Integrating Natural and Social Indicators for Enhanced Insurance Underwriting Strategies

Pengcheng Shen *

College of Sciences, China Jiliang University, Hangzhou, China, 310018

* Corresponding Author Email: pcshen@cjlu.edu.cn

Abstract. Facing extreme weather's growing impact, this study introduces a novel decision-making framework for the insurance sector. We use the Entropy Technique for Order of Preference by Similarity to the Ideal Solution (EW-TOPSIS) Model and Logistic Regression to assess underwriting risks and probabilities, underpinned by a Multi-Objective Planning Model that aims to optimize revenue and minimize churn. Case studies in Kumamoto Prefecture, Japan, and Texas, USA—areas affected by earthquakes and hurricanes—demonstrate the model's application. In 2022, it advised against underwriting in Kumamoto with profits of $1.325 billion and recommended a cautious strategy in Texas, yielding $3.544 billion. These results highlight the framework's utility in enhancing insurance strategies in extreme weather-prone areas for sustainability and profit.

Keywords: Extreme weather, Property insurance, Entropy weight-TOPSIS model, Logistic Regression.

1. Introduction

Against the backdrop of global climate change, the increasing frequency of extreme weather events not only has profound impacts on human societies but also presents unprecedented challenges to the sustainability of property insurance[1, 2]. The rise in natural disasters such as floods [3], hurricanes, and earthquakes escalates the risk and complexity of insurance payouts, necessitating the development of more sophisticated and scientific methods for risk assessment and management to ensure the healthy progression of the insurance sector and effective protection of properties [4].

In this work, we begin by employing the Entropy Technique for Order of Preference by Similarity to the Ideal Solution (EW-TOPSIS) Model to select indicators from both natural and social factors, thereby scoring the underwriting risk across different regions. Through Logistic Regression analysis, we further compute the probabilities of underwriting, setting the stage for a nuanced assessment of insurance risk. Subsequently, we construct a Multi-Objective Planning Model that incorporates constraints such as the underwriting probability and the annual interest rate of the insurance company's financial products. This model aims to optimize the insurance company's maximum revenue and minimum customer churn rate, thereby guiding strategic decision-making processes.

2. Data Collection and Processing

Firstly, focusing on the primary factors for insurance evaluation, we have collected various types of data from different sources, as shown in Table 1. These data are sourced from authoritative institutions or government databases, ensuring the scientific rigor and reliability of the information.

Table 1. Main Data Description and Source.

Data Description	Data Source
Data on agriculture, economy, earth sciences, and other fields	https://github.com/topics/awesome-public-datasets
Data on population, policies, economy, society, and other aspects	https://sedac.ciesin.columbia.edu/
Data on Remote sensing imagery, climate, ocean, and land	https://www.earthdata.nasa.gov/

By screening and examining the raw dataset, we identify outliers and interpolate missing values to improve the consistency and accuracy of the data. Specifically, we use the 3-sigma rule for outlier detection, which assumes that a dataset contains only random errors. We calculate the standard deviation of the raw data and then determine an interval based on a specified probability. Data points whose deviation from the mean exceeds this interval are considered outliers, as shown in equation (1).

$$P(x-\mu>3\sigma)\geq 0.003 \tag{1}$$

The data demonstrates a high level of integrity with only a minimal presence of null values. Any outliers have been filled and set to 0 to maintain consistency within the dataset.

After statistical enumeration and anomaly handling, the dataset may contain missing values that require attention. This paper utilizes the Lagrange interpolation method to impute missing data. The specific principles of this interpolation technique are detailed in Equation (2).

$$L(x_{defalt})=\sum_{j=0}^{kx} y_j l_j(x), l_j(x)=\prod_{i=0}^{k}\frac{x-x_i}{x_j-x_i} \tag{2}$$

Through the application of Lagrange interpolation, the dataset x_{defalt} has been effectively treated.

3. Comprehensive Risk Scoring for Insurance Companies

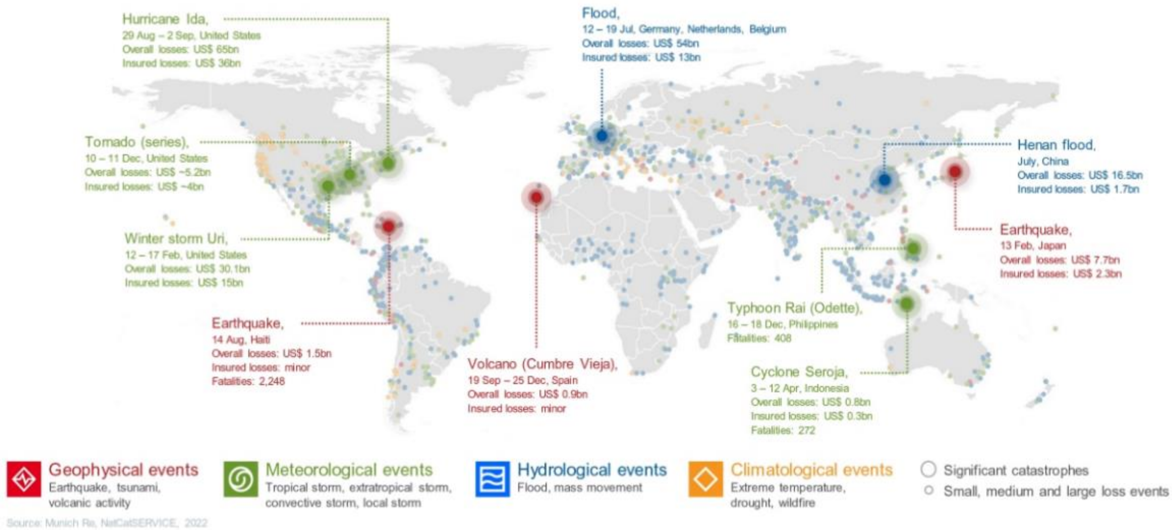

Figure 1. Relevant natural catastrophe loss events worldwide.

We have developed a comprehensive evaluation model to assess the underwriting risks for insurance companies. This model clarifies the impact factors of increasing extreme weather events in specific regions on the underwriting risks of insurance companies. It establishes an evaluation index system and calculates a comprehensive score.

Figure 1 depicts the regions where significant natural disasters have occurred. Taking Sichuan Province in China as a case study, we have collected the province's historical data for in-depth analysis. Using this data, we constructed a comprehensive evaluation model to ascertain the underwriting risks specific to this province. This provides insurance companies with targeted information, assisting them in understanding the potential risks associated with the region.

3.1. Construction of the Evaluation Index System

By the relevant conditions, we have identified natural and societal factors as the primary evaluation indices for the comprehensive assessment [5]. Drawing from the literature, we selected the frequency of extreme meteorological disasters, the intensity of such disasters, climate change trends, the susceptibility of geographical locations to disasters, and the historical claims records of natural disasters as secondary indices under natural factors. Population density, urbanization level, the competitive intensity of the insurance market, the policy and regulatory environment, and public awareness and demand for insurance are chosen as secondary indices under societal factors. The construction of the evaluation index system is illustrated in Figure 2:

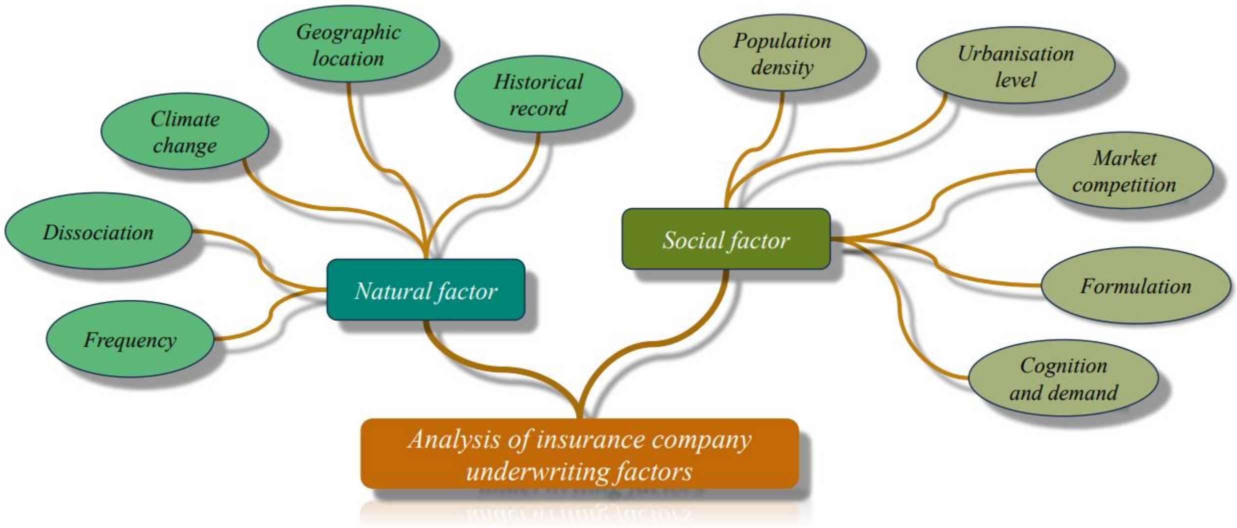

Figure 2. Evaluation Index System Diagram.

Next, we analyze the rationale behind the selection of the aforementioned indicators.

When evaluating natural factors, we consider the frequency (z_1) and intensity (z_2) of extreme weather events due to their direct link to increased claims and risks. The trend of climate change (z_3) is also analyzed for its future impact on weather severity. Additionally, the geographical susceptibility to disasters (z_4) and historical claim data from past natural disasters (z_5) provide insights into risk assessment.

Regarding social factors, we examine population density (z_6) and urbanization levels (z_7), reflecting potential risk exposure. The competitive landscape of the insurance market (z_8), influenced by premium pricing and profitability, the regulatory environment (z_9) affecting company operations, and public insurance awareness and demand (z_{10}) are also critical to understanding market dynamics and consumer behavior.

3.2. Weight Determination - Entropy Weight Method

Entropy is a measure of indicator dispersion. A smaller information entropy value indicates a greater degree of dispersion and, therefore, a higher weight for the indicator.

To calculate this, we first standardize the data for Sichuan Province, China, and the 10 evaluation indicators using the following formula, taking into account the variation in magnitude and dimensions across the indicators [6]:

Normalization is then performed to derive the standardized matrix Z_{ij}:

$$Z_{ij} = \frac{z_{ij} - \max(z_{kj})}{\max(z_{kj}) - \min(z_{kj})} \tag{3}$$

Where j=1,2,…,10 represents the 10 indicators, k=1,2,…,128 and i=1,2,…,21 represents the 21 cities in Sichuan Province.

Next, calculate the proportion p_{ij} of the i^{th} city under the j^{th} indicator, treating it as the probability used in relative entropy calculations to form the probability matrix P, as seen in Equation (5):

$$P_{ij}=\frac{z_{ij}}{\sum_{i=1}^{128}z_{ij}}(j=1,2,...,10) \tag{4}$$

Compute the information entropy e_j for the j^{th} indicator:

$$e_j=-\frac{1}{\ln 128}\sum_{i=1}^{128}P_{ij}\ln(p_{ij})(j=1,2,...,10) \tag{5}$$

Calculate the information utility value d_j and normalize to obtain the entropy weight W_j:

$$\begin{cases} d_j=1-e_j \\ W_j=\dfrac{d_j}{\sum_{j=1}^{10}d_j} \end{cases} \tag{6}$$

The entropy weight method has been applied to determine the weights for the 10 indicators, as detailed in Table 2:

Table 2. Weight coefficient table for evaluation indicators.

Ind.	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
W	0.1	0.05	0.15	0.12	0.08	0.09	0.11	0.13	0.07	0.1

3.3. Comprehensive Evaluation Score - TOPSIS Method

The TOPSIS algorithm is a widely used decision analysis method for multi-objective scenarios with a finite set of alternatives. Its core principle involves calculating the comprehensive evaluation scores of various alternatives and then ranking them based on these scores [7]. The steps of the method are as follows:

First, the data for the 21 cities in Sichuan Province and the 10 evaluation indicators are normalized to produce a standardized matrix Z. The extremum values of this standardized matrix are then defined.

The maximum $Z+$ and minimum $Z-$ values are defined, as shown in Equation (7):

$$\begin{cases} Z^+=(Z_1^+,Z_2^+,...,Z_j^+)=(\max\{z_{11},...,z_{n1}\},...,\max\{z_{1j},...,z_{nj}\}) \\ Z^-=(Z_1^-,Z_2^-,...,Z_j^-)=(\min\{z_{11},...,z_{n1}\},...,\min\{z_{1j},...,z_{nj}\}) \end{cases} \tag{7}$$

Calculate the distance of each evaluation object from the positive ideal solution (D_i^+) and the negative ideal solution (D_i^-), as shown in Equation (8):

$$\begin{cases} D_i^+=\sqrt{\sum_{j=1}^{10}W_j(Z_j^+-z_{ij})} \\ D_i^-=\sqrt{\sum_{j=1}^{10}W_j(Z_j^--z_{ij})} \end{cases}(i=1,2,...,128) \tag{8}$$

Calculate the comprehensive score C_i for the underwriting risk of the insurance company in the i^{th} city, as shown in Equation (9):

$$C_i=\frac{D_i^-}{D_i^++D_i^-} \tag{9}$$

4. Underwriting Strategy Based on Multi-Objective Programming

The underwriting strategy of an insurance company needs to address two primary objectives: the company's revenue and the customer attrition rate [8]. Insurance companies often use the underwriting probability in various regions to calculate expected revenue and make underwriting decisions [9]. The underwriting risk to some extent reflects the probability of the insurance company underwriting in a certain region. Therefore, we use the quantified underwriting risk to determine the underwriting probability for a particular region. As the underwriting probability for a region is a discrete variable, Logistic regression, a commonly used classification algorithm that provides the probability of classification, is adopted.

4.1. Determining Insurance Underwriting Probability Using Logistic Regression

Logistic regression is a generalized linear regression model. Let x_1 represent the underwriting risk of the insurance company. The logistic regression equation is derived as shown in Equations (10) and (11):

$$\log it(P) = \beta_0 + \beta_1 x_1 \tag{10}$$

$$\hat{y} = logit\left(P\right) = \ln \frac{p}{1-p} \tag{11}$$

The equation for calculating the underwriting probability p is obtained as follows:

$$p = \frac{\exp(\beta_0 + \beta_1 x_1)}{1 + \exp(\beta_0 + \beta_1 x_1)} \tag{12}$$

Utilizing the determined underwriting probabilities and data such as the historical insurance participation in the area, we can establish a multi-objective programming model. This model aims to maximize the insurance company's profits over the years and minimize the customer churn rate.

4.2. Development of a Multi-objective Programming Model

Given the extensive impact of extreme weather events, we assume that the underwriting probability of insured individuals in a city aligns with the underwriting probability of the city itself.

We categorize the insurance premiums paid by the insured into three tiers based on the amount.

We posit that the insurance company invests the collected premiums in financial products, deriving substantial returns by opting for financial products with suitable annual interest rates. Additionally, the annual interest rates of financial products are associated with specific ranges depending on the tier of the insurance premium.

Furthermore, we assume that if the insurance company decides to underwrite a policy, it will refund an amount 10 times the insurance premium to the policyholder.

(1) Decision Variable

$r_{ij}^{(k)}$ Is a binary variable.

When $r_{ij}^{(k)} = 0$ it represents that the annual interest rate of the financial product for the insurance premium at tier k does not correspond to the j^{th} annual interest rate.

When $r_{ij}^{(k)} = 1$ it signifies that the annual interest rate of the financial product for the insurance premium at tier k corresponds to the j^{th} annual interest rate.

(2) Objective Functions:

a). Maximize Profit:

$$\max z_1 = \sum_{k=1}^{3}\sum_{i=1}^{n^{(k)}}\left[\left(1 - p_i^{(k)}\right)x_i^{(k)}\sum_{j=1}^{30}(1 + r_{ij}^{(k)}c_j) - 10p_i^{(k)}x_i^{(k)}\right] \tag{13}$$

Where $p_i^{(k)}$ denotes the underwriting probability of the i^{th} policyholder at tier k;

$x_i^{(k)}$ Represents the insurance premium of the i^{th} policyholder at tier k;

c_j Denotes the annual interest rate of the j^{th} financial product.

b) Minimize Customer Churn Rate:

$$\min z_2 = \frac{\sum_{k=1}^{3}\sum_{i=1}^{n^{(k)}}\sum_{j=1}^{30}(r_{ij}^{(k)} l_j)}{N} \tag{14}$$

(3) Constraints

a). Constraint on Annual Interest Rates of Financial Products: Ensuring that each insurance premium has a corresponding annual interest rate for the financial product.

$$\sum_{i=1}^{30} r_{ij}^{(k)} = 1 (i = 1, 2, ..., n^{(k)})(k = 1, 2, 3) \tag{15}$$

b). Range Constraint on Annual Interest Rates of Financial Products: Ensuring that the insurance premiums of each tier have a corresponding range of annual interest rates for the financial products.

$$\begin{cases} r_{ij}^{(1)} = 0, (j \geq 10)(i = 1, 2, ..., n^{(1)}) \\ r_{ij}^{(2)} = 0, (10 \leq j \leq 20)(i = 1, 2, ..., n^{(2)}) \\ r_{ij}^{(3)} = 0, (j \leq 20)(i = 1, 2, ..., n^{(3)}) \end{cases} \tag{16}$$

4.3. Solving Model Using the Primary Objective Method

We designate the maximization of the insurance company's profit as the primary objective and the minimization of customer churn rate as the secondary objective. Appropriate boundary values are determined as shown in Equation (17):

$$\begin{aligned} &\max z_1 = \sum_{k=1}^{3}\sum_{i=1}^{n^{(k)}}\left[\left(1-p_i^{(k)}\right)x_i^{(k)}\sum_{j=1}^{30}(1+r_{ij}^{(k)}c_j) - 10p_i^{(k)}x_i^{(k)}\right] \\ &s.t. \begin{cases} \dfrac{\sum_{k=1}^{3}\sum_{i=1}^{n^{(k)}}\sum_{j=1}^{30}(r_{ij}^{(k)} l_j)}{N} < 0.1 \\ \sum_{i=1}^{30} r_{ij}^{(k)} = 1(i = 1, 2, ..., n^{(k)})(k = 1, 2, 3) \\ r_{ij}^{(1)} = 0, (j \geq 10)(i = 1, 2, ..., n^{(1)}) \\ r_{ij}^{(2)} = 0, (10 \leq j \leq 20)(i = 1, 2, ..., n^{(2)}) \\ r_{ij}^{(3)} = 0, (j \leq 20)(i = 1, 2, ..., n^{(3)}) \end{cases} \end{aligned} \tag{17}$$

4.4. Resilience and the Impact of Property Owners on Underwriting Strategies

(1) Model Resilience: To adapt to future claims, the model uses historical data to guide underwriting decisions: no underwriting below \$2 billion, cautious underwriting between \$2 and \$4 billion, and active underwriting above \$4 billion, ensuring financial readiness for extreme weather losses [10].

(2) Underwriting Criteria: Based on past performance, the insurance company avoids underwriting if predicted profits fall below \$2 billion. It considers underwriting for profits between \$2 billion and \$4 billion and commits to underwriting when profits exceed \$4 billion, balancing risk and potential gains.

(3) Influence of Property Owners: Property owners impact the company's strategy through their choice of high-value policies. The company evaluates past claims and property risk factors, favoring areas with lower claim histories and properties with reduced risks, guiding underwriting decisions toward safer investments.

4.5. Application and Demonstration of the Model

Leveraging the aforementioned models, we utilize the comprehensive evaluation model and the planning model to analyze the underwriting strategies for two regions: Kumamoto Prefecture in Japan and Texas in the United States. Figure 3 provides a clear and intuitive depiction of the general geographical locations and topographic features of these two areas.

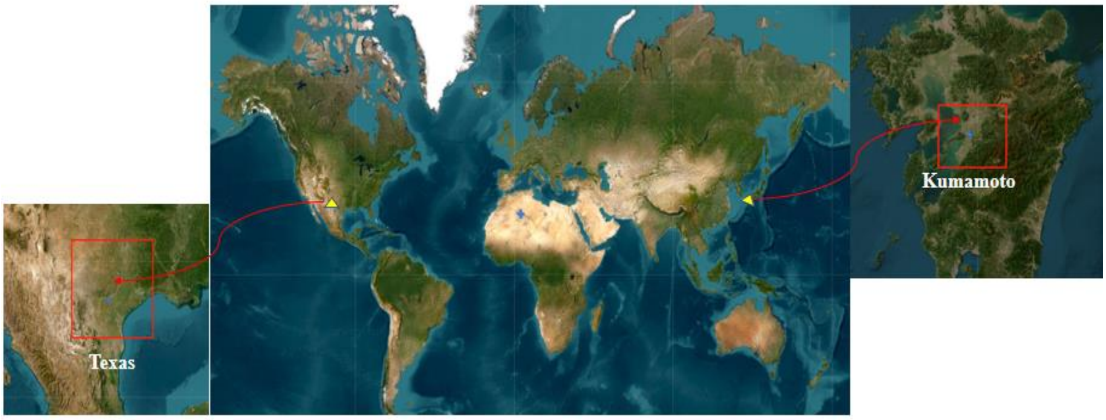

Figure 3. World Satellite Map and Location of Kumamoto and Texas.

Figure 3 highlights Kumamoto Prefecture and Texas, areas prone to earthquakes and hurricanes, respectively. The decision-making model, using 2022 data, suggests refraining from underwriting in Kumamoto due to a maximum profit of $1.325 billion, below the $2 billion threshold. In contrast, Texas, with a $3.544 billion profit, might see underwriting within acceptable risk levels. This approach underscores the model's utility in guiding underwriting strategies based on regional risk profiles and potential profitability, enhancing the sustainability of property insurance in extreme weather-prone areas.

5. Conclusions

In this study, we develop a decision-making framework for the insurance industry, targeting areas prone to extreme weather, by combining the Entropy Technique for Order of Preference by Similarity to Ideal Solution (EW-TOPSIS) Models with Logistic Regression to assess underwriting risks from natural and social factors. This integrated model calculates underwriting probabilities to guide a Multi-Objective Planning Model, optimizing for revenue maximization and customer retention. Applied to case studies in Kumamoto Prefecture and Texas, which face earthquakes and hurricanes respectively, the model suggests not underwriting in Kumamoto where expected profits fall below $2 billion, while recognizing Texas as a profitable yet risky market with potential earnings of $3.544 billion, demonstrating the model's ability to provide strategic insights into underwriting decisions in regions affected by severe weather conditions.

Herein, this study effectively guides insurance underwriting amid climate challenges by leveraging scientific analysis and emphasizing the role of geographic and historical data in risk assessment. Its versatility across regions underscores its global applicability and potential to bolster the resilience and financial stability of the insurance industry.

References

[1] NEWMAN R, NOY I. The global costs of extreme weather that are attributable to climate change [J]. Nature Communications, 2023, 14(1): 6103.

[2] ROBINSON A, LEHMANN J, BARRIOPEDRO D, et al. Increasing heat and rainfall extremes are now far outside the historical climate[J]. npj Climate and Atmospheric Science, 2021,4(1): 45.

[3] MERZ B, BLÖSCHL G, VOROGUSHYN S, et al. Causes, impacts, and patterns of disastrous river floods [J]. Nature Reviews Earth & Environment, 2021, 2(9): 592-609.

[4] KRICHEN M, ABDALZAHER M S, ELWEKEIL M, et al. Managing natural disasters: An analysis of technological advancements, opportunities, and challenges [J]. Internet of Things and Cyber-Physical Systems, 2024, 4: 99-109.

[5] MAJ M, van OS J, De HERT M, et al. The clinical characterization of the patient with primary psychosis aimed at personalization of management [J]. World Psychiatry, 2021, 20(1): 4-33.

[6] S. S, A. S, A. J, et al. Class-Agnostic Weighted Normalization of Staining in Histopathology Images Using a Spatially Constrained Mixture Model [J]. IEEE Transactions on Medical Imaging, 2020, 39(11): 3355-3366.

[7] SIREGAR V M M, SONANG S, PURBA A T, et al. Implementation of TOPSIS Algorithm for Selection of Prominent Student Class [J]. Journal of Physics: Conference Series, 2021, 1783(1): 12038.

[8] SMITH K A, WILLIS R J, BROOKS M. An analysis of customer retention and insurance claim patterns using data mining: a case study [J]. Journal of the Operational Research Society, 2000, 51(5): 532-541.

[9] TAYLOR G C. Underwriting strategy in a competitive insurance environment [J]. Insurance: Mathematics and Economics, 1986, 5(1): 59-77.

[10] SHEEHAN B, MURPHY F, KIA A N, et al. A quantitative bow-tie cyber risk classification and assessment framework [J]. Journal of Risk Research, 2021, 24(12): 1619-1638.