ORIGINAL PAPER

Spatio-temporal clustering analysis using generalized lasso with an application to reveal the spread of Covid-19 cases in Japan

Septian Rahardiantoro[1,2] · Wataru Sakamoto[1]

Abstract

This study addressed the issue of determining multiple potential clusters with regularization approaches for the purpose of spatio-temporal clustering. The generalized lasso framework has flexibility to incorporate adjacencies between objects in the penalty matrix and to detect multiple clusters. A generalized lasso model with two L_1 penalties is proposed, which can be separated into two generalized lasso models: trend filtering of temporal effect and fused lasso of spatial effect for each time point. To select the tuning parameters, the approximate leave-one-out cross-validation (ALOCV) and generalized cross-validation (GCV) are considered. A simulation study is conducted to evaluate the proposed method compared to other approaches in different problems and structures of multiple clusters. The generalized lasso with ALOCV and GCV provided smaller MSE in estimating the temporal and spatial effect compared to unpenalized method, ridge, lasso, and generalized ridge. In temporal effects detection, the generalized lasso with ALOCV and GCV provided relatively smaller and more stable MSE than other methods, for different structure of true risk values. In spatial effects detection, the generalized lasso with ALOCV provided higher index of edges detection accuracy. The simulation also suggested using a common tuning parameter over all time points in spatial clustering. Finally, the proposed method was applied to the weekly Covid-19 data in Japan form March 21, 2020, to September 11, 2021, along with the interpretation of dynamic behavior of multiple clusters.

Keywords Clustering · Cross-validation · Generalized lasso · Covid-19 · Spatio-temporal effect

✉ Septian Rahardiantoro
   septianrahardiantoro@apps.ipb.ac.id

1 Department of Human Ecology, Graduate School of Environmental and Life Science, Okayama University, Okayama 700-8350, Japan

2 Department of Statistics, Faculty of Mathematics and Natural Science, IPB University, Bogor 16680, Indonesia

1 Introduction

Spatio-temporal data consists of information about objects or events located in space over a period of time. Ansari et al. (2020) classified spatio-temporal data into five types and provided a review on spatio-temporal clustering. Our study focuses on geo-referenced time series clustering, which aims to identify dynamic behavior of clusters of objects over time. The preceding literature on geo-referenced time series clustering includes fuzzy clustering method (Izakian et al. 2013, 2015), NeuCube spiking neural network architecture for brain data (Doborjeh and Kasabov 2015; Doborjeh et al. 2018), and Correlation-based Clustering of Big Spatiotemporal Datasets (CorClustST) (Husch et al. 2020).

Recently, regularization approaches have attracted attention in the spatio-temporal analysis. One approach that has great potential is the generalized lasso. The generalized lasso (Tibshirani and Taylor 2011; Arnold and Tibshirani 2016), a general form of lasso (Tibshirani 1996), makes constraints on regression coefficients based on the general structure or geometry using the L_1 penalty. Let $y \in \mathbb{R}^n$ be a response vector, $X \in \mathbb{R}^{n \times p}$ be a predictor matrix, and $\theta \in \mathbb{R}^p$ be a parameter vector. Then the generalized lasso can be formulated as

$$\arg \min_{\theta} \left\{ \|y - X\theta\|_2^2 + \lambda \|D\theta\|_1 \right\}, \tag{1}$$

where $\lambda (\geq 0)$ is a tuning parameter, and $D \in \mathbb{R}^{m \times p}$ is a penalty matrix, of which each row constructs a linear combination of θ to define the desired structural or geometric property of the problem. If $D = I$, then the problem (1) becomes the ordinary lasso.

The generalized lasso has various applications by considering different forms of the penalty matrix D and the predictor matrix X in the model. If we specify the predictor matrix as $X = I$, then (1) becomes the coefficient smoothing problem, widely known as the fused lasso (Tibshirani et al. 2005; Tibshirani and Wang 2008), trend filtering (Kim et al. 2009; Tibshirani 2014), and the wavelet smoothing (Donoho and Johnstone 1995), according to the specified structure in D. In contrast, in the case $X \neq I$, the applications are extended to the modeling problems, such as a modeling for MRI image data (Tibshirani and Taylor 2011), spatially varying coefficient models (Zhao and Bondell 2020; Rahardiantoro and Sakamoto 2021, 2022b), and outlier detection (She and Owen 2010).

The generalized lasso has been applied to spatial data and time series data by determining the penalty matrix D appropriately. For spatial clustering analysis, a special form of the generalized lasso is the fused lasso on an irregular graph (Tibshirani and Taylor 2011; Arnold and Tibshirani 2016). In this case, the penalty matrix D shows the structure of the graph, so that its each row corresponds to the difference of coefficients between each pair of nodes connected by an edge. A collection of nodes on which the coefficients are estimated as common is considered to form a cluster. An application of the generalized lasso to the time series is the trend filtering (Tibshirani 2014). In this case, the penalty matrix D contains discrete difference operators of a specified order, that is, the first-order difference for estimating a piecewise constant structure, the second-order difference for estimating a piecewise linear, etc.

In the preceding literature on spatio-temporal clustering, ordinary lasso approaches have been mainly used in combination with existing clustering methods. Kamenetsky et al. (2022) proposed the lasso approach to detect the potential cluster using a scan statistic by implementing the sparse matrix representation of the effects of potential clusters. Chen et al. (2018) built separate lasso sub-models at each time point to detect influenced predictors for different historical lags up to 8-time points and included the neighborhood between objects in the specified radius as one of the predictors. However, these methods have limitation in determining multiple potential clusters, because they are highly dependent on the specified radius of the neighborhood.

In this study, we propose a more flexible approach for spatio-temporal clustering, using the generalized lasso framework with two L_1 penalties, in which one penalty corresponds to roughness on the temporal scale, and the other penalty for fusion of adjacent locations at each time point. The proposed model can be separated into the two generalized lasso problems: trend filtering on the temporal scale and fused lasso for spatial clustering at each time point. In the trend filtering problem, smoothed temporal pattern is estimated from the average value over all locations at each time point. In the fused lasso problem, clusters are constructed at each time point and their relative magnitude can be compared. Therefore, our proposed method can reveal dynamic behavior of spatial clusters as time proceeds. One advantage of our proposed method is its flexibility, that is, we can incorporate adjacencies between objects in the penalty matrix, and it is possible to detect multiple clusters.

An essential aspect to obtain appropriate estimates of parameters is to select the optimum tuning parameter. The most common method is the k-fold cross-validation. For example, Zhao and Bondell (2020) applied the 10-fold cross-validation to select the tuning parameter in the generalized lasso problem. However, it is known that the k-fold cross-validation suffers from large biases in estimation of the out-of-sample prediction error (Rad and Maleki 2020; Rad et al. 2020). In addition, spatio-temporal data have time-dependent and neighbor-dependent structures, so splitting such data into the test and training sets may fail in estimating some coefficients depending on the structures of the penalty matrix D. The leave-one-out cross-validation (LOOCV), which is the case $k = n$, can reduce the bias in estimating the out-of-sample prediction error, but it requires high computational cost. We consider using the approximate leave-one-out cross-validation (ALOCV) (Rad and Maleki 2020), and its generalized cross-validation (GCV) (Craven and Wahba 1979) version. The ALOCV and GCV provide an approximation of the leave-one-out predicted values based on the primal and dual formulations of the general regularization problems, and can be used formally without breaking the structures of penalty matrix D.

Then, we apply the proposed method to weekly Covid-19 data in Japan as a real data application. Many studies on spatio-temporal clustering intended to reveal the pattern of Covid-19 outbreaks in some countries, such as, in Brazil (Castro et al. 2021), in the United States (Wang et al. 2021b), and in China (Wang et al. 2021a). Takemura et al. (2022) applied an adjusted Echelon scan method to detect multiple space–time clusters of daily Covid-19 cases in Japan, in which they detected time intervals and regions with significantly higher risk of infections than their surrounding ones, and considered the factors that caused them

and affected the changes in a cluster's shape. In contrast, our proposed method can reveal the overall trend of the temporal effect and detect dynamic behavior of spatial clusters.

The motivation for this study lies in three aspects, based on its purpose, contribution to the generalized lasso studies, and real data application. The first is to propose an approach to estimate temporal effect and detect multiple clusters by using generalized lasso. Secondly, this paper contributes to the ongoing generalized lasso application by extending the application to spatio-temporal data, which can be used to estimate the smoothed temporal pattern and identify dynamic behavior of spatial clusters according to adjacent locations over time. Finally, we are interested in revealing the pattern of temporal effect and spatial clusters weekly in the Covid-19 case in Japan. Our source code is available via Supplementary Information which can be accessed online at the link https://github.com/Rahardiantoro/Spatiotemporal-Generalized-Lasso-.

The sections are arranged as follows. Section 2 explains our proposed method on the generalized lasso for spatio-temporal clustering. Section 3 contains the methods for selecting the optimum tuning parameter. In Sect. 4, we perform the simulation study to investigate the performances of the proposed method compared to some existing methods. Section 5 contains the real data application of the Covid-19 data in Japan. Finally, Sect. 6 is the conclusion of this study.

2 The generalized lasso for spatio-temporal clustering

We explain our proposed method for applying to spatio-temporal data by combining two types of the generalized lasso, the trend filtering on the temporal scale and the fused lasso on an irregular graph for spatial clustering. We consider the spatio-temporal observations as y_{it} with locations indexed by $i = 1, 2, \ldots, S$ and time points indexed by $t = 1, 2, \ldots, T$. We represent y_{it}, using temporal effect α_t and the spatial effect β_{it} at each time point, as a linear model

$$y_{it} = \alpha_t + \beta_{it} + \varepsilon_{it}, \ i = 1, 2, \ldots, S, \ t = 1, 2, \ldots, T, \tag{2}$$

where ε_{it} indicates the noise at i-th location and t-th time point. To estimate $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)^T$ and $\boldsymbol{\beta}_t = (\beta_{1t}, \ldots, \beta_{St})^T$, we use the regularization method to obtain the smoothed of temporal effect and clusters in space over time points, by minimizing

$$\sum_{t=1}^{T} \sum_{i=1}^{S} (y_{it} - \alpha_t - \beta_{it})^2 + \lambda_T P_T(\boldsymbol{\alpha}) + \sum_{t=1}^{T} \lambda_{S,t} P_S(\boldsymbol{\beta}_t), \tag{3}$$

where $P_T(\boldsymbol{\alpha})$ and $P_S(\boldsymbol{\beta}_t)$ indicate the penalty terms of temporal effect and spatial effect, respectively, with corresponding tuning parameters λ_T and $\lambda_{S,t}$. In this case, we use the L_1 penalty term for 1-dimensional trend filtering (Tibshirani and Taylor 2011)

$$P_T(\boldsymbol{\alpha}) = \sum_{t=3}^{T} |\alpha_t - 2\alpha_{t-1} + \alpha_{t-2}|, \tag{4}$$

and the L_1 penalty term for the fused lasso on the graph (Tibshirani and Wang 2008; Tibshirani and Taylor 2011)

$$P_S(\boldsymbol{\beta}_t) = \sum_{(i,j)\in\mathcal{E}} \left|\beta_{it} - \beta_{jt}\right|, \tag{5}$$

where $\mathcal{E}$ is the set of edges on the graph defining adjacency.

We can rewrite the first term in (3) as

$$\sum_{t=1}^{T}\sum_{i=1}^{S} (y_{it} - \alpha_t - \beta_{it})^2 = \sum_{t=1}^{T}\sum_{i=1}^{S} \left(y_{it} - \bar{y}_{\cdot t} - \beta_{it} - \bar{\beta}_{\cdot t}\right)^2 + S\sum_{t=1}^{T} \left(\bar{y}_{\cdot t} - \alpha_t - \bar{\beta}_{\cdot t}\right)^2, \tag{6}$$

where $\bar{\beta}_{\cdot t} = S^{-1}\sum_{i=1}^{S}\beta_{it}$. If we put the constraint $\bar{\beta}_{\cdot t} = 0$ for identifiability of the parameters, the Eq. (6) can be expressed as

$$\sum_{t=1}^{T}\sum_{i=1}^{S} (y_{it} - \alpha_t - \beta_{it})^2 = \sum_{t=1}^{T}\sum_{i=1}^{S} \left(y_{it} - \bar{y}_{\cdot t} - \beta_{it}\right)^2 + S\sum_{t=1}^{T} \left(\bar{y}_{\cdot t} - \alpha_t\right)^2. \tag{7}$$

Thus, we can express the problem (3) as

$$\sum_{t=1}^{T}\sum_{i=1}^{S} \left(y_{it} - \bar{y}_{\cdot t} - \beta_{it}\right)^2 + S\sum_{t=1}^{T} \left(\bar{y}_{\cdot t} - \alpha_t\right)^2 + \lambda_T P_T(\boldsymbol{\alpha}) + \sum_{t=1}^{T} \lambda_{S,t} P_S(\boldsymbol{\beta}_t). \tag{8}$$

Therefore, we can solve the problem of minimizing (8) as separated minimization on the temporal effect and the spatial effects over time, that is, for estimating $\boldsymbol{\alpha}$ we can only minimize

$$S\sum_{t=1}^{T} \left(\bar{y}_{\cdot t} - \alpha_t\right)^2 + \lambda_T P_T(\boldsymbol{\alpha}), \tag{9}$$

as a 1-dimensional trend filtering problem, and for estimating $\boldsymbol{\beta}_t(t = 1, 2, \ldots, T)$, we can only minimize, for each $t = 1, .., T$

$$\sum_{i=1}^{S} \left(y_{it} - \bar{y}_{\cdot t} - \beta_{it}\right)^2 + \lambda_{S,t} P_S(\boldsymbol{\beta}_t). \tag{10}$$

as a fused lasso problem on the graph. In this study, both the problems (9) and (10) have the form of the generalized lasso (1), and we can apply the R package "genlasso" (Arnold and Tibshirani 2016).

3 Methods for selecting optimum tuning parameters

LOOCV evaluates the mean square prediction error of each one observation in fitting the model using training set of rest $n - 1$ observations (Stone 1974). In the context of the generalized lasso problem (1), the LOOCV error for a specified λ can be stated as

$$LOOCV(\lambda) = \frac{1}{n} \sum_{c=1}^{n} \left(y_c - \boldsymbol{x}_c^T \widehat{\boldsymbol{\theta}}^{/c} \right)^2, \tag{11}$$

where $\widehat{\boldsymbol{\theta}}^{/c}$ represents the leave-one-out estimate of $\boldsymbol{\theta}$ when the c-th observation is omitted.

In the case of generalized lasso using the L_1 penalty, we have no exact explicit form of the leave-one-out estimate $\widehat{\boldsymbol{\theta}}^{/c}$ in (11), and solving the generalized lasso problem for each c requires high computational cost. However, we can apply the approximate leave-one-out cross-validation (ALOCV) to reduce computation time, which is based on the primal and dual formulations of non-differentiable regularization problems (Wang et al. 2018; Rad and Maleki 2020). In the context of generalized lasso problem (1), for each given λ, the algorithm of ALOCV can be described as follows (Wang et al. 2018).

(a) Estimate $\boldsymbol{\theta}$ as a solution of the primal problem (1).
(b) Estimate $\boldsymbol{u}$ as a solution of the dual problem of (1), which can be expressed as:

$$\arg \min_{\boldsymbol{\gamma}, \boldsymbol{u}} \frac{1}{2} \|\boldsymbol{\gamma} - \boldsymbol{y}\|_2^2 \, s.t. \|\boldsymbol{u}\|_\infty \leq \lambda \quad \text{and} \quad \boldsymbol{X}^T \boldsymbol{\gamma} = \boldsymbol{D}^T \boldsymbol{u}. \tag{12}$$

(c) Remove the rows of $\boldsymbol{D}$ belonging to the index set $E = \{ s = 1, \ldots, m : |\widehat{u}_s| = \lambda \}$, to construct a submatrix $\boldsymbol{D}_{-E}$.
(d) Construct the matrix $\boldsymbol{A} = \boldsymbol{XB}$, where $\boldsymbol{B}$ has columns span the null space of $\boldsymbol{D}_{-E}$.
(e) Compute $\boldsymbol{H}^* = \boldsymbol{A}\boldsymbol{A}^+$, where $\boldsymbol{A}^+$ represents the Moore–Penrose pseudoinverse of $\boldsymbol{A}$.
(f) Calculate the ALOCV error as

$$\frac{1}{n} \sum_{c=1}^{n} \left(\frac{y_c - \boldsymbol{x}_c^T \widehat{\boldsymbol{\theta}}}{1 - h_{cc}^*} \right)^2, \tag{13}$$

where h_{cc}^* is the c-th diagonal component of $\boldsymbol{H}^*$.

Then, the optimum tuning parameter λ can be selected as the one minimizing ALOCV error (13). Our simulation study (Rahardiantoro and Sakamoto 2022a) suggested that, in the context of spatial clustering with spatially varying coefficient models, the ALOCV could yield slightly smaller out-of-sample prediction error and could detect edges in a graph with differences shrunk more appropriately, compared to k-fold cross-validation.

In practical computation, we may fail to obtain the ALOCV error for very small λ values. Since $h_{cc}^* \to 1$ as $\lambda \to 0$, the denominator $1 - h_{cc}^*$ for some c in (13) may become close to zero, and then computation of ALOCV may be unstable, as

illustrated in our application to real data (Sect. 5). Rad and Maleki (2020) suggested the generalized cross-validation (GCV) approach, which is to approximate as $h_{cc}^* \approx tr\left(\boldsymbol{H}^*\right)/n$, to obtain the following score

$$\frac{1}{n} \sum_{c=1}^{n} \left(\frac{y_c - \boldsymbol{x}_c^T \widehat{\boldsymbol{\theta}}}{1 - tr\left(\boldsymbol{H}^*\right)/n} \right)^2. \tag{14}$$

However, its performance does not seem to have been well investigated. We also adopt the GCV approach in our simulation study and application to real data.

4 Simulation study

In this simulation study, we investigate the performance of our proposed method with generalized lasso compared to some existing regularization methods. The problem of minimizing (3) consists of two penalties, but as explained in the Sect. 2, it can be separated into the two generalized lasso problems with each single penalty. Therefore, we compare our proposed methods with the regularization methods which consist of single penalty, such as lasso (Tibshirani 1996), ridge (Hoerl and Kennard 1970), and generalized ridge (Zhao and Bondell 2020). Table 1 shows the corresponding penalties for $P_T(\boldsymbol{\alpha})$ and $P_S\left(\boldsymbol{\beta}_t\right)$.

For identifiability issues in lasso and ridge, we make some groups of adjacent time points (or spatial locations), and pool temporal effects α_t (or the spatial effects β_{it}) in the same group. Let $\boldsymbol{\alpha}^* = \left(\alpha_1^*, \ldots, \alpha_{T_1}^* \right)^T$ be the vector of pooled temporal effects, $\bar{\boldsymbol{y}} = \left(\bar{y}_{\cdot 1}, \ldots, \bar{y}_{\cdot T} \right)^T$, and $X_1 \in \mathbb{R}^{T \times T_1}$ be a block-diagonal predictor matrix with a vector ones for each block, representing how the elements are pooled. For example, suppose that we have 12 time points, grouped into $T_1 = 3$ groups, each containing 4 adjacent time points. In this case, $\alpha_1^*, \alpha_2^*, \alpha_3^*$ are the coefficients for group 1, 2, and 3, respectively, and the predictor matrix can be stated as $X_1 = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}$, where $\boldsymbol{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ and $\boldsymbol{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ are vectors of length 4.

For ridge problems, we can obtain the solution in the close form. Then, the problem of minimizing (9) using the ridge penalty for temporal effect is rewritten as

Table 1 The penalties $P_T(\boldsymbol{\alpha})$ and $P_S\left(\boldsymbol{\beta}_t\right)$ used in the simulation study

Type	$P_T(\boldsymbol{\alpha})$	$P_S\left(\boldsymbol{\beta}_t\right)$
Generalized lasso	$\sum_{t=3}^{T} \lvert \alpha_t - 2\alpha_{t-1} + \alpha_{t-2} \rvert$	$\sum_{(i,j) \in \mathcal{E}} \lvert \beta_{it} - \beta_{jt} \rvert$
Generalized ridge	$\sum_{t=3}^{T} \left(\alpha_t - 2\alpha_{t-1} + \alpha_{t-2} \right)^2$	$\sum_{(i,j) \in \mathcal{E}} \left(\beta_{it} - \beta_{jt} \right)^2$
Lasso	$\sum_{t=1}^{T_1} \lvert \alpha_t \rvert$	$\sum_{\forall i}^{S_1} \lvert \beta_{it} \rvert$
Ridge	$\sum_{t=1}^{T_1} \left(\alpha_t \right)^2$	$\sum_{\forall i}^{S_1} \left(\beta_{it} \right)^2$

$$S\left\|\bar{y} - X_1 \alpha^*\right\|_2^2 + \lambda_T \left\|\alpha^*\right\|_2^2, \tag{15}$$

and the solution of minimizing (15) is $\widehat{\alpha^*} = \left(X_1^T X_1 + \lambda_T I\right)^{-1} X_1^T \bar{y}$.

Similarly, let $\boldsymbol{\beta}_t^* = \left(\beta_{1t}^*, \ldots, \beta_{S_1 t}^*\right)^T$ be the vector of pooled spatial effect, $\tilde{y}_t$ be the vector of $y_{it} - \bar{y}_{\cdot t}$, and $X_2 \in \mathbb{R}^{S \times S_1}$ be a block-diagonal predictor matrix representing how the elements are pooled. Then, the problem of minimizing (10) using the ridge penalty for spatial effect over time is rewritten as

$$\left\|\tilde{y}_t - X_2 \boldsymbol{\beta}_t^*\right\|_2^2 + \lambda_{S,t} \left\|\boldsymbol{\beta}_t^*\right\|_2^2 \tag{16}$$

for $t = 1, \ldots, T$, and the solution of minimizing (16) is $\widehat{\boldsymbol{\beta}_t^*} = \left(X_2^T X_2 + \lambda_{S,t} I\right)^{-1} X_2^T \tilde{y}_t$. For lasso problems, the penalties in (15) and (16) are replaced with L_1 penalties. In this simulation study, we used the R package "glmnet" to solve the lasso.

In the generalized ridge problem, we can also obtain the solution in the close form as follows. The problem of minimizing (9) using the generalized ridge penalty for temporal effect can be expressed in the matrix form as:

$$S\left\|\bar{y} - I \alpha\right\|_2^2 + \lambda_T \left\|D_1 \alpha\right\|_2^2, \tag{17}$$

where $D_1 \in \mathbb{R}^{m_1 \times T}$ is the penalty matrix forming the second-order difference. The solution for α is written as $\hat{\alpha} = \left(I^T I + \lambda_T D_1^T D_1\right)^{-1} I^T \bar{y}$. In contrast, the problem of minimizing (10) using the generalized ridge penalty for spatial effect over time can be expressed in matrix form as:

$$\left\|\tilde{y}_t - I \boldsymbol{\beta}_t\right\|_2^2 + \lambda_{S,t} \left\|D_2 \boldsymbol{\beta}_t\right\|_2^2 \tag{18}$$

for $t = 1, \ldots, T$, where $D_2 \in \mathbb{R}^{m_2 \times S}$ is the penalty matrix forming the first-order difference on the set of edges $\mathcal{E}$. The solution for $\boldsymbol{\beta}_t$ is written as $\hat{\boldsymbol{\beta}}_t = \left(I^T I + \lambda_{S,t} D_2^T D_2\right)^{-1} I^T \tilde{y}_t$. We also compare the proposed method with the unpenalized estimation in (2), that is $\hat{\alpha}_t = \bar{y}_{\cdot t}$ and $\hat{\beta}_{it} = y_{it} - \bar{y}_{\cdot t}$.

We applied LOOCV using the R function "cv.glmnet" to select the tuning parameter in the lasso problem. For ridge and generalized ridge problems, we applied the efficient LOOCV (Meijer 2010). In the case of minimizing (17), an efficient formula of the LOOCV error for given λ_T is represented in a closed form as

$$\frac{1}{T} \sum_{t=1}^{T} \left(\frac{\bar{y}_{\cdot t} - \hat{\alpha}_t}{1 - h_{tt}}\right)^2, \tag{19}$$

where h_{tt} is the t-th diagonal element of the hat-matrix $H_T = \left(I^T I + \lambda_T D_1^T D_1\right)^{-1}$. Then, we select λ_T minimizing LOOCV error (19). Similarly, in the case of minimizing (18), the LOOCV error for given $\lambda_{S,t}$ can be expressed as

$$\frac{1}{S} \sum_{i=1}^{S} \left(\frac{y_{it} - \bar{y}_{\cdot t} - \hat{\beta}_{it}}{1 - h_{ii}}\right)^2 \tag{20}$$

for $t = 1, \ldots, T$, where h_{ii} is the i-th diagonal element of the hat-matrix $\boldsymbol{H}_{S,t} = \left(\boldsymbol{I}^T \boldsymbol{I} + \lambda_{S,t} \boldsymbol{D}_2{}^T \boldsymbol{D}_2\right)^{-1}$. We select $\lambda_{S,t}$ minimizing LOOCV error (20) for each t, or a common $\lambda_S \equiv \lambda_{S,t}$ minimizing the sum of (20) for $t = 1, \ldots, T$.

In this simulation study, we assessed the performance of the regularization methods explained above by using the mean square error (MSE) of the coefficients, which indicates the closeness between estimates and the true coefficients. We computed the MSE of the estimated temporal effect and estimated spatial effect. Moreover, to assess the accuracy for detecting clusters in the estimated spatial effect, we used the index of edges detection accuracy (*IEDA*) to evaluate the accuracy of detecting edges with zero differences, that is, zero elements of the vector $\boldsymbol{D}_2 \boldsymbol{\beta}_t$. For 100 data replications, *IEDA* can be stated as

$$IEDA = \frac{1}{100} \sum_{z=1}^{100} \frac{2 \times Sens_z^E \times PPV_z^E}{Sens_z^E + PPV_z^E}, \tag{21}$$

where $Sens_z^E$ and PPV_z^E indicate the sensitivity and PPV (positive prediction value) to detect the edges with zero differences, respectively, in the z-th replication. They are calculated as

$$Sens_z^E = \frac{len\left(\left\{s : \left(\boldsymbol{D}_2 \widehat{\boldsymbol{\beta}}_t\right)_s = 0\right\} \cap \left\{s : \left(\boldsymbol{D}_2 \boldsymbol{\beta}_t\right)_s = 0\right\}\right)}{len\left(\left\{s : \left(\boldsymbol{D}_2 \boldsymbol{\beta}_t\right)_s = 0\right\}\right)}, \tag{22}$$

$$PPV_z^E = \frac{len\left(\left\{s : \left(\boldsymbol{D}_2 \widehat{\boldsymbol{\beta}}_t\right)_s = 0\right\} \cap \left\{s : \left(\boldsymbol{D}_2 \boldsymbol{\beta}_t\right)_s = 0\right\}\right)}{len\left(\left\{s : \left(\boldsymbol{D}_2 \widehat{\boldsymbol{\beta}}_t\right)_s = 0\right\}\right)}. \tag{23}$$

for $t = 1, \ldots, T$, where *len* shows the length of a vector, $\left\{s : \left(\boldsymbol{D}_2 \widehat{\boldsymbol{\beta}}_t\right)_s = 0\right\}$ is the index vector of estimated edges with zero differences, and $\left\{s : \left(\boldsymbol{D}_2 \boldsymbol{\beta}_t\right)_s = 0\right\}$ is the index vector of actual edges with zero differences. The *IEDA* close to 1 means that the estimates can detect edges with zero differences appropriately which indicates the objects are clustered correctly. We calculated *IEDA* when selecting a common tuning parameter (λ_S) over all time points and when selecting a different tuning parameter ($\lambda_{S,t}$) for each time point. We also calculated the averages of sensitivity and PPV over 100 replications, that is,

$$\overline{Sens^E} = \frac{1}{100} \sum_{z=1}^{100} Sens_z^E, \quad \overline{PPV^E} = \frac{1}{100} \sum_{z=1}^{100} PPV_z^E.$$

Because we are motivated by revealing the spread of Covid-19 positive cases in Japan, we constructed data simulating cases for each prefecture in Japan. Japan consists of 47 prefectures, with code 1–47 assigned roughly from north to south, and is grouped into 8 regions: Hokkaido (1), Tohoku (2–7), Kanto (8–14), Chubu (15–23), Kansai (24–30), Chugoku (31–35), Shikoku (36–39), Kyushu and Okinawa (40–47). We

suppose that the adjacency between each pair of prefectures is defined based whether they are connected by land, bridges/tunnels, or ocean transportation (National Statistics Center 2016). We set weekly time points as $T = 25$ to represent about 6 months.

We generated new positive cases o_{it} in the i-th prefecture ($i = 1, 2, \ldots, 47$) at t-th week ($t = 1, 2, \ldots, 25$) from Poisson distribution with mean $\mu_{it} \times N_i$, where $\ln(\mu_{it}) = \alpha_t + \beta_{it} + \varepsilon_{it}$. Where, the noise ε_{it} was generated independently by following a normal distribution with mean 0 and standard deviation 3. N_i is the population in the i-th prefecture, which was obtained form the 2020 Japan's Population Census (Portal Site of Official Statistics of Japan (e-Stat), 2021). We define three cases of the true $\alpha_t + \beta_{it}$ with values 1, 5, and 10 as shown in the Fig. 1, to represent different problems and structures of clusters as follows.

(**a**) The Case 1 represents that one aggregated region of higher risk moves as time goes by as in Fig. 1a. In this case, we simulated the cluster of prefectures in Tohoku, Kanto, and Chubu Regions, which have a higher risk steady for four weeks. Then, the higher risk region moves to southwest prefectures within four weeks and become steady on most prefectures in Chubu, Kansai, and Chugoku Regions for eight weeks. After that, the higher risk region moves again to southwest prefectures within four weeks and become steady in Kansai, Chugoku, Shikoku, and Kyushu Regions for remaining five weeks.

(**b**) The Case 2 represents that one aggregated region of higher risk increases and decreases in size as in Fig. 1b. In the first four weeks, the higher risk region keeps steady on several prefectures in Kanto, Chubu, Kansai, and Chugoku Regions. Then, the higher risk region spreads to other surrounding prefectures within four weeks until it becomes steady up to seven regions: Tohoku, Kanto, Chubu, Kansai, Chugoku, Shikoku, and Kyushu Regions within eight weeks. After that, the higher risk region decreases for four weeks, and then returns to the initial size and keeps steady for five weeks.
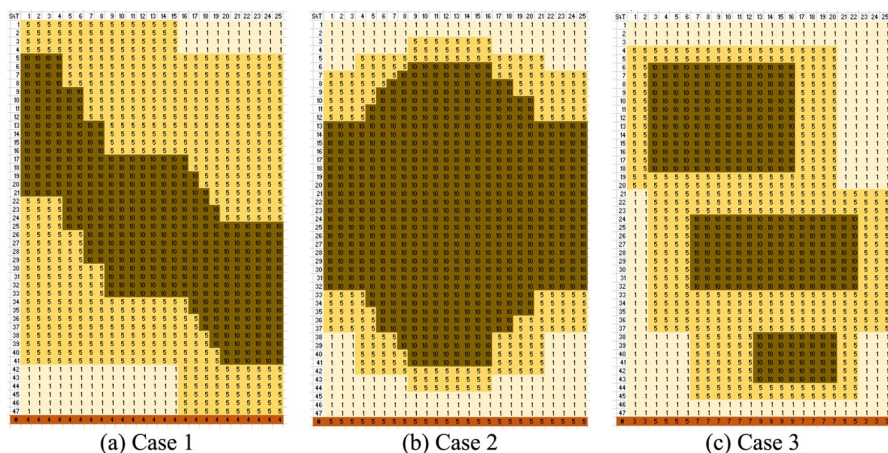

(a) Case 1	(b) Case 2	(c) Case 3

Fig. 1 True $\alpha_t + \beta_{it}; i = 1, 2, \ldots, 47, t = 1, 2, \ldots, 25$ for each case. The row label shows the prefecture code, and the column label shows the week. The last row indicates the number of separated regions for each time point

(**c**) The Case 3 represents that several aggregated regions of higher risk appears and disappears as in Fig. 1c. For the first two weeks, there are no region of higher risk. Then, a higher risk region appears on prefectures in Tohoku, Kanto, and Chubu Regions for 14 weeks. In week 7, the second higher risk region appears on prefectures in Kansai and Chugoku Regions for 16 weeks. Meanwhile, the third higher risk region appears on prefectures in Shikoku and Kyushu Regions from week 13 for 8 weeks.

The regions of adjacent prefectures with higher risk value means that they are clustered. The last row of Fig. 1 shows the number of regions separated by adjacency of prefectures and different level of $\alpha_t + \beta_{it}$.

For each case 1, 2, and 3, 100 data sets were replicated. Then, for each data set, we transformed as $y_{it} = \ln\left(\frac{o_{it}}{N_i}\right)$, and fitted several models explained above. For generalized lasso/ridge, we defined the second-order difference penalty matrix $\boldsymbol{D}_1 \in \mathbb{R}^{23\times25}$ for temporal effect. Moreover, the definition of adjacency between prefectures detects 93 edges, from which we obtained the penalty matrix $\boldsymbol{D}_2 \in \mathbb{R}^{93\times47}$ for spatial effect. For lasso and ridge, we pooled 4 successive coefficients for temporal effect to obtain $\alpha_1^*, \ldots, \alpha_6^*$ (the last one covers five weeks) and pooled coefficients based on 8 prefectural regions of Japan to obtain $\beta_{1t}^*, \ldots, \beta_{8t}^*$, so that $T_1 = 6$, $S_1 = 8$, $\boldsymbol{X}_1 \in \mathbb{R}^{25\times6}$ and $\boldsymbol{X}_2 \in \mathbb{R}^{47\times8}$.

Figure 2 shows the MSE for the estimate of the temporal effect α_t. In Case 1, the lasso had the smallest MSE than other methods, followed by the generalized lasso with ALOCV and GCV.

The true temporal effect α_t was almost constant, and so we guess that the lasso estimates of pooled coefficients might be more advantageous. In Case 2, the generalized lasso with ALOCV mainly provided the smallest MSE, followed by the generalized lasso with GCV. In this case, the MSEs of lasso and ridge were fluctuated highly, especially at the points where true risk values changed.

In Case 3, although the MSE of lasso had the smallest value for several weeks, but the generalized lasso with GCV's MSE was the most stable, followed by the generalized lasso with ALOCV. The MSEs of lasso and ridge were also fluctuated when the number of clusters increased or decreased. We guess that pooling of the coefficients might have caused poor performance of lasso and ridge. Generally, in the temporal effect estimation, the generalized lasso with ALOCV and GCV provided
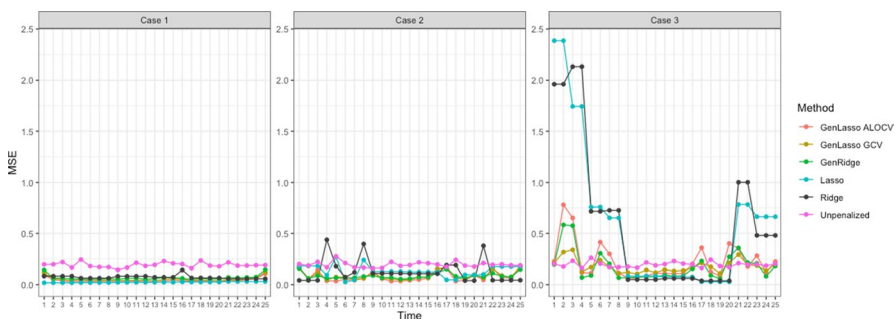

Fig. 2 Line-plots of MSE for coefficient α_t each time point from 100 replications. The Cases 1, 2, and 3 correspond to the true $\alpha_t + \beta_{it}$ displayed in Fig. 1

relatively smaller and more stable MSE than other methods, for different pattern of true risk values.

Figure 3 shows the average of MSE for the estimates of the spatial effect β_{it} over 47 prefectures for each time point. In the case of using a common tuning parameter λ_S (Fig. 3a), the generalized lasso with GCV mainly provided the minimum MSE in all cases, followed by the generalized lasso with ALOCV. In Case 1, when the true risk values were steady in weeks 9 to 16, the MSE of lasso was smaller than the generalized lasso with ALOCV. However, in Case 2 and Case 3, when the true risk values changed, the MSEs of other methods were relatively higher than the generalized lasso. The result in the case of using a different tuning parameter $\lambda_{S,t}$ (Fig. 3b) was slightly different. Mainly, the generalized ridge provided the smallest MSE. The MSEs of the generalized lasso with GCV and ALOCV was higher at several intermediate weeks in Case 1, and at the first and last several weeks in Case 2, but were smaller than lasso and ridge in other cases.

Figure 4 shows the plots of *IEDA* for all time points in clustering prefectures. We only show the result of the generalized lasso with ALOCV and GCV and lasso because all edges take non-zero differences for other methods. We can see that the generalized lasso with ALOCV outperformed, as indicated by the highest *IEDA* for most cases in Fig. 4a and b. The *IEDA* generally increased when the number of separated regions was small and decreased when the number of separated regions was large. Table 2 shows the averages of $\overline{Sens^E}$, $\overline{PPV^E}$ and *IEDA* over all time points. The generalized lasso with ALOCV provided higher sensitivity and *IEDA* than the generalized lasso with GCV and the lasso, although the coefficients of lasso were pooled in advance based on 8 prefectures regions. Moreover, we obtained slightly higher *IEDA* when using a common tuning parameter than using different tuning
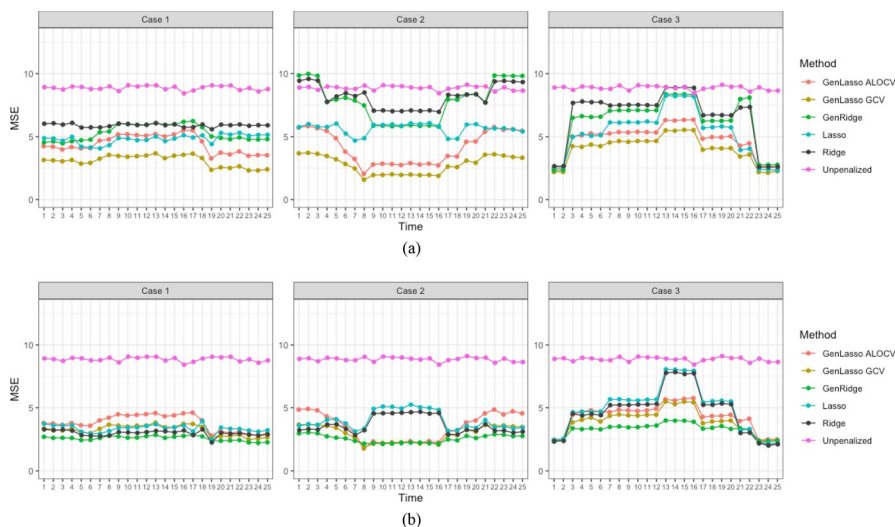

Fig. 3 Line-plots of average and the range of MSE from β_{it} over 47 prefectures for each time point, in the case of using (**a**) a common λ_S for all time points and (**b**) different $\lambda_{S,t}$ for each time point. The Cases 1, 2, and 3 correspond to the true $\alpha_t + \beta_{it}$ displayed in Fig. 1
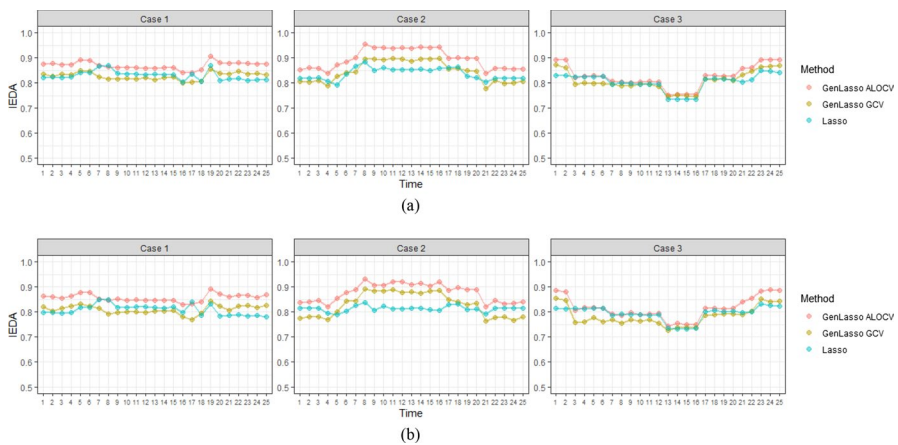
(a)

(b)

Fig. 4 Line-plots of *IEDA* for each time point, in the case of using (**a**) a common λ_S for all time points and (**b**) different $\lambda_{S,t}$ for each time point. The Cases 1, 2, and 3 correspond to the true $\alpha_t + \beta_{it}$ displayed in Fig. 1

parameters for all the cases and methods. If we use a common tuning parameter, the chosen tuning parameter value was not too small, so that the differences between coefficients on the edges tended to shrink to zero, which resulted in more accurate clustering. In contrast, if we use a different tuning parameter at each time point, the tuning parameter chosen varied greatly and was small for many time points. As a result, the differences between coefficients on the edges did not tend to shrink to zero, which decreased the clustering accuracy.

In summary, our simulation study showed that the proposed method performed well in estimating the temporal effect, as suggested by lower MSE. Moreover, the proposed method was also very flexible in detecting multiple clusters, as shown

Table 2 Averages of $\overline{Sens^E}$, $\overline{PPV^E}$, and *IEDA* over all time points

Case	Value	A common tuning parameter ($\lambda_{S,t} \equiv \lambda_S$)			Different tuning parameters ($\lambda_{S,t}$)		
		GenLasso ALOCV	GenLasso GCV	Lasso	GenLasso ALOCV	GenLasso GCV	Lasso
Case 1	$\overline{Sens^E}$	0.904	0.781	0.825	0.865	0.763	0.784
	$\overline{PPV^E}$	0.840	0.888	0.836	0.857	0.885	0.837
	IEDA	0.870	0.827	0.829	0.857	0.810	0.809
Case 2	$\overline{Sens^E}$	0.929	0.795	0.833	0.881	0.779	0.791
	$\overline{PPV^E}$	0.869	0.915	0.841	0.884	0.911	0.838
	IEDA	0.896	0.847	0.837	0.877	0.831	0.813
Case 3	$\overline{Sens^E}$	0.902	0.842	0.839	0.865	0.790	0.818
	$\overline{PPV^E}$	0.765	0.785	0.775	0.780	0.800	0.775
	IEDA	0.825	0.808	0.803	0.815	0.783	0.794

by high *IEDA* values. The generalized lasso with ALOCV outperformed in detecting clusters, while the generalized lasso with GCV performed well in estimating coefficients.

5 Real case data application: weekly Covid-19 cases in Japan

Since the first confirmed case was detected on January 16, 2020, Japan has experienced 5 major waves of the spread of Covid-19 until September 2021. As of September 11, 2021, the total number of Covid-19 cases in all prefectures in Japan was 1,627,898, with 98% recovered rate (Ministry of Health, Labor, and Welfare, 2021). At that time, the number of Japan's Covid-19 cumulative confirmed cases was the 26-th highest in the world (WHO 2021). In our study, we choose the start point on March 21, 2020, because on that date the total confirmed Covid-19 cases exceeded 1,000, spread in 39 of 47 (83%) prefectures of Japan.

Figure 5 shows daily reported Covid-19 cases in Japan, with (a)–(d) indicating periods of each declaration of emergency status respectively. To correspond with the first wave of Covid-19 spread, the first emergency status was declared on April 7, 2020, first in seven prefectures, and then it was expanded nationwide on April 16, 2020 (Fig. 5a). The second wave occurred in August 2020, but at that time the government didn't declare an emergency status until the end of the year. After the number of cases decreased in autumn, the third wave occurred at the end of 2020, in which the number of infections reached 230,000 people. The second emergency status was declared for Saitama, Chiba, Tokyo, and Kanagawa on January 8, 2021, and was expanded to 11 prefectures on January 13, 2021. The duration of this emergency status was until March 7, 2021 (Fig. 5b). The first dose of Covid-19 vaccination was implemented on April 1, 2021, while at that time, Japan was hit by the fourth wave of outbreak. The third emergency declaration was issued for Tokyo, Osaka, Kyoto, and Hyogo on April 25, 2021, and was expanded to five other prefectures on May 16,
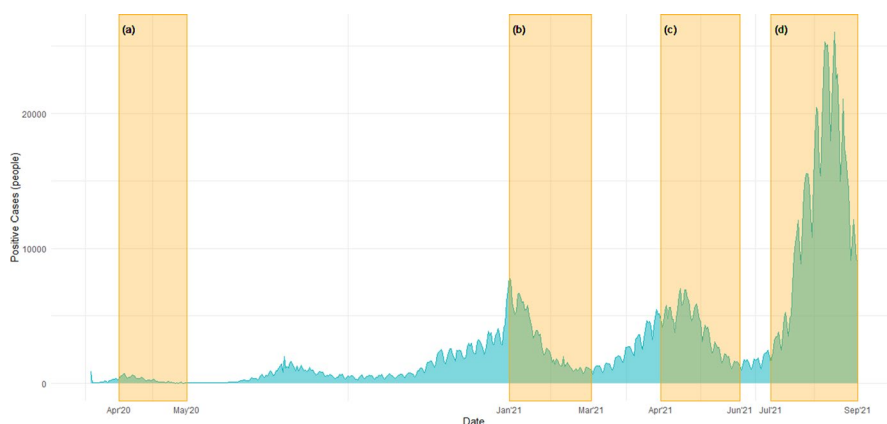

Fig. 5 Daily reported Covid-19 positive cases in Japan from March 18, 2020, to September 11, 2021, with emergency status periods (**a**)-(**d**)

2021, which was lifted on June 20, 2021, except in Okinawa (Fig. 5c). The fifth wave occurred around July to September 2021, during which the Olympic Summer Games was held in Tokyo, and the fourth state of emergency was declared in several prefectures, particularly to prevent spread of the highly contagious Delta variant (Fig. 5d).

We applied the minimization problem (3) for spatio-temporal analysis, which can be decomposed into the two generalized lasso problems (9) and (10), with tuning parameter λ_T and $\lambda_{S,t}$ selected by ALOCV and GCV, to understand the temporal effect and prefectural clusters constructed at each time point. We used the weekly Covid-19 positive case data for each prefecture in Japan from March 21, 2020, to September 11, 2021 (the data file covid_jpn_prefecture.csv in Takaya (2021)). Therefore, we have $S = 47$ and $T = 78$. Let y_{it}^* be the number of weekly positive cases in the i-th prefecture and at the t-th week, and N_i be the population in the i-th prefecture. Here, we used the log transformed positive case per population $y_{it} = \log\left(\frac{y_{it}^*}{N_i}\right)$ as the response variable in the generalized lasso problem. The adjacency between each pair of prefectures was introduced as constraints in the same way as in Sect. 4, and hence we have $\boldsymbol{D}_1 \in \mathbb{R}^{76 \times 78}$ and $\boldsymbol{D}_2 \in \mathbb{R}^{93 \times 47}$.

We calculated an unbiased estimate of the DF for λ_T or $\lambda_{S,t}$ to evaluate complexity of the model. According to Tibshirani and Taylor (2011), the DF for given λ in the generalized lasso (1) is defined as,

$$df = E\big[nullity\big(\boldsymbol{D}_{-E}\big)\big], \tag{24}$$

where $nullity\big(\boldsymbol{D}_{-E}\big)$ is the dimension of the null space of $\boldsymbol{D}_{-E}$, reduced rows of the penalty matrix $\boldsymbol{D}$ corresponding to the boundary index set E of a solution of the dual problem (12). The DF for the L_1 penalty suggests the number of fused groups. In estimating temporal effect with (9), we used the formula (24) to calculate DF for selected λ_T. In estimating spatial effect for each time point with (10), we selected a common λ_S for all $t = 1, 2, \ldots, 78$, based on the simulation study described in Sect. 4, in which a common tuning parameter resulted in higher *IEDA* values. In this case, the DF was calculated as the average value of DF over all $t = 1, 2, \ldots, 78$.

5.1 Result of estimating temporal effect

We considered minimizing (9) to estimate the temporal effect α_t. Table 3 contains selected λ_T and the DF based on the proposed method with ALOCV and GCV. We limited the maximum DF to $\frac{3}{4}(78) = 58.5$ to avoid extremely rough temporal effect. The generalized lasso with ALOCV selected higher λ_T and smaller DF than the results of the generalized lasso with GCV.

Figure 6 shows the estimated temporal effect $\widehat{\alpha}_t$ for each t based on selected λ_T using generalized lasso with ALOCV and GCV, with each emergency status period (a)-(d). The break points in the estimated trend should suggest some change of conditions such as emergency status declaration. The estimated trend using ALOCV for L_1 penalty has slightly fewer break points than the one using GCV for L_1 penalty.

Table 3 Selected λ_T and DF for estimating the temporal effect α_t

Method	Selected λ_T	DF
GenLasso with ALOCV	0.2759	37
GenLasso with GCV	0.0411	51

During the first emergency status period (a), the estimated temporal effect reached the first peak at first and then fell down. After the period (a) ended, it rose quickly and reached the second peak in summer of 2020. During the second emergency status period (b), it reached the third peak at first and then fell down again. During the third emergency status period (c), it slightly increased for a while, reached the fourth peak, and then decreased quickly. During the fourth emergency status period (d), it increased for more than one month and then reached the fifth peak, and then decreased.

5.2 Result of estimating spatial effect

We considered minimizing (10) to estimate the spatial effect β_{it}. In this case, we assumed that $\lambda_S \equiv \lambda_{S,t}$ for all $t = 1, 2, \ldots, 78$ and selected it by minimizing the sum of ALOCV and GCV errors over all t.

Table 4 contains selected λ_S and the DF. We limited the maximum DF to $\frac{3}{4}(47) = 35.25$ to avoid extremely rough spatial effect. ALOCV was not computable at lower λ_S (less than 3.47) for the reason of division-by-zero issue, described in the last paragraph of Sect. 3, so that ALOCV selected higher λ_S with lower DF compared to GCV. The estimates $\widehat{\beta}_{it}$ of spatial effect for these methods of selecting λ_S were plotted in Figs. 7, 8, and 9, in which the prefectures are plotted roughly from North (top) to South (bottom), and darker red color indicates higher values. Figure 7 shows unpenalized estimates of the spatial effect $\widehat{\beta}_{it} = y_{it} - \bar{y}_{.t}$. Based on this figure, we can see the relative spread of Covid-19 for each prefecture in every week. However, the unpenalized $\widehat{\beta}_{it}$ looks very rough, and hence it is very difficult to grasp commonalities and
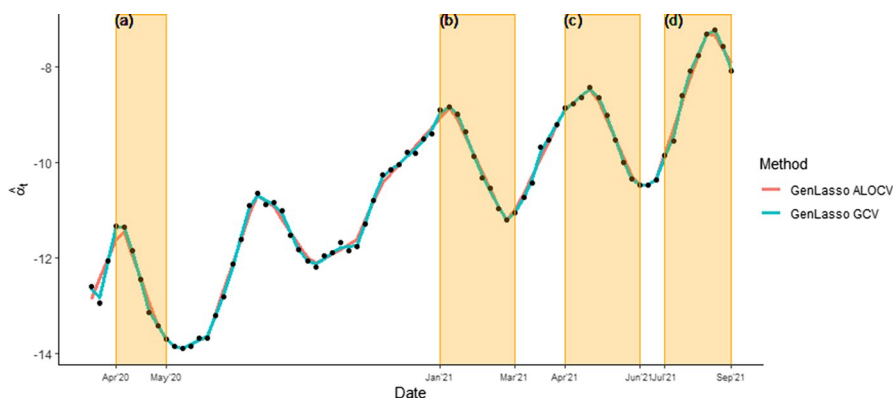

Fig. 6 Plot of estimated temporal effect $\widehat{\alpha}_t$ based on λ_T selected by using generalized lasso with ALOCV and GCV, with emergency status periods (**a**)-(**d**)

Table 4 Selected common λ_S and DF for estimating spatial effect

Method	Selected λ_S	DF
GenLasso with ALOCV	4.1400	1.29
GenLasso with GCV	0.8939	6.64

differences of spatial effect between regions for each week. Figure 8 shows the estimated spatial effect $\widehat{\beta}_{it}$ with λ_S selected by using ALOCV. It looks very smooth, and one or few clusters covered all prefectures at most of the weeks, and at some weeks the estimated spatial effect had large difference depending on the clustered regions. Figure 9 shows estimated spatial effect $\widehat{\beta}_{it}$ with λ_S selected by using GCV. It suggests that there were some clusters of prefectures with the same values. We can see that the largest cluster consisted of most of prefectures during the emergency status periods. However, in other period of weeks, the prefectures were divided into some clusters.

5.3 Clustering the regions based on the estimated spatial effect

Figure 10 shows the heatmap of the estimated spatial effect $\widehat{\beta}_{it}$ with a common λ_S selected by using GCV for generalized lasso as in Fig. 9, but prefectures have been arranged based on agglomerative hierarchical clustering. The heatmap after the arrangement can display relative infection risk, that is, how the infection occurred in a specific area and then spread to other areas.

Based on Fig. 10, we can detect six major clusters of prefectures from top to bottom: (1) All prefectures in Kyushu region, (2) All prefectures in Chugoku and Shikoku regions, (3) All prefectures in Kanto, Chubu, Kansai regions, and Fukushima prefecture (central part of Japan), (4) Prefectures in Tohoku region except Fukushima, (5) Hokkaido prefecture, and (6) Okinawa prefecture. We provide the following interpretation of the dynamic behavior of spatial clusters based on the result of generalized lasso clustering with separating into the five waves that Japan has experienced.

i.　First wave (March 21 to June 27, 2020)

　　During the first wave of infections, relative infection risk increased gradually in the central part of Japan (cluster 3) and Okinawa (cluster 6) and decreased in the remaining clusters. Then, while the first emergency status had been declared from mid-April to May 2020, relative infection risk was extremely higher in Hokkaido (cluster 5) and decreased gradually in the other clusters.

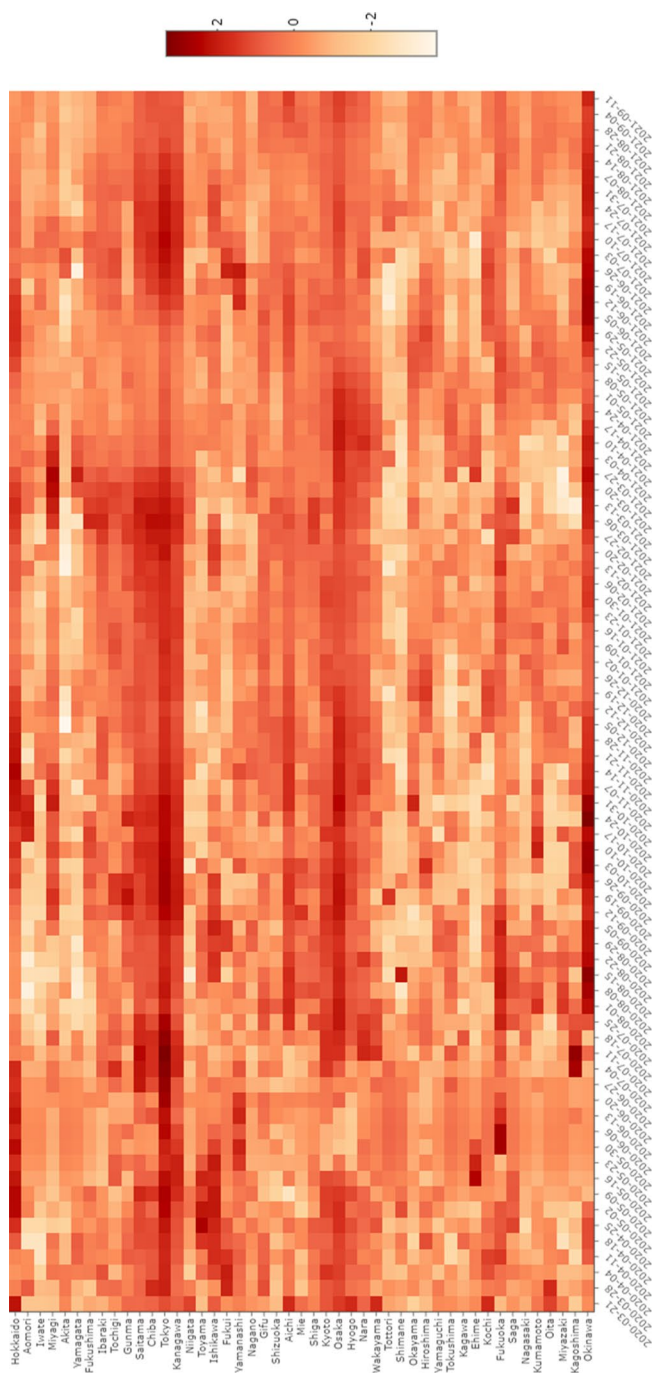ii.　Second wave (July 4 to October 17, 2020)

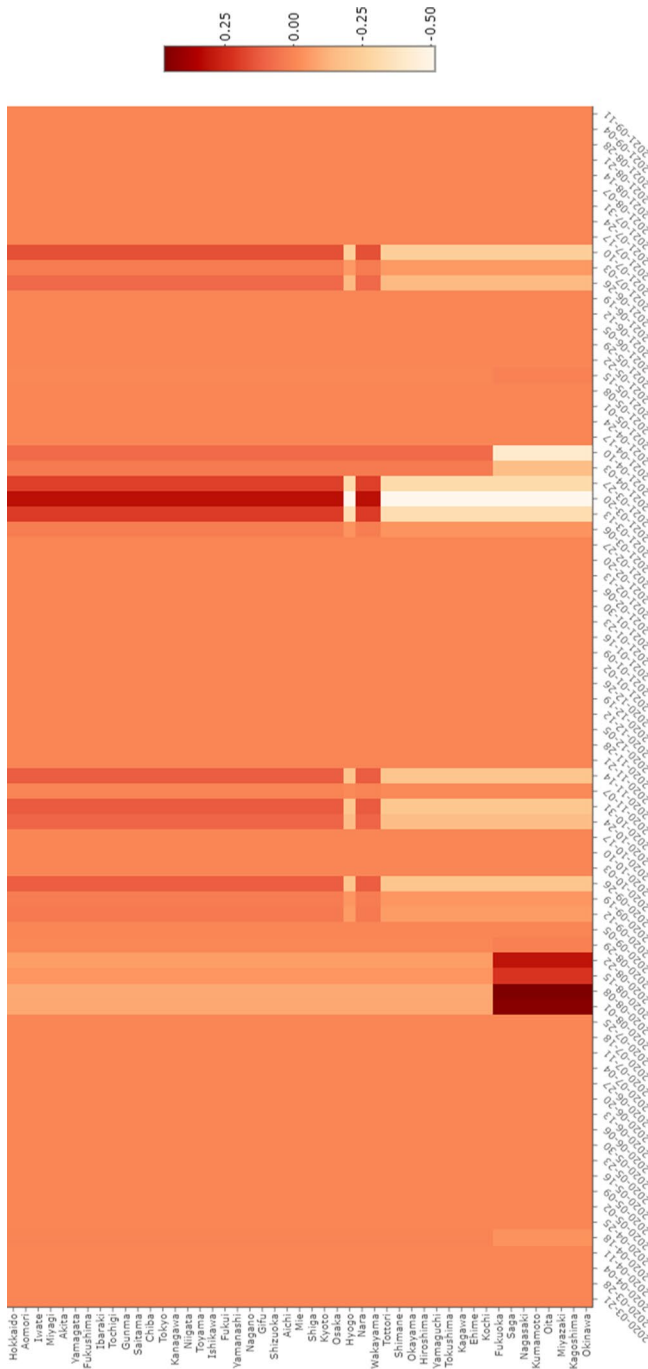Fig. 7 Heatmap of unpenalized estimated spatial effect $\widehat{\beta}_{it}$ with a common λ_S

Fig. 8 Heatmap of estimated spatial effect $\hat{\beta}_{it}$ with a common λ_S selected by ALOCV
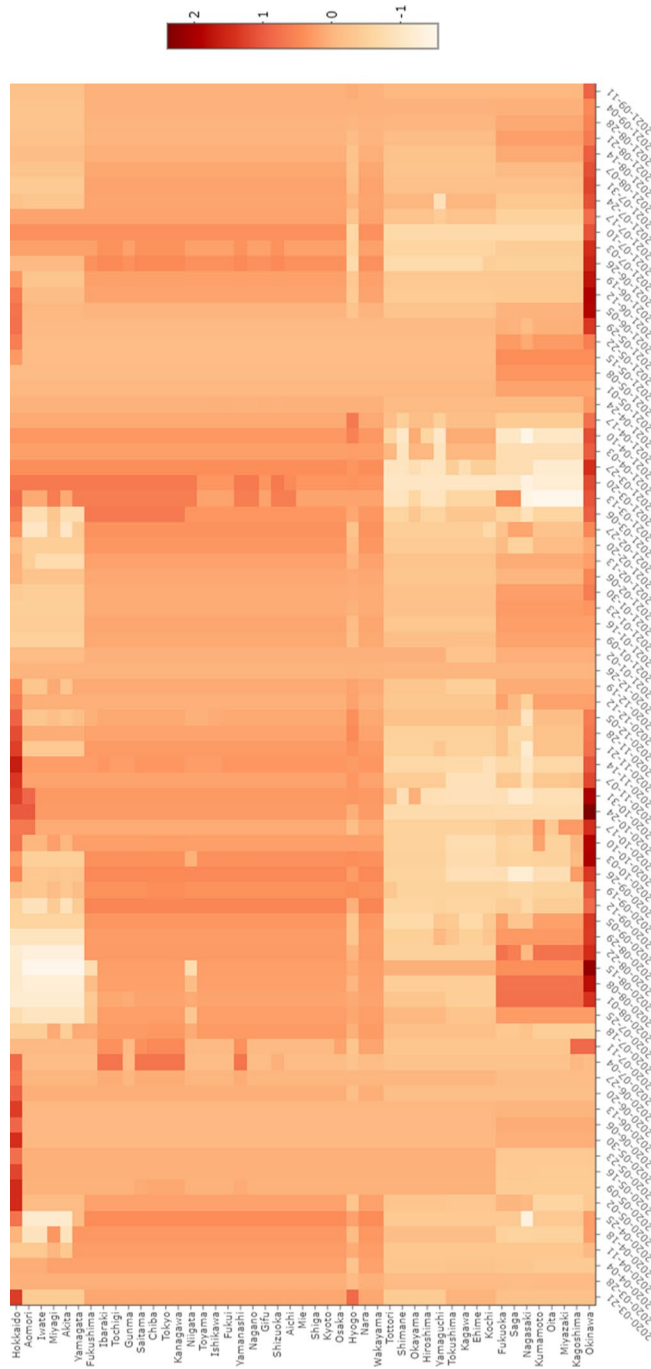
Fig. 9 Heatmap of estimated spatial effect $\hat{\beta}_{it}$ with a common λ_S selected by GCV
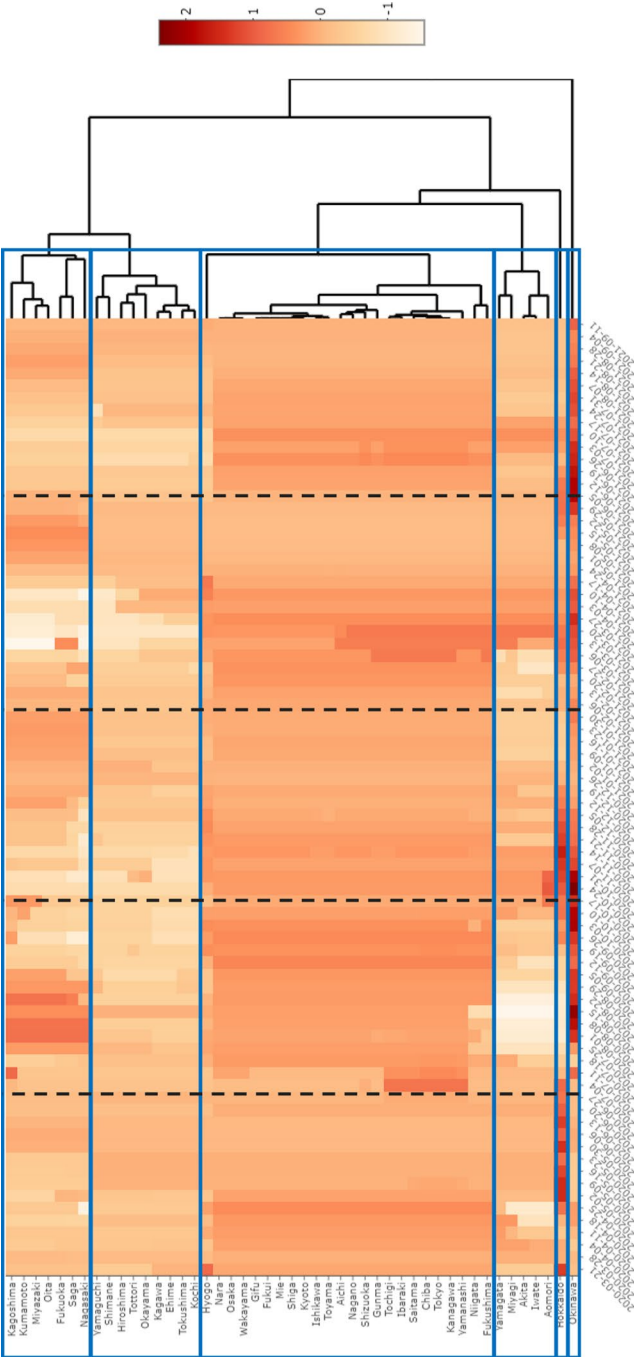
Fig. 10 Heatmap of estimated spatial effect $\widehat{\beta}_{it}$ with a common λ_S selected by using GCV for generalized lasso, in which prefectures arranged based on agglomerative hierarchical clustering. The blue horizontal boxes indicated separated clusters constructed and black vertical dashed lines indicated separator between waves

In the second wave of outbreaks, relative infection risk was higher in Kyushu (cluster 1), central Japan (cluster 3), and Okinawa (cluster 6), while lower in the other clusters. In cluster 1, the outbreak reached a peak in August 2020 and then decreased gradually. During the period, the relative risk increased gradually in cluster 3. It was the highest and stagnant in Okinawa.

iii.　Third wave (October 24, 2020, to February 6, 2021)

In the third wave, relative risk was higher in central and northern parts of Japan (clusters 3, 4, 5, and 6). Then, it decreased gradually while the second emergency status had been declared from January to March 2021.

iv.　Fourth wave (February 13 to June 5, 2021)

After higher risk in Kanto and Tohoku regions in March 2021, the fourth wave spread to other regions. While the third emergency status had been declared from April to June 2021, relative risk was higher in Okinawa (cluster 6) in April and in Kyushu region (cluster 1) in May but was lower in the other clusters.

v.　Fifth wave (June 12 to September 11)

In the fifth wave of infections, infection risk increased first in Okinawa (cluster 6), spread into central Japan (cluster 3), Tohoku (cluster 4), and Hokkaido (cluster 5), next into Chugoku and Shikoku (cluster 2), and then into Kyushu (cluster 1).

In summary, we can see that the outbreaks that occurred in central Japan (cluster 3) spread into outer regions such as Chugoku-Shikoku region (cluster 2) and Tohoku region (cluster 4) in one month, and then spread into Kyushu region (cluster 1) a few months late. We can also see that the outbreaks in some regions leaped into Hokkaido (cluster 5) and Okinawa (cluster 6) a few months late.

6 Conclusion

In this study, we proposed a regularization approach using a modified generalized lasso model with two L_1 penalties for temporal effect and spatial effect. Then, our proposed method can be separated into two generalized lasso problems: trend filtering to estimate smooth temporal effect and fused lasso to detect clusters of spatial location for each time point. Through our proposed method, we can understand dynamic behavior of spatial clusters over time more flexibly, based on relative magnitude of estimated spatial effect at each time point.

To select the appropriate tuning parameters in the generalized lasso, we considered using ALOCV and GCV. Our simulation study suggested that estimation of temporal and spatial effects using generalized lasso with ALOCV and GCV was comparable or superior in terms of MSE to existing regularization methods such as lasso, ridge, and generalized ridge. Also, we showed that the generalized lasso with ALOCV provided higher *IEDA*, the accuracy of detecting edges with non-zero difference. In addition, our simulation study suggested that a common tuning parameter over all time points was preferable in spatial clustering.

Then, through the analysis of weekly Japan's Covid-19 panel data, we illustrated how to understand the spread of Covid-19 infection using our modified generalized lasso model. In estimation of the spatial effect over weeks, the generalized lasso with a common tuning parameter over all time points selected by GCV, provided a reasonable result.

This study mainly used the "genlasso" package of R software to solve the generalized lasso problems using the dual path algorithm (Arnold and Tibshirani 2016). However, we may consider using the coordinate descent algorithm as suggested in Yamamura et al. (2021), which suggested to have better estimation accuracy and speed than the algorithm used in "genlasso". Moreover, to detect the spatial clusters in the spread of disease as a task in epidemiology studies, the response variable is often observed as count data. The application of modified generalized lasso for count data was proposed by Choi et al. (2018), to which we have a great attention in our future work.

References

Ansari MY, Ahmad A, Khan SS, Bhushan G, Mainuddin. (2020) Spatiotemporal clustering: a review. Artif Intell Rev 53(4):2381–2423. https://doi.org/10.1007/s10462-019-09736-1

Arnold TB, Tibshirani RJ (2016) Efficient implementations of the generalized lasso dual path algorithm. J Comput Graph Stat 25(1):1–27. https://doi.org/10.1080/10618600.2015.1008638

Castro MC, Kim S, Barberia L, Ribeiro AF, Gurzenda S, Ribeiro KB, Abbott E, Blossom J, Rache B, Singer BH (2021) Spatiotemporal pattern of COVID-19 spread in Brazil. Science 372(6544):821–826. https://doi.org/10.1126/science.abh1558

Chen Y, Ong JHY, Rajarethinam J, Yap G, Ng LC, Cook AR (2018) Neighbourhood level real-time forecasting of dengue cases in tropical urban Singapore. BMC Med 16(1):129. https://doi.org/10.1186/s12916-018-1108-5

Choi H, Song E, Hwang S, Lee W (2018) A modified generalized lasso algorithm to detect local spatial clusters for count data. AStA Adv Statis Anal 102(4):537–563. https://doi.org/10.1007/s10182-018-0318-7

Craven P, Wahba G (1979) Estimating the correct degree of smoothing by the method of generalized cross-validation. Numer Math 31:377–403

Doborjeh MG, Kasabov N (2015) Dynamic 3D clustering of spatio-temporal brain data in the neucube spiking neural network architecture on a case study of fMRI data. In: Arik S, Huang T, Lai WK, Liu Q (eds) Neural information processing. Springer International Publishing, Cham, pp 191–198. https://doi.org/10.1007/978-3-319-26561-2_23

Doborjeh MG, Kasabov N, Doborjeh ZG (2018) Evolving, dynamic clustering of spatio/spectro-temporal data in 3D spiking neural network models and a case study on EEG data. Evol Syst 9(3):195–211. https://doi.org/10.1007/s12530-017-9178-8

Donoho DL, Johnstone IM (1995) Adapting to unknown smoothness via wavelet shrinkage. J Am Stat Assoc 90(432):1200. https://doi.org/10.2307/2291512

Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12(1):55. https://doi.org/10.2307/1267351

Hüsch M, Schyska BU, von Bremen L (2020) CorClustST—Correlation-based clustering of big spatio-temporal datasets. Futur Gener Comput Syst 110:610–619. https://doi.org/10.1016/j.future.2018.04.002

Izakian H, Pedrycz W, Jamal I (2013) Clustering spatiotemporal data: an augmented fuzzy C-means. IEEE Trans Fuzzy Syst 21(5):855–868. https://doi.org/10.1109/TFUZZ.2012.2233479

Izakian H, Pedrycz W, Jamal I (2015) Fuzzy clustering of time series data using dynamic time warping distance. Eng Appl Artif Intell 39:235–244. https://doi.org/10.1016/j.engappai.2014.12.015

Kamenetsky ME, Lee J, Zhu J, Gangnon RE (2022) Regularized spatial and spatio-temporal cluster detection. Spatial Spatio-Temporal Epidemiol 41:100462. https://doi.org/10.1016/j.sste.2021.100462

Kim S-J, Koh K, Boyd S, Gorinevsky D (2009) l1 trend filtering. SIAM Rev 51(2):339–360. https://doi.org/10.1137/070690274

Ministry of Health, L. and W. (2021). Current situation in Japan. https://www.mhlw.go.jp/stf/covid-19/kokunainohasseijoukyou_00006.html

Meijer R (2010) Efficient approximate leave-one-out cross-validation for ridge and lasso. Delft University of Technology, Netherlands

National Statistics Center. (2016). Publication of counted and indexed lists of combined adjacent blocks of prefectures in Japan (in Japanese). https://www.nstac.go.jp/technology/research/prefcomp/

Portal Site of Official Statistics of Japan (e-Stat). (2021, Oct 24). *Population Census 2020*. https://www.e-stat.go.jp/

Rad KR, Zhou W, and Maleki A. (2020). Error bounds in estimating the out-of-sample prediction error using leave-one-out cross validation in high-dimensions. In: Proceedings of the 23rd international conference on artificial intelligence and statistics (AISTATS), pp 108

Rad KR, Maleki A (2020) A scalable estimate of the extra-sample prediction error via approximate leave-one-out. J R Statis Soc Series B. 82(4):965–996

Rahardiantoro S, Sakamoto W (2021) Clustering regions based on socio-economic factors which affected the number of COVID-19 cases in Java Island. J Phys: Conf Series 1863(1):012014. https://doi.org/10.1088/1742-6596/1863/1/012014

Rahardiantoro S, Sakamoto W (2022) optimum tuning parameter selection in generalized lasso for clustering with spatially varying coefficient models. IOP Conf Series: Earth Environ Sci 950(1):012093. https://doi.org/10.1088/1755-1315/950/1/012093

Rahardiantoro S, Sakamoto W (2022) Spatially varying coefficient modeling of numerical and categorical predictor variables in the generalized lasso. J Environ Sci Sustain Soc 11(Supplement PP05):16–19

She Y, and Owen AB. (2010). Outlier detection using nonconvex penalized regression. Unpublished manuscript. http://www-stat.stanford.edu/~owen/reports/theta-ipod.pdf

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. J R Statis Soc Series B (Methodol) 36:111–147

Takaya, H. (2021). COVID-19 dataset in Japan, Kaggle Dataset. https://www.kaggle.com/lisphilar/covid19-dataset-in-japan

Takemura Y, Ishioka F, Kurihara K (2022) Detection of space–time clusters using a topological hierarchy for geospatial data on COVID-19 in Japan. Japan J Statis Data Sci 5(1):279–301. https://doi.org/10.1007/s42081-022-00159-x

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Stat Methodol 58(1):267–288

Tibshirani RJ (2014) Adaptive piecewise polynomial estimation via trend filtering. Ann Statis. https://doi.org/10.1214/13-AOS1189

Tibshirani RJ, Taylor J (2011) The solution path of the generalized lasso. Ann Statis. https://doi.org/10.1214/11-AOS878

Tibshirani R, Wang P (2008) Spatial smoothing and hot spot detection for CGH data using the fused lasso. Biostatistics 9(1):18–29. https://doi.org/10.1093/biostatistics/kxm013

Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K (2005) Sparsity and smoothness via the fused lasso. J R Statis Soc: Series B (Statis Methodol) 67(1):91–108. https://doi.org/10.1111/j.1467-9868.2005.00490.x

Wang Q, Dong W, Yang K, Ren Z, Huang D, Zhang P, Wang J (2021a) Temporal and spatial analysis of COVID-19 transmission in China and its influencing factors. Int J Infect Dis 105:675–685. https://doi.org/10.1016/j.ijid.2021.03.014

Wang Y, Liu Y, Struthers J, Lian M (2021b) Spatiotemporal characteristics of the COVID-19 epidemic in the United States. Clin Infect Dis 72(4):643–651. https://doi.org/10.1093/cid/ciaa934

Wang, S., Zhou, W., Maleki, A., Lu, H., and Mirrokni, V. (2018). Approximate leave-One-Out for high-dimensional non-differentiable learning problems. arXiv:1810.02716v1 [cs.LG]

World Health Organization (WHO). (2021, October 23). *WHO Coronavirus (COVID-19) Dashboard*. https://covid19.who.int/table

Yamamura M, Ohishi M, Yanagihara H (2021) Spatio-temporal adaptive fused lasso for proportion data. In: Czarnowski I, Howlett RJ, Jain LC (eds) Intelligent Decision Technologies: Proceedings of the 13th KES-IDT 2021 Conference. Springer Singapore, Singapore, pp 479–489. https://doi.org/10.1007/978-981-16-2765-1_40

Zhao Y, Bondell H (2020) Solution paths for the generalized lasso with applications to spatially varying coefficients regression. Comput Statis Data Anal 142:106821. https://doi.org/10.1016/j.csda.2019.106821