



Published in final edited form as:

Stat Methods Med Res. 2016 August ; 25(4): 1101–1117. doi:10.1177/0962280216660422.

Spatial Bayesian surveillance for small area case event data

Chawarat Rotejanaprasert¹, Andrew Lawson¹, Susan Bolick-Aldrich², and Deborah Hurley²

¹Department of Public Health sciences, Medical University of South Carolina, USA

²South Carolina Central Cancer Registry, South Carolina Department of Health and Environmental Control, USA

Abstract

There has been little development of surveillance procedures for epidemiological data with fine spatial resolution such as case events at residential address locations. This is often due to difficulties of access when confidentiality of medical records is an issue. However, when such data are available, it is important to be able to affect an appropriate analysis strategy. We propose a model for point events in the context of prospective surveillance based on conditional logistic modeling. A weighted conditional autoregressive model is developed for irregular lattices to account for distance effects, and a Dirichlet tessellation is adopted to define the neighborhood structure. Localized clustering diagnostics are compared including the proposed local Kullback–Leibler information criterion. A simulation study is conducted to examine the surveillance and detection methods, and a data example is provided of non-Hodgkin’s lymphoma data in South Carolina.

Keywords

Bayesian; spatial; Kullback–Leibler; surveillance; case event data

1 Introduction

The analysis of disease maps has received considerable attention over the last two decades. The need to consider prospective changes in disease risk is great given increases in public awareness of disease risk related to environmental insults (such as cluster alarms) and increased risks from bioterrorism.^{1,2} This latter focus has led to an increase in availability of testing methodology but little evidence of novel developments in modeling.¹ This is particularly true of Bayesian approaches to map surveillance, while there has been some development focusing on temporal approaches.^{3,4} Recently, developments in this area have focused on small area count disease incidence.^{3,5}

Reprints and permissions: sagepub.co.uk/journalsPermissions.nav

Corresponding author: Chawarat Rotejanaprasert, Department of Public Health sciences, Medical University of South Carolina, Charleston, SC 29425, USA. chawarat.rot@biotec.or.th

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

While count data are an important focus area, it is also often important to consider finer spatial scale events such as events at residential addresses. This may be because of fine scale environmental effects acting over short distances (1–10 km) or to localized high impact insults such as a bioterrorism event (aerosol anthrax; nuclear radiation). Development in this area has been hampered by the difficulties related to patient confidentiality, when case addresses are used. Point process models have been examined previously.^{6,7} In those examples, point process models (log Gaussian Cox processes) are employed and a smoothly varying separable spatial and temporal model is assumed for background and covariance structure in space–time. Exceedence probability surfaces are estimated for examination of ‘unusual’ space–time risk anomalies. This approach has some limitations as it relies heavily on sophisticated modeling with attendant heavy computational load. It also focuses on overall modeling of risk without recourse to individual level modulation by covariates. Examination was not made as to the surveillance capabilities of the method in terms of average run length or false alarm rate.

In our approach, we exploit the fact that we have individual level information about cases but also wish to model overall variation in space–time risk. Unlike earlier work, we adopt a conditional logistic spatial model for the covariate linkage conditioning on the point process superposition of cases and controls. This allows for the rich structure of cancer registry information to be accommodated at the individual level. In addition, it allows the surveillance to be performed on the binary logistic model directly and this allows natural inclusion of individual level covariates.

2 Spatio-temporal modeling for case event data

When a bivariate realization of cases and controls are available, it is possible to make conditional inference on this joint realization (see Lawson,⁸ chapter 5). A usual choice of formulation for intensity is the conditional logistic model which leads to $\text{logit}(p_i) = \eta_i$, where p_i is the probability of i th case. Components in the endemic part include individual covariates and random effects as $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i + u_i + \lambda_i$ where \mathbf{x}_i is the covariate vector with the corresponding coefficient vector $\boldsymbol{\beta}$. To describe variation in the endemic part, the prior distributions are defined as follows. Each covariate coefficient is assumed to have a zero mean Gaussian prior distribution with variance τ_{β}^{-1} . u_i and v_i are employed to capture spatially correlated and unstructured extra variation in the model. It is often important to include both structured and unstructured random effects in a spatial analysis since without prior knowledge unobserved confounding can take various forms (see Lawson,⁸ chapter 5). The uncorrelated random effect is described by a zero mean Gaussian prior distribution with variance τ_u^{-1} . The spatially correlated effect is assumed to have the intrinsic conditional autoregressive model (ICAR) model proposed by Besag et al.⁹ That is, conditionally,

$u_i | \mathbf{u}_{-i} \sim N\left(\bar{u}_{\delta_i}, \tau_u^{-1} / n_{\delta_i}\right)$ where \mathbf{u}_{-i} is the vector containing the correlated effect of all regions except the i th area. δ_i , n_{δ_i} and \bar{u}_{δ_i} are a set of spatial neighbors, cardinality and the mean of the neighborhood of the i th tract, respectively, and τ_u^{-1} is the spatial component variance. The ICAR model requires the first-order neighboring information specified by a common boundary, i.e., adjacency information is needed. There are a number of ways to

define neighborhoods. The Dirichlet tessellation partitions a study region into polygons such that each divided small area contains only one data point and every point in a given area is closer to its generating point than to any other. The neighborhood set of each point can then be derived. The partition can be created by the R function DELDIR.¹⁰

To capture the temporal trend a random walk is assumed. In general, a random walk is assumed to have a prior distribution as a Gaussian distribution with mean the previous time point which can be both positive and negative. Then the temporal trend is specified as

$$\lambda_i \sim N(\lambda_{i-1}, \tau_\lambda^{-1})$$

which allows for a type of non-parametric temporal trend effect. All variances in Gaussian prior distributions are described by a relatively non-informative Uniform distribution as

$\tau_*^{-1} \sim 1/SD^2$; $SD \sim Unif(0, 10)$, where $\tau_*^{-1} = \tau_\beta^{-1}, \tau_v^{-1}, \tau_u^{-1}, \tau_\lambda^{-1}$. Hence, the full specification of the binomial model for case event surveillance is shown as follows

$$\begin{aligned} y_i &\sim Bern(p_i) \\ \text{logit}(p_i) &= \mathbf{x}_i^T \boldsymbol{\beta} + v_i + u_i + \lambda_i \\ \boldsymbol{\beta}_i &\sim N(0, \tau_\beta^{-1}) \\ v_i &\sim N(0, \tau_v^{-1}) \\ u_i | \mathbf{u}_{-i} &\sim N(\bar{u}_{\delta_i}, \frac{\tau_u^{-1}}{n_{\delta_i}}) \\ \lambda_i &\sim N(\lambda_{i-1}, \tau_\lambda^{-1}) \\ \tau_*^{-1} &\sim 1/SD^2; SD \sim Unif(0, 10) \\ \tau_*^{-1} &= \tau_\beta^{-1}, \tau_v^{-1}, \tau_u^{-1}, \tau_\lambda^{-1}. \end{aligned}$$

2.1 Weighted conditional auto regression

The convolution model proposed by Besag et al.⁹ including a correlated effect based on an ICAR model and uncorrelated Gaussian random effects is the most popular choice in Bayesian disease mapping for small area count data. Although the ICAR model has been criticized for over smoothing the risk surface, studies have shown the convolution model perform very well in many situations including disease clusters.¹¹⁻¹³ ICAR can have variants of correlation or covariance structure depending on neighborhood definitions. The traditional ICAR is defined based on adjacent neighboring areas only. Rodrigues and Assunção¹⁴ introduced a highly parameterized approach to allow the neighbor structure to be estimated in the parameter space.

For case event data, the convolution model was proposed by Lawson¹⁵ to be used with Dirichlet tessellation to determine adjacency. White and Ghosh¹⁶ suggested a stochastic neighborhood model for a irregular lattice which was extended from the family of Gaussian-Markov Random Fields by Hrafinkelsson and Cressie.¹⁷ The neighborhood sizes were estimated by assuming that there is an unknown cutoff distance. The weights within the proximity distance are equal and sum to one, and decrease exponentially outside the range. In this paper, we extend the ICAR specification to include distance information as follows.

Denote a set of irregular non-overlapping areas $i = 1, \dots, I$ within a study region. The general specification of a CAR model can be in a form of

$$\mathbf{u} \sim N_I(\mathbf{x}'\boldsymbol{\beta}, [I - C]^{-1}M)$$

provided $I - C$ is non-singular and $M = \text{diag}(\tau_1^{-1}, \dots, \tau_I^{-1})$. Denote $[I - C]^{-1}M = \Sigma_u$. We need to ensure that Σ_u is symmetric; i.e., $c_{ij}\tau_i^{-1} = c_{ji}\tau_j^{-1}$ where c_{ij} is an element in C . Often c_{ij} is parameterized as ρw_{ij} where w_{ij} is an element in W and ρ indicates spatial dependence. Note that ρ is not a correlation parameter. W is known as the proximity or weight matrix with $w_{ij} \geq 0$ for $i \neq j$ and $w_{ij} = 0$ for $i = j$. A common choice is to specify $w_{ij} = 1$ for adjacent neighbors. Denote W is the standardized weight matrix by the row sums (over its neighbors);

i.e., $\tilde{w}_{ij} = \frac{w_{ij}}{\sum_j w_{ij}} = \frac{1}{w_{i+}}$. Let $\tau_i^{-1} = \tau^{-1}/w_{i+}$. Then

$$\sum_u [I - C]^{-1}M = \tau^{-1}[D^{-1} - \rho W]^{-1} = \tau^{-1}[I - \rho W]^{-1}D$$

where $D = \text{diag}(1/w_{1+}, \dots, 1/w_{I+})$.

The intrinsic CAR is equivalent to specifying $\rho = 1$. Although the ICAR is an improper distribution, the specification has a practical meaning. The mean of u_i is an average of its neighbors with weights in the matrix W . Usually, the weights are assigned equally to be one. In contrast, adding the ρ , although it remedies singularity, we do not have the interpretation of the mean as an average, but proportional to the average. Furthermore, the ICAR is a more common choice of the prior distribution for a spatial effect due to its simplicity and ease of computation. We extend the ICAR formulation to account for the distance effect in irregular-shape or case event data. Here, we incorporated distance information into the weight matrix using a distance function, and define a weighted conditional auto regression (WCAR) model as

$$u_i | \mathbf{u}_{-i} \sim N\left(\bar{u}_{\delta_i}, \frac{\tau_u^{-1}}{w_{\delta_i}}\right)$$

where $\bar{u}_{\delta_i} = \sum_{j \in \delta_i} \frac{w_{ij}u_j}{w_{\delta_i}}$, $w_{\delta_i} = \sum_{j \in \delta_i} w_{ij}$. For instance, we can consider exponential decay as the distance function. That is $K(d_{ij}, \alpha) = K_{ij} = w_{ij} = e^{-\alpha_j d_{ij}}$, $\alpha_j > 0$, and $\tilde{w}_{ij} = K_{ij}/K_{i+}$. From the specification of WCAR, ICAR is a special case of WCAR when $w_{ij} = 1$. To demonstrate the WCAR model, two distance functions are assumed: an exponential decay (WCAR^{exp}; $K(d_{ij}, \alpha) = \exp(-\alpha_j d_{ij})$) and inverse distance (WCAR^{inv}; $K(d_{ij}, \alpha) = 1/\alpha_j d_{ij}$). Note that there are possibilities of distance specifications including kernel functions; the two functions are commonly used in practice. For simplicity of model demonstration, α_j is assumed to be one. However, it could also be treated as a random parameter and assumed to have a prior distribution.

3 Local concentration diagnostics

To evaluate clustering behavior, a number of local measures of goodness of fit are examined. These include leave-one-out cross-validation and individual information criteria.

3.1 Conditional predictive ordinate

An influential measure proposed by Geisser and Eddy¹⁸ using the leave-one-out cross-validation predictive density is known as the conditional predictive ordinate (CPO). The CPO is defined as $CPO_i = f(y_i | y_{\setminus i}) = \int f(y_i | \theta) f(\theta | y_{\setminus i}) d\theta$ and is estimated from a sampler by

$\left(\frac{1}{G} \sum_{g=1}^G \frac{1}{f(y_i | \theta^g)} \right)^{-1}$, where y_i is the i th observation of y and $y_{\setminus i}$ is y after omitting y_i and G is the sampler sample size. The CPO is a convenient posterior predictive checking tool because it is easily implemented using MCMC and can be used to identify outliers, influential observations across different non-nested models. A larger value implies a better fit of the model to the observation and a low value suggests that the observation is an outlier or influential observation.

3.2 Local information criteria

Information criteria are measures of the relative goodness-of-fit of a statistical model and developed with a target to select models with good predictive abilities. In general, we have only one data set, no validation data. Unlike in cross validation where the data are portioned and refitted, the data are used in both parameter estimation and prediction. This leads to the double use of the data. Akaike suggested the Akaike information criterion (AIC) in the form of deviance (predictive ability) plus a penalty term (bias correction) using the maximum likelihood estimator (MLE) as the plug-in estimator for prediction and introduced the penalty term as $2k$, where k is the number of parameters the model to asymptotically correct the bias from using the data twice. The AIC also shown to be equivalent to cross validation in a large sample setting.¹⁹ However, in hierarchical models, the prior distribution can reduce the degree of freedom of the number of parameters since the parameters tend to be correlated in the prior levels. Thus, the free number of parameters is not clear.

Deviance information criterion (DIC), an important variant of information criteria in hierarchical modeling, was introduced by Spiegelhalter et al.²⁰ DIC can be seen as a Bayesian version of AIC²¹ by replacing the MLE with the Bayes estimator and replacing k with pD , the effective number of parameters. Although it was a significant advance and has received a credible response from Bayesian and BUGS community, there are criticisms on DIC due to theoretical justification.^{20,22,23} Many information criteria have been developed in the context of hierarchical modeling; the most promising invent perhaps is the Watanabe Akaike (or widely applicable) information criterion (WAIC).²²

WAIC was introduced by Watanabe,^{24,25} who calls it the WAIC. This measure is a more fully Bayesian approach for estimating the out-of-sample expectation²¹ because it is based on an appropriate predictive distribution, not a plug-in. WAIC starts with the computed log pointwise posterior predictive density and then adding a correction for effective number of parameters to adjust for overfitting.^{21,24} The bias adjustment used in this paper is variance of

individual terms in the log density summed over the areas: $p\text{WAIC} = \sum_i \text{var}_{\text{post}}(\log(p(y_i|\boldsymbol{\theta})))$. Then WAIC can be computed as $\text{WAIC} = lppd - p\text{WAIC}$ where

$lppd = \sum_i \log\left(\frac{1}{G} \sum_{g=1}^G p(y_i|\boldsymbol{\theta}^g)\right)$ and G is the sampler sample size. One advantage of WAIC is that it can be applied to both singular and non-singular models while AIC and DIC only work for singular models.^{24,25} Although some development has been made to deal with singular models for those criteria,²³ WAIC appears to be less computationally intensive.

However, DIC can be assessed locally providing a finer level of cluster diagnostics. The partitioning of the DIC can be computed to investigate contribution of each individual observation as²⁰ $\text{Local DIC}_i = \text{LDIC}_i = \bar{D}_i + pD_i$, where \bar{D}_i is the average deviance of the observation i , and pD_i is the effective number of parameters or the information contributing to its own fitted value of the observation i . One can also estimate the pD_i from the local average deviance and the local deviance of the mean parameters. Because WAIC is pointwise computing, it is very convenient to obtain local WAIC (LWAIC) from the definition as $\text{Local WAIC} = \text{LWAIC}_i = lppd_i - p\text{WAIC}_i$.

3.3 Local Kullback–Leibler divergence

We may consider the outcome of a Bayesian analysis to be the distribution of some quantity given the model choice and observed data. We particularly can consider the posterior of model parameters in context of leave-one-out cross-validation. The Kullback–Leibler (KL) divergence is usually used as a measure of how different two distributions are. We propose to define the surveillance KL divergence to evaluate influence of the observation at location i as

$$\text{LKL}_i = \text{KL}(P_i, P_{-i}) = \int \dots \int P_i \log\left(\frac{P_i}{P_{-i}}\right) d\boldsymbol{\theta}$$

where $P_i = p(\boldsymbol{\theta}|\mathbf{y}_i)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_i)$, $\mathbf{y}_i = (y_1, \dots, y_i)$ and $P_{-i} = p(\boldsymbol{\theta}|\mathbf{y}_{-i})$, $\mathbf{y}_{-i} = (y_1, \dots, y_{i-1})$. P_i is the joint posterior distribution of all parameters given full data. P_{-i} is the joint posterior distribution of all parameters given the data without the i th area. Therefore,

$\text{LKL}_i = E_{\boldsymbol{\theta}|\mathbf{y}_i}[\log p(y_i|\boldsymbol{\theta})] + \log E_{\boldsymbol{\theta}|\mathbf{y}_i}\left[\frac{1}{p(y_i|\boldsymbol{\theta})}\right] = E_{\boldsymbol{\theta}|\mathbf{y}_i}[\log p(y_i|\boldsymbol{\theta})] - \log(CPO_i)$. The algebraic detail of the proof is offered in Appendix 1. The second term in the above express is a function of the log of CPO. Similar formulations were proposed in the context of outlier detection.^{26,27}

4 Simulation study

4.1 Simulated data

A simulation experiment is conducted to examine distance specifications of the weight matrix in WCAR and compare to the ICAR and compare different cluster detection methods. A simulation map used as a basis for our evaluation and assessment of outbreak detection was the map of the state of South Carolina (SC), USA. The locations of 300 events are independently simulated from a point process whereby point events occur within a given

map in a completely random fashion. A Bernoulli sampling scheme is employed to generate the nature of each event (case or control) is also simulated: the probability of a case event is 0.8 inside a cluster and the probability of being a control is 0.1 outside a cluster. Three circular clusters are simulated with radius of 1, 5 and 10 km. Figure 1 shows a realization of simulated cases and controls using CSR function in Splancs package in *R*.²⁸

Each simulated dataset was fitted sequentially using the model described in “Spatiotemporal Modeling for Case Event Data” section and localized concentration detection in “Local Concentration Diagnostics” section. The weight functions are chosen to be the exponential decay and the inverse of distance with alpha in both functions fixed to be one. We carried out the posterior sampling with the mixed Gibbs–Metropolis algorithms implemented in WinBUGS. The burn-in period of 10,000 samples with two chains was used to evaluate the convergence of the chains, and, after the burn-in, a sample of 10,000 iterations was obtained as the converged posterior sample. To account for simulation artifacts, we simulated 50 datasets and the results presented are averaged over those datasets. The three weight functions and surveillance methods are evaluated using the simulated data.

4.2 Simulation results

The concepts of false positive and false negative rates are readily usable to measure the performance of the cluster diagnostics. An ROC curve is a way to illustrate the performance of a surveillance system as its discrimination threshold is varied. It is created by plotting the fraction of true sensitivity against the fraction of 1-specificity at various threshold settings. To construct a ROC curve, we can connect all the point obtained from the all possible chosen thresholds. A true positive rate for a measure was calculated by averaging the proportion of times when the posterior sampler of the measure was greater than a threshold during the convergent period over the clusters. A false positive rate was computed in a similar way but averaged over the non-cluster areas. A threshold was picked in a range of possible values of the measure. Then, the threshold was varied over the range to get a curve by plotting the true positive rates, proportion of high risk areas exceeding a threshold, against false positive rates, proportion of low risk areas exceeding a threshold. These curves are seen in Figure 2.

We examined a range of localized concentration detections including case probability, CPO, and local goodness of fit measures. The convolution models with ICAR and WCAR with the exponential decay ($WCAR^{exp}$) have a better performance in terms of ROC curves; however, the WCAR with the exponential decay has the most area under curve in the case of case probability. From the Figure 3, ICAR appears to produce an overly smooth case probability surface where WCAR which accounts for distance information estimate the surface more reasonably. However, WCAR with the inverse distance function seems to be too localized and yields more false positives than WCAR with the exponential decay. CPO (Figure 3) and local information criteria (Figure 4) produce similar results as they appear to be sensitive to noises.

5 Data example

Non-Hodgkin's lymphoma (NHL) is cancer that originates in the lymphatic system and the tumors develop from a type of white blood cell. NHL is more common than the other general type of lymphoma, Hodgkin lymphoma. We analyzed the incidence of NHL for South Carolina over a period of 15 years (Figure 5). We have requested all SC registry patients diagnosed with NHL as primary diagnosis. Locational information down to census block level is available for the period 1996–2010. These data are at a spatial and temporal scale fine enough to ensure that point process methodology is appropriate and there are no multiple occurrences within any spatial unit.

Since the data are of patients diagnosed with the disease, we do not have information about population at risk. In this analysis, we compare between patients diagnosed stage 1 versus patients diagnosed in the later stages, and define the label as stage 1 to be 0 and later stages to be 1. We adopted the model described in “Spatiotemporal Modeling for Case Event Data” section using the CAR model with the exponential decay weight with $\alpha = 1$ because it has the best method in the simulation study. We carried out the posterior sampling with the mixed Gibbs–Metropolis algorithms implemented in WinBUGS. The burn-in period of 10,000 samples with two chains was used to evaluate the convergence of the chains, and, after the burn-in, a sample of 10,000 iterations was obtained as the converged posterior sample.

Figure 6 displays the probability maps of being diagnosed in the early stage (control) comparing to late stages (case) with $WCAR^{exp}$ of South Carolina NHL incidences during January until June in 1996. It seems that some areas in the north have particularly high probabilities to be diagnosed with late stages during first four months but there is a possible cluster of incidence of late stage NHL in a southeast region in June 1996. However, interpretation and decision making with the incidence probabilities should be done with knowledge of their nature and quality or reliability.

Figure 7 shows the temporal profile of exceedence probabilities for NHL South Carolina incidences during April–May 1996. Circles are labels for early (0) and late (1) stages. Red and orange dots represent observations with exceedence probabilities greater than 0.95 and 0.9, respectively. The cut-off value of 0.9 may be used as an indicator for a moderate temporal effect of incidences whereas the threshold at 0.95 provides a stronger alarm. Note that although it is a temporal profile, the model also accounts for the underlying distribution of baseline locations. Therefore, if a case happens among surrounding controls, that observation would send a weaker signal than the ones within a cluster of cases. This paradigm is sensible for case event data and an advantage of our model.

Richardson et al.²⁹ proposed in the context of disease counts that posterior probability values can be used to support interpretation of areas with actual excess risk. Regions where the relative risk is more than 1 which is a null value and the posterior probabilities are greater than a high (nearly one) number can be more assuredly considered as having high risk. For binary outcomes, we set a cut off at 0.5 and a region with exceedence larger than 0.95 are considered to be an incidence hot spot. Figure 8 shows the maps of exceedence probability

of South Carolina NHL incidences. Nonetheless, exceedence probabilities can be sensitive to model specification and Rotejanaprasert¹³ suggested incorporating residual information into consideration because the cluster information may be also left in the residuals instead of posterior mean estimates. However, classical dichotomous residuals can be difficult to work with due to the discrete nature of the response variable. The exceedence probability may not be appropriate in the context of case event data.

The Bayesian approach provides a rich output associated to the estimated case probability and allows us to measure its reliability in terms of credibility intervals. The one-sided credible intervals of case probability at each location are calculated. If the lower limit of the interval excludes the value of 0.5, that location would be determined as signaling. There are two levels of credible intervals considered: 90% and 95%. Figures 9 and 10 show the maps of 90% and 95% lower limits of the NHL incidence probability estimates. The 90% credible interval may be used as an indicator for clusters of moderate probability of incidence while 95% credible intervals provide a strong signal of high localized concentration where some intervention perhaps is needed. These proposed spatial and temporal surveillance procedures can aid public health practitioners to monitor unusual spatial case events.

6 Conclusion and discussion

Although count data are an important focus area, it is also often important to consider finer spatial scale events such as events at residential addresses. In our approach, we exploit the fact that we have individual level information about cases but also wish to model overall variation in space–time risk. Unlike earlier work, we adopt a conditional logistic spatial model for the covariate linkage conditioning on the point process superposition of cases and controls. This allows for the rich structure of cancer registry information to be accommodated at the individual level. In addition, it lets the surveillance to be performed directly on the binary logistic model which is natural inclusion of individual level covariates and less computationally intensive.

The convolution model proposed by Besag et al.⁹ including a correlated effect based on an ICAR model and uncorrelated Gaussian random effects is perhaps the most popular choice in Bayesian disease mapping for small area count data. For case event data, it is sensible to include distance information into the CAR structure. Thus, we extend the ICAR specification to account for the distance effect in irregular-shape or case event data as the WCAR using two distance functions, the inverse distance ($WCAR^{inv}$) and exponential decay ($WCAR^{exp}$).

To evaluate clustering behavior, a number of local measures of goodness of fit are examined. These include exceedence probabilities, CPO and local information criteria. We also propose an alternative to detect localized concentration based on KL divergence criteria which is related to CPO. To compare mathematical expressions of the local measures, we can see that the proposed local KL is a function of CPO. Furthermore, the formulation of the local DIC and local KL are in a similar form. They both include deviance but have different penalty terms. However, DIC, which is an analog of AIC, is based on point estimation, the Bayes estimate in this case. This can make DIC or perhaps also AIC an inappropriate measure of

goodness of fit for singular models such as mixture models.^{22–24} On the other hand, CPO or WAIC which are based on pointwise computing are more suitable when singular models are of interest.^{24,25} Nonetheless, the proposed local KL may be useful as a global goodness of fit as the summation of the partition, $\sum_j KL_j$, which is easily computed.

$$\begin{aligned} CPO_i &= f(y_i | \mathbf{y}_{-i}) \\ DIC_i &= -2E_{\theta | \mathbf{y}}(\log f(y_i | \theta)) + 2\log f(y_i | \hat{\theta}) \\ WAIC_i &= \log(E_{\theta | \mathbf{y}} f(y_i | \theta)) + \text{var}_{\theta | \mathbf{y}}(\log(f(y_i | \theta))) \\ KL_i &= -2E_{\theta | \mathbf{y}}(\log f(y_i | \theta)) + 2\log(CPO_i). \end{aligned}$$

With the simulated data, the convolution models with ICAR and WCAR with the exponential decay have a better performance in terms of ROC curves (Figure 2); however, the WCAR with the exponential decay has the most area under curve in the case of case probability. Figure 3 shows that ICAR appears to produce an overly smooth case probability surface where WCAR which accounts for distance information estimate the surface more reasonably. However, WCAR with the inverse distance function seems to be too localized and yields more false positives than WCAR with the exponential decay. CPO (Figure 3) and local information criteria (Figure 4) produce similar results as they appear to be sensitive to noises.

We also demonstrated the performance of surveillance techniques with real data. The application concerns with the distribution of patients diagnosed with NHL. Locational information down to census block level is available for the period 1996–2010. These data are at a spatial and temporal scale fine enough to ensure that point process methodology is appropriate. Interpretation and decision making with the incidence probabilities should be done with knowledge of their nature and quality or reliability. The one-sided credible intervals of case probability estimate and exceedence probability at each location are calculated as a measure of localized concentration of moderate and high NHL incidences. This proposed surveillance system can aid for public health practitioners to monitor unusual spatial case events.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grant number R03CA179665 from the National Cancer Institute and the data were provided by South Carolina Central Cancer Registry, South Carolina Department of Health and Environmental Control.

Appendix 1

$$KL(P_i, P_{-i}) = \int \dots \int P_i \log \left(\frac{P_i}{P_{-i}} \right) d\theta_i$$

where

$$P_i = p(\boldsymbol{\theta} | \mathbf{y}_i), \mathbf{y}_i = (y_1, \dots, y_i) \text{ and } P_{-i} = p(\boldsymbol{\theta} | \mathbf{y}_{-i}), \mathbf{y}_{-i} = (y_1, \dots, y_{i-1})$$

By Bayes' theorem

$$p(\boldsymbol{\theta} | \mathbf{y}_i) = \frac{p(\mathbf{y}_i | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y}_i)} \text{ and } p(\boldsymbol{\theta} | \mathbf{y}_{i-1}) = \frac{p(\mathbf{y}_{i-1} | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y}_{i-1})}$$

Then

$$\frac{p(\boldsymbol{\theta} | \mathbf{y}_i)}{p(\boldsymbol{\theta} | \mathbf{y}_{i-1})} = \frac{p(\mathbf{y}_i | \boldsymbol{\theta})}{p(\mathbf{y}_{i-1} | \boldsymbol{\theta})} \times \frac{p(\mathbf{y}_{i-1})}{p(\mathbf{y}_i)} \times \frac{p(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} \text{ and } \log \left[\frac{p(\boldsymbol{\theta} | \mathbf{y}_i)}{p(\boldsymbol{\theta} | \mathbf{y}_{i-1})} \right] = \log \left[\frac{p(\mathbf{y}_i | \boldsymbol{\theta})}{p(\mathbf{y}_{i-1} | \boldsymbol{\theta})} \right] + \log \left[\frac{p(\mathbf{y}_{i-1})}{p(\mathbf{y}_i)} \right]$$

For the second term with conditional independence in likelihood

$$\begin{aligned} \log \left[\frac{p(\mathbf{y}_{i-1})}{p(\mathbf{y}_i)} \right] &= \log \left[\frac{\int \dots \int \frac{p(\mathbf{y}_{i-1} | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y}_i)} d\boldsymbol{\theta}}{p(\mathbf{y}_i)} \right] \\ &= \log \left[\int \dots \int \frac{p(\mathbf{y}_{i-1} | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y}_i)} d\boldsymbol{\theta} \right] \\ &= \log \left[\int \dots \int \frac{p(\mathbf{y}_{i-1} | \boldsymbol{\theta})}{p(\mathbf{y}_i | \boldsymbol{\theta})} \frac{p(\mathbf{y}_i | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y}_i)} d\boldsymbol{\theta} \right] \\ &= \log \left[\int \dots \int \frac{p(\mathbf{y}_{i-1} | \boldsymbol{\theta})}{p(\mathbf{y}_i | \boldsymbol{\theta})} p(\boldsymbol{\theta} | \mathbf{y}_i) d\boldsymbol{\theta} \right] \\ &= \log E_{\boldsymbol{\theta} | \mathbf{y}_i} \left[\frac{p(\mathbf{y}_{i-1} | \boldsymbol{\theta})}{p(\mathbf{y}_i | \boldsymbol{\theta})} \right] = \log E_{\boldsymbol{\theta} | \mathbf{y}_i} \left[\frac{1}{p(\mathbf{y}_i | \boldsymbol{\theta})} \right] \end{aligned}$$

Thus,

$$\begin{aligned} &LKL(P_i, P_{i-1}) \\ &= \int \dots \int p(\boldsymbol{\theta} | \mathbf{y}_i) \times \left[\log \frac{p(\mathbf{y}_i | \boldsymbol{\theta})}{p(\mathbf{y}_{i-1} | \boldsymbol{\theta})} + \log E_{\boldsymbol{\theta} | \mathbf{y}_i} \left[\frac{p(\mathbf{y}_{i-1} | \boldsymbol{\theta})}{p(\mathbf{y}_i | \boldsymbol{\theta})} \right] \right] d\boldsymbol{\theta} \\ &= \int \dots \int p(\boldsymbol{\theta} | \mathbf{y}_i) \times \left[\log \frac{p(\mathbf{y}_i | \boldsymbol{\theta})}{p(\mathbf{y}_{i-1} | \boldsymbol{\theta})} \right] d\boldsymbol{\theta} + \log E_{\boldsymbol{\theta} | \mathbf{y}_i} \left[\frac{p(\mathbf{y}_{i-1} | \boldsymbol{\theta})}{p(\mathbf{y}_i | \boldsymbol{\theta})} \right] \\ &= E_{\boldsymbol{\theta} | \mathbf{y}_i} [\log p(\mathbf{y}_i | \boldsymbol{\theta})] + \log E_{\boldsymbol{\theta} | \mathbf{y}_i} \left[\frac{1}{p(\mathbf{y}_i | \boldsymbol{\theta})} \right] \\ &= E_{\boldsymbol{\theta} | \mathbf{y}_i} [\log p(\mathbf{y}_i | \boldsymbol{\theta})] - \log(CPO_i). \end{aligned}$$

References

1. Buehler JW, Berkelman RL, Hartley DM, et al. Syndromic surveillance and bioterrorism-related epidemics. *Emerg Infect Dis.* 2003; 9:1197. [PubMed: 14609452]
2. Cooper, GF., Dash, DH., Levander, JD., et al. Proceedings of the 20th conference on uncertainty in artificial intelligence. Corvallis, OR: AUAI Press; 2004. Bayesian biosurveillance of disease outbreaks; p. 94-103.
3. Zou J, Karr AF, Datta G, et al. A Bayesian spatio-temporal approach for real-time detection of disease outbreaks: a case study. *BMC Med Inform Decis Mak.* 2014; 14:108. [PubMed: 25476843]
4. Jiang X, Cooper GF. A Bayesian spatio-temporal method for disease outbreak detection. *J Am Med Inform Assoc.* 2010; 17:462-471. [PubMed: 20595315]
5. Corberán-Vallet A, Lawson AB. Prospective analysis of infectious disease surveillance data using syndromic information. *Stat Meth Med Res.* 2014; 23:572-590.
6. Diggle P, Rowlingson B, Su TI. Point process methodology for on – line spatio – temporal disease surveillance. *Environmetrics.* 2005; 16:423-434.

7. Brix A, Diggle PJ. Spatiotemporal prediction for log – Gaussian Cox processes. *J R Stat Soc B StatMethodol.* 2001; 63:823–841.
8. Lawson, AB. Bayesian disease mapping: hierarchical modeling in spatial epidemiology. 2nd. Boca Raton, FL: CRC Press; 2013.
9. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math.* 1991; 43:1–20.
10. Turner, R. deldir: Delaunay Triangulation and Dirichlet (Voronoi) Tessellation. 2014. <http://CRANR-project.org/package=deldir> R package version 01–7
11. Lawson A, Biggeri A, Boehning D, et al. Disease mapping models: an empirical evaluation. Disease Mapping Collaborative Group. *Stat Med.* 2000; 19:2217–2241. [PubMed: 10960849]
12. Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. *Stat Meth Med Res.* 2005; 14:35–59.
13. Rotejanaprasert C. Evaluation of cluster recovery for small area relative risk models. *Stat Meth Med Res.* 2014; 23:531–551.
14. Rodrigues EC, Assunção R. Bayesian spatial models with a mixture neighborhood structure. *J Multivar Analy.* 2012; 109:88–102.
15. Lawson AB. Bayesian point event modeling in spatial and environmental epidemiology. *Stat Meth Med Res.* 2012; 21:509–529.
16. White G, Ghosh SK. A stochastic neighborhood conditional autoregressive model for spatial data. *Computat Stat Data Anal.* 2009; 53:3033–3046.
17. Hrafnkelsson B, Cressie N. Hierarchical modeling of count data with application to nuclear fall-out. *Environ Ecol Stat.* 2003; 10:179–200.
18. Geisser S, Eddy WF. A predictive approach to model selection. *J Am Stat Assoc.* 1979; 74:153–160.
19. Stone M. An asymptotic equivalence of choice of model by cross-validation and Akaike’s criterion. *J R Stat Soc B Methodol.* 1977; 39:44–47.
20. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. *J R Stat Soc B StatMethodol.* 2002; 64:583–639.
21. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comput.* 2013; 24:1–20.
22. Spiegelhalter DJ, Best NG, Carlin BP, et al. The deviance information criterion: 12 years on. *J R Stat Soc B Stat Methodol.* 2014; 76:485–493.
23. Celeux G, Forbes F, Robert CP, et al. Deviance information criteria for missing data models. *Bayesian Analy.* 2006; 1:651–673.
24. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res.* 2010; 11:3571–3594.
25. Watanabe, S. Algebraic geometry and statistical learning theory. Cambridge: Cambridge University Press; 2009.
26. Carlin BP, Carlin BP. An expected utility approach to influence diagnostics. *J Am Stat Assoc.* 1991; 86:1013–1021.
27. Pettit L, Smith A. Outliers and influential observations in linear models. *Bayesian Stat.* 1985; 2:473–494.
28. Rowlingson B, Diggle P, Bivand R, et al. SplanCs: spatial and space-time point pattern analysis. R Package Version. 2013 2.01-33.
29. Richardson S, Thomson A, Best N, et al. Interpreting posterior relative risk estimates in disease-mapping studies. *Environ Health Perspect.* 2004; 112:1016. [PubMed: 15198922]

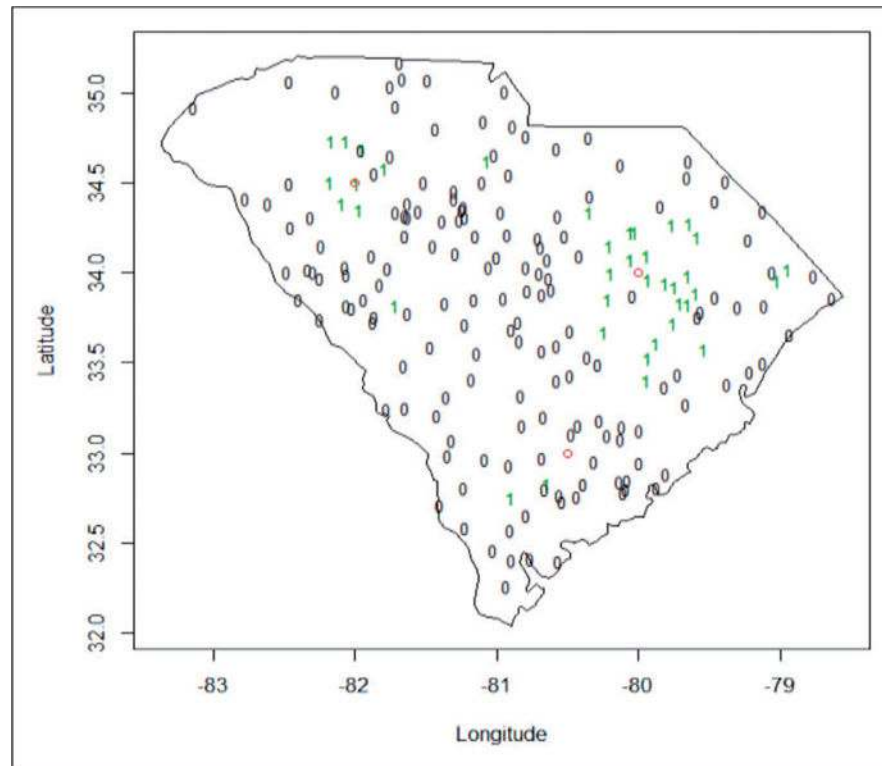


Figure 1.
A realization of cases (green) and controls (black) simulated from a completely random spatial process. Red dots represent the centroid of the three cluster locations.

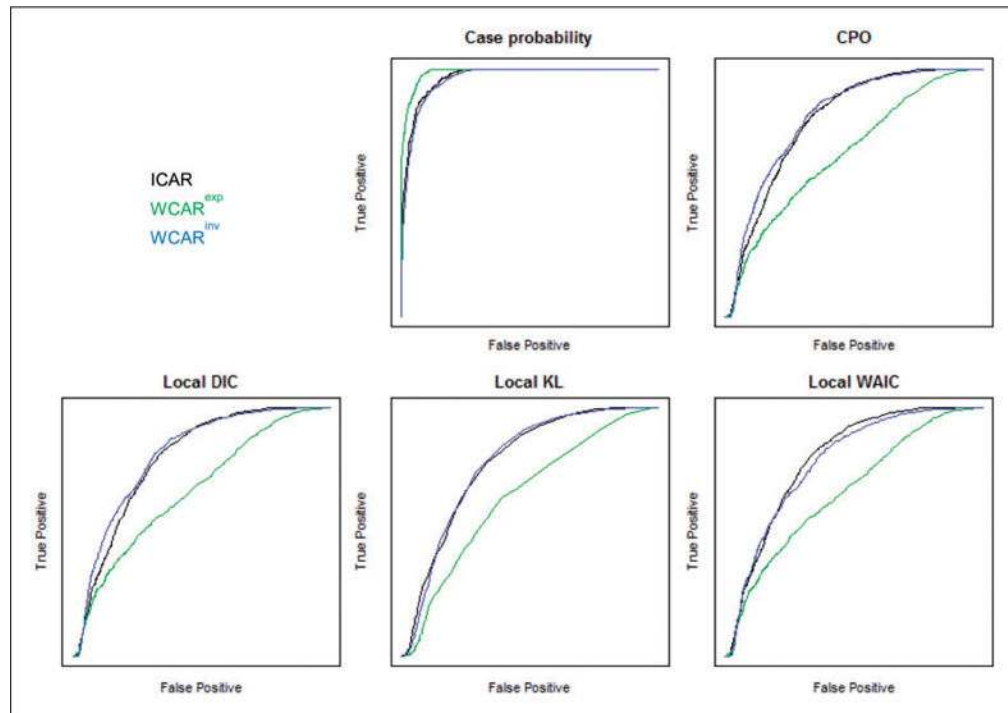


Figure 2. ROC curves of different cluster diagnostics with ICAR (black), $WCAR^{exp}$ (exponential decay, green) and $WCAR^{inv}$ (inverse distance, blue).

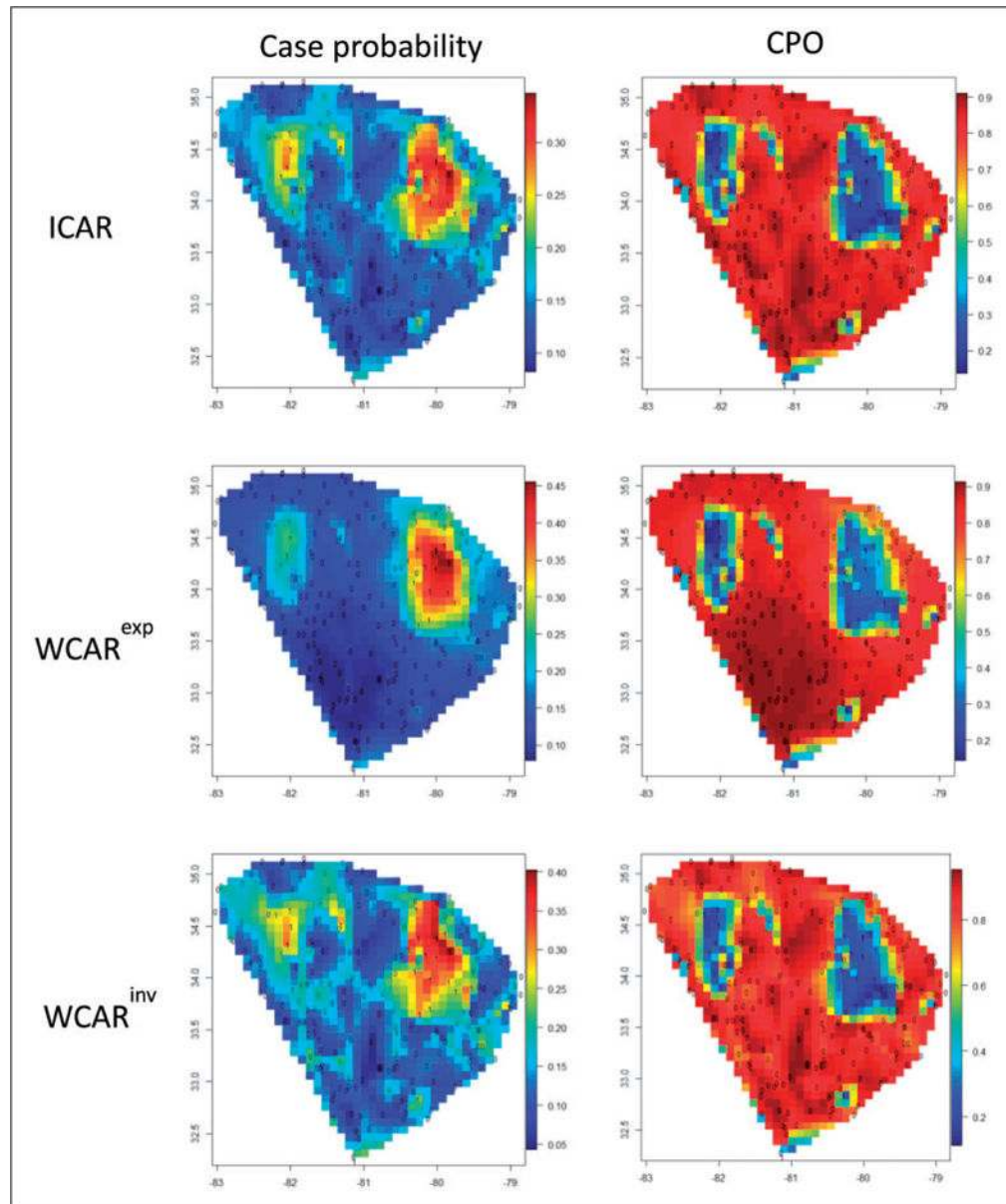


Figure 3. Maps of estimated case probability (p) and CPO during day 200 from a simulated dataset modeled using ICAR, WCAR with exponential decay (WCAR^{exp}) and WCAR with inverse distance (WCAR^{inv}).

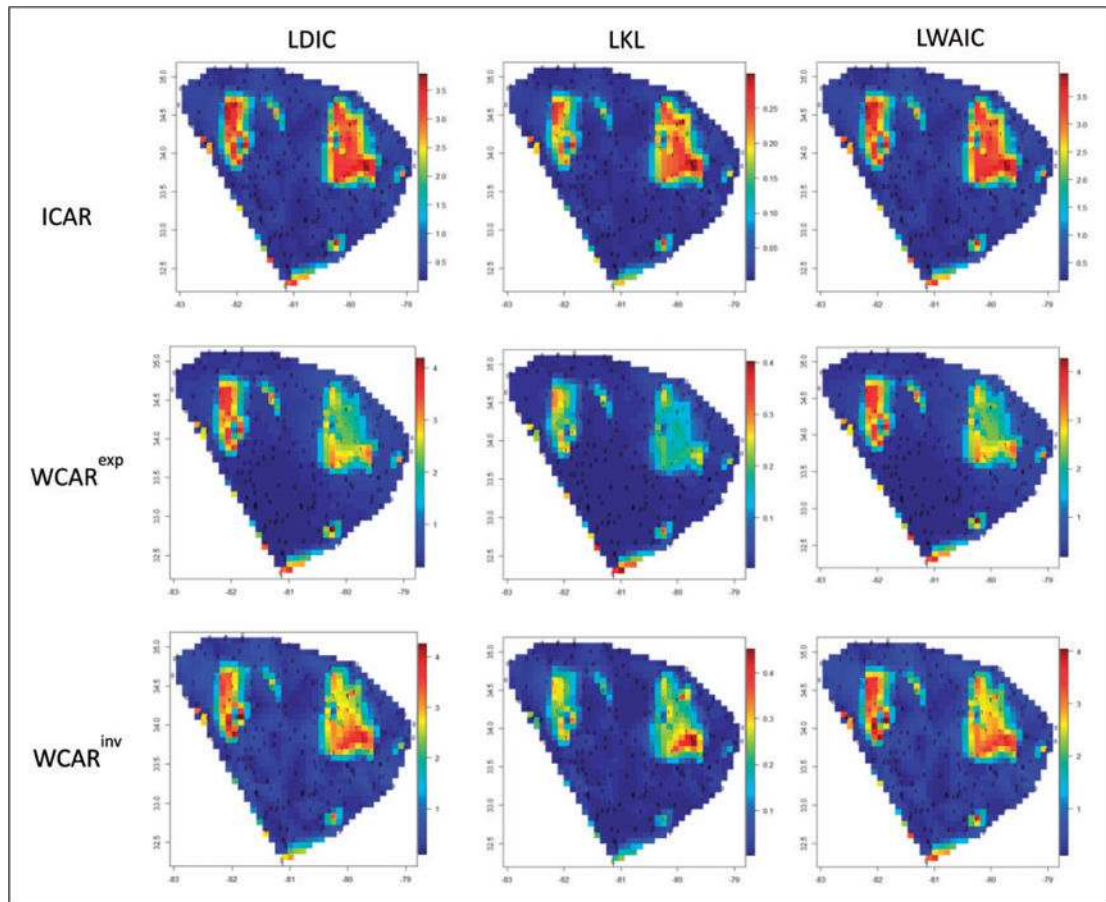


Figure 4. Maps of LDIC, LKL and LWAIC during day 200 from a simulated dataset modeled using ICAR, WCAR with exponential decay ($WCAR^{exp}$) and WCAR with inverse distance ($WCAR^{inv}$).

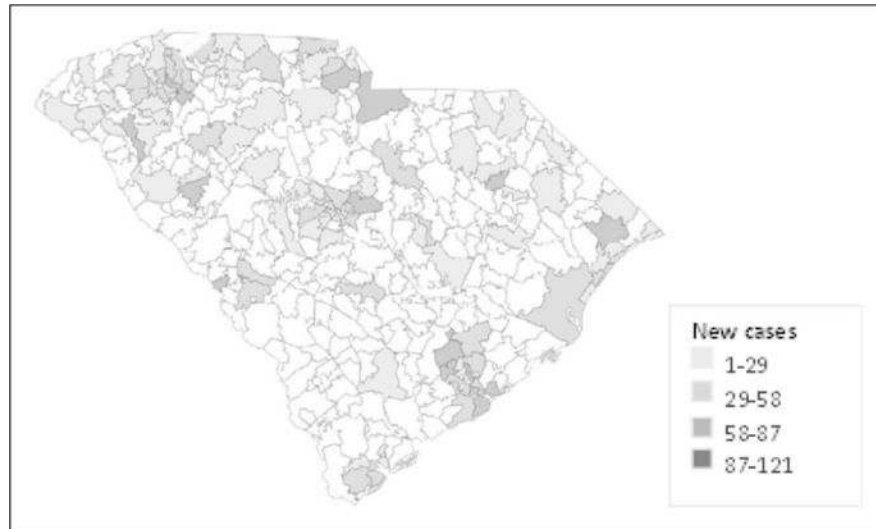


Figure 5. The zip code map of Non-Hodgkin's lymphoma incidences during 1990–1996 in South Carolina.

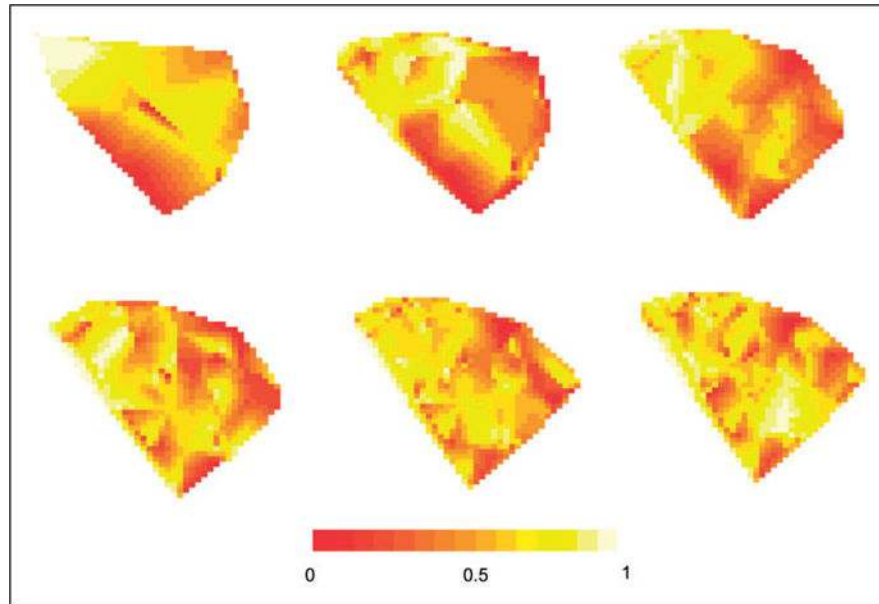


Figure 6. The probability maps of being diagnosed in the early stage vs. late stages with $WCAR^{exp}$ of South Carolina NHL incidences during January (top left) – June (bottom right) in 1996.

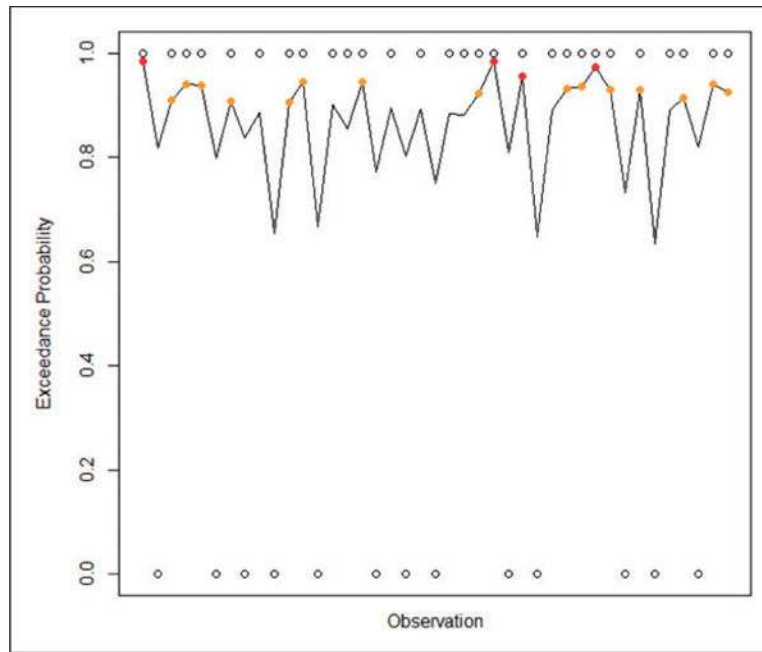


Figure 7. The temporal profile of exceedance probabilities for NHL South Carolina incidences during April–May 1996. Circles are labels for early (0) and late (1) stages. Red and orange dots represent observations with exceedance probabilities greater than 0.95 and 0.9, respectively.

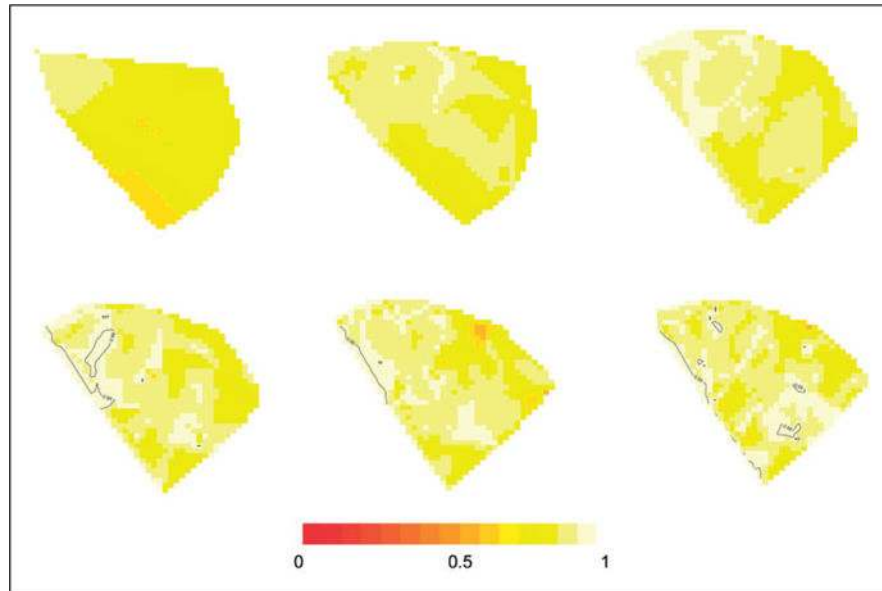


Figure 8. The maps of exceedence probability of South Carolina NHL incidences during January (top left) – June (bottom right) in 1996. The contours indicate the areas with exceedence probability greater than 0.95.

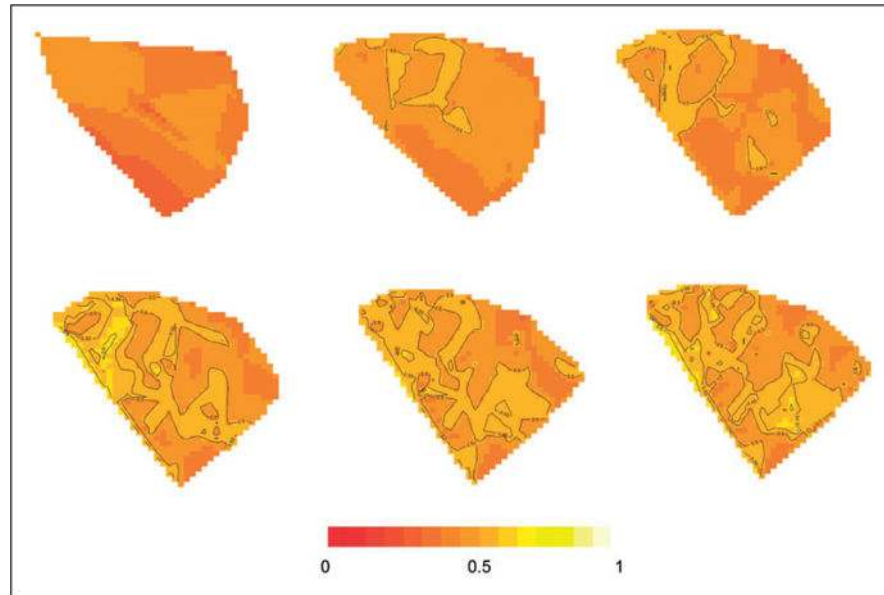


Figure 9. The maps of lower limit of 90% credible interval of incidence probability being diagnosed in the early stage vs. late stages with $WCAR^{exp}$ of South Carolina NHL incidences during January (top left) – June (bottom right) in 1996. The contours indicate the areas with the lower limit greater than 0.5.

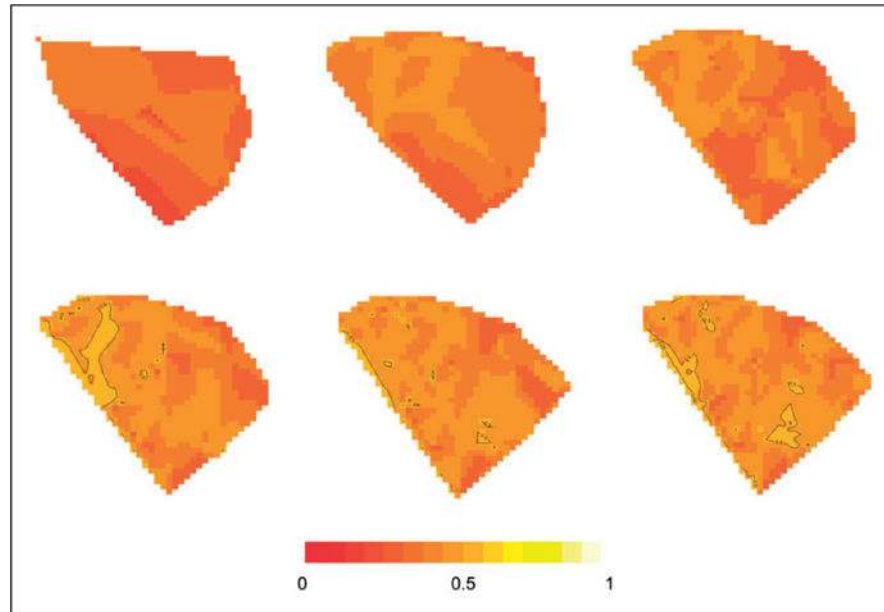


Figure 10. The maps of lower limit of 95% credible interval of incidence probability being diagnosed in the early stage vs. late stages with $WCAR^{exp}$ of South Carolina NHL incidences during January (top left) – June (bottom right) in 1996. The contours indicate the areas with the lower limit greater than 0.5.