International Journal for Uncertainty Quantification, 11(5):39–55 (2021)

SURROGATE BASED MUTUAL INFORMATION APPROXIMATION AND OPTIMIZATION FOR URBAN SOURCE LOCALIZATION

Andrew N. Hollis,[1,*] Ralph C. Smith,[2] & Alyson G. Wilson[1]

[1]*Department of Statistics, North Carolina State University Raleigh, North Carolina 27695, USA*

[2]*Department of Mathematics, North Carolina State University Raleigh, North Carolina 27695, USA*

*Address all correspondence to: Andrew N. Hollis, Department of Statistics, North Carolina State University Raleigh, North Carolina 27695, USA, E-mail: anhollis@ncsu.edu

Original Manuscript Submitted: 04/09/2020; Final Draft Received: 12/20/2020

The ability to efficiently and accurately localize potentially threatening nuclear radiation sources in urban environments is of critical importance to national security. Techniques to infer the location and intensity of a source using data from a configuration of radiation detectors, and the effectiveness of the source localization depends critically on how the detectors are configured. In this paper, we introduce a framework that uses surrogate models to efficiently compare and optimize different detector configurations. We compare our technique to others and demonstrate its effectiveness for selecting optimal detector configurations in the context of urban source localization.

KEY WORDS: *computational statistics, information theory, bayesian inference, spatial statistics*

1. INTRODUCTION

The interception of smuggled nuclear materials, prevention of nuclear terrorism, and recovery of hazardous radiological sources requires accurate and efficient methods for radiation source detection and localization. Hence, the development, implementation, and experimental validation of mathematical and statistical algorithms for detecting and locating nuclear devices in urban environments constitutes a critical nuclear security and nonproliferation problem.

Radiation source detection is the process of determining the presence or absence of a radiation source in a given area, whereas radiation source localization is the estimation of the location and intensity of one or more sources that are assumed to be present. Both localization and detection require computing the response induced at a detector by a source. Because of the computational cost associated with computing detector responses, surrogate models are employed to approximate detector responses or functions of detector responses. In this paper, we will consider an approach that optimizes surrogate models of design criteria to find optimal configurations of radiation detectors that will yield precise localization results for sources in urban environments. Whereas some authors address both detection and localization [1–3], we will focus on the localization problem assuming a radiation source is present. We will consider a single, fixed-location source, and refer readers to [3,4] for moving sources and [1] for multiple sources.

Radiation source localization employs a configuration of radiation detectors to collect data used to estimate a source's location and intensity parameters. If there are no obstructions between the source and the detector and no noise or error in detector readings, a simple four-detector method suffices to localize the source [5]. Obstructions or sources of random noise in the environment induce error and variation in detector measurements making it necessary to fuse data collected from several detectors to perform localization accurately.

In the absence of obstructions, Chin et al. [6] introduced methods that cluster the candidate source locations obtained using the ratios of squared distances computed from several three-detector subsets to identify the true source location. Another approach is to treat detector measurements as random variables generated by a statistical model parameterized by the source parameters. Maximum likelihood estimation (MLE) can be used to estimate the source

parameters [2,7,8]. While MLE methods provide point estimates for the source parameters, it can be difficult to obtain exact or asymptotic variances for these estimates, making uncertainty quantification of these estimates challenging. This has led many investigators to employ Bayesian inference methods, which provide a posterior distribution for source parameters that can be used to easily compute point estimates for the source parameters and quantify the uncertainty of these estimates [9–11]. We will employ a Bayesian formulation for source localization.

Radiation source localization is most important, and often most challenging, when the source is in an urban environment. The detonation of a nuclear device in a densely populated urban environment could cause major loss of life, serious infrastructural damage, and severe health hazards. Lost or stolen medical radiation sources or misplaced radioactive waste can also pose health threats to an urban population. Various characteristics of the urban environment, however, make source localization particularly difficult. In contrast to the noise-free, obstruction-free scenario, buildings with varying materials as well as moving objects such as vehicles can obstruct the source. The presence of background radiation, which may vary in time and space within a city, also introduces noise that may be difficult to separate from radiation emitted by the source of interest.

To account for fixed position attenuating objects, such as buildings, that may block the path between the detector and the source, a simplified version of the Boltzmann transport model can be employed [9,11]. To account for background radiation, Penny et al. [12] showed that the environment can be partitioned so that each region has approximately uniform background. Alternatively, by employing a hierarchical Bayesian model, background radiation estimation and source localization can be accomplished simultaneously [13,14]. For simplicity, we will treat background radiation observed at each detector as a Gaussian random variable with fixed parameters.

Another challenge of urban source localization is the sensitivity of the localization results to the employed configuration of the radiation detectors. In the Bayesian localization context, different detector configurations will produce different posterior distributions for the source parameters. An optimal configuration will produce a posterior distribution with minimal uncertainty. Michaud [13] and Schmidt [14] used mutual information to quantify the reduction in uncertainty for the source parameters achieved by knowing the responses observed at a detector configuration. Finding a detector configuration that maximizes mutual information is equivalent to finding a configuration that yields the greatest reduction in uncertainty about the source parameters. Michaud [13] and Schmidt [14] optimized mutual information over combinations of a finite subset of detectors. The cost of computing the response at hundreds or thousands of source detector pairs makes this approach appealing because it limits the number of detectors considered. The primary disadvantage of this approach is that the ideal configuration may not exist as a combination of some fixed set of possible detector locations.

Instead of computing mutual information explicitly for each detector configuration, we can construct a surrogate model between detector configurations and mutual information that does not require thousands of detector response computations. In this paper, we will present an approach for maximizing mutual information over all possible detector configurations by constructing a surrogate model between mutual information and detector configuration and then optimizing over this surrogate model. This approach is computationally efficient and considers general detector configurations instead of only a finite subset of possible configurations. The generality of this approach is particularly important in larger urban environments or three-dimensional space where small subsets of candidate detector locations may not be representative of the space.

The paper is organized as follows: in Section 2, we discuss the Bayesian framework used for source localization, mutual information estimation, and the combinatorial optimization of mutual information demonstrated in [13,14]. In Section 3, we develop a continuous optimization approach using surrogate models, which, in Section 4, we compare to the performance of the combinatorial optimization approach for a localization problem with movable detector networks. We present concluding remarks in Section 5.

2. BAYESIAN SOURCE LOCALIZATION AND DETECTOR CONFIGURATION SELECTION

2.1 Bayesian Framework for Constructing Posterior Distributions of Source Parameters

Our objective is to estimate the source parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, where θ_1 and θ_2 are the two-dimensional location coordinates of the source, and θ_3 is the source intensity. We assume the source is located in a bounded search space. We use the block of downtown Washington, DC described by Stefanescu et al. [11] as our search space.

To obtain estimates and perform uncertainty quantification for $\boldsymbol{\theta}$, we treat $\boldsymbol{\theta}$ as a random vector and compute the posterior distribution, $p(\boldsymbol{\theta}|\mathbf{Y})$, given the data $\mathbf{Y}$ where $\mathbf{Y} = (Y_1, ..., Y_n)$ are the radiation count measurements taken from a configuration of n detectors, $\mathbf{x} = \{d_{1j}, d_{2j}\}_{j=1}^{n}$, where d_{1j} and d_{2j} are the location coordinates of the jth detector. To compute $p(\boldsymbol{\theta}|\mathbf{Y})$, we require the probability of the data given $\boldsymbol{\theta}$, $p(\mathbf{Y}|\boldsymbol{\theta})$, and a prior distribution for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$. We employ delayed rejection adaptive Metropolis (DRAM) [15] Markov chain Monte Carlo (MCMC) methods to sample $p(\boldsymbol{\theta}|\mathbf{Y})$.

In specifying the $p(\mathbf{Y}|\boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$, we will describe the process we use to generate detector response data $\mathbf{Y}$ for a set of radiation sources. In practice, $\mathbf{Y}$ would be collected from real detectors. Instead of using real detector data as in [9] or expensive, high-fidelity radiation transport codes [16], we have simulated detector response using a statistical framework employing a simplification of the Boltzmann transport model. All detector response data used in this paper will be generated using this framework.

We begin by drawing N parameter vectors, $\boldsymbol{\theta}$, from the source parameter prior distribution, $p(\boldsymbol{\theta})$,

$$p(\theta_s) = \frac{I(\theta_{s,\text{min}} < \theta_s < \theta_{s,\text{max}})}{\theta_{s,\text{max}} - \theta_{s,\text{min}}}, \quad s = 1, 2, 3, \tag{1}$$

$$p(\boldsymbol{\theta}) = p(\theta_1) \cdot p(\theta_2) \cdot p(\theta_3). \tag{2}$$

Here $\theta_{s,\text{min}}$ and $\theta_{s,\text{max}}$ are, respectively, the smallest and largest values that θ_s can take for $s = 1, 2, 3$. Notice that the source parameters are assumed to be independent. Because our search space is 246.615 m $\times$ 176.333 m, we set $\theta_{1,\text{min}} = 0$, $\theta_{1,\text{max}} = 246.615$, $\theta_{2,\text{min}} = 0$, $\theta_{2,\text{max}} = 176.333$. We set $\theta_{3,\text{min}} = 10^9$, and $\theta_{3,\text{max}} = 5 \times 10^9$.

After generating N source parameter vectors, we evaluate the expected detector response due to the radiation source for each of the N sources at n detectors. The expected detector response γ_i due to a source located at (θ_1, θ_2) with intensity θ_3 for detector i located at $\mathbf{d}_i = \{d_{i1}, d_{i2}\}$ is calculated using an approximation to the Boltzmann transport model [11],

$$\gamma_i = \theta_3 \cdot \Delta t_i \cdot \epsilon_i \cdot \frac{A_i}{4\pi||\mathbf{d}_i - (\theta_1, \theta_2)||_2^2} \exp\left(- \sum_{p=1}^{N_\ell} \ell_p \cdot \Sigma_T^{(p)} \right), \tag{3}$$

Here ϵ_i is the efficiency of the detector, and A_i is the face area of the detector. We assume all the detectors are the same and have the same dwell time, Δt_i, so ϵ_i, A_i, and Δt_i will not vary with i. Here N_ℓ is the number of buildings in the bounded search space intersected by a ray from the source to the detector, ℓ_p is the length of the portion of the ray intersecting building p, and $\Sigma_T^{(p)}$ is the macroscopic cross section of building p. We note that γ_i quantifies uncollided flux, the flux of photons on the path from the source to the detector that are not absorbed or scattered.

To generate the observed radiation count at detector i, we draw B_i from a Gaussian distribution with mean 140.8 and variance 9.61, multiply it by Δt_i, and add this to γ_i to get

$$\lambda_i = \gamma_i + B_i \cdot \Delta t_i, \quad B_i \sim N(140.8, 9.61).$$

Here B_i simulates the background radiation in the urban environment in counts per second. The mean and variance based on radiation concentration measurements collected from a site replicating a small urban environment at Fort Indiantown Gap National Guard Training Center [17]. The concentration measurements were used to compute detector responses with the gamma detector response and analysis software detector response function (GADRAS-DRF) [18]. These mean and variannce parameters were also used in [13].

Observed radiation counts are often treated as Poisson random variables [19], with the Poisson accounting for detector measurement error. Thus, after computing λ_i, the expected radiation count due to the radiation source, and the random radiation count due to background, we draw the radiation count observed at detector i from a Poisson distribution with λ_i as the mean parameter. The probability of the data observed at n detectors, given θ and B_i is

$$p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{B}) = \frac{e^{-\sum_{i=1}^{n} \lambda_i} \prod_{i=1}^{n} \lambda_i^{y_i}}{\prod_{i=1}^{n} y_i!}. \tag{4}$$

2.2 Mutual Information Estimation

To select an optimal configuration, $\mathbf{x}^{\text{opt}}$, at which to collect the detector data, we maximize the mutual information between the source parameters $\boldsymbol{\theta}$ and the data collected at a particular detector configuration $\mathbf{Y}$. We denote this quantity as $I(\boldsymbol{\theta}, \mathbf{Y})$.

The mutual information between continuous random vectors $\mathbf{Y}$ and $\boldsymbol{\theta}$ is

$$I(\boldsymbol{\theta}, \mathbf{Y}) = \int_{\boldsymbol{\theta}} \int_{\mathbf{Y}} p(\boldsymbol{\theta}, \mathbf{Y}) \log\left(\frac{p(\boldsymbol{\theta}, \mathbf{Y})}{p(\boldsymbol{\theta})p(\mathbf{Y})}\right) d\boldsymbol{\theta} d\mathbf{Y}, \tag{5}$$

or equivalently,

$$I(\boldsymbol{\theta}, \mathbf{Y}) = H(\boldsymbol{\theta}) - H(\boldsymbol{\theta}|\mathbf{Y}),\ H(\boldsymbol{\theta}) = -\int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}) \log p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{6}$$

where $H(\boldsymbol{\theta})$ is the Shannon entropy of $\boldsymbol{\theta}$, a measure of the uncertainty of $\boldsymbol{\theta}$. Thus, mutual information between $\boldsymbol{\theta}$ and $\mathbf{Y}$ is the total uncertainty of $\boldsymbol{\theta}$ minus the uncertainty of $\boldsymbol{\theta}$ given we know $\mathbf{Y}$.

In general, Eq. (5) cannot be evaluated analytically. Consequently, many estimators of mutual information have been proposed to estimate this quantity [20–23]. We use the Kraskov-Stögbauer-Grassberger (KSG) k-nearest neighbors estimator [23]. Although biased, the KSG estimator is straighforward and efficient to compute and has been shown to perform well in selecting optimal configurations based on the maximization of the mutual information criterion [13].

The KSG mutual information estimator, employing N draws from $p(\boldsymbol{\theta})$, $\{\boldsymbol{\theta}_i\}_{i=1}^N$, and the corresponding realizations of data, $\{\mathbf{Y}_i\}_{i=1}^N$, from some detector configuration is

$$\hat{I}(\boldsymbol{\theta}, \mathbf{Y}) = \psi(N) + \psi(k) - \frac{1}{N}\sum_{i=1}^{N} \psi(n_{\boldsymbol{\theta}}(\epsilon_{\boldsymbol{\theta}}(i))) - \frac{1}{N}\sum_{i=1}^{N} \psi(n_{\mathbf{Y}}(\epsilon_{\mathbf{Y}}(i))) - \frac{1}{k}, \tag{7}$$

where

$$n_{\boldsymbol{\theta}}(\epsilon_{\boldsymbol{\theta}}(i)) = \sum_{j=1}^{N} I\left(||\boldsymbol{\theta}_j - \boldsymbol{\theta}_i||_\infty \leq \frac{\epsilon_{\boldsymbol{\theta}}(i)}{2}\right) \text{ and } n_{\mathbf{Y}}(\epsilon_{\mathbf{Y}}(i)) = \sum_{j=1}^{N} I\left(||\mathbf{Y}_i - \mathbf{Y}_j||_\infty \leq \frac{\epsilon_{\mathbf{Y}}(i)}{2}\right), \tag{8}$$

and where

$$\frac{\epsilon_{\boldsymbol{\theta}}(i)}{2} = \max_{1\leq j\leq k} ||\boldsymbol{\theta}_i^j - \boldsymbol{\theta}_i||_\infty \text{ and } \frac{\epsilon_{\mathbf{Y}}(i)}{2} = \max_{1\leq j\leq k} ||\mathbf{Y}_i^j - \mathbf{Y}_i||_\infty. \tag{9}$$

Here $(\boldsymbol{\theta}_i^j, \mathbf{Y}_i^j)$ is the jth nearest neighbor to $(\boldsymbol{\theta}_i, \mathbf{Y}_i)$ in the joint space for $\boldsymbol{\theta}$ and $\mathbf{Y}$, N is the number of $(\boldsymbol{\theta}_i, \mathbf{Y}_i)$ pairs in the data, and $\psi(\cdot)$ is the digamma function. The estimator uses the form of mutual information in (6). The average count of points in a ball with radius equal to a point's distance to its kth nearest neighbor is employed to estimate entropy. We direct the reader to [23] for a full derivation of this estimator and its connection to entropy.

For fixed $\{\boldsymbol{\theta}_i\}_{i=1}^N$, $\hat{I}(\mathbf{Y}, \boldsymbol{\theta})$ is a noisy function of detector configuration $\mathbf{x}$, requiring $N \times n$ evaluations of (3). The noise in $\hat{I}(\mathbf{Y}, \boldsymbol{\theta})$ is due to the random background radiation detector readings, B_i, and $\mathbf{Y}_i$ being a Poisson random variable. To emphasize the dependence on $\mathbf{x}$, we denote $\hat{I}(\mathbf{Y}, \boldsymbol{\theta})$ as $\hat{I}(\mathbf{Y}(\mathbf{x}), \boldsymbol{\theta})$.

Our objective is to determine the detector configuration that maximizes $E(\hat{I}(\mathbf{Y}(\mathbf{x}), \boldsymbol{\theta})|\boldsymbol{\theta})$, which we denote as $g(\mathbf{x})$, a function of $\mathbf{x}$ because the expectation taken over $\mathbf{Y}$ eliminates the dependence on $\mathbf{Y}$. Thus our objective is

$$\mathbf{x}^{\text{opt}} = \underset{\mathbf{x}\in\mathcal{D}}{\operatorname{argmax}}\ g(\mathbf{x}). \tag{10}$$

Here $\mathbf{x} \in \mathcal{D}$ if and only if $\theta_{1,\min} \leq d_{1j} \leq \theta_{1,\max}$ and $\theta_{2,\min} \leq d_{2j} \leq \theta_{2,\max}$ for all $j = 1, ..., n$.

In general, solving (10) requires K evaluations of $\hat{I}(\mathbf{Y}(\mathbf{x}), \boldsymbol{\theta})$ and $N \times n \times K$ evaluations of Eq. (3), for K on the order of 10^4 or more to insure the space over $\mathbf{x}$ is thoroughly searched. Because of the cost of evaluating (3), (10) can be prohibitively expensive to solve over all $\mathcal{D}$.

2.3 Combinatorial Optimization of Detector Configurations

To facilitate the solution of Eq. (10), Michaud [13] and Schmidt [14] proposed maximizing $\hat{I}(\mathbf{Y}(\mathbf{x}), \boldsymbol{\theta}|\boldsymbol{\theta})$ over a finite 2D set $\{d_{1m}, d_{2m}\}_{m=1}^{p}$ instead of $\mathcal{D}$. This strategy is illustrated for our scenario in Fig. 1. For this strategy, there are only p possible detector locations, and we only need to evaluate (3) $N \times p$ times. The data for any K configurations we wish to consider will be some subset of these $N \times p$ computations. Instead of using this combinatorial optimization approach, we will propose a new surrogate-model based approach that allows us to perform continuous optimization over the whole detector configuration space.

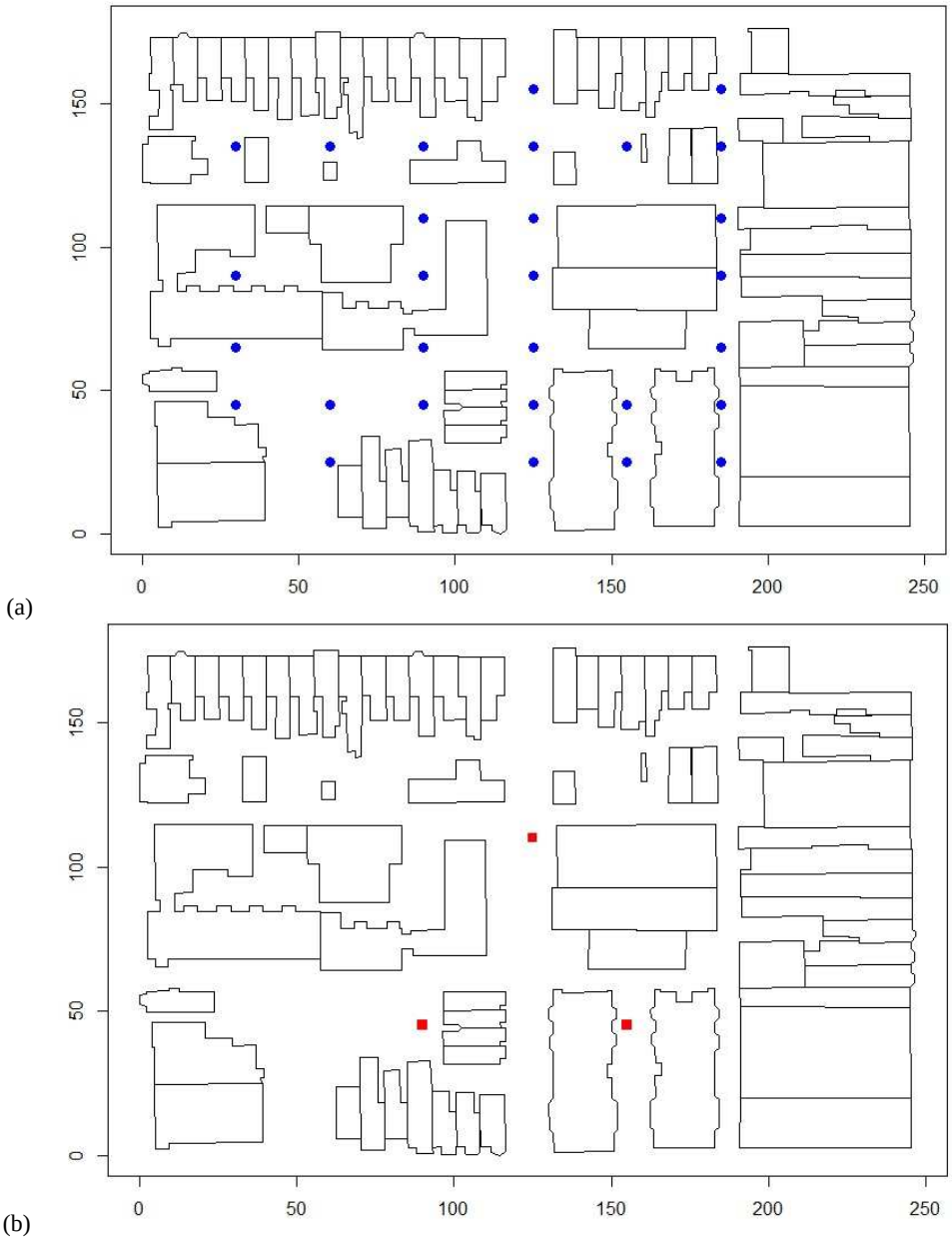


FIG. 1: Optimization of detector configurations over finite subset. (a) Finite set of p detector locations and (b) optimal three-detector configuration.

3. CONTINUOUS OPTIMIZATION OF DETECTOR CONFIGURATIONS

3.1 Surrogate Based Mutual Information Approximation and Optimization

Rather than constrain $\mathcal{D}$ to some finite set, we could find an approximate solution to Eq. (10) by creating a surrogate model that approximates $g(\mathbf{x})$ and is less expensive to compute. We first describe a method for constructing sets of input-output pairs that can be used to fit these models. We then review a variety of popular surrogate models that can be used for our application and compare the performance of these models in capturing the relationship between mutual information and detector configuration.

For some fixed $\{\boldsymbol{\theta}_j\}_{j=1}^N$, let $z_i = \hat{I}(\mathbf{Y}(\mathbf{x}_i), \boldsymbol{\theta})$ be the KSG estimated mutual information corresponding to the ith n-length detector configuration, $\mathbf{x}_i$. We assume

$$z_i = g(\mathbf{x}_i) + e_i. \tag{11}$$

Here $e_i \overset{iid}{\sim} N(0, \sigma^2)$ is noise due to random background radiation and the stochastic nature of the detector readings.

It is difficult to optimize $g(\mathbf{x})$ directly. We never observe it or compute it directly; instead we only compute the noisy z_i estimates. Each z_i is expensive to generate. A single z_i estimate from 1000 data points for a five-detector configuration can take 8–10 min to generate on a standard laptop computer. Rather than optimizing $g(\mathbf{x})$ directly, we will generate a set of outputs $\{z_i\}_{i=1}^M$ at a set of inputs $\{\mathbf{x}_i\}_{i=1}^M$, called design points, and use these design points to build a surrogate model $\hat{g}(\mathbf{x})$ that approximates $g(\mathbf{x})$. Surrogate models can be used to efficiently find approximate optima of expensive objective functions [24,25].

Under this surrogate based optimization framework we replace $g(\mathbf{x})$ with $\hat{g}(\mathbf{x})$ to obtain the new optimization problem,

$$\mathbf{x}^{\text{opt}} = \underset{\mathbf{x}\in\mathcal{D}}{\operatorname{argmax}}\, \hat{g}(\mathbf{x}). \tag{12}$$

When the surrogate model is built only once for a fixed set of design points prior to optimization, the surrogate-based optimization is called the one-shot solution [25]. This is the solution method we will employ to perform detector configuration optimization.

3.2 Design Point Selection

To construct $\hat{g}(\mathbf{x})$, we require a set of M input-output pairs, $\{z_i, \mathbf{x}_i\}_{i=1}^M$, from the model in Eq. (11). The set $\{\mathbf{x}_i\}_{i=1}^M$, called the design points set, should be space-filling, so that the design points are spread throughout the input space we are investigating to insure the surrogate model is accurate for the entire space. One popular space-filling design is the Latin hypercube design (LHD) [26]. A LHD of size M is generated by treating each input variable as an independent random variable and then dividing each variable's range into M intervals having equal probability. If U is an input variable with assumed distribution function $F_U(u)$, then a LHD selects interval boundaries $t_1 < ... < t_M$ in the range of U_k such that $F_U(t_1) = F_U(t_2) - F_U(t_1) = ... = F_U(t_M) - F_U(t_{M-1})$. After performing this procedure for all variables in the input, design points are selected so that every interval for each variable is represented in the design by exactly one design point.

A key advantage of using a LHD is that it allows us to manipulate the resolution at which different regions of the input space will be represented. If there are parts of the input space that we are less interested in, we can select distribution functions for the input variables that assign low probability to those regions and those regions will be sparsely sampled. Figure 2 shows the LHD from (a) two independent uniform (–3, 3) input variables and (b) two independent standard Gaussian random variables. Note that because the standard Gaussian has more probability mass concentrated near its mean of zero, the majority of the points in the Gaussian LHD are clustered in the box $[-1, 1] \times [-1, 1]$. The uniform LHD, however, distributes points evenly across the range of the two input variables. For our application, we wish to have more configuration design points with detectors in regions of the search space where the source location has high probability. To achieve this using the LHD method, we will use the marginal prior distribution functions of the source location coordinates as described in [27] for independent input variables. We denote the generated matrix of design points as $\mathbf{X}$.

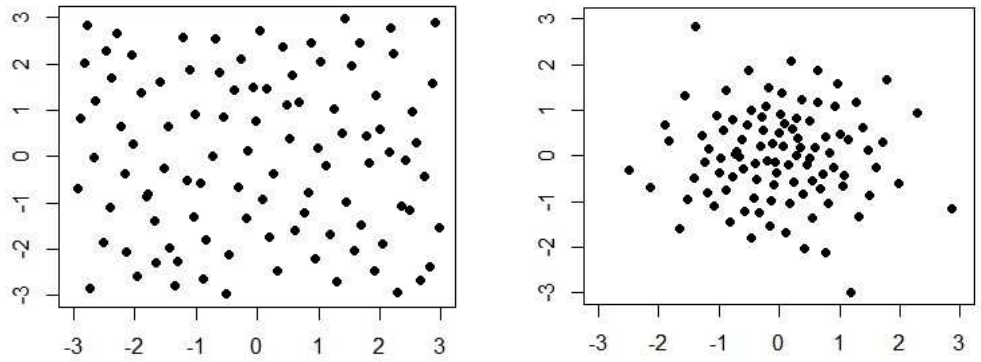


FIG. 2: Uniform versus Gaussian Latin hypercube design. LHD from two independent (a) uniform $(-3, 3)$ variables and (b) Gaussian $(0, 1)$ variables.

From Section 2.3, computing the mutual information for M design points, each with n detectors, requires evaluating (3) $N \times n \times M$ times, where N is the number of source locations used to approximate mutual information. We need $n \times M$ evaluations for each of the N source locations because each of the n detectors in the M design points is distinct and has distinct responses for each source. Instead of computing the response at the $n \times M$ distinct detectors in the design, we could reduce the computational cost by reducing the number of distinct detectors in the M point design to some $p < n \times M$, and then use the p distinct detectors to construct a design that approximates the original $n \times M$ design. In particular, we can generate a p point LHD across the space of possible detector locations: $P = \{(d'_{11}, d'_{21}), (d'_{12}, d'_{22}), ..., (d'_{1p}, d'_{2p})\}$ and compute the detector response at these locations in $p \times N$ model evaluations. We then approximate $\mathbf{X}$ by replacing each $(d_{1ij}, d_{2ij}) \in \mathbf{X}$ with $\underset{(d'_{1k}, d'_{2k}) \in P}{\arg\min} ||(d'_{1k}, d'_{2k}) - (d_{1ij}, d_{2ij})||_2$. This constructs the best elementwise ℓ_2-norm approximation to $\mathbf{X}$ with the elements of P. We denote this approximation as $\tilde{\mathbf{X}}$. Simulation studies show that while using $\tilde{\mathbf{X}}$ instead of $\mathbf{X}$ has a small impact on localization precision, it does not have a significant impact on localization accuracy (see Appendix A).

3.3 Surrogate Model Selection

We consider three surrogate models for $g(\mathbf{x})$: Gaussian processes (GP) (a.k.a. kriging), Bayesian Multivariate Adaptive Regression Splines (BMARS), and thin-plate spline radial basis functions (RBF). Each method is trained using M estimated mutual information values, $\{z_i\}_{i=1}^M = \mathbf{z}$, and a full LHD, $\mathbf{X}$, or an approximated LHD, $\tilde{\mathbf{X}}$.

3.3.1 Gaussian Processes

Gaussian processes are a class of nonparametric, nonlinear models. Consider an $M \times 1$ vector of KSG estimates, $\mathbf{z}$, and an $M \times 2n$ design matrix of detector configuration coordinates, $\mathbf{X}$. We assume that

$$z_i = g(\mathbf{x}_i) + e_i, \;\; e_i \overset{iid}{\sim} N(0, \sigma_e^2). \tag{13}$$

Gaussian processes approximate $g(\mathbf{x})$ for new configuration points $\mathbf{x}_i^*$ by constructing a conditional distribution for unobserved function values $z_i^* = g(\mathbf{x}_i^*)$ given $\mathbf{z}$ and $\mathbf{X}$, allowing us to make predictions for the underlying function and assess the uncertainty of these predictions.

Let $g(\mathbf{X})$ be the M-length vector of $g(\cdot)$ evaluated for each row of $\mathbf{X}$. A prior for $g(\mathbf{X})$ is

$$g(\mathbf{X}) \sim MVN(\beta_0 \mathbf{1}_M, \sigma^2 \mathbf{R}). \tag{14}$$

Here $\mathbf{1}_M$ is an M-length vector of ones, I is the $M \times M$ identity matrix, and $\mathbf{R}$ is an $M \times M$ matrix with $R_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$, for kernel function $K(\cdot, \cdot)$.

In predicting M^* new function values $\mathbf{z}^*$ for a set of new inputs $\mathbf{X}^*$ we derive the conditional expected value, which we use as the predicted value for the underlying function, and the conditional variance as

$$E[\mathbf{z}^*|\mathbf{z},\mathbf{X}] = \beta_0 + \sigma^2\mathbf{r}^{*T}(\sigma_e^2 I + \sigma^2\mathbf{R})^{-1}(\mathbf{z} - \beta_0\mathbf{1}_M), \tag{15}$$

$$Var[\mathbf{z}^*|\mathbf{z},\mathbf{X}] = \sigma^2\mathbf{R}^* - \sigma^4\mathbf{r}^{*T}(\sigma_e^2 I + \sigma^2\mathbf{R})^{-1}\mathbf{r}^*, \tag{16}$$

where $\mathbf{r}^*$ is an $M \times M^*$ matrix with $\mathbf{r}^*_{ij} = K(\mathbf{x}_i, \mathbf{x}^*_j)$, and $\mathbf{R}^*$ is an $M^* \times M^*$ matrix with $\mathbf{R}^*_{ij} = K(\mathbf{x}^*_i, \mathbf{x}^*_j)$.

The parameters β_0 and σ^2 are fit using maximum likelihood estimation. The noise variance, σ_e^2, is estimated by computing the sample variance of mutual information for several replications of the mutual information estimator at the same detector configuration.

Due to the presence of buildings and other obstacles in our input space, mutual information is expected to be a nonsmooth function of detector configuration. The kernel function $K(\cdot,\cdot)$ can be used to encode smoothness of the underlying function $g(\mathbf{x})$ being modeled [28]. The Matérn 3/2 or Matérn 5/2 kernels are often used for modeling nonsmooth functions. Simulation studies confirmed the superiority of the Matérn 3/2 kernel for our problem. The Matérn 3/2 kernel is

$$K(\mathbf{x}_i,\mathbf{x}_j) = \left(1 + \sqrt{3}h(\mathbf{x}_i,\mathbf{x}_j)\right)\exp\left\{-\sqrt{3}h(\mathbf{x}_i,\mathbf{x}_j)\right\}, \tag{17}$$

where $h(\mathbf{x}_i,\mathbf{x}_j) = \sqrt{\sum_{k=1}^{2n}[(x_{ik} - x_{jk})^2/\ell_k^2]}$, where each ℓ_k is a scaling factor that is fit using maximum likelihood. We use the DiceKriging package in R for fitting GP models [29].

3.3.2 Bayesian Multivariate Adaptive Regression Splines

One common alternative to GPs is the multivariate adaptive regression spline (MARS) [30]. MARS models are flexible, nonparametric models. They often scale better than GP models to training sets with high dimension or a large number of design points because they do not require the inversion and multiplication of large matrices [31,32]. As with GP models, we assume the observed mutual information values $\mathbf{z}$ represent noisy observations of some function of the input configuration coordinates $\mathbf{x}_i$,

$$z_i = g(\mathbf{x}_i) + e_i, e_i \overset{iid}{\sim} N(0,\sigma^2). \tag{18}$$

MARS approximates the underlying function $g(\cdot)$ with a basis function expansion,

$$\hat{g}(\mathbf{x}_i) = a_0 + \sum_{q=1}^{Q} a_q B_q(\mathbf{x}_i), \tag{19}$$

where the basis functions, $B_q(\mathbf{x}_i)$, are

$$B_q(\mathbf{x}_i) = \prod_{k=1}^{K_q}[s_{kq}(x_{v_{kq}} - t_{kq})]_+. \tag{20}$$

Here K_q is the maximum degree of interaction between terms, v_{kq} indexes a variable selected from the set of all input variables, t_{kq} is a knot location selected from the design points for $x_{v_{kq}}$, and s_{kq} is –1 or 1. The truncation function is $[x]_+ = \max\{0, x\}$. In Friedman's original paper [30], MARS models are fit using a two-stage process. In the first stage, the forward pass, the algorithm generates a large pool of candidate basis functions by adding to the model the basis function which most reduces the model fit error given the other basis functions already in the model. In the second stage, the backward pass, candidate basis functions are sequentially eliminated using the generalized cross-validation criterion until some stopping criterion is met. The forward pass insures that all relevant basis functions are considered as candidates for the model; the backward pass insures the model is not overfit to the training data.

The Bayesian formulation of MARS, first proposed in [33], samples the posterior distribution of the parameter vector $\{\sigma^2, Q, \mathbf{a}, \mathbf{K}, \mathbf{s}, \mathbf{v}, \mathbf{t}\}$ given the observed data $\mathbf{z}$. Here $\mathbf{a}$ is the vector of the basis function coefficients

$(a_0, ..., a_Q)$; **K** is the vector of degrees of interaction for each basis function; **s**, **v**, and **t** are the vectors of signs, variables, and knots used for each basis function, respectively. Similar to GP surrogates, BMARS constructs a posterior distribution over approximations of $g(\mathbf{x})$. Instead of a closed form posterior, however, we have samples from the posterior for $\hat{g}(\mathbf{x})$. The details for the parameter prior specification and the posterior sampling strategy used are found in [34]. Each Monte Carlo sample generated by BMARS represents a MARS model for $g(\mathbf{x})$. To predict response values $\mathbf{z}^*$ for new design points $\mathbf{X}^*$ we evaluate each of the sampled MARS models at $\mathbf{X}^*$ and take the sample mean of responses to be our vector of predicted responses. We use the BASS package in R for fitting the models [35].

3.3.3 Radial Basis Functions

The final surrogate model we consider is the radial basis function model. As with the previous surrogate models, the relationship between the mutual information estimates z_i and the detector configuration coordinate design points $\mathbf{x}_i$ is

$$z_i = g(\mathbf{x}_i) + e_i, e_i \overset{iid}{\sim} N(0, \sigma^2). \tag{21}$$

As with the MARS models, $g(\mathbf{x}_i)$ is modeled as a basis function expansion,

$$\hat{g}(\mathbf{x}_i) = a_0 + \sum_{q=1}^{Q} a_q \phi(||\mathbf{x}_i - \mathbf{x}^q||_2). \tag{22}$$

The M design points $\{\mathbf{x}_1, ..., \mathbf{x}_M\}$ are formed into Q clusters using the K-means algorithm. The point $\mathbf{x}^q$ is the cluster center of the qth cluster. There are several common choices for the basis function, $\phi(\cdot)$; for our analysis, we found the multiquadratic basis function performed the best. The form of the multiquadratic basis function is

$$\phi(||\mathbf{x}_i - \mathbf{x}^q||_2) = \sqrt{||\mathbf{x}_i - \mathbf{x}^q||_2^2 + s^2}. \tag{23}$$

Here s is a hyperparamter that must be selected by the user; we select this parameter using maximum likelihood. The model for $\hat{g}(\mathbf{x}_i)$ is fit by estimating the basis function coefficients $\mathbf{a} = \{a_0, ..., a_Q\}$ using ordinary least squares. We employ an implementation based on the radial basis function implementation by Neto [36].

3.3.4 Surrogate Model Comparisons

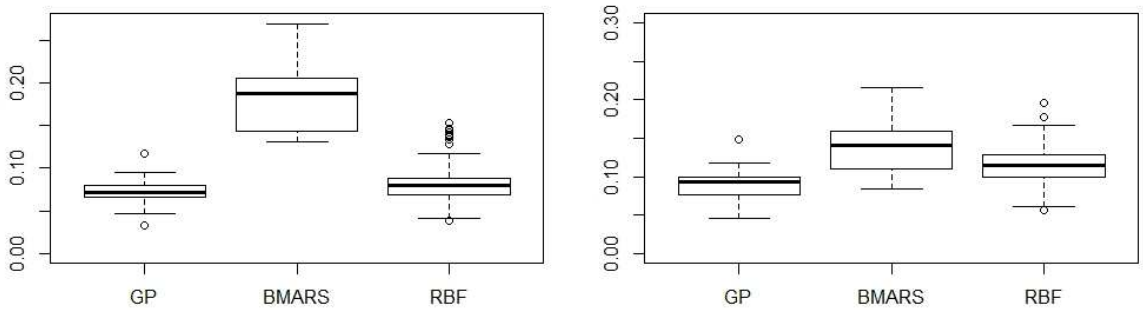
To compare the accuracy of these three surrogate models, we computed a training set of 200 LHD design points, $\{\mathbf{x}_i\}_{i=1}^{M=200}$. Using the data generating process described in Section 2.1, we compute the detector response y_{ij}, for detector j in configuration i, $i = 1, ..., M = 200$, $j = 1, ..., n$, given $N = 1000$ source parameters drawn from $p(\boldsymbol{\theta})$. We do this for $n = 3$ and $n = 5$, corresponding to three and five-detector configurations. The computed response vectors $\mathbf{y}_i$ as well as the 1000 source parameter vectors are used to compute KSG estimates z_i for each configuration. The set $\{z_i, \mathbf{x}_i\}_{i=1}^{M=200}$ is a training set for fitting $\hat{g}(\mathbf{x})$ using each of the surrogate models discussed in Sections 3.3.1–3.3.3. We repeat this process to form a test set of 200 input-output points. We randomly divide the set of 200 test points into 20 subsets of ten points each. For each test set, we use one of the three fitted surrogate models to predict the output for each input in the test set and compute the error between the surrogate model response and the test set response. For each test set, we fit the model and compute the error 15 times to account for the stochastic nature of each surrogate model-fitting process. We do this for all three surrogate models.

We measure surrogate model prediction error using root mean squared error (RMSE),

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{M}(z_i - \hat{z}_i)^2}{M}}, \tag{24}$$

where z_i is the test set mutual information output and $\hat{z}_i$ is the surrogate prediction for z_i.

Figure 3 presents a boxplot comparison of RMSE across the three methods for the three- and five-detector problem. For this application, the RBF and GP models appear to perform better than the BMARS model in both the three-



(a) (b)

FIG. 3: RMSE comparison for surrogate methods: (a) three-detectors and (b) five-detectors

and five-detector cases. For the 20 test sets, we reported the median RMSE of the 15 repeated fits for each of the surrogate model types.

For the three-detector case, we found that for for all test sets, GP and RBF both had a lower median RMSE than BMARS. For 70% of the test sets, GP had a lower median RMSE than RBF. Thus for this problem in the three-detector case, GP fits the best.

For the five-detector case, we found that for 80% of the test sets, RBF had a lower median RMSE than BMARS and for 90% of the test sets, GP had a lower median RMSE than BMARS. For 85% of the test sets GP had a lower RMSE than RBF. Thus, the GP model seems to also be the best choice for the five-detector case in this application.

While the GP model is best for this particular example, this may not be the case when the method is applied in different urban settings. For different applications, we encourage the comparison of different surrogate models to compare accuracy, precision, and computational performance.

4. MOVABLE DETECTOR PLACEMENT EXPERIMENT

To compare the performance of the surrogate based configuration optimization method described in Section 3 and the combinatorial optimization method described in Section 2.3, we compare the results of these two methods in localizing a source in the block of Washington, DC using movable detector placement. Movable detector networks are networks of radiation detectors where each detector can be moved so that the network is easily reconfigurable. We describe how to construct and execute movable detector experiments and metrics that are useful for analyzing the results of these experiments. We will compare the source localization performance of the continuous and combinatorial configuration optimization methods in both the three- and five-detector cases.

4.1 Movable Detector Placement

Because of the difficulty in localizing a source due to varying background and obstructions, obtaining data from only a single detector configuration may be insufficient for precisely localizing the source. One strategy to improve source localization is movable detector placement. After we obtain data from an initial detector configuration and then employ the data along with the framework in Section 2.1 to compute $p(\boldsymbol{\theta}|\mathbf{Y})$, we can find a subsequent detector configuration based on $p(\boldsymbol{\theta}|\mathbf{Y})$ to collect more data. This iterative process of constructing detector configurations, collecting data, and updating the source parameter distribution can be repeated until we obtain an estimate of the source location with sufficient precision.

In this section, we compare the performance of the combinatorial optimization method and the continuous surrogate based optimization method in the movable detector placement problem. For both methods, we start by drawing $N = 1000$ source parameter vectors from the source prior $p(\boldsymbol{\theta})$.

The combinatorial method generates detector response data $\mathbf{y}$ at each of 29 prespecified detector locations for each of the N source parameter vectors. These prespecified locations are depicted in Fig. 1(a) and are the same detector locations used in [13] for combinatorial optimization. We use a genetic algorithm [37] with the KSG mutual

information estimator as the fitness function to perform global optimization over the set of all possible three- and five-detector configurations composed of the 29 detector locations. We compute the detector response for each of the detectors in the selected optimal configuration using the data generating process in Section 2.1 for a true radiation source with location coordinates $\theta_1 = 118.073$ and $\theta_2 = 134.388$ and intensity parameter $\theta_3 = 3.472 \times 10^9$. We use these generated data as well as the Bayesian framework in Section 2.1 to draw MCMC samples from $p(\boldsymbol{\theta}|\mathbf{Y})$ using DRAM MCMC as implemented in the Python library pymcmcstat [38]. We find a subsequent configuration by drawing $N = 1000$ MCMC samples from $p(\boldsymbol{\theta}|\mathbf{Y})$, and then repeating the process of computing detector responses and finding a configuration that optimizes the KSG estimator. We draw samples from the updated posterior distribution using the data generated from the subsequent detector deployment using the same true source. We obtain an initial and subsequent deployment for the three- and five-detector scenarios.

For surrogate based continuous optimization, we generate a LHD $\mathbf{X}$ with $M = 200$ over the configuration space. This design is large enough to fill the configuration space, but is small enough to be approximated by $\tilde{\mathbf{X}}$. We approximate $\mathbf{X}$ with $\tilde{\mathbf{X}}$ where $p = 30$. To generate detector response data, we only generate detector responses at the 30 LHD detectors for each of the 1000 source parameter vectors. Thus, the time required to generate detector response data is comparable to the combinatorial case with 29 detector locations. We employ the 1000 source parameter vectors and the generated detector data to compute the set of KSG mutual information estimators $\mathbf{z}$ for each detector configuration in $\tilde{\mathbf{X}}$. We employ $\mathbf{z}$ and $\tilde{\mathbf{X}}$ to fit a GP surrogate model in the three-detector and five-detector case. We then employ a genetic algorithm [39] using the surrogate model as the fitness function to find a detector configuration that approximately maximizes mutual information. We use data generated from this optimal configuration and the same fixed source used in the combinatorial optimization procedure to draw MCMC samples from $p(\boldsymbol{\theta}|\mathbf{Y})$.

In the subsequent deployment, we sample $N = 1000$ source parameter vectors from $p(\boldsymbol{\theta}|\mathbf{Y})$. For the initial deployment, we used the marginal prior distributions of θ_1 and θ_2 to generate $\mathbf{X}$. In order to generate a LHD that adapts to our updated beliefs about source location, we should generate a LHD using the marginal posterior distributions of θ_1 and θ_2. We do not have these in closed form, but we obtain approximate closed form expressions by fitting two-component Gaussian mixture models to the MCMC samples from the posterior. Under the posterior, θ_1 and θ_2 may no longer be independent, an assumption for the procedure described in Section 3.2. We employ the method described in [27] for generating LHDs using dependent random variables. Since we are primarily interested in configurations containing detector locations that are close to the likely location of the source, we only need to accurately model the reduced portion of the configuration space with detector locations in this small, high-density region. Accurately modeling this reduced space requires fewer generated design points. Thus, we generate an $\mathbf{X}$ with $M = 50$ and $\tilde{\mathbf{X}}$ with $p = 30$. The rest of the procedure is the same way as for the initial deployment except that the search space for the genetic optimization algorithm is restricted to the box whose boundary is formed by the most extreme coordinates from the MCMC sample from the posterior distribution.

Both the combinatorial and continuous surrogate based optimization strategies yield a posterior distribution for the source parameters. In comparing the two approaches for configuration optimization, we compare the quality of the produced posterior distributions. We employ two metrics that quantify accuracy and precision. Maximum a posteriori (MAP) error compares the posterior mode of source location to the true location. The 95% highest-density region (HDR) is the smallest region containing 95% of the probability density of the posterior. For our application a higher-quality posterior distribution has a lower MAP error and a smaller HDR area.

4.2 Simulation Results: Initial Deployment

We compute an initial detector configuration deployment for both three- and five-detector configurations using both the combinatorial and continuous optimization approaches and generate the resulting posterior distributions using data generated from the optimal configurations as described in Section 4.1. We repeat this 30 times for each combination of optimization method and number of detectors to account for variation in the optimal configuration found using genetic algorithm optimization. We compute the MAP error and HDR area metrics for all 30 generated posterior samples.

We compare the distributions of the MAP error and HDR area metrics for the 30 runs by ranking the 30 metrics for the continuous and combinatorial approaches and then counting the number of times the metric for the combinatorial approach exceeds the metric of the same rank for the continuous approach.

Figure 4 shows the boxplots of the HDR area and MAP error for the combinatorial and continuous optimization methods in the three-detector case. For 29 of the 30 ranked MAP errors, the combinatorial method produced a higher error. In 16 of the 30 ranked HDR area metrics, the combinatorial method produced a posterior with higher HDR area. Thus, the continuous method is more accurate than the combinatorial approach and they have comparable precision. The superior accuracy of the continuous surrogate-based optimization method is likely due to its greater flexibility in selecting optimal detector configurations.

The same analysis is carried out for the five-detector case as depicted in Fig. 5. We found that for 17 of the 30 ranked MAP errors, the combinatorial method produced a higher error. Only three of the 30 ranked HDR area metrics, however, were higher for the combinatorial method than for the continuous method. Thus, while the continuous method still tends to be more accurate, it is less precise than the combinatorial metric in the five-detector case. This may be due to the greater flexibility gained by both methods in selecting five detectors instead of three. This would indicate that the continuous method has a greater advantage in situations where we have a limited number of detectors to deploy relative to our search space. Because we are dealing with a higher-dimensional input in the five-detector case but using the same training set size as in the three-detector case, it could be that the surrogate models are less accurate relative to those used in the three-detector case. We did find that the surrogates for the five-detector case had higher MSE for all surrogate modeling methods.

4.3 Simulation Results: Subsequent Deployment

After comparing the results of the two optimization methods for an initial deployment of the detectors, we find subsequent deployments of three- and five-detector configurations using the strategy described in Section 4.1. To find a subsequent deployment, we require a posterior distribution for source location from which to draw N source parameter vectors. From the initial deployment results, we have 30 posterior distributions for each combination of optimization method and configuration size. For each optimization combination, we choose a posterior distribution whose MAP error and HDR area is close to the median MAP error and median HDR area for the 30 posterior distributions so

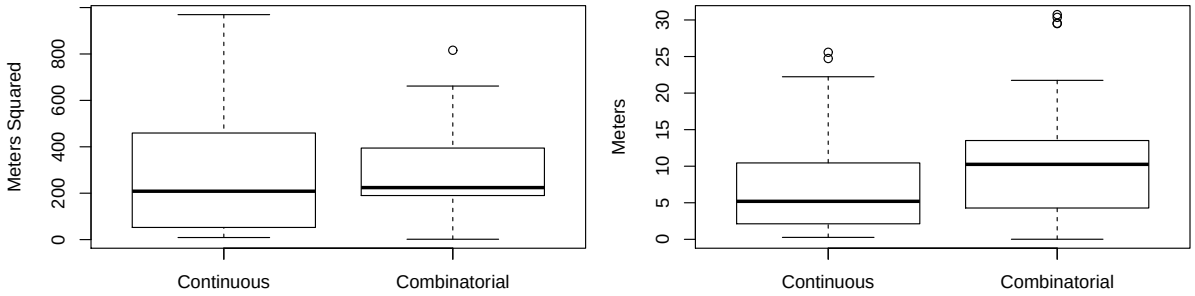


FIG. 4: Comparison of initial posterior distributions in three-detector case. (a) HDR area and (b) MAP error comparisons.

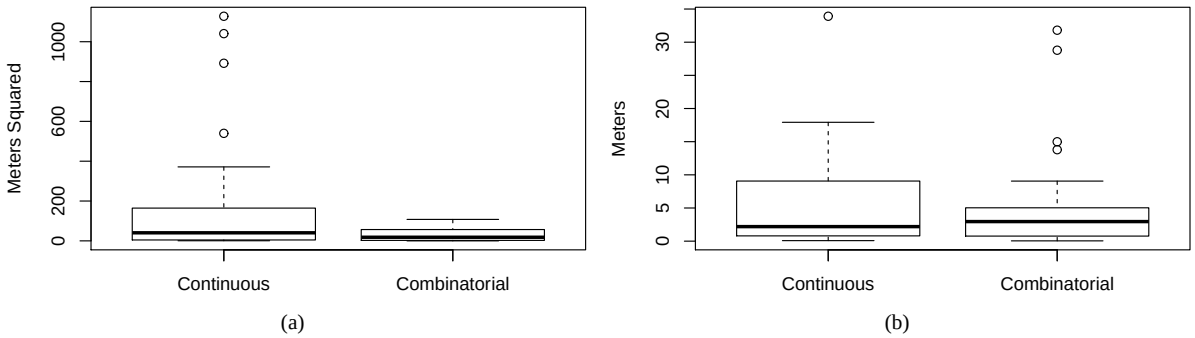


FIG. 5: Comparison of initial posterior distributions in five-detector case. (a) HDR area and (b) MAP error comparisons.

that the posteriors chosen are comparable relative to the optimization method that produced them. Once subsequent optimal configurations have been found, we generate the detector response data for each configuration using the same fixed source as in the initial deployment. We use these generated data to draw MCMC samples from the updated posterior distribution. We repeat this process 30 times for each combination of configuration size and optimization method and compute the associated MAP error and HDR area metrics.

Figure 6 shows the box plots for the MAP error and HDR area for the two methods in the three-detector case. For all 30 runs, the ranked MAP errors and ranked HDR area metrics were larger for the combinatorial method than for the continuous method. This in conjunction with Fig. 6 indicates the superiority across all metrics of the continuous optimization method for subsequent deployments of three detectors.

The strong performance of the continuous method in the subsequent deployment scenario is illustrated in Fig. 7. The combinatorial method can only place detectors at prespecified candidate locations, and thus can only place detectors as close as the closest candidate locations. As illustrated in Fig. 7, the continuous method can select configurations with detectors that are arbitrarily close to the true source location.

In the initial deployment, the search space is the entire city block because the source location prior is uniform across this whole region. In the subsequent deployment, the search space is reduced by the posterior distribution constructed in the initial deployment. Since the combinatorial method employs a set of candidate detector locations fixed for all deployments, the density of candidate detector locations available in the subsequent search space is substantially reduced relative to the initial search space. Thus, the flexibility of the combinatorial optimization approach in constructing configurations is even more limited than it was in the initial deployment. The continuous optimization approach, because it searches the continuous configuration space, automatically adjusts itself to the smaller search space. Resampling the design points using the posterior distribution and rebuilding the surrogate models insures the surrogates are still accurate in this smaller search space. Thus, the continuous optimization approach automatically adapts itself to search spaces of different resolutions.

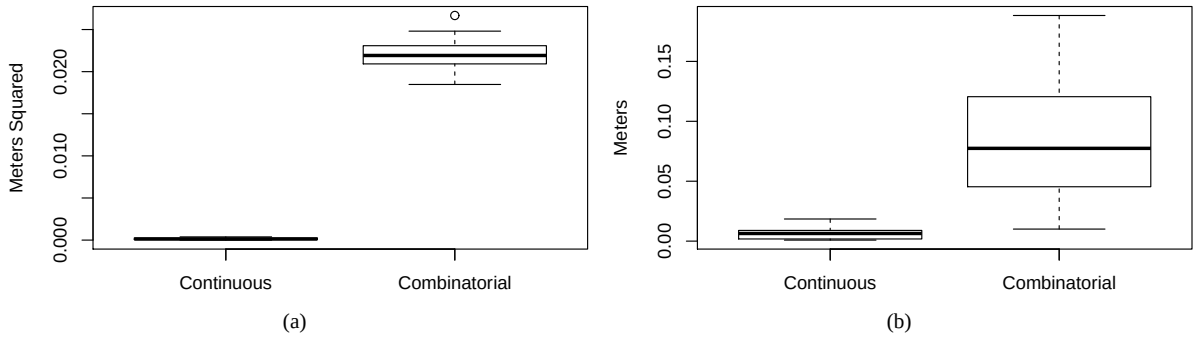


FIG. 6: Comparison of subsequent posterior distributions in three-detector case. (a) HDR area and (b) MAP error comparisons.

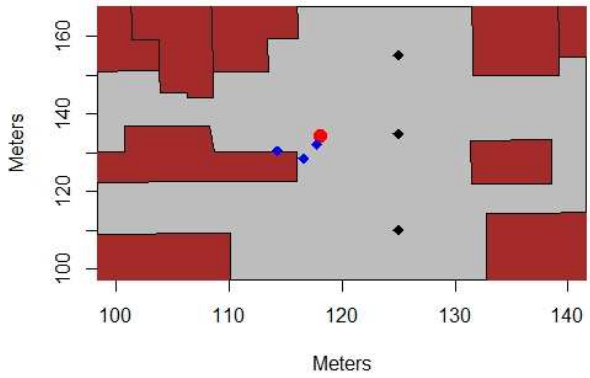


FIG. 7: Subsequent detector deployments for using continuous optimization and combinatorial optimization

Figure 8 displays analogous results for the five-detector case. For all 30 runs, the ranked HDR area metrics and the ranked MAP errors were larger for the combinatorial method than for the continuous method. Again, the continuous optimization method clearly outperforms the combinatorial method. Unlike in the initial deployment, in the subsequent deployment, the continuous optimization method for the five-detector case performs significantly better. Even with five detectors, the candidate location grid used by the combinatorial optimization approach is too coarse to produce the same quality results as the continuous method.

5. CONCLUSION AND FUTURE WORK

Localizing nuclear radiation sources in an urban environment is important for securing urban populations against nuclear hazards resulting from negligence or hostility. This task is particularly challenging in urban environments where obstructions and background radiation make localization difficult. In this paper, we have presented a framework for constructing configurations of radiation detectors that are optimal for performing radiation source localization. This framework optimizes mutual information to select the detector configuration that yields the greatest reduction in uncertainty about source location.

This framework constructs surrogate models of mutual information using Latin hypercube designs. We found that these surrogates are sufficiently accurate for identifying detector configurations that produce data and posterior distributions that are comparable or superior to the combinatorial optimization method. The advantage of this method over a combinatorial optimization method is that it optimizes over the continuous space of detector configurations while remaining computationally tractable because of the efficiency of surrogate models. The greater generality of this surrogate based optimization allows for greater flexibility when selecting optimal detector configurations. This greater flexibility is critically important in large search spaces or domains represented in three dimensions, where sets of candidate detector locations may not be representative of the space of possible detector configurations.

In future work, we will apply Bayesian optimization, a technique that simultaneously estimates and optimizes an objective function for which we only have input-output pairs, to detector configuration optimization. In our case, mutual information between source parameters and detector configuration data is an expensive unknown objective function we wish to optimize. The Bayesian optimization framework is thus a natural choice for solving this optimization problem.

We plan to use this Bayesian optimization framework in three-dimensional urban environments, rather than the two-dimensional urban environment used for illustration in this paper. In this case, the efficiency and flexibility of the continuous optimization method should produce significant advantages over the combinatorial optimization approach without incurring greater computational costs.

ACKNOWLEDGMENTS

This research was supported by the Department of Energy National Nuclear Security Administration (NNSA) under the Award Number DE-NA0002576 through the Consortium for Nonproliferation Enabling Capabilities (CNEC).

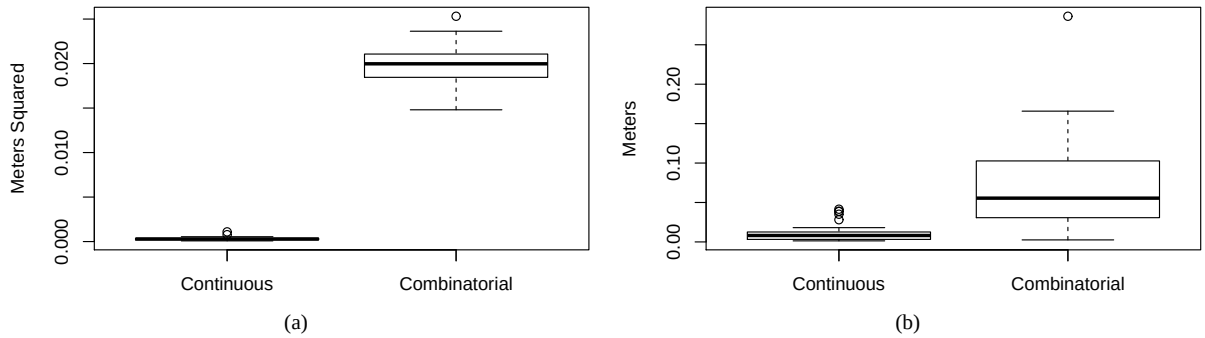


FIG. 8: Comparison of subsequent posterior distributions in five-detector case. (a) HDR area and (b) MAP error comparisons.

The authors would also like to thank Isaac Michaud of Los Alamos National Laboratory for his valuable input and perspective.

REFERENCES

1. Morelande, M., Ristic, B., and Gunatilaka, A., Detection and Parameter Estimation of Multiple Radioactive Sources, in *FUSION 2007–2007 10th Int. Conf. on Information Fusion*, Quebec City, Canada, 2007.
2. Rao, N.S., Shankar, M., Srivathsan, S., Iyengar, S.S., Chin, J.C., Yau, D.K., Yang, Y., and Hou, J.C., Identification of Low-Level Point Radiation Sources Using a Sensor Network, in *Proc. Int. Conf. on Information Processing in Sensor Networks, IPSN 2008*, St Louis, MO, IEEE, pp. 493–504, 2008.
3. Vilim, R. and Klann, R.T., RadTarc: A System for Detecting, Localizing, and Tracking Radioactive Sources in Real Time, *Nucl. Technol.*, **168**(1):61–73, 2009.
4. Chandy, M., Pilotto, C., and McLean, R., Networked Sensing Systems for Detecting People Carrying Radioactive Material, in *Proc. of INSS 2008—5th International Conference on Networked Sensing Systems*, Kanazawa, Japan, IEEE, pp. 148–155, 2008.
5. Cox, J. and Partensky, M.B., Spatial Localization Problem and the Circle of Apollonius, *Phys. Educ.*, arXiv:physics/0701146, 2007.
6. Chin, J.C., Yau, D.K., Rao, N.S., Yang, Y., Ma, C.Y., and Shankar, M., Accurate Localization of Low-Level Radioactive Source under Noise and Measurement Errors, in *SenSys'08—Proc. of the 6th ACM Conf. on Embedded Networked Sensor Systems*, Raleigh, NC, ACM, pp. 183–196, 2008.
7. Baidoo-Williams, H.E., Dasgupta, S., Mudumbai, R., and Bai, E., On the Gradient Descent Localization of Radioactive Sources, *IEEE Signal Process. Lett.*, **20**(11):1046–1049, 2013.
8. Gunatilaka, A., Ristic, B., and Gailis, R., On Localisation of a Radiological Point Source, in *Conf. Proc. of 2007 Information, Decision and Control, IDC*, New Orleans, LA, IEEE, pp. 236–241, 2007.
9. Hite, J., Mattingly, J., Archer, D., Willis, M., Rowe, A., Bray, K., Carter, J., and Ghawaly, J., Localization of a Radioactive Source in an Urban Environment Using Bayesian Metropolis Methods, *Nucl. Instrum. Methods Phys. Res., Sect. A*, **915**:82–93, 2019.
10. Jarman, K.D., Miller, E.A., Wittman, R.S., and Gesh, C.J., Bayesian Radiation Source Localization, *Nucl. Technol.*, **175**(1):326–334, 2011.
11. Ştefănescu, R., Schmidt, K., Hite, J., Smith, R.C., and Mattingly, J., Hybrid Optimization and Bayesian Inference Techniques for a Non-Smooth Radiation Detection Problem, *Int. J. Numer. Methods Eng.*, **111**(10):955–982, 2017.
12. Penny, R.D., Crowley, T.M., Gardner, B.M., Mandell, M.J., Guo, Y., Haas, E.B., Knize, D.J., Kuharski, R.A., Ranta, D., Shyffer, R., Labov, S., Nelson, K., Seilhan, B., and Valentine, J.D., Improved Radiological/Nuclear Source Localization in Variable NORM Background: An MLEM Approach with Segmentation Data, *Nucl. Instrum. Methods Phys. Res., Sect. A*, **784**:319–325, 2015.
13. Michaud, I.J., Simulation-Based Bayesian Experimental Design Using Mutual Information, PhD, North Carolina State University, 2019.
14. Schmidt, K., Uncertainty Quantification for Mixed-Effects Models with Applications in Nuclear Engineering, PhD, North Carolina State University, 2016.
15. Haario, H., Laine, M., Mira, A., and Saksman, E., DRAM: Efficient Adaptive MCMC, *Stat. Comput.*, **16**:339–354, 2006.
16. Briesmeister, J.F., MCNPTM—A General Monte Carlo N-Particle Transport Code, accessed from https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-13709-M, 2000.
17. Swinney, M.W., Peplow, D.E., Nicholson, A.D., and Patton, B.W., NORM Concentration Determination in Common Materials in an Urban Environment, *Trans. Am. Nucl. Soc.*, **114**:635–638, 2016.
18. Horne, S.M., Thoreson, G.G., Theisen, L.A., Mitchell, D.J., Harding, L., and Amai, W.A., GADRAS-DRF 18.5 User's Manual, Tech. Rep., Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA (United States), 2014.
19. Knoll, G.G., *Radiation Detection and Measuremen*, 4th ed., Hoboken, NJ: John Wiley and Sons, 2010.
20. Belghazi, M.I., Baratin, A., Rajeswar, S., Ozair, S., Bengio, Y., Courville, A., and Hjelm, R.D., Mutual Information Neural Estimation, in *ICML Proc. 35th Int. Conf. on Machine Learning*, pp. 531–540, Stockholm, Sweden, 2018.

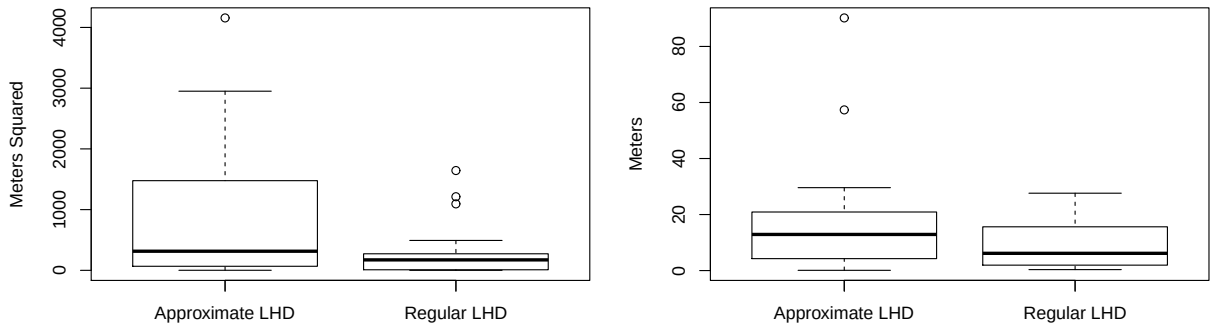
21. Gao, S., Ver Steeg, G., and Galstyan, A., Efficient Estimation of Mutual Information for Strongly Dependent Variables, in *Proc. of Artificial Intelligence and Statistics*, pp. 277–286, San Diego, CA, 2015.
22. Gao, S., Ver Steeg, G., and Galstyan, A., Estimating Mutual Information by Local Gaussian Approximation, in *Uncertainty in Artificial Intelligence, Proc. of the 31st Conf., UAI*, pp. 278–287, Amsterdam, Netherlands, 2015.
23. Kraskov, A., Stögbauer, H., and Grassberger, P., Estimating Mutual Information, *Phys. Rev. E*, **69**:066138, 2004.
24. Han, Z.H., Zhang, K.S., Song, W.P., and Qiao, Z.D., Optimization of Active Flow Control over an Airfoil Using a Surrogate-Management Framework, *J. Aircraft*, **47**(2):603–612, 2010.
25. Queipo, N.V., Haftka, R.T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, K.P., Surrogate-Based Analysis and Optimization, *Prog. Aerosp. Sci.*, **41**(1):1–28, 2005.
26. McKay, M.D., Beckman, R.J., and Conover, W.J., A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, **21**(2):239–245, 1979.
27. Stein, M., Large Sample Properties of Simulations Using Latin Hypercube Sampling, *Technometrics*, **29**(2):143–151, 1987.
28. Santner, T.J., Williams, B.J., and Notz, W.I., *The Design and Analysis of Computer Experiments*, 2nd ed., New York, NY: Springer, 2018.
29. Roustant, O., Ginsbourger, D., and Deville, Y., DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization, *J. Stat. Software*, **51**:1–55, 2012.
30. Friedman, J.H., Multivariate Adaptive Regression Splines, *Ann. Stat.*, **19**(1):1–67, 1991.
31. Banerjee, A., Dunson, D.B., and Tokdar, S.T., Efficient Gaussian Process Regression for Large Datasets, *Biometrika*, **100**(1):75–89, 2013.
32. Østergård, T., Jensen, R.L., and Maagaard, S.E., A Comparison of Six Metamodeling Techniques Applied to Building Performance Simulations, *Appl. Energy*, **211**:89–103, 2018.
33. Denison, D.G., Mallick, B.K., and Smith, A.F., Bayesian MARS, *Stat. Comput.*, **8**:337–346, 1998.
34. Francom, D., Sansó, B., Kupresanin, A., and Johannesson, G., Sensitivity Analysis and Emulation for Functional Data Using Bayesian Adaptive Splines, *Stat. Sin.*, **28**:791–816, 2018.
35. Francom, D. and Sansó, B., BASS: An R Package for Fitting and Performing Sensitivity Analysis of Bayesian Adaptive Spline Surfaces, accessed from https://cran.r-project.org/package=BASS, 2017.
36. Neto, J.P., Radial Basis Functions, accessed from https://github.com/jpneto/Markdowns/tree/master/rbf, 2013.
37. Wolters, M.A., A Genetic Algorithm for Selection of Fixed-Size Subsets with Application to Design Problems, *J. Stat. Software*, **68**:1–18, 2015.
38. Miles, P., pymcmcstat: A Python Package for Bayesian Inference Using Delayed Rejection Adaptive Metropolis, *J. Open Source Software*, **4**(38):1417, 2019.
39. Scrucca, L., GA: A Package for Genetic Algorithms in R, *J. Stat. Software*, **53**:1–37, 2013.

APPENDIX A. SIMULATION USING THE APPROXIMATE LATIN HYPERCUBE DESIGN

In order to compare the effect on source localization of using the approximate LHD, $\tilde{\mathbf{X}}$, instead of $\mathbf{X}$, the conventional LHD generated using the method in [27], we employ a similar procedure to the one employed for comparing the combinatorial and continuous detector configuration optimization methods. We use $\tilde{\mathbf{X}}$ and $\mathbf{X}$ to fit a surrogate between mutual information and detector configuration, and we optimize this surrogate to obtain an optimal detector configuration. We then use this configuration to obtain MCMC samples from $p(\boldsymbol{\theta}|\mathbf{Y})$. We repeat this 30 times. We compare the two procedures by using the HDR area and MAP error metrics.

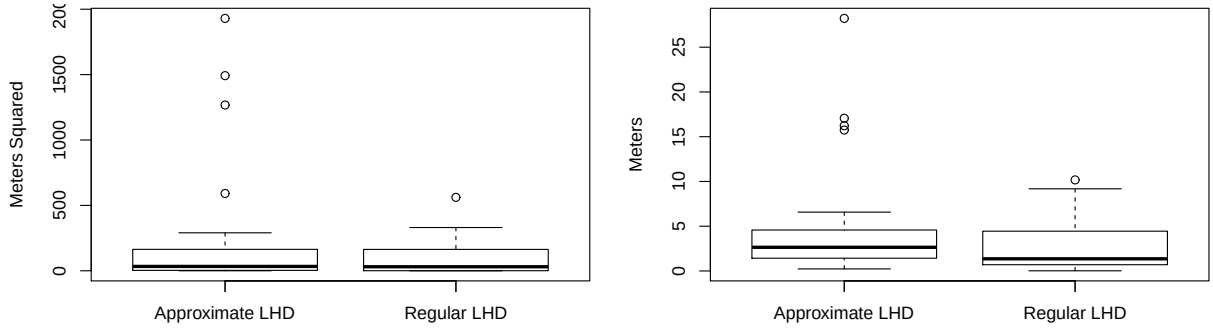
The procedure for generating the MCMC samples when using $\tilde{\mathbf{X}}$ is exactly the same as the procedure described in Section 4.1. The procedure when using $\mathbf{X}$ is also the same substituting $\mathbf{X}$ for $\tilde{\mathbf{X}}$.

Plots of the HDR area and MAP error for the three-detector and five-detector cases for initial detector configuration selection are shown in Figs. A1 and A2. For the three-detector case, the approximate LHD produces less precise results, but with the exception of the outliers produced when using the approximate LHD, the accuracy of the two approaches seems comparable. For the five-detector case, with the exception of the outliers, the HDR area and MAP error metrics appear comparable for the two approaches.



(a) (b)

FIG. A1: Comparison of posterior distributions in three-detector case. (a) HDR area and (b) MAP error comparisons.



(a) (b)

FIG. A2: Comparison of posterior distributions in five-detector case. (a) HDR area and (b) MAP error comparisons.