	0.492374892
RF	max_sens_spec	0.725745829	0.380363152	0.521307786
RF	max_sorensen	0.725745829	0.271024969	0.568871933
MeanW	equal_sens_spec	0.745306141	0.374121363	0.527618818
MeanW	max_sens_spec	0.745306141	0.41126667	0.529763553
MeanW	max_sorensen	0.745306141	0.361814927	0.595895734

注

1. MEM：最大熵模型；GAU：高斯过程模型；GLM：广义线性模型；RF：随机森林模型；MeanW：基于上述各模型性能的加权平均集成模型。
2. “阈值指标”指用于获取二元预测值（是 / 否为潜力区域）的阈值标准：max_sens_spec 表示最大化模型的灵敏度和特异性之和；equal_sens_spec 表示使灵敏度和特异性相等；max_sorensen 表示使 Sorensen 评价指标最高。

中部分区位呈现出的空间模式与一些现有研究结果相一致，验证了本研究的有效性。例如，现有研究使用了多准则决策分析法研究了武陵山区的生态旅游潜力，并通过结合AHP的专家打分法构建了评价体系，发现湖北域内的武陵山区的旅游服务潜力呈现“西高东低”特征[44]。本研究同样识别出这一特征，并且预测范围包含从武陵山区向江汉平原过渡的连续空间，回应了全域旅游对景区周边和景点之间的关切。另有一些全国范围的CES量化研究，如《中国陆地生态系统服务价值空间分布数据集》[45]对“美学景观价值”进行了可视化，其中十堰、神农架、恩施等地几乎所涵盖区域都具有较高价值。本研究的结果在鄂西区域尺度上与之相近，但进一步细化了游憩服务潜力的内部差异，使预测更有空间区分度和实用性。

5.3 模型运用

传统的游憩服务潜力预测方法多依托市场研究和经济指标，本研究主要运用环境特征变量进行空间建模，在实践中具有以下两大优势。其一，基于GIS的空间建模和评价能够在连续地理空间中识别出具有不同游憩服务潜力的区域，并在区域尺度中克服传统方法中常见的数据样本不平衡的问题[46]（仅少数城镇有游人或游憩经济数据，乡村区域数据匮乏）。其二，在实际应用中，全域旅游的开发可分步骤、有计划地连点成片。首先依托高游憩服务潜力区域建立优先开发区，整合基础资源较好的区域，完善区域性游憩品牌的建设。同时，空间建模通过识别高游憩服务潜力区域间的空间关系，分析出连片发展的挑战和障碍。这些障碍通常是基础设施（如道路、游憩服务设施）相对薄弱的区域，后期开

发建设可予以重点考虑。

然而，模型结果在运用过程中也存在限制。例如，模型的构建依托输入数据的类别、时间点，以及数据质量，因此为了保证研究结果的准确性，需要根据实证调研和相关文献构建数据集，并尽可能在研究初期分析较多的因素，包括环境、文化、政策等层面，并考虑以较新和较高分辨率的形式发展成指标变量。其次，当尝试将区域尺度的建模结果用于分析村镇或较小尺度时，应警惕跨尺度带来的不确定性。在小尺度范围，如城市内部和村镇，精细的游憩服务潜力分区可能取决于环境特征之外的因素，尤其是社会文化因素。因此，建模研究不能取代现场调研。在规划实践中，应把建模结果当作空间定位的参考，纳入相关工作流程。

5.4 研究不足

本研究主要有两点不足。首先，本研究的主要不足在于没有将更多人文因素纳入建模分析。人文因素在旅游相关研究中的重要性已得到广泛认可，也是连续空间建模研究中的难点。然而，由于数据和方法的限制，很多人文因素难以直接量化。本研究根据相关文献纳入了若干可以反映人文因素的环境特征，如建成区、农田、距离城镇距离等信息，部分弥补了该问题。作者曾尝试纳入更多的人文特征变量，如社会经济发展水平和少数民族分布等，但是仍受制于数据不足和无法精确空间化的掣肘。研究虽然采集了研究区域内的微博签到数据，但有效数据较少。其次，研究虽然纳入了不同类别的游憩服务热点，但遵循SDM模型的数据要求，仍将每一个发生点均等对待，未区分不同游憩服务热点的质性差异。在SDM模型框架下，如何分析多类别社会事件的发生并将结果聚合，将是未来研究的重点。

6 结论

本研究运用基于集成模型的机器学习方法，基于SDM模型提出全域旅游视角下区域游憩服务空间潜力预测的新研究路径。结果表明，在CES的理论框架内，该路径能够通过学习已知的游憩服务热点，通过多源社会—生态数据预测全域游憩服务潜力，有助于真正实现全域旅游中“连点成片”的目标。最终的集成模型在多个指标下的预测表现较好，在数据测试集中也有较高准确率。研究团队在比较了常用的算法和集成模型后发现，在使用相同数据的条件下，不同算法的差异可能导致预测结果的不同，并在空间表征有所差异。如果使用建模结果指导规划实践，则需要考虑当中的不确定性。集成模型可以提供了一种避免“将鸡蛋放在一个篮子里”的预测方法，具有实践指导价值。

鄂西地区的研究结果表明，区域尺度的综合游憩服务潜力预测结果与用地类型、景观构成、气候要素、交通便利等因素相关，但其评价体系的构建应该充分考虑区域内已受到官方认证或市场认可的文化游憩服务热点。在发展全域旅游的背景下，将通过热点“学习”到的环境特征推广到全域，能更好地指导“连点成片”，并通过可达性分析聚焦关键空间位置，助力区域发展。

图 1. 研究场地内受认证的游憩服务发生点
图 2. 鄂西游憩服务点示例
图 3. 环境特征的空间建模结果
图 4. 环境特征的数值分布
图 5. 基于不同算法的游憩服务潜力预测结果
图 6. 集成模型的预测结果和二元适用性分析