Spatio-temporal Analysis and Prediction of Convenience Store in Taipei

Thomas Chuang

April 2023

Abstract

This study aims to analyze areas with potential for convenience store expansion, focusing on Taipei City, which currently has the highest density of convenience stores. First, through a literature review, factors affecting convenience store location selection were identified, including store characteristics, demographic variables, and transportation factors. These were used as reference variables, and data from 2012 to 2019 were used to map the results. Based on the spatial autocorrelation of these factors, further filtering selected road nodes, population density, bus stops, metro stations, and public bike stations as independent variables, with convenience store locations (from 1997 to 2019) serving as the dependent variable.

According to previous studies, the grid was selected as the basic spatial unit for this study. Considering the dense development of convenience stores in Taipei, the grid size was adjusted to 200 meters to better reflect regional differences. Various models were employed to obtain predictions, including Ordinary Least Squares (OLS), suitability analysis, Geographically Weighted Regression (GWR), and the Space-Time Cube. The results were then validated using data from convenience store openings between 2019 and 2023, comparing the performance of different models. After recalibrating and adjusting the models, it was determined that the GWR model had the best explanatory power and was adopted as the final predictive model. Finally, this study identified areas in Taipei with potential for convenience store development, primarily in commercial zones. It also recognized the limitation that the availability of quantitative data remains a challenge and that some factors were not fully considered.

Chapter 1

Introduction

1.1 Research Motivation

From selling food and daily necessities to services like bill payments, ticket purchases, and parcel delivery, convenience stores have become a multifaceted and highly accessible medium catering to various aspects of daily life, including eating, living, transportation, and entertainment. In Taipei, where we attend school, the density of convenience stores ranks the highest in Taiwan (6.52 stores per square kilometer, according to the government's open data platform for the five major convenience store chains). This high density raises curiosity about how to find optimal locations for future store expansion and what factors should be considered, thus sparking this study.

1.2 Research Objectives

This study focuses on analyzing convenience store brands, including 7-11, FamilyMart, OK Mart, Hi-Life, and PX Mart. Through a literature review, key factors influencing location selection are identified. Using data from Taipei City, the study evaluates the impact of these factors to build a predictive model, and the results are used to explore location strategies for convenience stores in Taipei.

1.3 Research Scope

This research is centered on Taipei City, mainly because while this area has the highest density of convenience stores in Taiwan, it still appears to have potential for further development. Therefore, the aim is to uncover optimal locations for convenience store development in the region.

Chapter 2

Literature Review

2.1 Development of Convenience Stores

According to the Ministry of Economic Affairs' Department of Commerce, a convenience store is defined as "an industry providing convenience goods, such as fast food, beverages, daily necessities, and service-related products, to meet customers' immediate needs, excluding fresh food." Convenience stores are generally believed to have originated in the United States during the 1920s. In Taiwan, the first convenience store was introduced by Uni-President Enterprises Corporation, initially positioned as a small supermarket. However, due to underwhelming business performance, the marketing focus shifted to office workers and students (Chou, 2009). The product offerings became more diverse, and the stores' operational performance improved, attracting brands like FamilyMart, OK Mart, and Hi-Life to open locations rapidly. In recent years, PX Mart has oriented its services towards convenience, leading it to be categorized as one of the five major convenience store chains by the government. As of April 2023, there are 13,139 convenience stores across Taiwan, and the number continues to grow.

2.2 Factors Influencing Convenience Store Location

Based on the study by Wu (2017) on the location factors and decision-making systems for convenience stores, various factors from previous research were consolidated, analyzed, and categorized into investment returns, commercial district factors, store factors, demographic variables, competition, community development, and transportation factors. The following are geographic-related factors:

2.2.1 Store Factors

These include store visibility, rental costs, and obstacles. Among stores with similar location conditions, those located at "corner" spots (triangular corner sites) have the greatest influence on customer footfall (Li, 1991).

2.2.2 Demographic Variables

Key variables include population size, population density, population structure, population migration rate within the commercial district, and average consumer income and expenditure. These factors help assess the area's consumer potential and are commonly considered when selecting store locations.

2.2.3 Transportation Factors

These factors include proximity to train stations, bus stops, metro stations, as well as vehicle and pedestrian traffic flow. It is also noted that being located in the direction of pedestrian traffic flow is advantageous for increasing customer visits.

2.3 Research on Convenience Store Locations

2.3.1 Grid Data Analysis

For large-scale spatial analysis, while point data is highly accurate, it is primarily useful for visualizing clustering trends, identifying commonalities and anomalies, and lacks the ability to provide a detailed quantitative analysis or depict regional differences. This makes the analysis more complex and costly. Therefore, using polygon data as the basic spatial unit is more suitable for studying the spatiotemporal differences in convenience store development. To avoid the scale effect and zonation effect, which may distort statistical results, raster data with fixed, equally-sized boundaries is used for analysis (Su, 2019). The grid size is often determined by the service radius of a single convenience store, which previous research suggests is approximately 250 meters (Rong, 2005; Mei et al., 2009).

2.3.2 Spatial Distribution Trend Analysis

In Su's (2017) research on the spatial distribution trends of convenience stores, spatial autocorrelation analysis was employed. The global Moran's I index was used to observe the spatial clustering of convenience stores, while the local Getis-Ord Gi* statistic was applied to examine the temporal evolution and hot spot locations of new store openings. When analyzing the spatial distribution trends of geographic phenomena, spatial statistical methods are often used to construct predictive models. For instance, suitability analysis assigns weights to specific factors to determine the optimal locations for a geographic phenomenon in a given area, a method frequently used in ecological habitat analysis (Cheng et al., 2005).

Additionally, the Multi-scale Geographically Weighted Regression (MGWR) model offers greater explanatory power compared to the standard Geographically Weighted Regression (GWR) by considering the spatial scale variations between different variables. MGWR has been successfully applied to the analysis of public bike station locations (Zhong, 2022) and the spatial distribution of diseases (Mansour et al., 2021). Another method, the space-time cube (STC), is used to visualize spatial changes over time and has been utilized to model urban air pollution dynamics (Fang & Lu, 2011).

Chapter 3

Workflow and Methodology

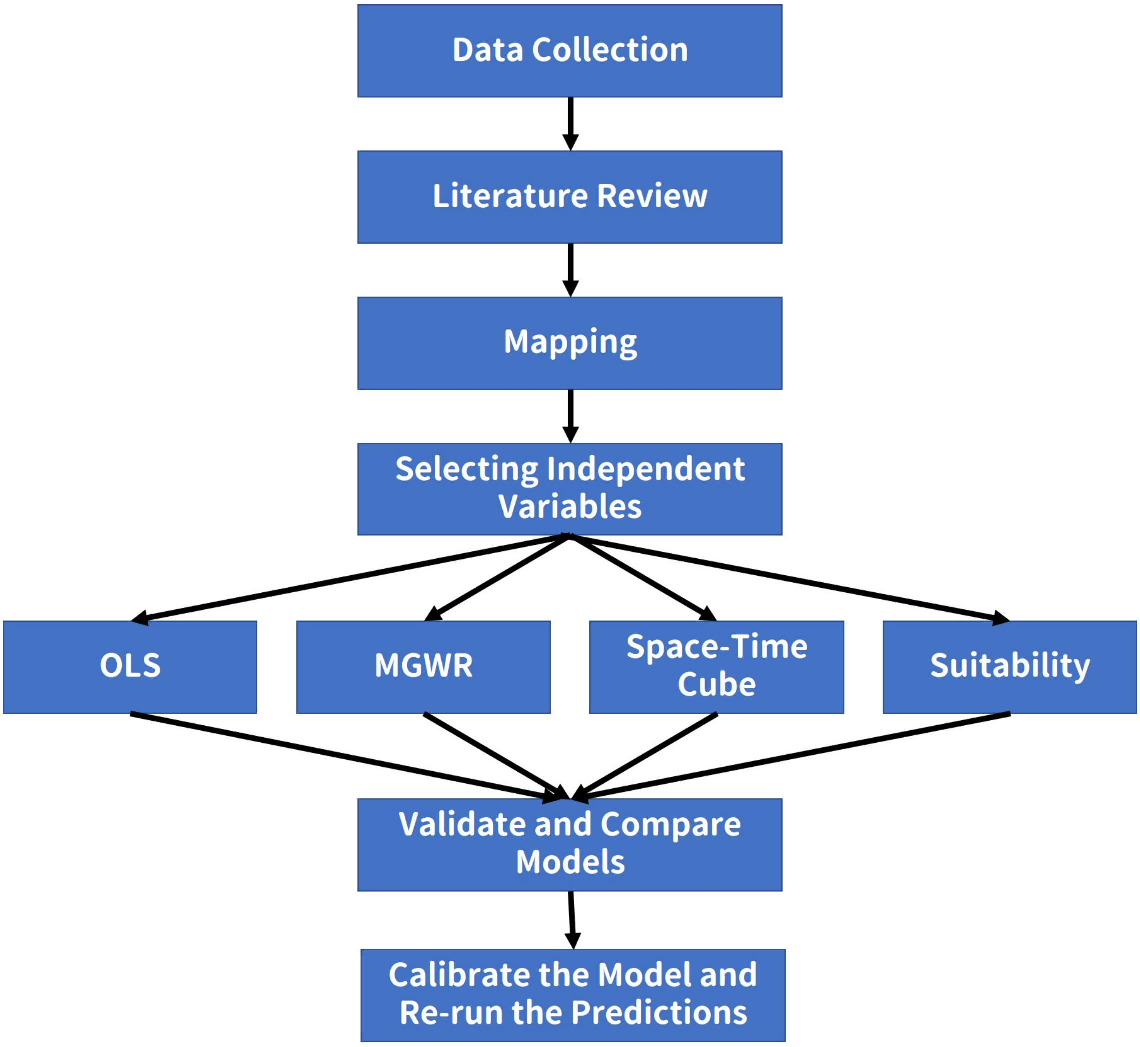

Figure 3.1: Research Workflow Diagram

3.1 Research Methodology

3.1.1 Literature Analysis

The literature analysis method refers to "the collection and comparative analysis of relevant data from journals, articles, books, dissertations, research reports, government publications, and newspaper reports related to a specific problem. This method aims to understand the possible causes of the problem, the resolution process, and potential outcomes" (Wu, 2003). This method is used in this study to identify the location factors and their respective weights.

3.1.2 Suitability Analysis

Suitability analysis is used to identify optimal locations. In this study, various independent variables were collected and scored. Variables with greater potential for convenience store development were assigned higher scores. These scores were then weighted according to the influence of each variable, and the total scores were summed. Areas with higher scores were identified as having greater potential for convenience store development (Wang & Yu, 2009).

3.1.3 Ordinary Least Squares Regression Model (OLS)

Linear regression is a statistical model used to analyze the relationship between independent and dependent variables. Through the ordinary least squares method, the relationship between each independent variable and the dependent variable is calculated to explain its influence. In this study, the OLS model is applied to select independent variables in the first stage (Zhong, 2022).

3.1.4 Multi-scale Geographically Weighted Regression (MGWR)

To address the limitation of OLS in representing the First Law of Geography: "Everything is related to everything else, but near things are more related than distant things," the MGWR model incorporates the spatial heterogeneity of geographic data into the prediction. Unlike classical Geographically Weighted Regression (GWR), which uses a single bandwidth for all variables, MGWR selects an optimal bandwidth for each variable, resulting in a better-fitting prediction model (Mansour et al., 2021).

3.1.5 Space-time Cube (STC)

The space-time cube model represents spatial locations in two dimensions, with the third dimension showing the temporal changes in geographic data. This model is useful for observing the trends of convenience store establishment over time and space. Additionally, the Getis-Ord Gi* statistic is employed to analyze spatial autocorrelation and identify cold spots and hot spots for convenience store locations (Zhu et al., 2019).

3.2 Research Design

After confirming the research topic, a literature review was conducted to select the relevant factors and research methods. Due to the difficulty in obtaining quantitative data, only store factors, demographic variables, and transportation factors were considered as independent variables, while the locations of convenience stores were used as the dependent variable (Table 3.1). Spatial data from 2012 to 2019 was used to create maps, and spatial autocorrelation was analyzed to select the independent variables. Considering the dense development of convenience stores in Taipei, the grid size for analysis was adjusted to 200 meters to better reflect regional differences. Ordinary least squares, suitability analysis, multi-scale geographically weighted regression (MGWR), and the space-time cube were used to obtain predictive results. These results were then validated and compared using data from convenience store openings in 2023, and the models were recalibrated and re-predicted accordingly, leading to final conclusions and recommendations.

Table 3.1: Comparison of Selected Variables and Actual Data

Category	**Sub-factor**	**Actual Data Conversion**
Store Factors	”Corner” store location	Road nodes[1]
	Rental cofsts	Real estate transaction prices[2]
Demographic Variables	Population density	Population density[3]
	Population change rate	Population change rate[4]
	Average consumer expenditure	Income tax data[5]
Transportation Factors	Bus stops	Bus stops[6]
	MRT stations	MRT stations[7]
	Public bike rental stations	YouBike rental stations[8]
Dependent Variable	Convenience store locations	Data from 1997 to 2023[9]

[1] National Land Surveying and Mapping Center
[2] Department of Land Administration, Ministry of the Interior
[3] Department of Civil Affairs, Taipei City Government
[4] Department of Civil Affairs, Taipei City Government
[5] Financial Data Center, Ministry of Finance
[6] Taipei Public Transportation Office
[7] National Land Surveying and Mapping Center
[8] Department of Transportation, Taipei City Government
[9] Department of Commerce, Ministry of Economic Affairs

Chapter 4

Research Results

4.1 Variable Selection

4.1.1 Mapping Data

Based on previous studies on convenience store location factors, variables were selected as independent variables, while the development of convenience store locations was used as the dependent variable. For subsequent analysis, the data for each variable was visualized as spatial maps, showing their spatial distribution (Table 4.1).

Table 4.1: Methods for Mapping Variables

Variable	**Mapping Method**
Convenience store data from 1997 to 2019	Mapped using Kriging.
Public bike accessibility	Mapped using Kriging.
Bus density	Mapped using kernel density estimation.
MRT density	Mapped using kernel density estimation.
Road node density	Mapped using kernel density estimation.
Rental cost	Spatially linked to the village boundary map using the median transaction price for second floors to basement floors from 2012 to 2019.
Average consumer income	Linked to the village boundary map using median income tax data.
Population change rate	Linked to the village boundary map using population change rate data.
Population density	Linked to the village boundary map using population density data.

4.1.2 Variable Selection

To address the issue of the curse of dimensionality, the variables listed earlier were selected using the ordinary least squares (OLS) method (Table 4.2). The forward selection procedure was applied, and the variance inflation factor (VIF) threshold was set to 5 (Chiu, 2014) to determine multicollinearity. The combination of variables with the highest model fit (Adj R^2) was used as the final set of independent variables.

Table 4.2: Results of Variable Selection

Adj R^2	AICc	Max VIF	Num Vars	Selected Variables
68.17%	18,480	3.99	5	Bus, MRT, Bike, Road, Population Density
67.93%	18,530	3.85	4	Bus, Bike, Road, Population Density
67.64%	18,592	3.62	3	Bike, Road, Population Density
67.43%	18,636	3.89	4	Bus, MRT, Bike, Road, Population Density
67.10%	18,705	2.69	4	MRT, Bike, Income Tax, Road

4.2 Suitability Analysis

4.2.1 Analysis Method

Following the variable selection process, the spatial data was analyzed using the Suitability Modeler in ArcGIS Pro. All independent variables were classified using natural breaks (Jenks) and divided into 1 to 100 classes to minimize the influence of outliers on the data. The importance of each variable for convenience store location was adjusted by assigning weights. Table 4.3 shows the weight adjustments and corresponding explanations:

4.2.2 Analysis Results

Figure 4.1 shows the predicted results from the suitability analysis model, and Figure 4.2 visualizes these results. The areas south of the Keelung River generally show higher suitability, while certain areas such as parks, universities, and highways also reflect specific suitability scores. On the other hand, in the regions north of the Keelung River, areas with higher suitability are mostly concentrated along MRT lines, which are also densely populated areas.

Table 4.3: Weight Adjustments for Independent Variables

Spatial Data	**Weight**	**Explanation**
Road Nodes	4	Representing the "corner" store location factor. According to previous literature, this factor has a significant influence (Li, 1991).
Population Density	4	According to prior research, population density is an important consideration for store location (Wu, 2017).
Bus Stops	2	Proximity to bus or MRT stations is one of the factors considered for store location, but Wu (2017) found its influence to be moderate, hence a medium weight is assigned.
MRT Stations	2	No specific explanation is provided, but this is considered along with bus stops for transportation-related factors.
Public Bike Stations	1	Although prior research did not include public bike stations, they are a common mode of transportation in Taipei City, so they are included with a lower weight.

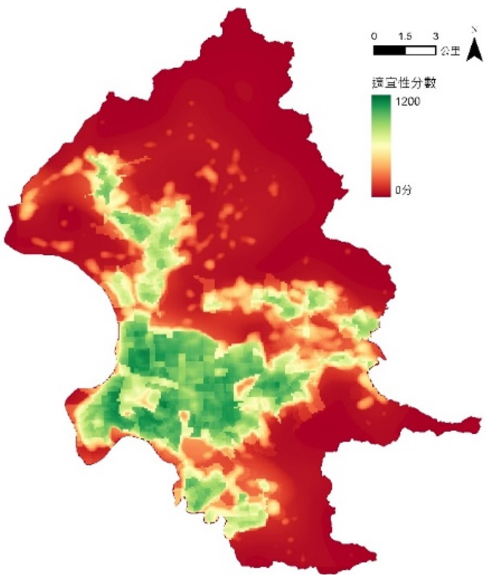

Figure 4.1: Results

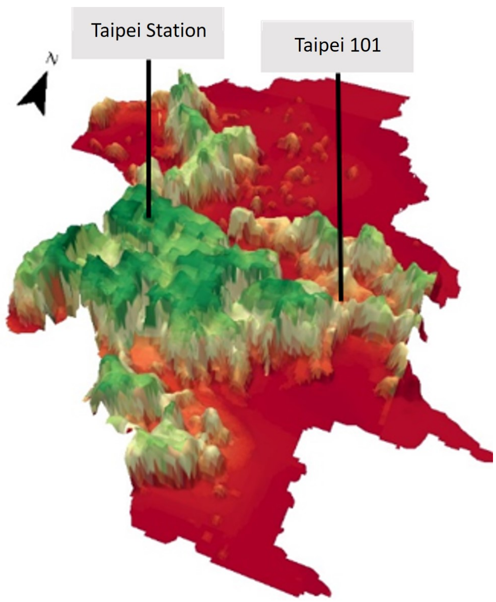

Figure 4.2: Visualization of the Suitability Analysis Results

4.3 Multi-Scale Geographically Weighted Regression (MGWR)

4.3.1 Analysis Method

The previously selected spatial data was used as independent variables for geographically weighted regression, with convenience store distribution data from 2019 serving as the dependent variable. The fit of the dependent variable was compared between the ordinary least squares (OLS) method and the multi-scale geographically weighted regression (MGWR) model. Finally, the residual spatial autocorrelation and R^2 values were observed for both models.

4.3.2 Analysis Results

Figures 4.3 and 4.4 show the statistical results obtained using the ordinary least squares method and the MGWR model, displayed side by side for easy comparison. The analysis reveals that the predicted distribution of the dependent variable shows no significant difference between the two models. Both models assign low values to areas such as campuses, mountainous regions, and locations not serviced by public transportation, while assigning high values to central business districts. As shown in Table 4.4, the MGWR model explains the data better with an R^2 of 0.90, which is higher than that of the OLS method. The residual Moran's I for MGWR is 0.34, which is lower than that of the OLS method, indicating a more random spatial distribution of residuals. Therefore, MGWR is considered the more appropriate model.

Table 4.4: Comparison of OLS and MGWR Results

Model	**R^2**	**Residual Moran's I**
OLS	0.85	0.45
MGWR	0.90	0.34

4.4 Space-Time Cube Analysis

4.4.1 Analysis Method

The Space Time Pattern Mining Tools built into ArcGIS Pro were used for the analysis. A space-time cube was created with convenience store hotspot maps as the spatial data plane and the time dimension as the third axis (with a time scale of 1 year, covering the period from 1996 to 2019). A cold and hot spot analysis using the Gi* statistic was performed. Subsequently, spatial samples were iterated

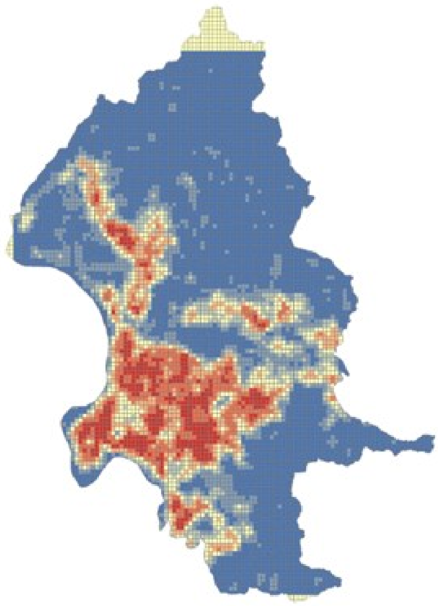

Figure 4.3: OLS Model Results

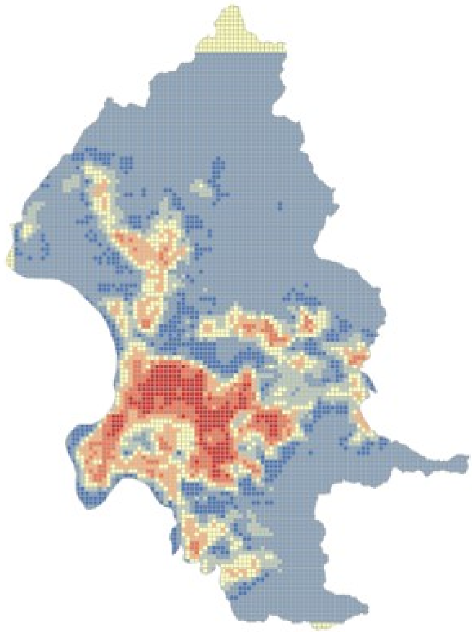

Figure 4.4: MGWR Model Results

using random forest, seasonal smoothing, and curve fitting methods to predict the distribution of convenience stores in 2023. The model with the best fit among the three was selected based on the fit quality of the above methods.

4.4.2 Analysis Results

Figure 4.5 shows the results of the space-time analysis for cold and hot spots. The results indicate that Neihu, Nangang, and the area surrounding Zhongxiao East Road have become hotspots in recent years. In contrast, areas like Heping East Road to Liuzhangli, and Muzha and Jingmei were previously hotspots, but their development has not been significant in recent years. The lower figure shows the predicted distribution three years later (missing values were interpolated using the Kriging method). According to the results, Zhongshan MRT station and the Neihu area have high development potential. Overall, the commercial clusters in Taipei City are not yet saturated according to the space-time cube analysis. The figure below also displays the distribution of different sample analysis methods: exponential smoothing accounted for half of the analysis, while random forest and curve fitting methods each accounted for about a quarter.

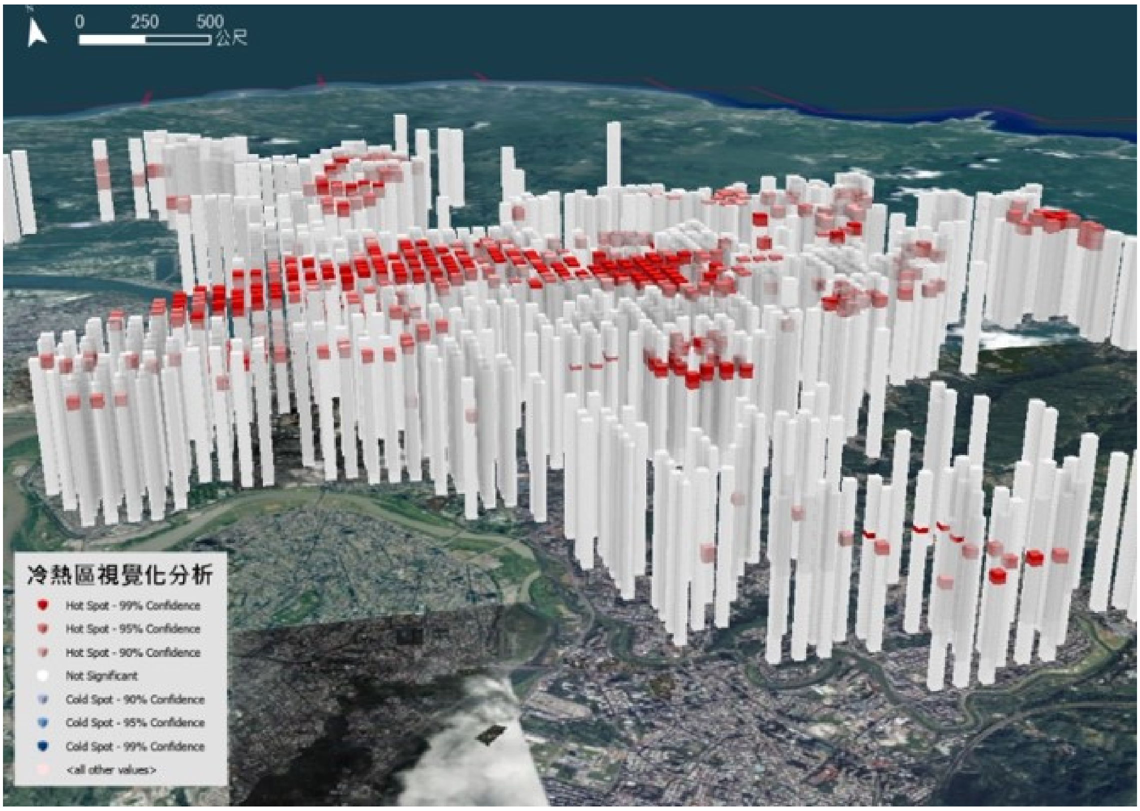

Figure 4.5: Cold and Hot Spot Analysis Results

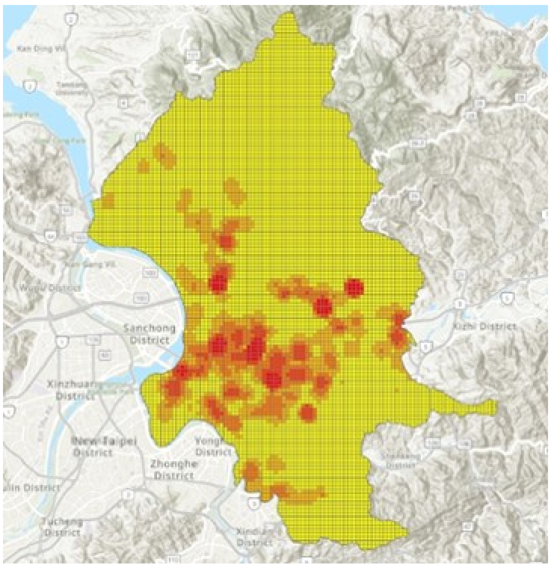

Figure 4.6: Predicted Distribution for 2023 (Kriging Method for Missing Value Interpolation)

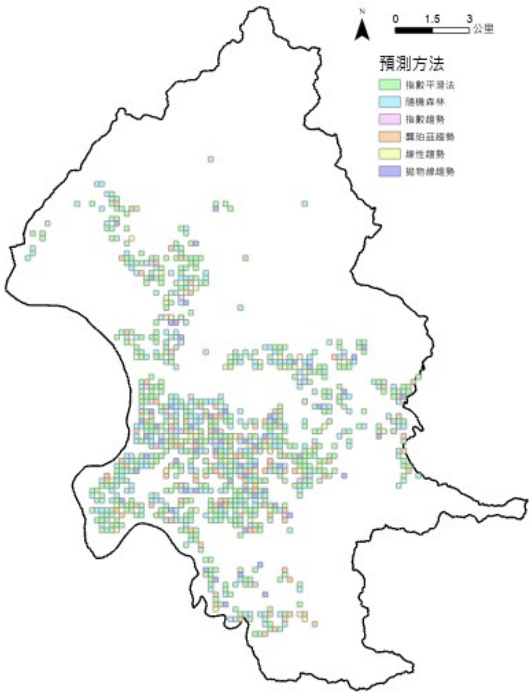

Figure 4.7: Distribution of Different Sample Analysis Methods

4.5 Model Comparison and Validation

Based on the prediction results of each model, verification and analysis were performed on the convenience stores that opened between 2019 and 2023. The study area contains 7,082 grid cells. Each model's suitability score, store opening probability, and fit results were sorted in descending order. A score weight column was added, with the highest value assigned as 7,082 and the lowest as 1. The final score was calculated by multiplying the score weight by the actual number of store openings. Additionally, the total score was divided by the highest possible score to obtain the model's accuracy. Considering that 264 convenience stores opened between 2019 and 2023, the highest possible score was 7,082 multiplied by 264, resulting in 1,869,468. Table 4.5 shows the scores of each prediction model.

Table 4.5: Scores of Each Prediction Model

Method	**Score**	**Accuracy**
Suitability Analysis	1,529,505	82%
MGWR	1,536,419	82%
OLS	1,522,914	81%
Space-Time Cube	1,454,732	78%

The analysis results show that all four models have good predictive capabilities. Among them, the multi-scale geographically weighted regression achieved the highest score and has the most rigorous and scientific calculation method. Statistically, the multi-scale geographically weighted regression also performed best, making it the preferred model for selection.

4.6 Model Calibration and Adjustment

In this study, convenience store data from 1997 to 2022 was used as the dependent variable, while buses, MRT stations, bike stations, roads, and population density were used as independent variables to recalibrate the multi-scale geographically weighted regression (MGWR) model. The results are shown in Figure 4.8. According to the model, the spatial autocorrelation coefficient of the residuals for the MGWR model is 0.23, with an R^2 value of 0.89. These statistics are not significantly different from the previous results, but the model avoids the issue where the development potential areas are overly concentrated in metropolitan regions, which was a limitation of the ordinary least squares (OLS) method. Therefore, the final model excludes the OLS method, and only the predictions from the MGWR are adopted.

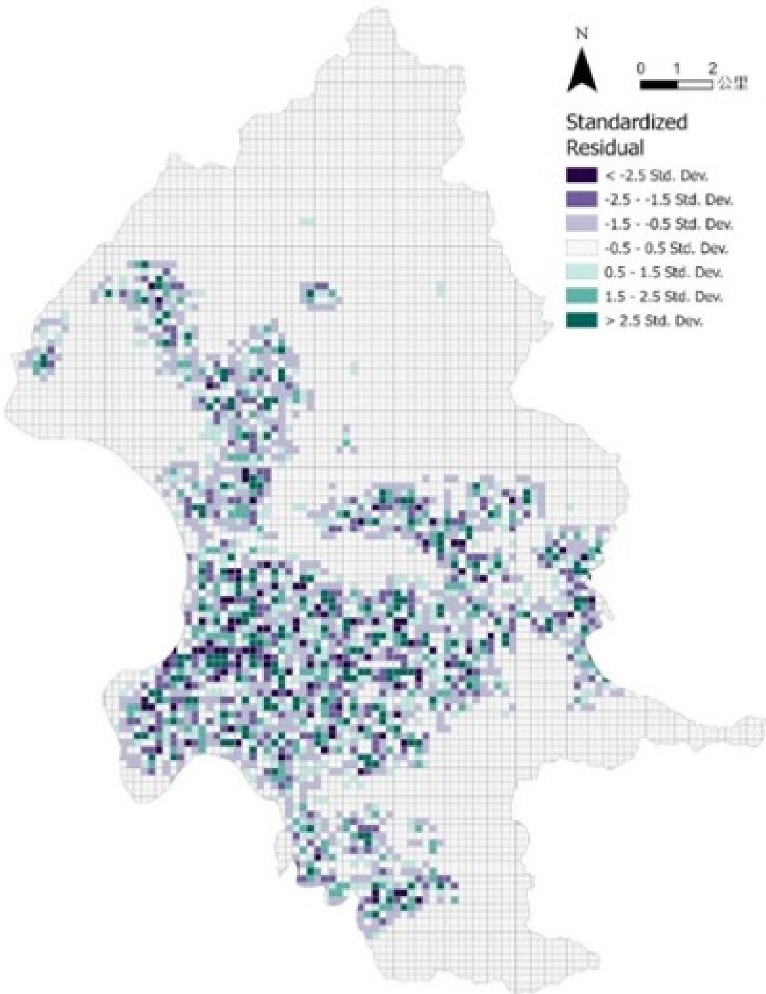

Figure 4.8: MGWR Model Prediction Results After Calibration

Chapter 5

Conclusion and Recommendations

5.1 Areas with Development Potential: Commercial Clusters

According to the prediction trends of the models, especially the calibrated multi-scale geographically weighted regression (MGWR) model, it was found that current commercial clusters, such as the Zhongshan Business District (from MRT Shuanglian Station to Taipei Main Station), the Eastern District Business Area along Zhongxiao East Road MRT line, and industrial parks in Neihu and Nangang, still exhibit development potential. Although these areas already have a higher number of convenience stores compared to other regions, due to excellent storefront locations, high pedestrian flow, and good accessibility to transportation hubs, they meet the key criteria for store expansion and thus remain potential areas for further development.

5.2 Key Factors and Limitations of the Model

The impact of each factor varies across different regions. This study applied MGWR to account for regional differences, assigning different bandwidths to various factors in different areas, which is the key to improving prediction accuracy. However, the predictive model is still limited by factors that cannot be quantified, such as the level of business district development, investment benefits, marketing strategies, competition among peers, and development plans. These unquantifiable factors could not be included in the model, which may cause deviations in the predictions. Future adjustments should consider these factors' impacts on the development potential of convenience stores.

Bibliography

[1] Shawky Mansour, Abdullah Al Kindi, Alkhattab Al-Said, Adham Al-Said, Peter Atkinson, Sociodemographic determinants of COVID-19 incidence rates in Oman: Geospatial modelling using multiscale geographically weighted regression (MGWR), 2021, Sustainable Cities and Society, Volume 65.

[2] Tianfang B. Fang, Yongmei Lu, Constructing a Near Real-time Space-time Cube to Depict Urban Ambient Air Pollution Scenario, 2011, Transactions in GIS, 15(5): 635–649.

[3] Wu Ding, Policy Management, 2003, Taipei: Lianjing Publishing.

[4] Wu Ruo-Jie, A Study on the Factors and Decision System of Convenience Store Location Selection, 2017, Master's Thesis, Department of Information Management, Tamkang University.

[5] Li Yang-Zhe, A Study on Location Selection for Chain Convenience Stores, 1991, Master's Thesis, Department of Business Administration, National Cheng Kung University.

[6] Su Xiang-Qing, The Evolution of Spatial Clustering and Channel Coexistence of Convenience Stores in Taipei, New Taipei, and Taoyuan, 2019, Environment and World, Issue 33: pp. 51–86.

[7] Su Xiang-Qing, The Spatial Distribution and Regional Differences of Convenience Stores in Taiwan: 1987-2016, 2017, Environment and World, Issue 32: pp. 23–60.

[8] Rong Jun-De, A Study on Spatial Data Mining of Census Data—Using Location Analysis of Convenience Stores in Taipei as an Example, 2005, Master's Thesis, Graduate Institute of Geography, National Taiwan University.

[9] Mei Ming-De, Xu Yu-Heng, Qiu Yu-Wen, Cai Jing-Hui, Using Geographic Information Systems to Assist Site Selection for Chain Stores, 2009, GIS Quarterly, Vol. 3, Issue 2.

[10] Zheng Qi-Quan, Chen Yong-Kuan, Zhan Jin-Fa, Wang Su-Fen, The Application of DTM, GIS, and DSS in Forest Ecosystem Classification and Suitability Analysis, 2005, Journal of Photogrammetry and Remote Sensing, Vol. 10, Issue 4, pp. 351–360.

[11] Zhong Kai-Ru, An Analysis of the Usage Characteristics and Spatial Distribution Patterns of Public Bikes in Taipei, 2022, Master's Thesis, Department of Land Economics, National Chengchi University.

[12] Zhu Yan-Li, Jing Chang-Feng, Fu Jia-Yun, Du Ming-Yi, Zhou Lei, Space-Time Cube Analysis of Spatial and Temporal Characteristics of Robbery Incidents, 2019, Journal of Surveying and Mapping Science, Vol. 44, Issue 9.

[13] Wang Shu-Hui, Yu Ru-Ling, GIS and Location Selection Analysis for Chain Businesses—A Case Study of Dumpling Restaurants in Tainan, 2009, Journal of Far East, Vol. 26, Issue 4, pp. 541–552.

[14] Yu Qing-Xiang, Liang Ying-Yi, Guo Rou-Yun, A Discussion on Visualization Analysis with Geographically Weighted Regression, 2022, Journal of Chinese Statistics, Issue 60, pp. 208–232.

[15] Chiu Hao-Zheng, A Comparison of Independent Variables in Multiple Regression and the Influence of Multicollinearity: Estimation and Application of Effect Sizes, Dominance, and Relative Weight Indices, 2014, NTU Management Review, Vol. 27, Issue 3, pp. 1–44.

[16] Zhou Xue-Zheng, Convenience Stores, 2009, Ministry of Culture: Taiwan Encyclopedia, retrieved online on 2023/03/30. URL: https://nrch.culture.tw/twpedia.aspx?id=4134.

[17] The Five Major Convenience Store Chains Nationwide, Government Open Data Platform, retrieved online on 2023/04/02. URL: https://data.gov.tw/dataset/32086.