ABSTRACT

HITE, JASON MICHAEL. Bayesian Parameter Estimation for the Localization of a Radioactive Source in a Heterogeneous Urban Environment. (Under the direction of John Mattingly).

This dissertation presents a new approach to localizing an unknown source of radiation in an urban environment using a distributed detector network. The method employs Bayesian statistical parameter estimation techniques, relying on an approximation for the response of a detector to the source using a simplified model of the underlying transport phenomena based on ray-tracing, combined with a Metropolis-type sampler that is modified to propagate the effect of fixed epistemic uncertainties in the material cross sections of objects in the scene. This method is first demonstrated with simulated measurements generated primarily using a geometry derived from a real city block, with the algorithm able to localize an 8.7 mCi source to within a few meters based on observations from 6 detectors with a 10 s count time. These results also show that the effect of uncertainties in the material cross sections on the precision of the localization is relatively weak, implying that precise estimates of the composition of objects in the scene are not necessary.

Results from these simulated tests were used to design a field measurement campaign conducted in cooperation with Oak Ridge National Laboratory in a scene mimicking a typical urban setting. After extending the simplified detector model to account for the orientation of the detectors, analysis of these measurements shows that the algorithm is able to localize a 37 mCi source to within $\sim$2 m for 30 min measurements in two separate trials, though the results also suggest that count times as short as 5 min are sufficient.

As an extension of this analysis, subsequent work also investigates several questions raised following these experiments. First, a method was developed to automatically detect and classifying anomalies in time-series count rate data caused by unexpected phenomena, e.g. objects moving in the scene. Next, a study of the relationship between the signal-to-noise ratio of the detectors and the precision of the localization was conducted in order to establish rough limits on the strength of the source required to achieve localization. Lastly, the results of localization using an even further simplified transport model for the detector counts that treats all objects in the scene as totally opaque are given. These results show that this detector model is still able to achieve acceptable localization in many cases, with the advantage that it does not require any estimation of the cross section of objects in the scene.

Bayesian Parameter Estimation for the Localization of a Radioactive Source in a Heterogeneous Urban Environment

by
Jason Michael Hite

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Nuclear Engineering

Raleigh, North Carolina

2019

APPROVED BY:

Ralph Smith	Yousry Azmy
Daniel Archer	John Mattingly Chair of Advisory Committee

DEDICATION

This work is dedicated to all of the family, friends, colleagues, and teachers who have supported (or at least, tolerated) me.

BIOGRAPHY

Jason Hite was born in Columbia, South Carolina. He earned his Bachelors degree in Applied Mathematics from North Carolina State University (NCSU) in 2012, and is he currently a PhD candidate in Nuclear Engineering at NCSU. His interest in nuclear engineering started with a summer science education program, and as an undergraduate he did research on reduced-order modeling of neutron cross sections for the DOE Consortium for Advanced Simulation of Light Water Reactors (CASL). He later joined the Consortium for Nonproliferation Enabling Capabilities (CNEC) as a graduate research assistant working on wide-area localization of radiation sources, advised by Prof. John Mattingly. In his limited spare time, Jason enjoys making things and studying Judo.

ACKNOWLEDGEMENTS

This would not have been possible without the input of everyone I have worked with throughout the years. Thanks especially to my committee and my fellow students, as well as my adviser John Mattingly, who stood up for me.

Thanks also to Oak Ridge National Laboratory for hosting the experiments described in this dissertation. The assistance of Dan Archer, Michael Willis, Andrew Rowe, Kayleigh Bray, Jake Carter, and James Ghawaly made these experiments run more smoothly than I thought possible.

TABLE OF CONTENTS

List of Tables . . . . . **vii**

List of Figures . . . . . **viii**

Chapter 1 Introduction **1**
1.1 Motivation and Applications 4
1.2 Prior and Related Work 7
1.3 Novel Contributions 9

Chapter 2 Bayesian Statistics and MCMC **10**
2.1 Bayes' Rule 10
2.2 Markov Chains 11
2.3 Markov Chain Monte Carlo 20
2.3.1 The Metropolis-Hastings Algorithm 21
2.4 Special Considerations 25
2.4.1 Convergence Rate and Burn-in 25

Chapter 3 Methodology for Source Localization **29**
3.1 Markov-Chain Monte Carlo Sampling 30
3.2 Statistical Model for Detector Response 32
3.3 Detector Response Model 33
3.3.1 Effect of Detector Orientation 36
3.4 Propagation of Cross Section Uncertainties 36
3.5 Adaptive Metropolis 37
3.6 Complete Algorithm for Source Localization 38
3.6.1 Relation to Maximum Likelihood Estimation 40
3.6.2 Difficulties with Direct Numerical Optimization 40

Chapter 4 Preliminary Experiments in Simulated Geometry **42**
4.1 Synthetic Model in Urban Geometry 43
4.2 Comparison to High-Fidelity Simulation 45
4.3 Effect of Cross Section Uncertainties on Localization Error 47
4.4 Informing the Design of a Field Experiment 50

Chapter 5 Field Test **52**
5.1 Description of Experiment 52
5.1.1 Detector Placement 54
5.1.2 Detector Characteristics 56
5.1.3 Site Characterization 58
5.2 Results and Analysis 60
5.2.1 Detector orientation 64
5.2.2 Count rate anomalies during measurements 66
5.2.3 Effect of Counting Time on Posterior Error 68

5.2.4 Other Sources of Error . . . 69
5.2.5 Trilateration . . . 70

Chapter 6 Effects of detector anomalies, signal-to-noise ratio, and alternative physics models . . . 72
6.1 Anomaly Classifier . . . 73
6.2 Dependence on Source Intensity . . . 80
6.3 Occluded Detector Model . . . 83

Chapter 7 Conclusions . . . 90
7.1 Summary of Major Results . . . 91

References . . . 93

LIST OF TABLES

Table 3.1 Model parameters. 34

Table 4.1 Predicted source locations and standard deviations for marginal posterior . 49

Table 5.1 Summary of actual source and detector placement during the experiment, including a qualitative description of the visibility of the source for each detector. 56

Table 5.2 Count rates recorded by the detectors. Entries marked with a (∗) denote measurements with anomalous count rates (see Section 5.2.2). Total-to-background ratios and SNRs are adjusted for live time. 61

Table 5.3 Predicted source locations using simple trilateration with error estimates without accounting for detector orientation. 71

Table 5.4 Predicted source locations using simple trilateration with error estimates including detector orientation. All distances are in meters. 71

Table 6.1 Classifier results for Experiment 1. 76

Table 6.2 Classifier results for Experiment 2. Highlighted rows correspond to the anomalous measurements identified in Section 5.2.2. 76

LIST OF FIGURES

Figure 1.1 Example source localization problem, with a detector network distributed over a city block. We wish to determine the location of the source using only count rates measured by the detectors. 2
Figure 1.2 Illustration of trilateration; source is located at the point where circles overlap. 3
Figure 1.3 Illustration of trilateration with uncertainty; source lies somewhere inside the black circle. 4
Figure 1.4 Count rates for various sources of interest, simulated using `GADRAS-DRF` for a 2" × 4" × 16" NaI detector. 6

Figure 2.1 States x_1 and x_2 are reachable from x_0, x_0 is not reachable from x_1 or x_2, x_1 and x_2 form a closed communicating class. 15
Figure 2.2 Graphs depicting reducible and irreducible Markov chains. 16
Figure 2.3 Example of a periodic chain, the example trajectory shows that states x_0 and x_3 are visited with a period of 3. 17
Figure 2.4 Limiting behavior of probability mass in a periodic chain. 18
Figure 2.5 Trace plots showing typical behavior of chains as they approach convergence 27
Figure 2.6 Behavior of Gelman-Rubin scores for chains in Fig. 2.5 28

Figure 3.1 Visual illustration of the ray tracing procedure 35
Figure 3.2 Sum-of-squares error as a function of source location for experiment 1 described in Chapter 5. The peaks are clipped to show detail of the error surface. 41

Figure 4.1 Satellite image of test location, with modeled geometry overlayed in green, as well as source (red) and detector (blue) locations. 44
Figure 4.2 Posterior density for source location in synthetic experiments. 45
Figure 4.3 Model geometry used for simulations using MCNP. 46
Figure 4.4 Posterior density for localization based on simulations with MCNP. 47
Figure 4.5 Posterior density for source location with 5% uncertainty in all cross sections. 48
Figure 4.6 Posterior density for source location with 50% uncertainty in all cross sections. 49

Figure 5.1 Overhead view of experiment site, the two source locations are shown in orange, while the overall experiment area is marked in red (satellite imagery courtesy of Google Maps). 53
Figure 5.2 Detector and source placements for each experiment (satellite imagery courtesy of Google Maps). 54
Figure 5.3 Illustration of the detectors used for the experiment. 57

Figure 5.4 Model geometry of the site, colored by the calculated mean free path. The DGPS reference points are also indicated by the black marks. Note that there are some small disagreements with the objects visible in the satellite imagery, this is due to the satellite photos being taken on a different day than that of the experiments. 59
Figure 5.5 Elevation difference in experiment 2. The source is located just behind the fence visible in the background. 60
Figure 5.6 Foreground and background count rates in both experiments. 62
Figure 5.7 Marginal posterior density for source location **r** in experiment 1 with 50% relative uncertainty in all cross-sections. 63
Figure 5.8 Marginal posterior density for source location **r** in experiment 2 with 50% relative uncertainty in all cross-sections. 64
Figure 5.9 Marginal posterior density for source location **r** in experiment 1 with 50% relative uncertainty in all cross-sections and including detector orientation. 65
Figure 5.10 Marginal posterior density for source location **r** in experiment 2 with 50% relative uncertainty in all cross-sections and including detector orientation. 65
Figure 5.11 Time series of fluctuating count rates recorded by three of the detectors in experiment 2 (blue) versus the average count rate over the entire measurement (green). 67
Figure 5.12 Marginal posterior density for source location **r** in experiment 2 with 50% relative uncertainty in all cross-sections, including detector orientation and with counts from anomalous detectors removed. 68
Figure 5.13 Error in estimated source location versus count time for experiment 1. . . 69

Figure 6.1 Illustration of count rate models for different types of anomalies. 74
Figure 6.2 1A-1, $K_p = 0.002$. Constant. 77
Figure 6.3 1B-5, $K_p = 1.557$. Constant, but borderline. Some linear trend is visible, but is not significant enough to reject a constant trend. 77
Figure 6.4 2A-1, $K_p = 2.321$. Linear. 78
Figure 6.5 2A-3, $K_p = 0.090$. Constant. Despite the clear drop in count rate near the start, classification indicates overall trend is constant. 78
Figure 6.6 2A-4, $K_p = 29.345$. Step. 79
Figure 6.7 2B-6, $K_p = 5.437$. Step. 79
Figure 6.8 Plot of localization error with various source activities. 81
Figure 6.9 Posterior density with varying source intensity. 83
Figure 6.10 Occluded model, 4 detectors with line-of-sight to the source. 86
Figure 6.11 Attenuating model, 4 detectors with line-of-sight to the source. 86
Figure 6.12 Occluded model, 2 detectors with line-of-sight to the source. 87
Figure 6.13 Attenuating model, 2 detectors with line-of-sight to the source. 87
Figure 6.14 Occluded model, 1 detector with line-of-sight to the source. 88
Figure 6.15 Attenuating model, 1 detector with line-of-sight to the source. 88
Figure 6.16 Occluded model, no detectors with line-of-sight to the source. 89
Figure 6.17 Attenuating model, no detectors with line-of-sight to the source. 89

Chapter 1

Introduction

This dissertation describes a new method for locating an unknown source of radiation in a heterogeneous environment. A source is assumed to exist in the search area[1] (referred to as the **scene**) and is observed by a network of radiation detectors, which records counts due to both the source and to natural background radiation present in the scene. Using these measurements, we seek to estimate the location of the source, as well as to quantify the uncertainty in this estimate. This process is often referred to as **source localization**, and I am especially concerned with localizing a source in a scene which is on the scale of a typical city block or larger. Figure 1.1 shows an example, where a detector network is sited in an urban scene with a complex, heterogeneous environment.

[1] It is worth highlighting that I am are pre-supposing the presence of an unknown source in the scene. *Detecting* the presence of a source is a distinct problem with a different set of challenges, and is constrained primarily by tradeoffs between the probability of detecting the presence of a source anywhere in the scene and the rate of false positives. Localization follows the detection of a source, and seeks to determine the exact position of the source with a given precision.

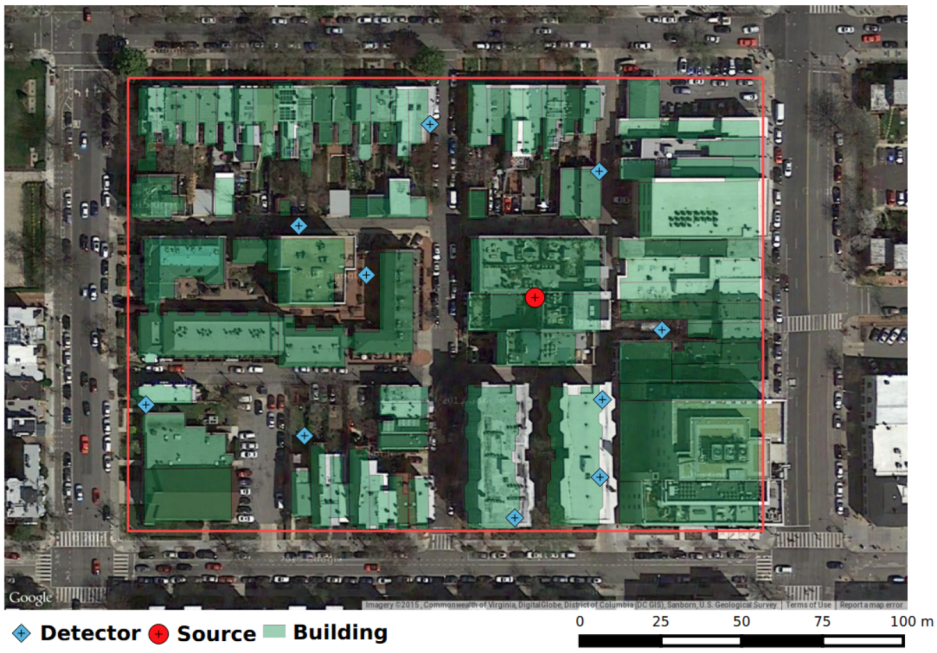


Figure 1.1: Example source localization problem, with a detector network distributed over a city block. We wish to determine the location of the source using only count rates measured by the detectors.

As a simple example, I will demonstrate what is perhaps the simplest method for localizing a source, generally known as *trilateration*. Assume for simplicity that the source and detectors are constrained to a plane and that the environment is a vacuum with no background. In this case, the mean count rate recorded by the i-th detector, d_i due to an isotropic and monoenergetic source of known activity I, located at position $\mathbf{r}$, is given by[2]

$$d_i(\mathbf{r}, I) = \epsilon_i \cdot \frac{I}{4\pi \|\mathbf{r} - \mathbf{r}_i\|_2^2}\,, \tag{1.1}$$

where $\mathbf{r}_i$ is the detector location and ϵ_i is the total efficiency of the detector, with $i = 1, 2, ..., N_D$. Let us denote the distance between the source and detector as $\delta_i = \|\mathbf{r} - \mathbf{r}_i\|_2$, where $\| \circ \|_2$ is the Euclidean norm of a vector. Given a particular $\hat{d}_i$ measured by the detector, we can solve Eq. (1.1) for $\delta_i = 2\sqrt{\pi \hat{d}_i / I \epsilon_i}$. Given several detectors, we can draw a set of circles, each centered on $\mathbf{r}_i$ with radius δ_i; the source is then localized by finding the intersection of these circles (Fig. 1.2). More precisely, we can solve the linear least-squares problem given

[2]We can extend the method shown here to estimate the activity of the source, but the system of equations which arises is non-linear and the resulting analysis is not as illustrative for the purposes of this basic example.

by Eq. (1.2), which will have a unique solution when $N_D \geq 3$ [3] [25].

$$2\left(\mathbf{r}_i - \mathbf{r}_{N_D}\right)^{\mathsf{T}} \mathbf{r} = \|\mathbf{r}_i\|_2^2 - \|\mathbf{r}_{N_D}\|_2^2 - \delta_i^2 + \delta_{N_D}^2 \quad i = 1, 2, \ldots, N_D - 1 \tag{1.2}$$

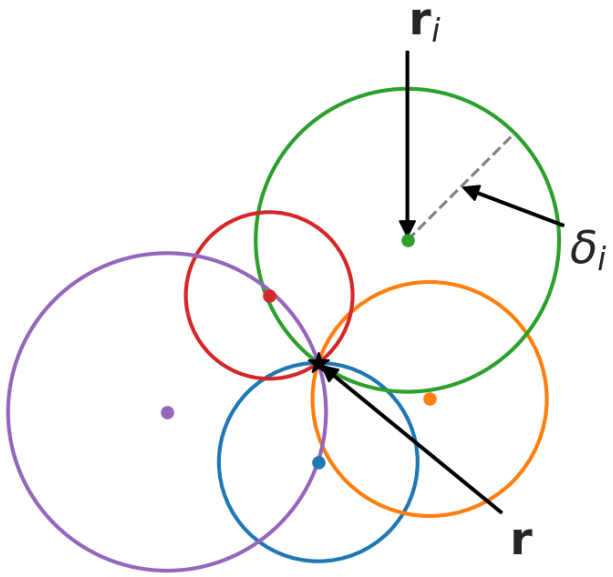


Figure 1.2: Illustration of trilateration; source is located at the point where circles overlap.

However, applications involving radioactive sources are subject to Poisson counting statistics, with the actual counts recorded by the detector varying about the mean value calculated from Eq. (1.1). This is illustrated in Fig. 1.3, where each circle is now less distinct, corresponding to statistical uncertainty in the measured count rates. The circles have no clear point of intersection and instead there is a region where the source lies. The least squares solution (and estimates of uncertainty) can still be found by solving Eq. (1.2), but will be sensitive to the random statistical fluctuations in the measurements.

[3] By calculating the numerical rank of the linear system defined from Eq. (1.2) we can show that the source must lie within the *convex hull* of the detector locations, defined as the union of all triangles with vertices belonging to the set of detector locations. Degenerate cases also exist, for example if all the detectors are placed at the same location.

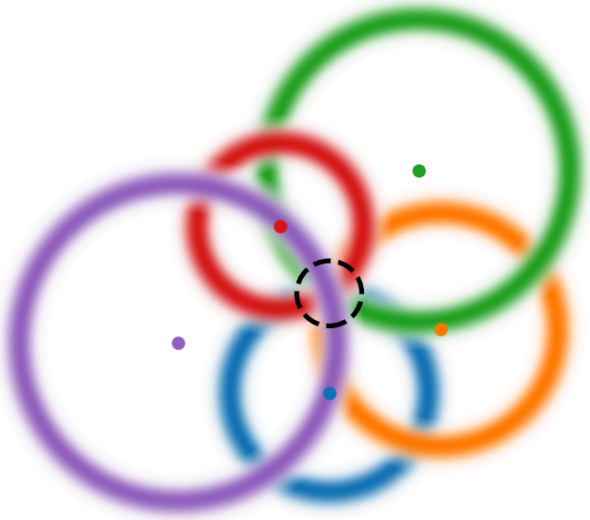


Figure 1.3: Illustration of trilateration with uncertainty; source lies somewhere inside the black circle.

This simple example also ignores the possibility of objects in the scene, which will attenuate the radiation emitted by the source at different rates, as well as the presence of background radiation, which adds random noise to the measurements. Accounting for the presence of attenuators greatly complicates estimation of the detector responses, generally resulting in a nonlinear (as well as non-smooth and non-convex) system of equations which must be solved using advanced iterative techniques [41]. Prior work in this area has tended to focus on source localization in the absence of attenuators (or ignoring their effect), while the method presented in this dissertation incorporates a simplified model for the transport of radiation in a heterogeneous medium.

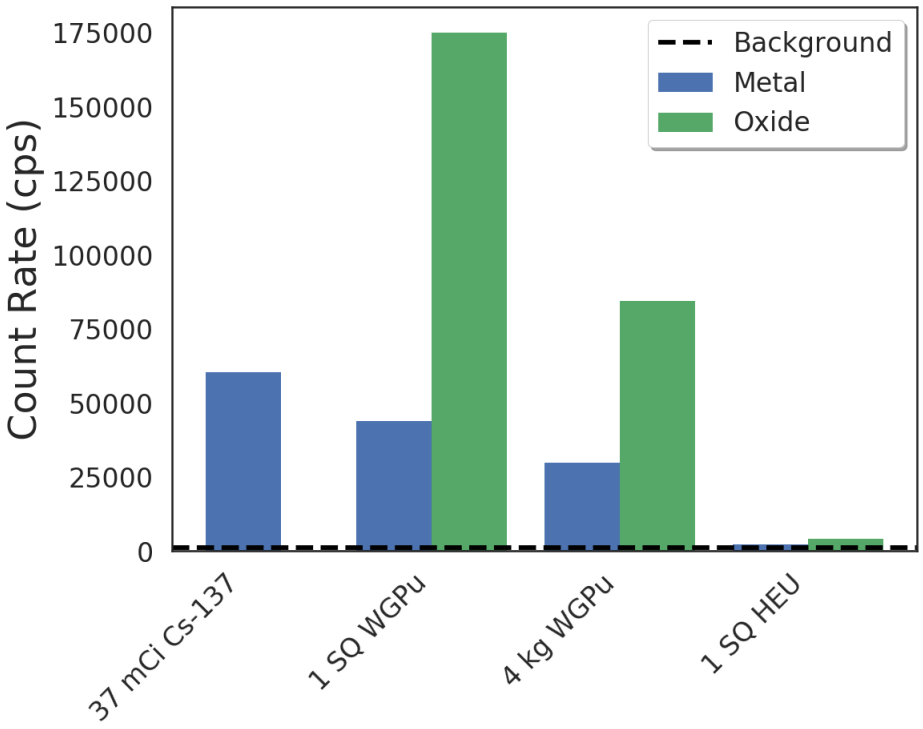
1.1 Motivation and Applications

The focus of this research is on applications in the area of nuclear security, particularly on the recovery of lost or stolen sources of ionizing radiation. A database compiled by the Nuclear Threat Initiative (NTI) [44] lists over 250 incidents related to the diversion or theft of nuclear material between 2005 and 2012 that were classified as proliferation-significant. Additional high-profile incidents such as the 1987 loss of a 1400 Ci cesium-137 source in Goiânia, Brazil [38] or

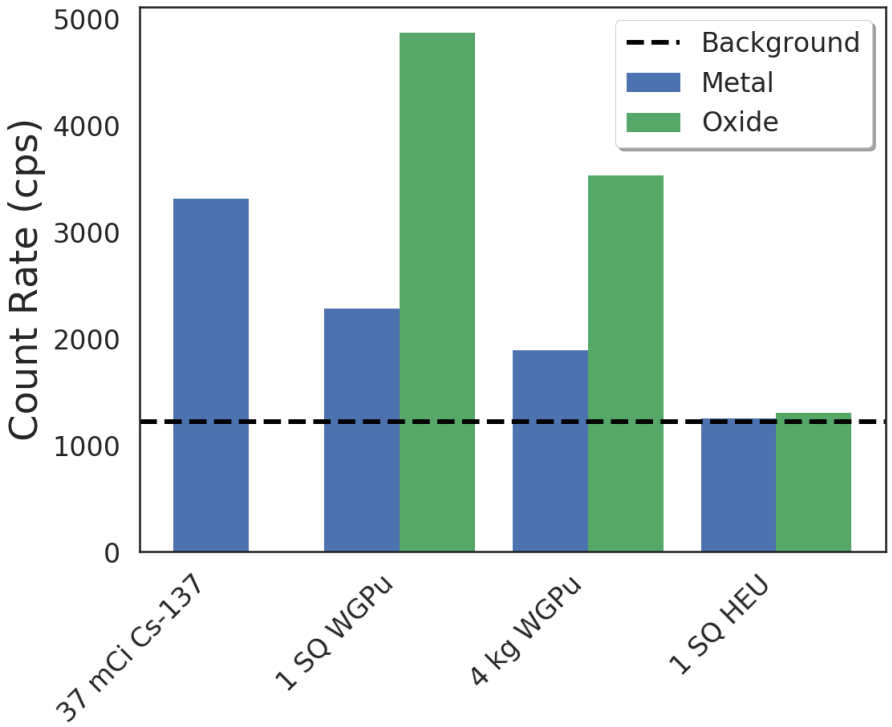
the more recent 2013 theft of a cobalt-60 source used for medical teletherapy near Hueypoxtla, Mexico [27] highlight the need for methods that can aid in the identification and location of radioactive sources over a large area.

My analysis is focused primarily on searching for sources which are significant sources of gamma radiation. This is due to the scale of interest; at the scale of a city block most other forms of radiation will quickly be attenuated to levels that are not measurable with readily-available detectors. The experiments described in subsequent chapters have generally employed millicurie amounts of cesium-137 with a characteristic gamma ray energy of 662 keV, arising from β-decay to an excited state of barium-137 and the subsequent relaxation to the ground state. While cesium-137 is itself a source of concern, with the NTI reporting at least 4 proliferation-significant incidents involving cesium-137 in 2012, I also consider it as a stand-in for other sources of interest, including special nuclear material (SNM). Figure 1.4 plots simulations of the count rates recorded in a 2" $\times$ 4" $\times$ 16" NaI detector at standoff distances of 10 m and 50 m, with simulations performed using `GADRAS-DRF` [31]. Count rates are shown for the 37 mCi source used for the experiments in Chapter 5, as well as for IAEA-defined significant quantities (SQ) of weapons-grade plutonium (WGPu) and highly-enriched uranium (HEU)[4], with all sources simulated without shielding. These result indicate that the test source used produces count rates comparable to 1 SQ of WGPu at both distances, though the activity of a similar significant quantity of HEU is much lower. Note, however, that unlike cesium-137, the gamma spectra of SNM sources are polyenergetic and so direct comparisons between the two are difficult. Nevertheless, the count rates shown in Fig. 1.4 provide a qualitative comparison between the cesium source used in the experiments and SNM.

[4] 1 SQ of WGPu is defined as 8 kg, while an SQ of HEU is 25 kg. These definitions are independent of the exact isotopic composition of the material [20].



(a) 10 m standoff distance



(b) 50 m standoff distance

Figure 1.4: Count rates for various sources of interest, simulated using `GADRAS-DRF` for a 2" × 4" × 16" NaI detector.

1.2 Prior and Related Work

Localizing an emissive source based on the measured signal strength of a network of detectors is a topic that has been widely investigated, with a large variety of algorithms described in the literature. As was briefly mentioned earlier, the simplest method is via **trilateration**, where the signal strength reported by each detector is used to infer a distance from the source to the detector. This creates a set of overlapping circles, each centered on a detector, with radius given by the predicted distance from the source to the corresponding detector. The source is then localized by finding the point of intersection for the set of circles.

This method is perhaps best illustrated in ref. [36], which uses trilateration with a $1/\text{distance}^2$ model for the detector count rates to estimate the source location, and a variance-weighted average to estimate source activity. The authors demonstrate that the algorithm is able to accurately localize a 0.95 µCi cesium source, though the distance scale of the experiments is small ($\sim$1 m) and the scene is free of obstructions. Trilateration is also well-known outside of the nuclear engineering community, with ref. [25] presenting an overview of applications to the localization of nodes in a wireless network; it also forms the basis of the Global Positioning System [23, Ch. 2].

Many other authors present work that can be interpreted in a similar framework to that of ref. [36], where localization is fundamentally achieved using a $1/\text{distance}^2$ model for the count rates recorded by each detector as a function of source location and intensity, see for example refs. [5, 32, 33, 34]. One shortcoming is that these formulations do not attempt to account for the presence of attenuating material in the scene, with the $1/\text{distance}^2$ model equivalent to assuming the scene is a homogeneous or a vacuum. More cluttered scenes may however contain objects such as buildings or vehicles, which significantly affect the count rates measured by detectors– an effect that is not accounted for by $1/\text{distance}^2$ models. To my knowledge, work on localization using models that account for variable attenuation is relatively sparse. An example that does attempt to incorporate attenuation is ref. [46], which discusses the effect of attenuators on the detector response in a localization scheme based on maximum likelihood estimation (discussed shortly), however it is unclear from the text how the authors calculated the corresponding attenuation coefficients.

Ref. [22] presents a more sophisticated approach using high-fidelity adjoint radiation transport to compute detector responses. Of particular interest, the authors compare the results using adjoint transport to those using an uncollided flux model equivalent to the one described in Section 3.3 for a cargo monitoring scenario. In contrast to the results we will see in following chapters, the authors conclude that the uncollided flux model is insufficient and produces erroneous localization. This is likely due to the scale of their test problem, which is on the order of 2 m×2 m, a scale on which the source and detector may not be treated as points and where the

contribution of scattered gammas to the detector response is substantial.

Another distinguishing characteristic of existing localization algorithms is in the statistical parameter estimation techniques used. The classic solution to the statistical parameter estimation problem is the **maximum likelihood estimate** (MLE), which is defined as the collection of source parameters that maximizes the statistical likelihood of observing the measured detector responses. MLE methods are very popular in many applications due to their robustness and the reliability of modern numerical optimization algorithms. They have seen wide application, with ref. [33] demonstrating the use of an MLE for a detector response model dependent on the source activity and the distance from the source to the detector. Recent work has also applied MLE to the complementary problem of estimating the spatial and temporal distribution of background and subsequently using this estimate to detect an anomalous source using a mobile detector network [26].

Bayesian techniques have also seen a growth in usage in recent years, enabled largely by the increasing computational resources available to researchers. Ref. [19] provides an early example, deriving a **maximum a posteriori** estimator (conceptually similar to an MLE, but formulated in a Bayesian context) of the source location intended for use in real-time tracking. Using a Gaussian noise state model, the authors show that their estimator is able to track a cesium-137 source in real-time as it moves throughout a $10\,\mathrm{m}\times15\,\mathrm{m}$ room. The results presented therein show the algorithm has good performance, being able to track the source to within a few inches as it is moved around, but does not include a detailed description of the experiments. Most notably, the source activity is not stated, though from the distance scale involved it would be reasonable to assume a source in the $100\,\mu\mathrm{Ci}$ to $500\,\mu\mathrm{Ci}$ range was used.

There also exists a separate class of Bayesian algorithms collectively known as **particle filters**. This approach can be understood as a type of genetic algorithm, where several "particles" are distributed throughout the search space, with each particle representing a candidate source location. Measured data is used to determine the fitness of each particle by computing its likelihood and the system is evolved to produce a new generation of particles; over several generations, the particles tend to cluster around the true source location. This method has proven especially useful in online monitoring, with ref. [35] demonstrating the algorithm applied to a vehicle portal monitoring scenario.

Other authors have chosen to focus on different aspects of the localization problem. One of the most interesting examples is given in refs. [32, 33], which describe the use of Bayesian model selection techniques to distinguish the presence of multiple sources in the scene and localize them via an algorithm known as **reversible-jump Markov chain Monte Carlo**. This algorithm can be understood as an extension of the algorithm described in Chapter 2, with an added model selection step that allows the chain to dynamically jump between models which include differing numbers of sources. In ref. [32], the authors demonstrate that their

method is able to detect and characterize 4 sources simultaneously without prior knowledge of the number of sources, though the presence of attenuating objects in the scene is not considered.

Ref. [42] also provides valuable insight into some of the practical aspects of operating a detector network for source localization. This work proposes methods for dealing with the spatial errors present in measurements recorded by a detector network, as well as for variations in the efficiency of individual detectors. Such errors arise from issues of time synchronization between detectors as well as errors in the position information derived from GPS measurements, and are expected to occur in any real-world measurements. These practical considerations are important to field deployment of any wide-area localization algorithm, particularly when using low-cost mobile radiation detectors where high-accuracy GPS location information is not available, as well as in sensor networks that rely on multiple different types of detector.

1.3 Novel Contributions

Section 1.2 shows that the use of Bayesian parameter estimation techniques, including the use of methods based on Markov chain Monte Carlo, are not new, though such work is at least fairly recent. The primary novelty of this work lies in the combination of such methods with a detector model that accounts for variable attenuation of gamma rays through a heterogeneous scene. Further, this dissertation also provides a methodology to account for arbitrary uncertainties in the composition of objects in the scene and their propagation onto the resulting prediction of the source location– a feature which is, to the best of my knowledge, entirely unique. This capability is extremely important to the practical deployment of such an algorithm, given the impracticality of precisely estimating the composition of all objects in a complex scene *in situ*.

Chapter 2

Bayesian Statistics and MCMC

In this chapter, I will outline the mathematical theory underlying the method for source localization that is the focus of this dissertation. I will begin with a description of Bayesian statistics while noting some of the difficulties involved with their direct application, followed by a description of the key properties of Markov chains. I will then show how these properties can be used to construct a procedure known as **Markov chain Monte Carlo** (MCMC), which enables the practical usage of Bayesian techniques, and conclude with some discussion of their convergence. In subsequent chapters I will specialize this methodology to the problem of source localization, as well as reexamine the material discussed here in more qualitative terms.

Note that the material presented in this chapter is not novel, and is included for the sake of completeness of presentation. In particular, I have drawn heavily on ref. [40] and ref. [24] in compiling the material shown here, while providing my own perspective on the portions relevant to the problem of source localization.

2.1 Bayes' Rule

Imagine that we are presented with a random variable X, with corresponding sample space $\mathcal{S}$. X is characterized by an *unknown* probability density function (PDF) $p(x)$. Here, $p(x)$ is a probability measure defined on event space $\mathcal{F}$, where $\mathcal{F}$ is the σ-algebra of $\mathcal{S}$ (the set of all subsets of $\mathcal{S}$ measurable under p). We will assume that we are provided with realizations of X, say $\{x_i\}_{i=0}^{N-1}$, which can be observed without knowing $p(x)$. Intuitively you might suspect that if we continue to observe realizations of X we may be able to reconstruct p in a form that is representative of the behavior we have observed.

Bayesian statistics provides a systematic approach for performing this reconstruction based on the interpretation of probability as representing our "state of knowledge" about a random variable. Each x_i is interpreted as providing information about $p(x)$ and for each observation

we "update" our state of knowledge to incorporate the new information. First, consider that we begin with an initial state of knowledge encoded as a PDF, which I will refer to as the **prior** or **prior density** and denote as $p_0(x)$. We seek the **posterior density**, $p(x \mid x_0)$, which is the conditional probability of x given that we have observed x_0, incorporating the observation x_0 into our knowledge about X.

To determine how we might compute $p(x \mid x_0)$, first recall one of the fundamental axioms of probability. Consider two arbitrary events $A, B \in \mathcal{S}$, then [3, Eq. 16.39]:

$$\begin{aligned} \Pr[A \wedge B] &= \Pr[A] \cdot \Pr[B \mid A] \\ &= \Pr[B] \cdot \Pr[A \mid B] \ . \end{aligned} \tag{2.1}$$

That is to say, the probability that A *and* B occur is the product of the probability of A and the conditional probability of B given A, i.e. the probability B occurs given that A has occurred (and vice-versa). We can equate the right-hand sides of Eq. (2.1) and solve for $\Pr[A|B]$:

$$\begin{aligned} \Pr[A] \cdot \Pr[B|A] &= \Pr[B] \cdot \Pr[A|B] \\ \Pr[A|B] &= \frac{\Pr[A] \cdot \Pr[B|A]}{\Pr[B]} \ . \end{aligned} \tag{2.2}$$

Equation (2.2) is often known as **Bayes' rule** and provides exactly what we are after– the conditional probability of observing A given that we have previously observed B. Expressed in terms of probability densities and substituting our prior and posterior for X:

$$p(x|x_0) = \frac{p(x_0|x) \cdot p_0(x)}{\int_{\mathcal{S}} p(x_0|x) \cdot p_0(x)\, dx} \ . \tag{2.3}$$

2.2 Markov Chains

The expression for the posterior density given in Eq. (2.3) can be calculated directly at least in principle, but historically the difficulty of integrating over the full event space has limited its direct application. Recent years have seen developments that enable direct evaluation of Eq. (2.3) (see [40, Ch. 10, 11]), however it has been more commonly addressed via Markov-chain Monte Carlo (MCMC). Instead of evaluating Eq. (2.3), MCMC methods perform a random walk which is constructed in such a way as to generate samples from the posterior distribution.

We will discuss MCMC in Section 2.3, but first we must cover some of the preliminary theory of Markov chains. We will start by defining one of the most fundamental concepts in this chapter, that of a **stochastic process**. This is an indexed and ordered sequence of random variables, where each random variable takes on values from a common state space.

Definition 2.2.1 (Stochastic process). Let $\mathcal{X} = \{X_t \,;\, t \in \mathcal{T}\}$ denote a sequence of random variables on a common probability space $(\mathcal{S}, \mathcal{F}, p)$, where $\mathcal{S}$ is the **sample space** of X, $\mathcal{F}$ is the σ-algebra of $\mathcal{S}$ (known as the **event space** of X), p is a probability measure over $\mathcal{F}$, and $\mathcal{T}$ is a totally ordered index set. The sequence $\mathcal{X}$ is known as a **stochastic process** or **random process**. If the index set $\mathcal{T}$ is taken as a subset of the integers then $\mathcal{X}$ is known as a **discrete time** stochastic process, whereas if it is taken as a sub-interval of the real numbers then $\mathcal{X}$ is referred to as a **continuous time** stochastic process. Similarly, if the state space $\mathcal{S}$ is finite or countably infinite then $\mathcal{X}$ is known as a **discrete state** stochastic process, otherwise it is typically referred to as a **continuous state** stochastic process.

We call each index in the set $\mathcal{T}$ a **step** in the process; this notation was selected because it is most common that these steps correspond to a procedure where we sequentially observe the evolution of a random system in time. In this context we typically refer to the actual value observed at step $t \in \mathcal{T}$, $x_t \in \mathcal{S}$ (a realization of X_t) as the **state** of $\mathcal{X}$ at step t, while $\mathcal{S}$ is the **state space** of $\mathcal{X}$.

In the interest of simplicity, we will limit our discussion here to *finite* state and *countable* time stochastic processes, where $\mathcal{S}$ is a finite set and $\mathcal{T}$ is countable. Similar results hold in in continuous time and countably infinite state spaces but the analysis is significantly more involved, so we will rely on the theory of discrete processes to illustrate the concepts [39]. We are also primarily interested in one particular type of stochastic process, the **Markov chain**. Loosely speaking, Markov chains are stochastic processes that lack "memory"– their state at time t depends *only* on the state at time $t-1$. More formally:

Definition 2.2.2 (Markov Chain). Let $\mathcal{X} = \{X_t \,;\, t \in \mathcal{T}\}$ be a stochastic process on a given probability space $(\mathcal{S}, \mathcal{F}, p)$. We consider here only the case that $\mathcal{X}$ is a discrete time and finite state stochastic process. $\mathcal{X}$ is called a **Markov chain** if it possesses the **Markov property**, namely that, for all $t \in \mathcal{T}$:

$$\Pr\left[X_t = x_t \,|\, X_0 = x_0, X_1 = x_1, \ldots, X_{t-1} = x_{t-1}\right] = \Pr\left[X_t = x_t \,|\, X_{t-1} = x_{t-1}\right] . \tag{2.4}$$

A Markov chain is a random sequence characterized by a state space $\mathcal{S}$, which is the common event space of the sequence of random variables, as well as a set of **transition probabilities**,

$g_{ij} := \Pr[X_{t+1} = x_j \mid X_t = x_i]$. These g_{ij}'s may in general depend on t but I will assume it does not, which is referred to as a **time-homogeneous Markov chain**. Intuitively, this implies that at any point we can "reset" the chain and begin again from any particular state, since

$$\begin{aligned} g_{ij} :&= \Pr[X_{t+1} = x_j \mid X_t = x_i] \\ &= \Pr[X_1 = x_j \mid X_0 = x_i] \,. \end{aligned} \tag{2.5}$$

Since we have assumed finite $\mathcal{S}$ we can form the matrix of g_{ij}'s, $\mathbf{G} = [g_{ij}]$, which is called the **transition kernel** of the chain. If we let $\mathbf{p}^{(0)}$ be a row vector[1] representing the probability mass function of X_0, then we can calculate $\mathbf{p}^{(1)} = \mathbf{p}^{(0)}\mathbf{G}$ to determine the probability mass function of X_1. If we further define the k-step transition probability $g_{ij}^k := \Pr[X_{t+k} = x_i \mid X_t = x_j]$, with corresponding $\mathbf{G}^k$, then we can calculate the probability mass function for any step X_t as follows:

$$\begin{aligned} \mathbf{p}^{(t)} &= \mathbf{p}^{(t-1)}\mathbf{G} \\ &= \left(\mathbf{p}^{(t-2)}\mathbf{G}\right)\mathbf{G} \\ &\dots \\ &= \mathbf{p}^{(0)}\mathbf{G}^t \,. \end{aligned} \tag{2.6}$$

We are most interested in the properties of the chain relating to the evolution of the probability mass as t is advanced. These are dictated primarily by the transition probabilities, hence we will study the relationship between $\mathbf{G}$ and $\mathbf{p}^{(t)}$ in some detail. Our main item of interest is the **stationary distribution** of the chain, which (if it exists) is a distribution that is a fixed point of $\mathbf{G}$.

Definition 2.2.3 (Stationary Distribution). For a Markov chain with state space $\mathcal{S}$ and transition kernel $\mathbf{G}$, a given probability mass function $\mathbf{p}$ is said to be a **stationary distribution** if it is an invariant probability measure under multiplication with $\mathbf{G}$, i.e. $\mathbf{pG} = \mathbf{p}$ and $\mathbf{p}$ is a valid probability distribution.

We are also interested in the limiting behavior of the chain and how $\mathbf{p}^{(t)}$ evolves as $t \to \infty$. Consider the sequence of state distributions for the chain, $\mathbf{p}^{(0)}, \mathbf{p}^{(1)}, \dots, \mathbf{p}^{(t)}, \dots$, with

$$\boldsymbol{\pi} := \lim_{t\to\infty} \mathbf{p}^{(t)} \,, \tag{2.7}$$

[1] Common notation is to represent probability densities as row vectors, hence the left-multiplication. I maintain this convention for compatibility with other references.

where convergence is understood to be in the weak sense (convergence in distribution). The existence of this limit is not guaranteed, though when it does exist the following relations must hold [40, Sec. 4.6]:

$$\begin{aligned}\boldsymbol{\pi} &= \lim_{t\to\infty} \mathbf{p}^{(t)} \\ &= \lim_{t\to\infty} \mathbf{p}^{(0)}\mathbf{G}^{(t)} \\ &= \lim_{t\to\infty} \mathbf{p}^{(0)}\mathbf{G}^{(t+1)} \\ &= \left(\lim_{t\to\infty} \mathbf{p}^{(0)}\mathbf{G}^{(t)}\right)\mathbf{G} \\ &= \boldsymbol{\pi}\mathbf{G}\,.\end{aligned} \tag{2.8}$$

Thus if the limit exists then $\boldsymbol{\pi}$ is also a stationary distribution of the chain. As mentioned, this limit is not guaranteed to exist and even when it does the resulting stationary distribution may not be unique. The remainder of our analysis will be spent establishing when the limit in Eq. (2.7) exists and when it converges to a *unique* stationary distribution. We will later exploit this relationship in Section 2.3 to construct a sampling process that is able to produce samples from the stationary distribution of the chain, given that it has been run long enough to converge.

To this end, we now introduce several key concepts for the analysis of Markov chains that we will use to characterize circumstances when the chain converges and when the resulting stationary distribution is unique. First is that of **reachability**:

Definition 2.2.4 (Reachability and communicating classes). For two states $x_i, x_j \in \mathcal{S}$, x_i is said to be **reachable** from x_j if there exists a positive integer K such that:

$$\Pr\left[X_{K+1} = x_j | X_0 = x_i\right] > 0\,. \tag{2.9}$$

For the discrete case, this implies $g_{ij}^K > 0$.

If state x_j is reachable from state x_i, they are said to **communicate**, denoted $x_i \to x_j$. If two states are mutually reachable it is often denoted $x_i \leftrightarrow x_j$. A subset of the state space $\mathcal{C} \subseteq \mathcal{S}$ is called a **communicating class** if $x_i, x_j \in \mathcal{C}$ implies $x_i \leftrightarrow x_j$. A communicating class is called **closed** if the probability of transitioning from state $x_i \in \mathcal{C}$ to any state $x_j \in \mathcal{S} \setminus \mathcal{C}$ is zero (that is, it is impossible to leave the class).

It is common to visualize discrete Markov chains as directed graphs, where nodes represent states and edges correspond to transition probabilities (an edge between x_i and x_j has weight

equal to g_{ij}). Figure 2.1 uses this technique to provide an illustrative example of Definition 2.2.4.

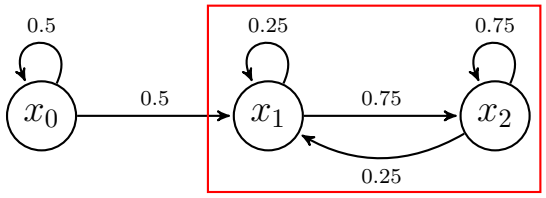
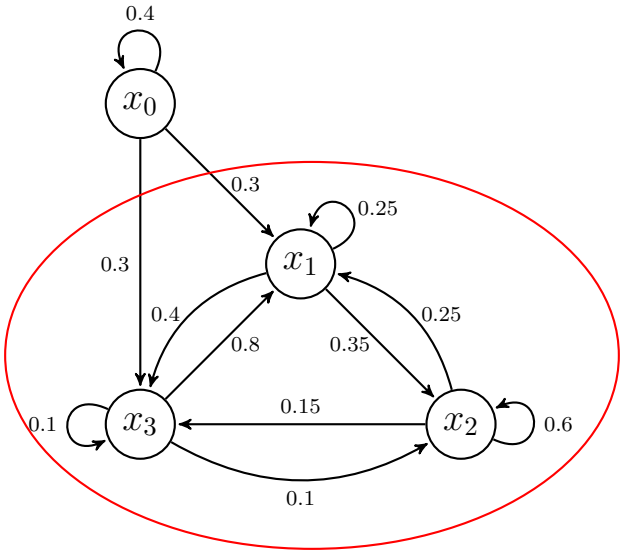


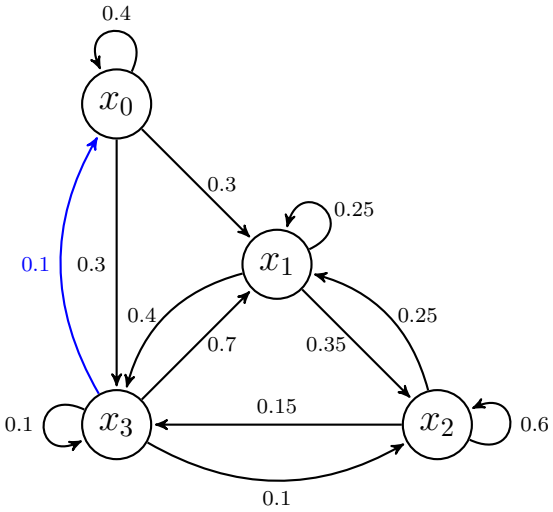
Figure 2.1: States x_1 and x_2 are reachable from x_0, x_0 is not reachable from x_1 or x_2, x_1 and x_2 form a closed communicating class.

When all states are mutually reachable, the chain is said to be **irreducible** (otherwise it is **reducible**). The reason for this naming is depicted in Fig. 2.2– once the chain enters the closed subcycle highlighted in red in Fig. 2.2a it becomes trapped and the chain is effectively equivalent to the subcycle. Since it is inevitable that the chain will eventually enter this closed subcycle as $t \to \infty$, it should be intuitive that the long term behavior of the chain is related to its reducibility. In contrast, the inclusion of a single return path as in Fig. 2.2b makes the entire chain irreducbile since there is now a nonzero probability of visiting any node.

Definition 2.2.5 (Irreducible). A Markov chain is said to be **irreducible** if, for every $x_i, x_j \in \mathcal{S}$, x_j is reachable from x_i, i.e. $x_i \leftrightarrow x_j$ (equivalently, $\mathcal{S}$ is its smallest possible closed communicating class). Otherwise it is **reducible**.



(a) Chain contains a closed subcycle (red), so it is *reducible*.



(b) Inclusion of a single return path (blue) makes the entire chain *irreducible*.

Figure 2.2: Graphs depicting reducible and irreducible Markov chains.

Our next consideration is the case that the chain contains pairs of states that are revisited at regular intervals, at which point it is said to be **periodic**. More formally, define **period** of a chain as follows:

Definition 2.2.6 (Period of a Markov chain). The **period** of a Markov chain, L is

$$\begin{aligned} L &:= \gcd\left\{l \; ; \; \Pr\left[X_l = x_i | X_0 = x_i\right] > 0\right\} \\ &= \gcd\left\{l \; ; \; g_{ii}^l > 0\right\} . \end{aligned} \tag{2.10}$$

For $L > 1$, the chain is said to be **periodic**; for $L = 1$ the chain is called **aperiodic**.

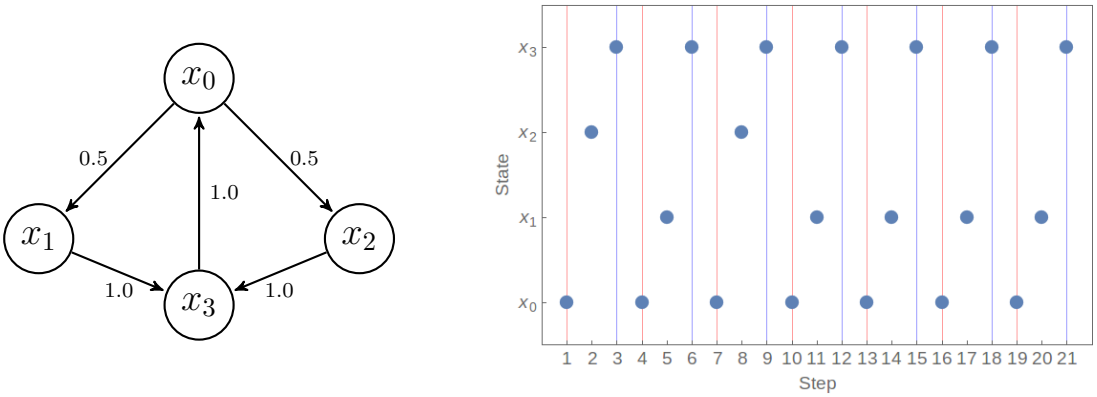


Figure 2.3: Example of a periodic chain, the example trajectory shows that states x_0 and x_3 are visited with a period of 3.

An example of a periodic chain is given in Fig. 2.3, where the transition probabilities are constructed in such a way as to ensure that two of the nodes will be visited at recurring intervals. If the chain contains periodic cycles then the limit in Eq. (2.7) will clearly not converge, hence we will require that a chain be aperiodic as a necessary condition for convergence [40]. The chain shown in Fig. 2.3 actually has a stationary distribution $\boldsymbol{\pi} = [\,\frac{1}{4}\;\frac{1}{4}\;\frac{1}{4}\;\frac{1}{4}\,]$, but we can see in Fig. 2.4 that the probability mass oscillates and does not converge to $\boldsymbol{\pi}$.

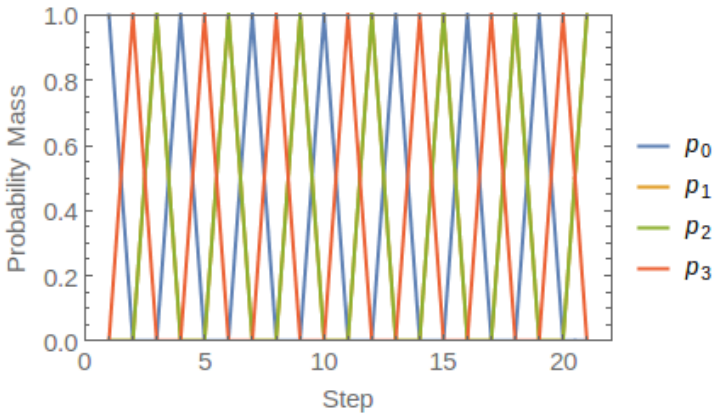


Figure 2.4: Limiting behavior of probability mass in a periodic chain.

Combined, these two properties establish when a chain has a unique stationary distribution, which also coincides with its limiting distribution:

Theorem 2.2.1 (Necessary and sufficient conditions for existence of and convergence to a unique stationary distribution). A Markov chain with transition kernel $\mathbf{G}$ that is both aperiodic and irreducible possesses a unique stationary distribution $\boldsymbol{\pi}$. Further, $\boldsymbol{\pi}$ is also the unique limiting distribution for the chain, that is

$$\lim_{t\to\infty} \mathbf{p}^{(t)} = \boldsymbol{\pi}\,,$$

for every initial probability distribution $\mathbf{p}^{(0)}$.

Proof. The classic proof for this theorem in the finite case is based on Perron's theorem for positive matrices [30, Sec. 8.2] and can be found in multiple texts, e.g. ref. [40, Thm. 4.61][2]. □

As a final consideration, note that while Theorem 2.2.1 establishes the conditions when the stationary distribution exists and is convergent, it does not provide a convenient means to actually construct this distribution. We will instead define a specific subclass of Markov chains called **reversible Markov chains**. These are chains that satisfy the **detailed balance**

[2]Note also that a more general version of this theorem also holds for countably infinite state spaces if one also adds the condition of **positive recurrence** for all states. Loosely speaking, this requires that, on average, the chain visits any given state within a finite number of steps. Ref. [24, Thm. 21.12] establishes that a positive recurrent, aperiodic and irreducible Markov chain with countably infinite state space converges to a unique stationary distribution $\boldsymbol{\pi}(x)$.

condition of Definition 2.2.7 and are called reversible because their definition implies that their behavior is independent of the "direction" that steps are taken (e.g., t increasing or t decreasing). Proposition 2.2.1 establishes that irreducible and aperiodic chains satisfying detailed balance will converge to a unique stationary distribution[3]. In Section 2.3, we will exploit this fact to construct a Markov chain that is guaranteed to converge uniquely to the posterior distribution we seek in Eq. (2.2).

Definition 2.2.7 (Reversible Markov chain). A Markov chain with transition kernel $\mathbf{G}$ (and corresponding transition probabilities g_{ij}) is said to be **reversible** if there exists a probability distribution[4]

$$\mathbf{p} = \begin{bmatrix} p_0 & p_1 & \dots & p_{|\mathcal{S}|-1} \end{bmatrix}, \tag{2.11}$$

such that, for all steps t and all states $x_i, x_j \in \mathcal{S}$, the chain satisfies the **detailed balance** condition:

$$p_i \cdot g_{ij} = p_j \cdot g_{ji}\,. \tag{2.12}$$

---

[3]This relationship is not symmetric, there exist irreducible and aperiodic chains with a unique stationary distribution that do not obey detailed balance, however detailed balance is a sufficient condition for our purposes.

[4]$|\mathcal{S}|$ denotes the cardinality of $\mathcal{S}$.

Proposition 2.2.1 (Alternative conditions for existence of and convergence to a unique stationary distribution). An irreducible, aperiodic and reversible Markov chain possesses a unique stationary distribution $\boldsymbol{\pi}$. $\boldsymbol{\pi}$ is also the unique limiting distribution of the chain.

Proof. Reversibility of the chain implies that there exists a probability distribution $\boldsymbol{\pi}$ that satisfies the criteria of detailed balance. Summing probabilities, we have:

$$\begin{aligned}\sum_i \pi_i \cdot g_{ij} &= \sum_i \pi_j \cdot g_{ji} \\ &= \pi_j \sum_i g_{ji} \\ &= \pi_j \,,\end{aligned}$$

hence $\boldsymbol{\pi}$ is a stationary distribution of the chain. Theorem 2.2.1 shows that the stationary distribution of an aperiodic and irreducbile chain is unique and is also the unique limiting distribution of the chain. □

2.3 Markov Chain Monte Carlo

The idea of using Markov chains constructed in such a way as to have a specific stationary distribution, a process known as **Markov chain Monte Carlo** (MCMC), can be traced to the same origin as standard Monte Carlo techniques: the Manhattan Project. The first article describing what we now know as an MCMC algorithm is generally credited to Metropolis *et. al* [29], and describes a random walk method for computing expectations of the Boltzmann distribution. In 1970, the original statement of the algorithm was generalized by Hastings in ref. [15], an article which is usually considered as the progenitor of all modern MCMC methods.

Later authors would continue to develop the theory underlying MCMC, but by-and-large it would languish for over 30 years, due mostly to the limited computational resources available to researchers during this period. In 1990, Gelfand and Smith published *Sampling-based approaches to calculating marginal densities* [7], which proved to be the genesis of a massive wave of applications using MCMC. It was also realized that MCMC techniques were applicable to the computation of Bayesian posterior densities of the form discussed in Section 2.1, sparking a similar revolution in the practical usage of Bayesian statistics, which had also suffered from the limited computational resources available up to that point. It is from this lineage that the methods outlined in this dissertation originate.

Note: The information in this brief historical overview is sourced primarily from ref. [37]. Interested readers should refer to that article for a more detailed overview of the history of MCMC.

2.3.1 The Metropolis-Hastings Algorithm

With the preliminaries out of the way, we now turn to constructing a procedure that will allow us to draw samples from the posterior distribution in Eq. (2.3), called the **Metropolis-Hastings** algorithm (MH). To see how the MH algorithm comes about, we will begin with a Markov chain that is irreducible, aperiodic and reversible. We know from Proposition 2.2.1 that this chain must possess a stationary distribution, which we have taken to calling $\boldsymbol{\pi}$. We can rearrange the detailed balance condition of Definition 2.2.7 to find:

$$\begin{aligned} \pi_i \cdot \Pr\left[X_1 = x_j \,|\, X_0 = x_i\right] &= \pi_j \cdot \Pr\left[X_1 = x_i \,|\, X_0 = x_j\right] \\ \frac{\Pr\left[X_1 = x_j \,|\, X_0 = x_0\right]}{\Pr\left[X_1 = x_i \,|\, X_0 = x_j\right]} &= \frac{\pi_j}{\pi_i} \\ \frac{g_{ij}}{g_{ji}} &= \frac{\pi_j}{\pi_i}\,. \end{aligned} \tag{2.13}$$

We assume that the transition probabilities can be factored into the product of a **proposal probability** $Q(x_i \,|\, x_j)$ and an **acceptance function** $\alpha(x_i, x_j)$, i.e.

$$g_{ij} = Q(x_i \,|\, x_j) \cdot \alpha(x_i, x_j)\,. \tag{2.14}$$

Inserting this into Eq. (2.13) yields:

$$\begin{aligned} \frac{Q(x_i \,|\, x_j)}{Q(x_j \,|\, x_i)} \frac{\alpha(x_i, x_j)}{\alpha(x_j, x_i)} &= \frac{\pi_j}{\pi_i} \\ \therefore \quad \frac{\alpha(x_i, x_j)}{\alpha(x_j, x_i)} &= \frac{\pi_j}{\pi_i} \frac{Q(x_j \,|\, x_i)}{Q(x_i \,|\, x_j)}\,. \end{aligned} \tag{2.15}$$

For the MH algorithm, choose $\alpha(x_i, x_j)$ as

$$\alpha(x_i, x_j) = \min\left\{1, \frac{\pi_j}{\pi_i} \frac{Q(x_j \,|\, x_i)}{Q(x_i \,|\, x_j)}\right\}\,. \tag{2.16}$$

I have described the MH algorithm as a **rejection sampler**; at each step we propose a new value by randomly sampling it from Q. We then compute the **Metropolis ratio** via Eq. (2.16) and if $\alpha = 1$ we automatically accept the proposed value. Otherwise, we reject the proposed value with probability $1 - \alpha$ and begin again. This procedure is detailed in Algorithm 1.

Algorithm 1 The Metropolis-Hastings Algorithm

Sample x_{t+1} from π, given current sample x_t and proposal distribution $Q(x \mid x_t)$.

procedure METROPOLIS($x_t; \pi, Q(x \mid x_t)$)
  $x^* \leftsquigarrow Q(x, x_t)$ ▷ Draw candidate x^* from Q
  $\alpha_1 \leftarrow \frac{\pi(x^*)}{\pi(x_t)}$ ▷ Compute Metropolis ratio[5]
  $\alpha_2 \leftarrow \frac{Q(x_t \mid x^*)}{Q(x^* \mid x_t)}$ ▷ Correction for asymmetry in Q
  $\alpha \leftarrow \alpha_1 \cdot \alpha_2$
  if $\alpha \geq 1$ **then**
    $x_{t+1} \leftarrow x^*$ ▷ Increase in probability - accept unconditionally
    return x_{t+1}
  else
    $\beta \leftsquigarrow \mathcal{U}(0, 1)$
    if $\beta \leq \alpha$ **then** ▷ Accept with probability proportional to α
      $x_{t+1} \leftarrow x^*$; Accept x^*
    else
      Reject x^*; Restart
    end if
  end if
end procedure

Proposition 2.3.1. The sequence $\mathcal{X} = \{x_t\}_{t=1}^{T}$ generated by the procedure in Algorithm 1 forms a Markov chain.

Proof. It should be clear that Algorithm 1 forms a random process. Let X_t denote the random variable corresponding to the state at time t. Slightly abusing notation, also let Q_t refer to the random variable with probability distribution $Q(x \mid x_t)$. Consider the probability that Algorithm 1 jumps from state x_t at step t to x_{t+1} at step $t+1$:

$$\Pr[X_{t+1} = x_{t+1} | X_t = x_t] = \Pr[X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_1 = x_1] \cdot \Pr[\text{Sample } x_{t+1} \text{ from } Q_t] \cdot \Pr[\text{Accept } x_{t+1}] . \quad (2.17)$$

It is also clear from the statement of Algorithm 1 that $\Pr[\text{Sample } x_{t+1} \text{ from } Q_t]$ and $\Pr[\text{Accept } x_{t+1}]$ are independent of $X_{t-1}, \ldots, X_1$ since α depends only on evaluating π and Q pointwise. We

[5] $\pi(x)$ is a slight abuse of notation - in the discrete case it denotes the entry of $\boldsymbol{\pi}$ corresponding to the probability of state x.

will use induction to show that the remaining term is independent of the states before step t:

- *Base case*

  Assume the process begins at initial state x_0 for $t = 0$ and moves to state x_1 at $t = 1$. By Eq. (2.17),

  $$\Pr[X_1 = x_1 \mid X_0 = x_0] = \Pr[X_0 = x_0] \cdot \Pr[\text{Sample } x_1 \text{ from } Q_t] \cdot \Pr[\text{Accept } x_1] \ , \quad (2.18)$$

  which obviously depends only on x_0

- *Inductive case*

  For the general case, we wish to show:

  $$\Pr[X_t = x_t \mid X_{t-1} = x_t, \ldots, X_0 = x_0] = \Pr[X_t = x_t \mid X_{t-1} = x_{t-1}] \stackrel{?}{\Rightarrow}$$
  $$\Pr[X_{t+1} = x_{t+1} \mid X_t = x_t, \ldots, X_0 = x_0] = \Pr[X_{t+1} = x_{t+1} \mid X_t = x_t] \ . \quad (2.19)$$

  Applying Eq. (2.17):

  $$\Pr[X_{t+1} = x_{t+1} \mid X_t = x_t, \ldots, X_1 = x_1] = \Pr[X_t = x_t \mid X_{t-1} = x_{t-1}, \ldots, X_1 = x_1]$$
  $$\cdot \Pr[\text{Sample } x_2 \text{ from } Q_t] \cdot \Pr[\text{Accept } x_2] \ . \quad (2.20)$$

  We have assumed $\Pr[X_t = x_t \mid X_{t-1} = x_t, \ldots, X_1 = x_1] = \Pr[X_t = x_t \mid X_{t-1} = x_{t-1}]$, hence the conclusion follows immediately.

□

It remains to show that Algorithm 1 produces samples from the stationary distribution of the chain, which established in Proposition 2.3.2. Note that for Proposition 2.3.2 to hold, we also require that the chain be aperiodic and irreducible, which suggests certain restrictions on acceptable proposal distributions. Specifically, the matrix $\mathbf{G}$ of transition probabilities defined from Eq. (2.14) must represent an irreducible and aperiodic stochastic process. Ref. [40, Thm. 4.58-4.60] establishes that if $\mathbf{G}$ strictly positive, then it is reversible and aperiodic. If we assume that $\alpha(x_i, x_j)$ defined in Eq. (2.16) is strictly positive, our primary requirement is that the proposal distribution $Q(x_i|x_j)$ is strictly positive at all points in the domain ($Q(x_i|x_j) > 0$ for all x_i, x_j). This is obviously a very mild restriction, hence we are free to use almost any proposal distribution in Algorithm 1.

Proposition 2.3.2. Algorithm 1 converges to the unique stationary distribution $\boldsymbol{\pi}$.

Proof. Observe that the transition probability in Algorithm 1 can be expressed as follows:

$$\begin{aligned} g_{ij} &= \Pr\left[X_j = x_j | X_i = x_i\right] \\ &= Q(x_j \,|\, x_i) \cdot \min\left\{1, \frac{\pi_j}{\pi_i} \frac{Q(x_i \,|\, x_j)}{Q(x_j \,|\, x_i)}\right\} . \end{aligned} \tag{2.21}$$

We proceed by showing that the detailed balance condition of Definition 2.2.7 is satisfied for $\boldsymbol{\pi}$: for every i, j,

$$\pi_i \cdot g_{ij} \stackrel{?}{=} \pi_j \cdot g_{ji} \,.$$

Recall the identity:

$$a \cdot \min\left\{1, \frac{b}{a}\right\} = \min\{a, b\} = b \cdot \min\left\{1, \frac{a}{b}\right\} ,$$

and hence

$$\begin{gathered} \pi_i \cdot Q(x_j, x_i) \cdot \min\left\{1, \frac{\pi_j}{\pi_i} \frac{Q(x_i \,|\, x_j)}{Q(x_j \,|\, x_i)}\right\} = \pi_j \cdot Q(x_i, x_j) \cdot \min\left\{1, \frac{\pi_i}{\pi_j} \frac{Q(x_j \,|\, x_i)}{Q(x_i \,|\, x_j)}\right\} \\ \pi_i \cdot g_{ij} = \pi_j \cdot g_{ji} \,. \end{gathered} \tag{2.22}$$

The conclusion then follows from Theorem 2.2.1. □

Algorithm 1 describes a scheme for sampling from some arbitrary distribution $\pi(x)$ by sampling from another distribution, Q, and then evaluating π pointwise. While potentially useful in its own regard, we still have not addressed how it might be applied to evaluating Bayes rule. The difficulty of evaluating Eq. (2.3) lies in the normalization term, $\int_{\mathcal{S}} p(x_0|x) \cdot p_0(x)\, dx$, which requires integration over the entire parameter space. Direct computation of this integral typically suffers from a variety of issues, foremost being that it is often numerically challenging.

Instead of evaluating Eq. (2.3) directly, consider that instead of $\pi(x)$, we substitute some function F in Algorithm 1 that is *proportional to* $\pi(x)$:

$$F(x) = \epsilon \cdot \pi(x) \,, \tag{2.23}$$

for some unknown constant $\epsilon \neq 0$.

Proposition 2.3.3. Replacing π with F in the calculation of α_1 in Algorithm 1 does not change the stationary distribution.

Proof. The following relation holds from the same identity used in Proposition 2.3.2:

$$\begin{aligned}
F(x_i) \cdot Q(x_j|x_i) \cdot \min\left\{1, \frac{F(x_j)Q(x_i|x_j)}{F(x_i)Q(x_j|x_i)}\right\} &= F(x_j) \cdot Q(x_i|x_j) \cdot \min\left\{1, \frac{F(x_i)Q(x_j|x_i)}{F(x_j)Q(x_i|x_j)}\right\} \\
\not{\epsilon} \cdot \pi(x_i) \cdot Q(x_j|x_i) \cdot \min\left\{1, \frac{\not{\epsilon}\pi(x_j)Q(x_i|x_j)}{\not{\epsilon}\pi(x_i)Q(x_j|x_i)}\right\} &= \not{\epsilon} \cdot \pi(x_j) \cdot Q(x_i|x_j) \cdot \min\left\{1, \frac{\not{\epsilon}\pi(x_i)Q(x_j|x_i)}{\not{\epsilon}\pi(x_j)Q(x_i|x_j)}\right\} \\
\pi(x_i) \cdot Q(x_j|x_i) \cdot \min\left\{1, \frac{\pi(x_j)Q(x_i|x_j)}{\pi(x_i)Q(x_j|x_i)}\right\} &= \pi(x_j) \cdot Q(x_i|x_j) \cdot \min\left\{1, \frac{\pi(x_i)Q(x_j|x_i)}{\pi(x_j)Q(x_i|x_j)}\right\}.
\end{aligned} \tag{2.24}$$

Thus $\pi(x)$ satisfies the conditions for detailed balance and is the stationary distribution for the chain. □

Proposition 2.3.3 shows that if we are able to construct a function that is proportional to $\pi(x)$ then we can produce a sequence of samples from π, even when the proportionality constant is unknown. In Chapter 3 we will examine how to construct such a function in a way that allows us to draw samples from the posterior distribution of Eq. (2.3) while avoiding the computation of the costly normalization term.

2.4 Special Considerations

In this section we will discuss two additional topics relating to the practical application of MCMC methods. We will briefly examine the convergence rate for the sample chains and the associated difficulties of assessing convergence, followed by an outline of an extension of Algorithm 1 that improved performance. Owing to the depth of these topics, we will only briefly summarize these issues and I will refer the reader to appropriate literature for further information.

2.4.1 Convergence Rate and Burn-in

We have seen that Algorithm 1 is guaranteed to converge to its stationary distribution from any initial distribution under fairly weak assumptions, however what is not guaranteed is the *rate* at which it converges. In general it is in fact impossible to rigorously assess convergence and instead it is typical to rely on heuristic monitoring of the chains combined with statistical analysis of the samples.

To provide some intuition of how to judge convergence, we will consider a few examples by examining their **trace plots**, which plot the time step versus the sampled state. For a chain

that is properly converged to the stationary distribution, we expect that the samples are being drawn from a distribution that is by definition independent of the sample number.

Figure 2.5a shows the expected behavior when a chain has converged– samples are being drawn from a constant distribution. In contrast, Fig. 2.5b shows a chain that has not yet converged, with a clear trend to the sample trajectory[6], while Fig. 2.5c illustrates the behavior as the chain reaches convergence (around sample number 8000). Visual analysis of trace plots is often used for heuristically assessing the convergence of a chain, with the chain being judged to have converged when the sample trajectories appear similar to Fig. 2.5a. Figure 2.5d indicates a potential pitfall of this method (and with any global method)– the chain can become trapped in a local minimum and appear to have converged, only to abruptly jump to a different minimum. Despite this, visual inspection is likely the most common approach used in practice and, when combined with specific knowledge of the problem at hand, is reasonably effective.

[6] A **trajectory** is simply a sequence of realizations drawn from a particular stochastic process.

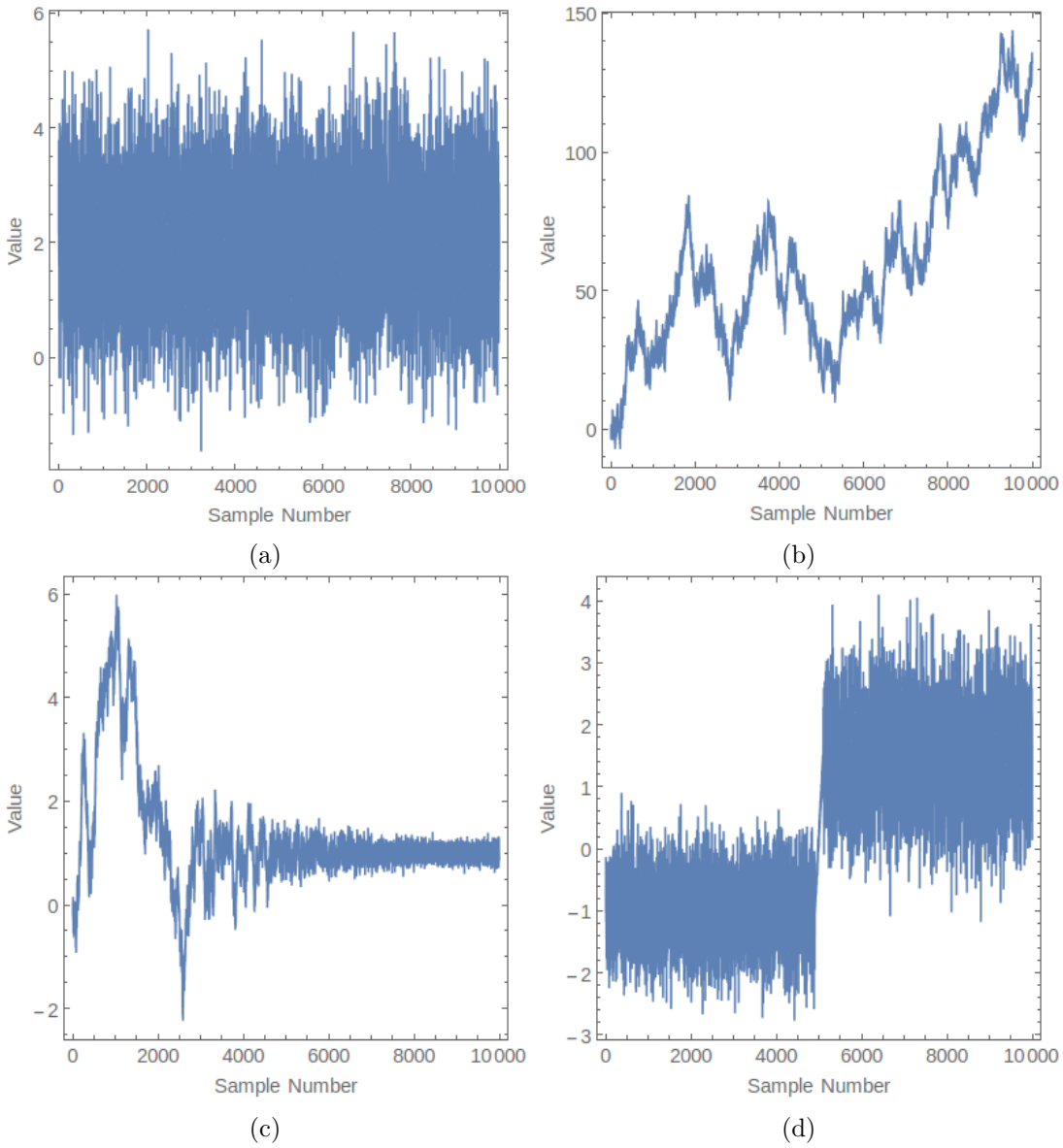
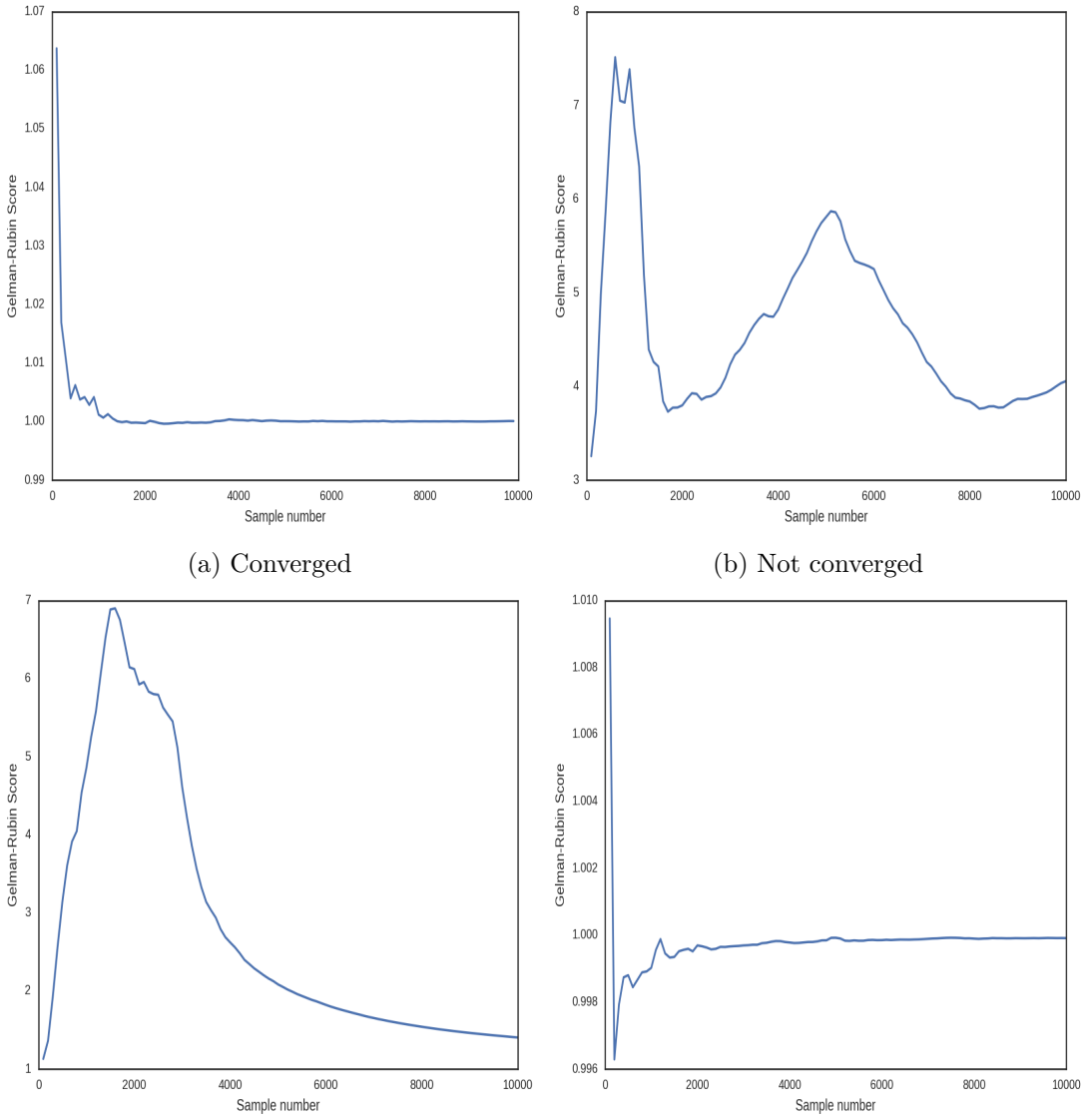


Figure 2.5: Trace plots showing typical behavior of chains as they approach convergence

A more quantitative approach is the **Gelman-Rubin test**, which provides a metric that can be used to determine convergence [10]. This metric is computed by examining several independent trajectories drawn from the same chain with random initial states and subsequently comparing the intra-chain variances. These variances are used to compute a **potential scale reduction factor**, $\hat{R}$. If enough samples have been drawn that the independent chains have

mixed (forgotten their starting states and reached the stationary distribution) then $\hat{R}$ will have a value near 1; otherwise the chains have not converged. A common heuristic is that chains have converged when $\hat{R} < 1.1$ [11].



(a) Converged

(b) Not converged

(c) Converges around 8000 samples

(d) Pathological case, local extrema

Figure 2.6: Behavior of Gelman-Rubin scores for chains in Fig. 2.5

Chapter 3

Methodology for Source Localization

In this chapter, I will formulate the problem of estimating the *true* source location $\mathbf{r}_{\text{true}}$ and intensity I_{true} in the context of Bayesian parameter estimation methods. Assume that we have a network of N_D detectors with known and fixed positions. The response of this detector network is considered to be a vector-valued random variable $\mathcal{D}$, where we assume that each $\mathcal{D}_i$, corresponding to the counts measured by the i-th detector ($1 \leq i \leq N_D$), is an independent Poisson-distributed random variable. We are provided with a vector of count data that is measured in the field by the detector network, denoted $\mathbf{D}$, which is a single realization of $\mathcal{D}$. We then seek to compute the **posterior density** $P(\mathbf{r}, I \,|\, \mathbf{D})$ (henceforth, the **posterior**), which is the probability that a given $(\mathbf{r}, I)$ is the true source location and intensity, conditioned on the observations $\mathbf{D}$. It is desirable that this method also account for the presence of heterogeneous attenuators in the scene, hence it must include a dependence on the composition of objects in the scene and their effect on the count rates measured by the detectors.

The posterior can be expressed directly via Bayes' formula as a normalized product of a **prior density** $P_0(\mathbf{r}, I)$ and a **likelihood function** $\mathcal{L}_{\mathbf{D}}(\mathbf{r}, I\,;\,\Sigma)$ [40]:

$$P(\mathbf{r}, I \,|\, \mathbf{D}) = \frac{\mathcal{L}_{\mathbf{D}}(\mathbf{r}, I\,;\,\Sigma) \cdot P_0(\mathbf{r}, I)}{\int_{I_{\min}}^{I_{\max}} \int_{\mathbb{X}} \mathcal{L}_{\mathbf{D}}(\mathbf{r}', I'\,;\,\Sigma) \cdot P_0(\mathbf{r}', I')\, d\mathbf{r}'\, dI'}\,. \tag{3.1}$$

$P_0(\mathbf{r}, I)$ is a probability distribution representing our knowledge about the source location and intensity before any measurements are recorded, while $\mathbb{X}$ and $(I_{\min}, I_{\max})$ denote the limits for the position and intensity of the source. In all of the experiments described in this dissertation I will use a uniform prior distribution, corresponding to the assumption that initially we only know a general region of interest where the source is located and nothing more. Such a prior is usually referred to as **uninformative** since it assumes that all points in the search space are equally probable. $\mathcal{L}_{\mathbf{D}}(\mathbf{r}, I\,;\,\Sigma)$ is called the likelihood function and employs a statistical model of the detector response to approximate the probability that the measurements $\mathbf{D}$ would be

observed if the source is located at $\mathbf{r}$ with intensity I. In this model I assume that the detector responses are governed by Poisson counting statistics and depend on the source location and intensity, as well as $\Sigma(\mathbf{r}')$, the total macroscopic cross-section of materials in the scene at location $\mathbf{r}'$.

For the remainder of this chapter we will focus on the various components required to construct $P(\mathbf{r}, I \,|\, \mathbf{D})$ in practice. I will begin with a summary of the material presented in Chapter 2, which describes a technique to produce samples from an arbitrary probability distribution using rejection sampling in a process called Markov-chain Monte Carlo (MCMC). Practically, Eq. (3.1) is difficult to evaluate directly, so instead I will show how to use MCMC to draw a sequence of samples from the posterior. MCMC requires that we construct a statistical model of the counts measured by the detector network to compute the likelihood, hence in the following two sections I will present a simplified deterministic transport model for the detector counts and subsequently use this to construct a statistical model for the measurement data with corresponding likelihood function. I will also describe how to extend the MCMC algorithm to account for fixed epistemic uncertainties in the material cross sections and propagate these onto the posterior estimate for the source location, an effect that has not be accounted for in prior work. Finally, I will give the complete form of the source localization algorithm, including an extended version of the Metropolis algorithm that improves the performance of the sampling process.

3.1 Markov-Chain Monte Carlo Sampling

Direct evaluation of Eq. (3.1) is typically impractical due to the difficulty of evaluating the integral normalization term in the denominator. Instead, we can employ a Metropolis sampler of the type described in Chapter 2 to draw samples from the posterior via rejection, a process typically referred to as Markov Chain Monte Carlo. This sampler avoids computing the integral in Eq. (3.1) by only evaluating ratios of $P(\mathbf{r}, I \,|\, \mathbf{D})$ at different points in the domain, causing the integral terms to cancel.

The basic form of the localization algorithm is listed in Algorithm 2, which uses the Metropolis-Hastings algorithm to generate a sequence of N_S samples from the posterior distribution in Eq. (3.1), where the t-th sample is referred to as the state of the sampler at step t. Metropolis-type samplers allow us to produce samples from an arbitrary probability distribution (called the **target distribution**) by performing rejection sampling on an unrelated **proposal distribution** that depends only on the state of the chain at step t, which for the purposes of this dissertation I will restrict to be normal[1]. It is clear that we need to construct

[1] As shown in Chapter 2, the proposal distribution is not generally required to be normal. A normally distributed proposal is however sufficient for our needs and is also a natural choice when using the Adaptive Metropolis algorithm discussed in Section 3.5.

a sampler where the target distribution is the posterior density $P(\mathbf{r}, I \,|\, \mathbf{D})$ in Eq. (3.1), which can be done by exploiting the properties of Markov chains. At each step in the chain, a candidate pair $x^* = (\mathbf{r}^*, I^*)$ is drawn from the proposal distribution, then accepted or rejected based on the ratio of P evaluated at the candidate point x^* versus at the last accepted sample x_t. I have previously shown from the theory of Markov chains that this procedure will produce samples from $P(\mathbf{r}, I \,|\, \mathbf{D})$ after an initial settling period colloquially known as "burn-in", referring samples that are drawn before the chain has had time to forget its initial state (see ch. 4, 8 of ref. [40]). We can then draw as many samples as needed in order to reconstruct $P(\mathbf{r}, I \,|\, \mathbf{D})$.

Algorithm 2 The Metropolis Algorithm for a Bayesian Posterior

1: **procedure** SAMPLE($N_S, \mathbf{r}_0, I_0, \mathbf{C}$)
2: $\quad \mathbf{r} \leftarrow \mathbf{r}_0, I \leftarrow I_0$
3: $\quad \mathcal{S} \leftarrow \{\}, t \leftarrow 1$
4: $\quad$ **while** $t \leq N_s$ **do**
5: $\quad\quad \mathbf{r}^*, I^* \rightsquigarrow \mathcal{N}\left[\left[\begin{smallmatrix}\mathbf{r}\\ I\end{smallmatrix}\right], \mathbf{C}\right]$ [2]
6: $\quad\quad \alpha \leftarrow \min\left\{1, \frac{P_0(\mathbf{r}^*, I^*)}{P_0(\mathbf{r}, I)} \cdot \frac{\mathcal{L}_{\mathbf{D}}(\mathbf{r}^*, I^*; \boldsymbol{\Sigma})}{\mathcal{L}_{\mathbf{D}}(\mathbf{r}, I; \boldsymbol{\Sigma})}\right\}$
7: $\quad\quad \beta \rightsquigarrow \mathcal{U}[0, 1]$
8: $\quad\quad$ **if** $\beta \leq \alpha$ **then**
9: $\quad\quad\quad \mathcal{S} \leftarrow \texttt{append}\,(\mathcal{S}, \{\mathbf{r}^*, I^*\})$
10: $\quad\quad\quad \mathbf{r} \leftarrow \mathbf{r}^*, I \leftarrow I^*$
11: $\quad\quad\quad t \leftarrow t + 1$
12: $\quad\quad$ **else**
13: $\quad\quad\quad$ **continue**
14: $\quad\quad$ **end if**
15: $\quad$ **end while**
16: $\quad$ **return** $\mathcal{S}$
17: **end procedure**

As an alternative to the more formal presentation in Chapter 2, to intuitively understand why this produces samples from the target posterior distribution given by Eq. (3.1), let us further examine the value of the Metropolis ratio α on line 6 of Algorithm 2. If we solve Eq. (3.1)

[2] $\mathcal{N}[\mathbf{r}, \mathbf{C}]$ denotes a (multi-)normal distribution with mean $\mathbf{r}$ and (co-)variance $\mathbf{C}$. Similarly, $\mathcal{U}[a, b]$ denotes a uniform distribution on the interval $[a, b]$.

for $\mathcal{L}_{\mathbf{D}}(\mathbf{r}, I\,;\,\boldsymbol{\Sigma})$ and substitute into the expression for α:

$$\begin{aligned}
\alpha &= \min\left\{1, \frac{P_0(\mathbf{r}^*, I^*)}{P_0(\mathbf{r}, I)} \cdot \frac{\mathcal{L}_{\mathbf{D}}(\mathbf{r}^*, I^*;\,\boldsymbol{\Sigma})}{\mathcal{L}_{\mathbf{D}}(\mathbf{r}, I;\,\boldsymbol{\Sigma})}\right\} \\
&= \min\left\{1, \frac{\cancel{P_0(\mathbf{r}^*, I^*)}}{\cancel{P_0(\mathbf{r}, I)}} \cdot \frac{\cancel{P_0(\mathbf{r}, I)}}{\cancel{P_0(\mathbf{r}^*, I^*)}} \cdot \frac{P(\mathbf{r}^*, I^* \,|\, \mathbf{D})}{P(\mathbf{r}, I \,|\, \mathbf{D})}\right\} \\
&= \min\left\{1, \frac{P(\mathbf{r}^*, I^* \,|\, \mathbf{D})}{P(\mathbf{r}, I \,|\, \mathbf{D})}\right\} .
\end{aligned}$$

Therefore accepting $\mathbf{r}^*, I^*$ with probability α is equivalent to accepting with a probability equal to the ratio of the posterior distribution at the candidate values versus at the previous sample. This allows us to avoid directly calculating $P(\mathbf{r}, I \,|\, \mathbf{D})$ (which is unknown) by instead evaluating the likelihood and prior, which can be computed using a statistical model for the observations and the measurement data. When the proposed candidates are more probable than the current state ($\alpha = 1$) the candidates are always accepted and the chain advanced. Conversely, if the candidates are less probable then the current state ($\alpha < 1$) we do not immediately reject them; instead we reject with a probability that is inversely proportional to the decrease in the posterior probability. This process therefore tends to accept samples from areas of the posterior which are higher in probability (at a rate that is proportional to $P(\mathbf{r}, I \,|\, \mathbf{D})$), while still allowing the chain to occasionally explore the lower probability areas of the distribution. Ultimately, the theory of Markov chains guarantees that the resulting samples will, after an initial period while the chain stabilizes, be distributed according to the posterior.

3.2 Statistical Model for Detector Response

In the formulation for the statistical model of the detector counts, I assume that the response of the detector to the source is a Poisson-distributed random variable with mean $d_i(\mathbf{r}, I\,;\,\Sigma)$, where the exact form of $d_i(\mathbf{r}, I\,;\,\Sigma)$ will be discussed in Section 3.3. I also allow for a random contribution from background to the counts measured by the i-th detector that is independent of the source, represented by the random variable $\mathcal{B}_i$. I typically assume $\mathcal{B}_i$ is also Poisson distributed with a mean b_i, which is allowed to vary with detector location[3], and hence the statistical model for the counts measured by the i-th detector $\mathcal{D}_i$ is

$$\begin{aligned}
\mathcal{D}_i &= \mathrm{Po}\left[d_i(\mathbf{r}, I;\,\Sigma)\right] + \mathcal{B}_i \\
&= \mathrm{Po}\left[d_i(\mathbf{r}, I\,;\,\Sigma)\right] + \mathrm{Po}\left[b_i\right] \\
&= \mathrm{Po}\left[d_i(\mathbf{r}, I\,;\,\Sigma) + b_i\right] ,
\end{aligned}$$

[3]That is, b_i is the mean of the background at the location of the i-th detector. It is only necessary to characterize background at the detector locations, not at all points in the scene.

with every $\mathcal{D}_i$ assumed to be mutually independent.

As shown in chapter 4 of ref. [40], when the measurements are mutually independent the likelihood function is simply the product of the probabilities of observing the individual measurements. Therefore

$$\begin{aligned}\mathcal{L}_{\mathbf{D}}(\mathbf{r}, I;\, \Sigma) &= \prod_{i=1}^{N_D} \Pr\left[\mathcal{D}_i\left[\mathbf{r}, I;\, \Sigma\right] = \mathbf{D}_i\right] \\ &= \prod_{i=1}^{N_D} \frac{\left(d_i(\mathbf{r}, I;\, \Sigma) + b_i\right)^{\mathbf{D}_i}}{\mathbf{D}_i!} \exp\left(-d_i(\mathbf{r}, I;\, \Sigma) - b_i\right) \\ &= \left(\prod_{i=1}^{N_D} \frac{\left(d_i(\mathbf{r}, I;\, \Sigma) + b_i\right)^{\mathbf{D}_i}}{\mathbf{D}_i!}\right) \cdot \exp\left(-\sum_{i=1}^{N_D} d_i\left(\mathbf{r}, I;\, \Sigma\right) + b_i\right) .\end{aligned} \tag{3.2}$$

If I further employ a normal approximation, such that $\mathcal{D}_i \xrightarrow{\text{dist}} \mathcal{N}\left[d_i + b_i, (d_i + b_i)^2\right]$, we obtain the likelihood function in Eq. (3.3):

$$\begin{aligned}\mathcal{L}_{\mathbf{D}}(\mathbf{r}, I;\, \Sigma) &= \prod_{i=1}^{N_D} \frac{1}{\sqrt{2\pi(d_i(\mathbf{r}, I;\, \Sigma) + b_i)}} \cdot \exp\left(-\frac{\left(\mathbf{D}_i - d_i(\mathbf{r}, I;\, \Sigma) - b_i\right)^2}{2(d_i(\mathbf{r}, I;\, \Sigma) + b_i)}\right) \\ &= \left(\prod_{i=1}^{N_D} \frac{1}{\sqrt{2\pi(d_i(\mathbf{r}, I;\, \Sigma) + b_i)}}\right) \cdot \exp\left(-\sum_{i=1}^{N_D} \frac{\left(\mathbf{D}_i - d_i(\mathbf{r}, I;\, \Sigma) - b_i\right)^2}{2(d_i(\mathbf{r}, I;\, \Sigma) + b_i)}\right) .\end{aligned} \tag{3.3}$$

A common rule is that such an approximation is valid when the expected value exceeds 30, which is the case in all of the experiments I will describe in subsequent chapters.

3.3 Detector Response Model

Evaluation of $\mathcal{L}_{\mathbf{D}}(\mathbf{r}, I\,;\, \Sigma)$ during the sampling process requires a computational model of the counts recorded by the detector network as a function of the source location and intensity. This model will be evaluated repeatedly during the sampling procedure (hundreds of thousands of times in our application), and in a way that is primarily serial by construction[4]. This requires a computationally inexpensive model for the underlying transport process; I choose to employ a point source/detector transport model, which only computes the uncollided flux arriving at the detector [17]. I assume that the detectors are sufficiently small that at typical standoff distances (tens to hundreds of meters), a gamma ray that is initially emitted within the detector solid angle and subsequently scatters is very unlikely to reach the detector, hence the detector response is

[4]There has been some success in designing parallel Metropolis samplers beyond simply running multiple independent chains, for example refs. [43, 21]. However, many of these modifications either achieve only modest parallel utilization or rely on assumptions that are not valid in this application.

dominated by the uncollided flux arriving at the detector. Gamma rays may backscatter off of materials near the detector, however the backscatter contribution will be roughly proportional to the uncollided contribution.

Under these assumptions, the simplified detector response model takes the form of exponential attenuation given in Eq. (3.4), with parameters defined in Table 3.1 [6, 16]. Note that this model includes a parametric dependence on $\Sigma(\mathbf{r}')$. Clearly the detector response will be influenced by these cross-sections, however as discussed Section 3.4 these are treated separately in the sampling algorithm since they are considered to be fixed epistemic uncertainties arising from our imprecise knowledge of the cross-section of materials in the scene.

$$d_i(\mathbf{r}, I\,;\,\Sigma) = I\Delta t_i \cdot \epsilon_i^{\text{int}} \cdot \frac{A_i}{4\pi\|\mathbf{r}-\mathbf{r}_i\|_2^2} \cdot \exp\left(-\int_{\mathbf{r}\to\mathbf{r}_i} \Sigma(\mathbf{r}')\,ds\right) \tag{3.4}$$

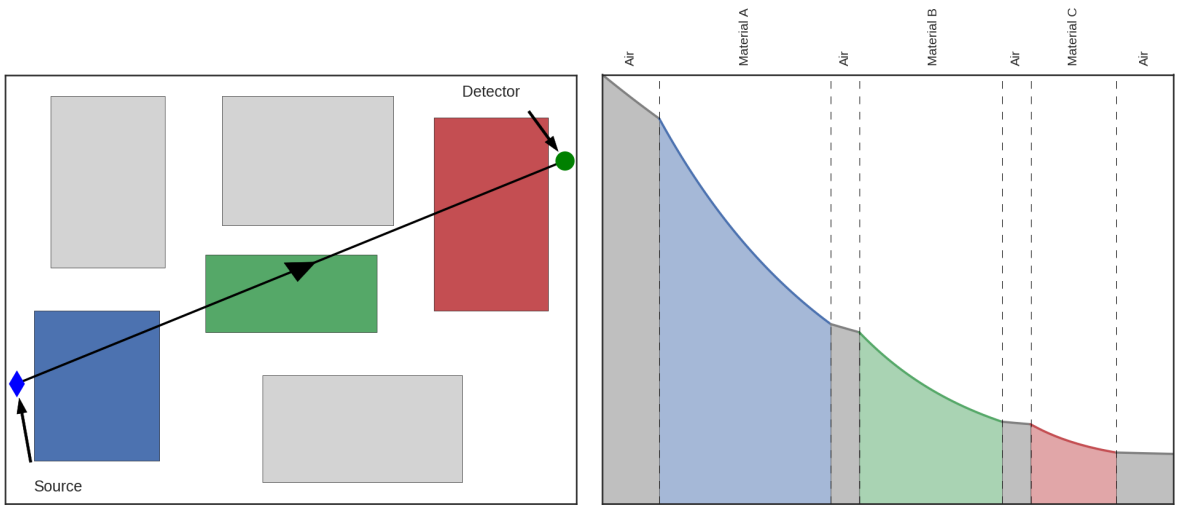
Table 3.1: Model parameters.

Parameter	Meaning
Δt_i	Measurement time (s)
ϵ_i^{int}	Detector intrinsic efficiency (%)
A_i	Detector face area (m^2)
$\mathbf{r}_i$	Detector position vector (m)
$\Sigma(\mathbf{r}')$	Total macroscopic cross-section at $\mathbf{r}'$ (1/m)

To implement Eq. (3.4), I employ a simple scheme based on ray tracing. Assume the geometry of the scene can be decomposed into a set of disjoint polygons, which we denote as $\mathcal{X}$. Each polygon corresponds to the exterior perimeter of an object in the scene, with the j-th object being assigned a constant macroscopic cross section Σ_j. This Σ_j is calculated using an estimate of the material composition of the object and homogenized over its volume while accounting for the estimated internal dimensions of the object; for the moment I will assume it is exact, though this will be addressed further in Section 3.4. The implementation then proceeds by drawing a ray between the source and detector locations, computing the intersection of this ray with the geometry, looking up the associated material and detector properties, and finally substituting into Eq. (3.4). This process is repeated for each detector in order to generate the response of the full network. The implementation of the ray-tracing calculations for the attenuation is specified in Algorithm 3, with a simple graphical illustration shown in Fig. 3.1.

Algorithm 3 Ray tracing algorithm for attenuation factor.

function RAYTRACE($\mathbf{r}, \mathbf{r}_i; \mathcal{X}, \mathbf{\Sigma}$)
    $L \leftarrow 0$
    for all $\mathcal{P}_j \in \mathcal{X}$ **do** ▷ Linear search over geometry[5]
        for all $\mathbf{p}_k \in \mathcal{P}_j \cap (\mathbf{r} \rightarrow \mathbf{r}_i)$ **do**
            $\ell \leftarrow \|\mathbf{p}_k\|_2$
            $L \leftarrow L + \ell \cdot \mathbf{\Sigma}_j$
        end for
    end for
    return L
end function



(a) Compute intersection of ray with geometry (b) Evaluation of Eq. (3.4)

Figure 3.1: Visual illustration of the ray tracing procedure

[5]I did not require the polygons corresponding to the objects in the scene to be convex, implying that the intersection of a ray with one of the polygons may contain several non-overlapping segments, hence the inner `for` loop in Algorithm 3.

3.3.1 Effect of Detector Orientation

Equation (3.4) relies on an approximation for the solid angle of the i-th detector, $\Omega_i(\mathbf{r})$, as viewed from the location of the source:

$$\Omega_i(\mathbf{r}) \approx \frac{A_i}{\|\mathbf{r} - \mathbf{r}_i\|_2^2} \,.$$

This assumes that the detector face area exposed to the source, A_i, is approximately constant regardless of detector orientation. Such an assumption is true for a spherical detector and approximately true for a cubic detector; however, the experiments described in Chapter 5 used NaI detectors with crystal dimensions of 2" × 4" × 16". The solid angle subtended by each of these detectors therefore depends strongly on the orientation of the detector in space about its own center and relative to the source location $\mathbf{r}$.

I will first demonstrate the simplest possible approach, using a constant face area computed by averaging the face areas of each detector; equivalently I assume that each detector has a face profile whose area is the average face area of that detector. In Chapter 5 we will see that this produces significant errors in the estimated source location, caused by variations in the measured count rates due to varying geometric efficiencies, $\epsilon_i^{\text{geo}} := \Omega_i / 4\pi$, an effect that cannot be accounted for by this simplistic model.

Therefore I will extend the detector model to also account for variations in the detector solid angle using the method of Van Oosterom and Strackee [45], which provides an exact expression for the solid angle subtended by an arbitrarily oriented right triangle as viewed from a specified position in free space. Equation (3.4) then becomes

$$d_i(\mathbf{r}, I\,;\,\Sigma) = I\Delta t_i \cdot \epsilon_i^{\text{int}} \cdot \frac{\Omega_i(\mathbf{r})}{4\pi} \cdot \exp\left(- \int_{\mathbf{r}\to\mathbf{r}_i} \Sigma(\mathbf{r}')\, ds \right) , \tag{3.5}$$

where $\Omega_i(\mathbf{r})$ is computed by summing the solid angles of each face of the detector exposed to the source, calculated using the expression given in eq. (8) of ref. [45].

3.4 Propagation of Cross Section Uncertainties

As shown in Section 3.3, the prediction for the counts observed by the detector network depends on the macroscopic cross sections of objects in the scene, which are typically poorly known, inaccurate and imprecise. Instead I must rely on estimates, $\mathbf{\Sigma}^0$, which are calculated from the best available estimates of the material composition and object dimensions. I then assign a wide uncertainty range onto the nominal cross sections, which is representative of our fundamental inability to provide an accurate estimate of their true values. The standard formulation of

the Metropolis sampler that we have discussed thus far is unable to account for these fixed uncertainties and would also attempt to "update" the uncertainty of $\mathbf{\Sigma}$ based on the measured count rates.

I propagate fixed uncertainties in the cross section values onto the posterior distributions for the source location and intensity by sampling cross section values from a user-specified proposal distribution, $\mathbf{f}_\Sigma$, accepting or rejecting sampled cross section values according to the Metropolis ratio α in the same manner as for source position and intensity. The user can provide an $\mathbf{f}_\Sigma$ that encodes their confidence in the estimated cross section values and see that confidence reflected in the resulting posterior. For the purposes of the work in this dissertation, I generally will choose $\mathbf{f}_\Sigma$ to be a broad uniform distribution, centered on $\mathbf{\Sigma}_0$. This is implemented in the Section 3.5 by using $\mathbf{f}_\Sigma$ as the proposal distribution when sampling new values for $\mathbf{\Sigma}$.

Note that due to the nature of the problem, true values of $\mathbf{\Sigma}$ are non-identifiable and hence the marginal posteriors for these parameters will not change during the calibration process. This is a consequence of the fact that multiple combinations of different cross section values can produce the same observed detector response, so there is no unique set of cross section values that can be identified based on measurements. The calibration process includes the effect of the cross section uncertainties encoded in the prior, but there will not be any refinement in their sampled posterior distribution since these values are non-identifiable. Thus, I choose to discard samples for the cross section that are drawn by the chain, since they do not provide any new information beyond what is already encoded in $\mathbf{f}_\Sigma$.

3.5 Adaptive Metropolis

I will also employ one further modification to the standard Metropolis algorithm in order to improve the convergence rate, which is known as the Adaptive Metropolis algorithm[6], and functions by adapting the proposal covariance using the sequence of previously accepted candidates. This method violates the typical assumptions of Markov chains ensure ensure convergence since the proposal distribution no longer depends only on the current state, however in refs. [13, 1] the authors use more sophisticated analysis based on the theory of mixingales to show that convergence is still asymptotically guaranteed.

Practically speaking this requires several modifications to the basic formulation described in Algorithm 2. Firstly, the adaptive method most often implies the use of a (multivariate) normal proposal distribution, which is initialized with covariance matrix $\mathbf{C}_0$. Each time a sample is accepted the proposal covariance is updated using the empirical covariance computed from the history of samples that have been accepted by the chain, so that as the chain evolves the

[6]There are several methods known as "adaptive Metropolis", but typically in literature "*the* Adaptive Metropolis algorithm" refers to the one described here, first presented in ref. [13].

proposal distribution adapts to the shape of the target distribution. To efficiently compute the empirical covariance a recursive expression is used to calculate the covariance online instead of using the standard offline formula, which is significantly more computationally intensive (see the definitions of `UpdateMean` and `UpdateCovariance` in Algorithm 4). In practice it is also common to delay the start of the adaptation process until after a certain number of samples have been accepted, as well as to limit the adaptation to only the k most recently accepted samples, but for simplicity this is not shown in Algorithm 4.

Observe also the presence of constant ϵ in the expression for the updated covariance (line 31 of Algorithm 4), which serves to guarantee that the resulting covariance matrix is non-singular[7]. Additionally, the constant s_p (line 30 of Algorithm 4) is a scaling parameter which controls chain mixing, with the specific value $s_p = 2.4^2/[\text{number of parameters}]$ being a common choice which optimizes performance when the target and proposal distributions are normally distributed [8, 13].

3.6 Complete Algorithm for Source Localization

I began by describing the basic implementation of the Metropolis algorithm in Algorithm 2, followed by a description of how each component is specialized to the problem of source localization. Now, in Algorithm 4 these results are combined into the complete form of the source localization algorithm.

The sampler begins by initializing the chain with starting values for the source location and intensity, $\mathbf{r}_0$ and I_0, as well as providing an initial proposal covariance $\mathbf{C}_0$ and an uncertainty distribution for the material macroscopic cross sections $\mathbf{f}_\Sigma$. The algorithm then proceeds by randomly drawing new sample candidates for the location and intensity from the proposal distribution, as well as sampling new values for the cross sections. It then calculates the Metropolis ratio α by comparing the likelihood at the most recently accepted location and intensity to the likelihood at the proposed values, with the likelihood computed using the ray tracing scheme in Algorithm 3 and the statistical model in Eq. (3.3). Lastly, the algorithm accepts the candidate values and updates the proposal covariance with probability proportional to α, otherwise the candidates are rejected and the procedure repeated. This process continues until the desired number of samples has been drawn by the chain, after which the posterior distribution can be reconstructed from the samples.

[7] It is often safe to assume $\epsilon = 0$ [40].

Algorithm 4 Adaptive Metropolis sampler for source location and activity

procedure SAMPLE($N_S, \mathbf{r}_0, I_0, \mathbf{C}_0, \mathbf{f}_\Sigma$)
  $\mathbf{r} \leftarrow \mathbf{r}_0, I \leftarrow I_0$
  $\bar{\mathbf{r}} \leftarrow \mathbf{r}_0, \bar{I} \leftarrow I_0, \mathbf{C} \leftarrow \mathbf{C}_0$
  $\mathcal{S} \leftarrow \{\}, t \leftarrow 1$
  $\boldsymbol{\Sigma} \leftsquigarrow \mathbf{f}_\Sigma$
  while $t \leq N_s$ **do**
    $\mathbf{r}^*, I^* \leftsquigarrow \mathcal{N}\left[\left[\begin{smallmatrix}\mathbf{r}\\ I\end{smallmatrix}\right], \mathbf{C}\right]$
    $\boldsymbol{\Sigma}^* \leftsquigarrow \mathbf{f}_\Sigma$
    $\alpha \leftarrow \min\left\{1, \frac{P_0(\mathbf{r}^*, I^*)\cdot \mathbf{f}_\Sigma(\boldsymbol{\Sigma}^*)}{P_0(\mathbf{r}, I)\cdot \mathbf{f}_\Sigma(\boldsymbol{\Sigma})} \cdot \frac{\mathcal{L}_{\mathbf{D}}(\mathbf{r}^*, I^*; \boldsymbol{\Sigma}^*)}{\mathcal{L}_{\mathbf{D}}(\mathbf{r}, I; \boldsymbol{\Sigma})}\right\}$
    $\beta \leftsquigarrow \mathcal{U}[0,1]$
    if $\beta \leq \alpha$ **then**
      $\mathcal{S} \leftarrow \texttt{append}\,(\mathcal{S}, \{\mathbf{r}^*, I^*\})$
      $\bar{\mathbf{r}}^*, \bar{I}^* \leftarrow$UPDATEMEAN$(t, \mathbf{r}^*, I^*, \bar{\mathbf{r}}, \bar{I})$
      $\mathbf{C} \leftarrow$UPDATECOVARIANCE$(t, \mathbf{r}, I, \bar{\mathbf{r}}, \bar{I}, \mathbf{r}^*, I^*, \bar{\mathbf{r}}^*, \bar{I}^*, \mathbf{C};\, \epsilon)$
      $\mathbf{r} \leftarrow \mathbf{r}^*, I \leftarrow I^*$
      $\bar{\mathbf{r}} \leftarrow \bar{\mathbf{r}}^*, \bar{I} \leftarrow \bar{I}^*$
      $\boldsymbol{\Sigma} \leftarrow \boldsymbol{\Sigma}^*$
      $t \leftarrow t + 1$
    else
      continue
    end if
  end while
  return $\mathcal{S}$
end procedure

function UPDATEMEAN$(t, \mathbf{r}^*, I^*, \bar{\mathbf{r}}, \bar{I})$
  $\bar{\mathbf{r}}^* \leftarrow \bar{\mathbf{r}} + \frac{1}{t}(\mathbf{r}^* - \bar{\mathbf{r}})$
  $\bar{I}^* \leftarrow \bar{I} + \frac{1}{t}\left(I^* - \bar{I}\right)$
  return $\bar{\mathbf{r}}^*, \bar{I}^*$
end function

function UPDATECOVARIANCE$(t, \mathbf{r}, I, \bar{\mathbf{r}}, \bar{I}, \mathbf{r}^*, I^*, \bar{\mathbf{r}}^*, \bar{I}^*, \mathbf{C};\, \epsilon)$
  $p \leftarrow \texttt{len}\,(\mathbf{r}) + 1$
  $s_p \leftarrow \frac{2.4^2}{p}$
  $\mathbf{C}^* \leftarrow \frac{t-1}{t}\mathbf{C} + \frac{s_p}{t}\left(t\left[\begin{smallmatrix}\bar{\mathbf{r}}\\ \bar{I}\end{smallmatrix}\right]\left[\begin{smallmatrix}\bar{\mathbf{r}} & \bar{I}\end{smallmatrix}\right] - (t+1)\left[\begin{smallmatrix}\bar{\mathbf{r}}^*\\ \bar{I}^*\end{smallmatrix}\right]\left[\begin{smallmatrix}\bar{\mathbf{r}}^* & \bar{I}^*\end{smallmatrix}\right] + \left[\begin{smallmatrix}\mathbf{r}^*\\ I^*\end{smallmatrix}\right]\left[\begin{smallmatrix}\mathbf{r}^* & I^*\end{smallmatrix}\right] + \epsilon\mathbb{I}_p\right)$
  return $\mathbf{C}^*$
end function

3.6.1 Relation to Maximum Likelihood Estimation

As mentioned in Chapter 1, a common alternative approach to source localization is to compute the **maximum likelihood estimate** (MLE), which is defined as the $\mathbf{r}, I$ that maximize the likelihood function $\mathcal{L}_{\mathbf{D}}$:

$$(\mathbf{r}_{\text{MLE}}, I_{\text{MLE}}) := \arg\max_{\mathbf{r},I} \mathcal{L}_{\mathbf{D}}(\mathbf{r}, I\,;\, \boldsymbol{\Sigma})\,. \tag{3.6}$$

In frequentist inference the MLE is often used to provide an estimate of the model parameters without invoking prior information. From a Bayesian perspective, the MLE is related to the **maximum a posteriori** (MAP) estimate, defined as the $\mathbf{r}, I$ that maximize the posterior given by Eq. (3.1):

$$\begin{aligned}(\mathbf{r}_{\text{MAP}}, I_{\text{MAP}}) &:= \arg\max_{\mathbf{r},I} P(\mathbf{r}, I \,|\, \mathbf{D}) \\ &= \arg\max_{\mathbf{r},I} \mathcal{L}_{\mathbf{D}}(\mathbf{r}, I\,;\, \boldsymbol{\Sigma}) \cdot P_0(\mathbf{r}, I)\,.\end{aligned} \tag{3.7}$$

As such, the MLE can be interpreted as a special case of the MAP estimate with a constant (uniform) prior distribution P_0, since multiplication by a constant does not change the location of the maximum in the parameter space.

The output of the source localization algorithm is the full posterior distribution, but for the sake of discussion I will typically identify the mode of the posterior as "the" predicted source location since it is, by definition, the source location we are most confident is the true one considering the count rates we observed. It is clear from Eq. (3.7) that the mode of the posterior is also the MAP estimate. As noted previously, for the analysis in this dissertation I assume uniformly distributed prior distributions for all parameters, hence the mode of the posterior distribution is also the MLE. The distinction is that the MCMC localization algorithm provides additional information since the full posterior distribution is available, whereas the MLE is only a point estimate. The MLE is also incompatible with other forms of prior distribution, which are of interest in situations where additional information about the source characteristics is available.

3.6.2 Difficulties with Direct Numerical Optimization

The consideration of attenuators significantly complicates the application of many of the alternative techniques for source localization described in Chapter 1, which generally do not account for variable attenuation in a heterogeneous environment. In particular, the most common approach to compute the MLE defined in Section 3.6.1 is to perform a direct numerical maximization of the likelihood function[8], but this becomes difficult when attenuators are included

[8]It is actually more common to minimize the negative log of the likelihood for numerical reasons, but the problem is equivalent because the log function is monotonic.

due to the non-smooth error surface that results. As an illustration, Fig. 3.2 plots the error surface in terms of the sum-of-squares error (SSE) for the measurements collected in the first of the experiments described in Chapter 5.

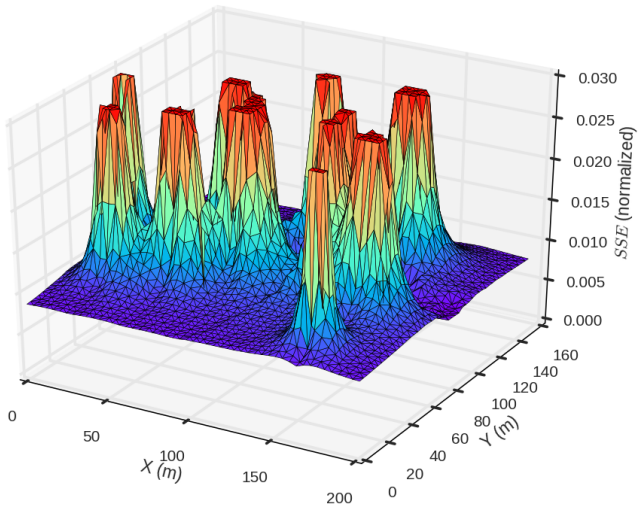


Figure 3.2: Sum-of-squares error as a function of source location for experiment 1 described in Chapter 5. The peaks are clipped to show detail of the error surface.

The complexity of this error surface, which includes discontinuities along every object boundary as well as singularities at the detector locations, makes it difficult to provide a direct comparison to methods based on numerical optimization. For instance, in principle the MLE produced by direct numerical optimization should agree with the mode of the posterior density produced by our algorithm (see Section 3.6.1); however, when I implemented this the search failed to converge in all cases using a variety of optimization algorithms including Nelder-Mead, Broyden-Fletcher-Goldfarb-Shanno (BFGS), and Simulated Annealing. Interested readers should see ref. [41] for more information regarding the numerical challenges involved, as well as the application of more advanced hybrid optimization techniques to the same model as in Section 3.3.

Chapter 4

Preliminary Experiments in Simulated Geometry

In this chapter I will present the results of applying the localization algorithm described in Chapter 3 to several test cases relying on simulated data. The objective of these studies is to investigate three fundamental questions regarding the localization algorithm:

1. Can the source be localized at all?
2. What effect does the simplified transport model have on the bias (accuracy) in the predicted source location?
3. What effect does imprecise knowledge of the composition of materials in the scene have on the uncertainty (precision) in the predicted source location?

Each test case directly addresses one of these questions, and the results form a basis for the design and execution of a measurement campaign in the field performed in collaboration with Oak Ridge National Laboratory, which will be described in Chapter 5.

The first test case uses a geometry derived from a real city block, with cross sections generated randomly based on the typical composition of wood and concrete buildings, and detector observations generated by randomly sampling directly from the statistical model in Section 3.2 with 300 counts per second background. This is the ideal case, where all phenomena affecting the "measured" count rates are accounted for by the statistical model of the detection process, hence it serves to demonstrate the basic feasibility of the localization algorithm.

We will then examine a situation where the measurement data is derived from high-fidelity simulation using analog Monte Carlo in a simplified geometry to generate the detector responses, while still relying on the ray-tracing model to predict the measured count rates. This introduces discrepancies in the measurement data based on the physical gamma transport phenomena that

are discounted by the model of the detector response, which the localization algorithm must be able to accommodate in practice since it is not possible to use a more complicated radiation transport model (see Section 3.1).

The final case introduces cross section uncertainties using the technique from Section 3.4. It is impossible to accurately determine the cross section of objects in the field and so in practice we must rely on highly uncertain estimates for cross sections of materials in the scene. The last test case demonstrates the effect that the cross section uncertainties have on the final predicted source location using the same simulated urban geometry as the first test.

4.1 Synthetic Model in Urban Geometry

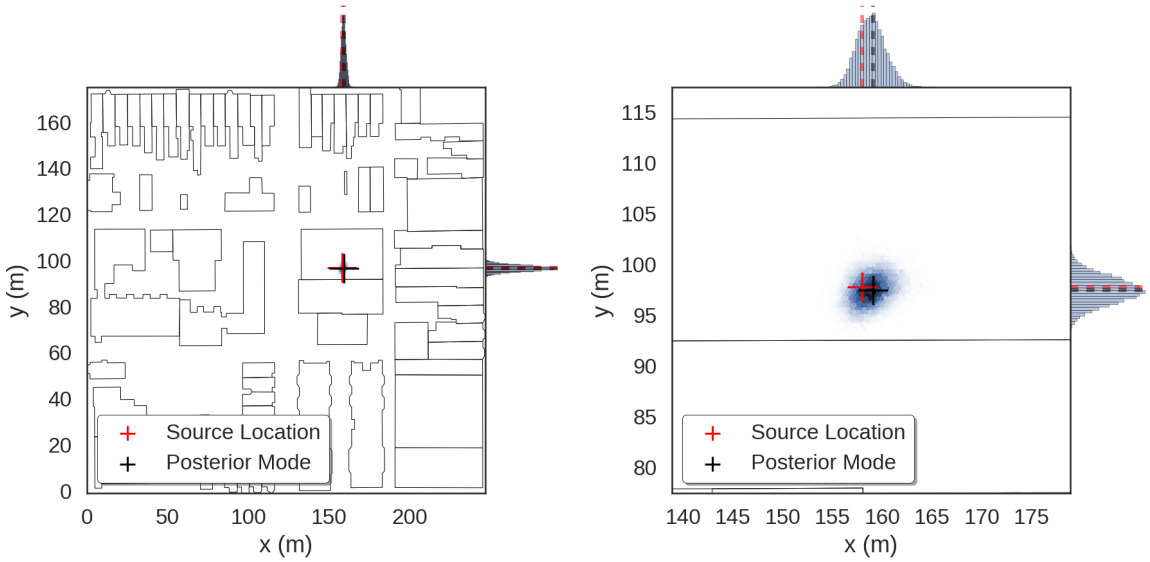
To illustrate the method, I constructed a synthetic test case based on a simulated urban geometry. Using geospatial data from the OpenStreetMaps database I randomly selected a city block in downtown Washington DC that was approximately 250 m×180 m, pictured in Fig. 4.1. This was then translated into a set of polygons representing each of the 70 buildings in the area. A single source with nominal intensity 8.7 mCi (∼100 µg of cesium-137) was placed in one of the buildings and 10 detectors were distributed randomly with uniform probability across the region. This arrangement is shown in Fig. 4.1. Cross section data was generated at random, constrained such that an average sized building was approximately 1-1.5 mean free paths thick[1], while synthetic detector measurements were generated by sampling directly from the statistical model in Section 3.2 (i.e., sampled from a Poisson distribution with mean computed using the ray-tracing model from Section 3.3 plus background). The detectors were modeled as 3" × 3" NaI scintillators ($A_i \sim 58\,\mathrm{cm}^2$) with counting times of 10 s, while background was taken as a Poisson distributed random variable with a mean intensity of 300 cps (typical of a 3" × 3" NaI detector in the Southeast United States [18]).

[1] This is typical of a wood and concrete building, based on estimates using information from ref. [4, 28].

Figure 4.1: Satellite image of test location, with modeled geometry overlayed in green, as well as source (red) and detector (blue) locations.

The MCMC process was used to draw 10^5 samples, with the first half discarded as burn-in. Analysis of the chains by visual inspection of the sample histories, autocorrelation plots, and computation of Gelman-Rubin scores indicated that this was a very conservative choice, providing a high level of confidence that the chains had converged and were producing samples from the posterior[2]. Figure 4.2 shows the results from the test problem using synthetic data, including marginal densities for $\mathbf{r}$. It can be seen that the method is able to localize the source to within roughly a $10\,\mathrm{m} \times 10\,\mathrm{m}$ region, with the mode of the posterior (corresponding to the most likely predicted source location) being within approximately 1.5 m of the true location.

[2] As mentioned in Chapter 3, the convergence of Metropolis samplers is only guaranteed asymptotically and it is not generally possible to determine with certainty whether a finite sub-sequence of a given chain has converged. This is true not only in esoteric situations but also in cases of practical interest [9], hence care is required to ensure that the chain has converged and is producing samples from its true stationary distribution (the posterior). Note that I performed similar convergence analysis for all results shown henceforth and in each case opted to draw at least an order of magnitude more samples than the convergence diagnostics indicated were necessary to provide a very high level of confidence that the samples were indeed drawn from the posterior.



(a) Full Scene

(b) Zoom

Figure 4.2: Posterior density for source location in synthetic experiments.

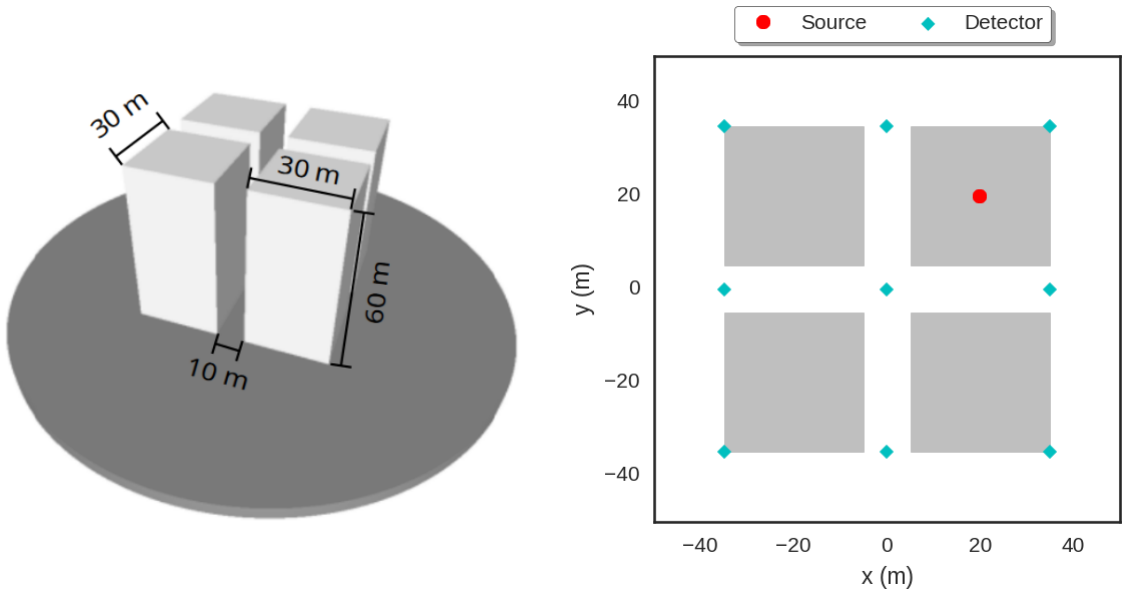
4.2 Comparison to High-Fidelity Simulation

I employed the simplified geometry shown in Fig. 4.3 for the generation of calibration data via high-fidelity Monte Carlo transport simulations. Buildings were modeled as rectangular prisms of concrete with densities adjusted to match an equivalent wall thickness of 0.5 m, while the ground was considered to be a 10 m thick concrete slab. Four identical buildings were arranged in a 2×2 grid, with all remaining interstitial space filled with dry air at standard temperature and pressure. A 3D rendering of the geometry with the building dimensions included is shown in Fig. 4.3a.

Nine point detectors were placed in the scene, with the source placed in the center of the northeast building. The source and detectors were all placed 1 m above ground according to the arrangement in Fig. 4.3b. This geometry was constructed and simulated in `MCNP6` [12] with standard photon physics at an energy of 662 keV using $2 \cdot 10^9$ photons. Particle histories were terminated when their energy fell below 10 keV or when they traveled farther than 95 m from the origin. Count rates were then computed from the detector tallies based on the same detector and source parameters as in Section 4.1.

Note that variations due to background are *not* included in this experiment; hence, the signal-to-noise ratio is effectively infinite, so a very narrow posterior is expected. The simulated measurements differ significantly from the model described in Chapter 3 due to the inclusion

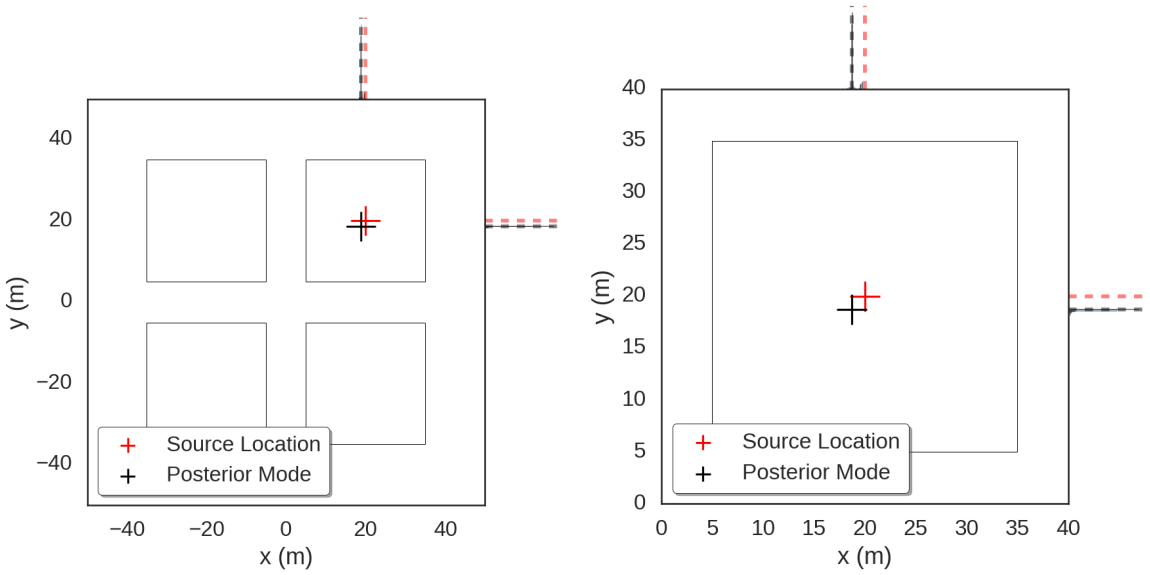
of the contribution of scattering, which is an effect that is not included in the model for the detector counts. The focus in this test is on the influence of the simplified physics model on the localization results and hence background was omitted in order to isolate these effects.



(a) 3D profile view with dimensions (b) Overhead view with source location indicated

Figure 4.3: Model geometry used for simulations using MCNP.

Figure 4.4 shows that the resulting posterior for the source location is still able to identify the source location to within approximately 5 m. We can, however, observe that there is a bias present, with the true source location falling outside of the uncertainty of the posterior distribution. This is due to the physical effects (principally Compton scattering) simulated by MCNP that are not represented by the detector model, which only accounts for the effect of uncollided gammas. The results suggest that this bias is relatively small (a few meters), which is on the same order of magnitude as other sources of error such as the cross section uncertainties that we examine in the next test.



(a) Full Scene (b) Zoom

Figure 4.4: Posterior density for localization based on simulations with MCNP.

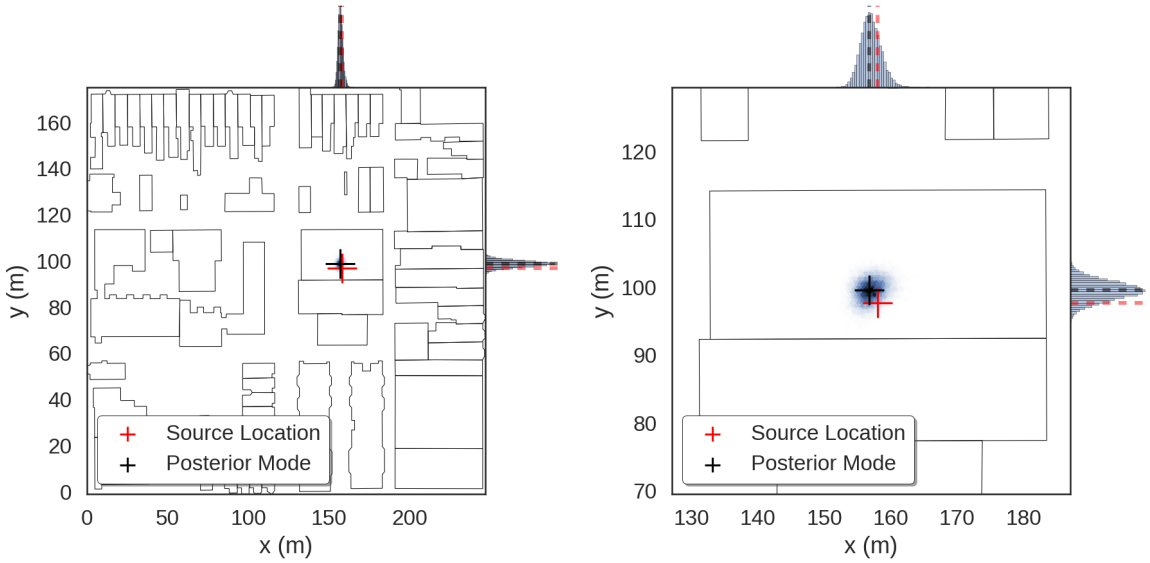
4.3 Effect of Cross Section Uncertainties on Localization Error

To test the impact of the cross section uncertainty on the localization of the source, I designed an experiment simulating a real-world search scenario in an urban environment using a scene identical to the one in Section 4.1. I then introduced fixed cross section uncertainties into the problem and propagated them using the method described in Section 3.4.

Synthetic detector responses were generated using the simplified ray tracing model and the expected values of the building cross sections, applying random variations to the computed responses according to Poisson counting statistics. These simulated measurements were provided to the MCMC simulation described in Chapter 3, and the chains were run sufficiently long to ensure convergence to the stationary distributions. The reference value Σ_i^0 for each cross section was taken to be the true value, and simulations were run with $\pm 5\%$ and $\pm 50\%$ uniformly distributed relative uncertainty in all cross sections. Priors were taken to be uninformative, i.e. uniform over the bounds of the problem. This corresponds to the case that we initially know nothing about the source beyond assuming it is located in the block shown in Fig. 4.1.

Figure 4.5 and Fig. 4.6 shows the resulting marginal posteriors for source location, with Fig. 4.5 showing the results with 5% uniform random variation in the building cross sections and Fig. 4.6 showing 50%. Table 4.1 also shows the corresponding uncertainty in the estimates for source location. We see that the method is able to correctly localize the source in both cases to within

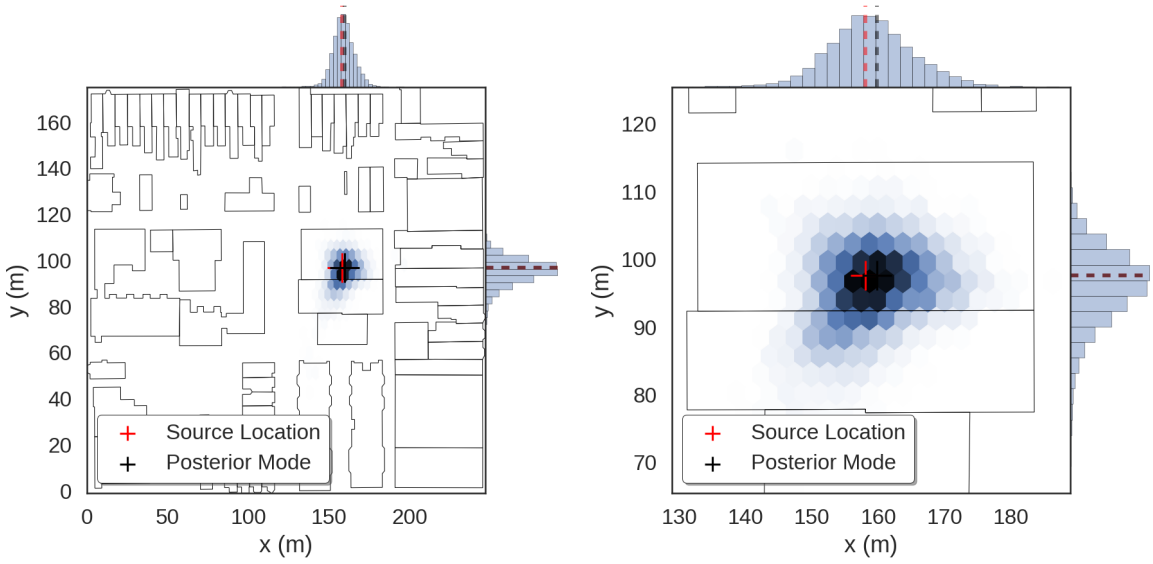
a few meters, with 95% confidence of ∼4 m for Fig. 4.5 and ∼14 m for Fig. 4.6, both sufficient to guide a subsequent, more targeted followup search.



(a) Full Scene

(b) Zoom

Figure 4.5: Posterior density for source location with 5% uncertainty in all cross sections.



(a) Full Scene (b) Zoom

Figure 4.6: Posterior density for source location with 50% uncertainty in all cross sections.

Table 4.1: Predicted source locations and standard deviations for marginal posterior

δ	$\mathbf{r}^{\text{true}}$ (m)	$\mathbf{r}^{\text{pred}}$ (m)	$\sigma_{\mathbf{r}}$ (m)
5%	$(158.0, 98.0)$	$(158.0, 99.7)$	$(1.55, 1.45)$
50%		$(160.2, 97.3)$	$(6.90, 7.19)$

Of particular interest is the relatively small increase in the uncertainty of the estimated source location as the cross section uncertainty is increased. The Bayesian method is still able to identify the building the source is located in despite the large degree of uncertainty in the cross sections of the buildings. This observation shows that the method is able to tolerate highly imprecise foreknowledge of the cross sections and still produce a usably precise estimate of the source location. This result is a direct consequence of the highly heterogeneous transport medium inherent to this problem. The buildings in the scene are highly attenuating, while the space between them is not, with the building cross sections exceeding the cross section of air by about 3 orders of magnitude. Consequently, the posterior estimate of source location is dominated by those detectors that have a near line-of-sight to the source, such that the width of the posterior; i.e., the uncertainty in the estimated source location, is moderately insensitive

to the large uncertainties in building composition and density. We will return to investigate this effect in Chapter 6, using an even further simplified transport model that models objects in the scene as entirely opaque, considering only the visibility of the source from the vantage point of the detectors.

4.4 Informing the Design of a Field Experiment

At the beginning of this chapter I presented three basic research questions, representing the most important issues that must be addressed in order to assess the localization algorithm proposed in Chapter 3. Before proceeding, it is worthwhile to consider how the results shown thus far have addressed these questions. It is important to emphasize that these results are based on synthetic measurement data, with the results in Sections 4.1 and 4.3 using a statistical and physical model that perfectly represents the simulated measurements, while those of Section 4.2 employ a greatly simplified geometry. As such, we cannot yet consider these questions satisfactorily answered based solely on these results. In later chapters, I will present the results of a more comprehensive test, using count rates recorded by real detectors for an unknown source in the field. Before proceeding to the description of these measurements in Chapter 5 and their results, let me first summarize the salient aspects of these synthetic tests with respect to how they influence the design of the field experiment.

Most importantly, we have seen that the algorithm is able to localize a source within a complex geometry in the presence of uncertainties in the count rates (including random contributions from background). This also holds true when simulating the effect of scattering, as well as when large cross section uncertainties are included, all of which are factors that will be relevant when using actual measurement data. Further, the uncertainties in the posterior densities show that the algorithm is able to localize a source with activity in the tens of mCi range with reasonable precision in a typical urban geometry, implying that a source of similar strength would be a reasonable choice for the field test.

The results in Section 4.2 show that the algorithm is able to localize a source even when the effect of scattering is included. Real measurements will include counts from gammas that are emitted by the source and subsequently undergo scattering, a physical effect that is entirely ignored by the detector model in Chapter 3. If this had not been the case, it would have been necessary to develop a more complex model for the transport of gamma rays in the scene, e.g., by approximating near-field scattering in the vicinity of the detector, further increasing the computational cost and complexity. However, the evidence presented thus far indicates that this is not necessary and so I will proceed with the uncollided flux model when analyzing the results of the field measurements.

Lastly, Section 4.3 demonstrates that the effect of uncertainties on the precision of the

localization is relatively weak. In practice, it is difficult to obtain accurate estimates of the cross section of objects like buildings or cars present in the scene, due primarily to poor knowledge of their exact dimensions and composition. The weak dependence of the posterior localization uncertainty on the cross section uncertainty provides evidence that the method is robust to even large uncertainties in the material cross sections, and as such it will not be necessary to perform a highly detailed characterization of the internal geometry and composition of objects present at the location of the experiment. Instead, I will rely on survey measurements to determine the exterior geometry of objects, combined with a rough estimate of material composition and interior dimensions.

Chapter 5

Field Test

In May 2017, a test measurement campaign was conducted at the Energy Systems Test Complex at ORNL, also known as the site of the former Experimental Gas-Cooled Reactor. The measurement campaign took place outdoors in a cluttered environment, chosen to mimic a real-world search in an urban setting. A Cs-137 source with a nominal activity of approximately 37 mCi was placed in two different locations, shown in orange in Fig. 5.1. These source placements divided the measurement campaign into two separate experiments, which I will refer to as experiment 1 and experiment 2.

5.1 Description of Experiment

For each experiment, a set of 6 networked mobile detectors (described in Section 5.1.2) was used to record count rates at 12 locations, both with and without the source present. Measurements with the source and of background were each taken for approximately 30 minutes real-time, and count rates were low enough that dead-time effects were negligible (typically $< 0.1\%$). The placement of these detectors is shown in Fig. 5.2, with all detector positions recorded using precision differential GPS. Source-to-detector distances ranged from approximately 20 m to 200 m. Additionally, all detectors were nominally oriented to be facing due north using hand compasses, since the source location is treated as unknown. Manual orientations were error-prone, with substantial deviations from north in many of the detectors; however, the actual orientations were recorded using the detectors' onboard compasses, mounted such that they record orientation relative to the longest axis of the detector.

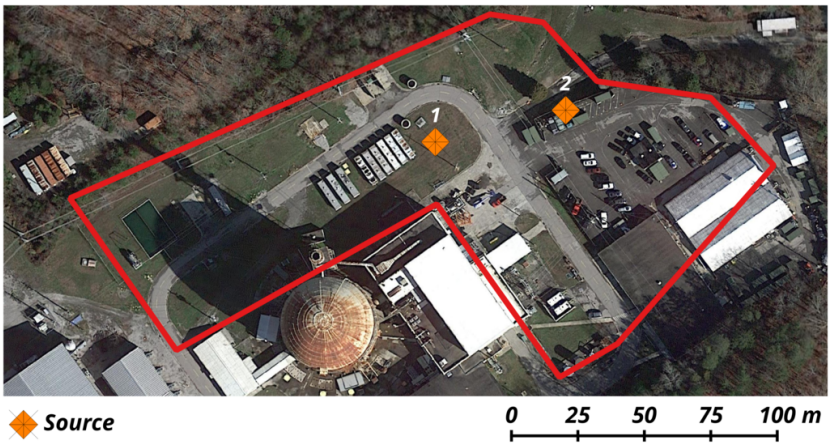
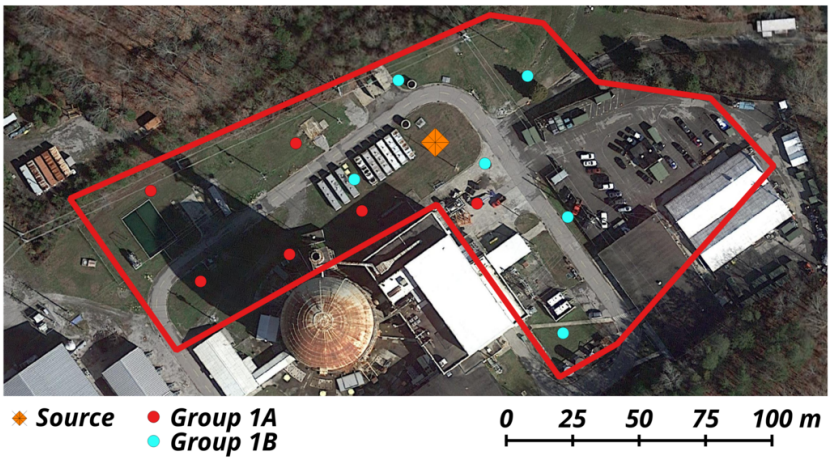
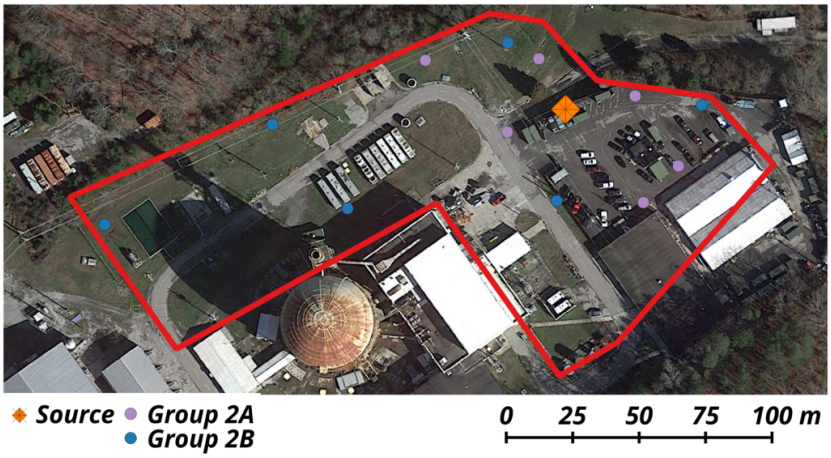


Figure 5.1: Overhead view of experiment site, the two source locations are shown in orange, while the overall experiment area is marked in red (satellite imagery courtesy of Google Maps).



(a) Experiment 1.



(b) Experiment 2.

Figure 5.2: Detector and source placements for each experiment (satellite imagery courtesy of Google Maps).

5.1.1 Detector Placement

The actual placements of the detectors in the field and their orientations as reported by the onboard compasses are summarized in Table 5.1. Each location is given a name denoting the experiment number, measurement set and detector number– for example, measurement 1A-2 refers to the position of detector number 2 (of 6) during the first set of measurements (set A) of experiment 1. x and y coordinates are reported in meters, relative to $35°56'12.1''N$ $84°16'36.8''W$ (WGS 84 Pseudo-Mercator/EPSG:3857 coordinate reference), with the positive x and y directions being east and north, respectively. Orientations are reported as the angle between the

longest axis of the detector and magnetic north, with positive angles in the counter-clockwise direction. Table 5.1 also provides a qualitative description of how obscured the source was from the position of each detector. These are organized into 4 categories, with “direct” indicating detectors that had an unobstructed view of the source, while “highly occluded“ indicates detectors with several mean free paths of material between themselves and the source. Detectors designated as “occluded” had approximately 1-2 mean free paths of material between themselves and the source, while “lightly occluded” refers to detectors with lines of sight interrupted by less than 1 mean free path of material.

Table 5.1: Summary of actual source and detector placement during the experiment, including a qualitative description of the visibility of the source for each detector.

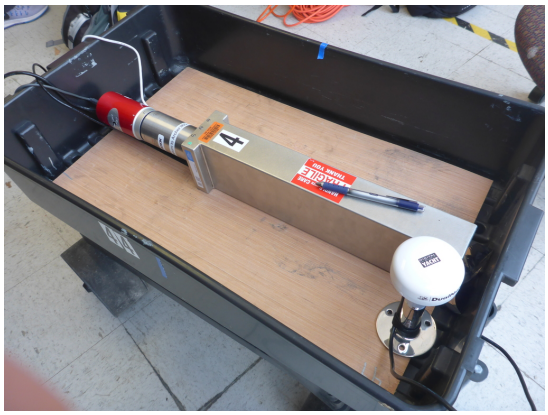
Name	x (m)	y (m)	Distance to Source (m)	Orientation (deg. N)	Source Visibility
Source	114.9	88.9	–	–	–
1A-1	7.84	70.71	108.61	61	Highly occluded
1A-2	26.61	36.56	102.67	-15	Highly occluded
1A-3	62.47	88.61	52.46	2	Occluded
1A-4	60.08	88.61	69.20	18	Occluded
1A-5	87.34	63.13	37.79	-2	Direct
1A-6	130.68	65.92	27.87	0	Lightly occluded
1B-1	84.50	74.99	33.49	33	Highly occluded
1B-2	101.08	112.38	27.22	19	Lightly occluded
1B-3	149.72	113.99	42.82	15	Lightly occluded
1B-4	133.66	89.00	20.32	84	Direct
1B-5	164.86	60.81	57.27	21	Direct
1B-6	163.09	17.05	86.49	0	Highly occluded
Source	164.1	100.8	–	–	–
2A-1	190.23	105.89	26.65	-68	Highly occluded
2A-2	206.51	79.54	47.51	-65	Highly occluded
2A-3	193.30	65.78	45.68	-27	Occluded
2A-4	142.84	92.36	23.56	-16	Lightly occluded
2A-5	153.99	119.94	21.54	-137	Lightly occluded
2A-6	111.22	119.34	55.94	-84	Lightly occluded
2B-1	-9.46	57.86	178.83	-57	Highly occluded
2B-2	53.72	95.20	110.45	-103	Lightly occluded
2B-3	81.98	63.52	90.16	-80	Lightly occluded
2B-4	142.28	125.97	33.19	-29	Direct
2B-5	215.31	102.44	51.26	-97	Highly occluded
2B-6	160.91	66.12	34.59	-63	Highly occluded

5.1.2 Detector Characteristics

The detectors used for the experiment were provided by ORNL and were comprised of a 2" × 4" × 16" NaI scintillator coupled to an integrated photomultiplier tube and Canberra Osprey

multichannel analyzer (MCA) (Fig. 5.3a). The output of the MCA was fed to an on board computer, which reported the measured spectra back to a central base station over a WiFi link. This wireless link also provided a centralized command and control system, allowing the detectors to be synchronized and monitored from a single location.

Each detector was housed in a wheeled cart (Fig. 5.3b) so that it could be moved into place between measurements. Every cart was equipped with a GPS antenna (the white disc in the bottom right of Fig. 5.3a) and sensors to record environmental data including temperature, pressure, and humidity, as well as the orientation of the cart. The carts were covered with a Mylar sheet to reflect sunlight and minimize heating of the electronics.



(a) Detector and integrated PMT/MCA, note the ink pen for scale.

(b) One of the detector carts during a measurement.

Figure 5.3: Illustration of the detectors used for the experiment.

Calibrations for every detector were taken at the beginning and end of each experiment using Eu-252, Co-60, Cs-137[1] and Ba-133 laboratory test sources, as well as a camera lens containing a significant amount of thorium. These were used to generate energy calibrations for each detector, as well as to provide reference points for the automatic gain stabilization algorithm used by the detectors. Examination of these calibrations after the conclusion of the measurement campaign indicated that they experienced negligible calibration drift over the course of the experiments.

When modeling the detectors an intrinsic efficiency of 29% was used, determined from the calibration spectra taken at the beginning and end of each experiment. Since the detectors were all placed facing north, they were all modeled as having a face area of 224 cm^2, corresponding to the average face area of the 2" × 4" × 16" crystal in the detector (see Section 3.3.1). Dwell times were taken to be the live-time recorded by each detector and were all approximately equal

[1]The Cs-137 calibration source was not the same as the target source used in the experiment.

to the nominal real-time for each measurement (1800 s).

5.1.3 Site Characterization

In order to use the technique described in Chapter 3, it was necessary to construct a computer model of the geometry of the scene. This was done by combining differential GPS (DGPS) measurements and georeferencing them with satellite photography of the scene. The DGPS system used for the measurements is notionally capable of sub-centimeter accuracy under ideal circumstances; however, for most measurements reported accuracies of 3 to 10 cm were typical. Manual measurements of several objects in the scene were also recorded using a tape measure to validate the dimensions computed from the DGPS points.

The geometry and DGPS reference points are shown in Fig. 5.4. When constructing the model, I decided to omit several objects in the scene. The detector response model of Section 3.3 only computes the uncollided flux arriving at the detectors, hence objects that do not lie between a source and a detector have no effect on the model's estimate of the count rate.

Material macroscopic cross-sections were assigned on a "best estimate" basis. Based on a photographic survey and measurements derived from the DGPS data, I computed dimensions for each object. Next, I consulted a standard handbook to determine material compositions [28] and then used the NIST XCOM database to compute the nominal microscopic cross-sections[2] [2]. Finally, the microscopic cross-sections were homogenized over the volume of the object to compute macroscopic cross-sections, accounting for the estimated internal dimensions and voids. These macroscopic cross-sections are visualized in terms of the corresponding mean free paths in Fig. 5.4, with the color corresponding to the mean free path in that object, categorized into four groups representing different qualitative levels of attenuation; e.g., black and red are strong attenuators, while yellow and blue are less so. I emphasize that accurate selection of these values is not critical since they only serve as a reference point for assigning the cross section uncertainties. The localization algorithm randomly varies the macroscopic cross section values over a wide range during the sampling process and does not rely on the reference values directly.

The model geometry is strictly two-dimensional, hence it does not account for changes in elevation across the scene. The site where the experiment was performed is approximately flat, with the exception of a downward drop-off in the grassy area located in the northeast section of the site. During the second experiment, one of the detectors was positioned on this hill, resulting in an elevation change of 1 to 2 m relative to the source position (this placement is

[2] Microscopic cross-sections were taken to be the total interaction cross-section at 662 keV as we are only concerned with computing the uncollided flux arriving at the detector. This violates our assumption that we know nothing about the source, however, the selection of this value is not critical since we will be treating them as highly uncertain in our later analysis. Furthermore, I analyzed the total counts acquired by each detector, not the counts under the 662 keV photopeak.

shown in Fig. 5.5). This was a deliberate choice, intended to introduce some of the variability expected in a real-world deployment. Other minor variations in elevation were also present and were similarly not accounted for in the geometry.

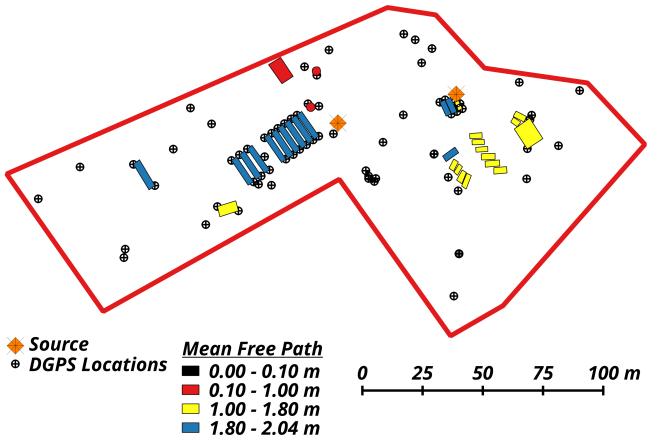


Figure 5.4: Model geometry of the site, colored by the calculated mean free path. The DGPS reference points are also indicated by the black marks. Note that there are some small disagreements with the objects visible in the satellite imagery, this is due to the satellite photos being taken on a different day than that of the experiments.

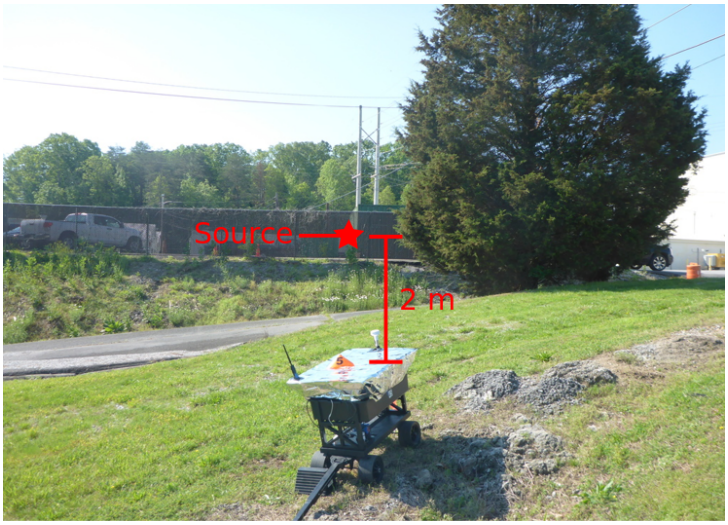


Figure 5.5: Elevation difference in experiment 2. The source is located just behind the fence visible in the background.

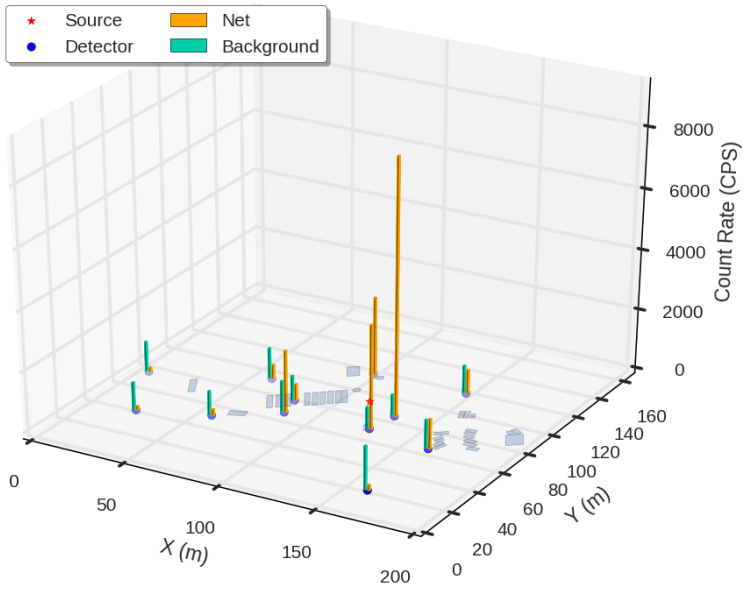
5.2 Results and Analysis

Table 5.2 summarizes the measurements recorded during the experiments, while Fig. 5.6 plots the net foreground and background count rates recorded during both experiments. Total counts refer to the total counts recorded during measurements with the source present, while background counts were accumulated from measurements with the source absent, and net counts are calculated from measurements with the source present subtracting the background counts after normalization for differences in live time. Table 5.2 also shows the ratio of total counts to background, which was approximately 3 to 13 for detectors near the source, while detectors on the perimeter had total to background ratios between 1 and 2. As can be seen in Fig. 5.6, detectors with a direct line-of-sight to the source recorded the highest number of counts and hence contribute the most to the localization, while the Bayesian formulation guarantees robustness against detectors with lower count rates since their contributions to the localization are weighted by their uncertainty.

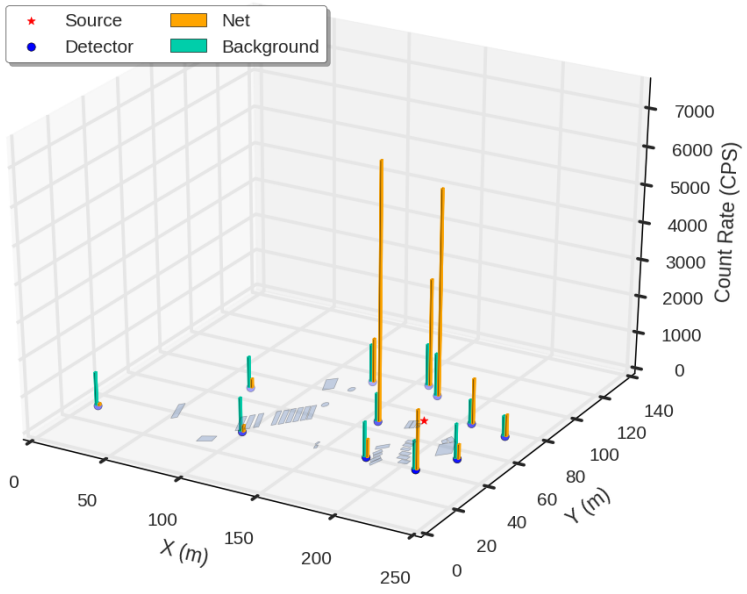
Table 5.2: Count rates recorded by the detectors. Entries marked with a (∗) denote measurements with anomalous count rates (see Section 5.2.2). Total-to-background ratios and SNRs are adjusted for live time.

Name	Total Live Time (s)	Total Counts	Background Live Time (s)	Background Counts	Total / Bg	SNR[3]
1A-1	1,789.13	2,008,213	1,716.42	1,702,957	1.13	175.0
1A-2	1,789.52	1,836,856	1,790.66	1,615,817	1.14	174.8
1A-3	1,785.44	2,656,026	1,789.15	1,815,188	1.47	627.5
1A-4	1,789.72	1,772,977	1,746.49	1,373,245	1.26	308.3
1A-5	1,770.83	5,497,133	1,790.07	1,841,017	3.02	2723.8
1A-6	1,756.45	7,310,350	1,739.25	1,221,574	5.93	5471.1
1B-1	1,787.34	2,407,008	1,791.92	1,430,501	1.69	820.6
1B-2	1,764.57	6,513,106	1,789.12	1,954,174	3.38	3303.2
1B-3	1,783.96	3,053,585	1,790.91	1,632,757	1.88	1119.1
1B-4	1,710.37	15,719,863	1,792.64	1,272,534	12.95	13164.6
1B-5	1,781.88	3,474,658	1,790.64	1,699,812	2.05	1371.1
1B-6	1,763.21	2,852,067	1,784.61	2,570,414	1.12	196.1
2A-1*	1,782.81	3,347,193	1,793.62	1,150,058	2.93	2061.5
2A-2	1,786.82	2,385,999	1,789.80	1,674,788	1.43	552.2
2A-3	1,776.18	4,276,542	1,790.73	1,398,868	3.08	2452.7
2A-4*	1,722.84	13,224,755	1,791.80	1,357,271	10.1	10434.1
2A-5	1,738.96	11,710,996	1,788.52	2,055,420	5.86	6870.4
2A-6	1,776.83	3,932,848	1,788.74	1,814,991	2.18	1586.3
2B-1	1,790.52	1,736,153	1,791.23	1,620,411	1.07	91.4
2B-2	1,789.14	1,945,903	1,791.20	1,501,151	1.30	364.6
2B-3	1,788.92	1,967,095	1,790.03	1,641,784	1.20	254.8
2B-4	1,759.72	7,098,048	1,788.51	2,005,485	3.60	3648.3
2B-5	1,789.42	2,050,384	1,794.64	983,992	2.09	1079.5
2B-6*	1,785.37	2,594,652	1,789.89	1,702,360	1.53	688.0

[3]The definition for signal-to-noise ratio in the context of radiation detector measurements is somewhat ambiguous. In Table 5.2, I choose to define the SNR as $C_{\text{Net}}/\sqrt{C_{\text{Bg}} \cdot \ell_{\text{Tot}}/\ell_{Bg}}$, where C denotes counts and ℓ the live time. That is, the ratio of the net counts to the standard deviation of the background noise, adjusted for differences in live time between the foreground and background measurements.



(a) Experiment 1.



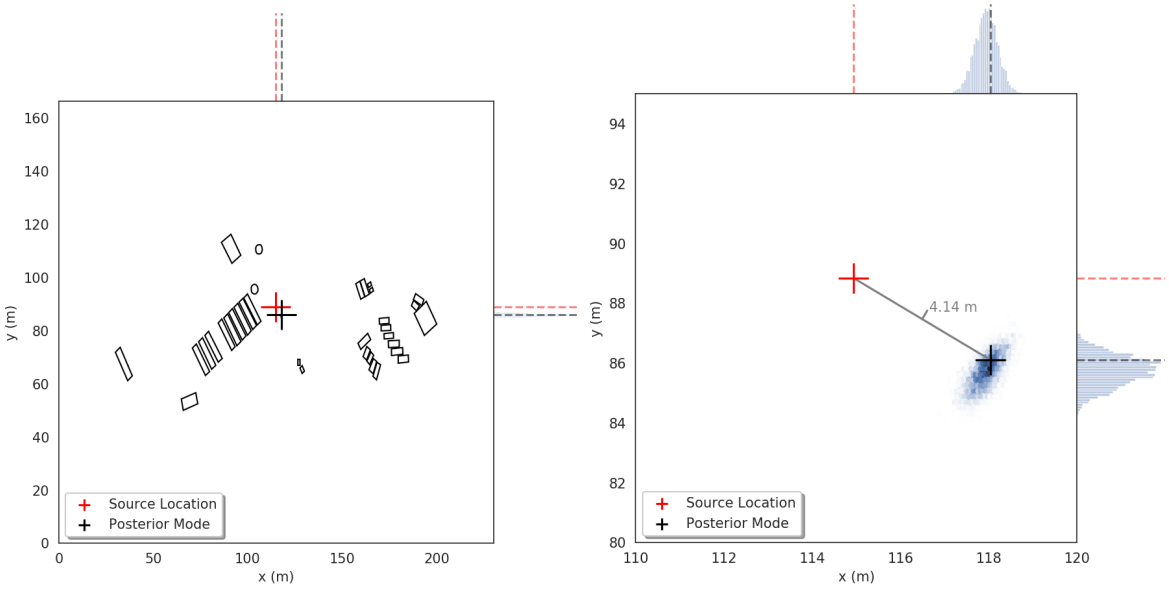
(b) Experiment 2.

Figure 5.6: Foreground and background count rates in both experiments.

Figures 5.7 to 5.8 show the marginal posterior densities for the source position from experiments 1 and 2 using the detector response model that does not account for detector orientation

(see Section 3.3.1). These posteriors were generated by drawing $5 \cdot 10^5$ samples[4], with convergence being determined by examining both the chain autocorrelation and Gelman-Rubin scores. The measurement data was taken to be the net counts recorded by each detector and did not exploit any spectral features as it would require a priori knowledge of the source to identify the relevant features of interest. Background was treated as a Poisson random field, with mean values at each detector location taken from the measurements made without the source present. All cross-section values were assigned a uniformly distributed uncertainty within $\pm 50\%$ of their respective nominal values.

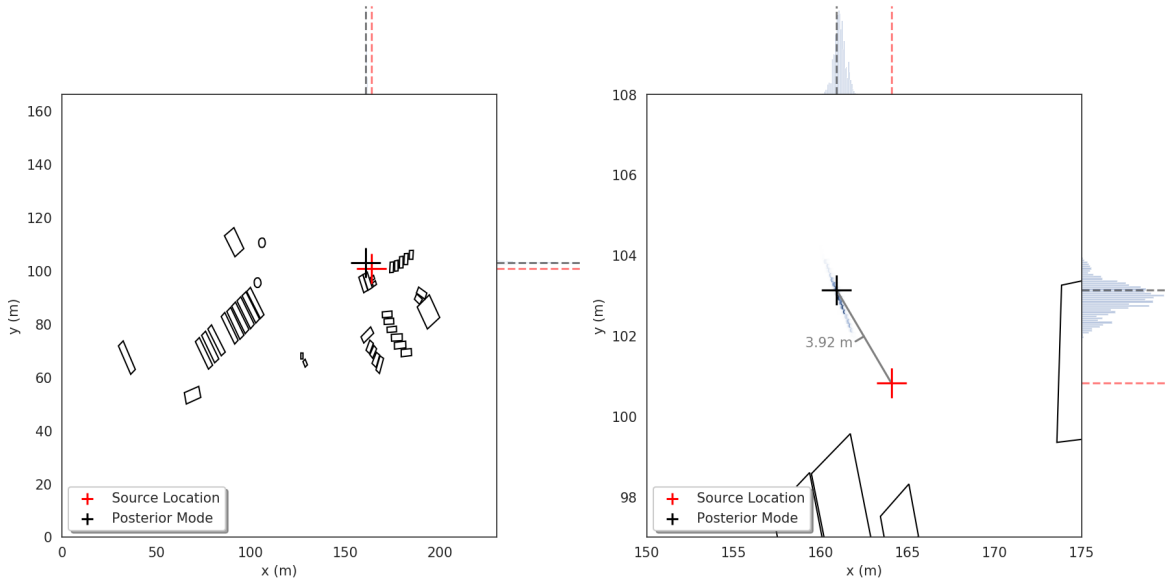
In Fig. 5.7a and Fig. 5.8a we see that the method is able to identify the source location with reasonable accuracy relative to the scale of the problem. On further inspection in Fig. 5.7b and Fig. 5.8b we can see that there is a distinct bias in the estimated source position, one which exceeds the posterior uncertainty that arises from the ad hoc estimation of the cross-sections and measurement uncertainties. These biases, which are on the order of 3 to 4 m, are small enough that the estimates are sufficiently accurate to guide a more targeted search, which is the primary objective. However, we shall proceed to examine them in more detail in order to account for detector orientation.



(a) View of full scene. (b) Detail.

Figure 5.7: Marginal posterior density for source location $\mathbf{r}$ in experiment 1 with 50% relative uncertainty in all cross-sections.

[4]All subsequent posteriors shown were also constructed using $5 \cdot 10^5$ samples.

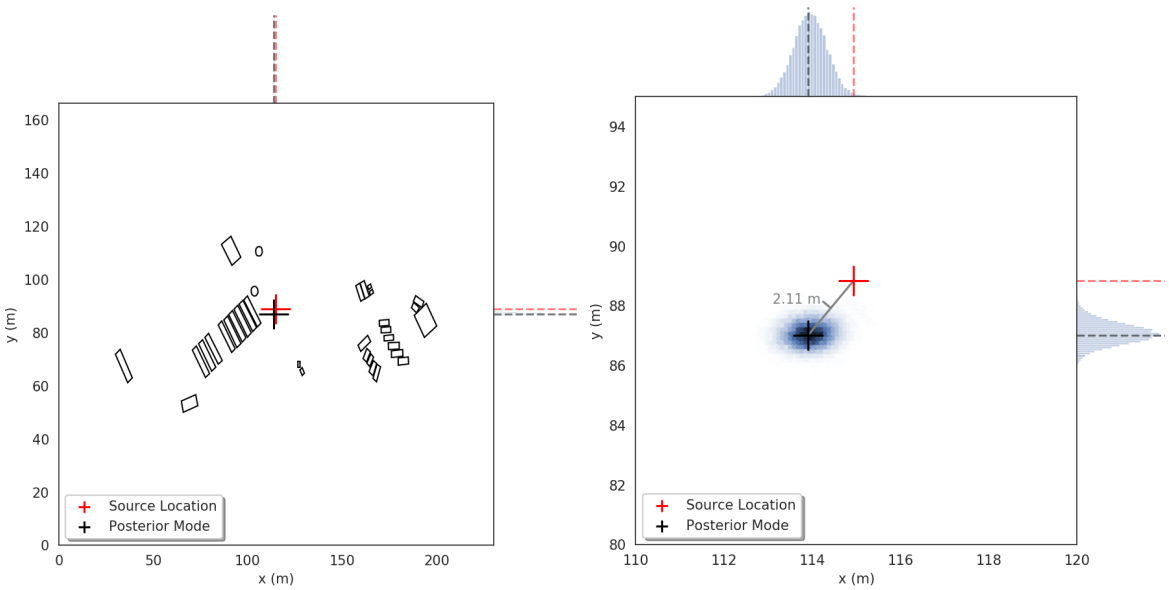


(a) View of full scene. (b) Detail.

Figure 5.8: Marginal posterior density for source location **r** in experiment 2 with 50% relative uncertainty in all cross-sections.

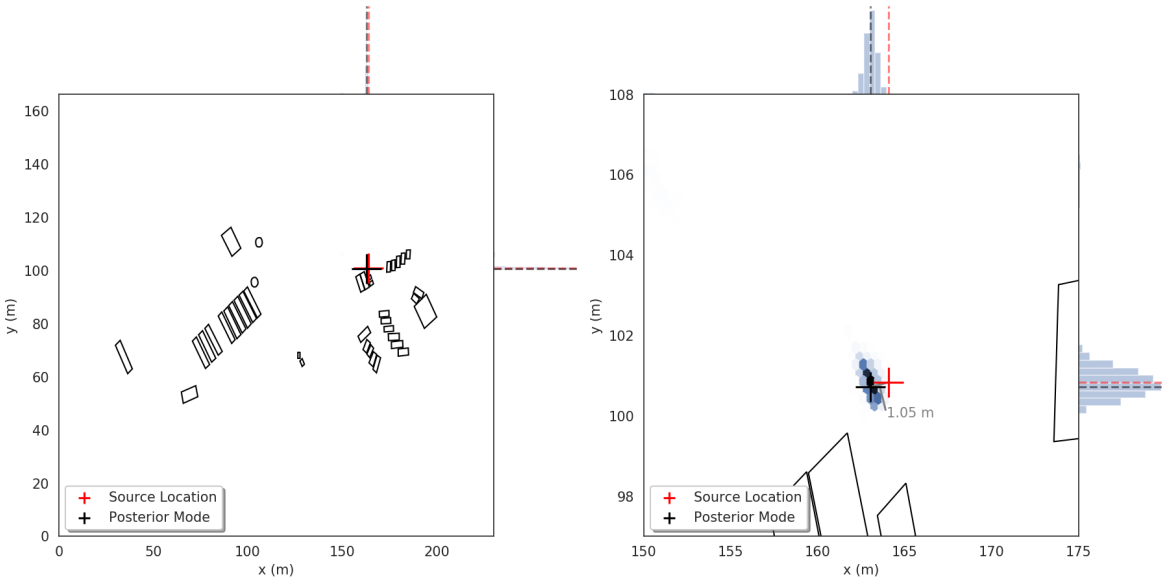
5.2.1 Detector orientation

As discussed in Section 3.3.1, the geometric efficiency of the detectors used for the measurements varies significantly with the orientation of the detector relative to the source. Figures 5.9 to 5.10 show the posterior densities obtained from the MCMC algorithm using the same parameters as the plots in Figs. 5.7 to 5.8, but with the correction for detector orientation described in Section 3.3.1 applied. Observe that the error in the source location is reduced to 1 m to 2 m when the effect of detector orientation is included. Additionally, the true source location is now much closer to lying within the range of uncertainty of the posterior distribution, indicating that any biases caused by the low fidelity of the physical model and scene geometry are of a similar magnitude to the uncertainty arising from counting statistics and material cross sections.



(a) View of full scene. (b) Detail.

Figure 5.9: Marginal posterior density for source location $\mathbf{r}$ in experiment 1 with 50% relative uncertainty in all cross-sections and including detector orientation.



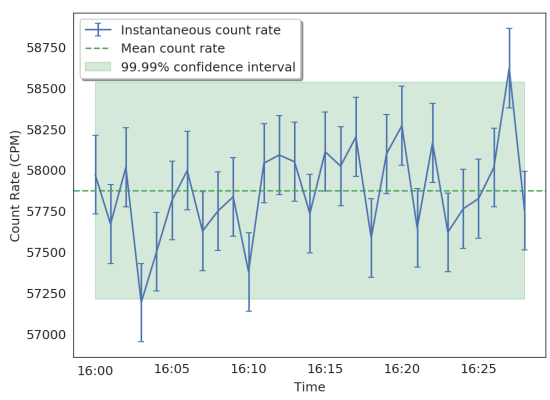
(a) View of full scene (b) Detail

Figure 5.10: Marginal posterior density for source location $\mathbf{r}$ in experiment 2 with 50% relative uncertainty in all cross-sections and including detector orientation.

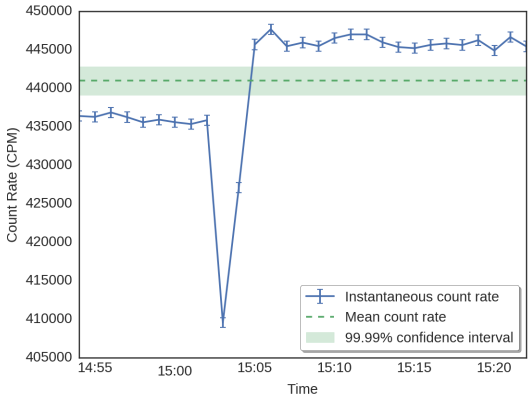
5.2.2 Count rate anomalies during measurements

By examining the footage taken from cameras situated throughout the experiment area, it was discovered that a large vehicle in the eastern parking lot moved during the measurements for experiment 2. The vehicle was located in close proximity to several of the detectors as well as the source, hence its movement produced significant changes in the count rates that were not included in the detector response model. Furthermore, the path traveled by the vehicle crossed between two of the detectors and the source, resulting in a large drop in the count rates measured when the vehicle interrupted their view of the source.

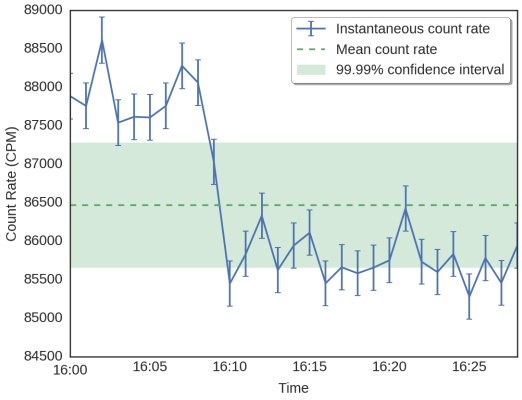
Figure 5.11 shows time series plots of the count rates measured by the three detectors that were affected by the vehicle's movement. Figure 5.11a shows the expected trend for a non-anomalous measurement, with fluctuations about the mean value, but ultimately with the majority of measurements falling within 1-2 standard deviations. Figure 5.11b shows the detector that was situated near the initial position of the vehicle, with a large drop in the count rate visible when the vehicle was passing between the source and detector, followed by an increase in the average count rate after the vehicle had passed due to reduced attenuation. Figure 5.11c shows the count rate recorded by the detector located near the final position of the vehicle, with a sharp drop in the count rate as the vehicle moved between the source and the detector, followed by an overall decrease in the count rate due to increased attenuation in the presence of the vehicle. Finally, Fig. 5.11d shows the detector that was located on the opposite side of the parking lot. Here the change in count rate is smaller than the other two detectors, nevertheless there is a clear trend visible in the recorded count rates. All three cases show statistically significant changes in the observed count rates when compared to the average count rate over the full measurement period and as such I consider them outliers.



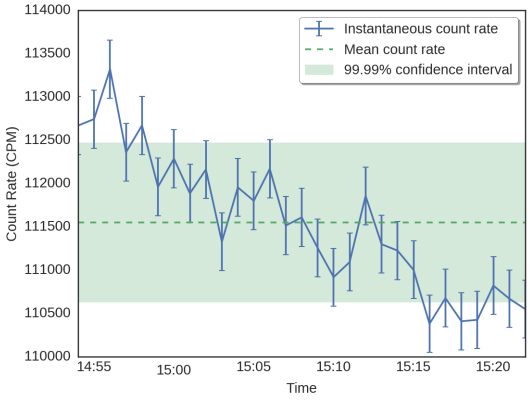
(a) Detector 2-B1.



(b) Detector 2-A4.



(c) Detector 2-B6.

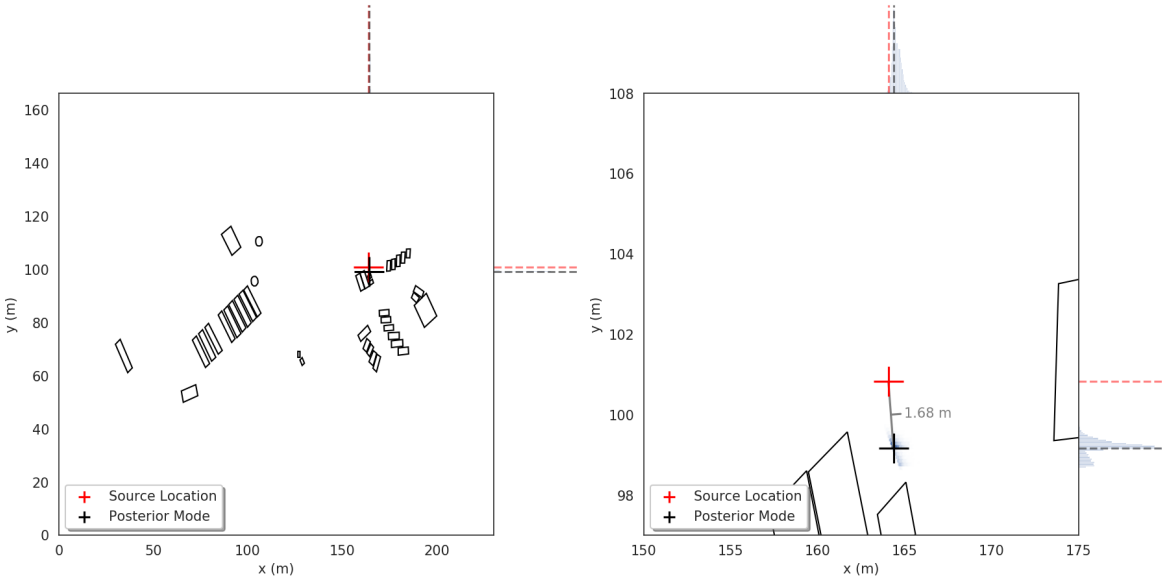


(d) Detector 2-A1.

Figure 5.11: Time series of fluctuating count rates recorded by three of the detectors in experiment 2 (blue) versus the average count rate over the entire measurement (green).

The current implementation of the detector response model does not allow for time-dependent changes in the geometry; instead I elected to discard the data recorded by these detectors. This naturally results in higher uncertainty due to the decreased number of counts recorded by the overall detector network, however the original experimental design included substantial redundancy in the number of detectors in case of such an abnormality so it is expected that localization is still possible even with the problematic measurements omitted. Figure 5.12 shows the marginal posterior that results when excluding these measurements. Interestingly, when compared to the results including the anomalous measurements, the accuracy of the localization is actually slightly *worse*. This suggests that improvement which might be expected from removing the anomalies is offset by the loss of information caused by the increased uncertainty

of having a smaller number of total counts (the detectors removed were among those with the highest total counts recorded). It also suggests that the method is fairly robust to these modeling discrepancies, though I caution that this may not be the case in other scenarios and further experiments are required in order to fully understand this effect. We will discuss these anomalies further in Chapter 6, where I propose a systematic methodology to detect and classify different types of anomaly.



(a) View of full scene (b) Detail

Figure 5.12: Marginal posterior density for source location **r** in experiment 2 with 50% relative uncertainty in all cross-sections, including detector orientation and with counts from anomalous detectors removed.

5.2.3 Effect of Counting Time on Posterior Error

All measurements were taken for approximately 30 min live-time, but shorter measurement times can be simulated by assuming the counts recorded follow a Poisson distribution and resampling measurement data based on the desired dwell time. I used this technique to run the MCMC sampler separately for a series of experiments with increasing counting time to examine its effect on the bias in the estimated source location.

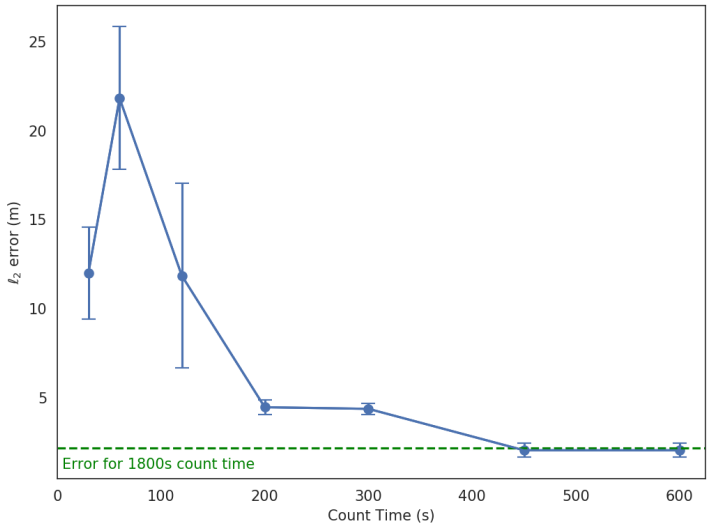


Figure 5.13: Error in estimated source location versus count time for experiment 1.

Figure 5.13 plots the Cartesian distance between the mode of the posterior distribution (the most probable source location) and the true source location versus the count time[5]. There is an overall trend of decreasing bias as the count time is increased, with a count time of 5 min resulting in a decrease in the bias to approximately 2 m. Alternatively, count times beyond 5 min only show modest improvements in accuracy of approximately 1 to 2 m, as well as little change in uncertainty. Together, these observations suggest that a significant portion of the bias observed is dependent on the counting time, but also that other independent sources of bias are present.

5.2.4 Other Sources of Error

Here, I will further identify several possible candidates as the cause(s) of the remaining bias. The first is in the georeferencing of the satellite photography with the DGPS control points taken. This process proceeds by referencing different locations in the overhead imagery with known coordinates, which were recorded using the DGPS system described in Section 5.1.3. An interpolating function is generated using these reference locations to associate geographic coordinates with every point in the image, followed by a transformation to align the image with the new coordinate system[6]. Since this transformation is generally non-linear (and non-affine)

[5]Error bars in Fig. 5.13 are estimated by taking the standard deviation of the distribution that results from propagating samples drawn from the posterior density through the equation for the Euclidian distance to the true source location. This is an imperfect measure of the error; as such, the error bars should be interpreted with caution, particularly in those for lower count times where the posterior densities tend to show multiple distinct modes.

[6]For an overview of the georeferencing process see ref. [14].

it can introduce non-uniform errors, particularly when computing real-world distances between different locations in the image. I hypothesize that small errors in the relative positions of objects in the scene have an effect on the estimated source location, causing a deviation in the location predicted by the posterior from the true source location relative to the position of objects in the scene.

The second potential source of error is in the reconstruction of the geometry and is also related to the georeferencing of the scene. The georeferencing error is an epistemic effect arising from interpolation of the reference points and in particular affects the distances computed between locations, however, there is also an aleatory uncertainty arising from imperfect knowledge of the dimension of objects in the scene. The exterior dimensions of objects were determined via a combination of the DGPS points recorded and manual measurements and have an associated uncertainty. Imperfect measurements may have yielded object dimensions that were too large or too small, resulting in errors in the estimated source position. The exact location of vehicles in the Eastern parking lot is especially relevant, as it was difficult to exactly determine their positions and I had to rely on best estimates.

Lastly, a portion of the bias almost certainly resulted from physical effects that are not accounted for in the detector response model. The model only computes the uncollided flux arriving at the detector, whereas we used net counts recorded as the measurement data. The net count rate obviously includes counts resulting from gammas that have undergone scattering, principally in the vicinity of the detector, and the resulting discrepancy would cause bias if its effect violated the assumption of independently distributed error in the statistical model.

5.2.5 Trilateration

For comparison, Tables 5.3 to 5.4 shows the results of applying trilateration to the data collected during each experiment using the least-squares formulation described in ref. [25]. This is applied without accounting for the presence of objects in the scene, equivalent to the model described in Section 3.3 excluding the exponential term. The resulting least-squares problems were then solved using a standard implementation of the Levenberg-Marquardt algorithm, with error estimates computed via the common method of linearizing the response surface. Table 5.3 shows the predicted source location without considering detector orientation, while the results in Table 5.4 are calculated using geometric efficiencies that have been adjusted for orientation. Run times for these calculations are negligible (less than a second), but the resulting predictions of the source location are significantly poorer than the results using MCMC.

Table 5.3: Predicted source locations using simple trilateration with error estimates without accounting for detector orientation.

Experiment	x (m)	y (m)	σ_x (m)	σ_y (m)	Residual norm (m)
1	148.9	83.5	8.3	1.8	34.4
2	176.0	76.3	9.1	21.0	13.3

Table 5.4: Predicted source locations using simple trilateration with error estimates including detector orientation. All distances are in meters.

Experiment	x (m)	y (m)	σ_x (m)	σ_y (m)	Residual norm (m)
1	110.3	76.3	9.1	21.0	13.3
2	145.7	119.4	4.6	6.0	26.2

Chapter 6

Effects of detector anomalies, signal-to-noise ratio, and alternative physics models

As is frequently the case, preliminary discussion of the results and analysis in Chapters 4 to 5 raised several research questions regarding the limits of the localization algorithm and practical aspects of dealing with unexpected phenomena encountered in the field. Specifically, the following three questions were identified:

1. How to detect the anomalous, non-stationary count rates during the measurements described in Section 5.2.2?

2. What is the effect of source intensity on the precision and accuracy of the posterior?

3. What is the impact of using of an even simpler detector model which does not require the estimation of the macroscopic cross sections of objects in the scene?

Section 6.1 will address the detection classification of count rate anomalies during measurements. I present a technique based on a likelihood ratio test, which is able to classify a given measurement according to a statistical model for ideal measurements versus models for different types of anomaly. Section 6.2 then uses the synthetic test problem from Section 4.3 to characterize the impact of source strength on the posterior density in order to estimate the behavior of the algorithm near the lower limits of detectability. Lastly, Section 6.3 studies the impact of a detector model that assumes all objects in the scene are fully opaque to gamma rays using a simplified test geometry.

6.1 Anomaly Classifier

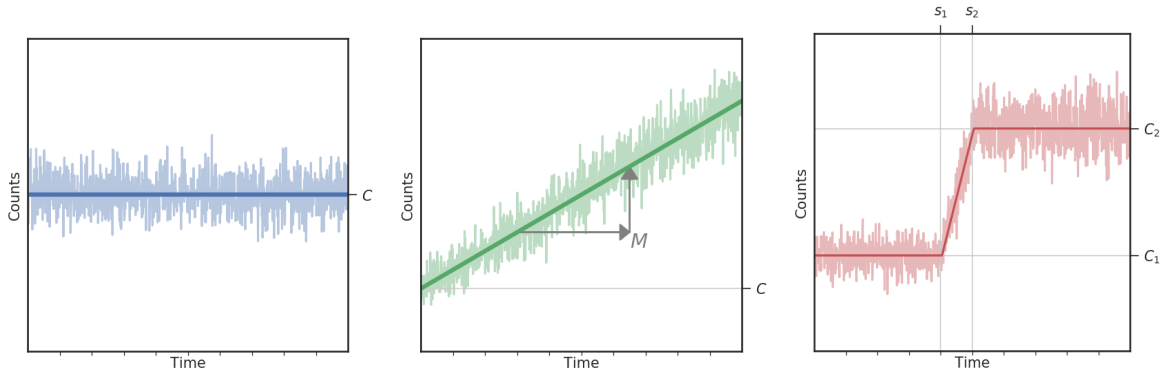
In Chapter 5, we saw that some of the detectors exhibited anomalies in the count rates they recorded. Under ideal conditions, we would expect that the mean count rate exhibited by the detector would remain constant for a static scene, however several of the detectors observed statistically significant changes in the mean count rate. Further, there were two distinct types of trend visible in the measurements. In the first case, a step change was visible in the count rate, caused by detectors being shielded from the source by vehicles moving through the scene during the measurements. In the second case, there is a clear linear trend in some measurements; the exact cause of these anomalies is unknown, but may be related to localized weather-dependent changes in background.

Equations (6.1) to (6.3) provide statistical models for each of these scenarios, modeling the counts collected by a detector with integration time Δt. Equation (6.1) represents the ideal case, where the count rate is fixed at a constant value of C. Equation (6.2) represents a count rate with a linear trend, beginning at rate C and changing at a rate of $M/\Delta t$. Finally, Eq. (6.3) represents a step change from rate C_1 to C_2 during the interval between time s_1 and s_2, assuming the change is linear between C_1 and C_2 for $s_1 < t < s_2$. The behavior of these statistical models and their parameters are illustrated in Fig. 6.1.

$$\begin{aligned} \mathcal{D}_i^{\text{const}}(t_j \,;\, C) &\sim \text{Po}\left[R^{\text{const}}(t_j) \cdot \Delta t\right] \\ R_i^{\text{const}}(t_j) &= C \end{aligned} \tag{6.1}$$

$$\begin{aligned} \mathcal{D}_i^{\text{line}}(t_j \,;\, M, C) &\sim \text{Po}\left[R^{\text{line}}(t_j) \cdot \Delta t\right] \\ R_i^{\text{const}}(t_j) &= M \cdot t_j + C \end{aligned} \tag{6.2}$$

$$\begin{aligned} \mathcal{D}_i^{\text{step}}(t_j \,;\, C_1, C_2, s_1, s_2) &\sim \text{Po}\left[R^{\text{step}}(t_j) \cdot \Delta t\right] \\ R_i^{\text{step}}(t_j) &= \begin{cases} C_1 & t_j \le s_1 \\ C_1 + \frac{C_2 - C_1}{s_2 - s_1}(t_j - s_1) & s_1 < t_j < s_2 \\ C_2 & t_j \ge s_2 \end{cases} \end{aligned} \tag{6.3}$$



(a) Constant. (b) Linear. (c) Step.

Figure 6.1: Illustration of count rate models for different types of anomalies.

We wish to devise a method to determine which of these statistical models best represents the counts recorded by a given detector. To do so, I will perform a **likelihood ratio test**, which compares the ideal model (constant count rate, $\mathcal{D}^{\text{const}}$) to the models for anomalies represented by $\mathcal{D}^{\text{line}}$ and $\mathcal{D}^{\text{step}}$. The case of constant mean count rate is what we would expect to see in the absence of anomalies and hence we will treat it as the null hypothesis model, while the alternative hypotheses are that the detector counts were generated from the anomalous models. Let $\boldsymbol{\theta}_{\text{MLE}}$ denote the parameter values for each of these models that maximize their respective likelihood functions. The log-likelihood ratio of the m-th alternative model, K^m, is defined in Eq. (6.4):

$$\begin{aligned} K^m &:= \ln\left(\frac{\mathcal{L}^m\left(\boldsymbol{\theta}^m_{\text{MLE}}\right)}{\mathcal{L}^{\text{const}}\left(\boldsymbol{\theta}^{\text{const}}_{\text{MLE}}\right)}\right) \quad m \in \{\text{line}, \text{step}\} \\ &= \ell^m\left(\boldsymbol{\theta}^m_{\text{MLE}}\right) - \ell^{\text{const}}\left(\boldsymbol{\theta}^{\text{const}}_{\text{MLE}}\right) . \end{aligned} \tag{6.4}$$[1]

$K^m > 1$ indicates that the likelihood of the alternative model is higher than the null model, while $0 < K^m < 1$ indicate that the null model is favored. Further, define:

$$\begin{aligned} K &:= \max\{K^m\} \\ K_p &:= \frac{K}{N_t} , \end{aligned} \tag{6.5}$$

where N_t is the total number of time bins for the data. K is thus the likelihood ratio for the best (most likely) of the alternative models, and K_p is simply normalized to the number of time

[1] ℓ denotes the log of the likelihood function, $\mathcal{L}$.

bins used when binning the count data so that scores may be compared across different bin widths.

We may now use the value of K_p as a criterion to classify a measurement as belonging to a particular type. Choose a threshold level ϵ (with $\epsilon > 1$), and if $K_p < \epsilon$ we classify a set of measurements according to the null model (constant count rate). Otherwise, we classify the measurements as belonging to the class of the best alternative hypothesis (one of the types of anomaly). Note that it is possible in principle to compute values of ϵ to rigorously guarantee a desired significance level, but this requires construction of the probability distribution for K_p, which is not straightforward in general. Based on the results of applying the method to the measurements from Chapter 5, I choose $\epsilon = 2$, which heuristically provides good discrimination of all anomalous cases that are readily identified by eye. Further analysis and experimental measurements is required in order to evaluate how well this choice of ϵ generalizes to measurements taken under different conditions.

The results of applying the classifier to the measurements from Chapter 5[2] are shown in Tables 6.1 to 6.2, while Figs. 6.2 to 6.7 show the model fits for several of the detectors along with the likelihoods. This includes all three of the anomalous cases described in Section 5.2.2, which are all clearly identified by the classification scheme. Figure 6.2 also shows a typical "ideal" measurement, where the trend is clearly constant. Figures 6.3 to 6.4 depict cases with K_p near the threshold value of $\epsilon = 2$; we can see by eye that a possible linear trend is visible, and the classification results clearly indicate that such cases are on the borderline between different classes. Finally, Fig. 6.5 is perhaps the most interesting case, with an apparent overall constant trend, but with significant outliers in two bins near the beginning and end of the measurement. These outliers can be deceptive to the eye, however the classification algorithm determines that these outliers are not significant enough to conclude that the measurements deviate from constant.

[2] All tests were performed using a time binning of 1 min, though the results were not overly sensitive to the choice of Δt. Δt should be selected to reflect the characteristic duration of the anomalies in question.

Table 6.1: Classifier results for Experiment 1.

Measurement	Best Alternative	K_p	Classification
1A-1	Linear	0.002	Constant
1A-2	Linear	0.021	Constant
1A-3	Linear	0.029	Constant
1A-4	Step	0.017	Constant
1A-5	Step	0.006	Constant
1A-6	Linear	0.035	Constant
1B-1	Step	1.198	Constant
1B-2	Step	0.236	Constant
1B-3	Step	1.585	Constant
1B-4	Linear	1.107	Constant
1B-5	Linear	1.557	Constant
1B-6	Step	0.799	Constant

Table 6.2: Classifier results for Experiment 2. Highlighted rows correspond to the anomalous measurements identified in Section 5.2.2.

Measurement	Best Alternative	K_p	Classification
2A-1	Linear	2.321	Linear
2A-2	Linear	1.947	Constant
2A-3	Linear	0.090	Constant
2A-4	Step	29.345	Step
2A-5	Linear	1.080	Constant
2A-6	Linear	0.777	Constant
2B-1	Linear	0.087	Constant
2B-2	Step	0.130	Constant
2B-3	Linear	0.111	Constant
2B-4	Step	0.032	Constant
2B-5	Linear	1.080	Constant
2B-6	Step	5.437	Step

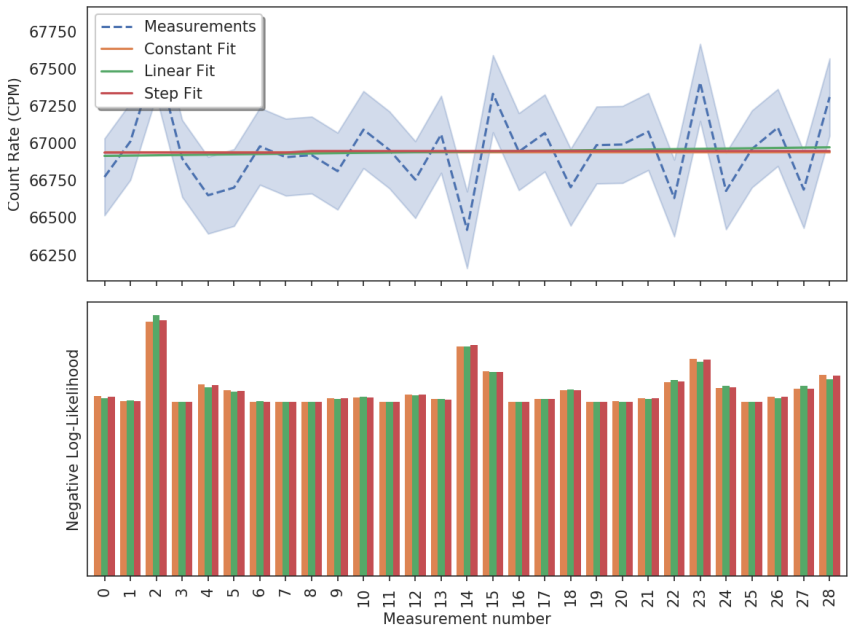


Figure 6.2: 1A-1, $K_p = 0.002$. Constant.

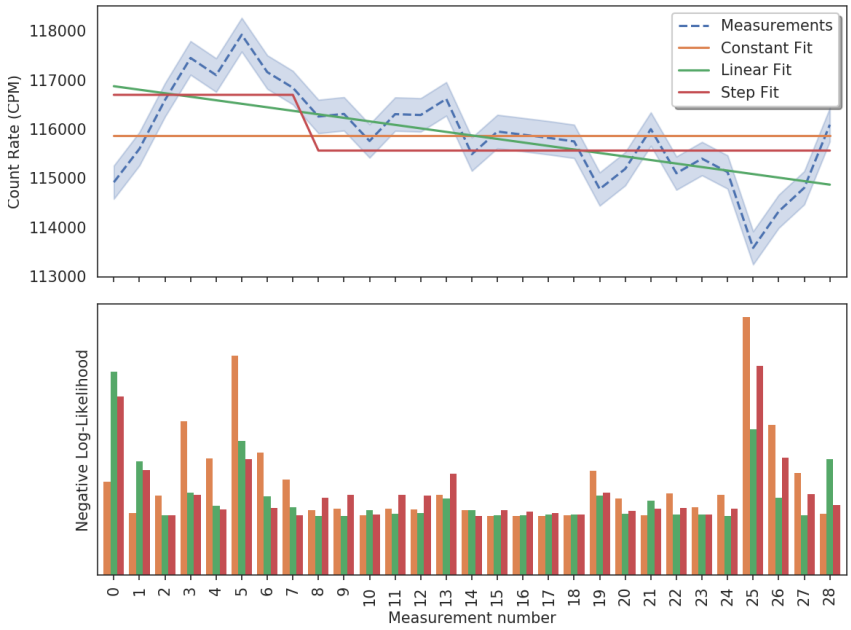


Figure 6.3: 1B-5, $K_p = 1.557$. Constant, but borderline. Some linear trend is visible, but is not significant enough to reject a constant trend.

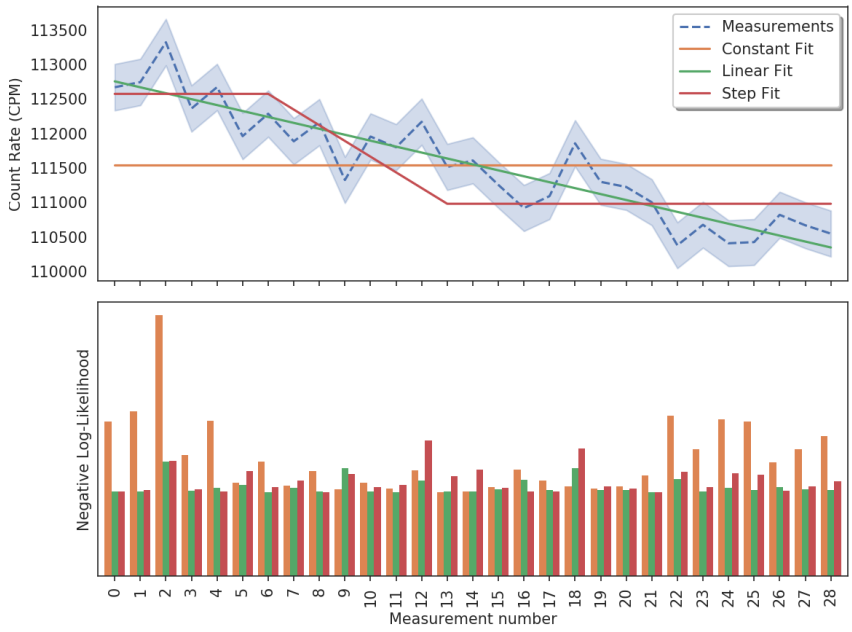


Figure 6.4: 2A-1, $K_p = 2.321$. Linear.

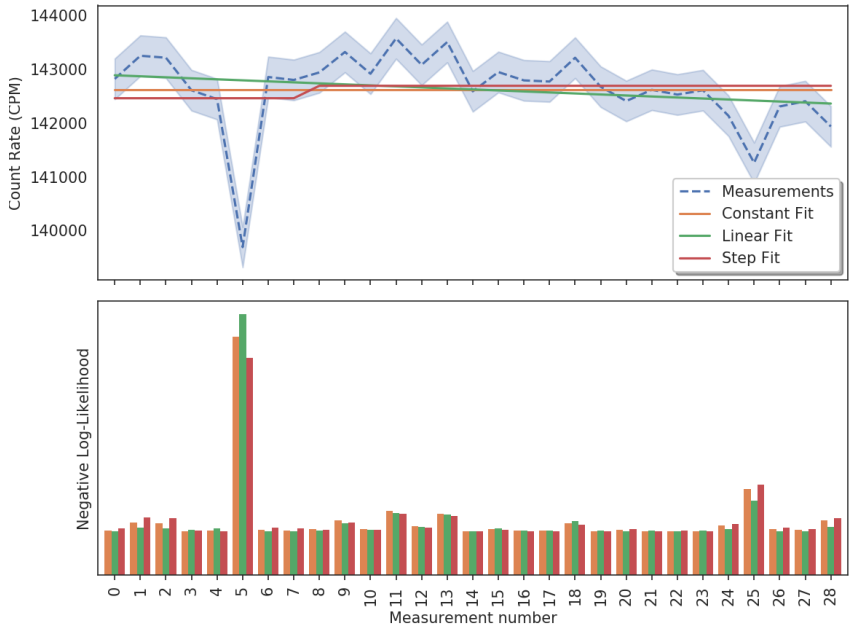


Figure 6.5: 2A-3, $K_p = 0.090$. Constant. Despite the clear drop in count rate near the start, classification indicates overall trend is constant.

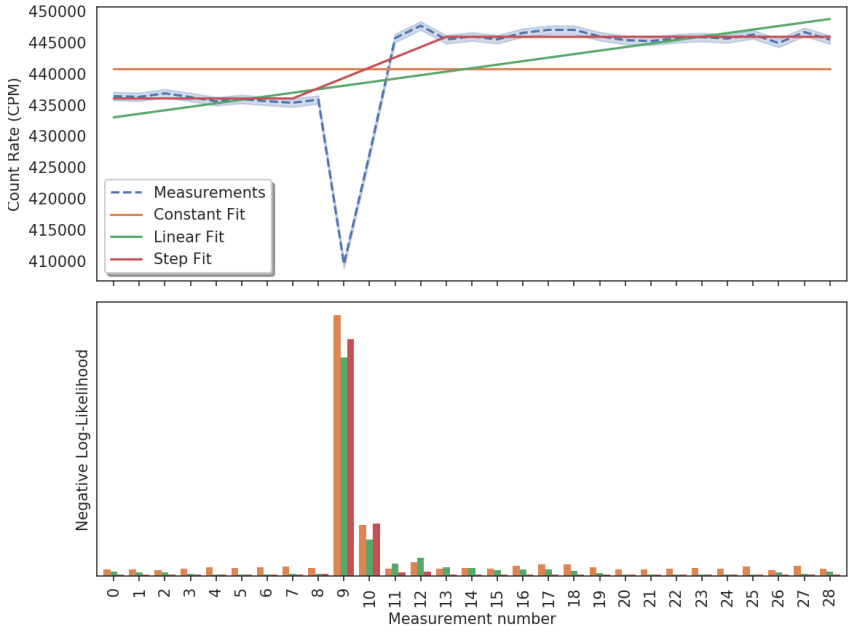


Figure 6.6: 2A-4, $K_p = 29.345$. Step.

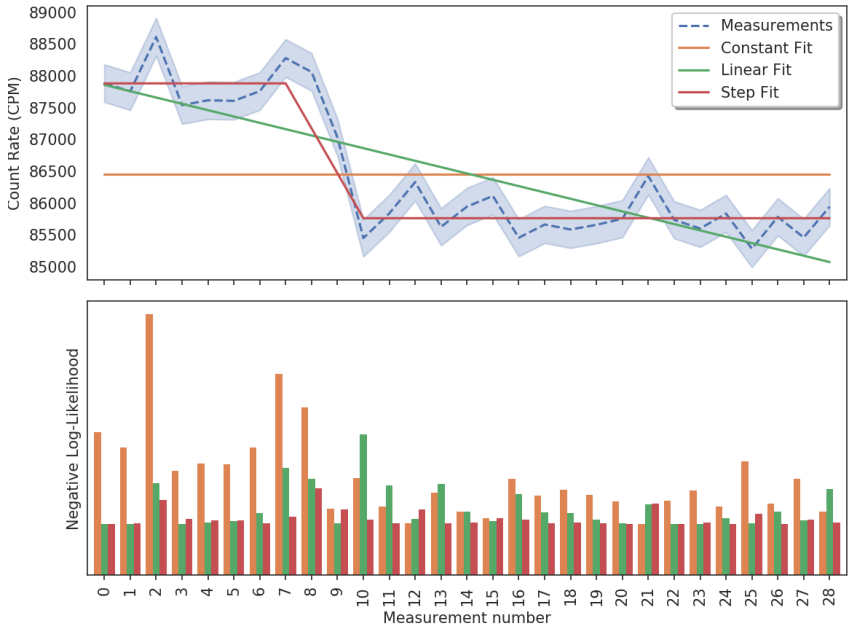


Figure 6.7: 2B-6, $K_p = 5.437$. Step.

6.2 Dependence on Source Intensity

Here, we will investigate the behavior of the posterior distribution as the source intensity is varied. I will use a test problem identical to the one described in Section 4.3, except that the activity of the source will be scaled to a range of values between 1% and 200% of the activity used in Section 4.3 (8.7 mCi) while background is kept constant with a fixed mean of 300 cps. This produces variations in the total signal-to-noise ratio (SNR), defined as the ratio of total counts from the source C_{src} over the uncertainty in the counts due to background[3] $\sqrt{C_{\text{bg}}}$, where counts are accumulated across all detectors:

$$\text{SNR} := \frac{C_{\text{src}}}{\sqrt{C_{\text{bg}}}} \,. \tag{6.6}$$

Figure 6.8 plots the localization accuracy versus source activity, while Fig. 6.9 plots the corresponding posterior distributions. We can see that reasonable localization is achieved down to an activity of approximately 0.87 mCi, while below that level the dispersion of the posterior makes it difficult to resolve a distinct source location. Conversely, for source activities above ∼6.5 mCi there is little improvement in the localization as the uncertainty is driven primarily by the material cross section uncertainties.

[3]The definition of SNR in the context of radiation detection is somewhat ambiguous, with multiple possible interpretations available. Equation (6.6) is derived assuming that the number of counts due to background is known exactly, which is the case here since the measurements are generated from simulation.

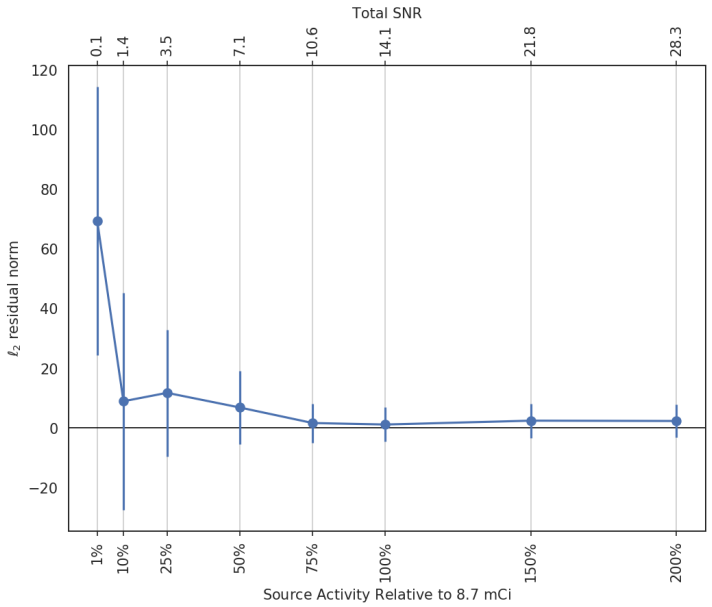
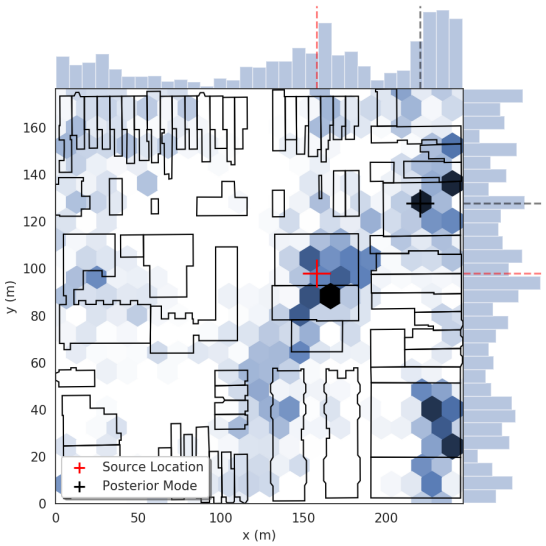
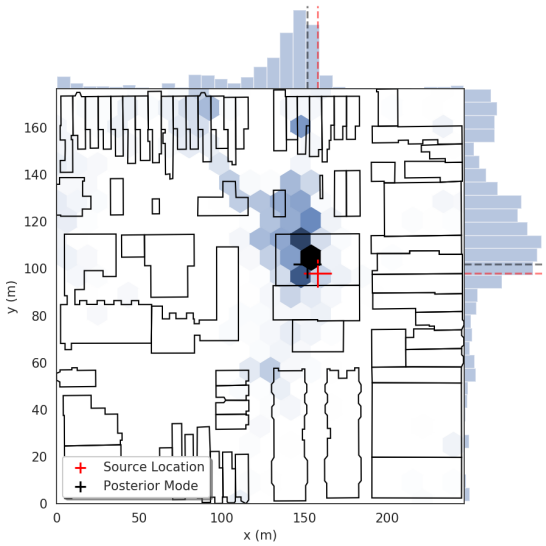


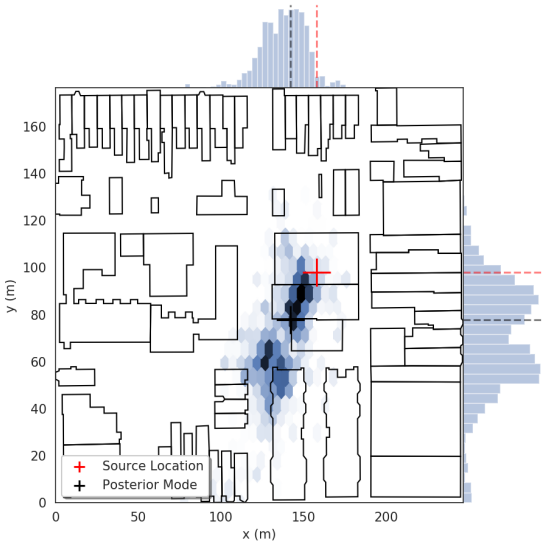
Figure 6.8: Plot of localization error with various source activities.



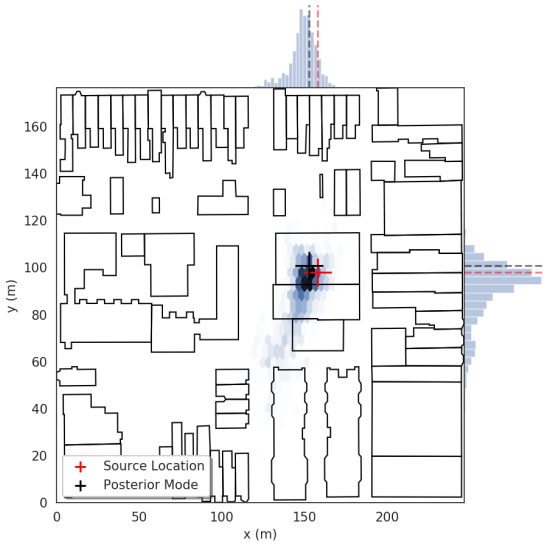
(a) 1% (SNR: 0.14)



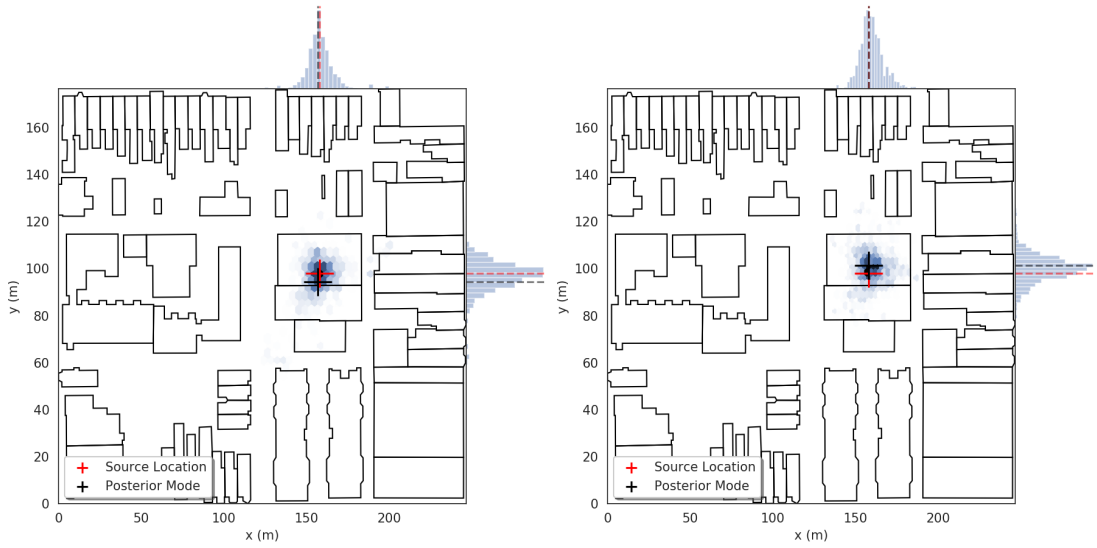
(b) 10% (SNR: 1.41)



(c) 25% (SNR: 3.54)

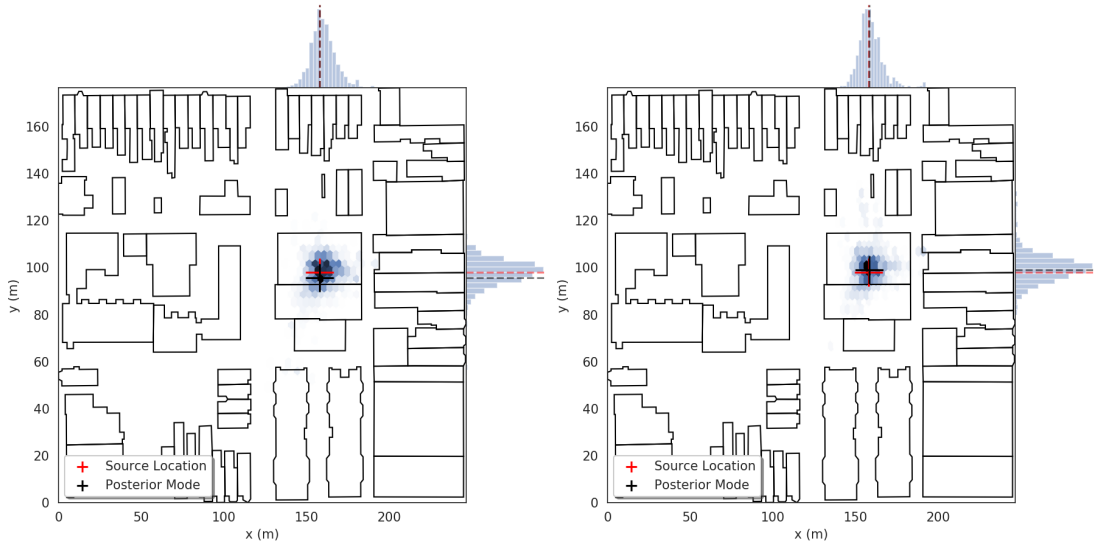


(d) 50% (SNR: 7.07)



(e) 75% (SNR: 10.61)

(f) 100% (SNR: 14.14)



(g) 150% (SNR: 21.76)

(h) 200% (SNR: 28.29)

Figure 6.9: Posterior density with varying source intensity.

6.3 Occluded Detector Model

In Section 3.4, I noted that it is difficult to precisely determine the composition of objects in the scene and developed a method to propagate uncertainties in the estimated cross sections of objects through the MCMC sampler. In this section, I propose an alternative approach using a detector response model that treats all objects in the scene as completely opaque to gamma

rays emitted by the source:

$$d_i^{\text{occ}}(\mathbf{r}, I) = \begin{cases} I\Delta t_i \cdot \epsilon_i^{\text{int}} \cdot \frac{A_i}{\|\mathbf{r}-\mathbf{r}_i\|_2^2} \cdot \exp\left(-\Sigma^{\text{air}} \cdot \|\mathbf{r}-\mathbf{r}_i\|_2\right) & \mathbf{r}_i \text{ visible from } \mathbf{r} \\ 0 & \text{otherwise}\,. \end{cases} \tag{6.7}$$

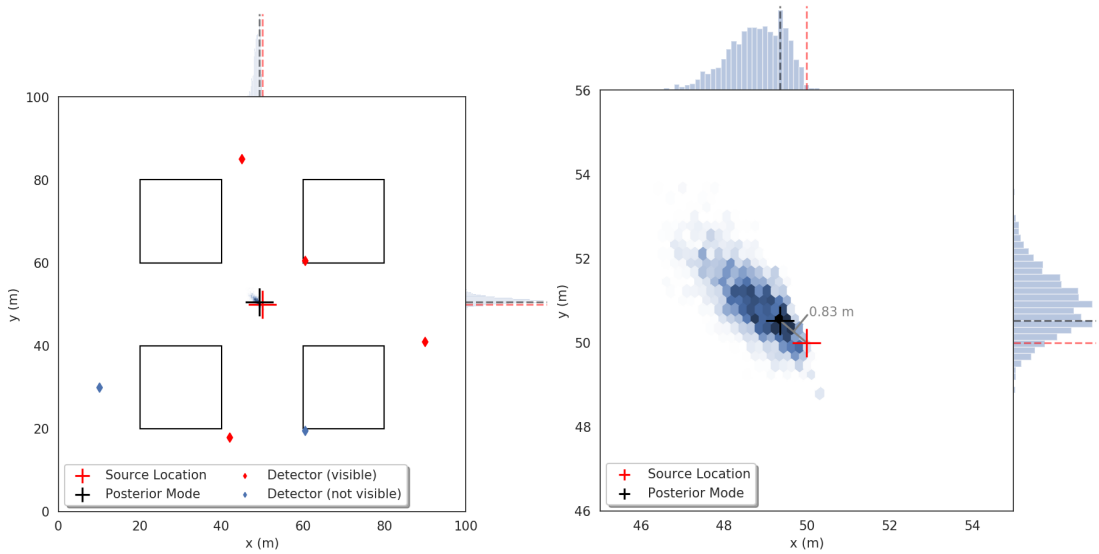
Equation (6.7) uses exponential attenuation in air to compute the response of detectors with line-of-sight to the proposed source location, while detectors that are obscured are assumed to record no counts from the source. A ray tracing algorithm similar to that in Section 3.3 is used to determine the thickness of attenuator interposed between $\mathbf{r}_i$ and $\mathbf{r}$, with a detector being considered visible when this thickness is less than a chosen threshold. In the results that follow, this threshold is chosen as 1 m to allow for detectors that are only lightly obscured to still be considered as visible.

The model in Eq. (6.7) has the advantage that it only requires an estimate for the total cross section of air, which is relatively easy in comparison to estimating the cross section of arbitrary objects in the scene. The primary drawback is that this model will obviously not be able to locate sources situated inside of any buildings or objects in the scene as all detectors will have zero response, regardless of the source location. However, we will see that this model still produces good localization for relatively uncluttered scenes where the source is placed outside.

For the purposes of discussion, I will refer to the model in Eq. (6.7) as the occlusion or occluded model, while the original uncollided flux model from Section 3.3 will be referred to as the attenuation model. Figures 6.10 to 6.17 compare the results using the occlusion model to the attenuation model for a simplified geometry using 6 detectors, with detector and measurement parameters similar to those in the synthetic studies in Section 4.1. Cases using the attenuation model assume cross sections are known exactly for the sake of comparison, with the cross sections of the objects in the scene chosen to be typical of concrete buildings. Note also that the scene geometry is static, however the positions of the source and detectors vary slightly between cases in order to achieve differing degrees of obscuration.

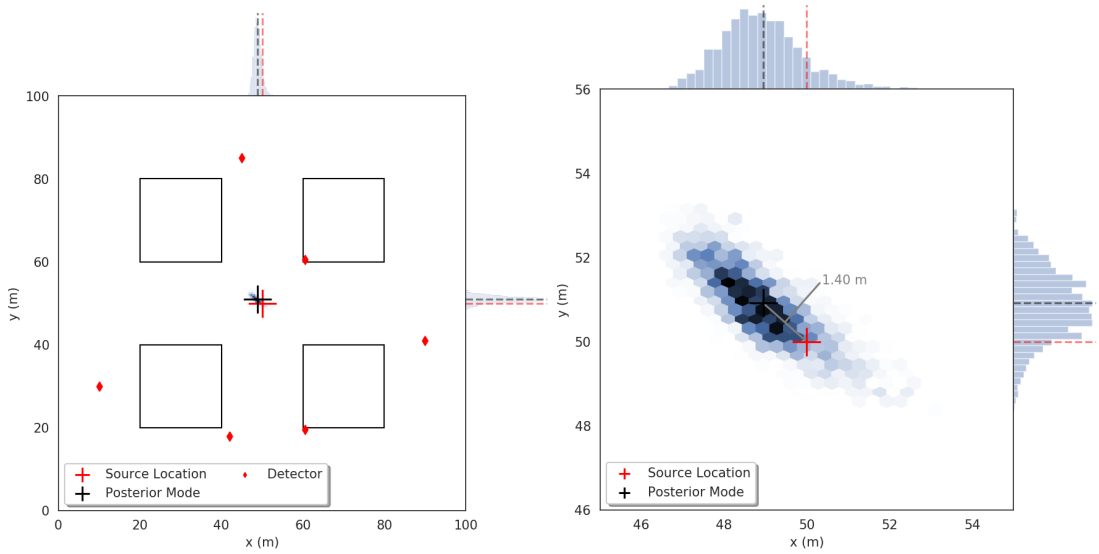
Figures 6.10 and 6.11 show the posteriors resulting from an arrangement where 4 detectors have line-of-sight to the true source location. The posterior generated using the attenuation model is very similar to that which results when using occlusion model. In this case, the occlusion model and attenuation model both predict similar count rates for the detectors which can see the source, and the source lies within the area enclosed by these detectors, and so they produce similar results. Figures 6.12 to 6.13 and Figs. 6.14 to 6.15 show more extreme cases, with only 2 and 1 detector(s) having line-of-sight, respectively. In these cases the difference between the occlusion model and the attenuation model is more pronounced, though the occlusion model is still able to provide some localization of the source, identifying that it is located between the two western buildings. Perhaps most interesting of all is the case shown in Figs. 6.16 to 6.17, where

no detectors are able to see the source directly. Despite this, the posterior using the occlusion model is still able to provide some degree of localization, again constraining the source location to the western area of the domain. In this pathological case, the occlusion model is still able to exclude certain areas of the domain strictly on the basis of visibility.



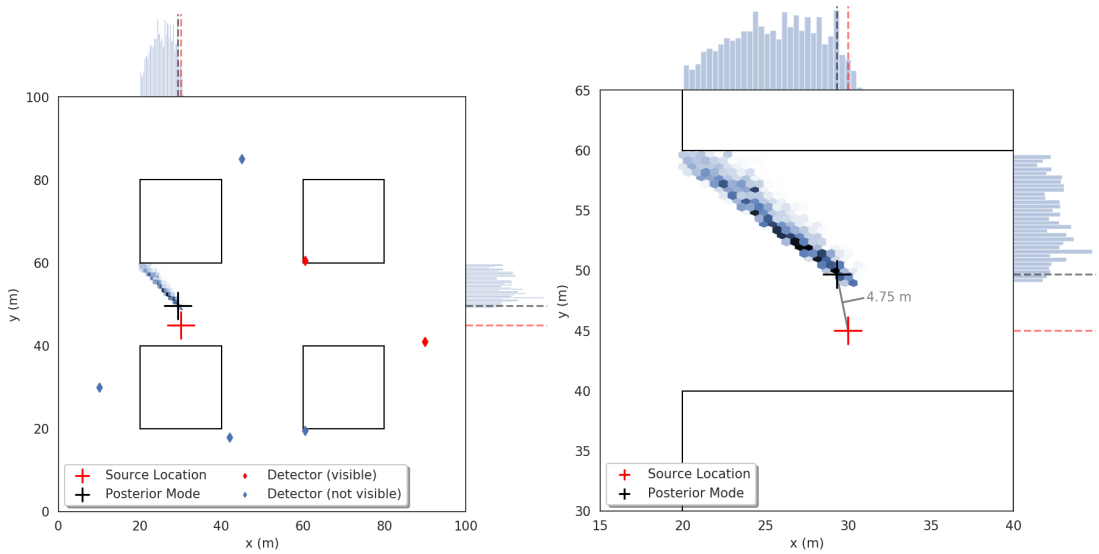
(a) Full Scene. (b) Zoom.

Figure 6.10: Occluded model, 4 detectors with line-of-sight to the source.



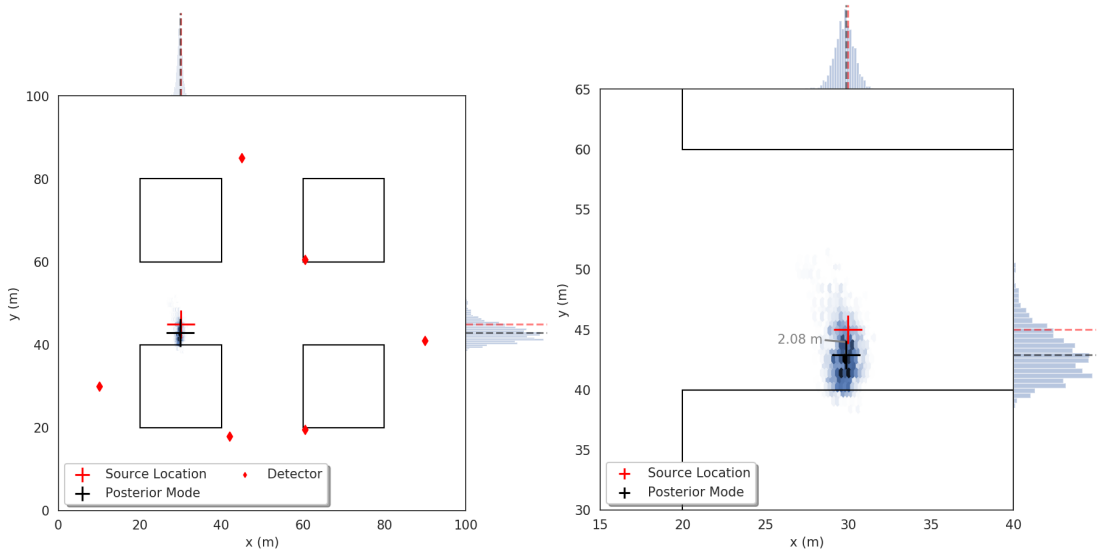
(a) Full Scene. (b) Zoom.

Figure 6.11: Attenuating model, 4 detectors with line-of-sight to the source.



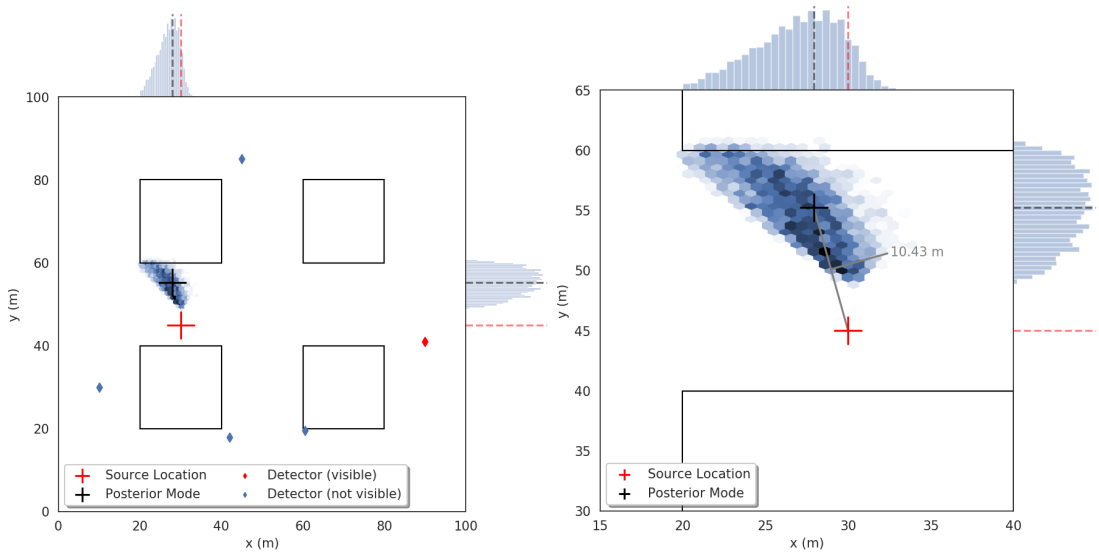
(a) Full Scene. (b) Zoom.

Figure 6.12: Occluded model, 2 detectors with line-of-sight to the source.



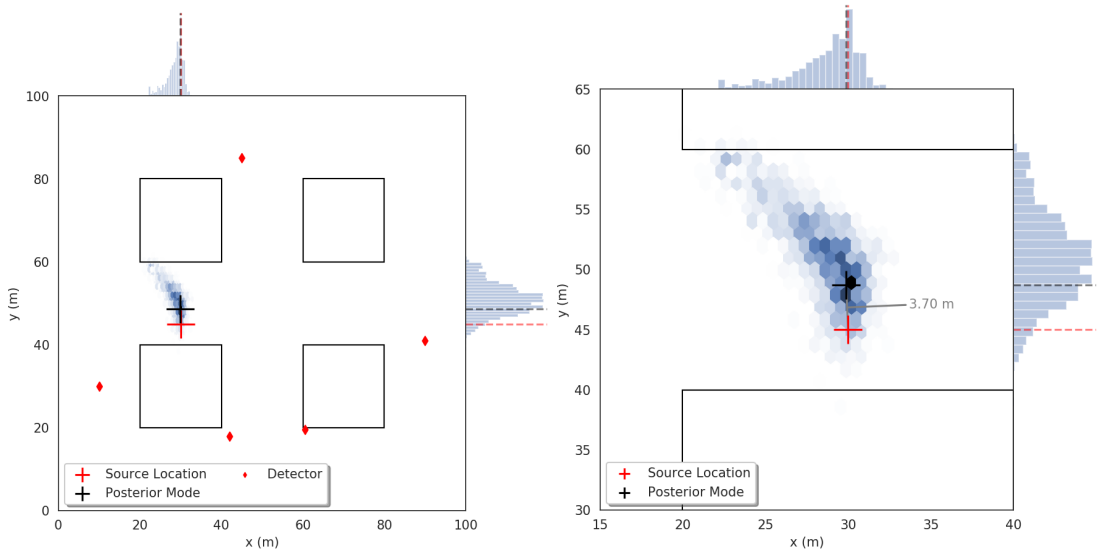
(a) Full Scene. (b) Zoom.

Figure 6.13: Attenuating model, 2 detectors with line-of-sight to the source.



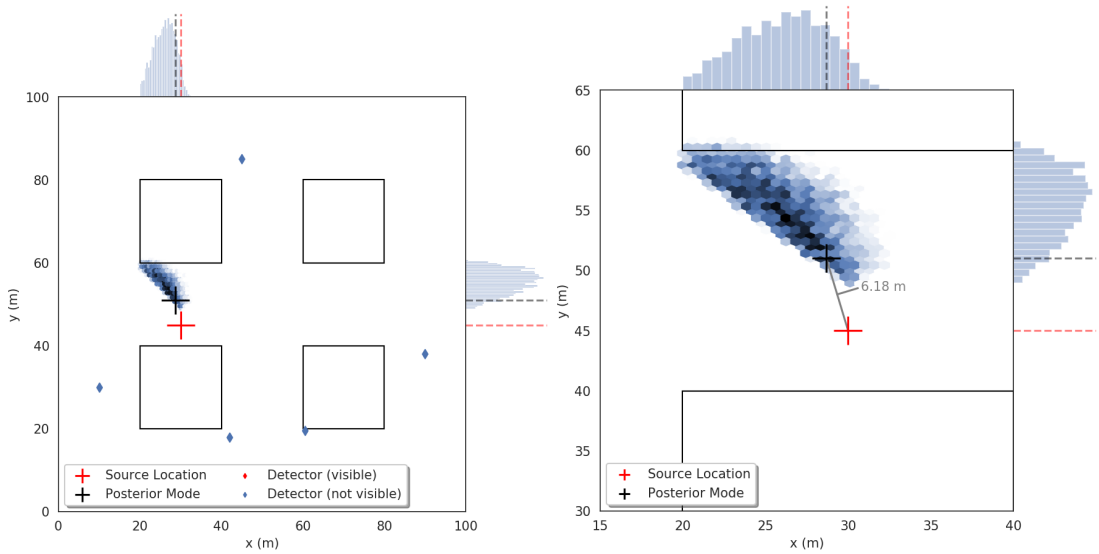
(a) Full Scene. (b) Zoom.

Figure 6.14: Occluded model, 1 detector with line-of-sight to the source.



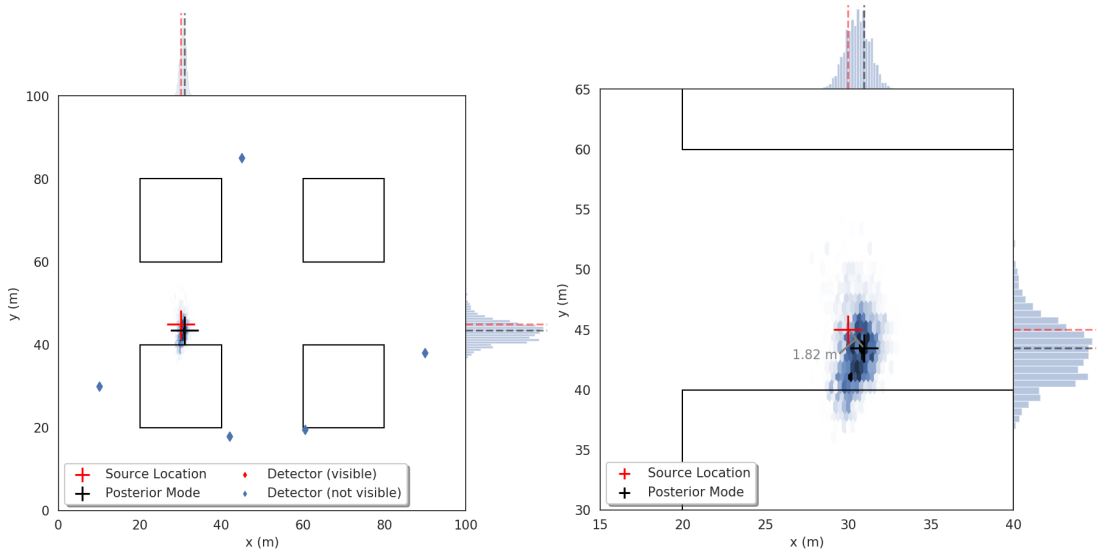
(a) Full Scene. (b) Zoom.

Figure 6.15: Attenuating model, 1 detector with line-of-sight to the source.



(a) Full Scene. (b) Zoom.

Figure 6.16: Occluded model, no detectors with line-of-sight to the source.



(a) Full Scene. (b) Zoom.

Figure 6.17: Attenuating model, no detectors with line-of-sight to the source.

Chapter 7

Conclusions

In this dissertation, I have described a new method for localizing an unknown source of radiation in a heterogeneous environment. I have provided a detailed description of the proposed methodology, which was implemented into a reusable tool for performing source localization. I used this tool to test the limits of the algorithm in several different simulated scenarios, with the results being used to design an experimental test in the field. These measurements were carried out in cooperation with Oak Ridge National Laboratory in May of 2017, and they demonstrated that the algorithm was capable of localizing a 37 mCi source in a scene with composition and scale comparable to a typical city block using real measurements.

The initial work of this research emphasized demonstrating the viability of the method and exploration of its characteristics and limitations. This was performed primarily using synthetic measurements generated using a geometry constructed based on real a city block. These synthetic measurements were generated using the same statistical model for the detector response that was used to model them in the localization algorithm and hence did not include the sort of deviations one would expect in the field; nevertheless the results proved sufficient for demonstrating the basic viability of the algorithm. Analysis was also performed using measurements generated using MCNP (and so, with some of the expected features of real-world measurements), which implied that these un-modeled phenomena would have a minor impact on the results of the localization.

Results from the initial experiments were used to plan an experiment, which was performed in cooperation with Oak Ridge National Laboratory at the site of the former Experimental Gas-Cooled Reactor. A 37 mCi source was placed at two locations in the scene and measured by a set of six mobile detectors, resulting in two independent datasets that were used to test the efficacy of the algorithm. Subsequent analysis suggested that an extension to the detector model was required in order to account for the changes in geometric efficiency due to detector orientation, and after implementing this correction the algorithm was able to localize the source to within

$\sim$2 m for both cases using a 30 min count time. Analysis also suggested that shorter count times could be used, with count times in excess of 450 s producing little change in the posterior density.

Discussion of the results of the experiments in the field led to several further points of investigation. Firstly, it was discovered that during the second set of measurements interference by vehicles moving through the scene caused unaccounted-for variations in the detector count rates. These variations were shown to have little effect on the localization, however as part of this investigation a method was developed to automatically detect and classify various types of anomaly. Second, studies were performed to examine the accuracy and precision of the localization versus the activity of the source relative to background. These studies suggest that the algorithm performs well for signal-to-noise ratio values down to ~ 0.5. Finally, the last case examined the use of a detector model which only modeled the occlusion of sources by objects in the scene, treating any such objects as fully opaque. This model has the advantage of not requiring estimates of the cross section of interposed objects and, despite some limitations, it was shown that it is still able to achieve useful localization of the source.

7.1 Summary of Major Results

The focus of this research is on the formulation a Bayesian approach for the localization of a source of gamma radiation in a heterogeneous urban environment. The major results of this research are:

- Formulated a method for source localization based on Bayesian parameter estimation via Markov-chain Monte Carlo sampling.
- Implemented a general framework for predicting the response of a network of detectors in a heterogeneous geometry by ray-tracing.
- Extended the Metropolis sampler to propagate fixed parameter uncertainties in the material cross sections onto the posterior density for the source location.
- Demonstrated the localization algorithm in two test problems with synthetic measurement data and measurement data simulated using MCNP.
- Performed a study on the error introduced by uncertainty in the building cross sections using synthetic measurement data, with results indicating that the uncertainty in the predicted source location remains reasonable even with highly uncertain cross section data.
- Designed and performed an experiment to test the localization algorithm in the field using a Cs-137 source.

- Implemented the effect of physical orientation of the detectors into the model for the detector counts. Results using this model show a significant improvement in localization accuracy.
- Analyzed the effect of count rate anomalies in some of the experimental measurements.
- Implemented a technique to automatically detect and classify count rate anomalies.
- Evaluated the effect of source activity on the accuracy and precision of localization as a function of signal-to-noise ratio.
- Implemented a further simplified detector model that treats objects as fully opaque and demonstrated that this model is still capable of localizing a source in some situations.

REFERENCES

[1] Y. Atchadé, G. Fort, E. Moulines, and P. Priouret. Adaptive Markov chain Monte Carlo: theory and methods. *Bayesian Time Series Models*, pages 32–51, 2011.

[2] M. Berger, J. Hubbell, S. Seltzer, J. Chang, J. Coursey, R. Sukumar, D. Zucker, and K. Olsen. XCOM: photon cross sections database (NIST standard reference database 8). Online database, National Institute of Standards and Technology, 2010.

[3] I. Bronshteĭn. *Handbook of mathematics*. Springer, Berlin New York, 2007.

[4] R. Chudley. *Building construction handbook*. Routledge, London New York, 2014.

[5] B. Deb. Iterative estimation of location and trajectory of radioactive sources with a networked system of detectors. *IEEE Transactions on Nuclear Science*, 60(2):1315–1326, 2013.

[6] J. Favorite, K. Bledsoe, and D. Ketcheson. Surface and volume integrals of uncollided adjoint fluxes and forward-adjoint flux products. *Nuclear Science and Engineering*, 163:73–84, 2009.

[7] A. Gelfand and A. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

[8] A. Gelman, G. Roberts, and W. Gilks. Efficient Metropolis jumping rules. In *Bayesian statistics*, volume 5, pages 599–607. Clarendon Press, 1996.

[9] A. Gelman and D. Rubin. A single series from the Gibbs sampler provides a false sense of security. *Bayesian statistics*, 4:625–631, 1992.

[10] A. Gelman and D. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

[11] A. Gelman and K. Shirley. Inference from simulations and monitoring convergence. In *Handbook of Markov Chain Monte Carlo*, pages 163–174. CRC Press/Taylor & Francis, 2011.

[12] T. Goorley, M. James, T. Booth, F. Brown, J. Bull, L.J. Cox, J. Durkee, J. Elson, M. Fensin, R.A. Forster, et al. Initial MCNP6 release overview. *Nuclear Technology*, 180(3):298–315, 2012.

[13] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2), 2001.

[14] A. Hackeloeer, K. Klasing, J. Krisp, and L. Meng. Georeferencing: a review of methods and applications. *Annals of GIS*, 20(1):61–69, 2014.

[15] W. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[16] J. Hite and J. Mattingly. Bayesian Metropolis methods for source localization in an urban environment. *Radiation Physics and Chemistry*, 155:271–274, 2019.

[17] J. Hite, J. Mattingly, K. Schmidt, R. Stefanescu, and R. Smith. Bayesian Metropolis Techniques for Source Localization (Poster). In *IEEE Symposium on Radiation Measurements and Applications*, Berkeley, CA, May 22-27 2016.

[18] S. Horne, G. Thoreson, L. Theisen, D. Mitchell, L. Harding, and W. Amai. GADRAS-DRF 18.5 user's manual (SAND2014-20625). *Sandia National Laboratory*, 23, 2014.

[19] J. Howse, L. Ticknor, and K. Muske. Least squares estimation techniques for position tracking of radioactive sources. *Chemical Engineering*, 37:1–15, 2001.

[20] IAEA. Safeguards glossary, 2001 edition. *International Nuclear Verification Series*, 3:23, 2002.

[21] P. Jacob, C. Robert, and M. Smith. Using parallel computation to improve independent Metropolis-Hastings based estimation. *Journal of Computational and Graphical Statistics*, 20(3):616–635, 2011.

[22] K. Jarman, E. Miller, R. Wittman, and C. Gesh. Bayesian radiation source localization. *Nuclear Technology*, 175(1):326–334, 2011.

[23] E. Kaplan and C. Hegarty. *Understanding GPS: principles and applications*. Artech House, 2005.

[24] D. Levin and Y. Peres. *Markov chains and mixing times*, volume 107. American Mathematical Socety, 2017.

[25] Y. Liu, Z. Yang, X. Wang, and L. Jian. Location, localization, and localizability. *Journal of Computer Science and Technology*, 25(2):274–297, 2010.

[26] Z. Liu. *Mobile radiation sensor networks for source detection in a fluctuating background using geo-tagged count rate data*. M.s., University of Illinois at Urbana-Champaign, 2016.

[27] G. Martinez and J. Partlow. Stolen cobalt-60 found in Mexico; thieves may be doomed. *Washington Post*, Dec 2013.

[28] R. McConn, C. Gesh, R. Pagh, R. Rucker, and R. Williams. Compendium of material composition data for radiation transport modeling. Technical report, Pacific Northwest National Laboratory, Richland, WA (US), 2011.

[29] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[30] C. Meyer. *Matrix analysis and applied linear algebra*. Society for Industrial and Applied Mathematics, Philadelphia, 2000.

[31] D. Mitchell and J. Mattingly. Gamma detector response and analysis software (GADRAS) v. 16.0. Technical report, Sandia National Laboratories, 2009.

[32] M. Morelande and B. Ristic. Radiological source detection and localisation using Bayesian techniques. *IEEE Transactions on Signal Processing*, 57(11):4220–4231, 2009.

[33] M. Morelande, B. Ristic, and A. Gunatilaka. Detection and parameter estimation of multiple radioactive sources. *FUSION 2007 - 10th International Conference on Information Fusion*, 2007.

[34] C. Pahlajani, I. Poulakakis, and H. Tanner. Networked decision making for poisson processes with applications to nuclear detection. *IEEE Transactions on Automatic Control*, 59(1):193–198, 2014.

[35] N. Rao, S. Sen, N. Prins, D. Cooper, R. Ledoux, J. Costales, K. Kamieniecki, S. Korbly, J. Thompson, J. Batcheler, R. Brooks, and C. Wu. Network algorithms for detection of radiation sources. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 784:326–331, 2015.

[36] N. Rao, M. Shankar, J. Chin, D. Yau, S. Srivathsan, S. Iyengar, Y. Yang, and J. Hou. Identification of low-level point radiation sources using a sensor network. *International Conference on Information Processing in Sensor Networks*, 7(3):1–35, 2008.

[37] C. Robert and G. Casella. A short history of Markov chain Monte Carlo: subjective recollections from incomplete data. *Statistical Science*, 26(1):102–115, feb 2011.

[38] L. Roberts. Radiation accident grips Goiania. *Science*, 238(4830):1028–1031, 1987.

[39] R. Serfozo. *Basics of applied stochastic processes*. Springer, Berlin Heidelberg, 2009.

[40] R. Smith. *Uncertainty quantification: theory, implementation, and applications*. Society for Industrial and Applied Mathematics, Philadelphia, 2014.

[41] R. Stefanescu, K. Schmidt, J. Hite, R. Smith, and J. Mattingly. Hybrid optimization and Bayesian inference techniques for a non-smooth radiation detection problem. *International Journal for Numerical Methods in Engineering*, pages 1–36, 2017.

[42] C. Sullivan. Radioactive source localization in urban environments with sensor networks and the Internet of Things. In *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 384–388. IEEE, 2016.

[43] C. Ter Braak. Genetic algorithms and Markov chain Monte Carlo: differential evolution Markov Chain makes Bayesian computing easy (revised). Technical report, Wageningen UR, Biometris, 2005.

[44] Author Unknown. NIS Nuclear Trafficking Collection. `http://www.nti.org/analysis/reports/nis-nuclear-trafficking-database/`. Accessed 2018-09-14.

[45] A. Van Oosterom and J. Strackee. The solid angle of a plane triangle. *IEEE Transactions on Biomedical Engineering*, BME-30(2):125–126, 1983.

[46] R. Vilim and R. Klann. RADTRAC: a system for detecting, localizing, and tracking radioactive sources in real time. *Nuclear Technology*, 168:61–73, 2009.