Spatial Statistics 70 (2025) 100933

Contents lists available at ScienceDirect

Spatial Statistics

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

Spatial and spatio-temporal cluster detection using stacking

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

[a] Department of Population Health Sciences, University of Wisconsin-Madison, Madison WI, USA
[b] Department of Statistics, University of Wisconsin-Madison, Madison WI, USA
[c] Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison WI, USA

ARTICLE INFO

Keywords:
Model-averaging
Multi-model inference
Spatial cluster detection
Spatial scan statistic
Spatio-temporal cluster detection
Stacking

ABSTRACT

Patterns in disease across space and time are important to epidemiologists and health professionals because they may indicate underlying elevated disease risk. In some cases, elevated risk may be driven by environmental exposures, infectious diseases or other factors where timely public health interventions are important. The spatial and spatio-temporal scan statistics identify a single most likely cluster or equivalently select a single correct model. We instead consider an ensemble of single cluster models. We use stacking, a model-averaging technique, to combine relative risk estimates from all of the single cluster models into a sequence of meta-models indexed by the effective number of parameters/clusters. The number of parameters/spatio-temporal clusters is chosen using information criteria. A simulation study is conducted to demonstrate the statistical properties of the stacking method. The method is illustrated using a dataset of female breast cancer incidence data at the municipality level in Japan.

1. Introduction

Identifying disease clusters across space and time is an important aspect of public health surveillance. In epidemiology, standardized incidence rates (SIRs) are commonly used to visualize risk across a study region. SIRs, however, result in crude maps where it can be challenging to identify subregions and timeframes with elevated incidence or prevalence. Due to privacy and identifiability issues, public health data are often aggregated according to administrative boundaries, or cells, such as municipalities or census tracts. Focusing on such aggregated data, we propose a stacking approach based on model-averaging for spatial and spatio-temporal cluster detection.

Spatial clustering is an alternative to other spatial smoothing approaches that have been proposed. Early efforts to identify spatial clusters used moving windows, or potential clusters, centered at cell centroids. As the windows moved across the study region, they were evaluated based on the case counts inside them. Formalizing this approach, Openshaw et al. (1988) proposed the geographic analysis machine (GAM), which evaluated many overlapping circles with varying radii as potential clusters. Turnbull et al. (1990), Besag and Newell (1991), and Kulldorff and Nagarwalla (1995) proposed rigorous approaches to hypothesis testing based on the GAM. Scan statistics have been frequently used in epidemiology to detect disease clusters (Abolhassani and Prates, 2021). The likelihood-based spatial scan approach of Kulldorff and Nagarwalla (1995) has been extended to other shapes (Kulldorff et al., 2007, 2006; Duczmal and Assuncáo, 2004; Tango and Takahashi, 2005; Takahashi et al., 2008) and other distributions (Huang et al., 2007; Kulldorff et al., 2009; Jung et al., 2007, 2010).

Adaptations of scan statistics to identify multiple clusters in a study region have been proposed, including sequential detection (Zhang et al., 2010) and formal incorporation of multiple clusters into the alternative hypothesis (Li et al., 2011). In the frequentist

* Corresponding author.
E-mail addresses: maria.kamenetsky@nih.gov (M.E. Kamenetsky), junzhu@wisc.edu (J. Zhu), ronald@biostat.wisc.edu (R.E. Gangnon).

https://doi.org/10.1016/j.spasta.2025.100933
Received 5 December 2024; Received in revised form 27 June 2025; Accepted 17 September 2025
Available online 24 September 2025

paradigm, Xu and Gangnon (2016) and Kamenetsky et al. (2022) recast the spatial cluster detection problem as a high-dimensional variable selection problem; Xu and Gangnon used a forward stagewise algorithm, while Kamenetsky et al. used Lasso regularization to select a parsimonious set of clusters. These approaches, however, were still limited to identifying a single correct model and failed to account for model uncertainty.

While Bayesian approaches have primarily focused on spatial smoothing for both spatial and spatio-temporal disease mapping, some approaches have been developed for cluster detection. In the spatial setting, several methods have been proposed for identifying clusters with elevated disease risk relative to neighboring risk patterns. Approaches such as the one by Wakefield and Kim (2013), specify wide priors for the cluster relative risks narrow priors on the background rates and then use Markov chain Monte Carlo to compute the posterior relative risks. Several Bayesian methods for spatial cluster detection have been developed using reversible jump Markov chain Monte Carlo to detect abrupt changes from the background to the cluster relative risk (Gangnon and Clayton, 2003; Charras-Garrido et al., 2013; Hossain and Lawson, 2005). Knorr-Held and Raßer (2000) considered cluster detection as a change-point problem and proposed a nonparametric Bayesian approach. Anderson et al. (2014) proposed a two-stage solution, first using a spatially-adjusted hierarchical agglomerative clustering algorithm and next fitting a separate Bayesian model to the candidate cluster structure and later an approach using adjacency modeling (Anderson et al., 2016). Other approaches have incorporated the temporal dimension using Bayesian hierarchical mixture models (Napier et al., 2019; Lee and Lawson, 2016), including latent structure modeling approaches (Lawson et al., 2010; Choi et al., 2011).

The weighting of the likelihood based on the strength of evidence has previously been introduced for multi-model inference. Stone (1974) originally proposed combining variables based on cross-validation by weighting them in his "model-mix" method. Wolpert (1992) introduced stacking or stacked generalization as a way to minimize the error rate of any single estimator by combining estimators. Breiman (1996) extended Wolpert's idea to the regression setting. Leblanc and Tibshirani (1996) explored combining regression coefficient estimates using cross-validation, and Burnham and Anderson (2010) introduced an information-theoretic approach to model-averaging using AIC-based weights to combine coefficient estimates. Dudoit and Laan (2005) further developed the theoretical foundations of model-averaging demonstrating that the overall predictor performs asymptotically as well as the best possible weighted combination of models if the true model is included. Bayesian model averaging (BMA) (Leamer, 1978; Kass and Raftery, 1995; Hoeting et al., 1999; Raftery et al., 1993; Chatfield, 1995; Raftery and Richardson, 1996) also considered an entire set of models, using posterior model probabilities as weights.

Here, we develop a model-averaging approach for spatial and spatio-temporal cluster detection using stacking. Specifically, using the information in all single cluster models, we can obtain a more accurate estimate of the relative risk inside the cluster(s) and more fully assess uncertainty about the cluster risk, size, and shape.

The structure of the paper is as follows. The statistical model is developed in Section 2. Stacking methodology is introduced in Section 3. In Section 4 we evaluate the statistical properties of our method using a simulation study and in Section 5 we illustrate our method by the analysis of female breast cancer incidence in Japan. We summarize our findings in a discussion in Section 6.

2. Modeling

2.1. Statistical model

Consider a study region in $\mathbb{R}^2$, divided into N cells over T time periods. Let y_{it} be the observed number of cases of disease for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Then

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

where

$$\mu_{it} = \rho_{it} E_{it}, \tag{2}$$

where ρ_{it} is the unobserved relative risk; E_{it} is the expected number of cases from a null (no cluster) model that can account for any known confounding variables and can be obtained from any regression model, including spatial smoothing regression models.

We define a series of single cluster models where each has one unique relative risk estimate different from the background. Each model serves as a potential candidate for a single contiguous region of differing risk. We model the relative risk in cell i at time t as:

$$\log \rho_{it} = \theta_j \mathbb{1}\{d(\boldsymbol{x}_i, \boldsymbol{c}_j) < r_j, l_j \le t \le u_j\}, \tag{3}$$

θ_j is the log relative risk in single cluster model j, and $\mathbb{1}\{d(\boldsymbol{x}_i, \boldsymbol{c}_j) \le r_j, l_j \le t \le u_j\}$ is an indicator function defining each jth single cluster model. It takes a 1 if the Euclidean distance $d(\cdot)$ between a cell with center $\boldsymbol{x}_i$ and cluster centered at $\boldsymbol{c}_j$ is less than or equal to radius r_j, and cell i is inside the (closed) time interval between l_j and u_j, and is 0 otherwise. To simplify notation, we let $x_{it}^{(j)} = \mathbb{1}\{d(\boldsymbol{x}_i, \boldsymbol{c}_j) \le r_j, l_j \le t \le u_j\}$. Since disease counts in public health are frequently overdispersed, we adopt a quasi-likelihood approach (McCullagh and Nelder, 1991) described in Supplementary Material.

For illustration, we use circles as a type of basis function to define the single cluster models, though the approach can be extended to other shapes. Circles are centered at cell centroids with varying radii up to a prespecified maximum radius, r_{max}, as spatial windows. When extended to time, these potential clusters become cylinders with elevated (reduced) risk applied to a continuous time interval.

The total number of single cluster models is $K = ST^*$, where S is the number of circles centered at each geographic centroid with radii that range from 0 to r_{max} and $T^* = T(T+1)/2$ is the number of time intervals.

Table 1
Observed and expected counts for four cells. A simplified example where we consider a study region with 4 cells, each with an observed (y_i) and expected (E_i) count per cell.

Cell i	y_i	E_i
A	20	10
B	15	10
C	9	15
D	5	15

Table 2
Calculations for each single cluster model. Each single cluster model comprised of the cells in Table 1. For each single cluster model, we calculate the observed case counts ($y_{\cdot\cdot}^{(j)}$), expected case counts ($E_{\cdot\cdot}^{(j)}$), the relative risk for each single cluster model ($\exp\hat{\theta}_j$), initial likelihood weights for each single cluster model (w_j) in the single stack, updated likelihood weights for each single cluster model in the single stack with localization ($w_j^{(1)}$), and updated likelihood weights for each single cluster model for the second stack (multiple stacks with localization, $w_j^{(2)}$).

Single cluster model (j)	Cells	y_j	E_j	$\exp\hat{\theta}_j$	w_j	$w_j^{(1)}$	$w_j^{(2)}$
1	A	20	10	2.00	9.60×10^{-6}	9.62×10^{-6}	0
2	B	15	10	1.50	7.15×10^{-8}	7.16×10^{-8}	0
3	A & B	35	20	1.75	9.11×10^{-1}	9.12×10^{-1}	0
4	C	9	10	0.90	1.05×10^{-8}	0	$5.96\times^{-6}$
5	D	5	10	0.50	3.02×10^{-7}	0	1.72×10^{-4}
6	C & D	14	20	0.70	1.76×10^{-3}	0	$9.99\times^{-1}$
7	A & C	29	25	1.16	5.54×10^{-2}	5.55×10^{-2}	0
8	B & D	20	25	0.96	3.19×10^{-2}	3.20×10^{-2}	0

2.2. Likelihood functions

Let the log relative risk in a given single cluster model be given as:

$$\hat{\theta}_j = \log\left(\frac{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}x_{it}^{(j)}}{\sum_{i=1}^{N}\sum_{t=1}^{T} E_{it}x_{it}^{(j)}}\right) = \log\left(\frac{y_{\cdot\cdot}^{(j)}}{E_{\cdot\cdot}^{(j)}}\right), \tag{4}$$

where $y_{\cdot\cdot}^{(j)} = \sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}x_{it}^{(j)}$ and $E_{\cdot\cdot}^{(j)} = \sum_{i=1}^{N}\sum_{t=1}^{T} E_{it}x_{it}^{(j)}$. Then the relative risk in cell i at time t in model j is:

$$\hat{\rho}_{it}^{(j)} = \exp\hat{\theta}_j x_{it}^{(j)}, \tag{5}$$

where $\hat{\rho}_{it}^{(j)}$ denotes the maximum likelihood estimate (MLE) for the jth cluster. The log-likelihood (or log quasi-likelihood) of the model parameters from a given single cluster model j is:

$$\log\mathcal{L}^{(j)} = \sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}\hat{\theta}_j x_{it}^{(j)} - e^{\hat{\theta}_j x_{it}^{(j)}}E_{it}. \tag{6}$$

For the quasi-Poisson case, the log quasi-likelihood is computed as $\log\mathcal{L}^{(j)}/\tilde{\phi}$, where the dispersion parameter ϕ can be estimated by $(N-1)^{-1}\sum_i^N\sum_t^T(y_{it}-\hat{\mu}_{it})^2/\hat{\mu}_{it}$, where $\hat{\mu}_{it} = \hat{\rho}_{it}\hat{E}_{it}$ and N is the number of cells (Burnham and Anderson, 2002).

3. Stacking

Our parameter of interest is ρ_{it}, the relative risk inside in cell i across time t. As an illustrative example, suppose that each of the four cells each with their own observed (y_i) and expected (E_i) counts (Table 1). Each cell represents a single cluster model, and also cells A and B represent a fifth model, cells C and D represent a sixth model, cells A and C represent a seventh model, and cells B and D represent the final eighth model. In this example, there are $K = 8$ eight single cluster models.

For each single cluster model j, we calculate the total observed counts ($y_{\cdot\cdot}^{(j)}$), expected counts ($E_{\cdot\cdot}^{(j)}$), and the estimated relative risk ($\exp\hat{\theta}_j x_{it}^{(j)}$) for that model. In the illustrative example, there are a total of $K = 8$ single cluster models and $y_{\cdot\cdot}^{(j)}$ and $E_{\cdot\cdot}^{(j)}$ are given in Table 2 as the sum of observed and expected counts for cells in that have membership in the single cluster model. The log relative risk, $\hat{\theta}_j$ is calculated as $\log(y_{\cdot\cdot}^{(j)}/E_{\cdot\cdot}^{(j)})$.

For each single cluster model j, we set the background relative risk to 1. The initial model weights, w_j, are the weight of evidence in favor of the jth single cluster model given by Eq. (7). The model weights reflect the uncertainty in cell membership in the cluster

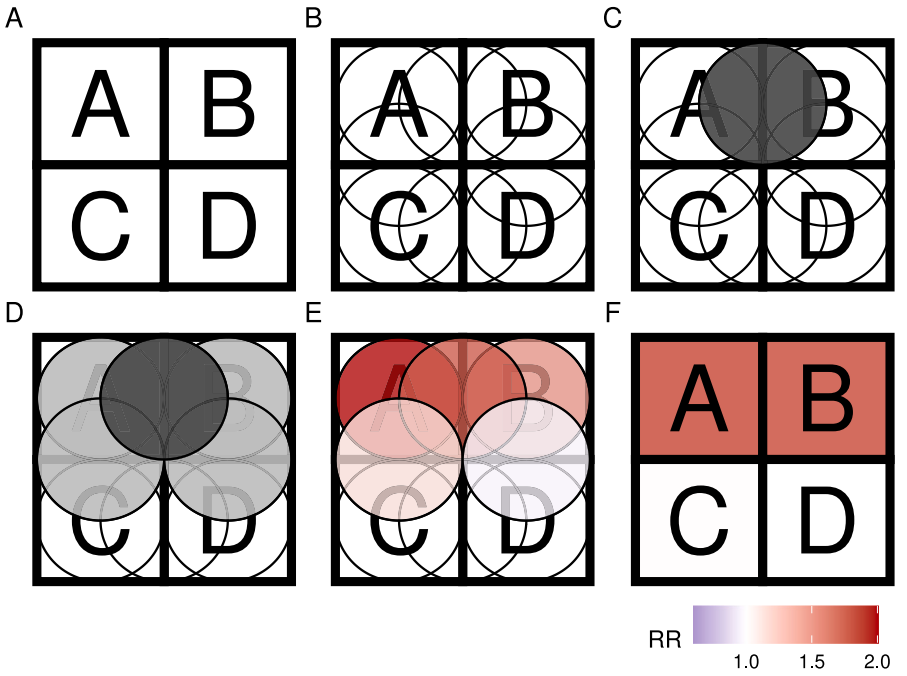

Fig. 1. (A) Study region comprised of four representative cells in illustrative example; (B) 8 single cluster models across the study region; (C) the single cluster model with the highest initial weight overlaps cells A and B and is indicated in dark gray; (D) the ensemble under stacking with localization consists of all single cluster models which overlap the dark gray cluster identified in (C); (E) each single cluster model has its own distinct relative risk; (F) the final stacked relative risk estimates for each cell with localization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Stacked relative risk estimates using a single stack, stacking with localization, and multiple stacks with localization for the illustrative example.

Cell	Single stack (SS):	Single stack with localization (SSL):	Multiple stacks with localization (MSL):	
	All single cluster models	Cluster ensemble 1	Cluster ensemble 2	Ensemble 1 & Ensemble 2
A	1.69	1.68	1.00	1.68
B	1.66	1.66	1.00	1.66
C	1.01	1.01	0.70	0.70
D	1.00	1.00	0.70	0.70

and will be larger for cells in the center of the cluster and smaller for cells along the cluster border.

$$w_j = \mathcal{L}^{(j)} / \sum_{j=1}^{K} \mathcal{L}^{(j)} \tag{7}$$

3.1. Single stack

Using the initial likelihood-based weights w_j (Table 2), we calculate a stacked relative risk estimate for a single ensemble without localization. We propose a single stacked (SS) estimate, $\hat{\rho}_{it}^{SS}$, where $\hat{\rho}_{it}^{(j)}$ is averaged over all j models, is given by:

$$\hat{\rho}_{it}^{SS} = \sum_{j=1}^{K} w_j \hat{\rho}_{it}^{(j)}, \tag{8}$$

Model selection bias is mitigated by the weights w_j shrinking $\hat{\rho}_{it}^{SS}$ towards the null, particularly for clusters with weak evidence (Lukacs et al., 2010). The single stack estimates consider all single cluster models to be in the ensemble and are shown in Table 3 for the illustrative example.

3.2. Single stack with localization

Without localization, all single cluster models are considered to be part of the cluster ensemble. In the case of two clusters, no localization would result in two half-masses in either cluster, thereby masking the distinction between the clusters. This may lead

to a noisier result without clear boundaries for a cluster. Localization, however, not only allows us to detect a cluster ensemble with distinct borders, but also allows for the detection of multiple cluster ensembles in the study region.

Let $\hat{\rho}_{it}^{SSL}$ be the single stack with localization (SSL) estimate and $j^* = \arg\max_j w_j$. A focal point is needed for localization. First, we consider an ensemble to consist of all single cluster models which overlap the maximum weight, or $S_1 = \{j | \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}^{(j)} x_{it}^{(j^*)}\}$ (*stacking by cluster*). Second, an ensemble can also consist of all single cluster models which overlap the cluster center with the largest weight at each cell, or $S_2 = \{j | \mathbb{1}\{d(c_{j^*}, c_j) < r_j, l_j \leq u_{j^*}, l_{j^*} \leq u_j\}\}$ (*stacking by cluster center*). For localization, we re-weight the likelihood weights such that the weights sum to 1 inside the identified cluster ensemble:

$$w_j^{(1)} = w_j \mathbb{1}\{j \in S\} / \sum_{j=1}^{K} w_j \mathbb{1}\{j \in S\}. \tag{9}$$

By imposing this constraint, we are able to identify a cluster ensemble. The single stacked relative risk estimate is $\hat{\rho}_{it}^{SSL} = \sum_{j=1}^{K} w_j^{(1)} \hat{\rho}_{it}^{(j)}$. The process is depicted in Fig. 1. The process is depicted in Fig. 1. We first assume a study region consisting of four cells. Over this study region, we create our set of single cluster models, where each individual cell can have membership in multiple models. To identify the first cluster ensemble, we select the single cluster model with the largest model weight. The first cluster ensemble then consists of this single cluster model with the largest weight and all of the other single cluster models which overlap it. The individual relative risks from each of the overlapping single cluster models are then stacked (with localization) where the weights sum to 1, resulting in stacked relative risk estimates for each cell.

In the illustrative example (Table 2), the model with the largest initial weight is model 3, comprised of cells A and B with a weight of 9.11×10^{-1}. The models which overlap model 3 include model 1 (cell A), model 2 (cell B), model 7 (cells A and C), and model 8 (cells B and D), these five models constitute the first cluster ensemble. We re-weight the weights in these five models such that they sum to 1, giving us $w_j^{(1)}$. Table 3 shows the stacked relative risk estimates for the first cluster ensemble only considering those models which overlap the model with the largest evidence (model 3 in the example).

3.3. Multiple stacks: Detection of multiple cluster ensembles

Let $m = \{1, \ldots, M\}$ be the number of cluster ensembles identified in the study region and $\hat{\rho}_{it}^{MSL}$ be the estimate from multiple stacks with localization (MSL). The procedure for identifying multiple cluster ensembles using stacking with localization by cluster is as follows:

1. Each single cluster model is defined by a cluster center, radius, and time span. For each single cluster model j, calculate the initial likelihood-based weight, $w_j = \mathcal{L}^{(j)} / \sum_{j=1}^{K} \mathcal{L}^{(j)}$. Normalize the weights such that $\sum_{j=1}^{K} w_j = 1$.
2. Order the single cluster models weights, $w_1, \ldots, w_K$, from largest to smallest. Identify the largest model weight, j^*.
3. Use localization by cluster or localization by cluster center to define the set of single cluster models in the mth ensemble. Let $w_j^{(m)}$ be the set of weights which satisfies the criteria for localization by cluster (S_1) or localization by cluster center (S_2), where $w_j^{(m)} = \max\left\{ w_j^{(m-1)}, w_j \mathbb{1}\{j \in S\} / \sum_{j=1}^{K} w_j \mathbb{1}\{j \in S\} \right\}$.
4. Calculate the stacked relative risk estimate as $\hat{\rho}_{it}^{MSL(m)} = \sum_{j=1}^{K} w_j^{(m)} \hat{\rho}_{it}^{(j)}$
5. Remove single cluster models previously identified in a cluster ensemble by re-defining the model weights as $w_j = w_j \mathbb{1}\{j \notin S\}$.

Using either stacking with localization by cluster or localization by cluster center, we identify the single cluster models which overlap the cluster or center with the largest initial weight, w_j. We estimate the relative risk from multiple stacks with localization for the mth ensemble by taking a weighted average of the weights (subject to a sum to 1 constraint) and the estimated cluster relative risks. Finally, after identifying the mth cluster ensemble, we remove the single cluster models identified in that ensemble by re-defining the weights w_j. Though this leads to earlier stacks containing more single cluster models than subsequent stacks in the iterative procedure, this difference in the numbers of candidate models within the stacks does not matter as the extra models will all receive negligible weight asymptotically. The process is repeated until the set of updated weights w_j is null. While single cluster models are allowed to overlap, the identified cluster ensembles will not. While single cluster models are allowed to overlap, the identified cluster ensembles will not. If the true clusters do not overlap, the weights will coalesce to 1 on the single cluster model within each stack, and our procedure will identify the correct model asymptotically. If multiple true clusters overlap, then one stack could potentially contain multiple true clusters, such that the weights within a stack would likely be split between the two (or more) true clusters and our approach will no longer be guaranteed to identify all of the correct clusters.

In the illustrative example, the stacked relative risk estimates are shown using a single stack (SS), single stack with localization (SSL), and with multiple stacks with localization (MSL). In Table 3, we note that the cells identified in cluster ensemble 1 do not have membership in cluster ensemble 2.

3.4. Selection of the number of cluster ensembles

Based on the procedure above, $\hat{\rho}_{it}^{MSL(m)}$ are the cell-specific stacked relative risk estimates for each mth stack. We select the optimal ensemble using Bayesian information criterion (BIC) (Schwarz, 1978) and advocate for the use of BIC over AIC, as AIC has been shown to have an unacceptably high false positive rate with respect to model selection for spatial cluster detection (Kamenetsky et al., 2022):

$$\text{BIC}(m) = -2\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}\log\hat{\rho}_{it}^{MSL(m)} - \hat{\rho}_{it}^{MSL(m)}E_{it} + m\log(y_{..}),$$

where $\hat{\rho}_{it}^{MSL(m)}$ is the stacked relative risk estimate associated with ensemble m using localization by cluster or localization by cluster center and $y_{..} = \sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}$.

For overdispersed counts, such that quasi-BIC (QBIC) (Lebreton et al., 1992) is defined as:

$$\text{BIC}(m) = -2\left(\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}\log\hat{\rho}_{it}^{MSL(m)} - \hat{\rho}_{it}^{MSL(m)}E_{it}\right)/\tilde{\phi} + m\log(y_{..}),$$

4. Simulation study

4.1. Data

We conducted a simulation study to evaluate the performance of stacking. We simulated data on breast cancer incidence in three neighboring prefectures of 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). Data on the female population size at risk and the number of deaths due to breast cancer were stratified by age for each of the 208 municipalities or cells. There were 20 age groups ranging from 0–4 through 95 to 99 years old. Females under 40 or over 74 years old were excluded to avoid misdiagnosis or potential comorbidities. Breast cancer counts and age-standardized expected counts were available across 5 time periods: 1975–1978, 1979–1982, 1983–1986, 1987–1990, and 1991–1994. Observed counts ranged from 0 to 82 in any given cell across the 5 time periods. Dirichelet tesselations were used to obtain approximate borders. We set the maximum spatial cluster radius, r_{max}, to 20 km and considered all possible time intervals of lengths 1 to 5 periods. This resulted in 66,870 single cluster spatio-temporal models.

4.2. Comparison methods

Instead of limiting our scope to selecting a single model, our stacking approach to spatial and spatio-temporal cluster detection considers all single-cluster models. This approach is flexible enough to detect multiple clusters without running into multiple testing issues, as well as to capture fine spatial granularity of the study region by estimating cell-specific stacked relative risks.

A regularized approach to spatial and spatio-temporal cluster using the least absolute shrinkage and selection operator (Lasso) has been shown to be an efficient method in detecting multiple overlapping clusters (Kamenetsky et al., 2022). Starting with the null model with no clusters, the Lasso regularization performs variable selection by shrinking θ_j's to zero as they reach the L_1 penalty. These clusters are then dropped from the active set, resulting in a coefficient path for each potential cluster θ_j. As the tuning parameter λ goes to 0, more clusters are allowed into the active set until we have the least squares estimates at $\lambda = 0$. To select the number of clusters, (Q)BIC was used.

Similar to the Lasso approach, the forward stagewise method (Xu and Gangnon, 2016) obtains the full coefficient path using incremental steps along the gradient. This approach starts with the null model with no clusters. At the current model, it identifies the cluster with the largest absolute value of the gradient and updates the coefficient with a small step size, ε. The process is repeated and the final selection of clusters along the path is determined also using BIC. The Lasso will usually shrink cluster coefficients estimates to zero and drop them from the active set whereas forward stagewise will flatten the coefficients along the path. By dropping the coefficients from the active set, the Lasso is more efficient than the forward stagewise approach (Kamenetsky et al., 2022). Kamenetsky et al. (2022) had previously found that the forward stepwise spatial scan, a variant of the spatial scan statistic which considers previously detected clusters by updating the expected number of cases, had an unacceptably high false positive rate and we therefore do not consider it as a comparison method in this study.

The spatial scan is a global cluster detection test statistic which identifies the single most likely cluster using the maximum likelihood ratio test statistic over all potential clusters. The statistical significance of this cluster is then evaluated using Monte Carlo simulations performed under the null hypothesis of constant disease risk. To identify multiple most likely clusters, methods have been proposed for sequential deletion (Zhang et al., 2010) and re-evaluation or hypothesis-based procedures that specify the number of clusters in the alternative hypothesis (Li et al., 2011). However the spatial scan is limited in its ability to detect multiple overlapping clusters. By applying a forward stepwise testing approach, we modify the standard spatial scan. We consider previously detected clusters in a Poisson model by updating the expected number of cases when an additional cluster is identified and call this the forward stepwise spatial scan.

Table 4
Detection results reported as: false positive rate % (power %). Simulation results for spatio-temporal cluster detection by stacking without overdispersion comparing stacking with localization by cluster versus localization by cluster center. 100 random datasets were generated for each simulation.

Radius/relative risk	Large population center		Small population center	
	By cluster	By cluster center	By cluster	By cluster center
9 km				
1.1	0% (1%)	0% (0%)	0% (0%)	0% (0%)
1.5	0% (74%)	0% (73%)	0% (2%)	0% (2%)
2.0	0% (100%)	0% (100%)	0% (14%)	0% (14%)
11 km				
1.1	0% (0%)	0% (0%)	0% (0%)	0% (0%)
1.5	0% (100%)	0% (100%)	0% (0%)	0% (0%)
2.0	0% (100%)	0% (100%)	0% (14%)	1% (15%)
18 km				
1.1	0% (1%)	0% (1%)	0% (0%)	0% (0%)
1.5	0% (100%)	0% (100%)	0% (23%)	0% (22%)
2.0	0% (100%)	0% (100%)	0% (99%)	0% (99%)

4.3. Simulation setup

In the simulation study, we characterized the performance of stacking for spatio-temporal cluster detection and cell-specific uncertainty quantification. Performance for spatio-temporal cluster detection was evaluated using the false positive rate and power across simulations. The false positive rate is defined as the proportion of simulations where at least one detected cluster did not overlap the true cluster. Power is defined as the proportion of simulations where at least one cell in the true cluster belonged to a detected cluster. One advantage of the single cell overlap is that a lack of any overlap with the true cluster(s) is definitely a false detection. Previous work used a simulation study to show that detection of a single cell inside the cluster was nearly equivalent to detection of the cluster center cell (Kamenetsky et al., 2022).

We compared stacking by model to two other methods: a regularized approach using the least absolute shrinkage and selection operator (Lasso) and forward stagewise methods. We explored the effects of cluster characteristics on detection, including three cluster radii (9, 11, and 18 km) and three cluster relative risks (1.1, 1.5, and 2) under 100 realizations of both Poisson and overdispersed Poisson counts (with overdispersion parameter estimated from the data). We also explored cluster detection with a cluster centered in a large versus small population center.

4.4. Stacking with localization by cluster yields similar results to localization by cluster center

We compared stacking by model to stacking with localization by cluster to localization by cluster center under two data-generating processes: (1) Poisson distribution without overdispersion and (2) quasi-Poisson with overdispersion. For the first scenario, we simulate 100 replications of Poisson distributed counts without overdispersion. For the second scenario, we set the shape parameter based on the estimate from the data.

Stacking with localization by cluster and localization by cluster center resulted in similar false positive rates and power. With no overdispersion, under the null model with no cluster (baseline) the false positive rates were 0% for both approaches. While a limitation of cluster detection studies is identifying clusters in areas with few case counts, stacking by both model and location is able to identify clusters in a small population setting once the cluster radius is large enough. In our simulation study, once the cluster radius was 18 km, stacking by model and by location was able to identify a cluster with 2 relative risk with 0% false positive rate in both large and small population centers (Table 4).

Stacking with localization by cluster and localization by cluster center resulted in near identical false positive rates across simulation settings. There was one setting under which stacking by model had a lower false positive rate than by location. In a small population center with a cluster radius of 11 km and relative risk of 2, the false positive rate by model was 0% and was 1% by location. Localization by cluster resulted in a slight increase in power over localization by cluster center in only select few simulation settings. With a relative risk of 1.5 and cluster radius of 9 km, localization by cluster had higher power (74%) than by cluster center (73%) in a large population center. In the small population center, power was slightly lower with localization by cluster (23%) than by cluster center (24%). Stacking with localization by cluster considers all single cluster models that overlap the single cluster model with the maximum weight, whereas stacking with localization by cluster center considers all single cluster models that overlap the cluster center with the maximum weight. Depending on the difference between the size of the cluster, this may lead to slight differences in the false positive rate. Since stacking with localization by cluster is more generalizable and we therefore proceed with it.

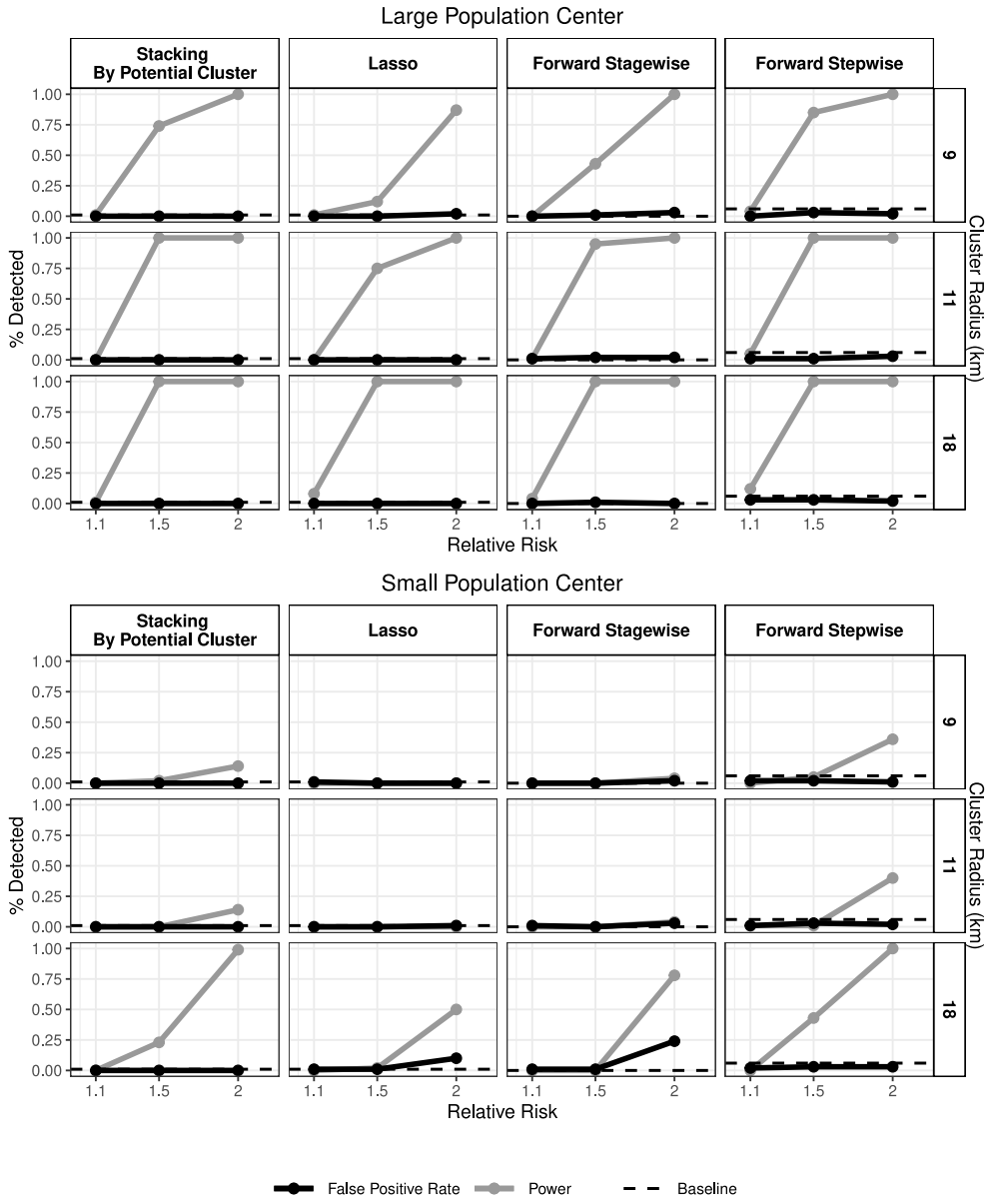

Fig. 2. Simulation results comparing the false positive rate and power in a large and small population center for stacking (by model), Lasso, forward stagewise, and forward stepwise. 100 random datasets were simulated without overdispersion and BIC was used to identify cluster ensembles by stacking with localization by cluster, the Lasso, forward stagewise, and forward stepwise. The false positive rate was calculated based on the percent of simulations which had no overlap with the true cluster. Percent detected was calculated as the number of simulations in which the detected cluster had overlap with the true cluster (power).

4.5. Stacking detects spatio-temporal clusters

Stacking maintained the false positive rate (Fig. 2) while exhibiting better power than comparison methods. Stacking had highest power while maintaining a low false positive rate across all methods. Without overdispersion in the large population center with higher observed and expected counts (more information), all three methods had high power (> 75%) once the cluster radius was at least 11 km and the cluster relative risk was 1.5 or greater. In the small population center with fewer observed and expected counts (less information), there was more differentiation across the methods. While all methods struggled, with enough information (large cluster radius of 18 km and large cluster relative risk of 2), stacking had enough power to detect the cluster (99%) while maintaining the false positive rate at 0%, while Lasso (power: 50%, false positive rate: 10%), forward stagewise (power: 78%, false positive rate: 24%), and forward stepwise spatial scan (power: 100%, false positive rate: 3%) struggled.

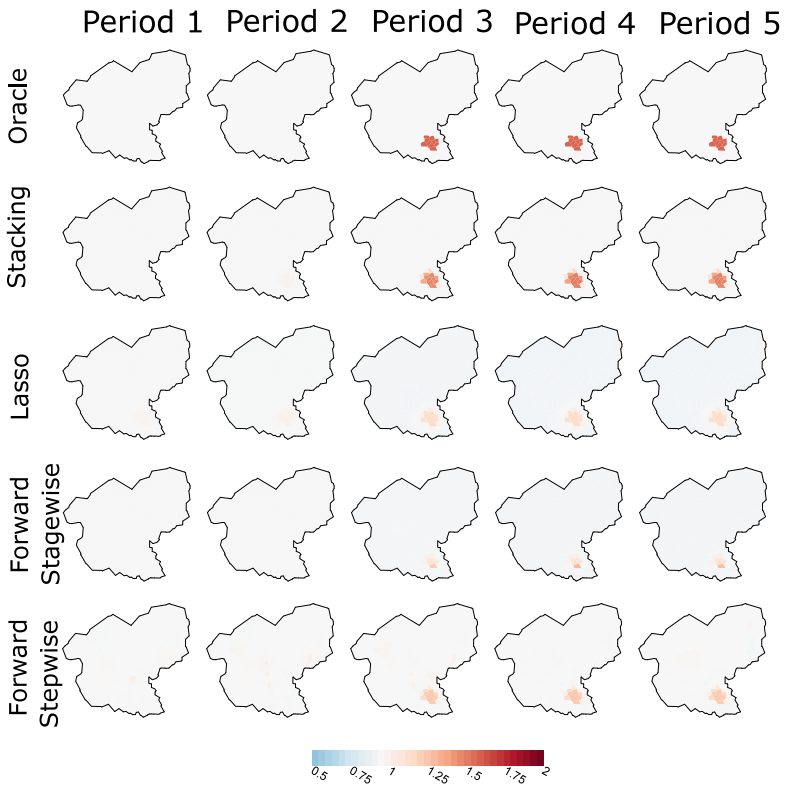

Fig. 3. Relative risk estimates using 100 random datasets simulated with overdispersion with a cluster radius of 11 km, 1.5 relative risk, in a large population center using stacking, Lasso, forward stagewise, and forward stepwise spatial scan.

Overdispersion made detection of the cluster more challenging, though stacking still outperformed the other comparison methods. With overdispersion in the large population center, stacking maintained the false positive rate near baseline. With a small radius of 9 km, stacking had better power and false positive rates (100% power; 0% false positive rate) at a relative risk of 2, compared to Lasso (power: 86%, false positive rate: 2%), forward stagewise (power: 99%, false positive rate: 12%), and forward stepwise (power: 100%, false positive rate: 26%). In the small population center with large 18 km radius, all methods struggled though stacking was able to detect the cluster with much higher power (100%) and lower false positive rate (1%) compared to Lasso (power: 38%, false positive rate: 4%), forward stagewise (power: 41%, false positive rate 31%), and forward stepwise (power: 100%, false positive rate: 46%).

With overdispersion present, both stacking and the Lasso approaches held a low false positive rate, while the forward stagewise. With overdispersion, cluster detection in the small population center was challenging for all methods. Fig. 3 shows the average stacked relative risk estimates across 100 simulations for each of the four methods for an 11 km cluster with 1.5 relative risk in a large population center under overdispersed Poisson counts. The maximum stacked relative risk estimate in the center of the cluster was 1.44 and the cluster relative risk estimate by forward stepwise regression was 1.22 (Lasso: 1.14; Forward Stagewise: 1.25). With clusters that were in a small population center and along the border of the study region (Table S.1, Figure S2), all methods had higher false positive rates. For a cluster with a relative risk of 2, radius of 15 km, and in time periods 2 through 4, across 100 simulations stacking had the highest maximum relative risk estimate (1.93) closest to truth. Stacking was followed by the stepwise spatial scan (1.81), and then the Lasso (1.05) and forward stagewise (1.08) approaches. While all methods identified the cluster, stacking was closest to recovering the cluster relative risk.

5. Data example

We show results for breast cancer incidence rates using the same study region in Japan as in the simulation study (see Section 4.1). In Japan, breast cancer incidence began increasing in the 1960's and plateaued around 2008 (Katanoda et al., 2015). These changing patterns of breast cancer are thought to be driven by changes in lifestyle and reproductive patterns such as later age at first birth, lower parity, and earlier menarche (Minami et al., 2004; Tamakoshi et al., 2005; Kawai et al., 2010). Our research goal is to identify spatio-temporal clusters of breast cancer in three prefectures in Japan: Tochigi, Gunma, and Saitama. Setting the maximum cluster radius to 20 km results in $K = 66{,}870$ single cluster models across the three prefectures.

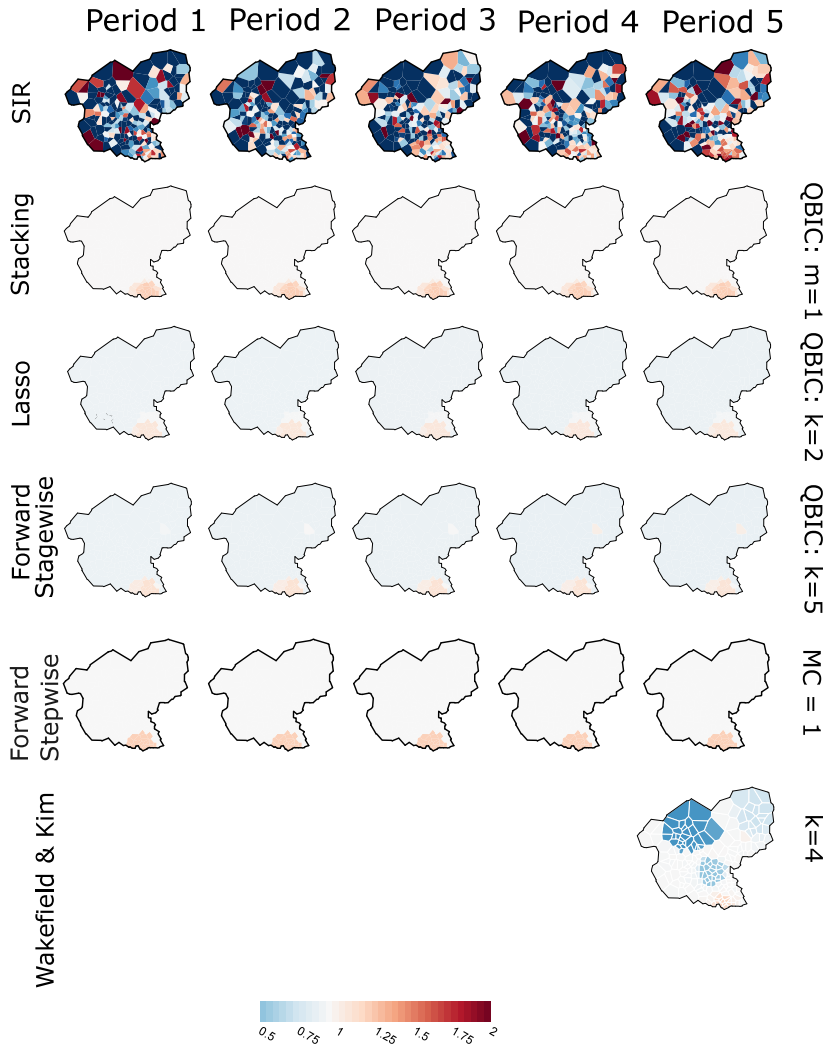

Fig. 4. Cluster identification in Japan across methods. Stacking by model using QBIC identified a single cluster. All four methods identified a cluster of elevated relative risk in the southern part of the Saitama prefecture.

Using stacking, we estimated the cluster relative risk and mapped the results based on QBIC to adjust for overdispersed counts. A single cluster was identified using stacking by model, comprised of 27 cells (Fig. 4). Cells in the center (1.18) and near-center cells (1.18) of the cluster had higher relative risk estimates, representing the cluster core. Cells further from the center had lower relative risks ranging from 1.06 (95%) to 1.01 (95%), representing the periphery around the cluster core. Cells closest to the cluster center have the highest stacked relative risk estimates of 1.18, and then estimates fall towards 1 as cells move further from the cluster center and outside of the cluster. Cell-specific relative risk estimates decreased as the cells moved from the cluster center to the cluster periphery and finally to the border-outside cell. The behavior along the border can represent uncertainty about cluster membership given a fixed cluster and risk, a gradient of risk lower on the margin, or both. Model weights reflect this membership uncertainty where cells along the border will have smaller weights in comparison to cells in the cluster center (Figure S2).

We next compare our stacking by model approach results to the Lasso-based approach, forward stagewise, forward stepwise spatial scan, and Bayesian cluster detection approach of Wakefield and Kim (2013) (Fig. 4). Using the Lasso-based approach, two clusters were identified. The estimated relative risks inside the cluster ranged from 1.03 near the cluster border to 1.12 in the cluster core. With forward stagewise using BIC, 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 the forward stepwise spatial scan, one cluster was identified with a relative risk of 1.18 (p-value: 0) using 1,000 Monte Carlo simulations. No additional clusters were selected after adjustment for multiple testing at the 5% level.

The Wakefield and Kim approach is limited to a space-only analysis and so we collapse counts across time. In this approach, a wide prior is used on the cluster relative risks and a narrow prior is used for the background relative risks. For both the narrow and wide priors, we require the mode to be 1 and for the upper 95% of points for the wide prior to be 2. For the wide prior, we require the upper 50% of points to be 1. This specification leads to choices $(\alpha_n, \beta_n) = (1000, 999)$ and $(\alpha_w, \beta_w) = (6.76, 5.76)$, and we

set the maximum number of potential clusters at 15. We calibrated the prior of no clusters as $\pi_0 = 0.95$ and ran the importance sampling algorithm using 10^5 points and the Markov chain algorithm for estimating the prior and posterior probabilities for 10^5 and 10^6 iterations, respectively. We set the maximum proportion of the study region's population that each potential cluster can contain to 0.15.

All methods identified a signal in the southern tip of the study region. The Lasso approach and forward stagewise, however, identified several additional overlapping clusters in the region. Since both of these approaches identify single clusters that are overlapping, they are similar to a partial stack, though possibly less efficient than the proposed stacking approach. In contrast, based on our results using stacking, we consider these individual clusters to be part of a larger single cluster in the region. The Wakefield and Kim spatial approach identified 4 clusters (posterior probability: 0.562), with one elevated cluster in the southern tip of the study region with posterior relative risk estimates ranging from 1.02 to 1.17 and three clusters of reduced relative risks with posterior relative risk estimates ranging from 0.66 to 0.97. This approach, however, is limited to space only and can be sensitive to the initial parameter estimates. In contrast, our stacking approach shrinks these clusters of reduced relative risk to the background and generates a smoother map with the one cluster of elevated risk. With stacking, we are not only able to better estimate a relative risks for cells in the cluster, but also can calculate confidence bounds for specific cells.

6. Conclusions and discussion

Identifying spatial and spatio-temporal clusters is important to public health surveillance and when considering possible interventions. With the goal of identifying contiguous areas different from the rest, tools in spatial cluster detection have considered sets of potential clusters and then identified the most likely cluster from the set. While procedures have been developed to identify multiple clusters, they have still been limited to estimating individual clusters.

With our stacking approach, we are no longer limited to identifying a single best model, but can use all of the information available. We are still able to estimate cell and cluster-specific relative risks as the weights shrink non-informative models to zero and allocate most of the weight on the set of models which overlap the cluster with the most evidence. By stacking single cluster models, we are able to better estimate the fine geographic risk pattern.

Our stacking approach to spatio-temporal cluster detection is limited by the case counts. With few observed and expected case counts, there is not enough information to identify a cluster from random noise. Our simulation results show that our stacking approach does not consider a cluster unless there is strong evidence, which can be seen in the cases with relatively low relative risks and small radii. We observed that clusters with small relative risks, small radii, and in small population centers were not detected by BIC. However when the cluster radius was relatively large, stacking was able to not only recover the cluster across various relative risks in the small population center, but also maintained the false positive rate at baseline, unlike the other comparison methods. If multiple adjacent clusters are identified, background knowledge should be used to determine the appropriate maximum radius for the set of single cluster models. Since the guaranteed validity of our approach requires that the true clusters not overlap, it may be advisable to use a relatively small maximum radius for the set of single cluster models to minimized the potential for such problematic overlap.

We are able to use all information available in single cluster models and are better able to map more granular geographic patterns and thereby identify areas of elevated relative risk that may be in need of further epidemiologic investigation. We have implemented this stacking approach in the `clustack` R package, available from https://mkamenet3.github.io/clustack/. The original data are not publicly available due to privacy considerations.

This work can be extended in several ways. Stacking and the larger literature on model-averaging has a natural Bayesian interpretation, where we can consider an informative or non-informative prior on the model weights. In this way, any additional information about the cluster location or model could be incorporated into the weighting. Also, uncertainty in case counts can be incorporated into the weights. Additionally, we could consider different formulations for potential clusters and subsequent single cluster models. While we create the models with radii from 0 to r_{max}, we could instead create them based on case counts inside each model (Turnbull et al., 1990). Identifying new methods for identifying spatial clusters can further the tools public health professionals have at their disposal for public health surveillance.

The reader is referred to the online Supplementary Materials for additional results.

CRediT authorship contribution statement

Maria E. Kamenetsky: Literature review, Developed the analysis plan, Conducted the statistical analysis, Wrote the original manuscript draft. **Jun Zhu:** Analysis plan and revised the manuscript up to its final form. **Ronald E. Gangnon:** Analysis plan and revised the manuscript draft up to its final form.

Financial disclosure

None reported.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was performed using the compute resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) (Center for High Throughput Computing, 2006) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, United States, the Wisconsin Institutes for Discovery, and the National Science Foundation, United States, and is an active member of the Open Science Grid, which is supported by the National Science Foundation, United States and the U.S. Department of Energy's Office of Science.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spasta.2025.100933.

References

Abolhassani, A., Prates, M.O., 2021. An up-to-date review of scan statistics. Stat. Surv. 15, 111–153. http://dx.doi.org/10.1214/21-SS132.

Anderson, C., Lee, D., Dean, N., 2014. Identifying clusters in Bayesian disease mapping. Biostatistics 15 (3), 457–469. http://dx.doi.org/10.1093/biostatistics/kxu005, URL https://academic.oup.com/biostatistics/article-lookup/doi/10.1093/biostatistics/kxu005.

Anderson, C., Lee, D., Dean, N., 2016. Bayesian cluster detection via adjacency modelling. Spat. Spatio-Temporal Epidemiol. 16, 11–20. http://dx.doi.org/10.1016/j.sste.2015.11.005, URL https://linkinghub.elsevier.com/retrieve/pii/S1877584515000507.

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

Breiman, L., 1996. Stacked regressions. Mach. Learn. 24, 49–64.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference, second ed. Springer, New York.

Burnham, K.P., Anderson, D.R., 2010. Model Selection and Multimodel Inference, second ed. Springer-Verlag, New York.

Center for High Throughput Computing, 2006. Center for high throughput computing. http://dx.doi.org/10.21231/GNT1-HW21, URL https://chtc.cs.wisc.edu/.

Charras-Garrido, M., Azizi, L., Forbes, F., Doyle, S., Peyrard, N., Abrial, D., 2013. On the difficulty to delimit disease risk hot spots. Int. J. Appl. Earth Obs. Geoinf. 22, 99–105. http://dx.doi.org/10.1016/j.jag.2012.04.005, URL https://linkinghub.elsevier.com/retrieve/pii/S0303243412000682.

Chatfield, C., 1995. Model uncertainty , data mining and statistical inference. J. R. Stat. Soc. . Ser. A (Stat. Soc.) 158 (3), 419–466.

Choi, J., Lawson, A.B., Cai, B., Hossain, M.M., 2011. Evaluation of Bayesian spatiotemporal latent models in small area health data: Evolution of Bayesian spatiotemporal latent models. Environmetrics 22 (8), 1008–1022. http://dx.doi.org/10.1002/env.1127, URL https://onlinelibrary.wiley.com/doi/10.1002/env.1127.

Duczmal, L., Assuncáo, R., 2004. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. Comput. Statist. Data Anal. 45 (2), 269–286. http://dx.doi.org/10.1016/S0167-9473(02)00302-X, ISBN: 0167-9473.

Dudoit, S., Laan, M.J.V.D., 2005. Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. Stat. Methodol. 2, 131–154. http://dx.doi.org/10.1016/j.stamet.2005.02.003.

Gangnon, R.E., Clayton, M.K., 2003. A hierarchical model for spatially clustered disease rates. Stat. Med. 3228 (20), 3213–3228. http://dx.doi.org/10.1002/sim.1570, ISBN: 0277-6715 (Print). 0277-6715 (Linking).

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging : A tutorial. Statist. Sci. 14 (4), 382–401.

Hossain, M.M., Lawson, A.B., 2005. Local likelihood disease clustering: development and evaluation. Environ. Ecol. Stat. 12, 259–273.

Huang, L., Kulldorff, M., Gregorio, D., 2007. A spatial scan statistic for survival data. Biom. J. Int. Biom. Soc. 63 (1), 109–118. http://dx.doi.org/10.1111/j.1541-0420.2006.00661.x.

Jung, I., Kulldorff, M., Klassen, A.C., 2007. A spatial scan statistic for ordinal data. Stat. Med. (June 2006), 1594–1607. http://dx.doi.org/10.1002/sim.

Jung, I., Kulldorff, M., Richard, O.J., 2010. A spatial scan statistic for multinomial data. Stat. Med. 29, 1910–1918. http://dx.doi.org/10.1002/sim.3951.

Kamenetsky, M.E., Lee, J., Zhu, J., Gangnon, R.E., 2022. Regularized spatial and spatio-temporal cluster detection. Spat. Spatio-Temporal Epidemiol. 41, http://dx.doi.org/10.1016/j.sste.2021.100462.

Kass, R.E., Raftery, A.E., 1995. Bayes factors bayes factors. J. Amer. Statist. Assoc. 90 (430), 773–795.

Katanoda, K., Hori, M., Matsuda, T., Shibata, A., Nishino, Y., Hattori, M., Soda, M., Ioka, A., Sobue, T., Nishimoto, H., 2015. An updated report on the trends in cancer incidence and mortality in Japan, 1958–2013. Jpn. J. Clin. Oncol. 45 (4), 390–401. http://dx.doi.org/10.1093/jjco/hyv002, URL https://academic.oup.com/jjco/article-lookup/doi/10.1093/jjco/hyv002.

Kawai, M., Minami, Y., Kuriyama, S., Kakizaki, M., Kakugawa, Y., Nishino, Y., Ishida, T., Fukao, A., Tsuji, I., Ohuchi, N., 2010. Reproductive factors, exogenous female hormone use and breast cancer risk in Japanese: the Miyagi Cohort Study. Cancer Causes Control. 21 (1), 135–145. http://dx.doi.org/10.1007/s10552-009-9443-7, URL http://link.springer.com/10.1007/s10552-009-9443-7.

Knorr-Held, L., Raßer, G., 2000. Bayesian detection of clusters and discontinuities in disease maps. Biometrics 56 (1), 13–21. http://dx.doi.org/10.1111/j.0006-341X.2000.00013.x, URL https://onlinelibrary.wiley.com/doi/10.1111/j.0006-341X.2000.00013.x.

Kulldorff, M., Huang, L., Konty, K., 2009. A scan statistic for continuous data based on the normal probability model. Int. J. Heal. Geogr. 9, 1–9. http://dx.doi.org/10.1186/1476-072X-8-58.

Kulldorff, M., Huang, L., Pickle, L., Duczmal, L., 2006. An elliptic spatial scan statistic. Stat. Med. 25 (22), 3929–3943. http://dx.doi.org/10.1002/sim.2490, ISBN: 0277-6715 (Print)\r0277-6715 (Linking).

Kulldorff, M., Mostashari, F., Duczmal, L., Yih, W.K., Kleinman, K., Platt, R., 2007. Multivariate scan statistics for disease surveillance. Stat. Med. (January), 1824–1833. http://dx.doi.org/10.1002/sim.

Kulldorff, M., Nagarwalla, N., 1995. Spatial disease clusters: Detection and inference. Stat. Med. 14 (8), 799–810. http://dx.doi.org/10.1002/sim.4780140809, ISBN: 0277-6715 (Print).

Lawson, A.B., Song, H.R., Cai, B., Hossain, M.M., Huang, K., 2010. Space-time latent component modeling of geo-referenced health data. Stat. Med. 29 (19), 2012–2027. http://dx.doi.org/10.1002/sim.3917, URL https://onlinelibrary.wiley.com/doi/10.1002/sim.3917.

Leamer, E.E., 1978. Specification Searches: Ad Hoc Inference with Nonexperimental Data. John Wiley & Sons, Inc..

Leblanc, M., Tibshirani, R., 1996. Combining estimates in regression and classification. J. Am. Stat. Assoc. 91 (436), 1641–1650.

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

Lee, D., Lawson, A., 2016. Quantifying the spatial inequality and temporal trends in maternal smoking rates in Glasgow. Ann. Appl. Stat. 10 (3), http://dx.doi.org/10.1214/16-AOAS941, URL https://projecteuclid.org/journals/annals-of-applied-statistics/volume-10/issue-3/Quantifying-the-spatial-inequality-and-temporal-trends-in-maternal-smoking/10.1214/16-AOAS941.full.

Li, X.z.Z., Wang, J.f.F., Yang, W.z.Z., Li, Z.j.J., Lai, S.j.J., 2011. A spatial scan statistic for multiple clusters. Math. Biosci. 233 (2), 135–142. http://dx.doi.org/10.1016/j.mbs.2011.07.004, ISBN: 1879-3134 (Electronic)\r0025-5564 (Linking) Publisher: Elsevier Inc..

Lukacs, P.M., Burnham, K.P., Anderson, D.R., 2010. Model selection bias and Freedman ' s paradox. Ann. Inst. Statist. Math. 62, 117–125. http://dx.doi.org/10.1007/s10463-009-0234-4.

McCullagh, P., Nelder, J., 1991. Generalized Linear Models, second ed. Chapman and Hall, London, Pages: 194–200, 323–339.

Minami, Y., Tsubono, Y., Nishino, Y., Ohuchi, N., Shibuya, D., Hisamichi, S., 2004. The increase of female breast cancer incidence in Japan: Emergence of birth cohort effect. Int. J. Cancer 108 (6), 901–906. http://dx.doi.org/10.1002/ijc.11661, URL https://onlinelibrary.wiley.com/doi/10.1002/ijc.11661.

Napier, G., Lee, D., Robertson, C., Lawson, A., 2019. A Bayesian space–time model for clustering areal units based on their disease trends. Biostatistics 20 (4), 681–697. http://dx.doi.org/10.1093/biostatistics/kxy024, URL https://academic.oup.com/biostatistics/article/20/4/681/5039880.

Openshaw, S., Craft, A.A., Charlton, M., Birch, J.J., Craft, A.A., Birch, J.J., 1988. Investigation of leukaemia clusters by use of a geographical analysis machine. Lancet 331 (8580), 272–273. http://dx.doi.org/10.1016/S0140-6736(88)90352-2, URL http://linkinghub.elsevier.com/retrieve/pii/S0140673688903522.

Raftery, A., Madigan, D., Hoeting, J., 1993. Model Selection and Accounting for Model Uncertainty in Linear Regression Models. Technical Report, University of Washington.

Raftery, A., Richardson, S., 1996. Model Selection for Generalized Linear Models via GLIB , with Application to Epidemiology 1 University of Washington. Technical Report April 2014.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464. http://dx.doi.org/10.1214/aos/1176344136, URL http://projecteuclid.org/euclid.aos/1176344136, arXiv:arXiv:1011.1669v3 ISBN: 0780394224 tex.arxivid: arXiv:1011.1669v3.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. J. R. Stat. Soc. Ser. B (Methodological) 36 (2), 111–147.

Takahashi, K., Kulldorff, M., Tango, T., Yih, K., 2008. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring.. Int. J. Heal. Geogr. 7 (1), 14. http://dx.doi.org/10.1186/1476-072X-7-14, URL http://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-7-14, ISBN: 1476-072X (Electronic)\n1476-072X (Linking).

Tamakoshi, K., Yatsuya, H., Wakai, K., Suzuki, S., Nishio, K., Lin, Y., Niwa, Y., Kondo, T., Yamamoto, A., Tokudome, S., Toyoshima, H., Tamakoshi, A., for the JACC Study Group, 2005. Impact of menstrual and reproductive factors on breast cancer risk in Japan: Results of the JACC study. Cancer Sci. 96 (1), 57–62. http://dx.doi.org/10.1111/j.1349-7006.2005.00010.x, URL https://onlinelibrary.wiley.com/doi/10.1111/j.1349-7006.2005.00010.x.

Tango, T., Takahashi, K., 2005. A flexibly shaped spatial scan statistic for detecting clusters. Int. J. Heal. Geogr. 4 (1), 11. http://dx.doi.org/10.1186/1476-072X-4-11, URL http://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-4-11.

Turnbull, B.W.B., Iwano, E.E.J., Burnett, W.S.W., Howe, H.L.H., Clark, L.L.C., 1990. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. Am. J. Epidemiol. 132 (supp1), 136–143, arXiv:/dx.doi.org.ezproxy.library.wisc.edu/10.1093/oxfordjournals.aje.a115775 [http:] tex.arxivid: http://dx.doi.org.ezproxy.library.wisc.edu/10.1093/oxfordjournals.aje.a115775.

Wakefield, J., Kim, A., 2013. A Bayesian model for cluster detection. Biostatistics 14 (4), 752–765. http://dx.doi.org/10.1093/biostatistics/kxt001, URL https://academic.oup.com/biostatistics/article-lookup/doi/10.1093/biostatistics/kxt001.

Wolpert, D.H., 1992. Stacked Generalization. Technical Report 505, Complex Systems Group, Theoretical Division, Center for Non-linear Studies, Los Alamos National Laboratory, pp. 241–259, Publication title: Neural networks Volume: 87545.

Xu, J., Gangnon, R.E., 2016. Stepwise and stagewise approaches for spatial cluster detection. Spat. Spatio-Temporal Epidemiology 17, 59–74. http://dx.doi.org/10.1016/j.sste.2016.04.007, Publisher: Elsevier Ltd.

Yan, P., Clayton, M.K., 2006. A cluster model for space – time disease counts. Stat. Med. 25 (5), 867–881. http://dx.doi.org/10.1002/sim.2424.

Zhang, Z., Assunção, R., Kulldorff, M., 2010. Spatial scan statistics adjusted for multiple clusters. J. Probab. Stat. 2010, http://dx.doi.org/10.1155/2010/642379.