



Article Point Event Cluster Detection via the Bayesian Generalized Fused Lasso

Ryo Masuda 🕩 and Ryo Inoue *🕩

Department of Human-Social Information Sciences, Graduate School of Information Sciences, Tohoku University, Miyagi 980-8577, Japan; ryo.masuda.t8@alumni.tohoku.ac.jp * Correspondence: ripoue@tohoku.ac.jp

* Correspondence: rinoue@tohoku.ac.jp

Abstract: Spatial cluster detection is one of the focus areas of spatial analysis, whose objective is the identification of clusters from spatial distributions of point events aggregated in districts with small areas. Choi et al. (2018) formulated cluster detection as a parameter estimation problem to leverage the parameter selection capability of the sparse modeling method called the generalized fused lasso. Although this work is superior to conventional methods for detecting multiple clusters, its estimation results are limited to point estimates. This study therefore extended the above work as a Bayesian cluster detection method to describe the probabilistic variations of clustering results. The proposed method combines multiple sparsity-inducing priors and encourages sparse solutions induced by the generalized fused lasso. Evaluations were performed with simulated and real-world distributions of point events to demonstrate that the proposed method provides new information on the quantified reliabilities of clustering results at the district level while achieving comparable detection performances to that of the previous work.

Keywords: spatial cluster detection; point event; Bayesian inference; sparse modeling; generalized fused lasso

1. Introduction

Deeper understanding of the socio-economic activities in small areas is often essential for discussing and planning regional strategies. As open data policies have become increasingly popular in many public sectors, a broad range of geospatial data are also made publicly available with high spatial resolutions. The available geospatial data, when combined with effective spatial analysis methods, are expected to provide detailed insights into the geographical aspects of socio-economic activities.

One of the common forms of geospatial data is point event data, which are the focus of the present study. Point event data are used to record event occurrences, such as crimes or infectious diseases, along with their locations. The crucial aspect of point event data is the unevenness of its spatial distribution. If we were able to know whether there are any areas with elevated risk for these events and their corresponding locations, it would be possible to implement effective measures against such events. Therefore, numerous studies have proposed cluster detection methods to identify sets of subregions (referred to as clusters) that are distinguished by higher event occurrences. These methods aim to identify broadly local spatial dissimilarities and can be viewed as special versions of the general hotspot/coldspot analyses such as the Getis–Ord Gi* [1] and Anselin local Moran's I [2].

The basic motivation of cluster detection methods is to obtain both reliable and spatially detailed cluster information from point event data. Achieving this is challenging because clusters exhibit spatially flexible shapes and we often have little prior knowledge about them. With the ability to provide statistical evidence for detected clusters, the spatial scan statistic [3] is a conventional and widely used (e.g., a freely available software SaTScan [4]) detection approach. This method has been extended to deal with various



Citation: Masuda, R.; Inoue, R. Point Event Cluster Detection via the Bayesian Generalized Fused Lasso. *ISPRS Int. J. Geo-Inf.* 2022, *11*, 187. https://doi.org/10.3390/ ijgi11030187

Academic Editor: Wolfgang Kainz

Received: 28 January 2022 Accepted: 10 March 2022 Published: 11 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). types of point event data, such as Poisson [3], exponential [5], and case–control data [6]. However, in exchange for securing statistical validity, the spatial scan statistic outputs single clusters rather than multiple clusters to avoid multiple testing problems that occur in the statistical procedure. Moreover, the clustering results of this method depend on the shape of predefined scanning windows, which sweep across the entire study area. As these requirements prevent the detection of clusters with flexible shapes, some studies extended the spatial scan statistic and enabled the detection of multiple clusters [7] and irregularly shaped clusters [8,9]. Although these efforts alleviated the limitations of the spatial scan statistic, multiple cluster detection faces computational difficulties, and prior settings are still needed for the shape of scanning windows.

Among the other methods, the false-discovery-rate (FDR)-based approach [10,11] applies the false discovery rate controlling technique in statistical testing theory for cluster detection. This approach allows the inclusion of a certain number of false discoveries, thereby enabling simultaneous detection of multiple clusters. However, the FDR-based approach fails to provide the intensity of the concentration of event occurrences at the sub-region level, which is particularly essential for identifying clusters that comprise groupings of spatially contiguous subregions.

To develop a cluster detection method that outputs spatially flexible clusters, a sparsemodeling-based method [12] is proposed to overcome some major drawbacks of conventional methods. Choi et al. [12] formulated cluster detection as a maximum likelihood estimation problem, where the likelihood function is derived from a Poisson regression model with generalized fused lasso penalties [13]. In the proposed regression model, each subregion-based intensity parameter represents the degree of concentration in that subregion, while the covariate vector of parameters adjusts for the observed covariates. The generalized fused lasso penalties induce zero values for the intensity parameters and identical values for adjacent pairs of intensity parameters. By introducing regularization penalties, the method achieves cluster detection through the estimated values of the intensity parameters. Although Choi et al. [12] successfully reported overcoming the existing limitations around multiple cluster detection by introducing parameter selection, there is room for additional improvement in the parameter estimations.

Choi et al. [12] proposed a computational procedure to output point estimates of the subregion-based parameters using the majorize–minimization (MM) algorithm [14]. Although the point estimates are the most fundamental parameter estimates, they do not indicate the degree of reliability of the estimated parameters. As reliability assessments are crucial to cluster detection results that vary across subregions, the estimation method should be capable of handling uncertainties.

Bayesian estimation is a statistical framework that expresses parameter uncertainties as probability distributions; it treats all the parameters of a statistical model as stochastic variables and estimates the probability distribution of each parameter. In the Bayesian framework, a probability distribution expressing prior knowledge can be considered as a prior distribution. Among the different types of probability distributions, the Laplace distribution [15,16] is known to encourage sparsity in the estimated parameters and is viewed as a Bayesian extension of lasso. Kyung et al. [17] further formulated the Bayesian counterpart of the generalized fused lasso for linear regression models as a prior distribution.

This study aimed to improve the spatial granularity of reliability assessment by combining a sparse-modeling-based approach with a Bayesian framework. We developed a new cluster detection method that extends the approach of Choi et al. [12] to the Bayesian framework through a prior distribution equivalent of the generalized fused lasso penalties and offers new information on the reliability of the clustering results. Analyses were performed using simulated distributions and real data of crime incidents to illustrate detection performance and improved reliability assessments of the proposed method.

3 of 14

2. Sparse-Modeling-Based Cluster Detection

Sparse-modeling-based cluster detection [12] leverages the parameter selection capability of the generalized fused lasso. This section first introduces the idea of the fused lasso and generalized fused lasso, followed by that of the cluster detection method.

2.1. Fused Lasso and Generalized Fused Lasso

The fused lasso is a sparse modeling method proposed by Tibshirani et al. [18] for detecting change points in time series data. Its core idea is to select parameters and identify consecutive pairs of parameters that share the same value. This is realized by introducing an L1 regularization term in both the parameter values and differences between consecutive pairs of parameters.

The minimization problem for a linear regression model with the fused lasso is formulated as:

$$\min_{\beta} \left[\| \boldsymbol{y} - \boldsymbol{X} \boldsymbol{\beta} \|_{2}^{2} + \lambda_{1} \sum_{i=1}^{p-1} |\beta_{i+1} - \beta_{i}| + \lambda_{2} \sum_{k=1}^{p} |\beta_{k}| \right]$$
(1)

where $\|\cdot\|_2$ is the L2 norm, $\boldsymbol{y} = (y_1, \dots, y_n)^\top$ is a dependent variable vector, and $\boldsymbol{X} = (x_1, \dots, x_p)$ is a design matrix; λ_1 and λ_2 are hyperparameters that govern the degrees of the L1 regularizations.

The generalized fused lasso extends the concept of the fused lasso by offering more flexibility to adjacency constraints. This extension broadened the applicable scope of sparse-modeling-based methods to spatial analysis [19,20].

The generalized fused lasso is written as:

$$\min_{\boldsymbol{\beta}} \left[\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_{2}^{2} + \lambda_{1} \sum_{(i,j) \in \mathcal{C}} |\beta_{i} - \beta_{j}| + \lambda_{2} \sum_{k=1}^{p} |\beta_{k}| \right]$$
(2)

where C is the set of adjacent pairs of parameters.

As the optimal values of the hyperparameters in Equations (1) and (2) are generally unknown, information criteria such as the Akaike information criterion (AIC) [21] or Bayesian information criterion (BIC) [22] are frequently used to compare and determine the combination of hyperparameters.

2.2. Sparse-Modeling-Based Cluster Detection

Choi et al. [12] formulated cluster detection from the spatial distribution of point events aggregated over small areas by introducing the generalized fused lasso penalty in the Poisson regression model. First, the number of point events recorded in a subregion i (i = 1, ..., n) is expressed as:

$$y_i \sim \text{Poisson}(\mu_i)$$
 (3)

$$\log E(y_i) = \log \mu_i = \log e_i + \alpha_i + \mathbf{x}^\top \boldsymbol{\beta}$$
(4)

where e_i is an offset term for subregion i, $x_i = (1, x_{i1}, ..., x_{ip})^\top$ is a covariate vector, and $\boldsymbol{\beta} = (\beta_0, ..., \beta_p)^\top$ is the corresponding parameter vector shared by the entire study region; $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)^\top$ denotes a vector consisting of subregion-based intensity parameters that represent the degree of concentration for each subregion. If the estimated value of α_i is equal to zero, then subregion *i* does not constitute a cluster, and if the estimated values of α_i are greater than zero, then subregion *i* constitutes a cluster.

Here, the Poisson likelihood function $L(\alpha, \beta | X, Y)$ and Poisson log-likelihood function $l(\alpha, \beta | X, Y)$ are given by:

$$L(\boldsymbol{\alpha},\boldsymbol{\beta}|\boldsymbol{X},\boldsymbol{Y}) = \prod_{i=1}^{n} \mu_{i}^{y_{i}} \frac{1}{y_{i}!} \exp(-\mu_{i})$$
(5)

$$l(\boldsymbol{\alpha}, \boldsymbol{\beta} | \boldsymbol{X}, \boldsymbol{Y}) = \sum_{i=1}^{n} y_i \log \mu_i - \mu_i - \log y_i!$$
(6)

where *X* and *Y* are sets of observed data defined as $X = (x_1, ..., x_n)^\top$ and $Y = (y_1, ..., y_n)^\top$, respectively.

By introducing the generalized fused lasso penalty to the Poisson log-likelihood function, the cluster detection problem of [12] can be formulated as:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[l(\boldsymbol{\alpha},\boldsymbol{\beta}|\boldsymbol{X},\boldsymbol{Y}) + \lambda_1 \sum_{(i,j)\in\mathcal{C}} |\alpha_i - \alpha_j| + \lambda_2 \sum_{k=1}^p |\alpha_k| + \lambda_3 \sum_{l=1}^p |\beta_l| \right]$$
(7)

where C is a set of adjacent pairs of parameters and λ_1 , λ_2 , and λ_3 are the hyperparameters.

Clusters likely constitute small portions of the entire study area and are made up of spatially contiguous subregions, which is considered in their detection. The generalized fused lasso penalty fits this purpose as the penalty imposes constraints on both the parameters themselves and adjacent pairs of parameters simultaneously. Therefore, Choi et al. [12] introduced the generalized fused lasso penalty to the subregion-based intensity parameters and realized cluster detection using the estimated intensity parameters.

As Equation (7) includes an L1 regularization term that is non-differentiable, Choi et al. [12] proposed a computational procedure that outputs point estimates using the MM algorithm [14]. The MM algorithm is a parameter estimation technique that updates the parameters iteratively using a surrogate function for the objective function.

3. Previous Studies on Sparsity-Inducing Priors

In the Bayesian framework, some of the prior distributions are referred to as sparsityinducing priors because of their ability to induce sparse solutions in the posterior distributions. In particular, parts of the sparsity-inducing priors can offer approximate point estimates of the lasso or its extensions, in addition to quantitative descriptions of the uncertainties in the form of probability distributions. This section first explains the Bayesian counterpart of the lasso and then that of the generalized fused lasso.

3.1. Bayesian Lasso

Tibshirani [15] first suggested that in linear regression models, placing the independent Laplace distributions as the prior distributions for the regression coefficients can shrink the posterior distributions toward zero and yield lasso estimates in the posterior modes. Following this implication, Park and Casella [16] proposed the Gibbs sampling formulation for Bayesian models with a Laplace distribution and called it the "Bayesian lasso". Equation (8) shows the Laplace distribution as a sparsity-inducing prior, as proposed in [16]:

$$\pi \left(\boldsymbol{\beta} \middle| \lambda, \sigma^2 \right) = \prod_{i=1}^p \frac{\lambda}{2\sqrt{\sigma^2}} \exp\left(-\lambda |\boldsymbol{\beta}_i| / \sqrt{\sigma^2}\right)$$
$$\pi \left(\sigma^2\right) = \frac{1}{\sigma^2} \tag{8}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^{\top}$ is the covariate vector and λ is a hyperparameter comparable to the regularization parameter in lasso. Park and Casella [16] stated that assuming an improper prior on σ^2 can avoid multiple posterior modes in some cases.

3.2. Bayesian Generalized Fused Lasso

Kyung et al. [17] extended the Bayesian lasso and proposed the Bayesian fused lasso for linear regression models. The formulated prior on the regression coefficients is given by:

$$\pi\left(\boldsymbol{\beta}\Big|\lambda_{1},\lambda_{2},\sigma^{2}\right) \propto \prod_{i=1}^{p-1} \exp\left(-\frac{\lambda_{1}}{\sigma}|\beta_{i+1}-\beta_{i}|\right) \prod_{j=1}^{p} \exp\left(-\frac{\lambda_{2}}{\sigma}|\beta_{j}|\right)$$
(9)

where λ_1 and λ_2 are hyperparameters.

Equation (9) can easily be extended to the Bayesian generalized fused lasso, whose formulation includes multiple Laplace distributions and is written as:

$$\pi\left(\boldsymbol{\beta}\Big|\lambda_{1},\lambda_{2},\sigma^{2}\right) \propto \prod_{(i,j)\in\mathcal{C}} \exp\left(-\frac{\lambda_{1}}{\sigma}|\boldsymbol{\beta}_{i}-\boldsymbol{\beta}_{j}|\right) \prod_{k=1}^{p} \exp\left(-\frac{\lambda_{2}}{\sigma}|\boldsymbol{\beta}_{k}|\right)$$
(10)

where C is the set of adjacent pairs of parameters.

Equations (9) and (10) correspond to the Bayesian version of the fused lasso and the generalized fused lasso, respectively.

4. Proposed Method

4.1. Likelihood and Prior Distributions

This study extends sparse-modeling-based cluster detection to the Bayesian framework. The extension offers information on the reliabilities of all the estimated parameters. Consider a study region consisting of *n* subregions. Let y_i be the number of point events recorded in subregion *i* (*i* = 1,...,*n*) and e_i be the offset term for subregion *i*. $x_i = (1, x_{i1}, ..., x_{ip})^\top$ is a covariate vector, and $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)^\top = (\beta_0, \tilde{\boldsymbol{\beta}}^\top)^\top$ is its corresponding parameter vector. Then, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)^\top$ denotes a vector consisting of subregion-based intensity parameters.

Under the assumption of a Poisson point process, the number of points y_i is given by the following Poisson regression model:

$$\log E(y_i) = \log \mu_i = \log e_i + \alpha_i + \mathbf{x}^\top \boldsymbol{\beta}$$
(11)

Subsequently, the Poisson likelihood function of Equation (11) can be written as:

$$\pi(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{\alpha},\boldsymbol{\beta}) = \prod_{i=1}^{n} \mu_{i}^{y_{i}} \frac{1}{y_{i}!} \exp(-\mu_{i})$$
(12)

where *X* and *Y* are sets of observed data defined by $X = (x_1, ..., x_n)^\top$ and $Y = (y_1, ..., y_n)^\top$, respectively.

Now, we define the joint prior distribution for the intensity parameter vector α and covariate vector $\tilde{\beta}$ as:

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\tilde{\beta}}|\lambda_1, \lambda_2, \lambda_3) = \prod_{(i,j)\in\mathcal{C}} \exp\{-\lambda_1|\alpha_i - \alpha_j|\} \prod_{i=1}^n \exp\{-\lambda_2|\alpha_i|\} \prod_{i=1}^p \exp\{-\lambda_3|\beta_i|\}$$
(13)

where C is the set of adjacent pairs of parameters and λ_1 , λ_2 , and λ_3 are the hyperparameters. In this study, adjacency is defined as a pair of subregions that share geographical borders.

Equation (13) consists of multiple Laplace distributions and achieves a sparse solution encouraged by the generalized fused lasso in the posterior mode. Additionally, a non-informative prior $\pi(\beta_0)$ is placed on β_0 . The posterior distribution derived by Equations (12) and (13) is equal to the penalized likelihood function proposed in [12] after logarithmic transformation. Therefore, the maximum a posteriori (MAP) estimators obtained from Equations (12) and (13) correspond to the estimated values of the cluster detection problem in [12].

4.2. Tuning Hyperparameters with the Watanabe–Akaike Information Criterion

The proposed prior distribution $\pi(\alpha, \hat{\beta}|\lambda_1, \lambda_2, \lambda_3)$ includes the hyperparameters λ_1 , λ_2 , and λ_3 , whose value should be set before parameter estimation. The Watanabe–Akaike information criterion (WAIC) [23], an information criterion proven to be suitable for comparing Bayesian models [24], was used in this study to select the optimal set of hyperparameters. Tuning the hyperparameters starts with predefining multiple candidate values for each hyperparameter. Then, Bayesian estimations are performed for all combinations of the candidate values, and the WAIC value is calculated. Thereafter, a set of candidate values that minimizes the WAIC value and passes a convergence test is adopted as the optimal hyperparameters.

5. Evaluation

5.1. Evaluations with Simulated Distributions

This section illustrates the characteristics of the proposed method using simulated distributions. A performance comparison is also presented between the proposed method and [12].

5.1.1. Overview

We considered a two-dimensional grid-like study region with a cluster located at the center of the region, as in Figure 1. The study region consists of 17×17 discrete subregions, and the cluster consists of 5×5 subregions. The objective of cluster detection is to determine the locations of the clusters and their shapes, that is to list the sets of possible subregions that constitute the clusters. Hereafter, the set of all subregions in the study region is represented as N, the set of subregions inside the cluster by C, and the set of subregions outside the cluster by C^C .



Figure 1. Two-dimensional grid-like study region with randomly generated count data. The red subregions in the center constitute a cluster.

Assuming a Poisson point process, we randomly generated count data (i.e., number of point events) for each subregion from a Poisson distribution. To simulate the existence of the central cluster that is characterized by a higher event occurrence, the expected number of point events in the subregions in C was adjusted to several times higher than that in C^C . This was achieved by changing the parameter of the Poisson distributions for 15 scenarios by choosing the expected number of points outside C from {10, 20, 30} and point density ratio (i.e., ratio of expected numbers of points inside a cluster to outside a cluster) from

{1.25, 1.5, 2.0, 2.5, 3.0}. For each scenario, this evaluation repeated the parameter estimation process with the hyperparameter candidates listed in Table 1. This simulation did not incorporate covariates because the primary focus was on assessing the regularization mechanism for spatially adjacent parameters during cluster detection.

The parameter estimation process is as follows: A Bayesian modeling platform Stan samples the posterior distributions for all parameters through the Hamiltonian Monte Carlo algorithm [25]. Each estimation process independently generates four Markov chain sequences. The total number of iterations was 2000 for all parameters, where the first 500 burn-in iterations were discarded. After all iterations were complete, the Gelman–Rubin statistic \hat{R} [26] was used to check convergence of the posterior distributions. The threshold value of \hat{R} was set to 1.1.

The sampled posterior distributions of the intensity parameters determine if each subregion constitutes a cluster. The threshold probability p was set in advance, and if the lower p percentage point value of an estimated intensity parameter assigned to a subregion exceeds zero, then that subregion is detected as constituting a cluster. In this study, p = 0.1 was adopted for the threshold probability because cluster classification with this threshold produced results comparable to those of Choi's method. From the clustering results for all subregions, we calculated two performance measures, namely the power and false-positive rate, which generally have a trade-off relationship. Their definitions are as follows:

$$Power = \frac{Number of detected subregions inside C}{Number of subregions inside C}$$
(14)
False-positive rate =
$$\frac{Number of detected subregions inside C^{C}}{Number of subregions inside C^{C}}$$
(15)

In each scenario, the measures were averaged after repeating the data generation and cluster detection steps 100 times. In addition, we compared the performance with the method in Choi et al. [12] ("Choi's method") for the same dataset. Table 2 shows the candidate hyperparameters for Choi's method. The optimal combination of hyperparameters was selected on the basis of the BIC. In Choi's method, if the point estimate of the intensity parameter assigned to a subregion is greater than zero, the subregion is considered to be detected.

Table 1. Candidate values for tuning the hyperparameters.

| Hyperparameters (Equation (13)) | Candidate Values | | |
|--|---|--|--|
| $egin{array}{c} \lambda_1 \ \lambda_2 \ \lambda_3 \end{array}$ | $\{10^{0.0}, 10^{0.5}, 10^{1.0}, 10^{1.5}, 10^{2.0}\}\$ $\{10^{-1.0}, 10^{-0.5}, 10^{0.0}, 10^{0.5}, 10^{1.0}\}\$ (Unnecessary because this evaluation introduces no covariates.) | | |

Table 2. Candidate hyperparameter values for Choi's method.

| Hyperparameters (Equation (7)) | Candidate Values |
|---|---|
| $\begin{matrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{matrix}$ | $\{10^{0.0}, 10^{0.5}, 10^{1.0}, 10^{1.5}, 10^{2.0}\}\$ $\{10^{-5.5}, 10^{-5.0}, \cdots, 10^{1.0}, 10^{1.5}\}\$ (Unnecessary because this evaluation introduces no covariates.) |

5.1.2. Results

To provide visualization examples of the clustering results, we first applied the proposed method to a single simulated distribution (Figure 2). The distribution was generated under conditions where the expected number of points outside a cluster was 10 and the point density ratio was 2.0.



Figure 2. Spatial distribution of the simulated data. Each subregion is colored according to the number of point event occurrences.

The clustering results shown in Figure 3 involve two visualizations: estimated subregionbased coefficients and their standard deviations. In particular, Figure 3b is based on the proposed method. In Bayesian estimation, the standard deviation of a parameter is an indicator of the reliability of the estimated parameter results. Figure 3 confirms that the reliabilities were relatively lower in subregions closer to the boundaries of a cluster, thus implying that the cluster boundaries were not adequately identified from the given count data. These visualizations suggest that the proposed method provides information on the reliability of the detection results.



Figure 3. Results of the proposed method for Figure 2. (a) Estimated subregion-based coefficients. The color of each subregion corresponds to the degree of concentration, and subregions enclosed within bold lines are detected as constituting a cluster. (b) Standard deviation of the estimated subregion-based coefficients. The darker colors represent lower reliabilities for the estimated coefficients in the corresponding subregions.

Performance measures of the proposed and Choi's methods are summarized in Tables 3 and 4. Table 3 shows the power, and Table 4 shows the false-positive rates for all 15 scenarios. In both methods, the power (Table 3) increased similarly as the expected numbers of points outside the cluster or point density ratio increased and reached a value close to 1.000 when the point density ratio exceeded 2.0.

The false-positive rate remained low between 0.000 and 0.030, thereby showing the similarity between the two methods. The above results confirm that the proposed method detects clusters with an accuracy comparable to Choi's method and further provide new information on the probabilistic variations of clustering results.

Focusing on the differences of detection performances, we see that the false-positive rate (Table 4) for the proposed method did not decrease when the point density ratio

increased and the cluster was detached from the background, as opposed to Choi's method. This difference was attributed to the information criteria adopted in the respective methods. The BIC used in Choi's method favors a strongly penalized solution as its value improves if a single parameter or the difference of adjacent pairs of parameters is estimated to be zero. In contrast, the WAIC adopted in the proposed Bayesian model does not prefer a strongly penalized solution as simply estimating the parameters as zero does not improve the criterion. The weakly penalized solution determined by the WAIC possibly results in a constant false-positive rate for the proposed model, where the point density ratio is large. Given these differences between the two methods, this comparison confirms that the proposed method can detect clusters as accurately as Choi's method in the overall sense.

| Method | Expected Points Outside a Cluster | Point Density Ratio | | | | |
|-----------------|-----------------------------------|---------------------|-------|-------|-------|-------|
| | | 1.25 | 1.5 | 2.0 | 2.5 | 3.0 |
| | 10 | 0.035 | 0.733 | 0.997 | 1.000 | 1.000 |
| Choi's method | 20 | 0.096 | 0.927 | 0.999 | 1.000 | 1.000 |
| | 30 | 0.153 | 0.994 | 1.000 | 1.000 | 1.000 |
| Proposed method | 10 | 0.071 | 0.599 | 0.964 | 0.998 | 1.000 |
| | 20 | 0.258 | 0.882 | 0.998 | 1.000 | 1.000 |
| | 30 | 0.470 | 0.942 | 1.000 | 1.000 | 1.000 |

Table 3. Evaluation results for the power.

Table 4. Evaluation results for the false-positive rate.

| Mathad | Europeted Paints Outside a Chuston | | Point Density Ratio | | | | |
|-----------------|------------------------------------|-------|---------------------|-------|-------|-------|--|
| wiethod | Expected Points Outside a Cluster | 1.25 | 1.5 | 2.0 | 2.5 | 3.0 | |
| | 10 | 0.003 | 0.018 | 0.020 | 0.026 | 0.013 | |
| Choi's method | 20 | 0.003 | 0.011 | 0.006 | 0.003 | 0.001 | |
| | 30 | 0.006 | 0.009 | 0.008 | 0.008 | 0.012 | |
| | 10 | 0.005 | 0.009 | 0.018 | 0.020 | 0.017 | |
| Proposed method | 20 | 0.006 | 0.012 | 0.015 | 0.014 | 0.014 | |
| | 30 | 0.009 | 0.018 | 0.018 | 0.017 | 0.019 | |

5.2. Evaluations with Real-World Data

To examine the proposed method in practical settings, this section presents application of the proposed method to real-world crime data.

5.2.1. Target Area and Data Description

The target area of this analysis (Figure 4) was the central Tokyo region, which comprises five municipalities, namely Chiyoda, Chuo, Shinjuku, Minato, and Shibuya. The point event dataset used in this analysis was non-intrusive theft data from 2019. This point event dataset was chosen because one of the typical applications of cluster detection is crime analysis [8,27]. This publicly available dataset records the number of non-intrusive theft cases recognized by local police agencies for 2019 at the subregion level ("Cho" in Japanese). As shown in Figure 5, the target area has 546 districts, and the number of crimes in the corresponding area totals 12,396 cases. The average number is 22.7 cases per district, with the lowest being zero and highest being 837. This analysis used the areas of the districts for the offsets and number of employees engaged in the retail sector as the covariate.



Figure 4. Target area and its spatial partition into district units. Railway networks and several large stations are also shown on the map.



Figure 5. Choropleth map showing non-intrusive theft occurrences in the target area.

5.2.2. Estimation Settings

As with the simulation analysis in the previous section, practical analysis was performed to estimate the parameters via Monte Carlo sampling on Stan. The estimation process generated four Markov chain sequences independently. The total number of iterations per sequence was 2000, with the first 500 burn-in iterations being discarded. The threshold value of the Gelman–Rubin statistic was set to 1.1. Table 5 shows the candidate values of the hyperparameters. This analysis contained repeated estimations with all possible combinations of the candidate values and adopted the WAIC-minimizing combination for the optimal hyperparameters.

| Hyperparameters (in Equation (13)) | Candidate Values |
|------------------------------------|--|
| λ_1 | $\{10^{-1.5}, \mathbf{10^{-1.0}}, 10^{-0.5}, 10^{0.0}, 10^{0.5}, 10^{1.0}, 10^{1.5}, 10^{2.0}\}$ |
| λ_2 | $\{10^{-1.5}, 10^{-1.0}, \mathbf{10^{-0.5}}, 10^{0.0}, 10^{0.5}, 10^{1.0}, 10^{1.5}, 10^{2.0}\}$ |
| λ_3 | $\{10^{-1.5}, 10^{-1.0}, \mathbf{10^{-0.5}}, 10^{0.0}, 10^{0.5}, 10^{1.0}, 10^{1.5}, 10^{2.0}\}$ |

Table 5. Candidate values for tuning the hyperparameters. The numbers in bold font indicate the optimal combination in this example.

5.2.3. Results

Figures 6 and 7 show the estimated subregion-based coefficients and their standard deviations, respectively. We set the same threshold p = 0.1 as in the previous evaluation with simulated distributions. The proposed method detected 166 districts as clusters. The estimated parameter of the covariate was 0.30, and its standard deviation was 0.08. Figure 6 shows the spatial distribution of the clusters after adjusting for covariates. First, we confirmed that the bustling downtowns near Shinjuku and Shibuya stations were distinctly detected. Although this observation is easily assumed when we first see the choropleth map and may not be notable, Figure 6 detects several neighborhoods around Roppongi station that are not sufficiently highlighted in the choropleth map. This indicates the possible existence of spatially concentrated crime hotspots where the crime counts are relatively low. Figure 7 confirms that the proposed method estimated the standard deviations of subregion-based coefficients for different values across the target area. The spatial unevenness of the estimated standard deviations suggests the capability of the proposed method to identify spatially varying reliabilities in real-world data and enhance data interpretation.



Figure 6. Estimated subregion-based coefficients. Subregions bounded by bold lines are detected as constituting clusters.



Figure 7. Standard deviations of estimated subregion-based coefficients.

5.3. Discussion

This study evaluated the proposed method from the viewpoint of detection performance and reliability assessment capabilities. The first evaluation with the simulated dataset primarily compared the detection performance with Choi's method. This comparison confirmed that the proposed method maintains comparable performance to the baseline method. The following evaluation with the real-world crime dataset confirmed that the proposed method outputs the quantified reliability for each subregion.

These results support that the proposed method advances Choi's method by newly providing reliability assessments at the subregion level, the minimal spatial unit of analysis. The real-world application demonstrated that the reliability of detection results could be spatially varying at the subregion level rather than uniform. This fact can affect the interpretation of the results and thereby reveals the significance of reliability assessments with sufficient spatial granularity. However, as most existing detection methods including Choi's method provide no or spatially aggregated representation of reliability information, these findings might be overlooked without the proposed method. Thus, this study contributed to the improvement of reliability assessments, a crucial element of cluster detection, from the aspect of spatial granularity.

6. Conclusions

This study proposed a Bayesian cluster detection method that can provide information on the reliability of the clustering results. The proposed method extends the sparsemodeling-based cluster detection approach formulated by Choi et al. [12] to the Bayesian framework. The extension was achieved by constructing a Bayesian model using multiple sparsity-inducing priors to encourage a sparse solution equivalent to that obtained by the above work. This study first formulated the likelihood function and prior distribution of the proposed Bayesian model with mathematical equivalence to the model by Choi et al. [12] Then, analyses were performed using simulated distributions and real-world crime data of central Tokyo. The simulation analysis revealed that the proposed method could quantify the reliability of clustering results at the subregion level and detect clusters with an accuracy comparable to that of the previous work for most of the scenarios evaluated. The crime data analyses confirmed that the proposed method could capture spatially varying reliability in real-world data. These findings underpin the validity and contribution of the present study to the cluster detection methods by improving the spatial granularity of reliability assessments. This improvement led to obtaining a more reliable interpretation of the clusters.

The proposed method can be improved to incorporate the temporal nature of the point event data. Although the present study only considered the geographical aspect of the phenomena, the proposed method can model adjacency in a flexible fashion. This extension should be useful, especially for phenomena that change rapidly along time.

There are also possible improvements that may contribute to better performance. Fan and Li [28] demonstrated that lasso estimators are biased towards zero; as the proposed method introduces a Laplace distribution, a Bayesian counterpart of the lasso, biased estimations may worsen the detection accuracy. In the Bayesian context, some studies have proposed alternative sparsity-inducing priors for the Laplace distribution. Notably, the normal-exponential-gamma (NEG) distribution [29] and Horseshoe prior [30,31] have been theoretically shown to alleviate the biases of the lasso estimators. Adopting these instead of the Laplace distribution may be a worthy line of future investigations.

Author Contributions: Conceptualization, Ryo Masuda; methodology, Ryo Masuda; software, Ryo Masuda; validation, Ryo Masuda; formal analysis, Ryo Masuda; investigation, Ryo Masuda; resources, Ryo Inoue; data curation, Ryo Masuda; writing—original draft preparation, Ryo Masuda; writing—review and editing, Ryo Inoue and Ryo Masuda; visualization, Ryo Masuda; supervision, Ryo Inoue; project administration, Ryo Inoue; funding acquisition, Ryo Inoue. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Japan Society for the Promotion of Science KAKENHI, Grant Numbers JP18H01552 and JP21H01447.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://www.keishicho.metro.tokyo.lg.jp/about_mpd/jokyo_tokei/jokyo/ninchikensu. html (accessed on 20 January 2022) and https://www.e-stat.go.jp/en/stat-search/files?page=1& toukei=00200553&tstat=000001095895 (accessed on 20 January 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Getis, A.; Ord, J.K. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* 1992, 24, 189–206. [CrossRef]
- 2. Anselin, L. Local indicators of spatial association—LISA. *Geogr. Anal.* 1995, 27, 93–115. [CrossRef]
- 3. Kulldorff, M.; Nagarwalla, N. Spatial disease clusters: Detection and inference. Stat. Med. 1995, 14, 799–810. [CrossRef] [PubMed]
- 4. Kulldorff, M. SaTScan v10.0.2: Software for the Spatial, Temporal, and Space-Time Scan Statistics. 2022. Available online: https://www.satscan.org/ (accessed on 25 February 2022).
- 5. Huang, L.; Kulldorff, M.; Gregorio, D. A spatial scan statistic for survival data. *Biometrics* 2007, 63, 109–118. [CrossRef] [PubMed]
- 6. Jung, I. Spatial scan statistics for matched case–control data. *PLoS ONE* **2019**, *14*, e0221225. [CrossRef]
- Takahashi, K.; Shimadzu, H. Detecting multiple spatial disease clusters: Information criterion and scan statistic approach. *Int. J. Health Geogr.* 2020, 19, 1–11. [CrossRef]
- 8. Duczmal, L.; Assuncao, R. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Comput. Stat. Data Anal.* **2004**, *45*, 269–286. [CrossRef]
- 9. Duczmal, L.; Cançado, A.L.; Takahashi, R.H.; Bessegato, L.F. A genetic algorithm for irregularly shaped spatial scan statistics. *Comput. Stat. Data Anal.* 2007, 52, 43–52. [CrossRef]
- Caldas de Castro, M.; Singer, B.H. Controlling the false discovery rate: A new application to account for multiple and dependent tests in local statistics of spatial association. *Geogr. Anal.* 2006, *38*, 180–208. [CrossRef]
- 11. Brunsdon, C.; Charlton, M. An assessment of the effectiveness of multiple hypothesis testing for geographical anomaly detection. *Environ. Plan. Des.* **2011**, *38*, 216–230. [CrossRef]
- 12. Choi, H.; Song, E.; Hwang, S.S.; Lee, W. A modified generalized lasso algorithm to detect local spatial clusters for count data. *AStA Adv. Stat. Anal.* **2018**, *102*, 537–563. [CrossRef]
- 13. Tibshirani, R.J.; Taylor, J. The solution path of the generalized lasso. Ann. Stat. 2011, 39, 1335–1371. [CrossRef]
- 14. Hunter, D.R.; Li, R. Variable selection using MM algorithms. Ann. Stat. 2005, 33, 1617–1642. [CrossRef] [PubMed]
- 15. Tibshirani, R.J. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B (Methodol.) 1996, 58, 267–288. [CrossRef]
- 16. Park, T.; Casella, G. The Bayesian lasso. J. Am. Stat. Assoc. 2008, 103, 681–686. [CrossRef]

- 17. Kyung, M.; Gill, J.; Ghosh, M.; Casella, G. Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal.* 2010, *5*, 369–411.
- 18. Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J.; Knight, K. Sparsity and smoothness via the fused lasso. J. R. Stat. Soc. Ser. B (Stat. Methodol.) 2005, 67, 91–108. [CrossRef]
- 19. Inoue, R.; Ishiyama, R.; Sugiura, A. Identification of geographical segmentation of the rental housing market in the Tokyo metropolitan area by generalized fused lasso. *J. Jpn. Soc. Civ. Eng. Ser. D3 (Infrastruct. Plan. Manag.)* 2020, 76, 251–263. (In Japanese) [CrossRef]
- 20. Inoue, R.; Ishiyama, R.; Sugiura, A. Identifying local differences with fused-MCP: An apartment rental market case study on geographical segmentation detection. *Jpn. J. Stat. Data Sci.* 2020, *3*, 183–214. [CrossRef]
- 21. Akaike, H. A new look at the statistical model identification. IEEE Trans. Autom. Control. 1974, 19, 716–723. [CrossRef]
- 22. Schwarz, G. Estimating the dimension of a model. Ann. Stat. 1978, 6, 461–464. [CrossRef]
- 23. Watanabe, S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **2010**, *11*, 3571–3594.
- 24. Gelman, A.; Hwang, J.; Vehtari, A. Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **2014**, 24, 997–1016. [CrossRef]
- 25. Duane, S.; Kennedy, A.D.; Pendleton, B.J.; Roweth, D. Hybrid monte carlo. Phys. Lett. B 1987, 195, 216–222. [CrossRef]
- 26. Gelman, A.; Rubin, D.B. Inference from iterative simulation using multiple sequences. Stat. Sci. 1992, 7, 457–472. [CrossRef]
- 27. Shiode, S. Street-level spatial scan statistic and STAC for analysing street crime concentrations. *Trans. GIS* **2011**, *15*, 365–383. [CrossRef]
- 28. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* 2001, *96*, 1348–1360. [CrossRef]
- 29. Griffin, J.E.; Brown, P.J. Bayesian hyper-lassos with non-convex penalization. Aust. N. Z. J. Stat. 2011, 53, 423–442. [CrossRef]
- 30. Carvalho, C.M.; Polson, N.G.; Scott, J.G. Handling sparsity via the horseshoe. J. Mach. Learn. Res. 2009, 5, 73-80.
- 31. Carvalho, C.M.; Polson, N.G.; Scott, J.G. The horseshoe estimator for sparse signals. Biometrika 2010, 97, 465–480. [CrossRef]