Study On Real Estate Location Based on GIS Technology

Xuan Cai [1, #], Zhiyu Lin [2, #], Gang Yang [3], Bowen Liu [1, *, #]

[1] School of Finance, Nanjing Audit University, Nanjing, China, 211815

[2] School of Government Audit, Nanjing Audit University, Nanjing, China, 211815

[3] School of Applied Engineering, Henan University of Science and Technology, Luoyang, China, 471023

* Corresponding Author Email: bylorw@gmail.com

[#]These authors contributed equally.

Abstract. Through a review of relevant literature,this study categorizes factors affecting real estate site selection into three main categories: environmental, social, and economic, termed as the "triple bottom line." Based on the "triple bottom line," a PUM (Portfolio Underwriting Model) model was constructed. Meanwhile, GIS technology was utilized to identify the locations of public transportation stations, medical institutions, schools, and fire stations within Xiamen City. Additionally, different regions were delineated into suitable residential areas based on slope, aspect, land type, and per capita GDP. Combining the PUM model with GIS technology ultimately aids in real estate market site selection. The research findings indicate that integrating the PUM model and GIS technology for real estate site selection allows for a more scientific and comprehensive evaluation of various site selection factors, thereby providing important decision-making support for developers and investors.

Keywords: Real estate site selection, Triple bottom line, PUM model, GIS technology.

1. Introduction

Today, global climate change is becoming increasingly severe, characterized by rising global average temperatures, rising sea levels, acid rain, and an increase in the frequency of extreme weather events. Moreover, extreme weather and climate events such as heatwaves, heavy rainfall, sudden droughts or floods, wildfires, and dust storms are becoming more frequent [1]. These enormous losses brought the number of property insurance claims for insurance companies to unprecedented levels. Many insurance companies are also reluctant to bear the enormous risks brought to the real estate industry by unforeseen factors.Faced with the increasing investment risks, whether insurance companies are willing to underwrite the risks of real estate development, how to conduct real estate site selection is crucial for real estate enterprises.

Research on real estate project site selection has been conducted for quite some time. And scholars at home and abroad have conducted research from different perspectives and using different methods, forming some conventional site selection models. Liu Xingyu utilizes research methods such as literature review, questionnaire survey, expert interviews, and analytic hierarchy process. By applying relevant research theories and case studies of site selection, he comprehensively identifies significant factors influencing site selection and constructs an evaluation model [2]. Zhang Qi and Xie Peng analyze the advantages and disadvantages of transportation location, natural environment, topography, and geomorphology through qualitative and quantitative analysis. They establish evaluation criteria for different types of real estate projects, providing important site selection evidence for real estate project investments [3]y constructing a rough set BP neural network model, the indicators affecting real estate site selection decisions are simplified, and the main indicator factors affecting site evaluation are extracted using attribute reduction algorithms. This model provides real estate enterprises with a more effective and practical new method for site selection decisions in the real estate industry [4]. Therefore, drawing on the experience of predecessors, we first established a risk assessment model and an Operational Efficiency Model. By determining the risk index and Underwriting Index, we established a PUM (Portfolio Underwriting Model) model. We then combined the PUM model with GIS technology to form a real estate site selection model to assist in real estate site selection. Finally, taking Xiamen as an example, we selected suitable residential areas in Xiamen. Our workflow is shown in Figure 1.

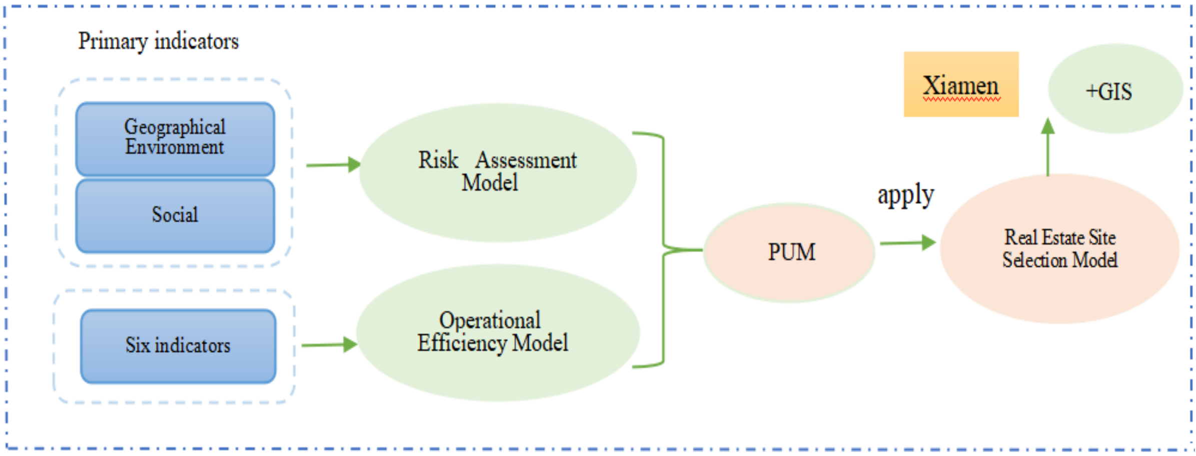

Figure 1 Our work

2. Real Estate Site Selection Model.

2.1. The structure of RAM

As shown in Figure **2, we** consulted the literature and selected 6 representative secondary indicators to construct our model,and nine representative countries as objects.After that, we applied the TOPSIS method to these nine countries.

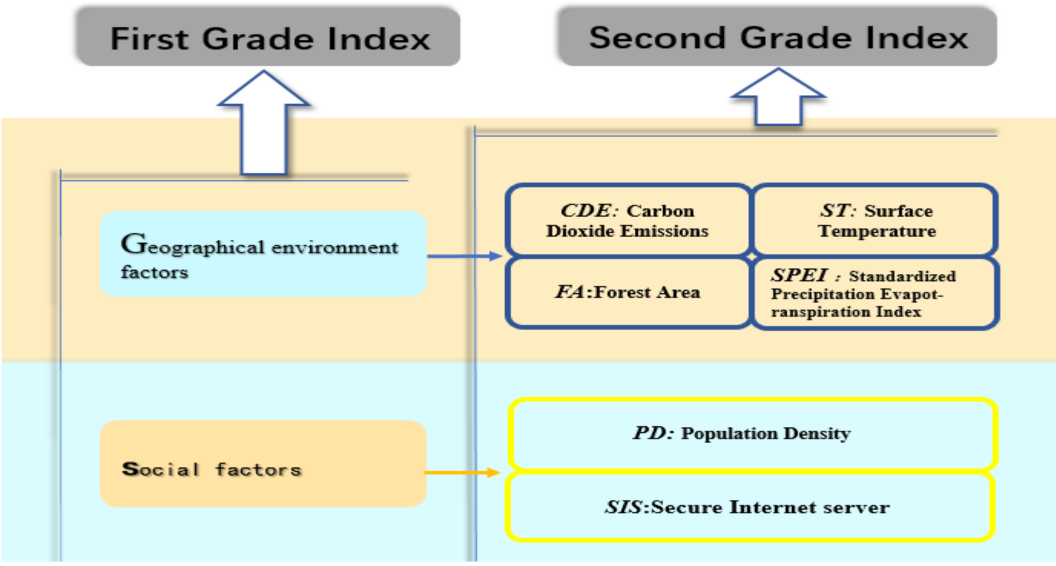

Figure 2 Indicators Selected

The key mathematical notations used in this model are listed in Table 1.

Table 1 Notations used in this model

Symbol	Description
D_i^-	the distance between z and the minimum value
D_i^+	the distance between z and the maximum value
S_i	non-normalized score
S_i'	normalized score
W_j	The weight of each indicator

Firstly, we need to normalize the original matrix in a forward manner. Transform the small-scale indicators into large-scale ones through the max-x method. The standard SPEI is an interval-type indicator, and we consider the range [-0.5, 0.5] as the normal interval. represents the indicator sequence of SPEI. We perform forward normalization on it:

M=max$\left\{-0.5-\min\{y_i\},\max\{y_i\}-0.5\right\}$

$$y_i' = \begin{cases} 1-\frac{0.5-y_i}{M}, y_i < -0.5 \\ 1, -0.5 \le y_i \le 0.5 \\ 1-\frac{y_i-0.5}{M}, y_i > 0.5 \end{cases} \tag{1}$$

Then,we standardize the results. The standardized matrix is denoted as Z, and each element in Z is:

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{9} x_{ij}^2}} (i = 1,2,\cdots,9; \mathrm{j} = 1,2,\cdots,6) \tag{2}$$

After processing, we obtain a non-negative matrix.

$$\mathrm{Z}=\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{16} \\ z_{21} & z_{22} & \cdots & z_{26} \\ \vdots & \vdots & \ddots & \vdots \\ z_{91} & z_{92} & \cdots & z_{96} \end{bmatrix} \tag{3}$$

Then, the non-normalized score for the i-th (i=1,2, ⋯,9) evaluation object can be calculated:

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{4}$$

Finally, we normalize the scores.:

$$S_i' = \frac{S_i}{\sum_{i=1}^{n} S_i} \tag{5}$$

Next, in order to assign weights to these 6 secondary indicators and obtain the risk assessment index, we use the entropy weight method to determine the weights of the 6 secondary indicators.

We calculate the probability matrix P, where the calculation formula for each element p_{ij} in P is as follows:

For the j-th indicator, calculate its information entropy using

$$p_{ij} = \frac{z_{ij}}{\sum_{i=1}^{9} z_{ij}} \tag{6}$$

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^{n} p_{ij} \ln(p_{ij}) (\mathrm{j}=1,2,\cdots,6) \tag{7}$$

Then define the information utility value as $d_j = 1 - e_j$. Finally, normalize the information utility values to obtain the entropy weight for each indicator:

$$W_j = \frac{d_j}{\sum_{i=1}^{6} d_j} \ (j = 1,2,\cdots,6) \tag{8}$$

Finally, we normalize the scores, and obtain the scores for each country, sort from large to small: 1.0000, 0.8389, 0.5606, 0.4366, 0.4128, 0.4097, 0.2833, 0.1783, 0.0000.

The weights for the 6 secondary indicators were determined using the entropy weight method, and the final results are as follows:Secure Internet server:0.2163; Carbon Dioxide missions:0.2679; Population density:0.0649; SPEI:0.0999; Surface Temperature:0.1287; Forest Area:0.2222.

2.2. The structure of OEM

By reading a large volume of literature, we chose to use the number of employees, capital investment, and operating expenses as input variables, while premium income, investment income, and claims expenses are selected as output variables.

To avoid the occurrence of multicollinearity in the model, we use cross-sectional data for modeling, with recent years' data serving as the prediction verification for the model. Then, the Analytic Hierarchy Process(AHP) is applied to measure the absolute operational efficiency of a company, enabling our model to be applied to any company. According to existing research, we use the Analytic Hierarchy Process (AHP) to establish a judgment matrix.As shown in Table 2.

Table 2 Operational Efficiency Model Judgment Matrix

	Premium Income	Claims Expenses	Capital Investment	Operating Expenses	Workforce Size	Investment Income
Premium Income	1	3	5	7	9	9
Claims Expenses	1/3	1	3	5	7	7
Capital Investment	1/5	1/3	1	3	5	5
Operating Expenses	1/5	1/3	1	3	3	3
Workforce Size	1/7	1/5	1/3	1	1	1
Investment Income	1/7	1/5	1/3	1	1	1

The consistency ratio for the judgment matrix is 0.0448, indicating an acceptable level of consistency. Moving forward, we compute the weights using the arithmetic mean method. Initial steps involve normalizing the matrix by columns and summing the normalized values for each column. Subsequently, each element in the resulting vector is divided by 'n' to derive the weights.The obtained weights are as follows: Premium Income (0.475), Claim Expenses (0.257), Capital Investment (0.135), Operating Expenses (0.068), Workforce Size (0.033), and Investment Income (0.033). Notably, Premium Income, serving as the primary revenue source, holds the highest weight among them.

2.3. Portfolio Underwriting Model（PUM）based on MCDM and OSM

To determine the feasibility of insurance underwriting in a specific region, we propose the OSM (Optimization-Selection-Monitoring) model. It consists of three parts: object(OEM-RAM), strategy(risk assessment and evaluation of operating efficiency), and measure. The main principles and functions are as follows.

Table 3 OSM Schematic

Object	Assessing whether regions experiencing increasingly extreme weather can be underwritten.	
Strategy	**Risk Assessment:** Utilizing the Entropy Weightand TOPSIS methods to evaluate and rank the risklevels of regions.	**Operational Efficiency:** Measuring operational efficiency using specified input and output variables.
Measure	(1)Developing Composite Indicators: Creating composite indicators for risk assessment andoperational efficiency. Weighting and combining the sub-indicators for each domain to form asingle risk score and a single efficiency score. (2)Combining Composite Indicators with MCDM: Applying MCDM criteria to integrate risk andefficiency scores into a single composite indicator. Insurance underwriting conditions shouldmeet the following requirements: a region must satisfy specific thresholds (e.g., risk level belowa certain threshold, efficiency above a certain threshold) to be considered for insuranceunderwriting, and the composite indicator should exceed a certain value.	

As shown in table 3, to assess whether an area is insurable, we integrate the models created for the two questions above to formulate a comprehensive indicator. We name it the Underwriting Index.. Its calculation is as follows:

$$\text{Underwriting Index } = \alpha_1 * \text{Risk Index} + \alpha_2 * \text{Operational Efficiency Index} \tag{9}$$

α_1 and α_2 represent the weights assigned to the Risk Index and Operational Efficiency Index respectively. The sum of the weights assigned to the Risk Index divided by the total weight for both indices is denoted as α_1, and the sum of the weights assigned to the Operational Efficiency Index divided by the total weight for both indices is denoted as α_2. After calculation, $\alpha_1 = \alpha_2 = 0.5$.

After defining the composite index, we establish a Multiple Criteria Decision Making (MCDM) model and set up two underwriting conditions:

$$\text{Underwriting Index } = \begin{cases} \text{Risk Index} > x_1 \\ \text{Underwriting Index} > x_2 \end{cases} \tag{10}$$

x_1 represents the threshold for the Risk Index. First, we calculate the Risk Index for nine countries. Based on literature review, we determine the threshold for the Risk Index as the 75th percentile. Therefore, we calculate the 75th percentile for these nine data points, resulting in x_1=2491.83.

x_2 is the threshold for the Underwriting Index. We select 18 property insurance companies within China with relatively complete data, calculate the weighted Underwriting Index for each company and obtain 18 data points. The 66th percentile of these 18 data points serves as the threshold x_2, yielding x_2= 3991.75.

When the Risk Index and the Underwriting Index for a region are greater than a threshold respectively, it indicates lower risk in the region and strong underwriting capacity for property insurance companies. In such cases, property insurance companies can underwrite in that region.

3. Real Estate Site Selection Model

3.1. Real estate site selection factors

By reviewing literature, we have categorized the factors influencing real estate site selection into three main categories: environmental, social, and economic factors, which we refer to as the "triple bottom line". For environmental factors, we consider slope, aspect, and land type. For social factors, we consider bus station, medical institutions, schools, and fire agencies. Economic factors are represented by per capita GDP. As shown in Figure 3:

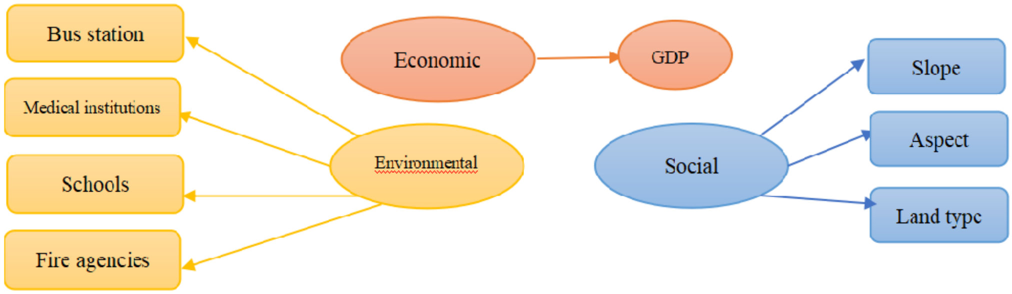

Figure 3 Triple bottom line

3.1.1 Environmental Factors

Geographic location is one of the most crucial factors. We usually consider whether there is bus station, medical institutions, schools, fire agencies around the residence. The choice of geographic location directly influences the market value and attractiveness of the project.

3.1.2 Economic Development

As the core of the national economic accounting system, GDP (Gross Domestic Product) is an important indicator reflecting the macroeconomic performance of a country [5] It often closely related to individuals' income levels. In regions with rapid GDP growth, people's purchasing power may correspondingly increase, driving the growth in demand for real estate. Additionally, the growth in GDP is usually accompanied by an increase in job opportunities, attracting population migration. In economically prosperous areas, people may be more willing to relocate, thereby increasing the demand for housing and commercial properties.

3.1.3 Social Factors

In areas with significant slope, soil erosion and stability can become issues, potentially leading to higher construction costs. Some steep slopes may be designated for green spaces or protected areas, while gentler slopes may be more suitable for residential or commercial purposes.

Land can be categorized into different types such as forests, grasslands, artificial surfaces, and bare ground. Among them, artificial surfaces and bare ground are suitable for building houses, while forests and grasslands are not suitable for building houses.

The aspect directly affects the lighting and ventilation of houses. Southern slopes generally receive more sunlight, which is advantageous for winter warmth, while northern slopes are relatively cooler and suitable for avoiding summer heat.

3.2. Improved PUM Combined with GIS

GIS (Geographic Information System) is a technology and tool used for capturing, storing, managing, analyzing, and displaying geographic spatial data. GIS integrates knowledge from various fields such as geography, cartography, remote sensing technology, database management, and computer science.It aims to process and analyze information related to geographical locations.

Regarding the application of GIS, we consulted literature and found that many scholars have previously researched GIS and combined it with other models and methods to assist in site selection. For example, Yang Li analyzed the role of GIS technology in risk identification for real estate investment projects. He established a commercial rent prediction model and studied the application points of a GIS-based real estate investment project risk identification system [6]. Li Heng utilized GIS combined with Data Envelopment Analysis (DEA) to objectively and intuitively determine the most suitable sites [7]. Oluwasola D. Taiwo employed a framework of Multi-Criteria Evaluation (MCE) analysis and integrated Analytic Hierarchy Process (AHP) into GIS to aid decision-making in site selection studies [8].

Therefore, leveraging the experience of predecessors, we integrate the Portfolio Underwriting Model (PUM) created in the previous question with GIS technology to assist in real estate site selection.

In real estate site selection, local economic factors are usually taken into consideration. Economic factors are one of the key influences on the real estate market, holding significant importance for developers and investors.And is a widely adopted macroeconomic indicator globally. In practice, Gross Domestic Product (GDP) is an important indicator for measuring the level of economic development in a country [9]. GDP is an indicator for assessing the overall economic performance of a country or region. Per capita GDP shows a strong linear relationship with insurance density, with insurance density increasing as per capita GDP rises. Regions with stronger economic capabilities tend to have relatively higher per capita incomes. And their level of insurance industry development far surpasses that of regions with comparatively lagging economies [10]. However, placing two indicators with a strong linear correlation into the same model may introduce multicollinearity issues. Multicollinearity will lead to unstable model coefficient estimates and difficulties in interpretation. Therefore, economic factors and insurance should not be included in the same model. Thus, when considering real estate site selection, the operational efficiency model should be excluded, and the risk assessment model should be combined with GIS.

We chose Xiamen as an example to illustrate the site selection process. Xiamen is a coastal city with diverse topography, including mountains, hills, and plains. The city experiences warm and humid summers, mild winters, and is prone to typhoon impacts. As the capital of Fujian Province, Xiamen has a well-developed economy with a relatively diverse industrial structure, including sectors such as electronic information and manufacturing. Xiamen is also known for its active foreign trade and efforts in promoting technological innovation.

In GIS and map services, POI (Points of Interest) refers to specific locations or places marked on a map that attract people's interest. Using POI technology, we identified the locations of bus stations, medical institutions, schools, and fire agencies in Xiamen on the map.They are shown in Figures 4, 5, 6 and 7 respectively.

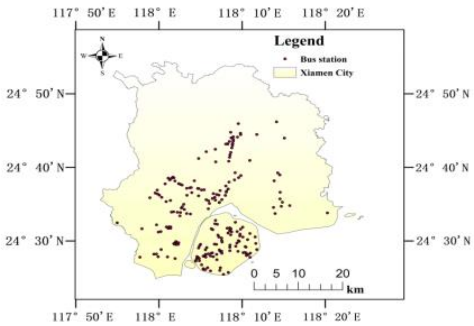

Figure 4 Bus station

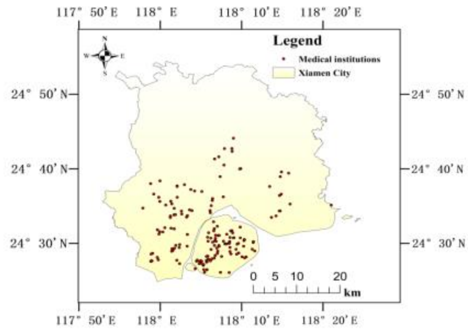

Figure 5 Medical institutions

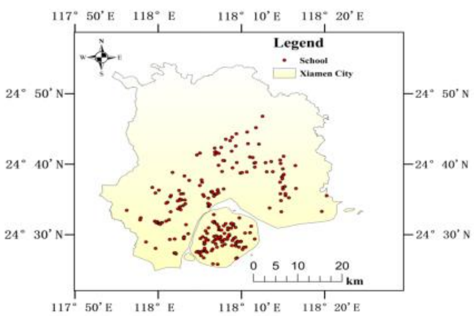

Figure 6 Schools

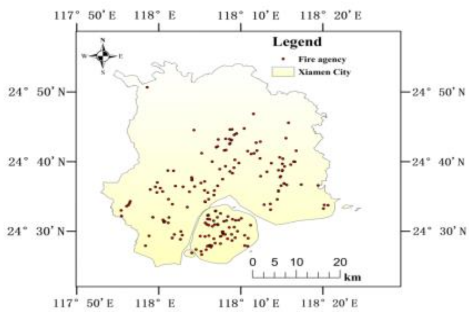

Figure 7 Fire agencies

We also utilized GIS to categorize different regions based on slope, aspect, land type and per capita GDP to determine which areas are suitable for residential purposes. They are shown in Figures 8, 9, 10 and 11 respectively.

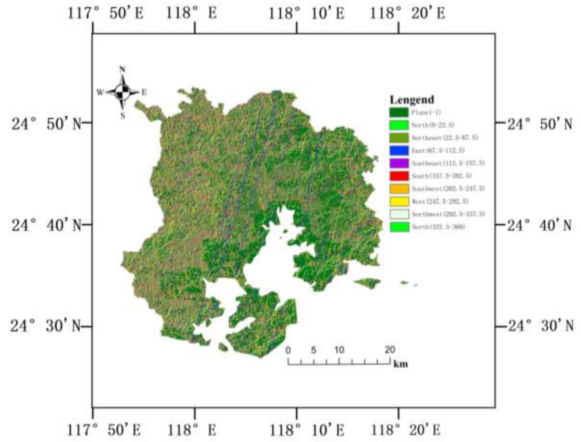

Figure 8 Slope

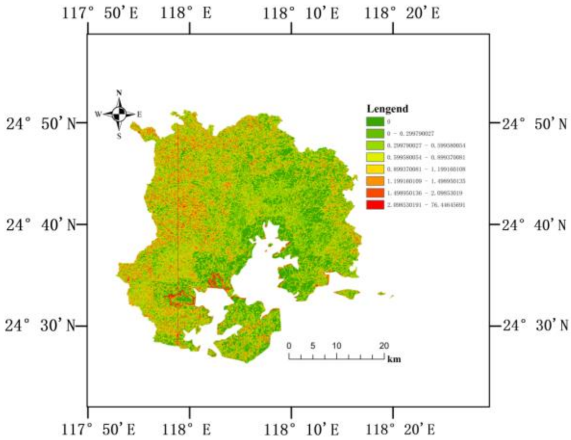

Figure 9 Aspect

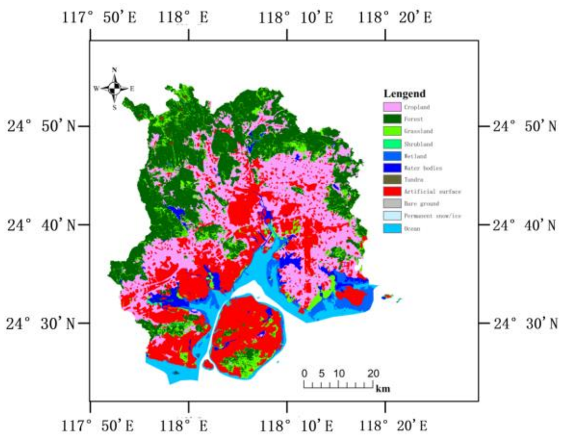

Figure 10 Land type

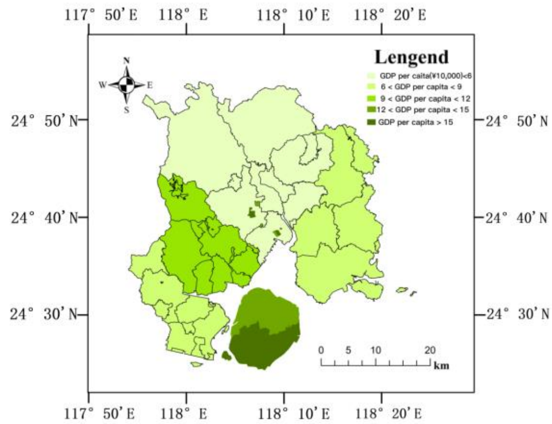

Figure 11 Percapita GDP

Finally, combining the eight graphs mentioned above, we calculated the Residential Suitability Index for each region using GIS, as shown in Figure 12:

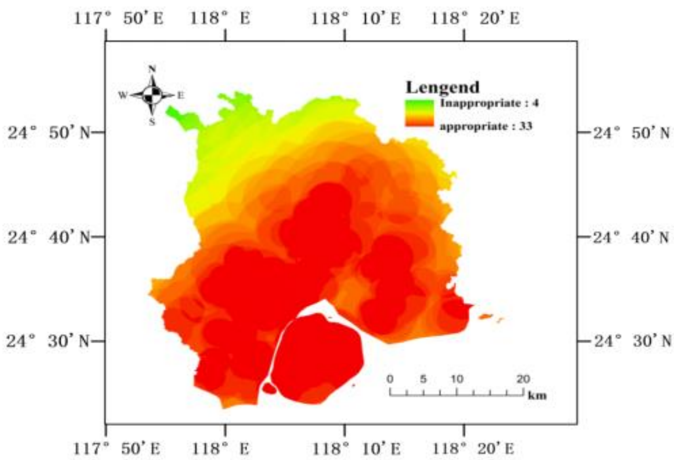

Figure 12 Residential suitability index

Applying the same methodology above to process the data for Xiamen City yields a risk assessment index of 3216.27. This index surpasses the risk threshold, leading to the classification of the region as a low-risk area. This designation supports informed site selection judgments when integrated with the real estate GIS site selection system.

4. Conclusions

Given the background of frequent extreme weather events, selecting a suitable residential area is a top priority for real estate enterprises. This paper categorizes the factors of real estate site selection into three categories, namely the "triple bottom line". And we combine our established PUM model with GIS technology to construct a real estate site selection model. Taking Xiamen City as an example, it assists the real estate market in selecting suitable residential addresses. The results indicate that combining the PUM model with GIS technology provides strong support for real estate site selection, helping developers to scientifically and effectively determine the best site selection scheme.

References

[1] Qi Fang. Research Finds: New Characteristics Emerge in Global Extreme Weather and Climate Events. Guangming Daily, 2024-04-18(008). DOI: 10.28273/n.cnki.ngmrb.2024.001716.

[2] Liu Xingyu. A Study on Site Selection of Residential Real Estate Projects for Enterprise A [D]. Xinjiang University, 2022. DOI: 10.27429/d.cnki.gxjdu.2021.000851.

[3] Zhang Qi, Xie Peng. A Study on Site Selection of Real Estate Development Projects in Jinan City. Residential and Real Estate, 2020, (05): 9.

[4] Shao Weishuang, Li Xiaohong, Zhang Tianshu, et al. Application of Data Mining Method Based on Rough Set Theory in Real Estate Site Selection Decisions. Mathematics in Practice and Understanding, 2019, 49(19): 97-103.

[5] Wang Shuping, Lu Zihan, Ji Chengxiu. Research on GDP Composite Forecasting Based on Web Search Data. Heilongjiang Science, 2024, 15(08): 44-48.

[6] Yang Li. Application Research on Real Estate Investment Project Risk Identification System based on GIS Technology[C]. Shenzhen. China Overseas Property Limited. 2023:1-4.

[7] Li, Heng; Yu, Ling; Cheng, Eddie W. L. (2005). A GIS-based site selection system for real estate projects. Construction Innovation: Information, Process, Management, 5(4), 231–241.

[8] Oluwasola D. Taiwo. GIS-MCE based suitability analysis for sustainable estate development in Ede North LGA Osun State, Nigeria. [J]. Journal of Geoinformatics and Environmental Research,2022,3(2):58-73

[9] Chen Manli, Zhang Huijuan, Jiao Nannan, et al. Research on Forecasting China's GDP Based on ARIMA Model using MATLAB. Chinese Market, 2024(12): 1-4. DOI: 10.13939/j.cnki.zgsc.2024.12.001.

[10] Li Zhongxian. Regional Disparities in the Development of China's Insurance Industry [N]. ChinaInsurance News, 2014-02-20.