A Latent Model to Detect Multiple Spatial Clusters with Application in a Mobile Sensor Network for Surveillance of Nuclear Materials

Jerry Cheng, Minge Xie, Rong Chen and Fred Roberts[1]

Abstract

Nuclear attacks are among the most devastating terrorist attacks, with severe loss of human lives as well as damage to infrastructure. To deter such threats, it becomes increasingly vital to have sophisticated nuclear surveillance and detection systems deployed in major cities in the U.S., such as New York City. In this paper, we design a mobile sensor network and develop statistical algorithms and models to provide consistent and pervasive surveillance of nuclear materials in major cities. The network consists of a large number of vehicles on which nuclear sensors and Global Position System (GPS) tracking devices are installed. Real time sensor readings and GPS information are transmitted and processed at a central surveillance center. Mathematical and statistical analyses are performed, in which we mimic a signal-generating process and develop a latent modeling framework to detect multiple spatial clusters. An MC-EM algorithm is developed to estimate model parameters, detect significant clusters, and identify their locations and sizes. We also determine the number of clusters using a modified AIC/BIC

[1] Jerry Cheng is Assistant Professor, Robert Wood Johnson Medical School, University of Medicine & Dentistry of New Jersey, New Brunswick, NJ 08901. Minge Xie is Professor of Statistics, Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854. Rong Chen is Professor of Statistics, Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854 and Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100871, China. Fred Roberts is Professor of Mathematics and Director of the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS), Rutgers University, Piscataway, NJ 08854. Email addresses are chengjq@umdnj.edu, mxie@stat.rutgers.edu, rongchen@stat.rutgers.edu, and froberts@dimacs.rutgers.edu. This research was supported by research grants from NSF and DHS (NSF-CBET 0736134, DHS-2008-DN-077-ARI012). The second author's research is also supported in part by NSF and NSA (NSF-DMS-0915139, NSF-SES-0851521, NSA-H98230-08-1-0104). The third author' research is also supported in part by NSF (NSF-DMS-0800183, NSF-DMS-0905763, NSF-DMS-0915139). The authors thank our colleagues in DIMACS and the DHS Center of Excellence for Command Control and Interoperability (CCICADA), as well as Drs. Simon Labov and Karl Nelson of Lawrence Livermore National Laboratory (LLNL), for their comments and suggestions. The authors would also like to thank Mr. Dong Dai for his help in developing the computing code for heteroscdastic sensors and Tsvetan Asamov for his help on developing a Visual Basic tool used in simulation. The research developed in this article is the authors' own and it does not reflect the policy or view of any sponsor agencies.

criterion. Simulation studies to evaluate the effectiveness and detection power of such a network are described for both stationary and moving nuclear sources.

KEYWORDS: AIC and BIC criteria, cluster detection, EM algorithm, latent model, likelihood inference, MCMC algorithm, nuclear detection and surveillance, scan statistics

1. INTRODUCTION

Since the September Eleventh attacks, national security has garnered increased attention of ordinary people and it has become one of the top priorities of the United States government. Among all the possible attacks by terrorists, nuclear attack is potentially the most devastating, and the global proliferation of nuclear weapon technology has made the threat increasingly serious. The U.S. government has made significant efforts to curb nuclear proliferation. In spite of many accomplishments, no effort can give full assurance against a clandestine delivery of a nuclear weapon for a terrorist attack. The graveness of such a cataclysmic possibility is apparent. As a part of the effort, the Domestic Nuclear Detection Office (DNDO) within the Department of Homeland Security (DHS) was established in 2005 to improve the nation's capability to detect and collect information on unauthorized attempts to import, possess, store, develop, or transport nuclear or radiological material for use against the United States. The DNDO, in partnership with the National Science Foundation (NSF), has supported the Academic Research Initiative (ARI) program in frontier research at academic institutions focusing on detection systems, individual sensors or other research that is potentially relevant to the detection of nuclear weapons, special nuclear material, radiation dispersal devices and related threats. The Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) of Rutgers University, with the collaboration of the DHS University Center of Excellence for Command, Control, and Interoperability (CCICADA), which is based at DIMACS, is one of the academic institutes that is leading the effort in a multi-institution research project on nuclear detection supported by DNDO.

As a part of the research at DIMACS, this paper focuses on one of the aspects of detecting nuclear materials using a fleet of mobile radiation sensors in metropolitan areas. Major cities are attractive targets for terrorists because of their dense population and economic importance. However, a major city spreads throughout a large geographic area and is difficult to monitor and patrol. It is important to develop effective detection and surveillance methods to overcome or mitigate such a difficulty. Throughout the paper, we will use the generic term "nuclear detection" to include detection of any radiation-emitting materials of concern. Often, the nuclear and radiological materials of particular concern are radiation dispersion devices (RDDs), more commonly known as dirty bombs, and special nuclear materials (particularly highly enriched uranium and weaponsgrade plutonium) that could provide the fissile material for a nuclear weapon.

When a hidden stationary or moving nuclear source emits radioactive energy to its immediately surrounding area, a functional sensor within a certain range of the source would be activated and send out warning signals, though it is possible for a sensor to send a false alarm. If there are multiple sensors nearby, a cluster of activated sensors will be formed around the source. The cluster is often visible when plotted on a map, and it can in turn help reveal the location of the hidden source with high accuracy. This consideration translates the hidden source detection problem into a visible cluster detection problem. Statistically, we inspect the entire region of interest, and test whether one or more spatial clusters exist in the region and whether or not the spatial clusters are statistically significant against random false alarms and noise in the signals.

Traditionally *Scan Statistics* is used to detect a cluster of events in spatial data (Glaz and Balakrishnan 1999; Glaz et al. 2001; Balakrishnan and Koutras 2001; Fu and Lou 2003). The most commonly used scan statistic is the maximum number of cases within a fixed size window that scans through the study area. Based on scan statistics, a generalized likelihood ratio test has been developed to test the null hypothesis that all signals are uniformly distributed in the area (the case of no cluster) (Naus 1966). Other scan statistics and related likelihood based tests for localized temporal or spatial clustering have been developed, often using a range of fixed window sizes or a range of fixed number of cases (Kulldorff and Nagarwalla 1995; Naus and Wallenstein 2004; Dembo and Karlin 1992; Su et al. 2001). Scan statistics methods have also been developed under the Bayesian framework (Lawson 1995; Gangnon and Clayton 2000; Denison and Holmes 2001; Gangnon and Clayton 2003).

Scan statistics procedures are very effective in detecting a single significant cluster, and they also have had some success in detecting multiple clusters of fixed sizes. But they are not particularly suitable for detecting multiple clusters of varying sizes (cf., e.g., Xie, Sun and Naus (2009)). In recent years, several model based procedures have been developed to detect multiple clusters of varying sizes. For instance, Demattei et al. (2006, 2007) proposed a stepwise regression model combined with model selection procedures to locate and determine the number of clusters. This method relies on a weighted least square formulation, although the response variable (gaps between incidents) is typically non-Gaussian. Xie et al. (2009) developed a latent cluster model for temporal data which utilizes standard likelihood inference for detecting multiple clusters. Sun (2008) extended the approach to spatial data. Based on likelihood inference, the latent model approach is more efficient in detecting clusters of varying sizes than the weighted least squares approaches. Another class of statistical methods for detecting a cluster of events are the "disease mapping" type of approaches both in Bayesian and frequentist paradigms; see, e.g., Lawson (1995), Waller et al. (1997), Ghosh et al. (1999), Knorr-Held and RaBer (2000), and Diggle et al. (2005). These approaches consider intensity functions (i.e., local density estimates of intensities) and they do not directly make inference about clusters. To detect clusters, the "disease mapping" approaches rely on a certain choice of threshold, which is often subjective.

In this paper, we further develop and expand the latent model method by Xie et al. (2009) and Sun (2008) for detecting nuclear sources and design a mobile sensor network for surveillance of nuclear materials in a metropolitan area. The network consists of a large number of moving vehicles, such as taxicabs, police cars, fire trucks and other participating private and public vehicles, on which nuclear sensors and Global Position System (GPS) tracking devices are installed. Real time sensor readings are processed at a central surveillance center, where the data is analyzed to detect significant clusters of signals which might indicate the locations of nuclear sources. Our method uses likelihood inference to detect multiple clusters simultaneously in a region. It is suitable to analyze signals from the proposed mobile sensor framework and reduce false alarms. It can also filter out known background and sources.

Note that it is often difficult to model accurately the movements of the nuclear sources and the sensors, since it is difficult to model or track the destination and intention of each driver or terrorist. In our data analysis, we do not consider motion models for either the nuclear sources (i.e., terrorist movement) or the sensors (i.e. the movements of taxicabs, etc.), and the statistical analysis is performed based on data collected at each fixed time point. A sequence of data analyses in consecutive times (e.g., every 30 seconds or every one minute, say) can then form a dynamic surveillance — just like the movement in a movie which is formed from stationary film frames. This approach is robust against potential model misspecification of the movements of the nuclear sources and the sensors. It is applicable to detect both stationary and moving sources.

The rest of the paper is arranged as follows. Section 2 begins with a prototype of a mobile sensor network and considers related models for nuclear intensity, sensor reading and detection. Section 3 provides a latent statistical model, a likelihood-inference-based methodology and an EM/MCMC algorithm to detect clusters and make inference based on information (sensor signals and their locations) collected at each fixed time point. Section 4 considers two additional developments in incorporation of background information and heteroscedastic detection. Section 5 describes simulation studies on several practical scenarios, where the detection powers of the network are estimated under various sets of parameters. In this section, in order to generate data that simulate city traffic, we introduce a movement model with a constraint that vehicles can only travel on street grids, though the data analysis is based on the development in Sections 3 and 4 and does not depend on the movement model. Section 6 concludes the paper with discussion and future research directions.

2. NUCLEAR DETECTION USING A MOBILE SENSOR NETWORK

2.1 Mobile Sensor Network Prototype

Proactive monitoring and detection via pervasive surveillance is crucial to detect and thwart malicious attacks in major cities. Here we propose a prototype of a mobile sensor network as follows:

- Inexpensive nuclear sensors and Global Position System (GPS) tracking devices are mounted on a large number of vehicles such as police cars, taxicabs, fire trucks, buses, and other participating vehicles.

- The sensors and GPS devices constantly send detection signals and location information to a central command center. These signals are marked onto a map of a metropolitan area under surveillance.

- Real time analysis is performed at the command center using sophisticated statistical algorithms including the latent modeling method discussed in this paper to detect and pinpoint nuclear sources.

- Sensors and the control center are periodically serviced and calibrated to ensure validity and accuracy.

The advantage of a mobile network over a static surveillance network is multifold. First, due to mobility, the failure of a portion of the sensors will not have significant impact on the surveillance coverage, while any sensor failure in a static network will create blind spots. Second, the locations of blind spots in a static network are unknown since a failed sensor often would send in negative signals. On the other hand, the blind spots of a mobile network (where vehicles can not reach) can be covered with a small number of static sensors. Third, mobile sensors may not need to be highly reliable and of large range, due to their mobility, while static sensors would need to be more reliable and powerful to maintain the coverage. Since our sensors are vehicle mounted, their size and power requirements may be not high. Hence, inexpensive sensors could be deployed. Fourth, mobile sensors can be inspected and calibrated during vehicle inspection and routine maintenance, while static sensors can only be inspected by visiting their physical locations. Fifth, it is almost impossible to tamper with a mobile sensor network while a static network can be an easy target.

Due to these attractive characteristics, there have been many studies and applications of the sensor networks in military and civil applications including surveillance, smart homes, remote environment monitoring (Akyildiz et al.2002a, 2002b). Much of the research is devoted to sensor placement, sensor reorganization and communications. In the area of radiation detection, the idea of using a network of mobile sensors has been adopted and tested by the Radiation Laboratory at Purdue University (2008). They used a network of cell phones with GPS capabilities to detect and track radiation. The noise and false positive detection problems were tackled by setting and tuning solid state devices. A multi-sensor nuclear threat detection problem was studied in Hochbaum (2009) using a combinatorial network flow algorithm, in which the main research objective is the development of a fast algorithm for detection with mobile sensors (such as taxicabs) instead of the efficient use of data information. In this paper, we use a latent modeling method for source detection. This statistical approach is effective at detecting true signals against random errors and thus reducing false alarms.

2.2 Nuclear Intensity, Sensor Reading and Detection Models

In this paper we consider the source as a portable nuclear device transported by an individual via vehicles or bags (FEMA 2008). As nuclear radiation emits from a source, its total energy remains constant due to the Energy Conservation Law. Following Wein et al. (2006) and many others, we assume that radiation travels in spherical waves. Let $z(r)$ denote the intensity at distance r from the source with a total energy of $4\pi r^2 z(r)$. Hence radiation intensity z decreases by the inverse square of distance r (Wein et al. 2006):

$$z(r) = c/r^2,$$

where c is a constant factor related to the total energy of the source.

Suppose a nuclear detection device is triggered when the radiation intensity it receives exceeds a certain threshold. Let S denote the reading status of the nuclear detector (sensor) with a value of 1 for a positive reading and 0 otherwise. We describe S with a threshold model:

$$S = \mathbf{1}_{\{z(r)\geq d\}} = \mathbf{1}_{\{c/r^2\geq d\}}, \tag{1}$$

where d is a threshold for detection and $\mathbf{1}_{\{.\}}$ is the indicator function. That is, if the intensity $z(r)$ at the sensor location is greater than or equal to the threshold d, the sensor reports a detection; otherwise the sensor reports a negative reading.

In the case when there are multiple nuclear sources, their energy levels and positions will jointly determine the reading status of a sensor. Let Γ be the number of sources, c_γ be the energy factor of the γth source. When sources have different energy spectra, they will activate a sensor independently. In this case, the threshold model is

$$S = \mathbf{1}_{\{\max_{\gamma\in\{1,..\Gamma\}} z_\gamma(r_\gamma)\geq d\}} = \mathbf{1}_{\{\max_{\gamma\in\{1,..\Gamma\}} c_\gamma/r_\gamma^2\geq d\}}, \tag{2}$$

where r_γ is the distance of the γ-th source to the sensor and $z_\gamma(r_\gamma)$ is its corresponding intensity. When the sources assume the same energy spectrum, intensities from all sources at the sensor location are aggregated: $z_{total} = \sum_{\gamma=1}^{\Gamma} c_\gamma/r_\gamma^2$. Then the reading S can be determined by:

$$S = \mathbf{1}_{\{z_{total}\geq d\}} = \mathbf{1}_{\{\sum_{\gamma=1}^{\Gamma} c_\gamma/r_\gamma^2\geq d\}}. \tag{3}$$

As with any detection device, a nuclear sensor may not be 100% accurate. A sensor might display positive readings when there is no nuclear source nearby (false positive), or fail to detect a real source (false negative). This is unavoidable even for expensive and highly accurate sensors and in particular for our study with a massive number of inexpensive sensors. The sensor errors can be from the variability in the manufacturing process, routine wear and tear, missing scheduled maintenance and calibrations, and other malfunctions. In addition, random traces of weak environmental nuclear signals can also trigger false alerts. Furthermore, wireless signals from a mobile sensor to the control center may incur transmission errors. We use two parameters, *sensitivity* η and *specificity* ζ, to measure the average performance of a sensor device.

Specially, let D be the binary indicator of a sensor signal, with a value of 1 for a positive detection and 0 otherwise. We have

$$\eta = P(D = 1|S = 1) \quad \text{and} \quad \zeta = P(D = 0|S = 0). \tag{4}$$

The terms $1 - \eta$ and $1 - \zeta$ correspond to false negative rate (FNR) and false positive rate (FPR), which are quality characteristics of a sensor. Then the probability of a positive detection is:

$$\begin{aligned} P(D = 1) &= P(D = 1|S = 1)P(S = 1) + P(D = 1|S = 0)P(S = 0) \\ &= \eta P(S = 1) + (1 - \zeta)(1 - P(S = 1)) \\ &= (1 - \zeta) + (\zeta + \eta - 1)P(S = 1). \end{aligned} \tag{5}$$

Under the perfect scenario where both η and ζ are 1, $P(D = 1)$ is the same as $P(S = 1)$.

2.3 Detection of Nuclear Sources and Statistical Problem of Cluster Detection

The sensor reading models in Section 2.2 relate to the cluster detection problem in statistics. In particular, with a single source, the threshold model (1) can be expressed as $S = \mathbf{1}_{\{I\}}$, where $I = \{r \leq (c/d)^{1/2}\}$ is a sphere, or a circle on a 2-dimensional map, centering at the nuclear source and with a radius $(c/d)^{1/2}$. From (4) the ratio of the probabilities of a positive reading inside and outside the set I is

$$\alpha = \frac{P(D = 1|I)}{P(D = 1|\bar{I})} = \frac{P(D = 1|S = 1)}{P(D = 1|S = 0)} = \frac{\eta}{1 - \zeta}. \tag{6}$$

It is easy to check that $\alpha > 1$ when $\eta + \zeta > 1$ (which is a condition satisfied by almost all commercially available detection devices). Formula (6) states that a sensor is α times more likely to report a positive signal ($D = 1$) inside I than outside I. This statement matches exactly the definition of a spatial cluster defined as an area within which an incident of interest is more likely to happen (i.e., with a higher probability of happening per squared unit) than outside of the area. Here, an incident of interest is an alert signal with $D = 1$. Thus, the sensor reading models presented in this section link us to a statistical cluster detection problem. A similar connection can be explored in the presence of multiple nuclear sources at different locations.

3. LATENT MODELING PROCEDURES FOR SOURCE DETECTION

Assume that there are k stationary or moving nuclear sources in a given two dimensional rectangular region, $\mathcal{I} = (0, L) \times (0, W)$, where L is the length and W the width. Suppose the jth source ($j = 1, ..., k$) is located at $\mathbf{o}_j$ and has a strength that can be effectively detected by a sensor within a distance of r_j. If the distance of a sensor to $\mathbf{o}_j$ is less than r_j, the sensor would generate a positive reading (subject to false negative errors). Denote by cluster I_j the circle centering at $\mathbf{o}_j$ with radius r_j and define $\mathbf{O} = (\mathbf{o}_1, ..., \mathbf{o}_k)$, $\mathbf{r} = (r_1, ..., r_k)$ as the collection of locations and sizes (radii) of clusters. For simplicity, we assume in this article that these k clusters are not overlapping; see Section 6 and Chapter 5 of Cheng (2010) for discussions of overlapping cases.

We treat $\mathbf{O}$ and $\mathbf{r}$ as latent random variables and assume that cluster centers follow a density function $\psi_{\mathbf{o}}(\cdot; \lambda_o)$ and radii a density function $\psi_r(\cdot; \lambda_r)$. Here, λ_o and λ_r are unknown parameters jointly denoted by $\lambda = (\lambda_o, \lambda_r)$, while $\psi_{\mathbf{o}}$ and ψ_r are known density functions. We use a uniform distribution on $\mathcal{I}$ for $\psi_{\mathbf{o}}(\cdot; \lambda_o)$ and a truncated exponential distribution for $\psi_r(\cdot; \lambda_r)$. Other choices for $\psi_r(\cdot; \lambda_r)$ are among inverse Gamma or log-normal distributions (see, e.g., Sun 2008). Let Ω_k be the set of ($\mathbf{O}$,$\mathbf{r}$) such that the k clusters specified by ($\mathbf{O}$,$\mathbf{r}$) are non-overlapping. Given that there are k non-overlapping clusters, the joint conditional likelihood function of $(\mathbf{O}, \mathbf{r})$ is

$$f_\lambda(\mathbf{O}, \mathbf{r}|k) = \frac{\prod_{j=1}^{k}\{\psi_o(\mathbf{o}_j; \lambda_o)\psi_r(r_j; \lambda_r)\}\mathbf{1}_{\{(\mathbf{O},\mathbf{r})\in\Omega_k\}}}{C_{\lambda,k}}, \tag{7}$$

where $C_{\lambda,k}$ is the normalization constant for the truncated density.

Assume that there are n sensors with positive readings. Let $\mathbf{y}_i$ be the location of the ith sensor ($i = 1, ..., n$) when it makes a positive reading and Y be the collection of these locations, $Y = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$. We assume that, given the locations $\mathbf{O}$ and the sizes (radii) $\mathbf{r}$ of the clusters corresponding to the k sources, the $\mathbf{y}_i$'s are independently identically distributed (iid) samples from a piecewise uniform distribution whose density function is

$$f_\alpha(\mathbf{y}|\mathbf{O},\mathbf{r},k) = \begin{cases} \frac{\alpha_1}{A+\sum_{j=1}^{k}(\alpha_j-1)A_j}, & \text{if } \mathbf{y} \in I_1 \\ \quad \cdots\cdots & \\ \frac{\alpha_k}{A+\sum_{j=1}^{k}(\alpha_j-1)A_j}, & \text{if } \mathbf{y} \in I_k \\ \frac{1}{A+\sum_{j=1}^{k}(\alpha_j-1)A_j}, & \text{if } \mathbf{y} \notin \bigcup_{j=1}^{k} I_j. \end{cases} \tag{8}$$

Here A=Area($\mathcal{I}$) is the total area of $\mathcal{I}$, A_j=Area($I_j \bigcap I$) is area of the jth cluster inside $\mathcal{I}$, and α_j is the relative density of positive readings in the jth cluster with respect to the background, defined in (6). Write $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)$; the α_j's may be different since sensors may have different sensitivity with respect to different type of sources. We treat them as unknown parameters to be estimated. Note that when $\alpha_1 = ... = \alpha_k = 1$ in (8), $\mathbf{y}_i$ follows a uniform distribution within $\mathcal{I}$, implying that there is no source in the region.

In the case of single cluster with $k = 1$, (8) is a one-step (single-disk) uniform distribution function, which is the same as the underlying inference model used in many spatial scan statistics approaches (Glaz et al. 2001). The formulation of model (8) extends the case to any given number of clusters, and it allows flexibility in estimating multiple clusters with different sizes. Also, for each given k clusters, the proposed latent model can be expressed using a hierarchical structure,

$$\begin{aligned} &\text{Level 1:} \quad \mathbf{y}_i|\mathbf{O},\mathbf{r},\alpha_1,...,\alpha_k, \sim \text{Model (8), for } i = 1,2,\ldots,n \\ &\text{Level 2:} \quad \mathbf{o}_j|\lambda_o \sim \psi_{\mathbf{o}}(\cdot) \text{ and } r_j|\lambda_r \sim \psi_r(\cdot), \text{ for } j = 1,2,\ldots,k \text{ and } (\mathbf{O},\mathbf{r}) \in \Omega_k. \end{aligned} \tag{9}$$

It is closely related to the following Bayesian hierarchical model, in which we further assume the unknown parameters α and λ are random and place priors on them,

$$\begin{aligned} &\text{Level 1:} \quad \mathbf{y}_i|\mathbf{O},\mathbf{r},\alpha_1,...,\alpha_k \sim \text{Model (8), for } i = 1,2,\ldots,n \\ &\text{Level 2:} \quad \mathbf{o}_j|\lambda_o \sim \psi_{\mathbf{o}}(\cdot),\ r_j|\lambda_r \sim \psi_r(\cdot), \text{ and } \alpha_j \sim \pi_\alpha(\cdot) \text{ for } j = 1,2,\ldots,k \text{ and } (\mathbf{O},\mathbf{r}) \in \Omega_k \\ &\text{Level 3:} \quad \lambda_o \sim \pi_{\lambda_o}(\cdot) \text{ and } \lambda_r \sim \pi_{\lambda_r}(\cdot), \end{aligned} \tag{10}$$

where π_α, π_{λ_o} and π_{λ_r} are the priors for α_j, λ_o and λ_r, respectively. If non-informative flat priors are used for these priors, the results from the likelihood inference using hierarchical model (9) are often the same as the ones using the Bayesian hierarchical model (10). Here, to simplify our inference, we first treat the number of clusters k as given and known; Section 3.2.2 provides modified AIC and BIC criteria to determine k when it is unknown. Of course, in the Bayesian approach, one can alternatively treat k as an unknown random integer and assume a Dirichlet prior on it (add to Level 3 model). However, as mentioned in Rodriguez et al. (2008), the outcome relies on the particular choice of the Dirichlet prior and caution should be applied. Finally, there are also other formulations for the same detection problem, e.g., those using Poisson assumptions or Bayesian state-space models. We use the current formulation of a step uniform function in order to highlight the interpretation of the parameters α_i's and to translate the nuclear detection problem into a formal statistical testing problem without involving thresholding.

From (8), the joint distribution function of $\mathbf{y}_1, ..., \mathbf{y}_n$, conditional on $\mathbf{O}$, $\mathbf{r}$ and $(\mathbf{O}, \mathbf{r}) \in \Omega_k$, is:

$$f_\alpha(Y|\mathbf{O}, \mathbf{r}, k) = \prod_{i=1}^{n} f_\alpha(\mathbf{y}_i|\mathbf{O}, \mathbf{r}, k) = \exp\Big\{\sum_{j=1}^{k} Z_j \log \alpha_j - n \log \Big[A + \sum_{j=1}^{k}(\alpha_j - 1)A_j\Big]\Big\}, \tag{11}$$

where $Z_j = Z_j(\mathbf{y}, \mathbf{O}, \mathbf{r}) = \sum_{i=1}^{n} \mathbf{1}_{\{\mathbf{y}_i \in I_j\}}$ is the number of positive readings inside the jth cluster. The observed likelihood function of the above model, for a fixed k, is

$$l(\theta, k) = f_\theta(Y, k) = \int ... \int f_\theta(Y, \mathbf{O}, \mathbf{r}, k) d\mathbf{O} d\mathbf{r}, \tag{12}$$

where $\theta = (\alpha, \lambda)$ and the integrand is the joint density of the complete data $(Y, \mathbf{O}, \mathbf{r}, k)$:

$$\begin{aligned} f_\theta(Y, \mathbf{O}, \mathbf{r}, k) &= f_\alpha(Y|\mathbf{O}, \mathbf{r}, k) f_\lambda(\mathbf{O}, \mathbf{r}, k) \\ &= \exp\Big\{\sum_{j=1}^{k} Z_j \log \alpha_j - n \log \Big[A + \sum_{j=1}^{k}(\alpha_j - 1)A_j\Big]\Big\} \prod_{j=1}^{k} \{\psi_o(\mathbf{o_j})\psi_r(r_j)\} \mathbf{1}_{\{(\mathbf{O}, \mathbf{r}) \in \Omega_k\}}. \end{aligned}$$

3.1 Estimation of Model Parameters and Locations of Clusters

3.1.1 Monte-Carlo EM algorithm for model estimation

Since the log-likelihood $l(\theta, k)$ from (12) involves multiple integrations, it is difficult to compute its value and its first and second derivatives directly, in order to obtain the maximum likelihood estimator (MLE) of the parameter θ. We instead develop a Monte Carlo Expectation-Maximization (MC-EM) algorithm (cf., e.g., Tanner 1993, Section 4.5) where we treat $(Y, \mathbf{O}, \mathbf{r}, k)$ as the complete responses and (Y, k) as the observed responses. The steps of the algorithm are as follows.

Step 0. Select a set of starting parameter values $\theta^{(0)} = (\alpha^{(0)}, \lambda^{(0)})$.

Step 1. (E-step). For $s = 0, 1, 2, \ldots$, calculate the conditional expectation of the complete log-likelihood function, given the observed observations and $\theta = \theta^{(s)}$:

$$Q(\theta|\theta^{(s)}) = Q_1(\alpha|\alpha^{(s)}) + Q_2(\lambda|\lambda^{(s)}),$$

where

$$Q_1(\alpha|\alpha^{(\mathbf{s})}) = \sum_{j=1}^{k} E\{Z_j|Y, k, \alpha^{(\mathbf{s})}\} \log \alpha_j - nE\{\log[A + \sum_{j=1}^{k}(\alpha_j - 1)A_j]|Y, k, \alpha^{(\mathbf{s})}\}, \quad (13)$$

$$Q_2(\lambda|\lambda^{(\mathbf{s})}) = \sum_{j=1}^{k} E\{\log \psi_o(o_j)|Y, k, \lambda^{(\mathbf{s})}\} + \sum_{j=1}^{k} E\{\log \psi_r(r_j)|Y, k, \lambda^{(\mathbf{s})}\}.$$

Step 2 (M-step). For $s = 0, 1, 2, \ldots$, update the parameter estimates: $\theta^{(s+1)} = (\alpha^{(s+1)}, \lambda^{(s+1)})'$, by maximizing the $Q_1(\alpha|\alpha^{(\mathbf{s})})$ and $Q_2(\lambda|\lambda^{(\mathbf{s})})$ functions: $\alpha^{(s+1)} = \operatorname{argmax} Q_1(\alpha|\alpha^{(\mathbf{s})})$ and $\lambda^{(s+1)} = \operatorname{argmax} Q_2(\lambda|\lambda^{(\mathbf{s})})$.

Step 3. Repeat Steps 2 and 3 until $\|\theta^{(s+1)} - \theta^{(s)}\|$ is very small.

In the E-step of the EM algorithm, the conditional expectations do not usually have explicit form. We use Monte Carlo estimation Gibbs sampler in association with importance sampling. The detailed sampling steps are in the Appendix.

3.1.2 Identification of cluster regions

In nuclear detection, a primary goal is to identify regions of the sources determined by the centers and radii: $\{(\mathbf{o_j}, r_j) : j = 1, .., k\}$. Their conditional expectations given (Y, k) (similar to "posterior means" in the context of Bayesian statistics) are $E\{\mathbf{o_j}|Y, k\}|_{\theta=\hat{\theta}}$ and $E\{r_j|Y, k\}|_{\theta=\hat{\theta}}$. They can be estimated by their sample means $\frac{1}{M}\sum_* \mathbf{o}_j^*$ and $\frac{1}{M}\sum_* r_j^*$, respectively. Here, $\sum_*$ is the summation over the M sets of Gibbs samples in the last iteration step in the EM algorithm. An alternative approach is to use the medians of the M sets of $\mathbf{o}_j^*$ and r_j^*, as the point estimators of the unknown center $\mathbf{o_j}$ and radius r_j, respectively. Since their distribution may not be symmetric, using medians may provide more accurate estimators.

3.2 Likelihood Ratio Test and Determination of the Unknown Number of Clusters

3.2.1 Likelihood ratio test of significant clusters

For a fixed k, we can obtain k clusters from the estimation procedures. An interesting problem is to check whether any of the k sources is significant, or equivalently, to determine whether any of the detected sources is due to random noise in the context of nuclear detection. In particular, we test a hypothesis about the parameter set α, i.e., H_0: $\alpha_1 = ... = \alpha_k = 1$ versus H_1: at least one $\alpha_j \neq 1$. We propose to use the log likelihood ratio test statistic

$$\begin{aligned} R &= \log \left\{ \frac{\max_{H_1 \cup H_0} f_\theta(Y, k)}{\max_{H_0} f_\theta(Y, k)} \right\} \\ &= \log \int\int f_{\hat{\alpha}}(\mathbf{y}|\mathbf{O}, \mathbf{r}, k) f_{\hat{\lambda}}(\mathbf{O}, \mathbf{r}|k) d\mathbf{O} d\mathbf{r} + \log C_{\lambda,k} + n \log A, \end{aligned}$$

where $\hat{\theta} = (\hat{\alpha}, \hat{\lambda})^T$ are non-restricted MLEs (under $H_1 \bigcup H_0$) estimated from the aforementioned MC-EM algorithm. This test statistic is difficult to evaluate and we again utilize a Monte Carlo method. In particular, we note that $f_{\hat{\lambda}}(\mathbf{O}, \mathbf{r}|k)/f_{\hat{\lambda}}(\mathbf{O}, \mathbf{r}) \leq 1/C_{\lambda,k} \equiv c_b$ and, given $\hat{\lambda}$, $f_{\hat{\lambda}}(\mathbf{O}, \mathbf{r})$ is easy to simulate from. We use a rejection sampling approach to simulate M sets of samples, say, $(\mathbf{O}^{(j)}, r^{(j)}), j = 1, .., M$, from $f_{\hat{\lambda}}(\mathbf{O}, \mathbf{r}|k)$, with $f_{\hat{\lambda}}(\mathbf{O}, \mathbf{r})$ being the candidate distribution. The acceptance rate of this rejection sampling method is $1/c_b = C_{\lambda,k}$, and its empirical acceptance rate, say r, can be used to estimate $C_{\lambda,k}$. Thus, we approximate R by

$$R^{**} = \log \Big[\frac{1}{M} \sum_{j=1}^{M} f_{\hat{\alpha}}(\mathbf{y}|\mathbf{O}^{(j)}, r^{(j)}, k) \Big] + \log r + n \log A. \tag{14}$$

Note that, the parameters $\lambda = (\lambda_0, \lambda_1)$ are nuisance parameters in the test and they only exist under the alternative hypothesis H_1. In this case, as reported in Davies (1977, 1987), the likelihood test statistic does not follow the asymptotic chi-squared distribution, and the usual chi-squared test (also called the large sample test) is no longer valid. Our numerical study later in Section 5 finds that the actual testing size is far below 5 percent if we blindly use the chi-squared test of size 5 percent. To overcome this difficulty, Davies (1987) suggested use of a Monte-Carlo simulation approach. We consider a simulation based Monte Carlo testing approach (Dwass 1957). Specifically, we sample L sets of k clusters under the null hypothesis (no true sources in the study region) and compute for each set the test statistics according to (14). When L is large, the empirical distribution of these values provides a good approximation to the theoretical distribution of the test statistics R^{**} under the null hypothesis. We thus obtain the critical value (for example, 5-th percentile) from the L values for our simulation-based Monte Carlo test.

3.2.2 Determination of the unknown number of clusters

The previous estimation and testing procedures are for a given number of sources k. In reality, we do not know how many nuclear sources exist in a study region, and we need to determine the number of clusters from the observed data. The determination of the number of clusters is a task of model selection, and the AIC criterion (Akaike 1974) and BIC criterion (Schwarz 1978) are two approaches that have been widely used for such practice. However, when a model selection problem involves missing (or latent) variables, a direct application of the usual AIC or BIC criterion can be problematic; see, Claeskens and Consentino (2008). Our numerical study (results not reported in the paper) also confirms such observations. Claeskens and Consentino (2008) proposed a modified AIC criterion and demonstrated that their modified AIC criterion can overcome the problem and has nice performance in the case involving missing (or latent) variables. In particular, using the Kullback-Leibler distance to measure the distance between the true data generating density and the model density used for describing the data, they derived a new criterion based on the Q function in the EM algorithm. This requires no additional computations since the quantity is available from EM algorithm in the model estimation step.

Following their approach, we use the modified AIC and BIC criteria as:

$$\text{AIC}(k)_{mod} = -2Q_1(\hat{\alpha}|\hat{\alpha}) + 2k, \tag{15}$$

$$\text{BIC}(k)_{mod} = -2Q_1(\hat{\alpha}|\hat{\alpha}) + k\log(n), \tag{16}$$

where $Q_1(\hat{\alpha}|\hat{\alpha})$ is the Q function in (13) at the value of the MLE $\hat{\alpha}$ from the EM algorithm. Given a data set, several competing models may be ranked according to their AIC or BIC values.

3.3 Summary of Cluster Detection Procedure

To conclude this section, we provide a summary of the approach for detecting clusters. Denote by $\mathcal{K}$ a preselected set of k's, which is computationally manageable but large enough to cover all potential choices of the correct number of clusters. For each k in $\mathcal{K}$, we apply the MC-EM algorithm to get the parameter estimates and conduct a hypothesis test to check whether clusters are significant. With the modified AIC or BIC values, we determine the number of clusters k. For the k chosen by AIC or BIC, the source locations and impact ranges are estimated.

4. ADDITIONAL CONSIDERATIONS IN THE LATENT MODELING METHOD

4.1 Incorporation of Background Information

In Section 3, for simplicity, we only used positive signals in our analysis and the information of negative sensors was not used. This procedure is valid, if we assume that the mobile sensors are uniformly distributed throughout the surveillance region and the background signal is also uniform. In this highly idealized case the negative sensor signals do not provide any extra information for cluster detection. However, the assumption of evenly distributed sensors may not be valid in many situations. For example, some parts of the city at certain times of a day may have more vehicles due to higher concentration of business and commerce activities. With more sensors in these areas, we are more likely to receive more positive readings, even if there is no elevated nuclear radiation in their surrounding areas. Indeed, when the sensors are not uniformly spreading through the surveillance region, the proportion of positive sensors in a local area, instead of the number of positives in the area, is a better measure of the nuclear signal. In our analysis, it is necessary to extend our approach to include the information of negative signals. This can be achieved by introducing a background function to the piecewise-uniform density function (8).

Let $B(\mathbf{y})$ denote the background information of sensors, which can be sensors' density such as the total number of mobile sensors in the surrounding area (e.g., the same block or the same segment of the street). The piecewise uniform density function (8) is now modified as

$$f_\alpha(\mathbf{y}|\mathbf{O}, \mathbf{r}, k) = \begin{cases} \frac{\alpha_1 B(\mathbf{y})}{\tilde{A}+\sum_{j=1}^k (\alpha_j - 1)\tilde{A}_j}, & \text{if } \mathbf{y} \in I_1 \\ \quad \cdots\cdots & \\ \frac{\alpha_k B(\mathbf{y})}{\tilde{A}+\sum_{j=1}^k (\alpha_j - 1)\tilde{A}_j}, & \text{if } \mathbf{y} \in I_k \\ \frac{B(\mathbf{y})}{\tilde{A}+\sum_{j=1}^k (\alpha_j - 1)\tilde{A}_j}, & \text{if } \mathbf{y} \notin \bigcup_{j=1}^k I_j, \end{cases} \tag{17}$$

where $\tilde{A} = \int_{\mathcal{I}} B(\mathbf{y}) d\mathbf{y}$, $\tilde{A}_j = \int_{\mathcal{I} \bigcap I_j} B(\mathbf{y}) d\mathbf{y}$ are the areas weighted by the background density function. Note that the likelihood of having a positive signal at location $\mathbf{y}$ is proportional to the background density function $B(\mathbf{y})$, across all clusters and the background. When $B(\mathbf{y}) \propto 1$ (uniform/constant background), model (17) reduces to (8). The statistical analysis and the cluster detection steps based on this model are the same as those in Section 3 for model (8), except that A is replaced by $\tilde{A}$ and A_j is replaced by $\tilde{A}_j$, respectively. Details are omitted to save space.

In addition to incorporate the instant information of the distribution of mobile sensors (taxicabs, etc.), the background function $B(\mathbf{y})$ introduced in (17) can be used to model complications of other types of non-uniform background in the region of interest. For instance, we can use the function $B(\mathbf{y})$ to filter out known benign sources of nuclear materials, such as hospitals or pre-registered shipments of nuclear materials for medical or other uses. We could learn a radiation profile at each given time of day from past (historical) data in the region of interest, and feed this information to our surveillance to cut down unnecessary false alarms. We could also calculate and compute a $B(\mathbf{y})$ function for a pre-registered delivery of radiation materials along its route at each given time, and add this one-time information to the surveillance system.

4.2 Detection with Heteroscedastic Sensors

So far, we have considered sensors of the same characteristics such as error rates, detection ranges, etc. In practice, we may use different types of sensors in a mobile sensor network with different characteristics such as error rates, detection ranges, etc. For example, conversations with law enforcement officials suggested a possible combination of a large number of lower quality sensors on a fleet of taxicabs and a limited number of higher quality sensors on police vehicles. If there are m types of sensors, the impact range of the jth nuclear source on the ith type sensors is assumed to be

$$r_j^{(i)} = c^{(i)} r_j^{(1)}, \quad i = 1, \ldots, m; \quad j = 1, \ldots, k,$$

where $r_j^{(1)}$ is the impact range for sensors of a reference type, $c^{(i)}$ is the ratio of the source impact range for the ith type sensors over that for the reference type sensors. We assume that $c^{(i)}$ is given and can be obtained from manufacturer's specification. The event data from m types of sensors is $Y = (Y^{(1)}, \ldots, Y^{(m)})$, where $Y^{(i)} = (\mathbf{y}_1^{(i)}, \ldots, \mathbf{y}_{n_i}^{(i)})$ is a $2 \times n_i$ matrix denoting the set of n_i event data from the ith type sensors. For the ith type sensors, with samples from a piecewise uniform distribution, the density function from (8) is modified as

$$f_\theta^{(i)}(\mathbf{y}|\mathbf{O}, \mathbf{r}, k) = \begin{cases} \frac{\alpha_1^{(i)}}{A + \sum_{j=1}^k (\alpha_j^{(i)} - 1) A_j^{(i)}}, & \text{if } \mathbf{y} \in I_1^{(i)} \\ \quad \cdots\cdots & \\ \frac{\alpha_k^{(i)}}{A + \sum_{j=1}^k (\alpha_j^{(i)} - 1) A_j^{(i)}}, & \text{if } \mathbf{y} \in I_k^{(i)} \\ \frac{1}{A + \sum_{j=1}^k (\alpha_j^{(i)} - 1) A_j^{(i)}}, & \text{if } \mathbf{y} \notin \bigcup_{j=1}^k I_j^{(i)}, \end{cases} \tag{18}$$

where $\theta = (\alpha; \lambda) = (\alpha^{(1)}, \dots, \alpha^{(m)}; \lambda)$ with $\alpha^{(i)} = (\alpha_1^{(i)}, \dots, \alpha_k^{(i)}), \lambda = (\lambda_o, \lambda_r), A_j^{(i)} = \pi(r_j^{(i)})^2$ is the source impact area of the jth cluster for the ith sensor. Hence for each type sensor, there is a set of parameters as in the model (8).

The conditional joint density function of Y given $\mathbf{O}, \mathbf{r}$ and k is

$$\begin{aligned} f_\theta(Y|\mathbf{O}, \mathbf{r}, k) &= \prod_{i=1}^{m} f_\theta(Y^{(i)}|\mathbf{O}, \mathbf{r}, k) \\ &= \exp\Big\{ \sum_{i=1}^{m} \Big[\sum_{j=1}^{k} Z_j^{(i)} \log \alpha_j^{(i)} - n_i \log(A + \sum_{j=1}^{k} (\alpha_j^{(i)} - 1) A_j^{(i)}) \Big] \Big\}, \end{aligned}$$

where $Z_j^{(i)} = Z_j^{(i)}(\mathbf{O}, \mathbf{r}, k) = \sum_{l=1}^{n_i} \mathbf{1}_{\{\mathbf{y}_l^{(i)} \in I_j^{(i)}\}}$ is the number of positive readings from the ith type sensors inside the jth cluster.

The cluster detection steps using this model are similar to those in Section 3 for the one-type sensor model (8).

5. SIMULATION STUDIES

In this section we present simulation studies to demonstrate that a mobile sensor network with analysis procedures using the proposed latent modeling method can effectively detect single and multiple nuclear sources. Intuitively, factors such as range, error rates and number of sensors will directly affect how likely nuclear sources can be detected. These factors form a set of network parameters and we study how they relate to the performance in detecting nuclear sources. In Section 5.1, we describe a Visual Basic graphical simulation tool that we used to generate our data. In Sections 5.2 and 5.3, we apply the cluster detection methods developed in Sections 3 and 4 to study the signal data simulated using the graphical tool under various scenarios.

5.1 Simulation Design and Evaluation

We designed a mobile sensor network in an area of similar size as downtown Manhattan in New York City. The study region was set to an area of 25 by 25 blocks. Each block was a square with side length of 200 which represents 200 feet in real distance. To achieve this, a street grid-based graphical tool was developed using Visual Basic Version 9.03 (Microsoft, Inc.) and utilized to simulate traffic patterns in a metropolitan area. The tool supports multiple control parameters such as numbers of streets in either horizontal or vertical direction, size of street blocks and types, quantities and speed

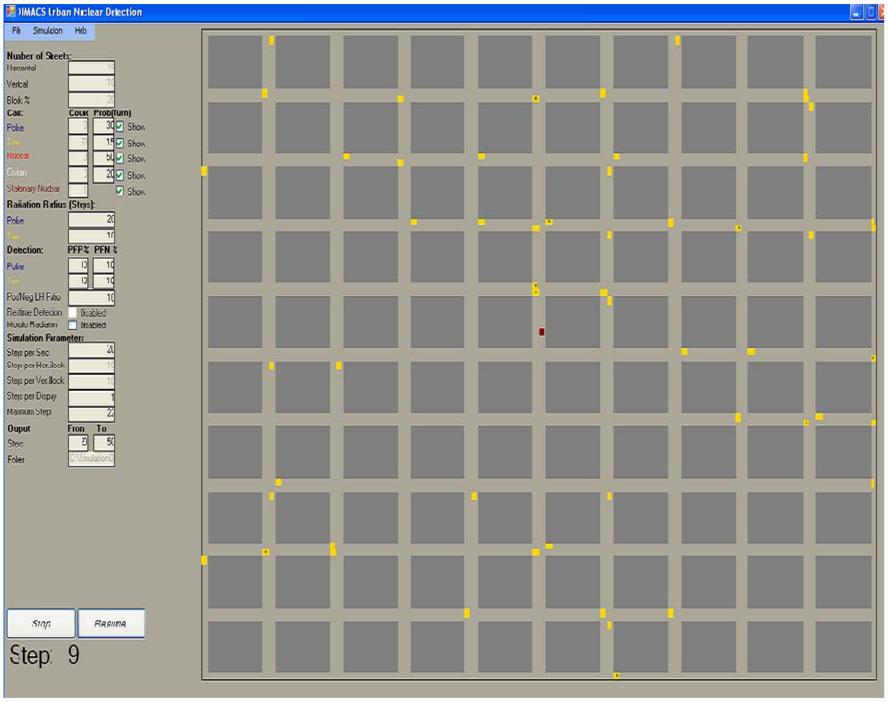


Figure 1: Snapshot of the simulation tool

of vehicles, a probability of turning upon reaching an intersection, etc. A snapshot of the simulation tool is in Figure 1. Since, by registered taxicabs alone, the number is more than thirteen thousand in New York City (New York City Taxi & Limousine Commission 2005), it seems reasonable to assume that there are 1,500 to 2,000 participating vehicles (such as taxicabs, police cars, fire trucks, security cars, etc.) in the downtown area. In the simulation tool, all vehicles are confined within the street grids.

To study the detection power of a sensor network, the set of network parameters was selected as: 1) source impact range: 150, 200 and 250 feet, roughly corresponding to 0.75, 1.0 and 1.25 of a block length, respectively; 2) sensitivity and specificity: $(\eta, \zeta) = (0.95, 0.95)$ and $(0.98, 0.98)$; 3) number of vehicles with sensors: from 500 to 3000 with increments of 500. Note that current technology does not achieve such high source impact range numbers, but we are seeking to understand the impact of potential future technology. For each of the $3 \times 2 \times 6 = 36$ sets of network parameters, we used the graphic tool to generate random positions (on the street grids) of participating vehicles and randomly placed a single or multiple sources anywhere in the study region. For multiple sources, we assumed that they had the same energy spectrum. Based on models (3) and (5), we assigned a probability of positive detection for each sensor and activated it accordingly. The positions of the sensors and their nuclear detection signals, either positive or negative, were collected. These geographic positions and the detection signals formed an observed data set. We then applied the proposed latent modeling approach to detect clusters. Such simulations were repeated 500 times in each setting of network parameters.

In our simulation study, two versions of powers were considered, along with the corresponding empirical sizes (type I errors). The hypothesis testing power (abbreviated as "testing power") was computed as the percentage of times that clusters are statistically significant at the significance level 5 percent for the test described in Section 3.2.1. We also computed the "detection power" from the percentage of times that the proposed algorithm correctly detects randomly placed nuclear sources. Here, a correct detection refers to the case that meets the following criteria: 1) detected cluster regions cover true nuclear sources; 2) clusters are statistically significant at the significance level 5 percent. In addition, to compute the empirical size (type I error) of the testing method, we repeated the same simulations for power computations but with no nuclear source placed in the study region.

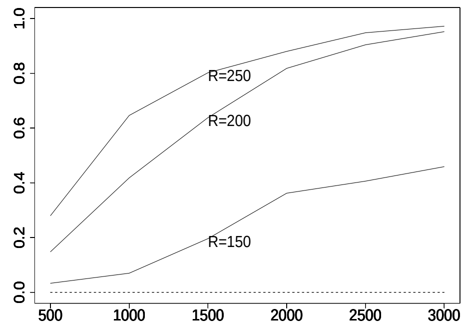
We started the simulations with a single set of readings over a period of time. Extension to multiple readings is discussed in Section 5.3.

5.2 Study I: Performance of Mobile Sensor Networks

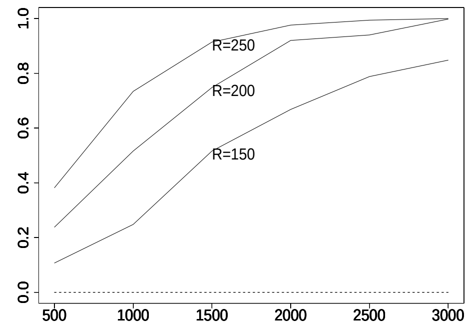
5.2.1 Powers for detecting one source

We first studied the effectiveness of the long source impact range (250 feet) against the shorter ones (200 feet and 150 feet) with a single random source in the region. Sensitivity and specificity were set at (0.95, 0.95). With level 5% tests, we plot the detection powers in Figure 2(a). The result shows that we need one thousand 250-foot sensors to achieve 60 percent detection power, while one thousand 200-foot sensors yield about 40 percent and 1,000 150-foot sensors less than 10 percent detection power. When we increased the number of sensors, the detection power of 200-foot sensors got close to that of 250-foot ones.

We repeated the same simulations with sensitivity and specificity of (0.98, 0.98) and obtained



(a) η=0.95, ζ=0.95



(b) η=0.98, ζ=0.98

Figure 2: Detection power of a mobile sensor network with one source. X-axis represents the numbers of sensors and Y-axis the detection power. (a) Three different sensor ranges with assumed η=0.95, ζ=0.95. (b) Three different sensor ranges with assumed η=0.98, ζ=0.98.

the power curves in Figure 2(b). This network with higher quality sensors achieved much higher power. For example, the detection power increased from 64% to 82% when both networks have 1,500 sensors of 200 feet range.

Table 1 presents detection performance under six network settings (from two error rates and three source impact ranges), with the number of sensors fixed at 1,500. We report detection powers, testing powers, and sizes of hypothesis testing using either the Monte Carlo method or the method of blindly applying the large sample chi-squared approximation (see Section 3.2.1). We observe that the detection powers from both methods are roughly the same. This is because the criterion to cover true nuclear sources for a correct detection is more stringent than the one for model parameters to be statistically significant. Without this coverage requirement, testing powers are always higher than detection powers. It also appears that the Monte Carlo method achieves a slightly higher testing power and is able to maintain Type 1 error around 5% in all settings. But the sizes obtained using the chi-squared approximation are far less than 5%.

5.2.2 Powers for detecting two sources

To demonstrate the performance of the mobile sensor network when there is more than one source, we studied the same network settings as in Section 5.2.1 but with two sources. Here, the two sources were placed randomly in the study region. We report the detection performance in Table 1.

Table 1: Summary of detection powers, hypothesis testing powers and sizes in various cases when true k=1 and 2 and the number of sensors is 1,500. The notation A_m is defined with an action A of either D=Detection or T=Testing, and a method m of either MC=the Monte Carlo test or LS=the Large sample test.

Simulation Setting			True k=1				True k=2			
(η, ζ)	Detection Range		D_{MC}	D_{LS}	T_{MC}	T_{LS}	D_{MC}	D_{LS}	T_{MC}	T_{LS}
(0.95,0.95)	150	Power	15.5%	15.5%	25.5%	19.8%	5.9%	5.9%	27.8%	22.7%
		Size	-	-	6.1%	0.0%	-	-	5.6%	0.0%
(0.95,0.95)	200	Power	63.8%	63.8%	67.2%	66.0%	56.0%	56.0%	77.7%	69.4%
		Size	-	-	4.4%	0.0%	-	-	6.0%	0.0%
(0.95,0.95)	250	Power	83.2%	83.2%	86.2%	85.0%	82.0%	82.0%	98.2%	96.6%
		Size	-	-	5.6%	0.0%	-	-	6.2%	0.0%
(0.98,0.98)	150	Power	35.5%	35.5%	44.1%	38.7%	27.0%	27.0%	61.5%	58.9%
		Size	-	-	5.5%	0.2%	-	-	5.8%	0.2%
(0.98,0.98)	200	Power	81.6%	81.6%	85.2%	83.2%	74.7%	74.7%	97.0%	93.8%
		Size	-	-	5.4%	0.2%	-	-	5.6%	0.0%
(0.98,0.98)	250	Power	97.4%	97.4%	97.8%	97.8%	96.4%	96.4%	100.0%	100.0%
		Size	-	-	5.2%	0.0%	-	-	5.2%	0.0%

As in the case of one source, we observe that the Monte Carlo method achieves slightly higher testing powers and is able to control the sizes (Type 1 error) of the hypothesis test around 5%, while the large sample method tends to produce more conservative results with sizes close to zero. Furthermore, when we have two clusters to detect, the detection power is lower than that for only one cluster under the same network setting while the testing power is higher in contrast.

5.2.3 Power for detection using heteroscedastic type sensors

In what follows, we selected the setting of 1,500 sensors with detection range of 200 feet and 95% for both η and ζ as a benchmark in further studies. This setting achieved detection powers of 64% in one cluster and 56% in two clusters.

In addition to this benchmark network setting, we added a small number (50, 75, 100) of higher quality sensors with both sensitivity and specificity at 0.98. The source impact ranges of the new sensors were set at 240, 300 and 360 feet.

The detection results of the combination of the two types of sensors for one cluster are listed in Table 2. We observed a significant improvement in detection power as compared to the results in Table 1. For example, with 50 additional high quality sensors of 240-foot detection range (3.3% increase in total number), the detection power increases from 64% to 76%.

5.2.4 Determining the number of sources

To evaluate the proposed AIC and BIC criteria in determining the number of clusters, we used the benchmark setting of the mobile sensor network and put 0 to 4 sources in the study region. Then we followed the procedures in Section 3.3 and defined $\mathcal{K}$={0, 1, 2, 3, 4} as the pre-selected set of ks. For each $k \in \mathcal{K}$, we calculated the values of the modified version $AIC(k)_{mod}$ and $BIC(k)_{mod}$ from (15) and (16), and estimated the number of clusters accordingly.

Table 3 summarizes the model selection results. The modified AIC and BIC methods seem to work fine, picking up the number of clusters correctly for a majority of times in all settings. In addition, the BIC, with a larger penalty term, appears to perform a little better than the method of AIC.

5.3 Study II: Improvement of Detection Power from Combinations of Event Data

Detection power is directly related to event density (i.e., detection signals) around sources. The higher the density, the higher the power to detect the sources at any given time. Higher detection

Table 2: Power improvement with additional higher quality sensors. Notations D_{MC}, D_{LS}, T_{MC} and T_{LS} are defined in Table 1.

Number of Additional Sensors		Detection Range (feet) 240	300	360
50	D_{MC}	76.0	77.4	79.4
	D_{LS}	75.5	77.4	78.8
	T_{MC}	94.8	95.6	96.6
	T_{LS}	81.2	83.0	83.8
75	D_{MC}	76.6	77.8	80.6
	D_{LS}	76.2	77.6	80.4
	T_{MC}	95.2	96.2	97.1
	T_{LS}	81.6	84.2	85.4
100	D_{MC}	78.6	79.8	81.0
	D_{LS}	78.4	79.4	81.0
	T_{MC}	96.3	97.0	98.2
	T_{LS}	83.4	84.8	85.8

Table 3: Model selection evaluation. Reported in the table are the percentages of times that the AIC or BIC criterion selects for clusters numbers. The bold entry indicates the most selected cluster numbers in each case.

	AIC					BIC				
	Estimated					Estimated				
True k	k=0	k=1	k=2	k=3	k=4	k=0	k=1	k=2	k=3	k=4
0	**100.0**	0.0	0.0	0.0	0.0	**100.0**	0.0	0.0	0.0	0.0
1	29.4	**63.6**	6.4	0.6	0.0	29.6	**69.2**	1.0	0.2	0.0
2	7.4	11.4	**40.2**	31.6	0.4	7.6	19.4	**66.4**	6.2	0.4
3	2.2	1.2	10.8	**43.4**	42.4	2.2	2.8	20.6	**64.0**	10.4

power can be achieved at higher cost either by using higher quality sensors or using a larger number of vehicles. Therefore, this operation is often restricted within some budget limit. Alternatively, we can collect and combine event data at short time intervals to achieve similar results using the same number of sensors. We study this idea in this subsection and analyze the combined data in various scenarios by directly using our developed algorithm. Even though some data from neighboring time slices may be correlated, the simulation study suggests that combinations can help increase detection power, especially when sources are stationary at fixed locations or only move slowly.

In this simulation, we set the speed of vehicles to be around 25 miles per hour so that the vehicles would take 2 minutes to go through a block of 200 feet. Upon reaching an intersection, a vehicle was given an equal probability of making a turn or going straight. Under the benchmark network and when we have one stationary nuclear source, we set the sampling frequency to every 0.5, 1.0, 1.5 and 2.0 minutes to reflect the time durations that sensors need for sufficient radiation readings. At each frequency, different numbers of sample combinations were studied. For moving sources, sensor data accumulated over time may adversely impact the accuracy of detection. To quantify the impact of source movement on the detection power, we collected positive sensor positions over time, conducted the cluster detection process, and computed the percentage of times that the detected clusters cover the source. We reported the average of these percentages to give the detection power in this case. In this simulation, the source was assumed to move as fast as vehicles with sensors.

Table 4 shows the detection performance under various settings. The detection powers from both the Monte Carlo test and the large sample test are almost the same as in Section 5.2. For the fixed source case, detection power from 4 or more combinations increases to more than 98.0 percent. When the sampling frequency is 1.0 minute or longer, the power is already more than 92 percent just from two sample combinations. It is a sharp increase from 64% in detection power in the benchmark setting without any combination. For the fixed source case, the combinations of data are equivalent to adding more sensors. Therefore the detection powers increase accordingly.

The powers from a moving source are similar to those from a stationary source when the sampling frequency is 0.5 minute. When we increased the frequency and number of combinations, the powers in the moving source case are much lower than those in the stationary source case. For example, the combination did worse than the benchmark setting when we combined 6 or 7 samples

Table 4: Detection power improvement via combinations over time in the benchmark setting (1,500 vehicles, range=200 feet, (η, ζ)=(0.95,0.95)). Reference power is 65% with no combination. The powers are based on level 5% tests by the Monte Carlo method.

Sampling Frequency		Number of Combinations					
(minutes)	Type of Source	2	3	4	5	6	7
0.5	Fixed	77.0	93.5	98.0	98.5	99.0	99.5
	Moving	86.2	91.3	93.1	96.2	96.7	97.1
1.0	Fixed	92.5	98.0	99.5	99.5	99.5	100.0
	Moving	85.5	92.5	91.6	91.9	91.9	88.2
1.5	Fixed	93.5	99.5	99.5	99.5	99.5	100.0
	Moving	89.5	89.2	85.3	79.5	73.6	63.8
2.0	Fixed	94.0	99.5	99.5	99.5	99.5	100.0
	Moving	88.0	87.6	81.2	71.8	60.6	47.0

at the frequency of 2 minutes. With a moving source, the combinations with observations further apart in time allowed the source to move away and thus lowered the detection power.

This study shows that with a small number of combinations of event data at short time intervals, we can improve detection powers noticeably for both stationary and moving source. Note that here, in the data analysis, we did not use any information about vehicle or source movements. This approach is robust against model misidentification, although it may compromise some detection power compared to other methods with an accurate model of vehicle movement.

6. DISCUSSION

This paper proposes a robust mobile sensor network and develops a statistical algorithm to provide consistent and pervasive surveillance for nuclear materials in major cities. Simulation studies under the settings of Section 5 suggest that the proposed network and statistical methods can provide an effective tool to detect nuclear signals placed in a spatial region. Since software and algorithm research is a critical component to breakthroughs in nuclear detection, this study aims to provide forward-looking design and implementation of a detection capability from a sensor management point of view.

In this study, we assume that the clusters are nonoverlapping. In practice, when two nuclear sources are close enough, a sensor can be activated by both of them. The clusters formed around the sources might be overlapping. Chapter 5 of Cheng (2010) relaxed the nonoverlapping assumption and studied detection of overlapping clusters. Though the detection power increases slightly when the two or more sources are nearby, the computation cost is much higher. Practically it seems reasonable to use one cluster to represent multiple adjacent sources.

The mobile sensor network proposed here can be supplemented by stationary sensors. In fact, in most cases such a supplement is necessary to cover locations with sparse or zero traffic, such as a large park in the city. Our detection algorithm can be extended to handle both types of networks since the stationary sensors' positions and states can be merged and processed with those of mobile sensors. We also extended our approach from one sample to several samples collected at consecutive time intervals. By stacking the detection results over time, our method can provide dynamic surveillance.

Our models have only considered 2-dimensional regions and have disregarded the possibility that a stationary source might be located above the ground in a building. This added level of complexity requires more research.

Finally, we have not considered in the paper the shielding factors of the nuclear energy by the buildings or other materials. The shielding issue is very complex. It depends on the geographic distribution and shapes of the buildings and other large objects in the region. It also depends on the packing materials used to transport the nuclear materials. Nelson and Sokkappa (2008) provided a detailed report and developed a generic packaging model to study the impact of shielding for several nuclear and packing materials. Much study is still needed to understand the shielding impact especially in metropolitan areas, an interesting and challenging topic for future research.

APPENDIX. GIBBS/IMPORTANCE SAMPLING ALGORITHMS

To carry out the EM computation in Section 3.1.1, we need to obtain samples of $(\mathbf{O}, \mathbf{r})$ - the k clusters from their conditional density functions. We use Gibbs and importance sampling methods to generate clusters one at a time. Specifically, we need to generate samples of the jth cluster from its full conditional distribution given the rest of the clusters:

$$\begin{aligned} f_\theta((\mathbf{o_j}, r_j)|(\mathbf{o_l}, r_l), l = 1, .., k, l \neq j, Y, k) &\propto f_\theta(\mathbf{O}, \mathbf{r}, Y, k) \\ &\propto \frac{\alpha_j^{Z_j(\mathbf{y},\mathbf{O},\mathbf{r})} \psi_{\lambda_o}(\mathbf{o_j}) \psi_{\lambda_r}(r_j) \mathbf{1}_{\{\mathbf{O},\mathbf{r} \in \Omega_k\}}}{[A + \sum_{j=1}^k (\alpha_j - 1) A_j]^n}. \end{aligned} \tag{A.1}$$

It is almost impossible to use Equation A.1 directly to generate samples. Since it is straightforward to simulate a $(\mathbf{o_j}, r_j)$ from $\varphi_{\lambda_o}(\mathbf{o_j})\varphi_{\lambda_r}(r_j)$, we adopt an importance sampling approach as follows:

- Step 1. Simulate a large number, say S, of random deviates $(\mathbf{o}_j, r_j)^{[s]}(s = 1, .., S)$ from a distribution $\varphi_{\lambda_o}(\mathbf{o_j})\varphi_{\lambda_r}(r_j)$. For each s, compute weight

$$w_s = \frac{\alpha_j^{Z_j^{[s]}} \mathbf{1}_{((\mathbf{O^{[s]}}, \mathbf{r^{[s]}}) \in \Omega_k)}}{[A + \sum_{l \neq j} (\alpha_l - 1) D_l + (\alpha_j - 1) D_j^{[s]}]^n},$$

  where $Z_j^{[s]}$, $D_j^{[s]}$ and $\mathbf{1}_{((\mathbf{O^{[s]}}, \mathbf{r^{[s]}}) \in \Omega_k)}$ are calculated with $(\mathbf{o}_j, r_j)$ replaced by $(\mathbf{o}_j, r_j)^{[s]}$ and the rest of the $\mathbf{o}$'s and r's remain the same.

- Step 2. Simulate one set of $(\mathbf{o}_j, r_j)$ from the S sets $(\mathbf{o}_j, r_j)^{[1]}$,...,$(\mathbf{o}_j, r_j)^{[S]}$ with respective probabilities $(p_1, ..., p_S)$ where $p_s = ws / \sum_{s=1}^S ws$.

REFERENCES

Akaike, H. (1974), "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, 19, 716–723.

Akyildiz, I., Su, W., Sankarasubramaniam, Y. E., and Cayiric, E. (2002a), "Survey on sensor networks," *IEEE Communications Magazine*, August 2002, 102–114.

Akyildiz, I., Su, W., Sankarasubramaniam, Y. E., and Cayiric, E. (2002b), "Wireless sensor networks: a survey," *Computer Networks*, 38, 393–422.

Balakrishnan, N., and Koutras, M. (2001), *Runs and Scans with Applications*, New York: John Wiley & Sons.

Cheng, J. (2010), Bayesian Methods for Non-standard Missing Data Problems, PhD thesis, Department of Statistics, Rutgers University.

Claeskens, G., and Consentino, F. (2008), "Variable selection with incomplete covariate data," *Biometrics*, 64, 1062–1069.

Davies, R. B. (1977), "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 64, 247–254.

Davies, R. B. (1987), "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 74, 33–43.

Demattei, C., Molinari, N., and Daures, J. (2006), "Spatclas: an R package for arbitrarily shaped multiple spatial cluster detection for case event data," *Computer Methods and Programs in Biomedicine*, 84, 42–49.

Demattei, C., Molinari, N., and Daures, J. (2007), "Arbitrary shaped multiple spatial cluster detection for case event data," *Computational Statistics and Data Analysis*, 51, 3931–3945.

Dembo, A., and Karlin, S. (1992), "Poisson approximation for r-scan processes," *Annals of Applied Probability*, 2, 329–357.

Denison, D., and Holmes, C. (2001), "Bayesian partitioning for estimating disease risk," *Biometrics*, 57, 143–149.

Diggle, P., Rowlingson, B., and Sun, T. L. (2005), "Point processes methodology for on-line spatio-temporal disease surveillance," *Environmetrics*, 16, 423–434.

Dwass, M. (1957), "Modified randomization tests for nonparametric hypothesis," *Annals of Mathematical Statistics*, 28, 181–187.

FEMA (2008), "Are you ready?,". `http://www.fema.gov/areyouready/nuclear_blast.shtm`.

Fu, J. C., and Lou, W. Y. (2003), *Distribution theory of runs and patterns and its applications: A finite Markov Chain embedding approach*, Singapore: World Scientific.

Gangnon, R., and Clayton, M. (2000), "Bayesian detection and modeling of spatial disease maps," *Biometrics*, 56, 922–935.

Gangnon, R., and Clayton, M. (2003), "A hierarchical model for spatially clustered disease rates," *Statistics in Medicine*, 22, 3213–3228.

Ghosh, M., Natarajan, K., Waller, L. A., and Kim, D. (1999), "Hierarchical Bayes GLMs for the analysis of spatial data: an application to disease mapping," *Journal of Statistical Planning and Inference*, 75, 305–318.

Glaz, J., and Balakrishnan, N. (1999), *Scan Statistics and Applications*, Boston: Birkhauser.

Glaz, J., Naus, J., and Wallenstein, S. (2001), *Scan Statistics and Applications*, New York: Springer.

Hochbaum, D. (2009), "The multi-sensor nuclear threat detection problem," in *Proceedings of the Eleventh INFORMS Computing Society (ICS) Conference*, eds. J. Chinneck, B. Kristjansson, and M. Saltzman, New York: Springer, pp. 389–399.

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

Kulldorff, M., and Nagarwalla, N. (1995), "Spatial disease clusters: Detection and infection," *Statistics in Medicine*, 14, 799–810.

Lawson, A. (1995), "Markov Chain Monte Carlo methods for putative pollution source problems," *Environmental Epidemiology*, 14, 2473–2486.

Naus, J. (1966), "A power comparison of two sets of non-random clustering," *Technometrics*, 8, 493–517.

Naus, J., and Wallenstein, S. (2004), "Multiple window and cluster size scan procedures," *Methodology and Computing in Applied Probability*, 6, 389–400.

Nelson, K., and Sokkappa, P. (2008), "A statistical model for generating a population of unclassified objects and radiation signatures spanning nuclear threats," *Technical Report, Lawrence Livermore National Laboratory (LLNL)*, LLNL-TR-408407. http://www.osti.gov/energycitations/product.biblio.jsp?query_id=1&page=0&osti_id=947761.

New York City Taxi & Limousine Commission (2005), "TLC Press Release #05-04, September 8, 2005,". http://www.nyc.gov/html/tlc/html/news/press05_04.shtml.

Purdue University (2008), "Cell phone sensors detect radiation to thwart nuclear terrorism,". http://news.uns.purdue.edu/x/2008a/080122FischbachNuclear.html.

Rodriguez, A., Sunson, D. B., and Gelfand, A. E. (2008), "The nested Dirichlet process (with discussion)," *Journal of the American Statistical Association*, 103, 1131–1144.

Schwarz, G. (1978), "Estimating the dimension of a model," *Annals of Statistics*, 6, 461–464.

Su, X., Wallenstein, S., and Bishop, D. (2001), "Non-overlapping clusters: approximation distribution and application to molecular biology," *Biometrics*, 57, 420–426.

Sun, Q. L. (2008), Statistical modeling and inference for multiple temporal or spatial cluster detection, PhD thesis, Department of Statistics, Rutgers University.

Tanner, M. (1993), *Tools for Statistical Inference*, New York: Springer-Verlag.

Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997), "Hierarchical spatio-temporal mapping of disease rates," *Journal of the American Statistical Association*, 92, 607–617.

Wein, L., Wilkins, A., Baveja, M., and Flynn, S. (2006), "Preventing the importation of illicit nuclear materials in shipping containers," *Risk Analysis*, 26, 1377–1393.

Xie, M., Sun, Q., and Naus, J. (2009), "A latent model to detect multiple temporal clusters," *Biometrics*, 65, 1011–1020.