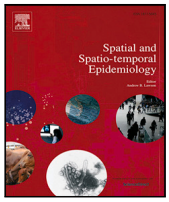
Spatial and Spatio-temporal Epidemiology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste



Original research

Regularized spatial and spatio-temporal cluster detection

Maria E. Kamenetsky [a], Junho Lee [b], Jun Zhu [c], Ronald E. Gangnon [a,d,*]

[a] Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI 53726, USA
[b] Department of Statistical Science, Baylor University, Waco, TX 76798, USA
[c] Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA
[d] Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI 53726, USA

ARTICLE INFO

Keywords:
Lasso
Poisson regression
Spatial cluster detection
Spatio-temporal cluster detection
Spatial scan statistic
Quasi-likelihood

ABSTRACT

Spatial and spatio-temporal cluster detection are important tools in public health and many other areas of application. Cluster detection can be approached as a multiple testing problem, typically using a space and time scan statistic. We recast the spatial and spatio-temporal cluster detection problem in a high-dimensional data analytical framework with Poisson or quasi-Poisson regression with the Lasso penalty. We adopt a fast and computationally-efficient method using a novel sparse matrix representation of the effects of potential clusters. The number of clusters and tuning parameters are selected based on (quasi-)information criteria. We evaluate the performance of our proposed method including the false positive detection rate and power using a simulation study. Application of the method is illustrated using breast cancer incidence data from three prefectures in Japan.

1. Introduction

Spatial and spatio-temporal cluster detection are important tools for public health surveillance and hypothesis generation. A spatial cluster is a geographic area with a distinct pattern of risk relative to the background across either space or space and time. Cluster detection methods are used in many fields such as epidemiology, ecology, and demography; the goal is to identify geographic subregions that differ from the rest of the study region. In public health, data are often aggregated to administrative geographic subregions, or cells. We focus on such areal data, aiming to identify clusters of elevated (or reduced) disease incidence or prevalence. We propose a new spatio-temporal clustering method to identify spatio-temporal clusters and characterize the risk inside identified clusters.

Traditional methodology, such as with standardized incidence rates (SIRs), results in noisy maps where trends across space and time are challenging to distinguish. To address the subjectivity of such noisy maps, early cluster detection methods using moving windows across the geographic area were proposed. These methods used overlapping circular windows centered at cell centroids within the study region and evaluated each circle using observed cases inside each window. Openshaw et al. (1988) developed a regular grid with circles of varying radii, centered at grid centroids and only significant circles were then identified as clusters. Bruce et al. (1990) and Besag and Newell (1991) formalized the idea of circular clusters and extended the set of potential clusters to be defined by fixed circle sizes or case counts, and facilitated formal testing of identified clusters.

In an effort to identify the single most likely cluster, Kulldorff and Nagarwalla (1995) and Kulldorff (1997) introduced the spatial scan statistic, which used a likelihood-based approach to compare observed to expected case counts using the maximum of the set of likelihood ratio tests. Significance of the identified most likely cluster was assessed using Monte Carlo-based hypothesis testing, under the null of a common rate in all cells. Further extensions of the spatial scan and likelihood-based testing for spatial clusters were introduced (Neill, 2012; Shu et al., 2012), such as methodology for other distributions (Huang et al., 2007; Kulldorff et al., 2009; Jung et al., 2010) and alternative shape variants (Duczmal and Assuncáo, 2004; Tango and Takahashi, 2005; Kulldorff et al., 2006; Takahashi et al., 2008), allowing for flexibility in the geographic definition of a cluster. Adjustments for multiplicity have also been proposed for the spatial setting (Gangnon and Clayton, 2004; Gangnon, 2010b) and the spatio-temporal framework (Kulldorff et al., 1998; Kulldorff, 2001). Methods in spatial cluster detection have been extended to allow for the detection of multiple clusters in a study area. Zhang et al. (2010) and Li et al. (2011) extended the spatial scan statistic to multiple clusters by either identifying a second cluster after sequential deletion of observed data in the first identified cluster or identification based on the likelihood ratio statistic of the second largest cluster that does not overlap with the first most likely cluster.

* Corresponding author at: Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI 53726, USA.
E-mail address: ronald@biostat.wisc.edu (R.E. Gangnon).

https://doi.org/10.1016/j.sste.2021.100462
Received 26 March 2021; Received in revised form 22 July 2021; Accepted 18 October 2021
Available online 1 November 2021

The spatial scan statistic and its variants have been implemented in the software, SaTScan (Kulldorff, 2015).

Spatial cluster detection can also be recast as a high-dimensional data problem. When the number of variables exceeds the number of observations, many methods are available for variable selection (Hastie et al., 2008; Efron et al., 2004) with high-dimensional data. Best subset selection methods, forward stepwise, and stagewise regression, for example, select a subset of variables from the set of all predictors and discard other variables. Xu and Gangnon (2016) developed a framework for identifying multiple clusters by considering spatial cluster detection to be a case of high-dimensional variable selection where $p \gg n$ and proposed the use of incremental forward stepwise and stagewise regression. The forward stepwise approach of Xu and Gangnon used a maximum likelihood ratio test statistic to find the most likely cluster. Though the forward stepwise approach is fast, it is known to behave poorly for highly-correlated predictors. In the case of highly-correlated predictors, forward stagewise regression was shown to be more suitable for cluster detection. Though it produces similar coefficient paths to the early stages of the least absolute shrinkage and selection operator (Lasso), it is computationally intensive as it requires evaluation of the set of predictors with the largest gradient and continuous updating and normalization at every small fixed step, ε. Other efficient regularization methods, such as ridge regression (Hoerl and Kennard, 1970), least angle regression (LARS) (Efron et al., 2004), and the Lasso (Tibshirani, 1996), instead efficiently shrink coefficient estimates according to a penalty. The Lasso also performs variable selection by shrinking most estimates to zero, thereby inducing sparsity.

In this paper, we develop a Lasso-based approach for efficiently detecting multiple spatial and spatio-temporal clusters. The shrinkage properties of the Lasso reduce most coefficient estimates to zero. Unlike the spatial scan or forward stepwise regression, our approach does not require sequential cluster deletion (Zhang et al., 2010; Li et al., 2011) and is computationally efficient. We propose a novel high-dimensional regression formulation of the cluster model and represent potential clusters as binary indicators, allowing us to use sparse matrices and gain computational efficiency. This approach also enables us to only estimate and keep coefficients which have a strong signal of an elevated relative risk compared to the estimated background rate and we determine the number of clusters by information criteria (Hurvich and Tsai, 1989; Cavanaugh and Shumway, 1997; Pan, 2004).

The structure of the paper is as follows. The statistical models are introduced in Section 2 and our methodology using the Lasso regularization is introduced in Section 3. In Section 4 we demonstrate the method properties by a simulation study and in Section 5 we present the analysis of a female breast cancer incidence dataset from Japan. Lastly, we summarize our findings in a discussion in Section 6.

2. Models and clusters

2.1. Statistical modeling

Suppose we have a study region in $\mathbb{R}^2$ which is divided into N cells and T time periods. For each cell i at time t, we observe the number of disease cases, y_{it}, the population at risk, n_{it}, and the expected number of cases, E_{it}. Our work here focuses on rare disease counts which can be approximated by the Poisson distribution, though it is easily extended to the binomial distribution. Let

$$y_{it} \sim \text{Poisson}(\mu_{it}), \tag{1}$$

where $\mu_{it} = \rho_{it} E_{it}$ is the expectation of y_{it} and ρ_{it} is the unobserved relative risk. To control for a confounding variable, either direct or indirect standardization can be used.

We model the relative risk, ρ_{it}, with a log-linear model (Gangnon and Clayton, 2003; Yan and Clayton, 2006):

$$\log(\rho_{it}) = \alpha_t + \sum_{j=1}^{k} \theta_j \mathbb{1}\{d(\boldsymbol{z}_i, \boldsymbol{c}_j) \le r_j, l_j \le t \le u_j\}, \tag{2}$$

The time-varying parameter α_t represents the risk for background cells (cells not belonging to any of the active clusters), t is the time period. The spatio-temporal clustering component is $\sum_{j=1}^{k} \theta_j \mathbb{1}\{d(\boldsymbol{z}_i, \boldsymbol{c}_j) \le r_j, l_j \le t \le u_j\}$, where k is the number of spatio-temporal clusters selected, θ_j is the log relative risk in cluster j, and $\mathbb{1}\{\cdot\}$ is the indicator function that takes 1 if the Euclidean distance $d(\cdot)$ between a cell with center $\boldsymbol{z}_i$ and cluster centered at $\boldsymbol{c}_j$ is less than or equal to radius r_j, and cell i is inside the (closed) time between l_j and u_j, and is 0 otherwise. For simplicity of notation, we let $x_{itj} = \mathbb{1}\{d(\boldsymbol{z}_i, \boldsymbol{c}_j) \le r_j, l_j \le t \le u_j\}$.

The spatial model is a special case of the spatio-temporal model Eq. (1)–(2), where $l_j \equiv 1$ and $u_j \equiv T$ for all choices of j. This framework can easily be extended to other distributions such as the binomial, normal, or multinomial, though not limited to the exponential family.

2.2. Potential clusters

Potential clusters in space are a collection of moving circular windows centered at geographic centroids of the cells as in the spatial scan (Kulldorff and Nagarwalla, 1995). When extended to space and time, potential clusters become cylinders. We consider the whole collection of potential clusters in the study region at once and allow for them to overlap spatially. We obtain the whole range of potential clusters by enumerating all combinations of potential clusters based on a radius specified by the user (r_{max}), centered at each centroid. Potential clusters are created using ordered distances from the centroid of a given cell i up to the centroids of all other cells until r_{max} truncates the process. Though setting the maximum cluster radius is an open research question (Ribeiro and Costa, 2012; Han et al., 2016), we use the general guideline of setting r_{max} *a priori* and such that larger clusters are not meaningful for public health. Since potential clusters are allowed to overlap, one or more circles are guaranteed to identify any cluster in the study region.

Let S be the number of potential 2-D circular spatial clusters centered at each geographic centroid with radii that range from 0 to r_{max} and let T be the number of time intervals, giving $T(T+1)/2$ potential 1-D temporal clusters. Then the total number of K spatio-temporal clusters is $ST(T+1)/2$, which remain space–time separable. The set of potential clusters are stored as a large sparse matrix, since each potential cluster will have a small number of non-zero elements. Our goal is to leverage regularization methods for high-dimensional data in order to select a parsimonious set of non-zero log relative risks while shrinking all other θ_j's to zero (see Section 3), thereby isolating the cluster(s) with relative risks that differ from the estimated background disease rate.

2.3. Likelihood functions

Our parameter of interest, θ_j, is the log relative risk inside the jth cluster for $j = 1, \ldots, K$ potential clusters. We estimate θ_j using Lasso regularization (see Section 3.1) and α_t as unpenalized terms. Based on the model specified in Eq. (2), the loglikelihood function evaluated at $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ simplifies to:

$$\log \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left[y_{it} \left(\alpha_t + \sum_{j=1}^{K} \theta_j x_{itj} + \log(E_{it}) \right) - \exp\left(\alpha_t + \sum_{j=1}^{K} \theta_j x_{itj} \right) E_{it} - \log y_{it}! \right]. \tag{3}$$

For Poisson, the assumption that the mean be equal to the variance may not be realistic when working with disease counts. An overdispersed Poisson distribution (quasi-Poisson) can be constructed using a gamma mixture of Poisson distributions which yields the negative binomial distribution (McCullagh and Nelder, 1991). For a single observation, y follows a Poisson distribution with mean Z. The parameter

Z, however, varies according to a gamma distribution with mean parameter m and shape parameter m/β. Then the Poisson-gamma mixture yields the negative binomial density:

$$p(y; m, \beta) = \frac{\Gamma(y + \beta m)\beta^{\beta m}}{y!\Gamma(\beta m)(1+\beta)^{y+\beta m}}, \tag{4}$$

where $E(y) = m$ and $Var(y) = m(1+\beta)/\beta$. As $\beta \to \infty$ the negative binomial distribution converges to the Poisson distribution.

Instead of assuming a specific distribution for y, we consider a quasi-likelihood-based approach (McCullagh and Nelder, 1991). Let μ be a single component from a mean vector μ, $\phi V(\mu)$ be a single component from a covariance matrix $\phi \boldsymbol{V}(\mu)$, and ϕ be an unknown constant, where $\phi > 1$ indicates overdispersion. Define $Q(\mu; y)$ to be the log quasi-likelihood, where

$$Q(\mu; y) = \int_y^{\mu} \frac{y - t}{\phi V(t)} dt \tag{5}$$

which behaves like a loglikelihood function for μ, assuming a constant ϕ. In the case of Poisson, $V(\mu) = \mu$, which gives us the log quasi-likelihood $Q(\mu; y) = y \log \mu - \mu$. To define the log quasi-likelihood for all data, we assume y_{it} observations are independent and define $Q(\mu, \boldsymbol{y}) = \sum_i \sum_t Q_{it}(\mu_{it}; y_{it})$. In the case of quasi-Poisson (Wedderburn, 1974), the mean–variance relationship is defined by $V(\mu) = \phi\mu$. The overdispersion parameter ϕ can be estimated by:

$$\tilde{\phi} = (n-p)^{-1} \sum_{i}^{N} \sum_{t}^{T} (y_{it} - \hat{\mu}_{it})^2 / V(\hat{\mu}_{it}) = X^2/(n-p),$$

where there are n observations, p parameters, and

$X^2 = \sum_{i=1}^{N} \sum_{t=1}^{T} \{y_{it} - \hat{\mu}_{it}\}^2 / \hat{\mu}_{it}$ is the Pearson statistic (Wedderburn, 1974; McCullagh and Nelder, 1991).

3. Regularization method

3.1. Lasso regularization

Xu and Gangnon (2016) proposed both forward stepwise and incremental forward stagewise approaches to identify spatial clusters. The forward stepwise approach iteratively added clusters into the active set based on the maximum likelihood ratio test. For highly correlated predictors, the incremental forward stagewise approach proposed sequential updating of the model, starting with the null model and gradually incrementing the cluster risk coefficients by a fixed step size ε until convergence. The incremental forward stagewise approach, however, can be very time-consuming and the tuning of ε can affect the convergence of the algorithm. It is also limited to the exponential family as it requires access to sufficient statistics for each potential cluster.

Therefore, we instead propose a regularization method based on the Lasso penalty. Both the Lasso and forward stagewise will decrease along the loss function, though the Lasso does so with respect to the ℓ_1 norm and forward stagewise does so with respect to the ℓ_1 arc length (Tibshirani, 2015). The ℓ_1 norm of the Lasso induces sparsity, which lends itself naturally to searching through a large number of potential clusters as we will demonstrate next. Forward stagewise regression and the Lasso result in similar (and under certain conditions, identical) coefficient paths (Hastie et al., 2007; Efron et al., 2004).

Let $-\log \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\theta})$ be the negative loglikelihood function, which is differentiable and strictly convex. The tuning parameter λ controls the amount of shrinkage and goes from 1 to L, where $\lambda_1, \ldots, \lambda_L$ are monotonically decreasing and K is the number of potential clusters. We propose to minimize the following penalized loglikelihood function:

$$f(\boldsymbol{\alpha}, \boldsymbol{\theta}) = -\log \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\theta}) + \lambda \sum_{j=1}^{K} |\theta_j|, \tag{6}$$

To minimize $f(\boldsymbol{\alpha}, \boldsymbol{\theta})$ in Eq. (6), we adopt a computationally-efficient algorithm and illustrate the idea by way of a linear model, which is a special case of the generalized linear model (GLM). The regularization for the linear model has the following steps:

1. Start with the null model such that $\theta_1 = \cdots = \theta_K = 0$.
2. Iterate $l = 1, \ldots, L$. At a given λ_l, find cluster θ_j most correlated with $\boldsymbol{y}$.
3. Let $\epsilon = \boldsymbol{y} - \hat{\boldsymbol{y}}$ be the residual. Increase the $|\theta_j|$ estimate closer to its maximum likelihood estimate until another cluster, $\theta_{j'}$, has more correlation with ϵ.
4. Add a new cluster, $\theta_{j'}$, to the active set, which is comprised of the clusters that have been selected thus far. If a non-zero $\theta_{j'}$ equals zero, drop the cluster from the active set. Increment l by 1.

The procedure will start with the null model with no clusters and the Lasso regularization performs variable selection by shrinking θ_j's to zero as they reach the penalty and are dropped from the active set. This results in a coefficient path for each θ_j over λ_l. The smaller λ gets, the more clusters are allowed to enter the model. When λ is 0, we are left with the least squares estimates. This approach leads to a large $K \times L$ matrix, where each row is a proposed coefficient path over λ_l. The number of clusters in the active set is the number of unique clusters identified at λ_l. We store only the non-zero entries in a sparse matrix. This not only reduces the amount of data to be stored and manipulated, but is also computationally more efficient. In the GLM case, cyclical coordinate descent is used to calculate the penalized weighted least squares estimates for the decreasing sequence of λ_l values along a grid, replacing steps (2–3) in the algorithm above (Friedman et al., 2010).

To correctly evaluate model fit using information criteria, we count the number of parameters in the active set at each λ_l by the degrees of freedom, which was shown to be approximately equal to the number of predictors in the model (Tibshirani and Taylor, 2012). We search through the sequence of λ's in order to identify when a new cluster enters the active set. In Fig. 1, we show the first six circular clusters and the corresponding coefficient paths. As $\log(\lambda)$ decreases, more clusters enter the active set.

3.2. Selection of multiple clusters

One limitation of the spatial scan is it can only identify a single most likely cluster at a time. Though methods such as sequential deletion or formally incorporating multiple clusters into the alternative have been proposed (Zhang et al., 2010; Li et al., 2011), they still require the recalculation of the loglikelihood for each subsequent cluster to be allowed into the model. In contrast, our Lasso-based approach identifies multiple overlapping spatial or spatio-temporal clusters simultaneously. At each λ, multiple clusters are allowed into the active set and clusters with estimated relative risks near zero are dropped from the active set.

The Lasso identifies a set of clusters for all choices of the penalty (λ). To select λ, we propose an information-theoretic approach allowing us to retain all observations in the original dataset. We avoid using cross-validation as it is computationally-costly and it is inconclusive as to how to split spatio-temporal data into test and train sets without breaking a potential cluster or multiple overlapping clusters when splitting the data spatially and temporally.

The first criterion we use to select the number of clusters is Bayesian information criterion (Schwarz, 1978). The Bayesian interpretation of the Lasso considers a Laplacian prior on θ_j (Tibshirani, 1996; Park and Casella, 2008), which pulls weakly-related clusters to 0, and the identified coefficients are then interpreted as posterior modes. We use Bayesian information criterion (BIC) as a selection tool, defined as:

$$\text{BIC}(k, \lambda) = -2 \log \mathcal{L}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}; \lambda) + k_\lambda \log(n), \tag{7}$$

where $\log \mathcal{L}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}})$ is the loglikelihood based on the active set for a given λ, k_λ is the number of clusters in the active set at each λ, and n is the total number of observations, calculated as the sum of observed cases.

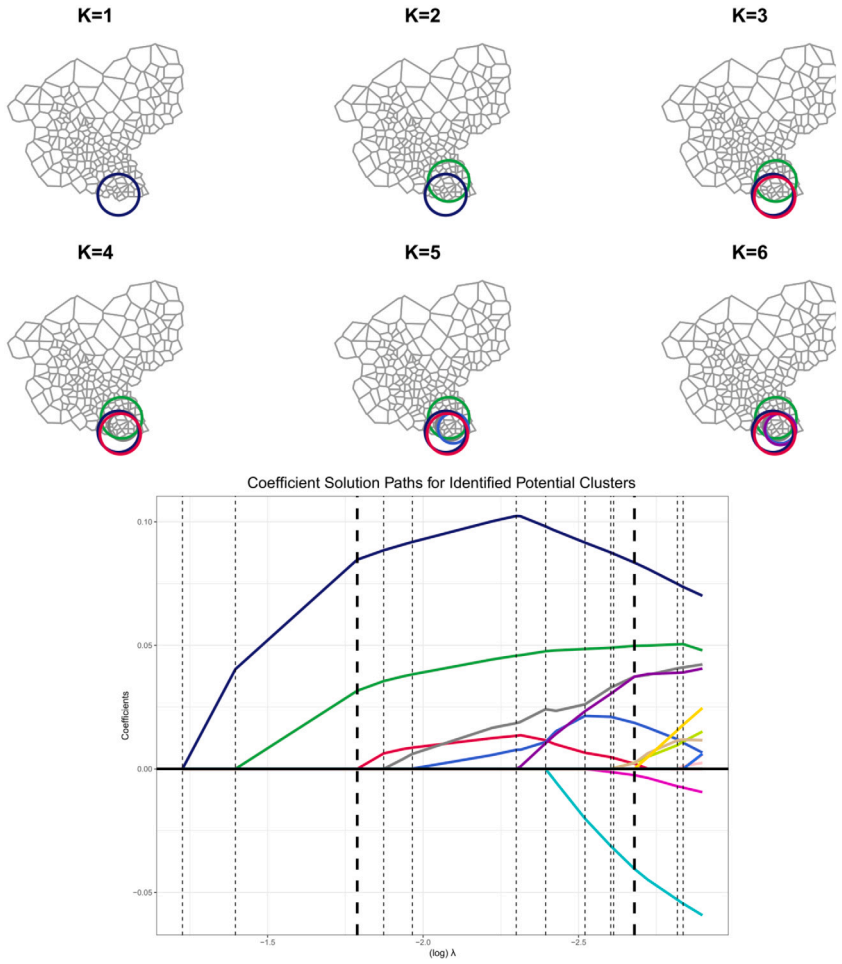


Fig. 1. The first $K = 1$ through $K = 6$ circular clusters across space and their corresponding coefficient paths from the spatio-temporal Lasso model with time interval covering all 5 time periods.

The second criterion used is Akaike's information criterion (Akaike, 1973) (AIC), which has been shown to be asymptotically equivalent to leave-one-out cross-validation (Stone, 1977). AIC is defined as:

$$\mathrm{AIC}(k, \lambda) = -2 \log \mathcal{L}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}; \lambda) + 2k_{\lambda}. \tag{8}$$

For quasi-Poisson, we scale the loglikelihood by the estimated overdispersion (Lebreton et al., 1992) parameter $\tilde{\phi}$, such that quasi-BIC (QBIC) is defined as:

$$\mathrm{QBIC}(k, \lambda) = -2 \log \mathcal{L}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}; \lambda)/\tilde{\phi} + (k_{\lambda} + 1) \log(n), \tag{9}$$

and quasi-AIC (QAIC) is defined as:

$$\mathrm{QAIC}(k, \lambda) = -2 \log \mathcal{L}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}; \lambda)/\tilde{\phi} + 2(k_{\lambda} + 1). \tag{10}$$

4. Simulation study

We first describe our simulation study setup and introduce two comparison methods to our Lasso approach: forward stagewise and forward stepwise spatial scan (Xu and Gangnon, 2016). After discussing the evaluation criteria for our simulation studies, we present our simulation results. We demonstrate the performance of our Lasso approach in spatio-temporal cluster detection as well as in detecting multiple clusters under both no overdispersion and under overdispersion, and finally summarize our findings. Table 1 summarizes the simulation study designs.

4.1. Simulation setup

We developed a simulation study based on breast cancer incidence counts in Japan (Yan and Clayton, 2006) to evaluate the performance of the Lasso in detecting spatio-temporal clusters, which we define to be rare and different from the background. We focus on three neighboring prefectures in Japan: Tochigi (49 municipalities), Gunma (70 municipalities), and Saitama (89 municipalities). The total area of Tochigi is 6 408 km^2, Gunma is 6 363 km^2, and Saitama is 3 767 km^2 (Yan and Clayton, 2006).

The original dataset contains the female population size at risk and number of deaths due to breast cancer, stratified by age, for each municipality in Japan between 1975–1994. Age strata consist of 20 groups from 0–4 through 95 to 99 years old. Females under 40 or over 74 years old were excluded to avoid potential misdiagnosis or comorbidities. Let y_{git} and n_{git} be the observed number of deaths and population at risk in the gth age group, ith municipality, and tth year, respectively. Then age-standardized expected number of deaths were calculated based on the population size at risk as $E_{it} = \sum_g n_{git} p_g$, and $p_g = \sum_{it} y_{git} / \sum_{it} n_{git}$. Expected and observed cases were aggregated within all 208 municipalities into five time periods: 1975–1978, 1979–1982, 1983–1986, 1987–1990, and 1991–1994. The observed counts ranged from 0 to 82 and expected counts ranged from 0.20 to 58.7 across the 5 time periods. Approximate municipality borders were obtained using the Dirichlet tessellation of municipality centroids.

Across the three prefectures, each municipality corresponds to a cell in our model. We set the maximum radius of the circular clusters centered on each municipality to be $r_{max} = 20$ km and we considered all

Table 1
Simulation settings.

	Model	Process	Method	Selection	Relative Risks, Radii		Population Centers	Time Duration
Method Comparisons	Spatio-Temporal	Poisson ($\beta = \infty$)	Lasso	BIC, AIC	(1.1, 9 km), (1.5, 11 km), (2, 18 km)		Large	Long (periods 3–5)
			Forward Stagewise	BIC, AIC				
			Forward Stepwise	MC				
		Quasi-Poisson ($\beta = 60$)	Lasso	QBIC, BIC, QAIC, AIC	(1.1, 9 km), (1.5, 11 km), (2, 18 km)		Large	Long (periods 3–5)
			Forward Stagewise	BIC, AIC				
			Forward Stepwise	MC				
					Fitted Models		True Models	
					Relative Risks	Radii	Population Centers	Time Duration
Lasso Simulations	Spatio-Temporal	Poisson ($\beta = \infty$)	Lasso	BIC, AIC	1.1, 1.5, 2	9, 11, 18 km	Small, Large	Short (periods 1, 2), Long (periods 3-5)
		Quasi-Poisson ($\beta = 100$)		QBIC, QAIC				
		Quasi-Poisson ($\beta = 60$)		QBIC, QAIC				
	Spatial + Constant Time	Poisson ($\beta = \infty$)	Lasso	BIC, AIC	1.1, 1.5, 2	9, 11, 18 km	Small, Large	Short (periods 1, 2), Long (periods 3–5)
		Quasi-Poisson ($\beta = 100$)		QBIC, QAIC				
		Quasi-Poisson ($\beta = 60$)		QBIC, QAIC				
	Multiple Clusters	Poisson ($\beta = \infty$)	Lasso	BIC, AIC	2 (large population centers), 4 (small population center)	18 km	2 Large, 1 Small	Long (periods 3–5)
		Quasi-Poisson ($\beta = 60$)		QBIC, QAIC				

Simulation results for baseline detection rates for both a model with simulated overdispersed Poisson data (Poisson-gamma mixture) and Poisson data. 100 random datasets were generated for each simulation. For method comparisons, the three relative risk, radii settings correspond to low, medium, and high information settings. For spatio-temporal and spatial + constant time Lasso simulations, the full factorial design across relative risks, radii, population centers, and time duration were explored. For multiple cluster Lasso simulations, the three relative risk, radii settings correspond to low, medium, and high information settings.
QAIC and QBIC in the overdispersed Poisson scenarios are both based on the log quasi-likelihood.

possible time intervals of lengths 1 to 5 periods. This resulted in 4 458 space-only clusters and 66 870 spatio-temporal clusters. As the spatial model is a special case of the spatio-temporal model, we focus our simulation results on the latter and relegate the former to supplemental material.

Let y_{it}^{sim} be the simulated case counts for cell i at time t. Expected counts were standardized to the simulated observed counts such that $E_i^* = \sum_i y_i^{sim} \times (E_i / \sum_i E_i)$ in the spatial setting or $E_{it}^* = \sum_i \sum_t y_{it}^{sim} \times (E_{it} / \sum_i \sum_t E_{it})$ in the spatio-temporal setting.

4.1.1. Forward stagewise and forward stepwise spatial scan methods

The forward stagewise method for detecting multiple clusters obtains the full coefficient path using small incremental steps along the gradient and also uses information criteria as final selection tools. Starting with the null model with no clusters, the cluster with the largest absolute value of the gradient, $\partial \log \mathcal{L}(\theta)/\partial\theta_j$, is evaluated at the current model. For a small step size ε, the cluster coefficient θ_j is updated such that $\theta_j \leftarrow \theta_j + \varepsilon \cdot \text{sign}(\partial \log \mathcal{L}(\theta)/\partial\theta_j)$. The expected number of cases is updated and normalized and the process is repeated. Not only is this gradient approach time-consuming, but the tuning of step size ε may need to be adjusted as the algorithm can fail to converge if the step size is too large and ε is caught between two local minima. Although under certain conditions forward stagewise regression has been shown to be equivalent to the monotonic Lasso (Hastie et al., 2008), the Lasso shrinks the cluster coefficient estimates to zero and drops them from the active set instead of flattening them along the coefficient path as forward stagewise regression does. In this way, the Lasso is more efficient in that it allows for sudden changes in the cluster estimates.

The spatial scan approach identifies the single most likely cluster using the maximum likelihood ratio test and then evaluating the statistical significance based on Monte Carlo simulations under the null hypothesis of constant disease risk in the study region. To identify additional clusters in the study region, previously proposed methods (Zhang et al., 2010; Li et al., 2011) have relied on sequential deletion or hypothesis-based procedures that specify the number of clusters in the alternative hypothesis. The constraint of having to sequentially identify multiple clusters by either removing previously identified clusters or having to specify the number of clusters in the alternative have prohibited the detection of overlapping clusters. Therefore, the standard spatial scan must be modified in order to fully consider previously detected clusters in the identification of subsequent clusters. We modify the standard spatial scan by applying a forward stepwise testing framework, and call this the forward stepwise spatial scan. The spatial scan statistic (the maximum likelihood ratio test statistic over all potential clusters $\text{LR}_{max} = \max_j \text{LR}_{A_j}$) is a global cluster detection test statistic. A p-value for the most likely cluster is calculated by comparing LR_{max} to simulated values under the null hypothesis that the sum of observed cases is a known constant and that the distribution of observed counts is multinomial and free of unknown parameters. To consider any previously detected clusters in a Poisson model, the expected number of cases are updated when an additional cluster is identified.

Table 2
Case counts inside cluster by radius and population.

Population	Radius (km)	Num. centroids	Cases in cluster
Large	9	7	266
Large	11	10	490
Large	18	33	1274
Small	9	2	24
Small	11	2	24
Small	18	8	209

Total number of centroids that make up the cluster (*Num. Centroids*) and observed case counts across two population settings at radii of 9, 11, and 18 kilometers (km).

Table 3
Baseline cluster detection under null (no cluster).

Model	β=60			$\beta = \infty$		
	(Q)AIC	(Q)BIC	MC	AIC	BIC	MC
Lasso	63%	0%	–	57%	0%	–
Forward Stagewise	92%	5%	–	58%	0%	–
Forward Stepwise Spatial Scan	–	–	25%	–	–	4%

Simulation results for spatio-temporal model only. Baseline detection rates for both a model with simulated overdispersed Poisson data (Poisson-gamma mixture, $\beta = 60$) and Poisson data ($\beta = \infty$). Percent of simulations that detect at least one cell in the null model are shown. QAIC and QBIC in the overdispersed Poisson scenarios are both based on the log quasi-likelihood in the Lasso-based approach. Selection tools for forward stagewise were BIC and AIC and for forward stepwise spatial scan 1 000 Monte Carlo (MC) simulation were performed.

4.1.2. Simulated data process & model fitting

We first compare our Lasso method to forward stagewise and forward stepwise spatial scan approaches to cluster detection (Xu and Gangnon, 2016). We explore how the three methods perform under three information settings: (1) cluster with 1.1 relative risk and 9 km radius (low information setting), (2) cluster with 1.5 relative risk and 11 km radius (medium information setting), (3) cluster with 2 relative risk and 18 km radius (high information setting). The three scenarios all had clusters placed in a large population center in long time periods (periods 3 through 5). We consider two data-generating processes: (1) Poisson distribution without overdispersion and (2) quasi-Poisson using the Poisson-gamma mixture, described in Eq. (4). We set $\beta = \infty$ to simulate Poisson distributed counts without overdispersion and $\beta = 60$ (estimated from the data) to simulate quasi-Poisson overdispersed counts. We simulate 100 random datasets for each settings. For the forward stepwise spatial scan, 1 000 Monte Carlo (MC) simulations were performed to calculate the p-value for each identified cluster. We also perform a secondary simulation to further explore detection across the methods. We use the low information and high information settings from above and consider two additional settings: (1) cluster with 1.1 relative risk and 18 km radius (low relative risk, high radius) and (2) cluster with 2 relative risk and 9 km radius (high relative risk, low radius).

We next explore our Lasso method in more depth. For the second data-generating process, we consider two shape parameter values where $\beta = 60$ and $\beta = 100$ with less overdispersion. For each simulation setting, 100 random datasets were fitted with a single cluster and both Poisson and quasi-Poisson models were fit to the data generated under their respective assumptions.

Finally, we examined how our Lasso-based method performs in detecting multiple clusters. We explored placing 1, 2, and 3 clusters into the study region, where two of the clusters were in large population centers and 1 cluster was in a small population centers. All clusters had a radius of 18 km and were placed in long time periods (periods 3 through 5). The two clusters in the large population center had a relative risk of 2 and the cluster in the small population center had a relative risk of 4, in order for it to be detected. We consider two data-generating processes: (1) Poisson distribution without overdispersion and (2) quasi-Poisson using the Poisson-gamma mixture. We set $\beta = \infty$ to simulate Poisson distributed counts without overdispersion and $\beta = 60$ (estimated from the data) to simulate quasi-Poisson overdispersed counts. We simulate 100 random datasets for each setting and both Poisson and quasi-Poisson models were fit to the data generated.

Fitted models were developed to assess the influence of cluster characteristics. The effects of three varying cluster radii (9, 11, and 18 km) and three varying cluster relative risks (1.1, 1.5, and 2) on detection were evaluated.

True models were developed to describe the influence of the region. To determine the effect of population size on cluster detection, clusters were placed in large and small population centers. The large population center corresponded to a densely-populated area of the study region whereas the small population center corresponded to a sparsely-populated area. Table 2 describes the large and small population centers based on the number of centroids and cases in the cluster as cluster radius increased from 9 km to 18 km. To determine the effect of cluster duration in time, clusters were placed in short time periods (periods 1 and 2) and long time periods (periods 3 through 5). The influence of population size and cluster duration were explored within the spatio-temporal model and across the fitted settings of the three varying cluster radii and three varying cluster relative risks.

4.1.3. Simulation evaluation

The same operating characteristics were used to evaluate both the three cluster detection methods and the fitted and true models using our Lasso method. For each simulation, performance was evaluated using two operating characteristics: (1) the proportion of simulations where at least one detected cluster contained no cells in the true cluster (false positive rate) and (2) the proportion of simulations where at least one cell of a detected cluster belonged to the true cluster (power). Across our simulation studies, we focus first on establishing control of the false positive rate. Then given a well-controlled rate, we explore detection based on power. There are scenarios where clusters will not be detected and due to public health implications and community concerns of identifying false positive clusters, we favor no detection over detection despite a high false positive rate. Operating characteristics were evaluated under the null hypothesis of no cluster in the study region to establish baseline performance in the region without any imposed clusters.

The null model corresponds to a large λ penalty that does not allow for any θ_j clusters into the model. In our comparisons to forward stagewise and the forward stepwise spatial scan, we also evaluate computational run time for 100 simulations under each setting.

4.2. Simulation results

Our simulation results demonstrate false positive rates and power of our Lasso approach to spatio-temporal cluster detection as it compares to the forward stepwise spatial scan and forward stagewise approach. Simulations further exploring our Lasso approach demonstrate the false positive rates and power under both true and fitted models. Under the null model with no cluster, we evaluated how both the Poisson under BIC and AIC and quasi-Poisson under QBIC and QAIC performed at baseline under the spatio-temporal model (Table 3).

4.2.1. Power

We calculate power as the proportion of simulations in which a detected potential cluster overlaps at least a single cell of the true cluster, allowing us to avoid arbitrary cut-offs for part of a cluster detected. If any cell of the cluster is to be detected, it is likely to be the cluster center. We explore this further across the following four cluster settings: (1) low information (9 km radius, 1.1 relative risk), (2) high information (18 km radius, 2 relative risk), (3) low radius, high relative risk (9 km radius, 2 relative risk), and 4) high radius, low relative risk (18 km radius, 1.1 relative risk). A single cluster was placed in a large population center in long time period (periods 3 through 5). 100

random datasets were simulated under quasi-Poisson (Poisson-gamma mixture with $\beta = 60$) and Poisson distribution ($\beta = \infty$). We compare our definition of power using a single cell inside the cluster (Power (any cell)) to the cluster center cell being detected in all three time periods (Power (center cell)). We find that if a single cell of the cluster is detected it will be the cluster center cell when using QBIC and BIC across both relative risks and cluster radii (see Fig. S1). Using QAIC and AIC, power as detected by the center cell in all 3 time periods was slightly lower (QAIC: 72%, AIC: 87%) than power detected by any cell (QAIC: 77%, AIC: 88%) when the relative risk was 1.1, which is to be expected because detection by QAIC and AIC consistently had a higher false positive rate. With larger cluster relative risks, power by any cell and power by center cell was identical by both selection criteria under both quasi-Poisson and Poisson.

4.2.2. Comparison to forward stagewise and stepwise

With no overdispersion ($\beta = \infty$), the false positive rate under the null with all three methods was 0% using BIC for both the Lasso (computation time: 24.29 min) and forward stagewise (computation time: 166.96 min) and was 4% using Monte Carlo simulation for the forward stepwise spatial scan (computation time: 780.2 min). Using AIC, the false positive rate was 57% for the Lasso approach and 58% for forward stagewise. In the quasi-Poisson scenario ($\beta = 60$), fitting the null model with no cluster using the spatio-temporal model with the Lasso, the false positive rate was 0% using QBIC for selection and 63% using QAIC for selection (computation time: 24.75 min). For the forward stagewise method, the false positive rate was 5% using BIC for selection and was 92% using AIC for selection (computation time: 166.66 min). With the forward stepwise spatial scan using a Monte Carlo procedure to select clusters, the false positive rate was 25% (computation time: 864.80 min). Fig. 2 shows cluster detection under the null model with no cluster across the two dispersion scenarios ($\beta = 60$, $\beta = \infty$) across the three methods. We observe that population size affected the detection of a cluster, as more potential clusters overlap these regions (Gangnon and Clayton, 2004). The same area that exhibited shadowing is the area that had the most potential clusters and this shadowing is observed under all three dispersion scenarios using QAIC and AIC for the Lasso and forward stagewise methods.

Under quasi-Poisson ($\beta = 60$) in the low information setting with a cluster with 1.1 relative risk and 9 km radius (in a large population center and long time periods (periods 3–5)), the Lasso and forward stagewise had the similarly low false positive rates using QBIC/BIC (1%/4%) compared to the forward stepwise spatial scan (2%). Power was similar using the forward stepwise spatial scan (2%) as compared to Lasso (1%) and forward stagewise (3%). With no overdispersion ($\beta = \infty$), the false positive rate was similar using BIC with both the Lasso and forward stagewise methods (false positive rates: 0%, power: 0%), though with the forward stepwise spatial scan the false positive rate was again higher and power lower (false positive rate: 4%, power: 4%) .

As the cluster increased in both relative risk and radius, the false positive rate remained near baseline or decreased and power increased across all three methods. Under quasi-Poisson ($\beta = 60$) for a cluster with a relative risk of 1.5 and radius of 11 km, using our Lasso approach the false positive rate remained near baseline at 1% (power: 69%) using QBIC. The false positive rates using BIC with forward stagewise and Monte Carlo simulations with forward stepwise spatial scan were both higher (8%, 25%), an power was higher for both forward stagewise (83%) and forward stepwise spatial scan (100%). However, computation time was also higher with the latter two methods (168.2 min, 1 573.21 min) than with the Lasso (35.18 min) for 100 simulations. Using the Lasso approach with QAIC, the false positive rate was much higher (27%, power: 100%) as well as with forward stagewise using AIC (false positive rate: 92%, power: 100%). In the highest information setting with a relative risk of 2 and cluster radius of 18 km, all three methods performed similarly. The false positive rate for our Lasso approach using QBIC was lowest at 0% (power: 100%), followed by the forward stepwise spatial scan at 6% (power: 100%), and then forward stagewise using BIC at 10% (power: 100%). With no overdispersion ($\beta = \infty$), the false positive rate was low using BIC with the Lasso-based method (false positive rate: 0%, power: 100%, computation time: 63.94 min) as well as with forward stagewise (false positive rate: 0%, power: 100%, computation time: 168.78 min) and forward stepwise spatial scan (false positive rate: 4%, power: 100%, computation time: 2 228.62 min). Fig. 3 illustrates the false positive rates, power, and computation time across all three methods and information settings and Table S2 in the supplementary material shows the detection results comparing our Lasso approach to forward stagewise and stepwise. We also further explored false positive rates and power across the methods in a follow-up simulation study. We found that detection using our Lasso approach by QBIC or BIC and using forward stagewise using BIC resulted in false positive rates near baseline and power increased as the cluster relative risk and radius increased. Conversely, the false positive

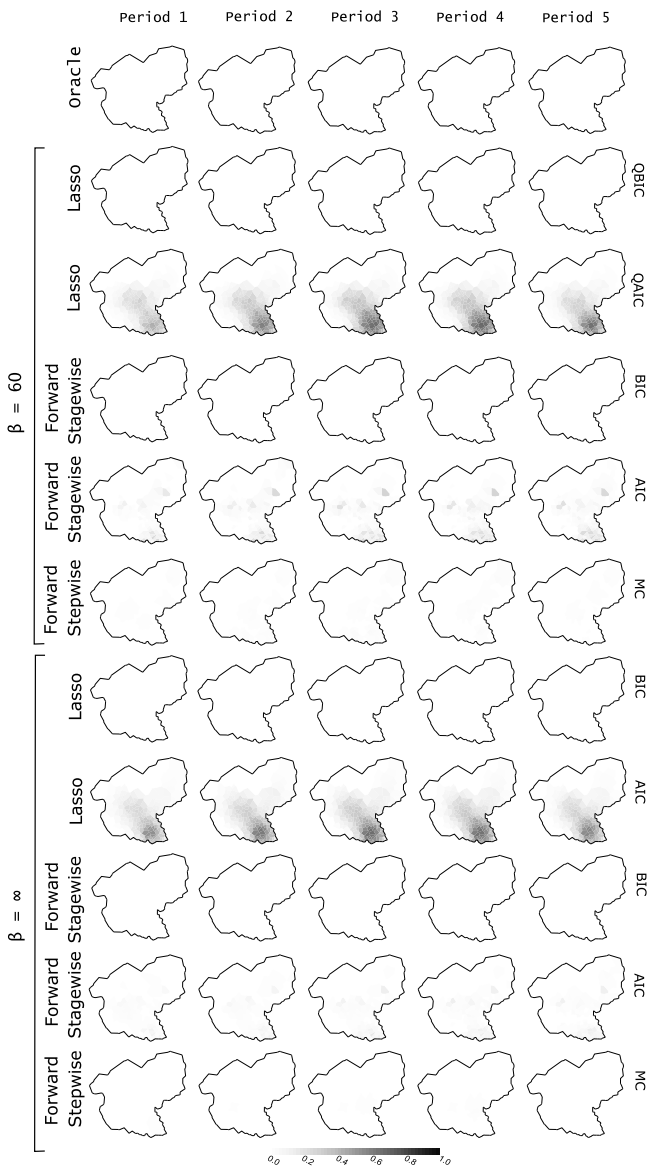


Fig. 2. False detection rate for null (no cluster) model. 100 random datasets were simulated under the quasi-Poisson (Poisson-gamma mixture with $\beta = 60$) and Poisson distribution ($\beta = \infty$). Detection probabilities for each cell are the proportion of simulations in which the cell belongs to one (or more) detected clusters. Under the null models, detection results were the same across both mixtures.

rate under forward stagewise using QAIC or AIC using or the forward stepwise spatial scan using 1000 Monte Carlo simulations resulted in consistently higher false positive rates with low information with either relative risk of 1.1. Fig. S2 compares detection results.

4.2.3. Spatio-temporal cluster detection

For our Lasso method, in the fitted models we explored the effects of cluster relative risks and cluster radii on operating characteristics. Clusters with larger relative risks had lower false positive rates and had higher power than clusters with smaller relative risks. Using BIC under Poisson ($\beta = \infty$), for a cluster with relative risk of 1.5, cluster radius of 11 km, in a large population center, in long time periods (3 through 5), the false positive rate was 0% (power: 93%), and remained at 0% (power: 0%) when relative risk was reduced to 1.1, holding all other parameters constant. Increasing the relative risk to 2 maintained the false positive rate at 0% and power rose to 100%. In contrast, using AIC the false positive rates were consistently higher (24% under a relative risk of 1.1 and 18% under a relative risk of 2), and increased as relative risks decreased, though power was higher (57% and 100%, respectively). Across all settings, increasing relative risk from 1.5 to 2 showed less drastic improvement than the improvement in detection from a relative risk of 1.1 to 1.5, indicating a threshold effect of relative risk. While the false positive rate remained near baseline using BIC, increasing the relative risk as well as the radius improved power once the relative risk was at 1.5 or greater (Fig. 4).

Identifying clusters with larger radii had lower false positive rates and had higher power than clusters with smaller cluster radii. Using BIC under Poisson ($\beta = \infty$), for a cluster with radius of 9 km, relative risk of 1.5, in a large population center, in long time periods (3 through 5), the false positive rate was 1% (power: 18%), and remained near baseline at 1% with a radius of 18 km (power: 100%), holding all other parameters constant. With little information at a 9 km cluster radius and 1.1 cluster relative risk, the false positive rate was 0% (power: 2%) and remained at 0% (power: 100%) even as the cluster radius grew to the largest information setting at an 18 km cluster radius and 2 relative risk. Using AIC, the false positive rates were consistently higher with a 9 km radius with 1.1 relative risk (false positive rate: 26%, power: 49%) and as cluster radius grew to 18 km with 1.1 relative risk (false positive rate: 18%, power: 88%). As cluster radius and relative risk inside the cluster increased, the false positive detection rate remained controlled near the baseline established under the null of 0% using BIC and did not increase as radius and relative risk varied.

In the true models, we explored the effects of population size and temporal duration on cluster detection. Clusters in the large population center had lower false positive rates and higher power than clusters in the small population center (Fig. 5). With relatively little information in a small population center, using BIC under Poisson ($\beta = \infty$), for a cluster in a small population center, with a relative risk of 1.5, cluster radius of 11 km, in long time periods (3 through 5), detection was challenging (power: 0%), but the false positive rate remained near baseline at 1%, as using BIC preferred to identify no clusters at the risk of identifying incorrect clusters. Increasing the relative risk to 2 and cluster radius to 18 km in the small population center, the false positive rate rose to 6% (power: 60%), but fell back to baseline at 0% (power: 100%) for the same setting in the large population center. Decreasing the relative risk to 1.1 and cluster radius to 9 km in the large population center, the false positive rate was still 0% (power: 2%), and remained at 0% (power: 0%) in the small population setting. In contrast, using AIC the false positive rates were much higher in the small population setting. For a cluster with 1.1 relative risk, 9 km radius, in long time periods (false positive rate: 47%, power: 0%), the false positive rate remained well above baseline (false positive rate: 70%, power: 98%) even when the cluster increased to a relative risk of 2 and 18 km radius in the small population center.

Clusters in long time periods (periods 3 through 5) had lower false positive rates and had higher power than clusters in short time periods (periods 1 and 2) (Fig. 6). Using BIC under Poisson ($\beta = \infty$), for a cluster in short time periods, with a relative risk of 1.5, cluster radius of 11 km, in a large population center, the false positive rate was 0% (power: 10%) and remained near baseline at 0% (power: 93%) in long time periods, holding all other parameters constant. Reducing the cluster to a low information setting with 1.1 cluster relative risk, 9 km cluster radius, in the large population setting, the false positive rate was 0% in both long (power: 2%) and short (power: 0%) time periods, with BIC opting to not identify any clusters with little information in both cluster time duration and fitted settings. With more information, increasing the relative risk to 2 and cluster radius to 18 km in the large population center, the false positive rate remained at 0% in both long time periods (power: 100%) and short time periods (power: 100%). In contrast, using AIC the false positive rates were higher in short time periods (false positive rate: 38%, power: 35%) with a 1.1 relative risk and 9 km cluster radius and remained above baseline with 2 relative risk and 18 km cluster radius (false positive rate: 12%, power: 100%). As the population center size grew from small to large and the cluster moved from short to longer time periods, the false positive rates using BIC remained near baseline detection established under the null.

4.2.4. Adjustment for overdispersion

In the fitted models, the false positive rates remained near baseline using QBIC even as overdispersion increased and cluster relative risks decreased. Clusters with larger relative risks continued to have lower false positive rates and higher power. Using QBIC for a cluster with relative risk of 1.5, cluster radius of 11 km, in a large population center, in long time periods (3 through 5), the false positive rates were 0% under both Poisson-gamma mixture with $\beta = 60$ (power: 58%) and with reduced overdispersion of $\beta = 100$ (power: 63%) (Fig. 4). Reducing the cluster relative risk to 1.1 and keeping all other parameters constant, the false positive rate remained at 0% (power: 1%) with both $\beta = 60$ and $\beta = 100$. Increasing the relative risk to 2 and keeping all other parameters constant, the false positive rate again remained near baseline at 0% (power: 100%) with $\beta = 60$ and at 2% (power: 100%) with $\beta = 100$. In contrast, using QAIC the false positive rates were again consistently higher, but decreased as overdispersion decreased from $\beta = 60$ (false positive rate: 21%, power: 100%) to $\beta = 100$ (false positive rate: 15%, power: 99%) for a cluster with relative risk of 1.5, cluster radius of 11 km, in a large population center, in long time periods (3 through 5). Smaller relative risks inside the cluster were more difficult to detect and had lower power and higher false positive rates across all scenarios.

As overdispersion increased and cluster radius decreased, QBIC maintained the false positive rate near 0%, though power was reduced. Using QBIC for a cluster with a radius of 9 km, relative risk of 1.5, in a large population center, in long time periods (3 through 5), the false positive rate was 0% with both $\beta = 60$ (power: 9%) and $\beta = 100$ (power: 12%). With a 9 km cluster radius and 1.1 cluster relative risk, the false positive rate was 0% (power: 0%) with $\beta = 60$ and remained at 0% (power: 1%) as overdispersion decreased to $\beta = 100$. With an 18 km cluster radius and 2 cluster relative risk, the false positive rate was still 0% (power: 100%) with both $\beta = 60$ and $\beta = 100$. Using QAIC, the false positive rates remained above baseline at 23% with both $\beta = 60$ (power: 94%) and $\beta = 100$ (power: 95%) for a cluster with 9 km radius, 1.5 relative risk, in a large population center, in long time periods (3 through 5).

In the true models, as overdispersion decreased and population size grew from small to large, the false positive rate remained near baseline while power increased. Using QBIC under Poisson-gamma mixtures with $\beta = 60$ and $\beta = 100$, for a cluster in a small population center, with a relative risk of 1.5, cluster radius of 11 km, in long time periods (3 through 5), the false positive rates were 0% (power: 0%). With more information, increasing the relative risk to 2 and cluster radius to 18 km in the small population center, the false positive rates were 4% (power: 30%) with $\beta = 60$ and 5% (power: 46%) with less overdispersion at

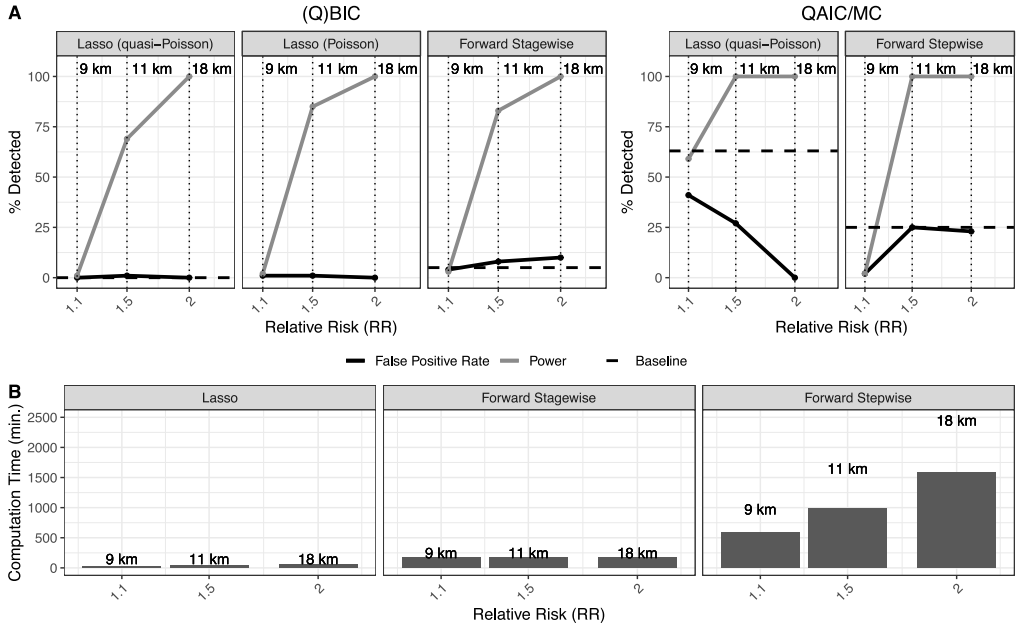


Fig. 3. A) Simulation results comparing the Lasso approach (using (Q)BIC/(Q)AIC) to forward stagewise (using BIC and AIC) and forward stepwise spatial scan (using MC). Baseline detection shows the false positive rate under the null model with no cluster. B) Computation time (in minutes) comparing the Lasso, forward stagewise, and forward stepwise spatial scan. 100 random datasets were simulated under quasi-Poisson (Poisson-gamma mixture with $\beta = 60$) under three information settings: low information (1.1 relative risk, 9 km radius), medium information (1.5 relative risk, 11 km radius), and high information (2 relative risk, 18 km radius).

$\beta = 100$. In the large population center, the false positive rates dropped back down to 0% (power: 100%) with both $\beta = 60$ and $\beta = 100$. Decreasing the relative risk to 1.1 and cluster radius to 9 km in the large population center, the false positive rates were still near baseline at 0% with $\beta = 60$ (power: 0%) and $\beta = 100$ (power: 1%) as well as in the small population center at 1% with $\beta = 60$ (power: 0%) and 0% with $\beta = 100$ (power: 0%). Using QBIC, overdispersion did not affect the false positive rate as much as power, though population size did affect both (Fig. 6). Using QAIC, the false positive rates were much higher in the small population setting with $\beta = 60$ (false positive rate: 44%, power: 2%) and decreased slightly with $\beta = 100$ (false positive rate: 32%, power: 0%) in a cluster with 1.1 relative risk, 9 km radius, in long time periods, and remained above baseline at $\beta = 60$ (false positive rate: 69%, power: 94%) and $\beta = 100$ (false positive rate: 59%, power: 99%) when the cluster increased to a relative risk of 2 and 18 km radius, all other parameters held constant.

Compared to clusters in short time periods, clusters in long time periods had lower false positive rates, which remained near baseline using QBIC (Fig. 6). Using QBIC for a cluster in short time periods (periods 1 and 2), with a relative risk of 1.5, cluster radius of 11 km, in a large population center, the false positive rate was 0% under the Poisson-gamma mixture with $\beta = 60$ (power: 6%) and $\beta = 100$ (power: 2%). Reducing the cluster radius to 9 km with a 1.1 cluster relative risk in both short and long time periods, the false positive rates were 0% with both $\beta = 60$ and $\beta = 100$. Increasing the relative risk to 2 and cluster radius to 18 km in the large population center, the false positive rate was 0% (power: 100%) with $\beta = 60$ and 1% (power: 100%) with $\beta = 100$ in short periods, similar to the rates in long time periods with $\beta = 60$ (false positive rate: 0%, power: 100%) and $\beta = 100$ (false positive rate: 0%, power: 100%). Using QAIC in a cluster with 1.1 relative risk and 9 km radius in short periods and in a large population center, the false positive rates were higher in short time periods with $\beta = 60$ (false positive rate: 25%, power: 26%) and increased as overdispersion decreased with $\beta = 100$ (false positive rate: 33%, power: 31%). In long time periods, false positive rates decreased or remained the same. Table S3 in the supplementary material shows the results for spatio-temporal cluster detection.

4.2.5. Detection of multiple clusters

We explored the false positive rate and power with multiple clusters in the study region under three clusters. The behavior of our Lasso method was consistent across both the single cluster and multiple cluster scenarios. Fig. 7 shows the three clusters considered and the detection probabilities associated with each cluster alone, 2 clusters with large population centers, 2 clusters each in one large and one small population center, and finally all three clusters in the study region as detected by QBIC under quasi-Poisson. Multiple clusters are distinctly detected similarly to how any single cluster.

Single clusters ($k = 1$) were all identified with some shadowing effect across time in the first two periods as well as in the southern tip, which we observed under the null model as well. Double clusters ($k = 2$) were distinctly regardless of if the clusters were in large or small population centers. Lastly, all three clusters ($k = 3$) were clearly identified with no additional cells detected outside of what was detected under the respective single cluster settings. This was consistent across detection both by QBIC under quasi-Poisson and BIC under Poisson, though there was slightly less of a shadowing effect under Poisson due to no overdispersion. Fig. 7 shows the false positive rates and power for detecting 1, 2, and 3 clusters in both large and small population centers by QBIC. Each of the three clusters was large enough that power was 100% across all settings, allowing us to better evaluate the false positive rate. As expected, using QBIC the single clusters in the small population center had a larger false positive rate (5%, power: 100%) than the two single clusters in large population centers (false positive rate 1: 0%, power 1: 100%; false positive rate 2: 0%, power 2: 100%). The false positive rate was otherwise below baseline across all cluster settings. In the Poisson case, the largest false positive rate was still near baseline and occurred with the large population center 2 and small population center clusters (false positive rate: 7%, power: 100%), which was larger than the false positive rate under quasi-Poisson (0%, power: 100%). Similarly, using both QAIC and AIC resulted in larger false positive rates, particularly in the settings with the small population center.

Similar to the single cluster detection, the Lasso will distinctly identify multiple clusters with enough information present. QBIC and BIC will minimize the false positive rate and detect the clusters. QAIC

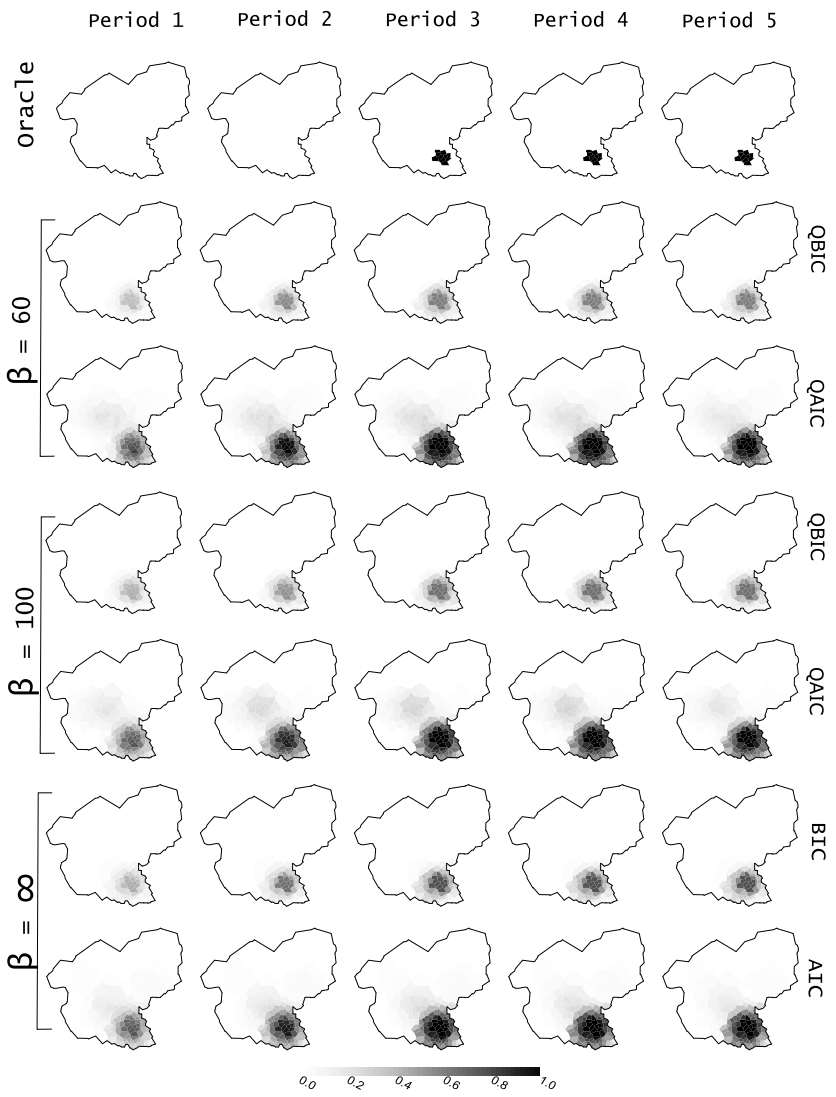


Fig. 4. Cluster detection with a cluster radius of 11 km, 1.5 relative risk, large population center, in longer time periods. 100 random datasets were simulated under quasi-Poisson (Poisson-gamma mixture with two shape parameters ($\beta = 60; \beta = 100$)) and Poisson distribution ($\beta = \infty$). Detection probabilities for each cell are the proportion of simulations in which the cell belongs to one (or more) detected clusters.

and AIC will have a higher false positive rate. Table S4 shows the detailed results for multiple cluster detection. Our Lasso approach distinctly identifies multiple clusters in the study region and behaves in the same way as under single cluster settings.

4.2.6. Summary

Under both overdispersion settings ($\beta = 60$, $\beta = \infty$), our Lasso approach outperformed both forward stagewise and forward stepwise spatial scan in minimizing the false positive rate and computation time. All three methods performed similarly in high information settings with low false positive rates and high power, but in low information settings forward stagewise and the forward stepwise spatial scan had higher false positive rates than our Lasso approach. We expected the shadowing effect from the null models (Fig. 2) to be persistent in any identified model and confirmed that the shadowing effect was driven by population and case counts, as shown in Table 2. Observed counts varied greatly by cell, where the larger population center had more counts to detect as the radius expanded, a problem common to many small area estimation studies.

As cluster radius and relative risk inside the cluster increased, power using QBIC under quasi-Poisson and using BIC under Poisson increased, as the false positive rates remained near baseline. Power was low when the cluster radius, cluster relative risk, or both were small, as well as in small population centers in short time periods. Power was highest for clusters with larger cluster radius, higher relative risk, in a large population center, with longer temporal duration. When the cluster radius or relative risk were large enough, using BIC and QBIC not only maintained the false positive rate near baseline, but also demonstrated power above 90% when the cluster radius was 18 km, relative risks were 2, and the cluster was in a large population center. As the information in the cluster increases either by the cluster radius, the relative risk, or the population size, the false positive rate will remain well-controlled even at smaller radii, while power will continue to increase (see Table S5). With less information, the Lasso method opts to not put a cluster into the model and minimize the false positive rate. This remained true even when detecting multiple clusters. Identifying multiple clusters with high information resulted in a false positive rate controlled at baseline and high power using QBIC and BIC. Given the public health implications of detecting cancer clusters and concerns this may raise in communities, we recommend using BIC and QBIC for selection to minimize false positive cluster detections.

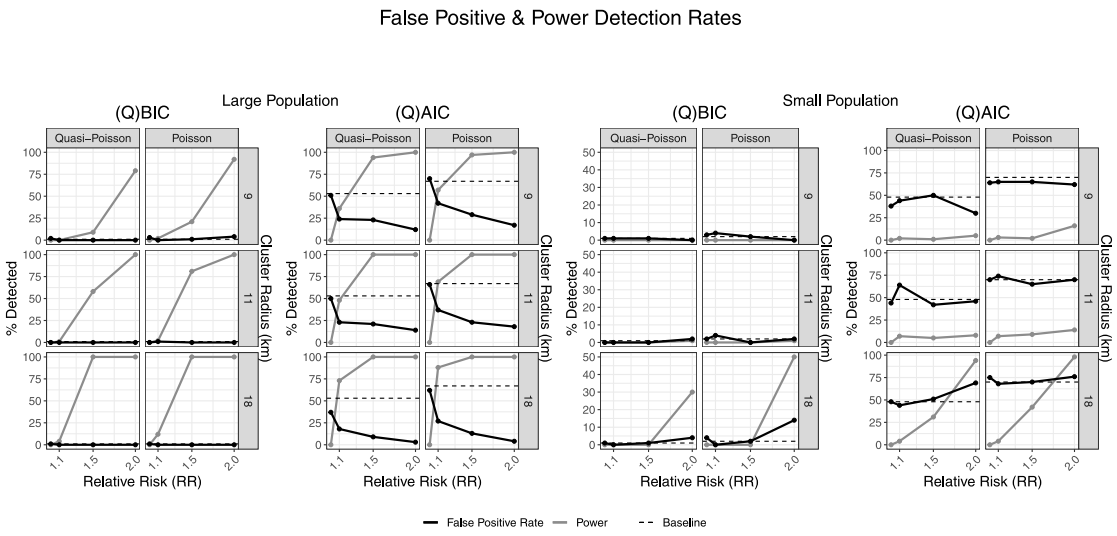


Fig. 5. Simulation results comparing power and false positive cluster detection rates in a large versus small population center scenarios across radii and relative risks. 100 random datasets were simulated under quasi-Poisson (Poisson-gamma mixture with $\beta = 60$) and Poisson distribution ($\beta = \infty$). Percent detected was calculated as the number of simulations in which the detected potential cluster had overlap with the true cluster (power). The false positive rate was calculated based on the percent of simulations which had no overlap with the true cluster. Results are shown by QBIC and QAIC (quasi-Poisson) and BIC and AIC (Poisson) where the cluster exists long time periods (periods 3 through 5).

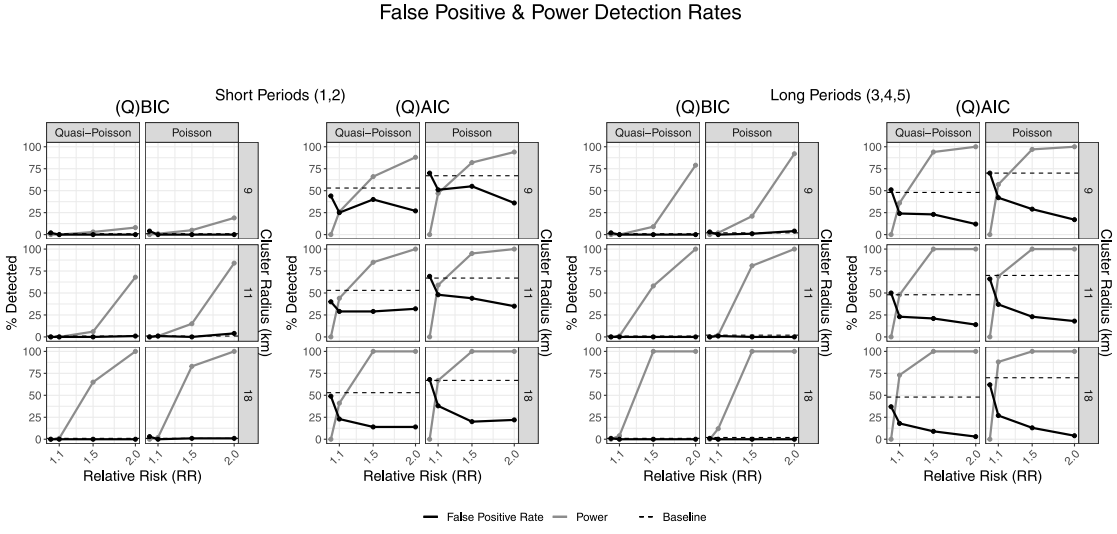


Fig. 6. Simulation results comparing power and false positive cluster detection rates in short (time periods 1 and 2) versus long time periods (time periods 3,4,5) across radii and relative risks. 100 random datasets were simulated under quasi-Poisson (Poisson-gamma mixture with $\beta = 60$) and Poisson distribution ($\beta = \infty$). Percent detected was calculated as the number of simulations in which the detected potential cluster had overlap with the true cluster (power). The false positive rate was calculated based on the percent of simulations which had no overlap with the true cluster. Results are shown by QBIC and QAIC (quasi-Poisson) and BIC and AIC (Poisson) where the cluster exists in the large population center.

5. Data example

We present results for the breast cancer incidence rates in the same study region of Japan as in the data-driven simulations. We set $r_{max} =$ 20 km, which results in $K = 66\,870$ potential spatio-temporal clusters centered at 208 cell centroids. We analyze the Japanese breast cancer data use our Lasso approach with a quasi-Poisson (using QBIC and QAIC for selection) and Poisson (using BIC and AIC) models, and compare our results for the forward stagewise approach (using BIC and AIC) and to the forward stepwise spatial scan (using Monte Carlo simulations) (Fig. 8).

Using the Lasso, we estimated the relative risks and mapped them based on QBIC or QAIC to adjust for overdispersed counts using quasi-Poisson or based on BIC or AIC for Poisson. Under both models, we identified clusters in the southern tip of the study region, shown in Table 4. Using both QBIC and BIC, we detected the same two overlapping-clusters in the southern tip of the study region. The overlap represents the core cluster of elevated risk and the non-overlapping edges are elevated risk areas surrounding the core cluster. The relative risk in the core of the two overlapping clusters was 1.12, and dropped to 1.03 and 1.09 moving from the core of the cluster outward. The estimated background relative risks in the 5 time periods were 0.958, 0.959, 0.961, 0.963, and 0.966, respectively.

Using both QAIC and AIC, we detected the same 10 clusters, including a reduced relative risk area in the central part of the study region. Eight of these clusters had an elevated relative risk and 2 clusters had reduced relative risks. The 8 clusters with elevated risk had relative

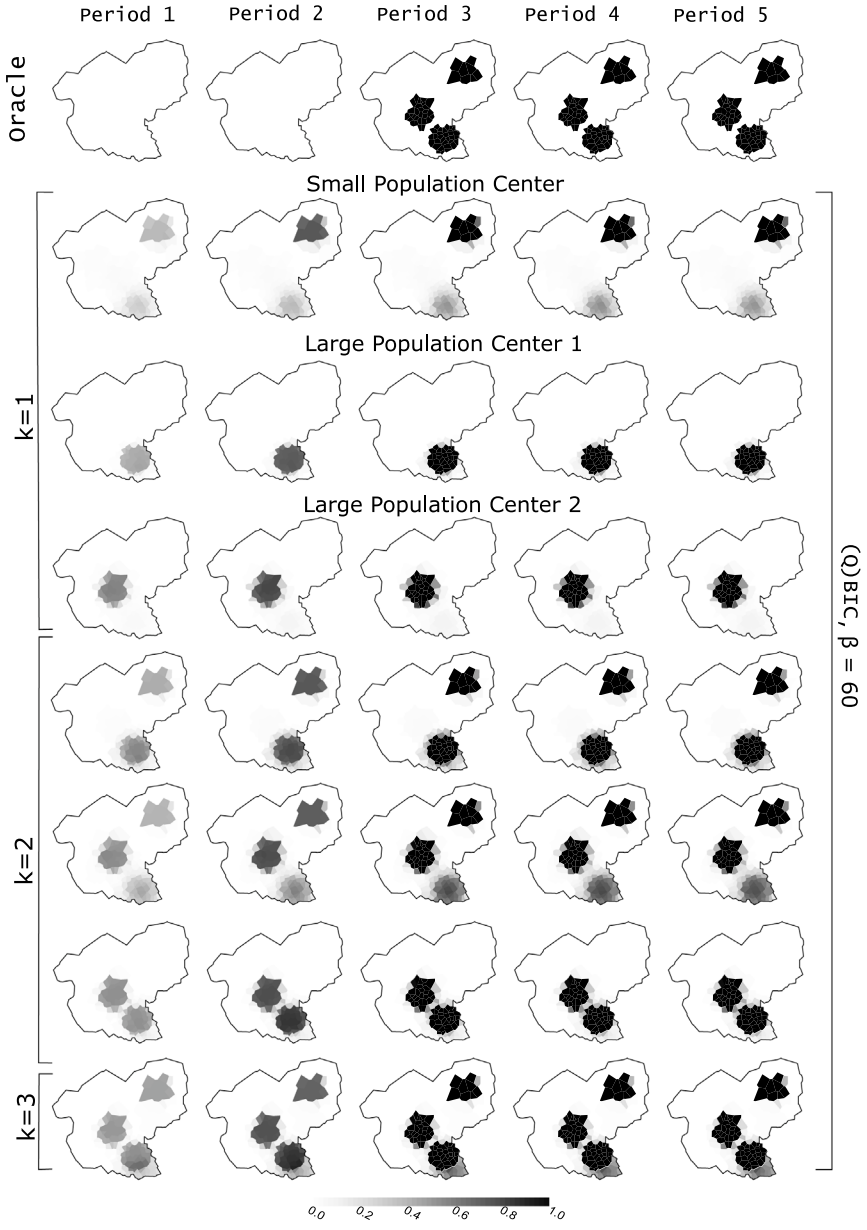


Fig. 7. Cluster detection with multiple clusters: two clusters centers in large population centers (2 relative risk) and one cluster center in a small population center (4 relative risk), all with an 18 km radius. 100 random datasets were simulated under quasi-Poisson (Poisson-gamma mixture with $\beta = 60$). Detection probabilities for each cell are the proportion of simulations in which the cell belongs to one (or more) detected clusters.

Table 4
Number of clusters detected in real data example.

Method	Scenario	(Q)BIC	(Q)AIC	Monte Carlo
Lasso	Poisson	2	10	–
Lasso	Quasi-Poisson	2	10	–
Forward Stagewise	Poisson	5	33	–
Forward Stepwise Spatial Scan	Poisson	–	–	1

The number of clusters found by model and criterion in the Japanese breast cancer example using the spatio-temporal model. For quasi-Poisson, the information criterion used are QAIC and QBIC. For Poisson, the information criterion used are AIC and BIC.

risks of 1.0096, 1.0116, 1.0120, 1.0158, 1.0397, 1.0415, 1.0517, and 1.0778. The 2 clusters with reduced risk had relative risks of 0.949 and 0.993. We also see that the background relative risks in the 5 time periods are reduced. In each period, the lower background relative risk is due to absorption of the other relative risks. With more clusters identified, the fewer elevated relative risks are available to smooth out the background relative closer to 1. Relative risks in the overlapping areas ranged from 1.01 to 1.29 (median: 1.13, mean: 1.15) in the elevated relative risk clusters and ranged from 0.942 to 0.993 (median: 0.942, mean: 0.951) in the reduced relative risk clusters, with overall mean relative risk and median relative risks of 1.00 and 1.02 in the study region. The estimated background relative risks in the 5 time periods were 0.928, 0.930, 0.934, 0.938, and 0.944.

Compared to other methods, using BIC as selection criteria with the forward stagewise method, 5 clusters were identified. The estimated background relative risks in the 5 time periods are 0.963, 0.960, 0.957, 0.954, and 0.953. The estimated relative risks in the overlapping clusters were 0.99, 1.01, 1.04, 1.05, 1.06, 1.10, and 1.11. Using AIC, 33 clusters were identified. The estimated background relative risks in the 5 time periods were 0.935, 0.915, 0.919, 0.904, 0.905. The estimated reduced relative risks in the overlapping clusters ranged from 0.835 to 0.997 (mean: 0.918, median: 0.915) and the estimated elevated relative risks in the overlapping clusters ranged from 1.00 to 1.23 (mean: 1.129,

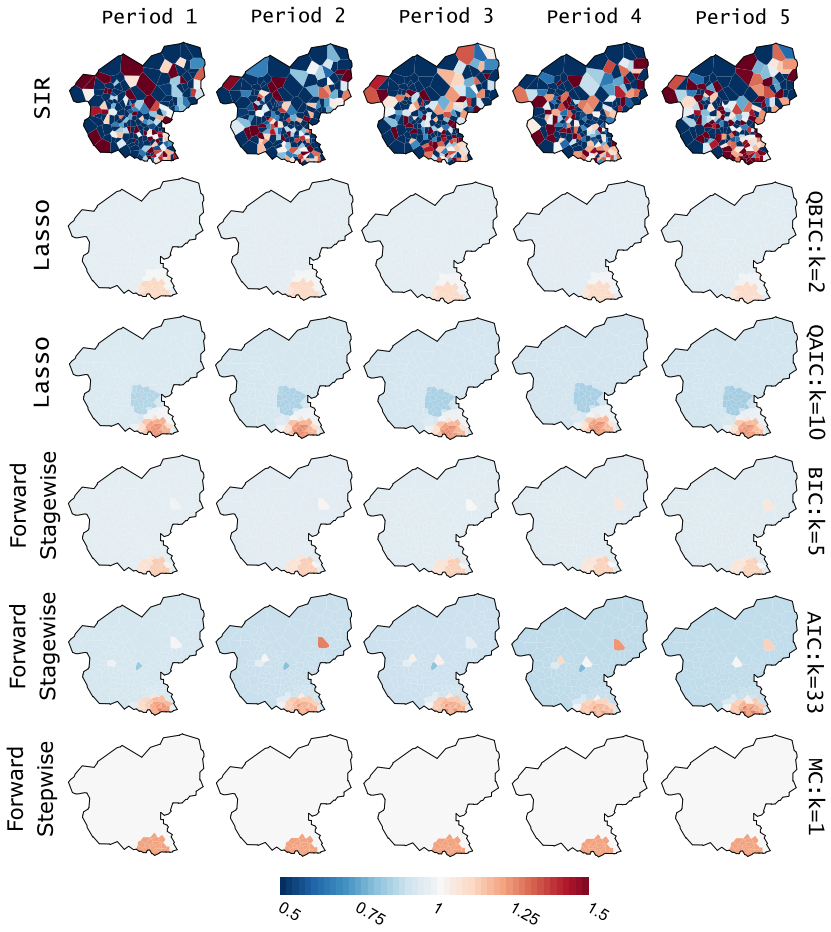


Fig. 8. Cluster identification in Japan across three methods. Using the Lasso-based method, both QBIC and BIC identified the same 2 clusters and both QAIC and AIC identified the same 10 clusters. All three methods identify a cluster of elevated geographic risk in the southern tip of the study region.

median: 1.136). Using the forward stepwise spatial scan, one clusters was identified with a relative risk of 1.1771 (p-value: 0), using 1 000 Monte Carlo simulations. Subsequent clusters were not selected after adjustment for multiple testing at the 5% level.

Our Lasso approach, forward stagewise, and forward stepwise spatial scan methods all identify the same cluster in the southern tip of the region. As expected, the forward stagewise method identified the same two persistent overlapping clusters in the southern region as our Lasso method. However it also identified a single cell in the eastern part of the region, which moved from reduced relative risk in the first three periods to elevated relative risk in the last two periods. This may be due to the tuning of ϵ (Section 4.1.1). The forward stepwise spatial scan does not identify the unique gradations of relative risks from the overlapping clusters and estimates a single relative risk in the most likely cluster. In contrast, our Lasso approach identifies two overlapping clusters in the same region and three unique relative risks. As demonstrated, the Lasso approach can be used with complicated spatial risk structures and can be used to characterize gradients of risk across a study region by using multiple overlapping clusters.

6. Discussion

Since Openshaw's geographical analysis machine, selecting disease clusters from a comprehensive set of potential clusters has been challenging in public health surveillance. Frequentist approaches have focused on hypothesis testing of multiple clusters, while Bayesian methods have instead focused on various prior spatial distribution specifications (Gangnon and Clayton, 2000, 2003, 2007; Lawson, 2000; Wakefield and Kim, 2013) and have likewise been extended to the spatio-temporal framework (Clark and Lawson, 2002; Yan and Clayton, 2006; Gangnon, 2010a). We do not formally consider a Bayesian approach here.

The innovation of using a likelihood-based approach enables new methods to be applied to the spatial cluster detection problem. For identifying multiple overlapping clusters, the forward stagewise method has previously been shown to have improved control of the false positive rate than the forward stepwise spatial scan variant. Our Lasso method produces a similar coefficient path to that of forward stagewise regression, but is not limited by the tuning of step size parameter ϵ and is computationally more efficient. By leveraging the sparsity of the Lasso, we are able to efficiently detect multiple overlapping clusters with complicated risk structures without sequential deletion. We have compared quasi-Poisson and Poisson and shown that cluster detection using QBIC and BIC better controlled the false positive rate near baseline. Detection using QAIC and AIC identified several additional clusters in addition to the clusters detected by QBIC and BIC, resulting in higher false positive rates, and given our knowledge of the area and that clusters are rare, fewer clusters are more reasonable for this study region. The number of clusters that QBIC and BIC detected is most plausible for breast cancer incidence in the given area. Where this rare assumption may not be reasonable, we recommend using AIC and QAIC, which will allow for more clusters into the model.

Our Lasso approach to spatial and spatio-temporal cluster detection is limited by the amount of information available in the cluster. When the relative risk, radius, population size, and temporal duration of the cluster were large enough, not only was the false positive rate near baseline but power was also above 90%. In clusters with little information, our method using QBIC and BIC as selection tools will opt to

not identify any cluster rather than identify a false positive cluster. We are also limited by the implementation of the Lasso regularization. In the `glmnet` implementation of the Lasso for GLM, a grid of decreasing λ values is proposed, in which cyclical coordinate descent is used to calculate cluster estimates. The exact λ at which a new cluster enters the model may be obscured by this gridding. However, we find from our empirical results that this does not affect detection of the cluster(s), though this leads to the shadowing effect observed around the cluster.

Through both the simulation study and the data example, we show that by using the Lasso with a sparse matrix representation of all potential clusters, we are able to identify when each new cluster enters the active set. By using information criteria instead of costly cross-validation, we are able to not only identify clusters due to the Lasso's shrinkage property, but also to estimate the relative risks inside those clusters, identifying geographic areas in need of more thorough epidemiologic investigation into possible mechanisms for differing risk patterns. We have implemented this method and creation of overlapping spatial and spatio-temporal clusters in the `clusso` R package, available from https://mkamenet3.github.io/clusso/. The original data are not publicly available due to privacy considerations, but have been discussed elsewhere (Yan and Clayton, 2006).

There are several ways by which this work can be extended. The formulation of potential clusters can further be expanded and constructed based on constant case counts inside each potential cluster (Bruce et al., 1990) instead of distance between cell centroids or use of the C information criterion (Takahashi and Shimadzu, 2020), which we leave for future work. Literature on post-selection inference (Lee et al., 2016; Tibshirani et al., 2016) using the Lasso can be used to develop confidence bounds and further extend this work. By defining a confidence set for the true cluster, statistical inference could be made about the estimated cluster covariates (Lee et al., 2017). We can also consider the full set of potential clusters across the study region in an ensemble, and use a model-averaging approach to estimate both the relative risks, but also calculate confidence bounds. Recently developed methods for high-dimensional data and multi-model inference can further be extended to the spatial cluster detection literature in order to continue developing tools available for public health surveillance.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.sste.2021.100462.

References

Akaike, Hirotugu, 1973. Information theory as an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademai Kaido, Budapest, pp. 267–281.

Besag, Julian, Newell, James, 1991. The detection of clusters in rare diseases. J. R. Stat. Soc. A 154 (1), 143–155.

Bruce, W., Turnbull, B.W., Eric, E.J., Iwano, J., William, S., Burnett, W.S., Holly, L., Howe, H.L., Larry, L.C., Clark, C., 1990. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. Am. J. Epidemiol. 132 (supp1), 136–143.

Cavanaugh, J.E., Shumway, R.H., 1997. A bootstrap variant of AIC for state-space model selection. Statist. Sinica 7 (2), 473–496.

Clark, A.B., Lawson, A.B., 2002. Spatio-temporal cluster modelling of small area health data. In: Lawson, A., Denison, D. (Eds.), Spatial Cluster Modelling. Chapman and Hall/CRC, pp. 235–258.

Duczmal, Luiz, Assuncáo, Renato, 2004. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. Comput. Statist. Data Anal. 45 (2), 269–286.

Efron, Bradley, Hastie, Trevor, Johnstone, Iain, Tibshirani, Robert, 2004. Least angle regression. Ann. Statist. 32 (2), 407–499.

Friedman, Jerome, Hastie, Trevor, Tibshirani, Robert, 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1–22.

Gangnon, Ronald E., 2010a. A model for space – time cluster detection using spatial clusters with flexible temporal risk patterns. Stat. Med. 29 (May), 2325–2337.

Gangnon, Ronald E., 2010b. Local multiplicity adjustments for spatial cluster detection. Environ. Ecol. Stat. 17 (1), 55–71.

Gangnon, Ronald E., Clayton, Murray K., 2000. BayesIan detection and modeling of spatial disease clustering. Biometrics 56 (3), 922–935.

Gangnon, Ronald E., Clayton, Murray K., 2003. A hierarchical model for spatially clustered disease rates. Stat. Med. 3228 (20), 3213–3228.

Gangnon, Ronald E., Clayton, Murray K., 2004. Likelihood-based tests for localized spatial clustering of disease. Environmetrics 15 (8), 797–810.

Gangnon, Ronald, Clayton, Murray K., 2007. Cluster detection using Bayes factors from overparameterized cluster models. Environ. Ecol. Stat. 14 (1), 69–82.

Han, Junhee, Zhu, Li, Kulldorff, Martin, Hostovich, Scott, Stinchcomb, David G., Tatalovich, Zaria, Lewis, Denise Riedel, Feuer, Eric J., 2016. Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics. Int. J. Health Geogr. 15 (27), 1–11.

Hastie, Trevor, Taylor, Jonathan, Tibshirani, Robert., Walther, Guenther, Boyd, Steven, Friedman, Jerome, Rosset, Saharon, Van Roy, Ben, Zhu, Ji, 2007. Forward stagewise regression and the monotone lasso. Electron. J. Stat. 1, 1–29.

Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome, 2008. The Elements of Statistical Learning, second ed. Springer, New York.

Hoerl, Arthur E., Kennard, Robert W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12 (1), 55–67.

Huang, Lan, Kulldorff, Martin, Gregorio, David, 2007. A spatial scan statistic for survival data. Biometrics 63 (1), 109–118.

Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small samples. Biometrika 76 (2), 297–307.

Jung, Inkyung, Kulldorff, Martin, Richard, Otukei John, 2010. A spatial scan statistic for multinomial data. Stat. Med. 29, 1910–1918.

Kulldorff, Martin., 1997. A spatial scan statistic. Comm. Statist. Theory Methods 26 (6), 1481–1496.

Kulldorff, Martin, 2001. Prospective time periodic geographical disease surveillance using a scan statistic. J. R. Stat. Soc. Ser. A 164, 61–72.

Kulldorff, Martin, 2015. SaTScan User Guide V9.4. Technical report, Harvard Medical School and Harvard Pilgrim Health Care Institute.

Kulldorff, Martin., Athas, W.F., Feuer, E.J., Miller, B.A., Key, C.R., 1998. Evaluating cluster alarms: a space–time scan statistic and brain cancer in Los Alamos, New Mexico. Am J Public Health 88 (C), 1377–1380.

Kulldorff, Martin, Huang, Lan, Konty, Kevin, 2009. A scan statistic for continuous data based on the normal probability model. Int. J. Health Geogr. 9, 1–9.

Kulldorff, Martin, Huang, Lan, Pickle, Linda, Duczmal, Luiz, 2006. An elliptic spatial scan statistic. Stat. Med. 25 (22), 3929–3943.

Kulldorff, Martin, Nagarwalla, Neville, 1995. Spatial disease clusters: Detection and inference. Stat. Med. 14 (8), 799–810.

Lawson, Andrew B., 2000. Cluster modelling of disease incidence via RJMCMC methods: A comparative evaluation. Stat. Med. 19 (17–18), 2361–2375.

Lebreton, Jean-Dominique, Burnham, Kenneth P., Clobert, Jean, Anderson, David R., 1992. Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. Ecol. Monograph 62 (1), 67–118.

Lee, Junho, Gangnon, Ronald E., Zhu, Jun, Liang, Jingjing, 2017. Uncertainty of a detected spatial cluster in 1D : quantification and visualization. Stat 6, 345–359.

Lee, Jason D., Sun, Dennis L., Sun, Yuekai, Taylor, Jonathon, 2016. Exact post-selection inference, with application to the lasso. Ann. Statist. 44 (3), 907–927.

Li, Xiao-Zhou, Wang, Jin-Feng, Yang, Wei-Zhong, Li, Zhong-Jie, Lai, Sheng-Jie, 2011. A spatial scan statistic for multiple clusters. Math. Biosci. 233 (2), 135–142.

McCullagh, P., Nelder, J.A., 1991. Generalized Linear Models, second ed. Chapman and Hall, London.

Neill, Daniel B., 2012. Fast subset scan for spatial pattern detection. J. R. Stat. Soc. B 74 (2), 337–360.

Openshaw, S., Craft, A.W.A.W, Charlton, M., Birch, J.M.J.M, Craft, A.W.A.W, Birch, J.M.J.M., 1988. Investigation of leukaemia clusters by use of a geographical analysis machine. Lancet 331 (8580), 272–273.

Pan, Wei, 2004. Akaike's information criterion in generalized estimating equations. Biometrics 57 (March), 120–125.

Park, Trevor, Casella, George, 2008. The Bayesian Lasso. J. Amer. Statist. Assoc. 103 (482), 681–686.

Ribeiro, Sérgio Henrique Rodrigues, Costa, Marcelo Azevedo, 2012. Spatial and Spatio-temporal Epidemiology Optimal selection of the spatial scan parameters for cluster detection : A simulation study. Spatial Spatio-Temp. Epidemiol. 3, 107–120.

Schwarz, Gideon, 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464.

Shu, Lianjie, Jiang, Wei, Tsui, Kwok-Leung, 2012. A standardized scan statistic for detecting spatial clusters with estimated parameters. Nav. Res. Logist. 59 (6), 397–410.

Stone, M., 1977. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1), 44–47.

Takahashi, Kunihiko, Kulldorff, Martin, Tango, Toshiro, Yih, Katherine, 2008. A flexibly shaped space–time scan statistic for disease outbreak detection and monitoring. Int. J. Health Geogr. 7 (1), 14.

Takahashi, Kunihiko, Shimadzu, Hideyasu, 2020. Detecting multiple spatial disease clusters : information criterion and scan statistic approach. Int. J. Health Geogr. 19 (33), 1–11.

Tango, Toshiro, Takahashi, Kunihiko, 2005. A flexibly shaped spatial scan statistic for detecting clusters. Int. J. Health Geogr. 4 (1), 11.

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

Tibshirani, Ryan J., 2015. A general framework for fast stagewise algorithms. J. Mach. Learn. Res. 16, 2543–2588.

Tibshirani, Ryan J., Taylor, Jonathan, 2012. Degrees of freedom in lasso problems. Ann. Statist. 40 (2), 1198–1232.

Tibshirani, Ryan J., Taylor, Jonathan, Lockhart, Richard, Tibshirani, Robert, 2016. Exact post-selection inference for sequential regression procedures. J. Amer. Statist. Assoc. 111 (514), 600–620.

Wakefield, Jonathon, Kim, Albert, 2013. A Bayesian model for cluster detection. Biostatistics 14 (4), 752.

Wedderburn, R.W.M., 1974. Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. Biometrika 61 (3), 439–447.

Xu, Jiale, Gangnon, Ronald E., 2016. Stepwise and stagewise approaches for spatial cluster detection. Spatial Spatio-Temp. Epidemiol. 17, 59–74.

Yan, Ping, Clayton, Murray K., 2006. A cluster model for space – time disease counts. Stat. Med. 25 (5), 867–881.

Zhang, Zhenkui, Assunção, Renato, Kulldorff, Martin, 2010. Spatial scan statistics adjusted for multiple clusters. J. Probab. Stat. 2010.