# Spatial Crime Analysis of Bicycle Theft: Regression Models and Nonlinear Effects

Naoki OHASHI [a], Masanori AJITO [b], Shintaro TERABE [c], Hideki YAGINUMA [d]

[a,b,c,d] *Faculty of Science and Technology, Tokyo University of Science, 2641 Yamazaki, Noda-shi, Chiba 278-8510, Japan*
[a] *E-mail: 7621033@ed.tus.ac.jp*
[b] *E-mail: ajito@rs.tus.ac.jp*

**Abstract**: Bicycle theft remains a concern in Japan, necessitating effective crime prevention. Although crime rates declined for years, recent increases highlight the need for proactive strategies. However, spatial crime analysis in Japan has room for further development. This study constructs a spatial regression model to better understand the factors influencing bicycle theft, contributing to crime prevention and urban planning. We applied Negative Binomial Regression (NBRM) and Generalized Additive Model (GAM) to analyze environmental factors. NBRM demonstrated strong predictive power, identifying key relationships between bicycle theft and proximity to railway stations and pachinko parlors. A novel finding was the positive correlation between hospital locations and bicycle theft, suggesting the importance of hospital bicycle parking management. GAM captured nonlinear influences on theft patterns, providing additional insights. These findings underscore the significance of spatial modeling in crime analysis and highlight the potential for data-driven improvements in urban safety.

*Keywords*: Negative Binomial Regression Model, Generalized Additive Model, Bicycle Theft, Nonlinear Effect, Spatial Correlation

## 1. INTRODUCTION

### 1.1 Background and Purpose

Criminal offenses in Japan exhibited a steady decline from 2002 to 2021, when they reached a post-war minimum. However, since 2021, crime rates have increased for three consecutive years. The body of literature on crime analysis in Japan remains limited, potentially due to the country's comparatively low crime rate relative to other developed nations. However, crime rates in Japan have been rising since the COVID-19 pandemic, necessitating urgent countermeasures. Therefore, this study aims to construct a model to analyze the factors contributing to bicycle theft and generate new knowledge that will contribute to crime prevention efforts by the police and bicycle theft prevention measures in urban development. Bicycle theft is characterized by a high incidence rate and a persistently low clearance rate, underscoring the urgent need for effective crime prevention measures. Therefore, this study examines the factors contributing to bicycle theft to inform targeted prevention strategies.

### 1.2 Literature Review and Position of This Study

Previous research on bicycle theft has employed various modeling approaches. Hashimoto *et al*. (2023) developed an incidence estimation model for Okayama City and Kurashiki City, Japan, utilizing seven years of bicycle theft data to investigate the relationship between theft occurrences and urban environmental elements. Similarly, Sugiura *et al*. (2022) examined the correlation between bicycle theft and environmental factors such as land use patterns, population density, and facility distribution near railway stations in Tokyo's 23 wards. Their findings revealed strong associations between bicycle theft and factors including the prevalence of abandoned bicycles, the availability of bicycle parking facilities, and the presence of recreational establishments.

When constructing a regression model with crime count data as the response variable, ordinary linear regression, which assumes normality, is inappropriate. To address this, Hashimoto *et al*. (2023) employed Negative Binomial Regression, while Sugiura *et al*. (2022) used kernel density estimates as the response variable. However, Tamesue and Tsutsumi (2017) identified a methodological gap in crime prediction research in Japan, emphasizing the lack of an established analytical approach that incorporates spatial correlation in count data. Although Oshita and Tarumi (2008) and Shimada *et al*. (2002) attempted to integrate spatial correlation, their methodologies had limitations. The former was constrained to a small-scale model with a single explanatory variable, while the latter solely analyzed spatial correlation using Moran's I statistic without constructing a regression model.

Given the limited amount of crime research that rigorously addresses spatial data characteristics, this study aims to develop a regression model that explicitly incorporates spatial correlation. Through systematic comparison and critical evaluation of parameter estimation results, we seek to generate novel insights that extend beyond the findings of existing research.

## 2. DATA OVERVIEW

### 2.1 Response Variable

This study utilizes bicycle theft data disaggregated by district and municipality, published by the Chiba Prefectural Police Department. The response variable constructed by aggregating bicycle theft incidents by district and municipality over a five-year period(2018-2022).

Within Chiba Prefecture, Chiba City recorded the highest incidence of bicycle theft during the study period (n=6,462), representing approximately 1.68 times the frequency observed in Funabashi City (n=3,855), which ranked second in prevalence, thus indicating a significant need for preventive interventions.

### 2.2 Explanatory Variables

Table 1 presents the explanatory variables and data sources used in this study.

Table 1. List of explanatory variables and sources

| | Explanatory Variables | Symbol | Created by |
|---|---|---|---|
| Land of Use | Area of Category 1 exclusively low-rise residential zone | lowres1 | MLIT |
| | Area of Category 2 Exclusively low-rise residential zone | lowres2 | MLIT |
| | Area of Category 1 mid/high-rise oriented residential zone | midres1 | MLIT |
| | Area of Category 2 mid/high-rise oriented residential zone | midres2 | MLIT |
| | Area of Category 1 residential zone | resarea1 | MLIT |
| | Area of Category 2 residential zone | resarea2 | MLIT |
| | Area of Quasi-residential zone | quasires | MLIT |

| | | | |
|---|---|---|---|
| | Area of Neighborhood commercial zone | comarea1 | MLIT |
| | Area of Commercial zone | comarea2 | MLIT |
| | Area of Quasi-industrial zone | quasiind | MLIT |
| | Area of Industrial zone | indarea1 | MLIT |
| | Area of Exclusively industrial zone | indarea2 | MLIT |
| Road | Length of roads with a width of 19.5m or more | w-19.5- | MLIT |
| | Length of roads with a width of 13m or more but less than 19.5m | w-13-19.5 | MLIT |
| | Length of roads with a width of 5.5m or more but less than 13m | w-5.5-13 | MLIT |
| | Length of road less than 5.5m wide | w-0-5.5 | MLIT |
| Population | Population under 15 years old | pop-0-14 | MIAC |
| | Population between 15 and 19 years old | pop-15-19 | MIAC |
| | Population over 20 years old | pop-20- | MIAC |
| Station | Distance to the nearest station | d-station | MLIT |
| | Number of passengers at the nearest station | passengers | MLIT |
| | Capacity of bicycle parking lot at the nearest station | c-bicycle | Chiba City |
| Facility | Distance to the nearest supermarket<br>Number of supermarkets | d-super<br>n-super | NTP |
| | Distance to the nearest pachinko parlor<br>Number of pachinko parlors | d-pachinko<br>n-pachinko | NTP |
| | Distance to the nearest karaoke room<br>Number of karaoke rooms | d-karaoke<br>n-karaoke | NTP |
| | Distance to the nearest convenience store<br>Number of convenience stores | d-convini<br>n-convini | NTP |
| | Distance to the nearest sports facility<br>Number of sports facilities | d-sfac<br>n-sfac | Chiba City |
| | Distance to the nearest tourist facility<br>Number of tourist facilities | d-tfac<br>n-tfac | Chiba City |
| | Distance to the nearest city park<br>Number of city parks | d-park<br>n-park | MLIT |
| | Distance to the nearest police station<br>Number of police stations | d-plst<br>n-plst | MLIT |
| | Distance to the nearest university<br>Number of universities | d-univ<br>n-univ | MLIT |
| | Distance to the nearest junior high school<br>Number of junior high schools | d-jhs<br>n-jhs | MLIT |
| | Distance to the nearest high school<br>Number of high schools | d-hs<br>n-hs | MLIT |
| | Distance to the nearest hospital<br>Number of hospitals | d-hospital<br>n-hospital | MLIT |
| | Distance to the nearest library<br>Number of libraries | d-lib<br>n-lib | Chiba City |
| Household | Number of head households | mainhh | MIAC |
| | Number of detached households | dhh | MIAC |
| | Number of households in apartment buildings | apt | MIAC |
| | Number of households in 1-2 story apartment buildings | apt-1-2 | MIAC |

| | | | |
|---|---|---|---|
| | Number of households in 3-5 story apartment buildings | apt-3-5 | MIAC |
| | Number of households in 6-10 story apartment buildings | apt-6-10 | MIAC |
| | Number of households in 11 story or more apartment buildings | apt-11- | MIAC |

MLIT: Ministry of Land, Infrastructure, Transport and Tourism
MIAC: Ministry of Internal Affairs and Communication
NTP: NTT Town Pages Co., Ltd.

- Land of Use

Land use zoning data produced by MLIT in 2019 was obtained for this analysis. As the original dataset was in grid-cell (mesh) format, areal weighting interpolation was applied to transform the data to district-level spatial units for analytical purposes (unit: $m^2$).

- Road

Road density and road length mesh data produced by MLIT in 2010 were obtained and categorized into four classes based on width thresholds of 5.5m, 13m, and 19.5m. These gridded datasets were subsequently transformed to district-level spatial units through areal weighting interpolation and incorporated as explanatory variables in the analysis (unit: m).

- Population

Age-specific population data for 2020, produced by MIAC, was obtained and classified into three categories: under 15 years old, 15 to 19 years old, and 20 years old and over. Each category was incorporated as an explanatory variable in the analysis.

- Station

Railway data and station-specific passenger volume data, produced by MLIT in 2020, were obtained. The distance to the nearest station and the number of passengers at the nearest station were computed for each district. Additionally, the number of bicycle parking spaces at the nearest station was compiled from the Chiba City website, published in April 2023.

- Facility

Facility data from MLIT (city parks: 2011, police stations: 2012, universities: 2021, junior high schools: 2021, high schools: 2021, hospitals: 2020) and open data released by Chiba City in June 2024 (sports facilities, tourist facilities, libraries) were obtained. These datasets were used to calculate both the distance to the nearest facility and the number of facilities per district. Moreover, variables for supermarkets, convenience stores, karaoke rooms, and pachinko parlors were compiled from NTT Town Page in February 2025.

- Household

Household data by housing construction type, produced by MIAC in 2020, was obtained. The number of head households, detached houses, and apartment buildings (1 - 2 stories, 3 - 5 stories, 6 - 10 stories, 11 or more stories) were aggregated by district and incorporated as explanatory variables.

### 2.3 Building a Regression Model

The model construction methodology in this study followed the framework proposed by Sugiura *et al*. (2022).

For explanatory variable selection, a bidirectional stepwise procedure was implemented using the stepAIC function from the MASS package in R statistical software, while Variance Inflation Factor (VIF) values were calculated to assess potential multicollinearity. Table 2 presents the final set of explanatory variables retained following the stepwise selection procedure, along with their corresponding VIF values. As all VIF values were below the threshold of 5, multicollinearity among the explanatory variables was deemed negligible.

Table 2. Explanatory variables and VIF values adopted by stepwise

| Variable | VIF |
|---|---|
| d-station | 2.041 |
| d-jhs | 1.247 |
| d-hs | 1.666 |
| d-tfac | 2.997 |
| d-sfac | 4.031 |
| d-lib | 3.189 |
| n-hospital | 1.648 |
| n-convini | 1.877 |
| d-pachi | 1.544 |
| lowres1 | 1.854 |
| midres1 | 2.110 |
| resarea1 | 1.718 |
| pop-0-14 | 2.077 |
| w-0-5.5 | 2.803 |
| w-13-19.5 | 2.387 |
| apt-1-2 | 2.291 |
| apt-3-5 | 1.759 |

Given the predominance of count data among the explanatory variables, robust Z-score standardization was employed to mitigate the potential influence of outliers on parameter estimates. Model performance was evaluated using Akaike's Information Criterion (AIC) and Nagelkerke's Pseudo $R^2$ as goodness-of-fit metrics.

## 3. ANALYSIS METHOD

### 3.1 Negative Binomial Regression Model

The Negative Binomial Regression Model (NBRM) is the most widely used model for count data with a high proportion of zeros (Lawless, 1987; Hilbe, 2007). In the NBRM, both the mean parameter $\lambda_i$ and the size parameter, which controls the degree of overdispersion, are estimated.

In this study, to assess the effect of standardizing event counts, we constructed three models: one without an offset term (an explanatory variable with a fixed regression coefficient of 1), one using population as an offset term, and one using area as an offset term. We then compared these models based on AIC and Pseudo $R^2$.

### 3.2 Generalized Additive Model

The Generalized Additive Model (GAM) (Hastie and Tibshirani, 1986), extends the Generalized Linear Model (GLM) framework to capture a broader spectrum of effects, including spatial and spatiotemporal correlations. It is defined as follows:

$$y_i \sim P(\theta_i)$$
$$\theta_i = \mathrm{g}\left(\sum_{k=1}^{K} x_{i,k}\beta_k + \sum_{l=1}^{L} f(z_{i,l})\right) \tag{1}$$

For this analysis, the probability distribution of parameter $\theta_i$ was specified as negative binomial, with an exponential function employed as the inverse link function $g(\cdot)$. The function $f(z_{i,l})$ models the effect of variable $z_{i,l}$, thereby accommodating nonlinear relationships in the analytical framework. It is defined as follows:

$$f(z_{i,l}) = \sum_{k_l=1}^{K_l} \varphi_{i,k_l}\gamma_{k_l}^{(l)} \tag{2}$$

The term $\varphi_{i,k_l}$ represents a basis function - a predefined functional form designed to capture specific effects - while $\gamma_{k_l}^{(l)}$ denotes the coefficient quantifying the contribution of each basis function to the overall model. The derivation of basis functions necessitates the specification of an appropriate spline function. To preserve model interpretability, thin-plate spline functions were implemented for basis function generation, as these minimize curvature in the resulting functional forms. One-dimensional basis functions facilitate the estimation of nonlinear effects from individual explanatory variables, while two-dimensional basis functions generated across latitude-longitude coordinates enable the quantification of spatial correlation patterns.

## 4. ANALYSIS RESULTS

### 4.1 Standardization of the Response Variable and Interpretation of NBRM Results

To investigate the effect of response variable standardization based on aggregation units, three distinct models were constructed and their parameters estimated (Table 3): Model 1 utilizes the absolute count of bicycle thefts as the response variable without an offset term; Model 2 incorporates population as an offset variable, effectively modeling bicycle theft rate (thefts per capita); and Model 3 employs area as an offset variable to analyze theft density (thefts per unit area). While Pseudo $R^2$ values were comparable across all three specifications, Model 1 exhibited the lowest Akaike Information Criterion (AIC), indicating superior goodness-of-fit and predictive performance. Statistical significance was established at the $\alpha = 0.05$ level for parameter evaluation. The subsequent analysis focuses exclusively on statistically significant parameters ($p < 0.05$) from Model 1.

The estimated coefficient for the distance to the nearest station(d-station) is −0.332, suggesting that bicycle theft decreases as the distance from the station increases. In other words, bicycle thefts are more frequent in areas closer to the station. This suggests that high pedestrian traffic and frequent use of bicycle parking facilities near stations increase opportunities for theft. The estimated coefficient for the number of hospitals(n-hospital) was 0.237, novel finding not reported in previous studies. The number of bicycle thefts exhibited a strong positive correlation with the number of hospitals (correlation coefficient $r$ = 0.71), suggesting that bicycle thefts may be more frequent in bicycle parking facilities within hospitals and their surrounding areas.

The estimated coefficients for category 1 low-rise residential areas(lowres1) and category 1 mid/high-rise residential areas(midres1), were −0.107 and −0.063, respectively, indicating that districts with higher proportion of these areas tended to experience fewer bicycle thefts. These areas are predominantly residential, characterized by high occupancy rates and a limited number of commercial facilities. This results in fewer non-residents and increased surveillance, which may reduce the likelihood of bicycle theft. Conversely, the estimated coefficient of category 1 residential area(resarea1) was 0.090, indicating that districts with a higher proportion of these areas are more susceptible to bicycle theft. Unlike lowres1 and midres1, resarea1 permits the construction of medium-sized commercial facilities (up to 3,000 $m^2$), making bicycle theft more likely in and around bicycle parking areas associated with these facilities. Thus, the difference in land use between lowres1, midres1, and resarea1 influence bicycle theft, suggesting that the risk of bicycle theft increases in districts with a higher concentration of commercial facilities.

The estimated coefficient for the distance to pachinko parlors (d-pachi) was −0.159, suggesting that bicycle theft decreases as the distance from pachinko parlors increases. This suggests that bicycle theft is more prevalent in areas closer to pachinko parlor. Since patrons of pachinko parlor tend to park their bicycles for extended periods, they may become prime targets for theft. Similarly, the estimated coefficient for the distance to the library(d-lib) was −0.130. Bicycle thefts were also frequent in the vicinity of the libraries. This may be because users tend to park their bicycles for extended periods and fail to manage them properly.

The estimated coefficient for the distance to tourist facilities (d-tfac) was −0.142, suggesting that bicycle theft is more frequent near tourist facilities. This finding suggests that visitors frequently use bicycles near tourist facilities and that bicycle parking facilities may be inadequately managed. The estimated coefficient of 0.031 for road widths less than 5.5 meters (w-0-5.5) suggests that bicycle theft is more likely to occur on narrow streets due to limited visibility.

These findings confirm that the risk of bicycle theft is higher around train stations, medical facilities, and pachinko parlors, whereas it is lower in residential areas. The findings also suggest that road conditions and facility management may influence the occurrence of bicycle theft. Based on these findings, strengthening bicycle parking management, particularly around train stations and medical facilities, and enhancing surveillance near pachinko parlors and libraries are recommended.

Table 3. Parameter estimation results for NBRM

| | Model 1 | | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|
| | Estimate | Z value | Estimate | Z value | Estimate | Z value |
| d-station | −0.332 | −5.528 *** | −0.464 | −7.003 *** | −0.498 | −7.721 *** |
| d-jhs | −0.089 | −1.877 * | 0.022 | 0.410 | −0.157 | −3.122 *** |
| d-hs | −0.134 | −2.395 ** | −0.049 | −0.794 | −0.266 | −4.307 *** |
| d-tfac | −0.142 | −2.361 ** | −0.229 | −3.428 *** | −0.219 | −3.350 *** |
| d-sfac | 0.091 | 1.883 * | 0.057 | 1.063 | 0.159 | 2.966 *** |
| d-lib | −0.130 | −2.530 ** | −0.082 | −1.432 | −0.270 | −4.793 *** |
| n-hospital | 0.237 | 7.694 *** | 0.256 | 7.514 *** | 0.230 | 6.822 *** |

| | | | | | | |
|---|---|---|---|---|---|---|
| n-convini | 0.059 | 1.592 | 0.098 | 2.431 ** | −0.009 | −0.214 |
| d-pachi | −0.159 | −3.008 *** | −0.076 | −1.300 | −0.252 | −4.448 *** |
| lowres1 | −0.107 | −2.937 *** | −0.091 | −2.301 ** | −0.093 | −2.345 ** |
| midres1 | −0.063 | −2.978 *** | −0.017 | −0.705 | −0.076 | −3.289 *** |
| resarea1 | 0.090 | 3.452 *** | 0.098 | 3.408 *** | 0.029 | 1.011 |
| pop-0-14 | 0.104 | 3.348 *** | −0.198 | −5.724 *** | −0.008 | −0.227 |
| w-0-5.5 | 0.031 | 3.458 *** | 0.021 | 2.207 ** | −0.033 | −3.274 *** |
| w-13-19.5 | −0.015 | −1.786 * | −0.015 | −1.612 | −0.017 | −1.875 * |
| apt-1-2 | 0.074 | 1.560 | −0.119 | −2.295 ** | 0.090 | 1.761 * |
| apt-3-5 | 0.157 | 6.116 *** | −0.001 | −0.039 | 0.182 | 6.490 *** |
| Pseudo $R^2$ | 0.84 | | 0.83 | | 0.88 | |
| AIC | 3243 | | 3340 | | 3323 | |

Statics significance; ‘***’: 1%, ‘**’: 5%, ‘*’: 10%
Model 1 uses the number of bicycle thefts as the objective variable.
Model 2 uses population as the offset term.
Model 3 uses area as the offset term.

### 4.2 Nonlinear Effects Analysis and Findings Using GAM

Table 4 presents the effective degrees of freedom of the explanatory variables and their test statistics (chi-square values), assuming nonlinear effects, as evaluated using GAM. However, the number of convenience stores(n-convini) and the number of hospitals(n-hospital) were count data with a small sample size, which prevented the generation of spline functions. Therefore, we assumed linear effects for these variables. Additionally, as the Edf for the distance to the nearest library(d-lib), Category 1 mid/high-rise residential areas(midres1), and the number of households in 1-2 story apartment(apt-1-2) was 1, suggesting a linear relationship, these variables were not included in the nonlinear terms and were assumed to have linear effects.

Table 4 shows that nonlinear effects are statistically significant for many of the explanatory variables, suggesting that the relationship between these variables and bicycle theft cannot be adequately captured by a simple linear model. In particular, factors such as the distance to the nearest station or specific facilities, population composition, and residential environment may exert complex influences on bicycle theft.

The AIC was approximately 40 points lower compared to Model 1 in Table 3. However, as several explanatory variables remained statistically insignificant, the improvement in model accuracy due to incorporating spatial correlation and assuming nonlinear effects was somewhat limited.

Table 4. Parameter estimation results for nonlinear effects of explanatory variables

| Smooth term | Edf | Chi.sq | Smooth term | Edf | Chi.sq |
|---|---|---|---|---|---|
| s(d-station) | 5.085 | 46.72 *** | s(resarea1) | 1.414 | 12.59 *** |
| s(d-jhs) | 2.672 | 8.902 ** | s(pop-0-14) | 2.457 | 16.72 *** |
| s(d-hs) | 1.055 | 6.686 ** | s(w-0-5.5) | 3.782 | 17.29 *** |
| s(d-tfac) | 3.419 | 9.211 * | s(w-13-19.5) | 1.293 | 0.297 |
| s(d-sfac) | 1.322 | 0.704 | s(apt-3-5) | 2.000 | 34.43 *** |
| s(d-pachi) | 3.896 | 14.20 ** | s(X, Y) | 5.150 | 11.23 |
| s(lowres1) | 1.005 | 9.578 *** | AIC | 3202 | |

Statics significance; '***': 1%, '**': 5%, '*': 10%

Figure 1 and Figure 2 illustrate the nonlinear effects of the distance to the nearest station(d-station) and the distance to the nearest pachinko parlor(d-pachi), where their Edf exceeded 3, indicating pronounced nonlinear effects. The horizontal axis ranges from 0 to 3000 meters, reflecting the assumption that the effect is limited within a 3 km radius. The vertical axis represents the effect magnitude estimated by the spline function. The figures also display the observed data points along with 95% confidence interval, represented by a dotted line.

Figure 1 illustrates that the effect of distance to the nearest station(d-station) on bicycle theft is strongest near the station and declines sharply up to approximately 500 meters. Subsequently, the effect remains relatively stable up to approximately 1500 meters but declines again beyond this threshold. This finding suggests that commercial facilities and pedestrian traffic around stations may influence bicycle theft. Beyond a certain distance, the factors contributing to bicycle theft may be more closely associated with land use characteristics, such as residential areas. However, it is important to note that data availability decreases beyond 1500 meters, leading to increased uncertainty in the estimates.

Figure 2 illustrates that the effect of the distance to the nearest pachinko parlor(d-pachi) on bicycle theft is strongest near the parlor and diminishes as the distance increases. In particular, the effect declined sharply up to approximately 500 meters, after which it remained relatively stable with no substantial fluctuations. This finding suggests that bicycle parking lots at pachinko parlors and adjacent commercial facilities may elevate the risk of bicycle theft. Conversely, beyond 1500 meters, the effect of pachinko parlors on bicycle theft becomes statistically uncertain, as the variation is minimal and the confidence interval widens. Therefore, the influence of pachinko parlors on bicycle theft may be confined to a radius of approximately 500 meters.

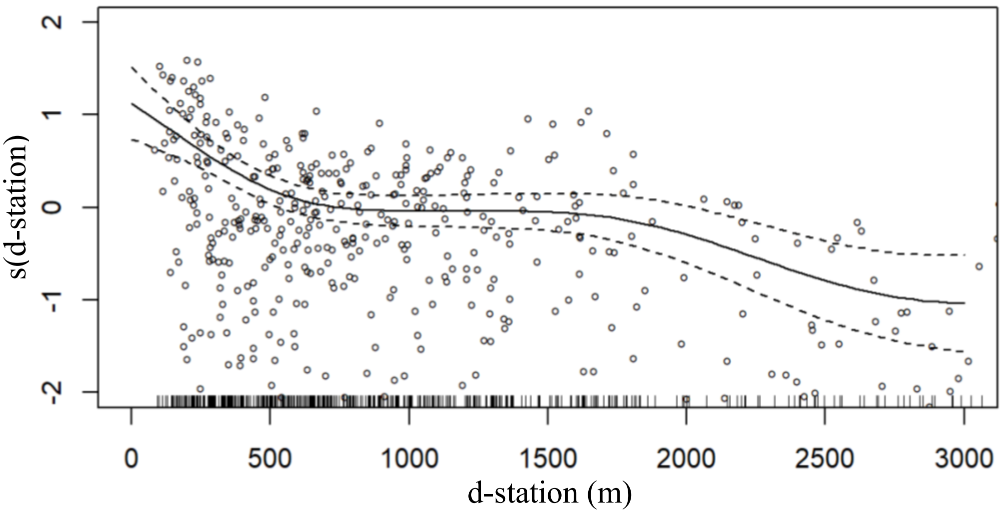

Figure 1. The effect of distance to the nearest station on bicycle theft

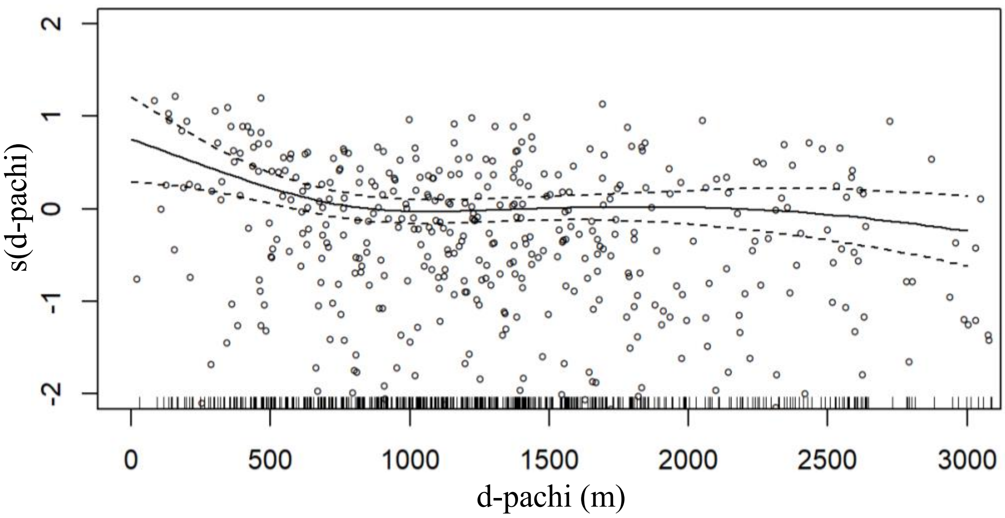

Figure 2. The effect of distance to the nearest pachinko parlor on bicycle theft

## 5. CONCLUSION

In this study, we developed NBRM and GAM to analyze the factors contributing to bicycle theft.

The NBRM analysis enabled us to identify the factors associated with bicycle theft with high accuracy. In particular, the model using the number of bicycle thefts as the response variable demonstrated high accuracy, suggesting standardizing theft counts may not always necessary to standardize the number of thefts. The analysis indicated a strong association between bicycle theft and the distance to the nearest station and pachinko parlor. Additionally, a positive correlation between the number of hospitals and bicycle theft was identified, representing a novel finding not documented in previous studies. This finding suggests that the use and management of bicycle parking facilities near hospitals may influence the incidence of bicycle theft, making it a potentially critical factor in future theft prevention strategies.

The GAM analysis identified nonlinear effects of explanatory variables on bicycle thefts incidence. In particular, the effect of distance to the nearest station and pachinko parlor were nonlinear, with the highest effect magnitude observed near these facilities, gradually diminishing as distance increased. These findings indicate that GAM enables a detailed assessment of factors that conventional linear regression models fail to capture, highlighting the effectiveness of nonlinear models in crime prediction.

One limitation of this study is that the accuracy of the GAM was not sufficiently high. While incorporating nonlinear relationships among explanatory variables resulted in some improvement in accuracy, the overall effect remained modest. Further accuracy improvements may be achieved by refining the selection of explanatory variables, optimizing the range of spline function applications, and incorporating interaction effects between variables. Another limitation was that, although spatial correlation was modeled, it did not yield statistically significant parameters. Although spatial effects were accounted for using a spline function derived from coordinates, the anticipated spatial correlation effects were not observed. This may due to the relatively small study area, which may have precluded the detection of significant spatial autocorrelation patterns. Generally, spatial autocorrelation tends to be more pronounced in datasets covering larger geographic areas, making it challenging to detect in localized analyses such as this study. Therefore, additional analysis encompassing a broader geographic scope is necessary to enhance the modeling of spatial correlations using GAM.

## REFERENCES


Hashimoto, S., Yada, A., Kudo, H., Uno, H., Higuchi, T. (2023) A Study on the Construction of Estimate Models for Bicycle Theft Frequent Area. *JSTE journal of traffic engineers*, 9(4A), 18-26

Hastie, T., Tibshirani, R. (1986) Generalized additive models. *Stat. Sci.*, 1(3), 297-318

Hilbe, J.M. (2007) *Negative Binomial Regression*, Cambridge University Press, Cambridge

Lawless, J. (1987) Negative binomial and mixed Poisson regression. *The Canadian Journal of Statics*, 15, 2099-225

Ohyama, T., Amemiya, M., Shimada, T., Nakaya, T. (2017) Recent Research Trends on Geographical Crime Prediction. *Theory and Applications of GIS*, 25(1), 33-43

Oshita, Y., Tarumi, T. (2008) Spatial analysis of crime data from Kawaguchi City. *Okayama University Faculty of Environmental Science and Engineering Research Report,* 13(1), 17-22

Shimada, T., Suzuki, M., Harada, Y. (2002) Analysis of crime distribution patterns using Moran's I statistics. *Theory and Applications of GIS*, 10(1), 49-57

Sugiura, H., Hino, K., Asami, Y., Yamada, I. (2022) Association between environmental factors and bicycle thefts around railway stations. J. *Archit. Plann., AIJ,* 87(791), 123-132

Tamesue, K., Tsutsumi, M. (2017) Spatial econometric models for count data. *Journal of Civil Engineering Planning*, 51(297), 1-5