Bayesian Scan Statistics

Daniel B. Neill

Contents

Introduction . . . 2
Univariate Bayesian Scan Statistics . . . 5
Multivariate Bayesian Scan Statistics . . . 7
Fast Subset Sums . . . 10
Learning Models for Bayesian Spatial Scanning . . . 12
Alternative Approaches to Bayesian Spatial Scanning . . . 13
Bayesian Network Scan Statistics . . . 14
Bayesian Cluster Detection and Modeling Approaches . . . 16
Summary and Future Directions . . . 18
References . . . 20

Abstract

In this chapter we describe Bayesian scan statistics, a class of methods which build both on the prior literature on scan statistics and on Bayesian approaches to cluster detection and modeling. We first compare and contrast the Bayesian scan to the traditional, frequentist hypothesis testing approach to scan statistics and summarize the advantages and disadvantages of each approach. We then focus on three different Bayesian scan statistic approaches: the Bayesian variable window scan statistic, the multivariate Bayesian scan statistic and extensions, and scan statistic approaches based on Bayesian networks. We describe each of these approaches in detail and compare these to related Bayesian scan methods and to the wider literature on Bayesian cluster detection and modeling. Finally, we discuss several promising areas for future work in Bayesian scan statistics, including multiple cluster detection, nonparametric Bayesian approaches, exten-

D. B. Neill (✉)
Center for Urban Science and Progress, New York University, New York, NY, USA
e-mail: daniel.neill@nyu.edu

© Springer Science+Business Media, LLC, part of Springer Nature 2019

J. Glaz, M.V. Koutras (eds.), *Handbook of Scan Statistics*,
https://doi.org/10.1007/978-1-4614-8414-1_28-1

sion of Bayesian spatial scan to nonspatial datasets, and computationally efficient methods for model learning and detection.

Keywords
Bayes' theorem · Posterior probability · Informative priors · Bayesian variable window scan statistic · Bayesian spatial scan · Multivariate Bayesian scan statistic · Fast subset sums · Bayesian network scan statistics

Introduction

Bayesian scan statistics are a relatively new class of methods that build both on the prior literature on scan statistics (Naus, 1965), particularly spatial and subset scanning (Kulldorff, 1997; Neill, 2012), as well as Bayesian approaches to cluster detection and modeling (Lawson and Denison, 2002). Bayesian scan methods view the detection problem from the perspective of Bayesian statistical inference, as opposed to the more traditional, frequentist hypothesis testing approaches pioneered by Naus (1965) and Kulldorff (1997). In practice, this differing perspective often results in three main distinctions between frequentist and Bayesian scans:

1. Bayesian scans tend to incorporate *informative prior information* about the expected size, shape, and other attributes of the affected subset of the data, as well as how this subset is affected by the event of interest. For example, the multivariate Bayesian scan statistic (Neill et al., 2007; Neill and Cooper, 2010), a Bayesian extension of Kulldorff's spatial scan statistic, models multiple event types, specifying for each event type both its prior distribution over space-time regions and its effects on the monitored data streams.
2. Bayesian scans tend to output probabilistic inferences about the *posterior probabilities* of different alternative hypotheses, thus quantifying their degree of uncertainty regarding the distribution over possible hypotheses. Frequentist scans, on the other hand, tend to identify a single, most likely alternative hypothesis, and then perform hypothesis testing to decide whether the null hypothesis can be rejected in favor of this alternative. For example, given a set of alternative hypotheses $H_1(S)$, each representing the occurrence of an event of interest (e.g., a spatial cluster of disease cases) in some subset of the data S, and the null hypothesis H_0 representing no events of interest, the Bayesian spatial scan (Neill et al., 2006a,b) computes the posterior probability $\Pr(H_1(S) \mid D)$, given the observed dataset D, for each alternative hypothesis. Applying Bayes' theorem, we can write $\Pr(H_1(S) \mid D) = \Pr(D \mid H_1(S))\Pr(H_1(S))/\Pr(D)$. The traditional spatial scan approach (Kulldorff, 1997), on the other hand, computes the hypothesis $H_1(S)$ which maximizes the likelihood ratio $\Pr(D|H_1(S))/\Pr(D|H_0)$.

   We note that both $\Pr(D)$ and $\Pr(D|H_0)$, in the expressions above, are independent of S. Thus, the subset S which maximizes posterior probability in the Bayesian setting, given a uniform prior over subsets $\Pr(H_1(S))$, is identical to the subset that maximizes the frequentist likelihood ratio statistic, assuming an identical model for the likelihood of the data in each case. Nevertheless, the frequentist and Bayesian approaches would provide different information about the alternative hypothesis $H_1(S)$: the former indicates whether $H_1(S)$ is sufficiently high scoring to reject the null hypothesis at a given significance level α, while the latter compares the posterior probability of $H_1(S)$ to other alternative hypotheses and to the null hypothesis H_0.

3. Bayesian scans tend to use a *marginal likelihood* approach, performing Bayesian model averaging to compute the total (summed) likelihood of multiple hypotheses of interest, or averaging over multiple parameter values for a given hypothesis. Frequentist scans, on the other hand, tend to use a *maximum likelihood* approach, computing the single most likely hypothesis and assuming the parameter values that maximize the likelihood of each hypothesis. More precisely, assuming that the null hypothesis H_0 and each alternative hypothesis $H_1(S)$ are point hypotheses with no free parameters, the frequentist scan uses the likelihood ratio $\Pr(D|H_1(S))/\Pr(D|H_0)$, as noted above. A more interesting situation arises when each hypothesis has some parameter space Θ: let $\theta_1(S) \in \Theta_1(S)$ denote parameters for the alternative hypothesis $H_1(S)$, and let $\theta_0 \in \Theta_0$ denote parameters for the null hypothesis H_0. For example, Kulldorff's spatial scan (Kulldorff, 1997) uses the maximum likelihood values of the relative risks q_{in} and q_{out} inside and outside region S, assuming $c_i \sim \text{Poisson}(q_{\text{in}} b_i)$ for locations $s_i \in S$ and $c_i \sim \text{Poisson}(q_{\text{out}} b_i)$ for locations $s_i \notin S$, respectively, and the maximum likelihood value of the relative risk q_{all} under the null hypothesis H_0, assuming $c_i \sim \text{Poisson}(q_{\text{all}} b_i)$. The typical, *maximum likelihood* framework uses the estimates of each set of parameters that maximize the likelihood of the data:

$$F(S) = \frac{\max_{\theta_1(S)\in\Theta_1(S)} \Pr(D|H_1(S), \theta_1(S))}{\max_{\theta_0\in\Theta_0} \Pr(D|H_0, \theta_0)}$$

The *marginal likelihood* framework instead averages over the possible values of each parameter:

$$F(S) = \frac{\int_{\theta_1(S)\in\Theta_1(S)} \Pr(D|H_1(S), \theta_1(S))\Pr(\theta_1(S))}{\int_{\theta_0\in\Theta_0} \Pr(D|H_0, \theta_0)\Pr(\theta_0)}$$

Both maximum likelihood and marginal likelihood approaches have certain advantages. Maximum likelihood leads to a generalized likelihood ratio test (GLRT) in the frequentist scan framework, and in certain cases, such as Kulldorff's spatial scan (Kulldorff, 1997), this leads to an individually most powerful statistical test under the given model assumptions. On the other hand, marginal likelihood tends to produce better posterior probability estimates for the Bayesian scan framework, since it incorporates the uncertainty (and if available, informative prior information) about the distribution of parameter values.

We note that each of these typical differences between frequentist and Bayesian scan statistics may be individually insufficient to distinguish the two classes of methods. Bayesian scan statistics may use an uninformative prior, while prior information can be incorporated into frequentist scan statistic approaches via hard constraints, e.g., on spatial proximity (Neill, 2012), or via maximization of a penalized likelihood ratio statistic (Gangnon and Clayton, 2004; Cancado et al., 2010; Speakman et al., 2013). In the latter case, penalties have been applied to counteract the inherent bias of spatial scanning toward finding clusters in areas with higher spatial resolution (Gangnon and Clayton, 2004), to penalize irregularly shaped and disconnected clusters (Cancado et al., 2010), and to reward dynamic clusters that change smoothly over time (Speakman et al., 2013). These penalties can be roughly interpreted as the prior log-odds of each alternative hypothesis (Speakman et al., 2013), resulting in a maximum a posteriori (MAP) estimate of the true affected subset. Similarly, marginal likelihood approaches have been used in a frequentist setting, either with informative prior weights (Gangnon and Clayton, 2001, 2004) or with uninformative priors, resulting in a simpler "average likelihood" approach (Chan, 2009). In either case, proponents of marginal likelihood argue that such approaches make better use of secondary cluster information as compared to the standard spatial scan, since the null hypothesis can be rejected based on multiple, moderately high-scoring clusters rather than a single, extremely high-scoring cluster (Chan, 2009). Nevertheless, these approaches generally do not compute the marginal likelihood over a continuous parameter space, in contrast to the Bayesian and multivariate Bayesian scan statistics (Neill et al., 2006a,b, 2007; Neill and Cooper, 2010) described below.

While we do not attempt to weigh in on the age-old debate between frequentist and Bayesian statistical methods in general (Neapolitan, 2008), we note that the Bayesian scan has both advantages and disadvantages compared to the more typical, frequentist scan approach:

- Bayesian methods can better quantify their uncertainty over alternative hypotheses, as well as integrate this information into interpretable graphical displays such as the *posterior probability map*. Frequentist methods instead draw conclusions, based on significance testing, as to whether to reject the null in favor of an alternative hypothesis $H_1(S)$ signifying an event of interest and characterizing the affected subset of the data.
- Bayesian scanning tends to have higher detection power when an informative prior can be accurately specified but can lose power for poorly chosen priors (Neill and Cooper, 2010). When expert knowledge is unavailable to specify the priors, this fact makes it essential to develop methods for efficiently learning models from data in the Bayesian framework (Makatchev and Neill, 2008a,b).
- With uninformative priors, maximization of the posterior probability in the Bayesian setting may reduce to likelihood ratio maximization, possibly with extra layers of hierarchy, e.g., a Gamma-Poisson statistic instead of a Poisson likelihood ratio (Neill et al., 2006a,b). Such hierarchical models may improve

detection or may simply complicate the model, making computation more difficult.

- Both frequentist and Bayesian approaches can be computationally expensive in some cases: frequentist methods may require randomization testing to determine statistical significance, while Bayesian methods may also require simulation (e.g., by Markov chain Monte Carlo) unless simple, conjugate priors are used. Nevertheless, both settings enable novel algorithmic approaches to efficiently search over the exponentially many subsets of the data, such as linear-time subset scanning (Neill, 2012) in the frequentist setting, or fast subset sums (Neill, 2011; Neill and Liu, 2011; Shao et al., 2011) in the Bayesian setting.
- Both frequentist and Bayesian approaches can be extended to integrate information from multiple data streams (Kulldorff et al., 2007; Neill et al., 2007), while approaches such as the multivariate Bayesian scan statistic can accurately model and distinguish between multiple event types (Neill and Cooper, 2010).

In the remainder of this chapter, we focus on three different Bayesian scan statistic approaches: the Bayesian variable window scan statistic (Zhang and Glaz, 2008), the multivariate Bayesian scan statistic and extensions (Neill et al., 2007; Neill and Cooper, 2010), and scan statistic approaches based on Bayesian networks (Neill et al., 2009). These approaches are described in some detail, as representatives of the larger class of Bayesian scan methods, and are compared to related methods as well as the large body of prior work on Bayesian mapping and modeling, which has also led to useful Bayesian approaches for cluster detection (Lawson and Denison, 2002).

Univariate Bayesian Scan Statistics

Work on Bayesian change-point detection in time series data has been ongoing since the 1970s (Smith, 1975; Barry and Hartigan, 1993). However, it is only in the last decade that researchers have integrated more general scan statistic approaches with Bayesian modeling and have applied these Bayesian scan statistic approaches to spatial (two-dimensional or higher-dimensional) data or to more general datasets. The first such approaches include the Bayesian variable window scan statistic (Zhang and Glaz, 2008) and the univariate Bayesian spatial scan (Neill et al., 2006a,b). Both approaches build on the two-dimensional discrete scan statistic setting (Chen and Glaz, 1996) and on Kulldorff's spatial scan (Kulldorff, 1997). They assume small-area count data mapped to a uniform grid and extend the frequentist (Poisson or binomial) likelihood ratio statistics through the development of Bayesian hierarchical models. The univariate Bayesian spatial scan is a special case of the multivariate Bayesian scan statistic (MBSS) approach (Neill et al., 2007; Neill and Cooper, 2010) described in detail below, so we focus mainly on the Bayesian variable window scan (Zhang and Glaz, 2008), which we denote by BVWS, in this section.

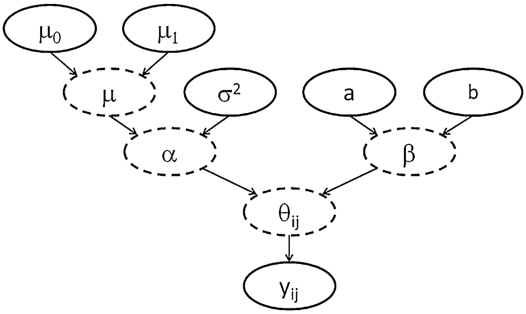

Fig. 1 Bayesian hierarchical model for the Bayesian variable window scan statistic (Zhang and Glaz, 2008). Solid lines denote observed variables or those with values assumed to be known. Dashed lines denote latent variables

As noted above, BVWS assumes small-area counts y_{ij} aggregated to an $N \times N$ grid. Following the two-dimensional discrete scan statistic setting of Chen and Glaz (1996), it scans over the set of alternative hypotheses $H_1(S)$, each assuming a cluster in some $m \times m$ square subregion S of the grid. The null hypothesis H_0 assumes that no clusters are present. The counts y_{ij} are modeled as independent Poisson or Bernoulli random variables. For each grid square (i, j), we have either $y_{ij} \sim \text{Poisson}(\theta_{ij})$ or $y_{ij} \sim \text{Bernoulli}(p_{ij})$, where the Poisson means θ_{ij} or the Bernoulli log-odds $\theta_{ij} = \log(p_{ij}/(1 - p_{ij}))$ are assumed to be latent variables with a two-stage prior (Fig. 1). The first-stage prior for θ_{ij} is assumed to be normally distributed with mean α and variance β. The second-stage prior for α is a normal distribution with mean μ and variance σ^2, where $\mu = \mu_1$ under the alternative hypothesis H_1 and $\mu = \mu_0$ under the null hypothesis H_0. The second-stage prior for β is an inverse Gamma distribution with shape parameter a and scale parameter b. The values of μ_1, μ_0, σ^2, a, and b are assumed to be known. Since a non-conjugate prior is used, the likelihoods of the data given each hypothesis cannot be computed in closed form. Instead, a Gibbs sampling approach with auxiliary variables is used to generate posterior samples $\theta_{ij}^{(0)}(n)$ and $\theta_{ij}^{(1)}(n)$, for $n = 1 \ldots N_{\text{samples}}$, under the null and alternative hypotheses, respectively. See Zhang and Glaz (2008) for further details. Then, following Kass and Raftery (1995), the Bayes factor $B_{01}(S)$ for a given subset S is defined to be the ratio of the harmonic means of the likelihood values given samples θ_{ij} from the null and alternative distributions, respectively:

$$B_{01}(S) = \frac{\left(\frac{1}{N_{\text{samples}}} \sum_{n=1 \ldots N_{\text{samples}}} \prod_{s_{ij} \in S} \Pr(y_{ij} | \theta_{ij}^{(0)}(n))^{-1}\right)^{-1}}{\left(\frac{1}{N_{\text{samples}}} \sum_{n=1 \ldots N_{\text{samples}}} \prod_{s_{ij} \in S} \Pr(y_{ij} | \theta_{ij}^{(1)}(n))^{-1}\right)^{-1}}$$

We can then compute the p-value $p(S)$ corresponding to the observed Bayes factor $B_{01}(S)$ for each S, comparing the observed value to the expected distribution of $B_{01}(S)$ under H_0, and use $p_{\min} = \min_S p(S)$ as a test statistic. Finally, in order to account for multiple hypothesis testing over the potentially large set of square regions S, statistical significance is computed by randomization, where the p-value is obtained by comparing $p_{\min}$ to its expected distribution under the null hypothesis H_0.

The univariate Bayesian spatial scan (denoted here as UBSS) is described below as a special case of MBSS with a single event type and single monitored data stream. UBSS differs from BVWS in several important ways. First, while UBSS also incorporates a hierarchical model (Gamma priors and Poisson counts), the parameters of the Gamma priors (α and β) are fit directly from data in an "empirical Bayes" approach. BVWS uses a more "fully Bayesian" approach with a two-stage hierarchical model, which may better capture the uncertainty in these parameter estimates. Second, UBSS uses a conjugate Gamma-Poisson prior, which enables the computation of an efficient, closed-form expression for the posterior probabilities. BVWS assumes a non-conjugate prior, necessitating the use of more computationally expensive Gibbs sampling techniques to compute posteriors. Finally, BVWS uses the Bayes factors $B_{01}(S)$ to compute p-values in a frequentist hypothesis testing approach, while UBSS incorporates the prior distribution over hypotheses $\Pr(H_1(S))$ and computes the posterior distribution $\Pr(H_1(S)|D)$.

Multivariate Bayesian Scan Statistics

Building on the univariate Bayesian spatial scan (Neill et al., 2006a,b), Neill et al. proposed the multivariate Bayesian scan statistic (MBSS) approach for event detection and characterization using multivariate spatial time series data (Neill et al., 2007; Neill and Cooper, 2010). The authors argue that MBSS has several advantages over the previously proposed, frequentist scan statistic approaches to event detection:

1. MBSS achieves high detection power, even when relatively uninformative priors are used, by combining information from multiple data streams, spatial locations, and time steps.
2. MBSS can incorporate informative priors, enabling much higher detection power for the specified and modeled event types. Priors can be pre-specified by expert knowledge or learned from labeled training data, as described below.
3. MBSS can accurately characterize events by specifying models for multiple event types and computing the probability that each type of event has occurred. This enables MBSS to model and distinguish between relevant events (e.g., a disease outbreak of interest to public health) and irrelevant events (e.g., a spike in over-the-counter medication sales that is due to a promotional sale rather than an outbreak).
4. MBSS is computationally efficient because of the use of conjugate priors. Unlike the frequentist approach, randomization testing is not necessary, which reduces runtime and leads to easier calibration of alerting thresholds.
5. MBSS results are easy to interpret, visualize, and use for decision-making. MBSS outputs the total posterior probability of each event type as well as the posterior probability that no events have occurred. For each event type, MBSS

provides the distribution of the posterior probability over space-time regions S. An intuitive way of viewing these results, the posterior probability map, is described below.

Given a set of space-time regions S to search and a set of event types E, MBSS computes the posterior probability $\Pr(H_1(S, E)|D)$ that each event type E has affected each space-time region S, given the observed dataset D consisting of multiple data streams $D_1 \ldots D_M$. Each data stream consists of spatial time series data collected at a set of spatial locations s_i, and for each combination of location s_i and stream D_m, we observe a time series of counts $c^t_{i,m}$. For example, in the multivariate disease surveillance problem, a given count $c^t_{i,m}$ could represent the number of emergency department visits with a specific symptom type D_m in zip code s_i on day t.

Given the multivariate, space-time count data, our task is threefold: to *detect* whether any events are occurring, *characterize* the event type, and *pinpoint* the affected space-time region (i.e., identifying both the affected subset of locations and the time duration for which these locations were affected). Thus, the MBSS approach has the goal of distinguishing between the set of alternative hypotheses $H_1(S, E)$, each representing the occurrence of an event of type E in a space-time region S, and the null hypothesis H_0 that no events have occurred. Each hypothesis $H_1(S, E)$ is assumed to be mutually exclusive, and thus we have $\Pr(H_0) + \sum_S \sum_E \Pr(H_1(S, E)) = 1$. Neill and Cooper (2010) assume a uniform prior over event types and space-time regions, i.e., $\Pr(H_1(S, E)) = \Pr(H_1)/(N_S N_E)$ for all S and E, where N_S and N_E, respectively, represent the numbers of space-time regions and event types under consideration. As described below, nonuniform priors can be estimated by various approaches for model learning from labeled (or partially labeled) training data.

Given the prior distribution over hypotheses, MBSS applies Bayes' theorem to compute the posterior probability of each hypothesis, integrating prior information about each event type with the observed multivariate dataset D:

$$\Pr(H_1(S, E)|D) = \frac{\Pr(D|H_1(S, E))\Pr(H_1(S, E))}{\Pr(D|H_0)\Pr(H_0) + \sum_S \sum_E \Pr(D|H_1(S, E))\Pr(H_1(S, E))}$$

The likelihood of the data given an alternative hypothesis $H_1(S, E)$ is computed assuming the Bayesian hierarchical model in Fig. 2. Observed counts $c^t_{i,m}$ are each assumed to have been drawn from a Poisson distribution with mean equal to the product of the expected count $b^t_{i,m}$ and the relative risk $q^t_{i,m}$, where the expected counts are learned from historical data by time series analysis. (One weakness of the MBSS approach is that it does not model the uncertainty introduced by estimating $b^t_{i,m}$ from data.) Relative risks $q^t_{i,m}$ are assumed to be drawn from a Gamma distribution with parameters $\alpha = x^t_{i,m}\alpha_m$ and $\beta = \beta_m$, where the parameters of the Gamma distribution for each stream under the null hypothesis (α_m, β_m) are estimated from historical data using an empirical Bayes approach. The

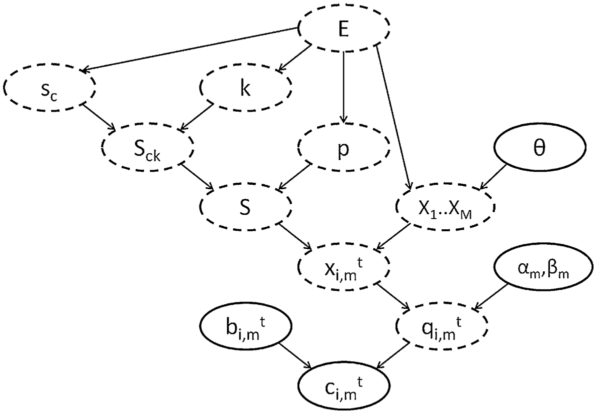

Fig. 2 Bayesian hierarchical model for the multivariate Bayesian scan statistic with generalized fast subset sums (Neill and Liu, 2011; Shao et al., 2011). Solid lines denote observed variables or those with values assumed to be known. Dashed lines denote latent variables. Note that the original multivariate Bayesian scan statistic (Neill et al., 2007; Neill and Cooper, 2010) and the original fast subset sums method (Neill, 2011) can be considered special cases of the generalized fast subset sums framework with sparsity parameter $p = 1$ and $p = 0.5$, respectively. The univariate Bayesian spatial scan can be considered a special case with sparsity parameter $p = 1$, a single event type E, and a single monitored data stream ($M = 1$)

$x^t_{i,m}$ represent the multiplicative effects of an event on the expected counts for each combination of location s_i, data stream D_m, and time step t. We note that $x^t_{i,m} = 1$ for all unaffected locations, streams, and time steps; under the null hypothesis H_0, $x^t_{i,m} = 1$ for all s_i, D_m, and t. Additionally, for a given occurrence of an event, $x^t_{i,m}$ is assumed to be uniform over the affected space-time region S for each data stream D_m. The effect X_m on a given data stream D_m is assumed to be a function of the event type E (which defines the "average" percent increase in each data stream given that event type) and the event severity (which multiplies the "average" percent increase for each data stream by the same constant θ). Effects of each event type on each data stream can be learned from labeled training data via maximum likelihood estimation, as described below. Alternatively, MBSS can be used as a "general" rather than "specific" event detector by defining $2^M - 1$ event models, each of which assumes that an event has uniform effects on some subset of the M monitored data streams (Neill and Cooper, 2010). As a "general" event detector, MBSS was able to achieve high detection power on a semisynthetic multivariate disease surveillance task and to identify the affected subset of data streams. When specific event models were learned from the data, these models dramatically increased detection power as well as enabling MBSS to distinguish between the multiple event types.

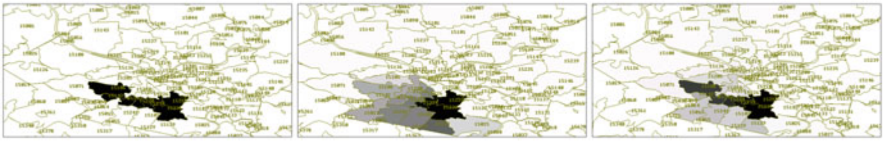

Fig. 3 Examples of posterior probability maps, from Neill (2011). The center and right panels are posterior probability maps formed by the multivariate Bayesian scan statistic (Neill et al., 2007; Neill and Cooper, 2010) and fast subset sums (Neill, 2011) methods, respectively, at the midpoint of a simulated disease outbreak. Darker shading denotes higher summed posterior probability $\Pr(H_1(s_i)|D) = \sum_{S:s_i \in S} \Pr(H_1(S)|D)$ for the given zip code s_i. Shaded zip codes in the left panel denote the true outbreak region

Fast Subset Sums

As discussed above, the multivariate Bayesian scan statistic (MBSS) can integrate information from multiple data streams and can model and distinguish between multiple event types. Given a set of space-time regions S, a set of modeled event types E, and the multivariate dataset D, MBSS calculates the posterior probabilities $\Pr(H_1(S, E)|D)$ that each event type has affected each space-time region. One useful and intuitive visualization of these outputs is the *posterior probability map*. Since the set of hypotheses $H_1(S, E)$ are assumed to be mutually exclusive, the total posterior probability that a given event type E has affected each spatial location s_i can be computed by summing the probabilities of all regions S containing s_i: $\Pr(H_1(s_i, E)|D) = \sum_{S:s_i \in S} \Pr(H_1(S, E)|D)$. These summed probabilities for each location can then be displayed on a map (Fig. 3), where darker shading corresponds to higher probability and different colors can be used for different event types. Unlike standard spatial scan visualizations, which do not compute probabilities but instead show the most likely cluster, this method is able to quantify its uncertainty about the spatial extent and type of events.

One disadvantage of the MBSS method, however, is the need to search over a typically very large number of space-time regions S, either to identify the hypothesis $H_1(S, E)$ with highest posterior probability, to enumerate all posterior probabilities $\Pr(H_1(S, E)|D)$ above some threshold value, or to compute the summed posterior probabilities $\Pr(H_1(s_i, E)|D)$ in order to display the posterior probability map. This limitation restricts the original MBSS approach to searching over regions of fixed shape, such as circles or rectangles, for computational feasibility. As a result, MBSS suffers from reduced power to detect elongated or irregular cluster shapes.

More recently, Neill (2011) proposed an efficient fast subset sums method which substantially improves detection power and accuracy for irregularly shaped regions. Fast subset sums extends the MBSS method by defining a hierarchical prior which assigns nonzero prior probabilities $\Pr(H_1(S, E))$ to every subset of locations while maintaining efficient computation of the posterior probability map. The key step is a computational shortcut that efficiently and exactly computes the summed posterior probability $\Pr(H_1(s_i, E)|D) = \sum_{S:s_i \in S} \Pr(H_1(S, E)|D)$

over all subsets containing location s_i, without computing the posterior probability of each individual subset. See Fig. 3 for an example comparing the posterior probability maps produced by MBSS (assuming a uniform prior over circular clusters) and fast subset sums for a simulated, irregularly shaped disease cluster. Fast subset sums is better able to capture the irregular shape of the cluster, enabling more timely and more accurate event detection.

This work was further extended to the generalized fast subset sums (GFSS) framework through incorporation of an additional parameter which allows the sparsity of the detected region to be controlled (Neill and Liu, 2011; Shao et al., 2011). As shown in Fig. 2, GFSS extends the Bayesian hierarchical model of MBSS (Neill and Cooper, 2010) by assuming that the affected spatial region S, for a given event of type E, is drawn from a hierarchical prior distribution with three steps. First, the center location s_c is drawn from a multinomial distribution. Second, the neighborhood size k is drawn from a multinomial distribution, thus defining the neighborhood S_{ck} consisting of location s_c and its $k-1$ nearest neighbors. Third, each location s_i in neighborhood S_{ck} is independently drawn from a Bernoulli distribution with parameter p, where s_i is included in the affected region S with probability p and excluded with probability $1-p$. The sparsity parameter p can be viewed as the expected proportion of locations affected within a given (circular) local neighborhood, and thus the original MBSS method (Neill and Cooper, 2010), assuming a uniform prior over circular regions, corresponds to a special case of GFSS with $p = 1$. The original fast subset sums method (Neill, 2011) does not include the sparsity parameter p, assuming uniform distributions over the center location s_c and the neighborhood size k and a uniform distribution over subsets $S \subseteq S_{ck}$. Shao et al. (2011) show that this is a special case of GFSS with $p = 0.5$. Additionally, they demonstrate that appropriate choice of the sparsity parameter p enables GFSS to achieve higher detection power and spatial accuracy than either MBSS or the original fast subset sums method. Moreover, they show that the distribution of the sparsity parameter can be accurately learned from a small amount of labeled training data, leading to improved detection, as described below.

Naive computation of the posterior probability map using GFSS would require computation of posterior probabilities for a number of subsets that scales exponentially with neighborhood size, which is computationally infeasible for $k > 25$. However, Shao et al. (2011) show that, for any value $0 < p \leq 1$, the posterior probability map can be computed without computing each individual region probability, thus reducing the run time from exponential to polynomial in k. The key trick is to note that the likelihood ratio of spatial region S (as compared to H_0) for a given event type E and event severity θ can be found by multiplying the individual likelihood ratios $LR(s_i | E, \theta)$ for all locations $s_i \in S$. Then the average likelihood ratio of the 2^k subsets for a given center s_c and neighborhood size k can be transformed from a sum of products to a product of sums, enabling us to write this quantity as the product of the smoothed likelihood ratios $(p \times LR(s_i | E, \theta) + (1-p))$ for all locations $s_i \in S_{ck}$. The contribution to the average likelihood ratio from the 2^{k-1} subsets containing a given location s_i can be found by computing this product of smoothed likelihood ratios for all locations $s_j \in S_{ck}$, $j \neq i$, and multiplying by $(p \times LR(s_i|E, \theta))$. We can then marginalize over the distributions of centers s_c, neighborhood sizes k, and severity values θ and normalize to compute the posterior probability map. More details are provided in Shao et al. (2011). In practice, this enables GFSS to run in time comparable to MBSS, i.e., computing the posterior probability map in seconds for each day of data, for the real-world disease surveillance tasks described by Shao et al. (2011).

Learning Models for Bayesian Spatial Scanning

As noted above, the multivariate Bayesian scan statistic can model and distinguish between multiple event types E. To do so, various parameters must be specified for each event model, including the prior probability $\Pr(H_1(E))$, the distribution of this prior probability over space-time regions $\Pr(H_1(S, E))$, and the average effects of event E on each of the M monitored data streams. The original MBSS approach (Neill et al., 2007; Neill and Cooper, 2010) assumes a uniform distribution over event types and space-time regions but learns the average effects $x_{\text{km,avg}}$ of each event type E_k on each monitored data stream D_m by maximum likelihood estimation. Assuming labeled training examples for which the event type E and affected subset S are known, $x_{\text{km,avg}}$ can be computed as the average ratio of the total count $\sum c^t_{i,m}$ to total baseline $\sum b^t_{i,m}$ for data stream D_m in regions affected by event type E_k. This approach was shown to improve detection power for the modeled event types as well as enabling MBSS to accurately determine which event type is occurring. Neill (2007) also proposes learning the prior probability of each event type $\Pr(H_1(E))$ and the conditional probability that the event occurs in each spatial region $\Pr(H_1(S, E)|E)$ by smoothed maximum likelihood estimation. However, the number of possible space-time regions is typically large, and a very large number of training examples are typically needed to accurately model a nonuniform distribution over regions.

An alternative approach is to assume a parameterized prior distribution over spatial regions S and to learn the parameters of that distribution for each event type E. Makatchev and Neill (2008a,b) propose a simple generative model that assumes a latent center location s_c and radius parameter r for each event. Each location is assumed to be affected with probability $(1 + \exp((d - r)/h))^{-1}$, where d is the location's distance from the center. The center location s_c for a given event type E follows a multinomial distribution. The radius r is assumed to follow a uniform or Gaussian distribution with mean μ learned from data, and the bandwidth h is also learned from data. Since each example specifies the affected spatial region S but not the underlying model parameters, parameter distributions are estimated using a generalized expectation-maximization (GEM) algorithm. Then the prior probabilities $\Pr(H_1(S, E))$ can be calculated directly from the learned models. Makatchev and Neill (2008b) show that event models can be accurately learned from a small number of labeled training examples and that the resulting models significantly improve detection performance as compared to MBSS with uninformative priors. Two disadvantages of this approach are the significant computational expense of the GEM algorithm and the restriction (as in the original MBSS approach) to a relatively small, exhaustively enumerable set of search regions S, such as circular or rectangular regions.

As noted above, the generalized fast subset sums framework (Neill and Liu, 2011; Shao et al., 2011) also proposes a parameterized prior distribution over the hypotheses $\Pr(H_1(S, E))$ for each event type E. However, this framework allows efficient computation of the posterior probability map, summing probabilities over the exponentially many subsets of the data S that contain a given location s_i, to calculate the total posterior $\Pr(H_1(s_i, E)|D)$. As noted above, the GFSS framework assumes a hierarchical model where the center location s_c and neighborhood size k are drawn from multinomial distributions, and then each location in the resulting neighborhood S_{ck} is either included with probability p or excluded with probability $1 - p$, for some sparsity parameter p. Shao et al. (2011) show that the distribution of the sparsity parameter p can be accurately learned from a small amount of labeled training data and that the resulting GFSS method with learned p distribution outperforms MBSS, the original fast subset sums method, and GFSS with a uniform p distribution. They also demonstrate that two otherwise identical event types with different sparsities can be reliably distinguished by learning each event's p distribution. Finally, they show that learning both an event's sparsity distribution and its relative effects on different data streams, as in Neill and Cooper (2010), leads to more timely detection and better characterization than learning either parameter on its own. Even better detection and characterization accuracy might be achieved in future work by jointly learning each event type's distribution over center locations s_c, neighborhood sizes k, and sparsity parameters p, as discussed below.

Alternative Approaches to Bayesian Spatial Scanning

We now consider how the multivariate Bayesian scan statistic framework described above differs from the previous work of Gangnon and Clayton (2001, 2004) on weighted average likelihood ratio (WALR) scan statistics, as well as describing several recent variants of Bayesian spatial scan. Gangnon and Clayton (2001) define the WALR statistic as a weighted average of the likelihood ratio statistics $F(S) = \Pr(D|H_1(S))/\Pr(D|H_0)$, i.e., $WALR = \sum \text{weight}(S)F(S)$, where $\text{weight}(S)$ corresponds to the (unnormalized) prior probability of $H_1(S)$. They then estimate the posterior probabilities $\Pr(H_1(S)|D) \propto \text{weight}(S)F(S)/WALR$. This approach differs from MBSS in three ways: first, it does not incorporate multiple data streams or multiple event types. Second, it uses maximum likelihood estimates of the relative risk parameters (q_{in}, q_{out}, q_{all}), rather than marginal likelihoods, thus presenting an upwardly biased estimate of each posterior probability. Third, it uses a hypothesis test to decide whether to reject H_0 in favor of H_1, instead of incorporating the prior probabilities $\Pr(H_1(S))$ and calculating the corresponding posterior probabilities. Thus, the WALR statistic can be thought of as a maximum likelihood approximation to the posterior probabilities $\Pr(H_1(S)|D)$ computed by the MBSS approach. Similarly, the WALRS statistic (Gangnon and Clayton, 2004) computes a weighted average of the likelihood ratios for regions containing a given location: $WALRS(s_i) = \sum_{S:s_i \in S} \text{weight}(S)F(S)$, with the maximum value $WALRS = \max_{s_i} WALRS(s_i)$ used as a frequentist test statistic. This approach can be considered a maximum likelihood-based approximation to the posterior probability map, $\Pr(H_1(s_i)|D) = \sum_{S:s_i \in S} \Pr(H_1(S)|D)$, computed by MBSS.

Recently proposed variants of UBSS and MBSS include the Bayesian beta-Bernoulli scan statistic (Read, 2011) and the rank-based scan statistic (Que and Tsui, 2008, 2011), as well as the Bayesian network scan statistics described below. Read (2011) proposes a straightforward variation of UBSS that substitutes a beta-Bernoulli model in place of the Gamma-Poisson model and argues that this approach is more appropriate for spatially distributed, binary labeled point data, as opposed to small-area count data as in the UBSS approach. Que and Tsui (2008, 2011) use the UBSS approach in two stages, first computing $\Pr(H_1(S)|D)$ for the single-element subsets S consisting of each individual location, and ranking the locations by these posterior probabilities. Then a greedy growth heuristic is used to form and evaluate clusters, where at each step the algorithm adds the highest-ranked adjacent location to the cluster and each such cluster is scored using the posterior computed by UBSS. Empirical results suggest that this approach is effective at identifying anomalous clusters. One disadvantage of the rank-based approach, as compared to UBSS, is that the "prior" distribution over clusters (assumed to be uniform over all clusters created by the algorithm) is specified after rather than prior to the search, and as such one would expect the resulting "posterior" probabilities to be upwardly biased, since higher priors are placed on subsets with higher observed likelihood given the data.

Bayesian Network Scan Statistics

Bayesian networks, a type of probabilistic graphical model, are a useful tool for modeling, inference, and learning from multivariate data. As described by Neill et al. (2009), several recent scan statistic approaches incorporate Bayesian networks either implicitly (e.g., the relationships between the variables in the MBSS approach can be described using a Bayesian network) or explicitly. Here we review several of the approaches described by Neill et al. (2009), including the entity-based scan statistic (Jiang et al., 2010) and anomalous group detection (Das et al., 2009), as well as several more recent methods (Jiang and Cooper, 2010; McFowland III et al., 2013).

Jiang et al. (2010) developed a Bayesian network scan statistic approach, the entity-based scan statistic (EBSS), which combines spatial and population-based approaches to detection. EBSS builds on both the multivariate Bayesian scan statistic (Neill et al., 2007; Neill and Cooper, 2010) and the Bayesian network model of PANDA (Cooper et al., 2004). PANDA models the relationships between variables including the presence, type, and severity of a disease outbreak, latent variables representing the underlying disease state

$D_r \in \{anthrax, influenza, \ldots, none\}$ of each individual r in the population, and observed variables $I_r \in \{cough, fever, chest\ \ pain, \ldots, other, no\ ED\}$ representing whether that individual visits the emergency department with a particular chief complaint type or does not visit the emergency department. The EBSS model adds a spatial component to the PANDA model, modeling the spatial region S affected by the outbreak as a latent variable (as in MBSS) and specifying the effects of an outbreak on individuals' disease states in the affected region. EBSS is similar to MBSS in that it uses a Bayesian model to differentiate between multiple event types and computes the posterior probabilities $\Pr(H_1(S, E)|D)$, but it models the effects of the event on each individual in a population rather than on a set of monitored data streams. This approach may be preferable to MBSS given detailed individual-level data, but it may be less useful when only aggregate count data is available. Jiang and Cooper (2010) further extend the approach of Jiang et al. (2010) by explicitly modeling the temporal trend of case counts given that an outbreak is occurring. This method assumes a linear increase in cases over time and models the number of days since the start of the outbreak as a latent variable.

Another recent set of approaches (Das et al., 2009; McFowland III et al., 2013) use Bayesian networks to detect patterns in general datasets, where each data record R_i has observed values v_{ij} for a set of categorical attributes A_j. These approaches first learn the structure and/or parameters of a Bayesian network model M_0 given the null hypothesis H_0, using "clean" training data that is assumed not to contain any patterns of interest. Given a separate set of test data, which may contain patterns of interest, the goal is to find related subsets of data records that are collectively anomalous given the null model M_0. The anomalous group detection (AGD) approach (Das et al., 2009; Neill et al., 2009) scans over related subsets of the data (as enumerated by a greedy search method), computes a likelihood ratio statistic for each subset, and reports the highest-scoring subsets. The novelty of this approach is that the likelihood ratio statistic $F(S)$ compares the likelihood of the observed data given a "local Bayesian network" (learned only from the given subset of the data S) to the likelihood of that data given the "global Bayesian network" learned from the entire training dataset. This method was demonstrated to accurately detect anomalous groups in disease surveillance and container shipping datasets but risks overfitting by learning a complex, multivariate model from a small subset of data records. It also has the disadvantage of high computational complexity, since a Bayesian network must be learned "on the fly" for each evaluated subset.

More recently, McFowland III et al. (2013) proposed a "Fast Generalized Subset Scan" approach for pattern detection. This approach consists of four steps: (1) efficiently learning a Bayesian network which represents the assumed null distribution of the data; (2) computing the conditional probability of each attribute value in the dataset given the Bayesian network, conditioned on the other attribute values for that record; (3) computing an empirical p-value corresponding to each attribute value by ranking the conditional probabilities, where under the null hypothesis we expect empirical p-values to be uniformly distributed on [0,1]; and (4) using a nonparametric scan statistic to detect subsets of records and attributes with an unexpectedly large number of low (significant) empirical p-values. The final step is computationally expensive (exponential in the numbers of records and attributes for a naive search), but the linear-time subset scanning property (Neill, 2012) can be used to speed up this search, converging to a local maximum of the score function and ensuring that each iteration step is linear (not exponential) in the number of records or attributes. FGSS was evaluated on multiple application domains, including early detection of simulated anthrax bio-attacks, discovery of patterns of illicit container shipments for customs monitoring, and network intrusion detection, demonstrating improved detection accuracy, efficient runtime, and ability to correctly characterize the affected subset of attributes in each domain. FGSS was shown to consistently outperform AGD and other previously proposed methods in terms of detection power and characterization accuracy and scales to much larger datasets. It is worth noting, however, that neither FGSS nor AGD are Bayesian approaches in the sense of incorporating priors over the possible alternative hypotheses $H_1(S)$ and computing the posterior probability of each hypothesis. Instead, Bayesian networks are used as a component of a frequentist approach that identifies high-scoring subsets and optionally computes their statistical significance by randomization testing. As discussed below, extension of truly "Bayesian" scan statistic approaches such as MBSS to more general datasets remains an interesting open problem.

Bayesian Cluster Detection and Modeling Approaches

Since the 1980s, the spatial epidemiology literature has developed a number of Bayesian spatial modeling approaches that focus on estimating and mapping spatially smoothed disease rates from small-area counts (Clayton and Kaldor, 1987; Waller et al., 1997; Knorr-Held and Raβer, 2000; Gangnon and Clayton, 2000). For example, Clayton and Kaldor (1987) assume a Gamma-Poisson model and estimate the parameters of the Gamma distribution using an empirical Bayes approach, while Waller et al. (1997) assume a log-linear model for location-specific disease rates. These models can incorporate both spatial autocorrelation and spatial heterogeneity but do not explicitly model cluster locations.

More recent approaches such as Knorr-Held and Raβer (2000) and Gangnon and Clayton (2000, 2003, 2007) propose Bayesian models that are more appropriate for cluster detection. These *spatial cluster modeling* methods attempt to combine the benefits of disease mapping and spatial cluster detection, by constructing a probabilistic model in which the underlying clusters are explicitly represented. For a more detailed discussion of spatial cluster modeling, see Lawson and Denison (2002). A typical approach is to assume that the observed counts are generated by some underlying process model which depends on a set of cluster centers, where the number and locations of cluster centers are unknown. Typically, a common disease rate for locations in the same cluster is assumed (Knorr-Held and Raβer, 2000; Gangnon and Clayton, 2000). Then we attempt to simultaneously infer all the parameters of the model, including the cluster centers and the disease risks in each cluster. Knorr-Held and Raβer (2000) assume that the study area is partitioned based on a set of latent center locations, where each location belongs to the partition with nearest center. Most similarly to the scan statistic approaches described above, Gangnon and Clayton (2000, 2003, 2007) assume a large background area and a small number of clusters, where the prior probability of a set of clusters is determined based on geographic characteristics such as size and shape.

These Bayesian cluster modeling approaches have many similarities to the Bayesian scan statistic methods described above, as well as some distinct advantages and disadvantages. Typically, precise cluster locations are inferred, and models with different numbers of cluster centers can be compared, giving an indication of both whether there are any clusters and where each cluster is located. Cluster modeling approaches can better model the presence of multiple clusters, as well as adjusting for observed covariates and accounting for spatial autocorrelation. Additionally, hidden Markov models can be used to model the underlying latent state of each location on each time step, thus allowing recently proposed Bayesian cluster modeling approaches such as Heaton et al. (2012) to capture the spatial spread of events over time. A similar generative model of event propagation was used in the frequentist, penalized likelihood ratio scan setting by Speakman et al. (2013), but incorporation of temporal dynamics into the Bayesian scan setting is still in its early stages.

One typical disadvantage of Bayesian spatial cluster modeling methods, as compared to Bayesian scan statistics, is their computational burden: the underlying models rarely have closed-form solutions, and the Markov chain Monte Carlo methods used to approximate the model parameters are often computationally intensive. In these models, the number of clusters or partitions is typically unknown, requiring the use of a reversible jump Markov chain Monte Carlo method (Green, 1995) which allows clusters to be added or deleted as part of the process of sampling from the posterior distribution. This approach is computationally expensive, but an alternative is to use a fixed, overly large number of cluster centers or partitions (Gangnon and Clayton, 2007). This alternative approach simplifies inference and leads to more efficient computation; though the identification of clusters is less clear, the method is still able to present evidence of local clustering through the use of Bayes factors. Finally, we note that, unlike the multivariate Bayesian scan statistic framework (Neill et al., 2007; Neill and Cooper, 2010) described above, Bayesian cluster modeling approaches are not typically able to model and distinguish between multiple event types or to integrate multiple data sources for detection. One exception is the recently proposed Bayesian conditional autoregressive model of Banks et al. (2012), which considers disease surveillance using multiple data streams, but it is unlikely that such an approach would scale to large numbers of locations and data streams without the expenditure of vast amounts of computing resources.

Summary and Future Directions

Bayesian scan statistics are a recent and promising new development in the scan statistics literature. These approaches can integrate prior information and multiple data sources for more accurate cluster detection in both spatial and nonspatial data and can model and distinguish between multiple event types. Bayesian scans can be used to detect and pinpoint clusters as well as quantify the amount of uncertainty in the spatial extent of each cluster, and the posterior probability map (representing the summed posterior probability of all subsets containing a given location) is an intuitive visual representation of the posterior probability distribution. Finally, the use of conjugate priors and efficient computational methods such as the generalized fast subset sums framework (Neill, 2011; Neill and Liu, 2011; Shao et al., 2011) can enable Bayesian scan statistics to scale to large numbers of locations and data streams while maintaining both flexible cluster models and computational feasibility.

Future research in Bayesian scan statistics might proceed in many directions, both addressing some of the current weaknesses of Bayesian scan approaches and building on their strengths. For example, most Bayesian scan approaches assume that at most one cluster is present in the data, comparing the alternative hypotheses $H_1(S)$ (where S is the affected subset of the data) to the null hypothesis H_0 of no clusters. The prior distribution $\Pr(H_1(S))$ assumes that these hypotheses are mutually exclusive, and thus the posterior distribution will often place all of its probability on a single cluster even if multiple distinct clusters are present. Several extensions of the Bayesian scan to multiple clusters might be possible. These range from simple approaches that are common in the frequentist setting (such as removing the most significant cluster and re-running the algorithm) to defining prior distributions over multiple clusters as in Bayesian cluster modeling approaches. However, the former approaches no longer produce a single, valid posterior probability distribution, while the latter approaches may lose the computational advantages of Bayesian scanning. For example, in the MBSS approach described above, an exhaustive computation of the probability of each alternative hypothesis $H_1(S_1, S_2, \ldots)$ would be difficult, since the number of hypotheses to be considered would scale exponentially with the maximum allowable number of clusters. It is an open question whether the posterior probability map (representing the summed posterior probabilities over all of these exponentially many hypotheses) can be efficiently computed in the generalized fast subset sums framework (Neill, 2011; Neill and Liu, 2011; Shao et al., 2011). One interesting approach to multiple cluster detection in the frequentist scan framework is the recently proposed latent source model of Cheng et al. (2013), which extends the temporal multiple cluster model of Xie et al. (2009) to spatial cluster detection. This approach demonstrated promising results for a mobile sensor network application to surveillance of nuclear materials, but it is not clear whether this hypothesis testing approach can be extended to compute posterior probabilities in a Bayesian scan framework.

Another interesting avenue for future research might be the extension of Bayesian scan statistics from parametric to nonparametric models. For example, Gaussian process regression is a useful representation that can be used for time series forecasting while accounting for multivariate correlations, and Dirichlet process priors can be useful for defining partition and cluster models. Finally, many interesting detection problems involve unstructured data such as text, for which Bayesian nonparametric models such as latent Dirichlet allocation (Blei et al., 2003) provide a useful representation and efficient inference methods for modeling "topics" (probability distributions over words). One recent approach that combines topic modeling with spatial scan is the semantic scan statistic (Liu and Neill, 2011). Semantic scan is able to detect novel disease outbreaks with previously unseen patterns of symptoms. To do so, it analyzes free-text chief complaint data from hospital emergency departments and identifies topics that are emerging in space and time.

As noted above, learning of models from labeled training data is a challenging but important aspect of the Bayesian scan framework. The incorporation of labeled data enables better modeling of multiple event types, allowing relevant patterns to be distinguished from irrelevant false positive clusters. While a variety of methods have been proposed to learn models from data, ranging from simple maximum likelihood to Bayesian network structure learning, few of these approaches have been integrated into the Bayesian scan framework for cluster detection. For example, expectation maximization (Dempster et al., 1977) is a useful approach to learning model parameters in the presence of latent variables and might be applied for joint learning of the multiple parameters (center location, neighborhood size, and sparsity parameter) in the generalized fast subset sums framework. Additional challenges arise when data is partially labeled (e.g., a training dataset might consist of multiple positive examples for which a cluster is present, but the cluster locations are not labeled), requiring the missing labels to be modeled as latent variables. The development of effective approaches for model learning from partially labeled data might enable incorporation of many more sources of data, leading to more accurate models and better detection.

Finally, there is an inherent tension in the Bayesian scan between computational efficiency (which often requires various simplifications and model assumptions) and more accurate representation of the underlying models of the real-world phenomena of interest. For example, Bayesian scan models typically assume a single affected subset and fail to model spatial and temporal variation in the effects of a cluster. Conjugate priors (such as the Gamma-Poisson model of MBSS) enable efficient computation but may lose the flexibility to account for spatial and temporal correlations, covariates, or other sources of variation in the data. Similarly, computationally efficient Bayesian scan methods have been developed only for spatiotemporal count data, but increased model flexibility (e.g., by the use of Bayesian networks to model the joint probability distribution) may allow approaches like MBSS to be extended to nonspatial datasets as well. The development of new Bayesian models that preserve the computational advantages of Bayesian scanning while incorporating more flexible models might enable these approaches to be useful for a wide array of new application domains.

References

Banks D, Datta G, Karr A, Lynch J, Niemi J, Vera F (2012) Bayesian CAR models for syndromic surveillance on multiple data streams: theory and practice. Inform Fusion 13:105–116

Barry D, Hartigan JA (1993) A Bayesian analysis of change point problems. J Am Stat Assoc 88:309–319

Blei D, Ng A, Jordan M (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022

Cancado ALF, Duarte AR, Duczmal LH, Ferreira SJ, Fonesca CM, Gontijo EC (2010) Penalized likelihood and multi-objective spatial scans for the detection and inference of irregular clusters. Int J Health Geogr 9:55

Chan HP (2009) Detection of spatial clustering with average likelihood ratio test statistics. Ann Stat 37:3985–4010

Chen J, Glaz J (1996) Two dimensional discrete scan statistics. Probab Stat Lett 31:59–68

Cheng JQ, Xie M, Chen R, Roberts F (2013) A latent source model to detect multiple spatial clusters with application in a mobile sensor network for surveillance of nuclear materials. J Am Stat Assoc 108(503):902–913

Clayton D, Kaldor J (1987) Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 43:671–681

Cooper GF, Dash D, Levander JD et al (2004) Bayesian biosurveillance of disease outbreaks. In: Maxwell D, Halpern J (eds) Proceedings of the Conference on Uncertainty in Artificial Intelligence, Banff, Canada, pp 94–103

Das K, Schneider J, Neill DB (2009) Detecting anomalous groups in categorical datasets. Technical report, Carnegie Mellon University, School of Computer Science

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B 39:1–38

Gangnon RE, Clayton MK (2000) Bayesian detection and modeling of spatial disease clustering. Biometrics 56(3):922–935

Gangnon RE, Clayton MK (2001) A weighted average likelihood ratio test for spatial disease clustering. Stat Med 20:2977–2987

Gangnon RE, Clayton MK (2003) A hierarchical model for spatial clustering of disease. Stat Med 22:3213–3228

Gangnon RE, Clayton MK (2004) Likelihood-based tests for detecting spatial clustering of disease. Environmetrics 15:797–810

Gangnon RE, Clayton MK (2007) Cluster detection using Bayes factors from overparameterized cluster models. Environ Ecol Stat 14:69–82

Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732

Heaton MJ, Banks DL, Zou J, Karr AF, Datta G, Lynch J, Vera F (2012) A spatio-temporal absorbing state model for disease and syndromic surveillance. Stat Med 31:2123–2136

Jiang X, Cooper GF (2010) A Bayesian spatio-temporal method for disease outbreak detection. J Am Med Inform Assoc 17:462–471

Jiang X, Neill DB, Cooper GF (2010) A Bayesian network model for spatial event surveillance. Int J Approx Reason 51:224–239

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90:773–795

Knorr-Held L, Raβer G (2000) Bayesian detection of clusters and discontinuities in disease maps. Biometrics 56:13–21

Kulldorff M (1997) A spatial scan statistic. Commun Stat Theory Methods 26(6):1481–1496

Kulldorff M, Mostashari F, Duczmal L, Yih WK, Kleinman K, Platt R (2007) Multivariate scan statistics for disease surveillance. Stat Med 26:1824–1833

Lawson AB, Denison DGT (eds) (2002) Spatial cluster modelling. Chapman & Hall/CRC, Boca Raton

Liu Y, Neill DB (2011) Detecting previously unseen outbreaks with novel symptom patterns. Emerg Health Threats J 4:11074

Makatchev M, Neill DB (2008a) Learning outbreak regions for Bayesian spatial biosurveillance. Adv Dis Surveill 5:45

Makatchev M, Neill DB (2008b) Learning outbreak regions in Bayesian spatial scan statistics. In: Proceedings of the ICML/UAI/COLT 2008 Workshop on Machine Learning for Health Care Applications. Helsinki, Finland

McFowland III E, Speakman S, Neill DB (2013) Fast generalized subset scan for anomalous pattern detection. J Mach Learn Res 14:1533–1561

Naus JI (1965) The distribution of the size of the maximum cluster of points on the line. J Am Stat Assoc 60:532–538

Neapolitan RE (2008) A polemic for Bayesian statistics. Innov Bayesian Netw Theory Appl 156: 7–32

Neill DB (2007) Incorporating learning into disease surveillance systems. Adv Dis Surveill 4:107

Neill DB (2011) Fast Bayesian scan statistics for multivariate event detection and visualization. Stat Med 30:455–469

Neill DB (2012) Fast subset scan for spatial pattern detection. J R Stat Soc (Ser B Stat Methodol) 74(2):337–360

Neill DB, Cooper GF (2010) A multivariate Bayesian scan statistic for early event detection and characterization. Mach Learn 79:261–282

Neill DB, Liu Y (2011) Generalized fast subset sums for Bayesian detection and visualization. Emerg Health Threats J 4:s43

Neill DB, Moore AW, Cooper GF (2006a) A Bayesian scan statistic for spatial cluster detection. Adv Dis Surveill 1:55

Neill DB, Moore AW, Cooper GF (2006b) A Bayesian spatial scan statistic. Adv Neural Inf Process Syst 18:1003–1010

Neill DB, Moore AW, Cooper GF (2007) A multivariate Bayesian scan statistic. Adv Dis Surveill 2:60

Neill DB, Cooper GF, Das K, Jiang X, Schneider J (2009) Bayesian network scan statistics for multivariate pattern detection. In: Glaz J, Pozdnyakov V, Wallenstein S (eds) Scan statistics: methods and applications. Birkhäuser, Boston, pp 221–250

Que J, Tsui FC (2008) A multi-level spatial clustering algorithm for detection of disease outbreaks. In: Proceedings of American Medical Informatics Association Annual Symposium, Washington, DC, pp 611–615

Que J, Tsui FC (2011) Rank-based spatial clustering: an algorithm for rapid outbreak detection. J Am Med Inform Assoc 18:218–224

Read S (2011) A Bayesian approach to the Bernoulli spatial scan statistic. Technical report, University of Sheffield

Shao K, Liu Y, Neill D (2011) A generalized fast subset sums framework for Bayesian event detection. In: Proceedings of the 11th IEEE International Conference on Data Mining, Vancouver, Canada, pp 617–625

Smith AFM (1975) A Bayesian approach to inference about a change-point in a sequence of random variables. Biometrika 62(2):407–416

Speakman S, Zhang Y, Neill DB (2013) Dynamic pattern detection with temporal consistency and connectivity constraints. In: 13th IEEE International Conference on Data Mining, Dallas, TX, pp 697–706

Waller L, Carlin B, Xia H, Gelfand A (1997) Hierarchical spatio-temporal mapping of disease rates. J Am Stat Assoc 92:607–617

Xie M, Sun Q, Naus J (2009) A latent model to detect multiple clusters of varying sizes. Biometrics 65:1011–1020

Zhang Z, Glaz J (2008) Bayesian variable window scan statistics. J Stat Plan Inference 138: 3561–3567