Uncertainty Quantification in Emission Quantitative Imaging

by

Aaron M. Bevill

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Nuclear Engineering and Radiological Sciences)
in The University of Michigan
2017

Doctoral Committee:

Professor William R. Martin, Chair
Keith Bledsoe, Oak Ridge National Laboratory
Professor Jeffrey A. Fessler
Professor Zhong He
Assistant Professor Brian Kiedrowski

ACKNOWLEDGEMENTS

This research was performed under appointment to the Nuclear Nonproliferation International Safeguards Graduate Fellowship Program sponsored by the National Nuclear Security Administration's Next Generation Safeguards Initiative (NGSI).

To Dr. Martin and my committee, for your guidance and for setting high standards; To Paul Hausladen, Matthew Blackston, and others at Oak Ridge National Laboratory, for considerable laboratory assistance; To my friends at University of Michigan, for years of wonderful memories and napkin sketches; To my past mentors at Oak Ridge National Laboratory, Oregon State University, and Texas A&M University, for your help founding my research career; To my parents, for perspective and support:

Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF ABBREVIATIONS

BCAT Glovebox Cleanout Assistance Tool [1, 2]

CPU central processing unit

DREAM DiffeRential Evolution Adaptive Metropolis MCMC algorithm [3, 4]

FNCA fast-neutron coded aperture radiation imaging system

GGH Generalized Geometry Holdup measurement technique [5]

GLLS Generalized Linear Least-Squares statistical inference

HMS4 Holdup Measurement System-4 software implementation of GGH [5]

ISOCS In-Situ Object Counting System measurement technique [6]

KKT Karush–Kuhn–Tucker conditions for constrained optimization [7, 8, 9]

MCLLS Monte Carlo library least-squares spectrum analysis approach [10]

MAP maximum a posteriori Bayesian estimator/objective

MC&A material control and accountability

MCMC Markov chain Monte Carlo

MCNP5 Monte Carlo N-Particle Transport Code version 5 [11]

MCNPX Monte Carlo N-Particle Transport Code version X [12]

ML maximum likelihood Bayesian estimator/objective

MLEM maximum-likelihood expectectation maximization iterative statistical reconstruction algorithm [13]

MURA modified uniformly redundant coded aperture array [14]

NDA Non-Destructive Assay

NWIS Nuclear Weapons Identification System [15]

OLLS Ordinary Linear Least-Squares statistical inference

P24 name of an FNCA detector at Oak Ridge National Laboratory with 24×24 scintillator pixels [16]

PDF probability density function

QI quantitative imaging

QOI quantity of interest—the ultimate goal of the inverse problem

QQ Quantile–Quantile plot

RCR Rapid Convergence Region where Newton's method converges rapidly

ROI region of interest—the spatial domain of the source of interest

RJMCMC reversible-jump Markov chain Monte Carlo [17]

SNM special nuclear material

SOE stochastic origin ensembles iterative statistical reconstruction framework

SVD singular value decomposition

UQ uncertainty quantification

ABSTRACT

Uncertainty Quantification for Emission Quantitative Imaging
by
Aaron M. Bevill

Chair: William R. Martin

Imaging detectors have potential to improve the reliability of plutonium holdup measurements. Holdup measurement is a significant challenge for nuclear safeguards and criticality safety. To infer holdup mass today, inspectors must combine data from counting (non-imaging) detectors with spatial measurements, process knowledge, and survey estimates. This process results in limited certainty about the holdup mass. Imaging detectors provide more information about the spatial distribution of the source, increasing certainty.

In this dissertation we focus on the emission quantitative imaging problem using a fast-neutron coded aperture detector. We seek a reliable way to infer the total intensity of a neutron source with an unknown spatial distribution. The source intensity can be combined with other measurements to infer the holdup mass.

To do this we first create and validate a model of the imager. This model solves the forward problem of estimating data given a known source distribution. We use cross-validation to show that the model reliably predicts new measurements (with predictable residuals).

We then demonstrate a non-Bayesian approach to process new imager data. The approach solves the inverse problem of inferring source intensity, given various sources of information (imager data, physical constraints) and uncertainty (measurement noise, modeling error, absence of information, etc). Bayesian approaches are also considered, but preliminary findings indicate the need for advanced Markov chain algorithms beyond the scope of this dissertation. The non-Bayesian results reliably provide confidence intervals for medium-scale problems, as demonstrated using a

blind-inspector measurement. However, the confidence interval is quite large, due chiefly to modeling error.

CHAPTER I

Introduction

Improvements in Non-Destructive Assay (NDA) techniques improve nuclear safeguards. Safeguards inspectors use mass-in/mass-out accounting methods to detect diversion of special nuclear material (SNM) from enrichment and reprocessing facilities in non-nuclear weapon states. These methods are confounded by SNM "held up" in pipes, ducts, tanks, and glove-boxes in the facilities. Methods to improve the precision of holdup measurements will improve inspectors' ability to detect small SNM diversions.

Holdup measurement is particularly challenging at plutonium powder facilities in Japan [18, 1, 2]. Analysts currently use the Glovebox Cleanout Assistance Tool [1, 2] (BCAT) system to estimate holdup in powder handling glove boxes. In this system, moderated helium-3 detectors measure the neutron fields near glove boxes. Careful analysis and modeling with Monte Carlo N-Particle Transport Code version X [12] (MCNPX) then yield holdup estimates with 15% uncertainty—a rather unsatisfactory result, given that the facility processes 8 tonnes of plutonium annually. Because the helium-3 detectors are integrating (not imaging), the glove box "voxels" on which mass estimates are made are meters in scale. More granular source information is not known except through "historical experience and process knowledge," and the holdup may be in "locations unknown to the operator" [1]. One speculates that the uncertainty is even larger than 15% due to coarse resolution.

Similar problems arise at uranium enrichment facilities. Water ingress into UF_6 process lines creates deposits of UO_2F_2 salt. For example, UO_2F_2 was distributed along kilometers of pipes at the K-25 site near Oak Ridge, Tennessee [19, 20]. Criticality safety and material control and accountability (MC&A) motivated a multi-year effort to analyze the deposits. The analysis included dual-particle tomography with the Nuclear Weapons Identification System (NWIS) to identify holdup geometry and

composition [15]. This approach is not feasible for many safeguards applications, since tomography requires access to many sides of the object.

When enrichment facilities fall under international safeguards, inspectors use Generalized Geometry Holdup [5] (GGH), In-Situ Object Counting System [6] (ISOCS), and destructive analyses. The latter two provide information about the chemical and isotopic composition of the holdup; GGH then determines the holdup mass. Unfortunately, GGH is predicated on the inspector's knowledge of the location, shape, and extent of the deposit. This knowledge is often imperfect, and inspectors may even analyze the wrong piece of equipment [21]. The resulting estimates typically have 25% to 50% statistical error and often have relative bias above 100% [5, 22]. One question effects large uncertainty with both GGH and BCAT: Where exactly is the holdup?

In 2012, Hausladen and colleagues demonstrated a fast-neutron coded aperture (FNCA) system to locate plutonium holdup [23]. Fast neutron imaging is less sensitive to inhomogeneity and self-shielding than photon imaging and GGH, and it provides much more granular spatial information than BCAT. The strong neutron signal from spontaneous fission of plutonium-240 makes FNCA a strong candidate for holdup measurement at reprocessing facilities. A FNCA estimate of neutron source strength paired with destructive-assay composition measurements would provide a precise and robust estimate of holdup SNM.

The final piece of the FNCA system is missing: an algorithm to quantify the source strength. Existing algorithms create a *qualitative* image of the source distribution in space. Usually the image is a maximum-likelihood estimate of source strength with noise reduction to aid interpretation. Jackson studied the source quantification problem using simulated coded aperture imaging, he did not analyze uncertainty in the estimate [24]. Although qualitative algorithms can be adapted to the quantitative source-strength problem, they do not provide uncertainty estimates.

Uncertainty quantification is critical for the applications listed above. International safeguards could ideally detect the diversion of a "significant quantity" of SNM—on the order of 10 kilograms. If 40 kg $\pm$5 kg of plutonium go unaccounted for in Japan, the international community has cause to investigate. However, when 40 kg $\pm$80 kg go missing, one can draw fewer conclusions. In domestic holdup measurement, one cannot eliminate the possibility of an accident without uncertainty quantification; see e.g., [15]. In both applications, uncertainty quantification (UQ) upgrades a best-guess estimate into an actionable statement.

In this dissertation we develop UQ algorithms for the FNCA emission quantitative imaging (QI) system. We identify three key sources of uncertainty affecting the

total-strength estimate: modeling approximations, counting statistics, and solution nonuniqueness. We minimize the uncertainty to the extent possible, and quantify the remaining uncertainty. Combined with the FNCA hardware, our algorithms will form a powerful system for robust plutonium holdup estimation, e.g., in Japan.

Looking forward, one envisions the ultimate radiological analysis system: fast, accurate, and portable. The system's hardware platform would likely resemble existing technology, including emission imaging, transmission tomography, and photon spectroscopy modalities. Its software would combine real-time data from all detection modalities with analyst intuition to provide qualitative output (3-dimensional images, spectra), quantitative estimates (mass, dimensions, composition, enrichment), and measurement optimization advice (detector relocation, settings adjustments, alternative modalities). Uncertainty estimates for the quantitative outputs are complete and robust. Existing algorithms like maximum-likelihood expectectation maximization [13] (MLEM) and Monte Carlo library least-squares [10] (MCLLS) can provide portions of the desired output; a comprehensive algorithm does not exist.

This dissertation provides one missing capability for that system—source strength quantification including uncertainty analysis. We specifically target the FNCA platform because of the advantages listed above, and we target the holdup application because of the need described above. At the same time, the algorithms developed here are an important step toward a comprehensive radiological analysis algorithm.

1.1 Dissertation Overview

This dissertation next provides background relevant to the FNCA QI problem. Section 2.1 overviews of common analysis techniques for problems like holdup measurement. Section 2.2 provides an illustration of the challenges associated with holdup measurement. Sections 2.3 and 2.4 describe common radiation imagers and techniques to reconstruct qualitative images of radiation fields. Section 2.5 reviews literature and approaches most closely related to the holdup measurement problem.

The dissertation then explores each source of uncertainty individually. Chapter III details a forward model of an FNCA imager and uses cross-valdiation to predict the typical discrepancy between the model and observed data. Chapter IV discusses aleatoric uncertainty, e.g., counting statistics. Chapter V discusses epistemic uncertainty, e.g., the limited ability to resolve the source geometry.

The dissertation then applies this understanding of the problem to create comprehensive estimates using a frequentist approach (Chapter VI). The frequentist ap-

proach is demonstrated in Chapter VII using an unknown distribution of sources.

Significant and original contributions of this dissertation include modeling and validation of the FNCA imager (Chapter III); creation of a non-Bayesian approach for quantifying uncertainty in total source strength (Chapter VI); enhancements to constrained optimization algorithms to make the numerical evaluation of the non-Bayesian problem tractable (Section 6.2); and demonstration of the approach using "blind" measurements (Chatper VII). Preliminary findings of the overall work [25] and of the modeling portions [26] were published in 2016.

CHAPTER II

Background

This chapter covers background useful for understanding the structure of the dissertation. The remaining chapters will include additional background where needed.

Section 2.1 casts emission QI into the archetype of an inverse UQ problem. In this archetype we can outline the major sources of uncertainty, each of which will be discussed further in later chapters. We can also define Bayesian and frequentist analysis approaches, which will also be discussed further in later chapters. In particular, the discussion of these approaches should clarify our decision to pursue a particular frequentist approach.

In Sections 2.2–2.5 we provide an illustration of inverse UQ, describe common radiation imagers, and describe common qualitative imaging algorithms.

2.1 Uncertainty Quantification for Inverse Problems

Typical dataflow patterns for inverse UQ problems are diagrammed in Fig. 2.1. We begin by considering a set of "hidden" parameters that cannot be observed directly. In other problems the hidden parameters may be fundamental quantities such as interaction cross-sections, thermohydraulic coefficients, etc. For FNCA emission QI, the hidden parameters are neutron source strength as a function of space: $s(\vec{x})$. We are chiefly concerned with sources inside a region of interest (ROI), but must consider other sources' effect on the data. We know physically that the source strength is non-negative everywhere: $s(\vec{x}) \geq 0 \; \forall \; \vec{x}$.

An analyst then experimentally acquires data to infer information about the parameters. The data is inherently noisy, which creates "aleatoric" uncertainty. Ideally the experiment is highly sensitive to the parameters—or at least a one-to-one (injective) function of the parameters. If the data are insensitive to the parameters, the inverse problem is ill-conditioned and the aleatoric uncertainty will be amplified

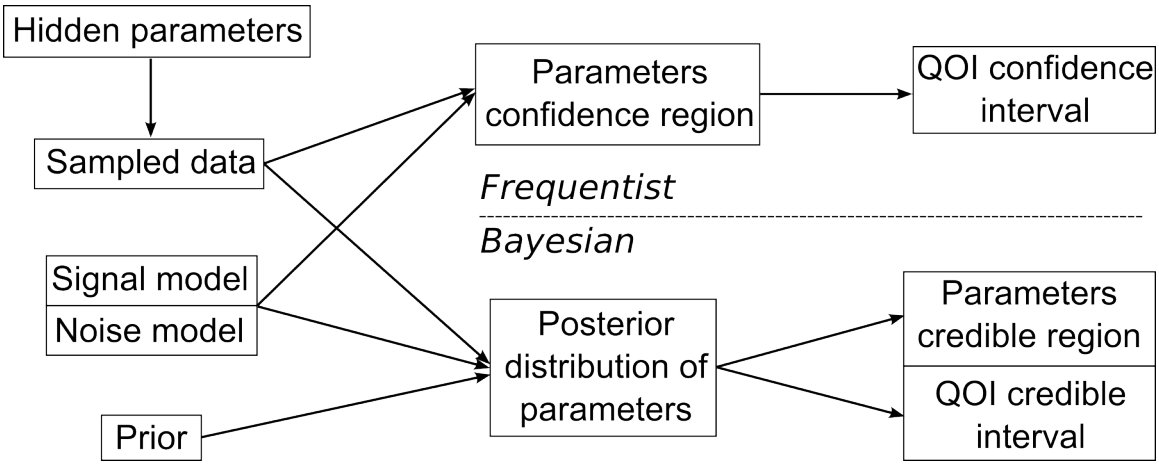


Figure 2.1: Typical dataflow for inverse UQ problems.

when estimating the parameters. If the data are not a one-to-one function of the parameters, the inverse problem is ill-posed—there are multiple parameter sets that explain the data equally well. This inability of the data to suggest a unique parameter solution is called "epistemic" uncertainty.

For FNCA QI, the data is an I-vector $\hat{d}$, where $\left[\hat{d}\right]_i$ is the number of neutron counts in pixel i. We assume photon counts and low-energy neutron counts have been removed using pulse-shape discrimination. The experiment is typically not one-to-one, because of the finite number of detector pixels and limited number of views from which the detector may observe the ROI.

The data must be compared to a forward model of the experiment to infer information about the parameters. The forward model predicts the data expectation ("signal") and distribution ("noise") as a function of the parameters. Forward models may incorporate physics simulations of varying fidelity and calibration data (i.e., "training data") of varying precision and completeness. These errors lead to "modeling" uncertainty, which we consider separately from epistemic and aleatoric uncertainty. (Some references classify modeling uncertainty as a type of epistemic uncertainty; in this dissertation it is useful to consider modeling uncertainty separately.)

Our signal model is imperfect due to numerical and model approximations. In FNCA QI, we discretize $s(\vec{x})$ using a finite number (J) of voxels or points. We solve the radiation transport problem with deterministic or Monte Carlo methods, resulting in truncation or stochastic errors. The geometry and composition of objects near the ROI may not be known exactly—and encoding these features may be prohibitively time consuming for the analyst. Scatter and attenuation by the deposit itself is almost

always neglected so that the signal model is linear. We model detector intrinsic efficiency, aperture opacity, and scatter effects by calibration.

Our noise model assumes that counts in each pixel will be independent and Poisson distributed. This neglects deadtime effects that are significant at high countrates [27]. We also occasionally approximate the Poisson distribution with a similar Gaussian distribution.

To summarize, the signal model predicts counts

$$d = As \quad , \tag{2.1}$$

where $[d]_i$ is the expected number of counts in pixel i of I, $[s]_j$ is the discretized source strength at voxel j of J, and A is the "system matrix." In radiation transport parlance, element $[A]_{i,j}$ is a Green's function from strength source voxel j to counts in detector pixel i. The noise model predicts that $\hat{d}$, the sampled data, will be distributed

$$\hat{d} \sim \mathbf{d} = \text{Poisson}(d) = \text{Gaussian}(d, \text{diag}(d)) + \mathcal{O}\left(d^{-1/2}\right) \quad , \tag{2.2}$$

where the Gaussian covariance matrix has elements

$$[\text{diag}(d)]_{i,i'} \equiv \begin{cases} [d]_i \,, & i = i' \\ 0, & \text{else} \end{cases} \quad . \tag{2.3}$$

This assumes that there is no covariance among the pixels, a common assumption.

Using the sampled data, forward model, and any other available information, our goal is to infer an interval estimate for some quantity of interest (QOI). The QOI is generally a function of the parameters, e.g., the depletion lifetime of a nuclear reactor configuration or the peak fuel centerline temperature in an accident. In our emission QI problem, the QOI is the total source within the ROI:

$$S \equiv \int\limits_{\vec{x} \in \text{ROI}} s(\vec{x}) \; dV \approx \sum_{j \in \text{ROI}} [s]_j \quad . \tag{2.4}$$

We can infer the QOI using techniques based in the Bayesian framework or frequentist (i.e., non-Bayesian) framework.

2.1.1 Bayesian Framework

Bayesian analysis views the parameters as unknown random variables. Within the Bayesian framework we can incorporate multiple sources of information into a distribution of the parameters. Typically the information sources are categorized as a "prior" distribution of the parameters ($p(s)$) and new data that will be distributed according to $p(d|s)$. Merging the prior with the new data gives a "posterior" distribution of the parameters

$$p(s|d) = \frac{p(s)p(d|s)}{p(d)} \quad . \tag{2.5}$$

(The data probabilty $p(d)$ is effectively a normalization factor.)

This posterior distribution represents our belief about what values of the parameters are reasonable. The parameters' distribution can be propagated into a credible interval for the QOI using several techniques. See Fig. 2.2.

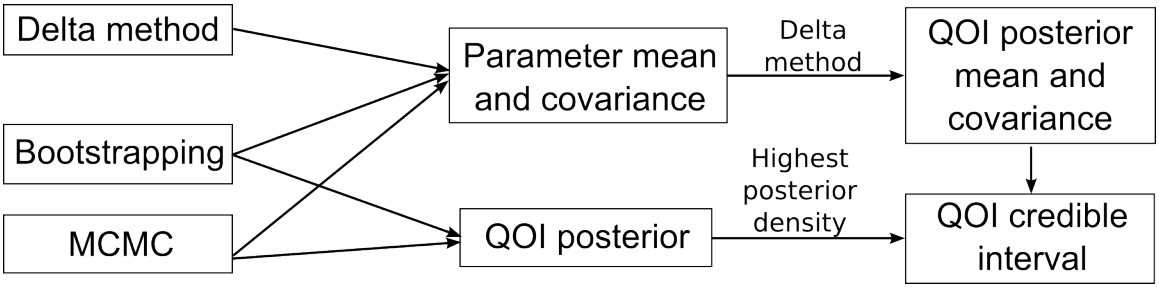


Figure 2.2: Techniques for Bayesian analysis; compare to Fig. 2.3.

An influential prior is often appropriate. For holdup QI, one expects many elements of s to be zero. A prior reflecting this expectation could take the form

$$p(s) = \prod_j \left\{ \begin{array}{cc} \zeta, & [s]_j > 0 \\ 1-\zeta, & [s]_j = 0 \\ 0, & \text{else} \end{array} \right\} \tag{2.6}$$

when we expect approximately ζJ nonzero elements in s. Note that this prior is not smooth, and it can make the posterior multimodal (i.e., it has multiple local maxima).

A Bayesian analysis would analyze the posterior of the parameters or the posterior of the QOI. A credible interval ($\mathbf{CI}_b$) for the QOI is the most common representation of the distribution. Multiple definitions exist for the credible interval. We prefer the

"highest posterior density" 95% credible interval:

$$\mathbf{CI}_b \equiv \left\{ S \;\middle|\; p\left(S \middle| \hat{d}\right) > p_{0.95} \right\} \tag{2.7}$$

with the posterior threshold $p_{0.95}$ defined such that

$$\int\limits_{S \in \mathbf{CI}_b} p\left(S \middle| \hat{d}\right) \, dS = 0.95 \quad . \tag{2.8}$$

Here we discuss notable Bayesian techniques: Markov chain Monte Carlo (MCMC), the delta method, Generalized Linear Least-Squares (GLLS), and bootstrapping.

2.1.1.1 Markov chain Monte Carlo

Markov chain Monte Carlo is a stochastic numerical technique to tally attributes (mean, variance, skew, kurtosis, ...) of a distribution. A specific MCMC algorithm, Metropolis–Hastings sampling, requires only a closed-form function that is proportional to the point density function of interest [28, 29, 30]. For this reason Metropolis sampling is popular for Bayesian posterior sampling: we can analyze the posterior

$$p(s|d) = \frac{p(s)p(d|s)}{p(d)} \propto p(s)p(d|s) \equiv \hat{p}(s|d) \tag{2.9}$$

without explicitly evaluating the data-dependent normalization $p(d)$.

To do this, the Metropolis algorithm performs a "random walk" through the solution space. Suppose at step k the origin points are summarized in a state vector s_k. A trial state s^*_{k+1} is sampled from a proposal distribution $Q(s^*_{k+1}|s_k)$ and accepted based on the Metropolis criterion:

$$s_{k+1} = \begin{cases} s^*_{k+1}, & Z_k < \frac{\hat{p}(s^*_{k+1}|d)}{\hat{p}(s_k|d)} \frac{Q(s_k|s^*_{k+1})}{Q(s^*_{k+1}|s_k)} \\ s_k, & \text{else} \end{cases} \quad , \tag{2.10}$$

where $Z_k \in [0, 1)$ is a uniform random real number. At each step quantities of interest may be tallied.

Sequential steps are correlated if $Q(s^*_{k+1}|s_k)$ is not independent of s_k or if the trial state is rejected. To maintain accuracy of the variance estimate, only every Nth step is tallied. An ideal $Q(s^*_{k+1}|s_k)$ will encourage trial steps that are large enough to quickly traverse the relevant domain but small enough to be accepted often.

Special propsal distributions are required to sample from multimodal posterior dis-

tributions. Previous work [3, 4] has demonstrated the DiffeRential Evolution Adaptive Metropolis [3, 4] (DREAM) solver for multimodal distributions. Unfortunately this technique is computationally expensive for problems with thousands of parameters. For this reason DREAM remains a candidate for future work.

An important generalization of MCMC is reversible-jump Markov chain Monte Carlo [17] (RJMCMC), which can analyze problems with variable solution-space dimensionality [17]. For FNCA posterior sampling, RJMCMC could evaluate solutions with a small variable number of point sources. Since this is particularly applicable for some QI problems, it should be considered in future work.

2.1.1.2 Delta Method

The delta method takes advantage of the fact that a linear transformation of a Gaussian distribution is also Gaussian. Suppose we have a random Gaussian process sampling $\hat{x}$

$$\hat{x} \sim \mathbf{x} = \text{Gaussian}(\text{mean}[\mathbf{x}], \text{covar}[\mathbf{x}]) \tag{2.11}$$

such that $\hat{x}$ is a realization from the distribution $\mathbf{x}$. If

$$\hat{y} \equiv f(\hat{x}) \quad , \tag{2.12}$$

for some linear operation f, then $\hat{y}$ is distributed

$$\hat{y} \sim \mathbf{y} = \text{Gaussian}(\text{mean}[\mathbf{y}], \text{covar}[\mathbf{y}]) \tag{2.13}$$

with

$$\text{mean}[\mathbf{y}] = f(\text{mean}[\mathbf{x}]) \tag{2.14}$$

and

$$[\text{covar}[\mathbf{y}]]_{i,j} = \sum_k \sum_l \left(\left. \frac{\partial [f]_i}{\partial [x]_k} \right|_{x=\text{mean}[\mathbf{x}]} \right) [\text{covar}[\mathbf{x}]]_{k,l} \left(\left. \frac{\partial [f]_j}{\partial [x]_l} \right|_{x=\text{mean}[\mathbf{x}]} \right) \quad . \tag{2.15}$$

Equation (2.15) is sometimes called the "sandwich equation." The Jacobian $\left. \frac{\partial [f]_i}{\partial [x]_k} \right|_{x=\text{mean}[\mathbf{x}]}$ is also called a sensitivity matrix—although it is entirely distinct from the detection sensitivity vector $A^\top \vec{1}$. This first-order propagation of uncertainty from $\mathbf{x}$ to $\mathbf{y}$ is

sometimes called propagation of uncertainty.

This estimation is useful as long as f is roughly linear and $\mathbf{x}$ is roughly Gaussian. A Poisson distribution Poisson(λ) approaches Gaussian($\lambda, \text{diag}(\lambda)$) as $\lambda \to \infty$. Therefore the delta method is useful in some reconstruction problems. As we show in Section 4.1, using the delta method with the maximum likelihood (ML) estimator is similar to GLLS.

2.1.1.3 Generalized Linear Least Squares Regression

Generalized linear least-squares regression takes advantage of principles similar to the delta method, but is derived within the framework of an inverse problem. This technique is general enough that it could be derived in a Bayesian or frequentist framework. Given a linear (or linearized) model, Gaussian-distributed data, and covariance estimates of the data we can estimate the parameters and the covariance associated with those estimates. This section derives the GLLS estimator and its covariance, then discusses its shortcomings for QI.

Assume $\hat{d}$ is drawn from a multivariate distribution

$$\hat{d} \sim \mathbf{d} \equiv \text{Gaussian}\left(d, C_{dd}\right) \tag{2.16}$$

with covariance matrix C_{dd} approximately constant with respect to the parameters Also assume that the model is perfect

In other words, we can calculate $d = As_{\text{true}}$ and C_{dd} exactly, if only s_{true} were provided. The response matrix A is also a sensitivity matrix in the sense that

$$[A]_{ij} = \frac{\partial [d]_i}{\partial [s]_j} \quad . \tag{2.17}$$

Recall that none of the measurements is perfectly precise, so there are some measurement residuals or discrepancies. The true (but unknown) measurement residuals are

$$\hat{r}(s_{\text{true}}) \equiv As_{\text{true}} - \hat{d} \quad . \tag{2.18}$$

Equation (2.16) implies that $\hat{r}(s_{\text{true}})$ is sampled from a multivariate Gaussian distribution:

$$\hat{r}(s_{\text{true}}) \sim \mathbf{r}(s_{\text{true}}) = \text{Gaussian}(0, C_{dd}) \quad . \tag{2.19}$$

One implication of this is that we expect $[\hat{r}(s_{\rm true})]_i^2$ to be approximately $[C_{dd}]_{ii}$.

More generally, the residuals associated with any hypothesized parameter set s are

$$\hat{r}(s) \equiv As - \hat{d} \quad \text{or} \tag{2.20}$$

$$As = \hat{d} + \hat{r}(s) \quad . \tag{2.21}$$

This is our linear model.

We could define an Ordinary Linear Least-Squares (OLLS) statistical metric for s [31]

$$\chi^2_{\rm OLLS}(s) \equiv \hat{r}(s)^\top \hat{r}(s) \quad . \tag{2.22}$$

Finding $s_{\rm OLLS} \equiv {\rm argmin}_s\, \chi^2_{\rm OLLS}(s)$ would be one way to infer an estimate of s. However, we expect that some values of $\hat{r}(s_{\rm true})$ are larger than others (based on C_{dd}), so weighting each element of $\hat{r}(s)$ equally is somewhat arbitrary.

To create better statistical metric, we first transform the system such that the residuals are serially uncorrelated with unity variance. Suppose that some $C_{dd}^{-1/2}$ exists such that

$$\left(C_{dd}^{-1/2}\right)^\top C_{dd}^{-1/2} C_{dd} = \mathbf{I}_I \quad . \tag{2.23}$$

This could be, for example, the Cholesky decomposition of C_{dd}^{-1}. Multiplying Eqs. (2.21) by $C_{dd}^{-1/2}$ gives

$$\tilde{A}s = \tilde{d} + \tilde{r}(s) \quad , \tag{2.24}$$

with $\tilde{A} \equiv C_{dd}^{-1/2} A$, etc. This achieves the desired property,

$$\tilde{r}(s_{\rm true}) \sim \text{Gaussian}\,(0, C_{\tilde{d}\tilde{d}}) \quad \text{with} \tag{2.25}$$

$$C_{\tilde{d}\tilde{d}} = \mathbf{I}_I \quad . \tag{2.26}$$

It is now appropriate to apply ordinary least squares to the transformed system.

Our generalized linear least-squares goodness-of-fit metric is

$$\chi^2_{\mathrm{GLLS}}(s) \equiv \tilde{r}(s)^\top \tilde{r}(s) \qquad = (\tilde{A}s - \tilde{d})^\top (\tilde{A}s - \tilde{d}) \tag{2.27}$$

$$... = \hat{r}(s)^\top C_{dd}^{-1} \hat{r}(s) \qquad = (As - \hat{d})^\top C_{dd}^{-1} (As - \hat{d}) \tag{2.28}$$

with maximum-likelihood estimator $s_{\mathrm{GLLS}} \equiv \mathrm{argmin}_s\, \chi^2_{\mathrm{GLLS}}(s)$. Solving for s_{GLLS} is a numeric problem

$$\tilde{A}^\top \tilde{A}\; s_{\mathrm{GLLS}} = \tilde{A}^\top \tilde{d} \tag{2.29}$$

$$A^\top C_{dd}^{-1} A\; s_{\mathrm{GLLS}} = A^\top C_{dd}^{-1} \hat{d} \tag{2.30}$$

We can also use the OLLS variance estimator to evaluate the covariance of s_{GLLS} among repeated samples of $\hat{d}$. In other words,

$$s_{\mathrm{GLLS}} \sim \mathrm{Gaussian}\left(s_{\mathrm{true}}, C_{s_{\mathrm{GLLS}} s_{\mathrm{GLLS}}}\right) \quad , \tag{2.31}$$

with $C_{s_{\mathrm{GLLS}} s_{\mathrm{GLLS}}}$ such that

$$\tilde{A}^\top \tilde{A}\; C_{s_{\mathrm{GLLS}} s_{\mathrm{GLLS}}} \equiv A^\top C_{dd}^{-1} A\; C_{s_{\mathrm{GLLS}} s_{\mathrm{GLLS}}} = \mathbf{I}_I \quad . \tag{2.32}$$

We note serious limitations of this approach for QI, namely its assumptions that s_{GLLS} follows a Gaussian distribution, that Eq. (2.29) has a unique solution, and that the prior is Gaussian.

The nonnegativity constraints provide information that that is not accounted for in $C_{s_{\mathrm{GLLS}} s_{\mathrm{GLLS}}}$. Even if s_{GLLS} is adjusted to meet the constraints, Eq. (2.32) is not affected by this adjustment. The GLLS Gaussian approximation results in wide estimates of $\mathbf{CI}_b$, since the Gaussian distribution is more broadly dispersed than a truncated Gaussian [32].

For many QI problems $I < J$, so a continuum of values of s_{GLLS} solve Eq. (2.29). Solutions may or may not exist for Eqs. (2.29) and (2.32). Regardless, the underdeterminancy of s_{GLLS} is a source of uncertainty not accounted for in $C_{s_{\mathrm{GLLS}} s_{\mathrm{GLLS}}}$; see Chapter V. Ignoring this underdeterminancy results in optimistically narrow estimates of $\mathbf{CI}_b$.

The Bayesian prior (Eq. (2.6)) is significant. Generalized Linear Least-Squares has no mechanism to incorporate a non-Gaussian prior. As Section 2.2 will show, neglecting the prior results in narrow estimates of $\mathbf{CI}_b$.

2.1.1.4 Bootstrapping

Bootstrapping, like the delta method, propagates the uncertainty through some function of the data. Unlike the delta method, the function need not be (nearly) linear and the data distribution can be arbitrary.

Bootstrapping approximates the data distribution by resampling datasets based on the measured data [33]. For example, if dataset $\hat{d}$ is measured from a Poisson distribution, the dataset can be resampled K times, with dataset k sampled

$$\tilde{d}_k \sim \text{Poisson}(\hat{d}) \quad . \tag{2.33}$$

Alternatively, the measurement can be subdivided into L identically distributed sub-measurements. Then $\tilde{d}_k$ is sampled with replacement from among the sub-measurements. This approach loses accuracy when the resampled distribution is far from the true distribution $\mathbf{d}$.

The function is applied to each d_k to estimate the distribution for the quantity of interest. For inverse problems, the function could be a reconstruction operation to estimate $s(d_k)$, followed by a QOI calculation ($S(d_k)$). This necessarily requires many reconstruction calculations.

Fleenor, Ziock, and Blackston applied bootstrapping to quantitatively image known-geometry sources [34, 35]. The geometry assumptions in that work reduce the number of unknown parameters, eliminating epistemic uncertainty. As this dissertation discusses in Chapter V, epistemic uncertainty is significant in QI holdup measurements (for which fewer geometry assumptions can be made). The bootstrapping procedure is designed to account for aleatoric uncertainty (the noisy distribution of the data), but procedures to account for epistemic uncertainty are not prominent in the literature.

2.1.2 Frequentist Framework

Frequentist techniques focus on the idea that some true source distribution s_{true} exists, and that we can use the data to generate interval estimates of S. A frequentist analysis could define a 95% confidence region in the parameter space ($\mathbf{CR}_f$). The confidence interval for the QOI ($\mathbf{CI}_f$) is then bounded by the extrema of S in the confidence region. This guarantees that the confidence interval will capture the true S in at least 95% of datasets. In practice, confidence intervals tend to be rather conservative, since the extrema of S may be uncharacteristic of the confidence region. This approach remains attractive, however, because it is fairly robust and computationally tractable for emission QI. This approach is detailed and implemented in Chapter VI.

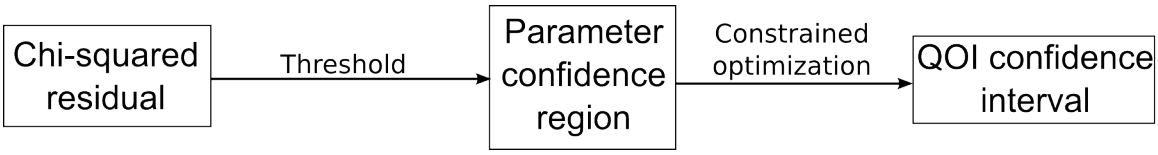


Figure 2.3: A technique for frequentist analysis; compare to Fig. 2.2.

More relaxed definitions of confidence interval exist. For example, Weerahandi proposed generalized confidence intervals in which the interval captures the true QOI in 95% of samples of both the parameters and the data [36]. (If the parameters are fixed at certain values, the generalized confidence interval may capture the QOI for less than 95% of sampled datasets.) Additional guarantees could be made as a 95%–95% confidence interval: for 95% of parameter sets, 95% of sampled datasets would result in confidence intervals that capture the true QOI. These approaches are attractive because they generate smaller intervals than basic confidence intervals, but it is difficult to speculate on the distribution of parameter sets (holdup distributions) among real holdup measurement scenarios.

2.2 Illustration

As a concrete example of the concepts introduced in Section 2.1, consider a measurement with two unknown parameters and one, two, or three datapoints.

Suppose laboratory inspectors needed to estimate the total intensity of two check sources for a safety audit. They needed to estimate the total strength of two americium–beryllium sources, numbered 0 and 1. The audit required an estimate of total-strength

$$S \equiv s_0 + s_1 \tag{2.34}$$

with a credible interval with a relative width less than 20%.

The inspectors first placed the sources on the table and set a counting (non-imaging) helium-3 detector to record counts for 3600 seconds (livetime). While acquiring data, they modeled the sources and detector using their favorite radiation transport software. The forward model indicated that the detector response is 0.01 counts per neutron emitted from source 0 and 0.005455 counts per neutron emitted from source 1. (Source 0 was closer to the detector than source 1. This is not ideal for reasons that become apparent momentarily, but it represents true holdup

measurements, in which deposits may exist at different distances from the imager.)

The inspectors then compared their data and model. Since the detector recorded for 3600 seconds, the model can be summarized as a system matrix with elements $A_{00} = 36$ count–seconds per neutron and $A_{01} = 19.6$ count–seconds per neutron. The detector measured $\hat{d}_0 = 175$ counts. Using brute-force calculations, they plotted $p(\hat{d}|s_0, s_1)$ and $p(\hat{d}|S)$, the Bayesian likelihood of acquiring $\hat{d}$ as a function of s_0 and s_1 or as a function of S. See Fig. 2.4.

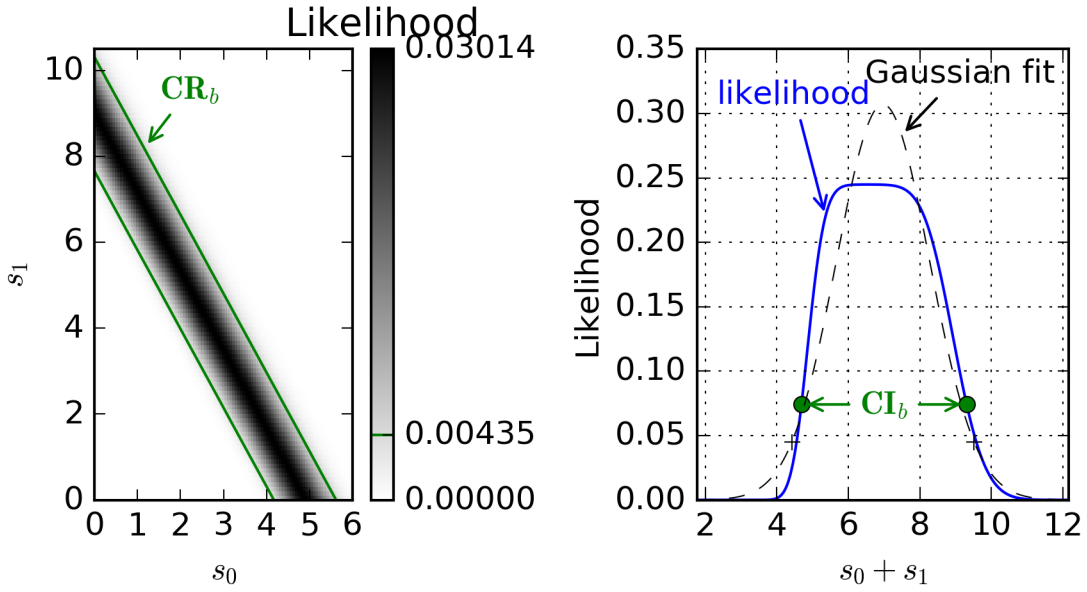


Figure 2.4: After the inspectors' first measurement, neglecting the prior: likelihood of s and its credible region (left); likelihood of S and its credible interval (right).

The resulting credible interval $(4.7, 9.3)$ was wider than the acceptable 20%. It was wide chiefly because of epistemic uncertainty—one cannot use a single datapoint to determine how many counts came from each of the two sources.

One must to determine how many counts came from each source because this affects our conversion from counts to total source intensity. A single count in the detector could be explained by $1/A_{00} \approx 0.03$ neutrons per second from source 0 or by $1/A_{01} \approx 0.05$ neutrons per second from source 1. The observed counts could be explained by a weak source 0, a strong source 1, or some combination of the two. The estimate of S was only bounded because neither source can have negative intensity.

The inspectors then noted that, based on past experience, 99.9% of the sources in the lab are dead. With this Bayesian prior ($\zeta = 0.001$), the posterior distributions of

s and S are bimodal, and the 95% credible interval is even wider. See Fig. 2.5.

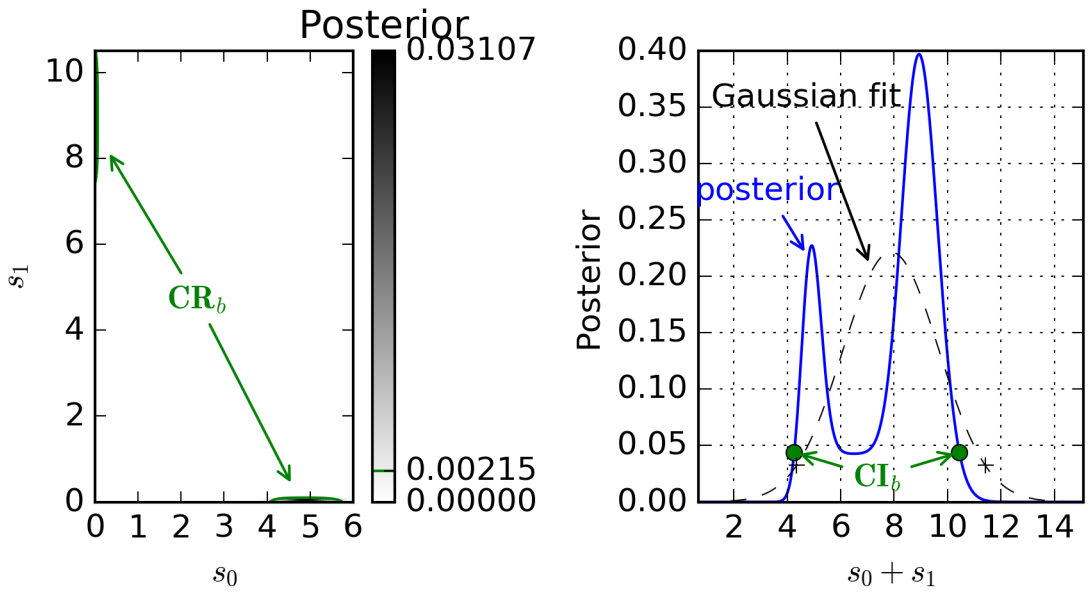


Figure 2.5: After the inspectors' first measurement: posterior of s and its credible region (left); posterior of S and its credible interval (right).

The inspectors then prepared a second measurement. This time they slightly shifted the sources, so that $A_{10} = 36.364$–count-seconds per neutron, and $A_{11} = 54.545$ count–seconds per neutron. The detector measured $\hat{d}_1 = 279$ counts.

The following morning, they updated their Bayesian posterior to include the new datapoint. See Fig. 2.6. The second measurement indicated that neither source is dead. However, because the second measurement was so similar to the first measurement, the inverse problem was well-posed but ill-conditioned. Ill-conditioning amplifies the measurement noise, so the credible interval was still too large.

The inspectors began to question their modeling work as well. They noticed that the table, floor, ceiling, and detector stand were all ignored in the original model. After adding these features, the elements of the system matrix increased 10%. In turn, this correction decreases the estimates of s_0, s_1, and S by 10%.

The inspectors now recognized that their third measurement should complement the first two measurements. In other words, the new measurement should improve the conditioning of the response matrix. To achieve this they put source 0 behind a thick neutron shield to isolate the counts from source 1. This yielded a total of $\hat{d}_2 = 187$ counts.

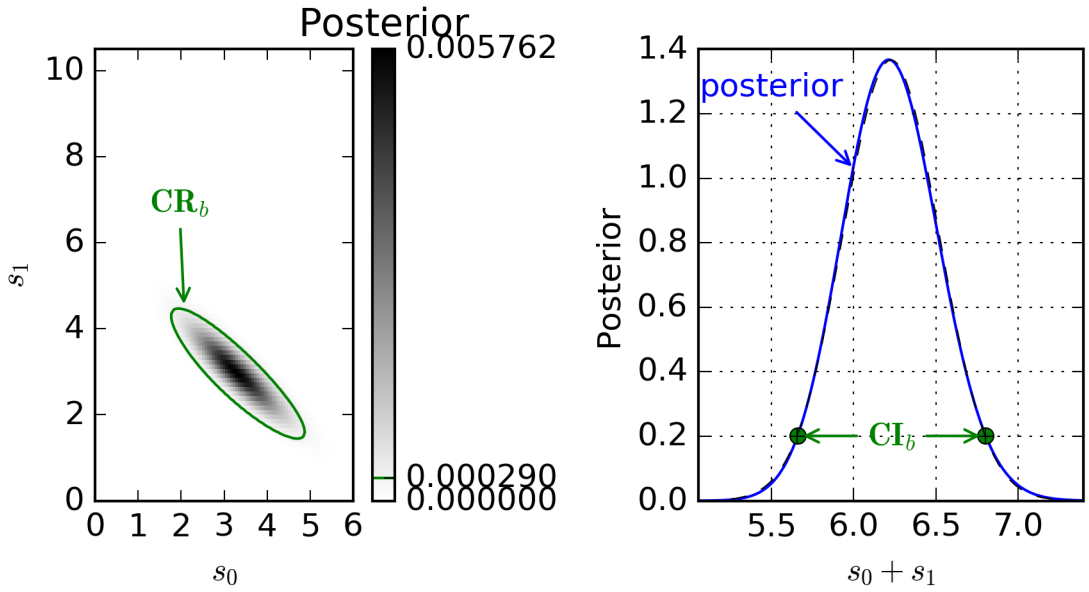


Figure 2.6: After the inspectors' second measurement: posterior of s and its credible region (left); posterior of S and its credible interval (right). The Gaussian fit overlaps.

With the refined model they re-calculated the entire system matrix:

$$A = \begin{bmatrix} 39.6 & 21.6 \\ 40 & 60 \\ 0.01 & 60 \end{bmatrix} \tag{2.35}$$

count–seconds per neutron, and summarized their data

$$\hat{d} = \begin{bmatrix} 175 \\ 279 \\ 187 \end{bmatrix} \tag{2.36}$$

counts.

The posterior calculated from this data, forward model, and prior is compactly dispersed. See Fig. 2.7. There is still significant covariance between s_0 and s_1, but the uncertainty on S is satisfactorily small. This result satisfies the audit requirements.

All of these effects are simultaneously present in typical FNCA emission QI problems. One should expect imperfect models, especially when analyst time is limited. Similarity among the datapoints will make the inverse problem ill-posed, which will amplify aleatoric uncertainty. If the source is finely discretized, one should expect

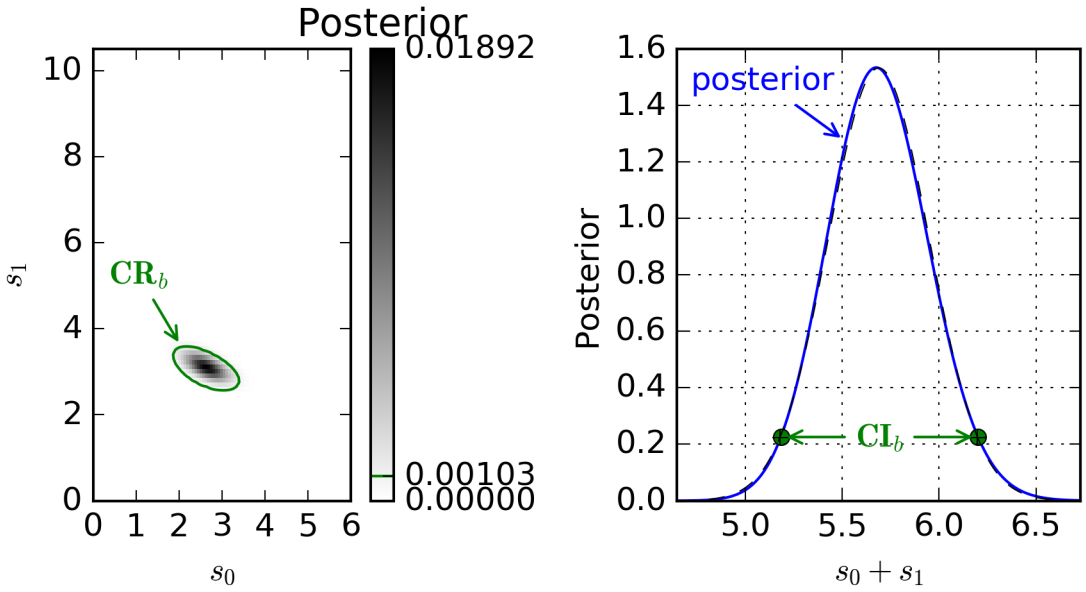


Figure 2.7: After the inspectors' third measurement: posterior of s and its credible region (left); posterior of S and its credible interval (right). The Gaussian fit overlaps.

more parameters than datapoints, leading to epistemic uncertainty. Given the time and access constraints of many inspection sites, even a diligently designed measurement will have these sources of uncertainty.

The nonnegativity constraint mitigates these effects to a limited extent. The prior belief that most elements of s are zero will further widen $\mathbf{CI}_b$. The nonnegativity constraint and the prior will each distort the posterior into a non-Gaussian distribution. This creates challenges for many popular UQ approaches like GLLS, which assume (nearly) Gaussian distributions.

From an experimental perspective, orthogonal measurements will decrease the QOI uncertainty most rapidly. This could take the form of a second measurement from a different location. Another example of orthogonal measurements is the selection of the coded aperture mask pattern—experts design patterns to make each detector pixel sensitive to different source voxels. Alternatively, the system matrix could be designed such that the QOI estimate is unaffected by the covariance among the parameters. This degree of control over the experiment is not always possible.

From a UQ perspective, the analysis should estimate modeling, aleatoric, and epistemic uncertainty. To neglect any of these sources is to express undue confidence in the estimate.

2.3 Imaging Radiation Detectors

Imaging radiation detectors collect information about the spatial or angular distribution of radiation at the detector.

For example, X-ray radiography and tomography detectors measure the spatial distribution of radiation from a small source. The data are interpreted to infer information about the materials between the source and detector. This process is called attenuation radiography or tomography, since the source is well characterized and the "unknown" parameters are attenuation factors. Associated particle imaging is an extension of attenuation tomography in which timing and angular information about source events is also collected [37, 38].

Angle-sensitive detectors—scatter cameras and coded aperture imagers—are more common in safeguards applications. Scatter cameras use the position and energy of two or more interactions to infer a region-of-response from which the particle originated. The response region for an event in a scatter camera is conical.

Coded aperture imagers use a patterned mask to encode the angular distribution of particles as a spatial distribution of detection events [39]. If we compare a small radiation source to a lightbulb, the mask casts a shadow on the detector. If the bulb moves down, the shadow moves up; up, down; left, right; etc. This shadow is apparent in a "hit pattern"—the spatial histogram of events recorded by the detector.

The mask should be designed so that the shadow is a unique function of the light's position. Other considerations like signal-to-noise ratio motivate the particular patterns employed in practice [40]. The modified uniformly redundant array [14] (MURA) pattern is ideal for some applications because its open elements are replaced with closed elements (and vice versa) when it is rotated 90°.

This property is useful because "mask–antimask" subtraction can mitigate some nuisance signals. Particles can reach the detector from angles outside the encoded portion of the mask, e.g., scatter off the floor near the imager. The resulting counts are difficult to interpret, especially using analytical reconstruction techniques. By subtracting twin mask- and antimask-measurements, the analyst can eliminate counts from these angles (in the mean). However, these particles still contribute to the measurement noise (since the variance of the counts sums when the counts are subtracted).

Angle-sensitive imagers have two key limitations when determining the spatial distribution of a source. First, particles scattered toward the imager appear to be additional sources. These scattered particles can sometimes be minimized using energy

thresholds or pulse-shape discrimination.

Second, angle-sensitive imagers provide limited information about the source–detector distance. For coded apertures, moving the radiation source toward the detector magnifies the mask shadow. The amount of information encoded in this shift is limited, especially at distances far from the detector. The imager resolution is much finer in azimuth and elevation than distance. Since distance is important for determining the source strength, we often collect multiple orthogonal views of the subject to infer the distance through parallax.

In this work we use a FNCA imager. See Section 3.1 for details on that imager.

2.4 Qualitative Reconstruction Solvers

Quantitative problems should be approached differently than qualitative problems. However, qualitative reconstruction methods can inform our approach to the quantitative problem. Here we provide a brief description of several solvers used for qualitative image reconstruction.

2.4.1 Analytical Solvers

Analytical solutions typically take advantage of the linearity of the convolution operator. Examples include filtered back-projection. These methods typically do not account for counting statistics and therefore provide limited-quality reconstructions for low-count problems.

For coded aperture reconstruction, this can be implemented by matched-shift filtering [39]. This facilitates estimation of pixel–pixel correlations in 2D reconstructions [41], which could estimate reconstruction uncertainty for strength estimates. This correlation method has not been extended to 3D reconstruction.

2.4.2 Posterior-maximizing Iterative Solvers

To account for the varying uncertainty among detector pixels, we may recast the deconvolution problem in a Bayesian framework. The Bayesian posterior combines the "likelihood" of obtaining the convoluted data from a hypothetical source with the analyst's "prior" beliefs about reasonable source distributions to form a "posterior" distribution of source configurations. The maximum a posteriori (MAP) solution is the source configuration that maximizes the posterior.

The prior in qualitative imaging is a form of regularization. Non-influential priors are common, but may yield excessive noise for low-count reconstructions. In these

cases an influential prior can be used to "smooth" the reconstructed image. When the prior is non-influential, the MAP is equivalent to ML.

We can gain an intuitive sense of the maximum likelihood solution for emission tomography by comparing it to the generalized least-squaress solution. The ML source satisfies a system similar to the generalized least-squares system with covar$[\hat{d}]$ = diag$(A\hat{s})$:

$$A^\top \text{diag}(A\hat{s})^{-1} A\hat{s} = A^\top \text{diag}(A\hat{s})^{-1}\hat{d} \quad \text{, or equivalently} \tag{2.37}$$

$$A^\top \vec{1} = A^\top \text{diag}(A\hat{s})^{-1}\hat{d} \quad . \tag{2.38}$$

Note that values of $\hat{s}$ satisfying this system typically have negative elements, which are prohibited on physical grounds [42]. The system is also not linear with respect to $\hat{s}$ and must be solved iteratively. The J-vector $A^\top \vec{1}$ is called the detector "sensitivity" vector—the expected total counts for each unit-strength source discretization basis.

The ML system is not always well-posed and well-conditioned. If we approximate diag$(A\hat{s})^{-1} \approx$ diag(As), then the reconstruction step is equivalent to solving the linear system

$$\left(A^\top \text{diag}(As)^{-1} A\right) \hat{s} = A^\top \text{diag}(As)^{-1}\hat{d} \tag{2.39}$$

for $\hat{s}$. The solution is only unique if $A^\top$diag$(As)^{-1}A$ is full rank; it cannot be full rank if $I < J$. Even if it is full rank, Poisson noise in $\hat{d}$ is amplified up to a factor

$$\kappa_{ML} \equiv \kappa\left(A^\top \text{diag}(As)^{-1} A\right) \equiv \left\|A^\top \text{diag}(As)^{-1} A\right\| \left\|(A^\top \text{diag}(As)^{-1} A)^{-1}\right\| \quad , \tag{2.40}$$

where $\|\cdot\|$ refers to the matrix Euclidian norm. This condition number characterizes the amount of reconstruction information the data contain.

Many algorithms exist to iteratively solve for the ML or MAP source, such as MLEM, OSEM, and PSCA. Here we consider the Maximum-likelihood expectation-maximization (MLEM) algorithm as a typical solver.

2.4.2.1 MLEM

The MLEM algrithm iteratively updates estimates of $\hat{s}$. This section connects the MLEM iteration step to the ML equation.

We first use the property

$$\text{diag}(A\hat{s})^{-1}A\hat{s} = \vec{1} \tag{2.41}$$

(since multiplying by $\text{diag}(\cdot)^{-1}$ is equivalent to elemental division) to simplify the left side of Eq. (2.37):

$$A^\top \vec{1} = A^\top \text{diag}(A\hat{s})^{-1}\hat{d} \quad , \tag{2.42}$$

where $\vec{1}$ is an I-vector of ones. This is a system of J equations

$$\sum_i [A]_{i,j} = \sum_i \frac{[A]_{i,j}[\hat{d}]_i}{\sum_{j'}[A]_{i,j'}[\hat{s}]_{j'}} \quad \forall j \in J \quad . \tag{2.43}$$

The MLEM algorithm iteratively solves the system of Eq. (2.43). We are searching for a solution $\hat{s}$ such that $[\hat{s}]_j \geq 0 \forall j$. The iteration prescription

$$[\hat{s}]_j^{(n+1)} = \frac{[\hat{s}]_j^{(n)}}{\sum_i [A]_{i,j}} \sum_i \frac{[A]_{i,j}[\hat{d}]_i}{\sum_{j'}[A]_{i,j'}[\hat{s}]_{j'}^{(n)}} \quad \forall j \in J \tag{2.44}$$

approaches a solution $\hat{s}$ as the iteration count n approaches ∞. At each iteration, $[\hat{s}]_j^{(n+1)} > 0$ if $[\hat{s}]_j^{(n)} > 0$ (since $[A]_{i,j} \geq 0$ and $[\hat{d}]_i \geq 0$).

The MLEM algorithm does not converge as quickly as competing algorithms, but is fairly robust and simple to code. It also allows only non-negative solutions. The solution converges roughly exponentially, as derived in the next section.

2.4.2.2 MLEM Convergence

In this section we show that MLEM iteration converges in norm as ρ_{MLEM}^n, where ρ_{MLEM} is the spectral radius of an iteration matrix. This convergence model assumes that the nonnegativity constraints are inactive. As we will demonstrate, the model breaks down when constraints are active.

Define an iteration's voxelwise error $\epsilon^{(n+1)} \equiv \hat{s}^{(n+1)} - \hat{s}$. Use a first-order Taylor expansion to approximate

$$[A(\hat{s}+\epsilon)]_i^{-1} = [A\hat{s}]_i^{-1} + \sum_j [\epsilon]_j \frac{\partial [A\hat{s}']_i^{-1}}{\partial [\hat{s}']_j}\bigg|_{\hat{s}'=\hat{s}} + O(\epsilon^2) = [A\hat{s}]_i^{-1} - \frac{[A\epsilon]_i}{[A\hat{s}]_i^2} + O(\epsilon^2) \quad , \tag{2.45}$$

so

$$\mathrm{diag}(A(\hat{s}+\epsilon))^{-1} = \mathrm{diag}(A\hat{s})^{-1} - \mathrm{diag}(A\epsilon)\mathrm{diag}(A\hat{s})^{-2} + O(\epsilon^2) \quad . \tag{2.46}$$

The MLEM step is

$$\hat{s}+\epsilon^{(n+1)} = \mathrm{diag}(\hat{s}+\epsilon^{(n)})\mathrm{diag}(A^\top\vec{1})^{-1}A^\top\mathrm{diag}(A(\hat{s}+\epsilon^{(n)}))^{-1}\hat{d} \tag{2.47}$$

$$\hat{s}+\epsilon^{(n+1)} = \mathrm{diag}(\hat{s}+\epsilon^{(n)})\mathrm{diag}(A^\top\vec{1})^{-1}A^\top\left(\mathrm{diag}(A\hat{s})^{-1} - \mathrm{diag}(A\epsilon^{(n)})\mathrm{diag}(A\hat{s})^{-2}\right)\hat{d} \tag{2.48}$$

after dropping the $O(\epsilon^2)$ terms. Substituting the re-arranged ML equation

$$\mathrm{diag}(A^\top\vec{1})^{-1}A^\top\mathrm{diag}(A\hat{s})^{-1}\hat{d} = \vec{1} \tag{2.49}$$

leaves

$$\hat{s}+\epsilon^{(n+1)} = \mathrm{diag}(\hat{s}+\epsilon^{(n)})\left(\vec{1} - \mathrm{diag}(A^\top\vec{1})^{-1}A^\top\mathrm{diag}(A\epsilon^{(n)})\mathrm{diag}(A\hat{s})^{-2}\hat{d}\right) \tag{2.50}$$

Neglecting $O(\epsilon^2)$ terms leaves

$$\hat{s}+\epsilon^{(n+1)} = \mathrm{diag}(\hat{s}+\epsilon^{(n)})\vec{1} - \mathrm{diag}(\hat{s})\mathrm{diag}(A^\top\vec{1})^{-1}A^\top\mathrm{diag}(A\epsilon^{(n)})\mathrm{diag}(A\hat{s})^{-2}\hat{d} \tag{2.51}$$

$$\hat{s}+\epsilon^{(n+1)} = \hat{s}+\epsilon^{(n)} - \mathrm{diag}(\hat{s})\mathrm{diag}(A^\top\vec{1})^{-1}A^\top\mathrm{diag}(A\hat{s})^{-2}\mathrm{diag}(\hat{d})A\epsilon^{(n)} \tag{2.52}$$

$$\epsilon^{(n+1)} = \left(\mathbf{I} - \mathrm{diag}(\hat{s})\mathrm{diag}(A^\top\vec{1})^{-1}A^\top\mathrm{diag}(A\hat{s})^{-2}\mathrm{diag}(\hat{d})A\right)\epsilon^{(n)} \tag{2.53}$$

$$\left\|\epsilon^{(n+1)}\right\| \leq \|\mathbf{I} - M\|\,\|\epsilon^{(n)}\| \quad , \tag{2.54}$$

with $M \equiv \mathrm{diag}(\hat{s})\mathrm{diag}(A^\top\vec{1})^{-1}A^\top\mathrm{diag}(A\hat{s})^{-2}\mathrm{diag}(\hat{d})A$. Hence $\epsilon^{(n)}$ decreases faster than ρ^n_{MLEM}, where the spectral radius

$$\rho_{MLEM} \equiv \rho(\mathbf{I} - M) \equiv \max_{\|u\|=1}\|(\mathbf{I} - M)\,u\| \quad . \tag{2.55}$$

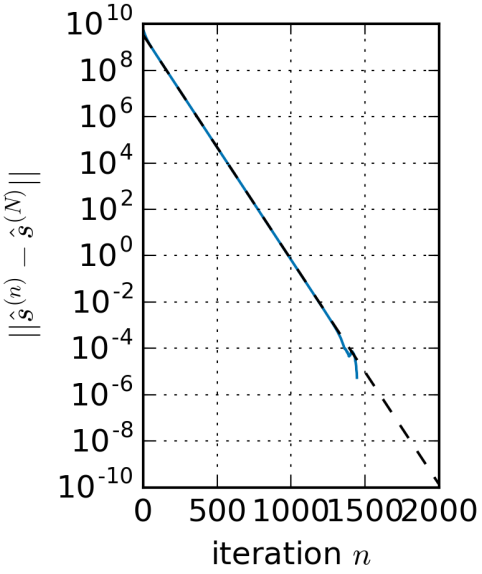


Figure 2.8: Convergence of MLEM (blue) compared to the prediction ρ_{MLEM} (black) when the constraints are inactive.

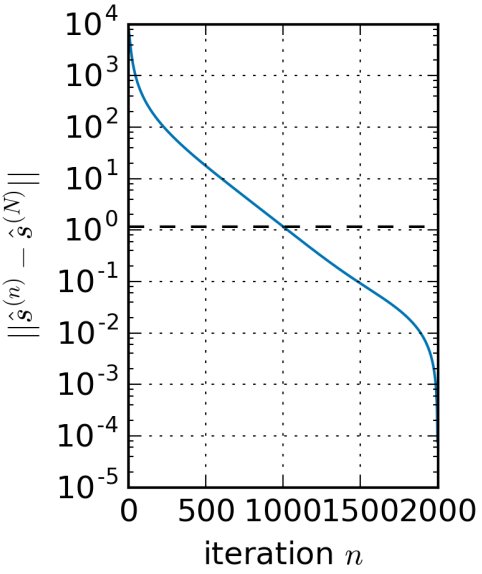


Figure 2.9: Convergence of MLEM (blue) compared to the prediction ρ_{MLEM} (black) when constraints are active.

This is similar to Hero and Fessler's finding

$$\left\|\ln \hat{s}^{(n+1)} - \ln \hat{s}\right\|_H \leq \rho(\mathbf{I} - M) \left\|\ln \hat{s}^{(n)} - \ln \hat{s}\right\|_H \quad , \tag{2.56}$$

where the logarithms are elemental and the norm is defined $\|u\|_H^2 \equiv u^\top \mathrm{diag}(A^\top \vec{1}) \mathrm{diag}(\hat{s}) u$ [43].

For well-posed problems $\alpha < 1$; see Section 2 of [43]. If M is not full rank, ϵ may be in the null space of M. This suggests that MLEM may converge on any solution of the ML equation, depending on $\hat{s}^{(0)}$. Underdetermined problems are discussed more in Chapter V.

We can demonstrate this convergence property using two problems from Section 4.2. The solution in Section 4.2.2 is far from the constraints; Section 4.2.4 is similar, but the solution is bounded by some of the constraints. The norm $\|\hat{s}^{(n)} - \hat{s}^{(N)}\|$ (where N is the final iteration) is plotted in Figs. 2.8 and 2.9. The spectral radius ρ_{MLEM} is predictive only for the inactive-constraint problem.

This observation about MLEM convergence suggests that MLEM may converge faster in some parts of the solution space than others. The convergence is faster when

the reconstruction problem is well-posed.

2.4.3 Origin Ensembles

Origin ensemble methods use an alternative source representation to solve the ML equation [44]. Rather than using basis functions, the origin ensemble represents the source using a set of origin points corresponding to the origin of each detection pulse. The stochastic origin ensembles (SOE) solver stochastically relocates these origin points to sample from the posterior. The iterative adjustment process is one example of a Markov chain random walk. Refer to Section 2.1.1.1 on Markov chain Monte Carlo.

2.5 Quantitative Imaging and Uncertainty Quantification

Several projects in the literature demonstrate different types of QI. These projects vary in their assumptions about the source and degree of UQ sophistication. We can broadly categorize the literature based on whether the source is assumed to be a small number of point sources or a more general distribution. We will begin with the point-source literature.

Hull and colleagues at PHDs locate and quantify source intensity [45, 46]. Their approach acknowledges that the imager data has limited ability to infer the source–detector distance, so the user is required to measure the distance separately [47]. Uncertainty of the source intensity is not estimated.

The SOE reconstruction method is a Markov chain process, so it can be naturally adapted into MCMC UQ. Goodman and He have demonstrated this approach for estimating uncertainty of the position of a localized source [48]. Source–detector distance and source intensity are not considered.

Generalizing somewhat from point-sources, Ziock and Blackston have used imaging data quantify the count and arrangement of block-shaped sources [35]. They estimated uncertainty on the blocks' positions using bootstrapping. Since this work is intended for arms-control, it does not address source–detector distance and source intensity.

Other authors have investigated non-imaging spectroscopy for inferring a few source and shielding parameters. Bledsoe and colleagues solve this problem using non-linear minimization of a goodness-of-fit metric [49, 50], and Mattingly and Mitchell propose similar [51]. Uncertainty is quantified using GLLS or MCMC. Streicher and colleagues demonstrated similar inference using bootstrapping [52].

Hausladen and colleagues inferred the source strength using FNCA with holdup-like geometries [23]. The work includes uncertainty estimates on the total source intensity. Their technique used a single view of the data, so the data contained limited information about the source–detector distance. To infer the distance, they calculated reconstructions at multiple distances and selected the distance with the reconstruction that best fit the data. The implicit assumption in this approach is that all of the sources are at the same distance; this assumption should be avoided for holdup quantification.

Four examples exist of QI or holdup measurement that do not assume the source is localized.

First, GGH (implemented by Holdup Measurement System-4 [5] (HMS4)) has been used in many holdup measurement applications, including demolition of the K-25 uranium processing facility [20]. The GGH approach assumes that the holdup is uniformly distributed as a point, line segment, or flat area [5]. The model is otherwise quite sophisticated, accounting for energy-dependent detector efficiency, attenuation, and self-shielding. However, the model requires the inspector to infer the holdup location and extent (linear or areal). The model then predicts count rates of a non-imaging detector as a function of the source mass. This inverse problem is well-posed because of the many modeling assumptions and approximations.

Uncertainty estimates using GGH are very limited. The HMS4 software automatically estimates uncertainty from counting statistics. It also allows the inspector to note uncertainty in the source extent, which it propagates to the mass estimate. Unfortunately, it is difficult to reliably estimate uncertainty in position, shape and extent using non-imaging survey techniques. In some cases inspectors even mis-identify which piece of equipment contains holdup [21]. Estimated uncertainty can be on the order of 25% to 50% [5], but measurement bias above 100% is common [22].

Second, dual-particle transmission tomography systems have been used to measure large holdup deposits [15, 53]. These measurements were analyzed to infer the holdup mass and composition with uncertainty, but the UQ procedures are not specified. One speculates that UQ based on propagating counting statistics is sufficient for this imaging modality, since modeling and epistemic uncertainties are minimal. Unfortunately transmission tomography systems require time and access that is impractical in safeguards inspection scenarios.

Third, Jackson performed source intensity quantification using FNCA data. Jackson performed MLEM reconstruction of a 2-dimensional (source–detector distance and one angular dimension) source distribution. The reconstruction allows for con-

tinuous source distributions, but the examples include only point sources. The source–detector distance is inferred using only the imager data, and Jackston notes considerably improved estimates using parallax measurements. Jackson does not quantify uncertainty on the intensity estimate.

Fourth, holdup measurements published by Nakamura and colleagues [1, 2] are similar to the measurements and goals of this dissertation. Nakamura performs emission tomography across a complex facility using a series of helium-3 neutron measurements. Although the detector is non-imaging, the measurements are spatially sensitive because the detector is repositioned. In this sense the holdup measurements are an example of QI and are mathematically similar to the FNCA measurements we analyze in this dissertation.

Nakamura estimates holdup mass uncertainty using GLLS [1, 54]. An MCNPX forward model calculates a 57×53 response matrix. One speculates that the selection of $I = 57$ datapoints to infer source strength in $J = 53$ voxels was chosen deliberately so that the GLLS problem is well-posed. This is an example of geometric assumptions concealing epistemic uncertainty: The data can only determine the source distribution with resolution on the scale of meters, but distribution changes of tens of centimeters would significantly affect the mass estimate. One therefore speculates that Nakamura's uncertainty estimate of $\sim 15\%$ is optimistic.

The literature survey did not turn up references on UQ for underdetermined QI problems. Emission tomography systems that are practical for safeguards holdup measurements have limited resolution, especially in the distance dimension. Uncertainty in the total source intensity—a product of modeling error, counting statistics, and geometry uncertainty—is not observed in the cited literature.

CHAPTER III

Forward Model

Accurate and expedient forward modeling is crucial for solving the inverse problem. The forward model is our mapping from the parameter space to the data space, so its errors are transferred to the inverse solution. The inverse solver usually evaluates the forward model serially, so the computation time per evaluation drastically increases the overall time-to-solution. The analyst may also have limited time to encode details of the model into an input file. These competing considerations force the developer and analyst to compromise forward-model accuracy to reduce the analyst burden and provide a timely solution.

Our first objective is a forward model that quickly and accurately predicts detector data. In this section we describe a forward model based on ray tracing. Parameters for the model are determined using calibration measurements. The model accuracy is evaluated using cross-validation.

Our second objective is a noise model that predicts the distribution of prediction–measurement residuals. Because of model errors and measurement statistics, the predictions cannot perfectly match measured data. The error model is important because it allows us to statistically compare measured data to hypothetical source distributions.

Phenomena that affect the FNCA emission QI problem are listed in Table 3.1. Several phenomena are "fundamental" to the coded aperture technique. These will be calculated using ray-tracing radiation transport.

Phenomena related to detector response are typically difficult to model. However, since these phenomena depend on the detector (not the particular measurement), the forward model can account for them using calibration. We make two exceptions: Pulse-shape discrimination techniques are sufficiently advanced that mis-classification of fast neutrons is negligible. We do not expect significant pulse pileup because of the low count rates in our measurements.

Table 3.1: Phenomena relevant to FNCA forward modeling.

category	phenomenon	approach
fundamental	tracklength through mask	ray trace
	detector pixel orientation	ray trace
	source–pixel distance	ray trace
near-detector	finite mask and scintillator thickness	ray trace
	attenuation in scintillator	ray trace
	scatter (mask and detector)	calibrate
	pixel mis-identification	calibrate
	intrinsic efficiency	calibrate
	mask penetration	calibrate
	energy-dependent effects	calibrate
	photon/neutron mis-classification	neglect
	pulse pileup	neglect
near-ROI	attenuation	evaluate
	scatter	evaluate
	self-attenuation and -scatter	evaluate
	nuisance sources	avoid
ambient	sources and scatter	nuisance parameters
	attenuation	irrelevant
numerical	detector discretization (sub-pixels)	evaluate
	source discretization (voxels)	evaluate

Phenomena near the ROI generally cannot be calibrated, but are also difficult to model. For example, neutron attenuation and scatter in glove box walls significantly affects the number of fast neutrons reaching the imager. To model attenuation, the analyst must encode the ROI geometry and materials for the ray tracer. To model scatter, ray tracing is insufficient; full-fidelity radiation transport must be used, which increases the computational burden. Self-attenuation and self-scatter further increase the burden, since they make the forward model non-linear; this precludes the usual approach of tabulating a system matrix. We will evaluate the effect of neglecting near-ROI phenomena to estimate their impact.

For good analysis, one must eliminate "nuisance" neutron sources that are near the ROI but excluded from the QOI. The imager has finite resolution, and cannot distinguish a source that is slightly inside the ROI from one slightly outside the ROI.

However, we anticipate ambient sources away from the ROI. The distribution of these ambient sources is a "nuisance parameter" that we must account for when interpreting the data. Therefore the forward model should include a smattering of point sources (in addition to the voxelized source in the ROI). The strength of each of these sources is an unknown parameter that is excluded when calculating the QOI. Ambient scatter is indistinguishable from a ambient sources, especially when pulse-shape discrimination eliminates non-fast neutrons from the hit pattern. Ambient attenuation only removes nuisance neutrons, so it is ignored.

Finally, we must consider numerical error. Discretizing the detector pixels and source is far more significant than other sources (e.g., roundoff). If the discretization is sufficiently fine, the numerical error will be trivial compared to the other approximations. We evaluate the pixel discretization error in this chapter and evaluate the effect of source discretization in later chapters.

In this section we will define and calibrate a model to predict FNCA data. We will also infer a distribution of the discrepancies between the model and measured data (which typically exceed counting statistics). We demonstrate the predictive power of the model using cross-validation, we show the need to model near-ROI materials, and we determine adequate discretization for the source and scintillator. A subset of these results were published by Bevill and Martin in late 2016 [26].

3.1 Equipment

We will model the P24 FNCA [16]. Hausladen and colleagues demonstrated this system for locating neutron sources hidden in mock holdup scenarios [55]. For calibra-

tion we will measure a small bare neutron source in multiple locations. See Fig. 3.1. This section describes that equipment and the pre-processing logic used to generate a hit pattern.

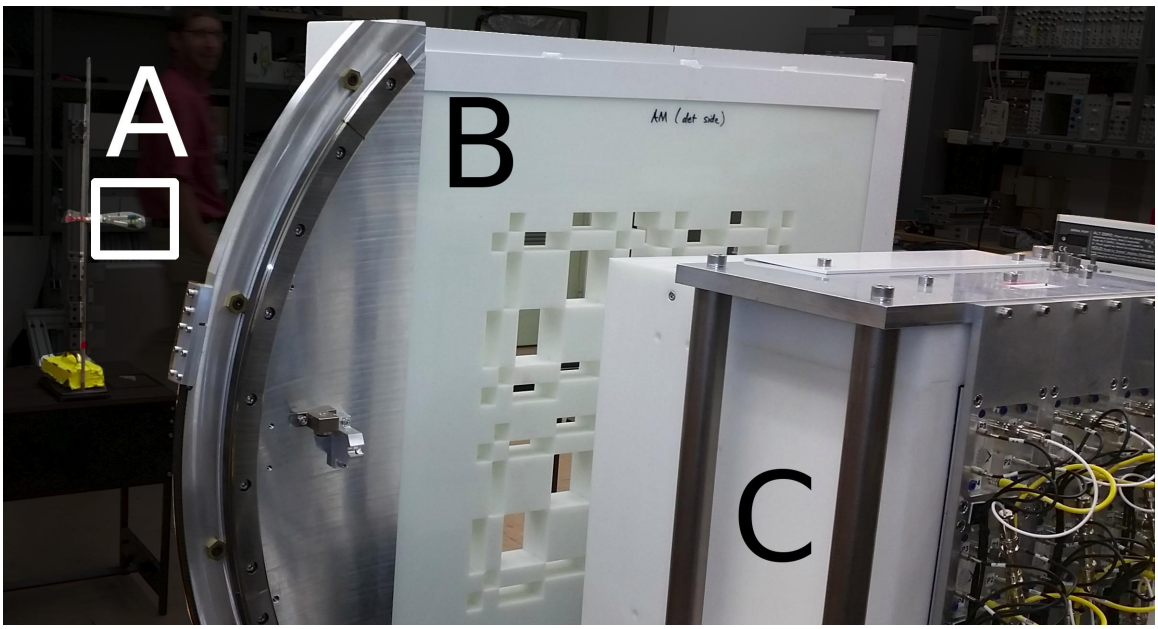


Figure 3.1: Typical setup of the P24 imager, including a neutron source (A), mask (in the anti-mask orientation, B), and scintillator blocks (C).

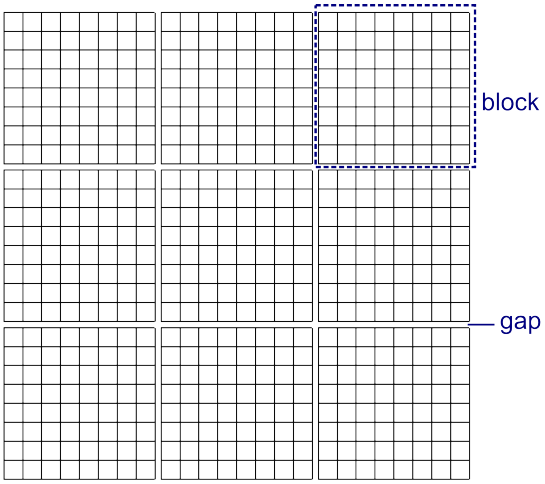
The P24 imager has $24 \times 24 = 576$ scintillator pixels. The sensitive region is 9 blocks of EJ-299-34 scintillator plastic, each subdivided into 64 optically isolated channels. See Fig. 3.2. Light from each block is collected in four photomultiplier tubes. Onboard electronics assign each event to a pixel using Anger logic [56]. Each pixel measures $1.35\,\text{cm} \times 1.35\,\text{cm} \times 5\,\text{cm}$, so each block measures $10.8\,\text{cm} \times 10.8\,\text{cm} \times 5\,\text{cm}$. Gaps of $0.4\,\text{cm}$ separates the blocks. The scintillators are centered $116.5\,\text{cm}$ above the floor.

The electronics report a digitized waveform for every recorded pulse. Photon events are eliminated using pulse-shape discrimination [57]. Low-energy neutron events are eliminated by setting a minimum waveform amplitude.

For measurements in this work, the P24 was outfitted with a polyethylene mask 5.08 cm thick. The mask is cut with a tiled base-11 MURA pattern. See Fig. 3.3. The pattern is $50.8\,\text{cm} \times 50.8\,\text{cm}$; including the unpatterned border, the mask is $81.28\,\text{cm} \times 81.28\,\text{cm}$. The mask center is 30.54 cm from the scintillator center (adjustable).

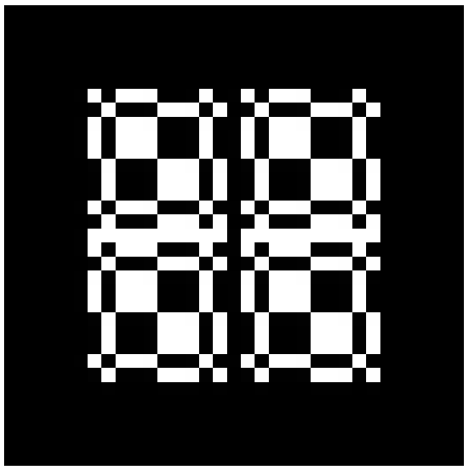
The neutron source is a californium-252 sample with identification number Cf-252-5557. The source is certified as $9.96 \times 10^{-4}\,\text{Ci}$ as of 2007-03-13, so its activity as

A single block of 8×8 pixels [16].

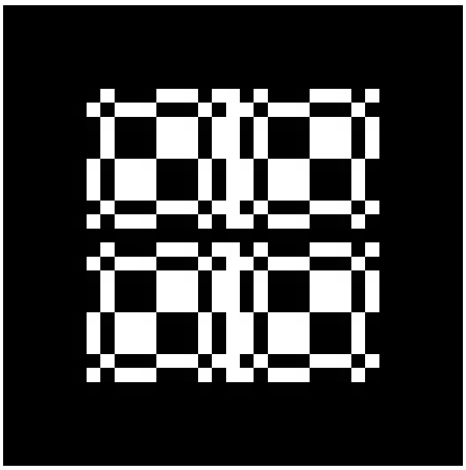


The 3×3 arrangement of blocks.

Figure 3.2: The scintillator geometry of the P24 imager.



mask orientation



antimask orientation

Figure 3.3: The P24 imager's mask pattern used in this work. Open elements are white.

of 2016-08-03 is

$$9.96 \times 10^{-4}\,\mathrm{Ci} \times \left(\frac{1}{2}\right)^{9.3938\,\mathrm{a}\ /\ 2.645\,\mathrm{a}} = 8.50 \times 10^{-4}\,\mathrm{Ci} \tag{3.1}$$

and its neutron emission rate is

$$\left(8.50 \times 10^{-4}\,\mathrm{Ci}\right)\left(\frac{3.7 \times 10^{10}\,\mathrm{decays/s}}{1\,\mathrm{Ci}}\right)(0.030\,92\,\mathrm{fiss\,/decay})\,(3.7509\,\mathrm{n/fiss}\,) = ... \tag{3.2}$$

$... = 3.65 \times 10^5$ neutrons/s. This source is positioned in the imager field of view for calibration measurements.

3.2 Model Specification

In this Section we define a ray-tracing model to predict counts measured by the (P24) imager. We also propose several "covariance models" that predict the covariance of prediction–measurement residuals. The first covariance model assumes that the data are Poisson-distributed with no covariance. The second covariance model assumes that modeling error creates relative uncertainty and covariance among the pixels. The third covariance model makes similar assumptions but uses the sampled data, not the predicted data, to predict covariance.

The forward model and the latter two covariance models require calibration, which will be described in Section 3.3.

3.2.1 Forward Model

The forward model solver uses ray-tracing to calculate fundamental detector effects. Consider the calculation of the expected counts in pixel i from a point source at x_j.

The solver first selects a set of discrete interaction sites in pixel i. Each site x_k ($k \in K_i$) represents a 3D sub-pixel of pixel i. Each pixel has the same quadrature size $|K_i|$. In this work K_i is a uniform Cartesian grid of $\sqrt[3]{|K_i|} \times \sqrt[3]{|K_i|} \times \sqrt[3]{|K_i|}$ points, so $|K_i|$ must be a perfect cube. The sub-pixels are a spatial quadrature for pixel i.

The solver then traces line segment $\overline{x_j x_k}$ from x_j to each x_k. If $\overline{x_j x_k}$ intersects a closed mask element, then the expected count rate in sub-pixel k decreases by $\exp(-\Sigma_m t_m)$. The coefficient Σ_m is an "effective" cross section determined through calibration; the closed-element tracklength t_m is calculated using structured-mesh ray

tracing. Interaction likelihood depends on the source spectrum, so the Σ_m depends on the source isotope. For the P24 imager with the californium-252 source, a typical $\Sigma_m \approx 0.15\,/\text{cm}$.

The model also accounts for angular effects and attenuation in the scintillator. Neglecting the mask, the probability an isotropically emitted neutron will interact in the subpixel around x_k is approximately proportional to

$$\exp(-\Sigma_s t_s)(1-\exp(-\Sigma_s t_k))\frac{a_k}{4\pi|x_k-x_j|^2} \quad , \tag{3.3}$$

where Σ_s is the effective cross section of the scintillator, t_s is the neutron tracklength to reach the subpixel, t_k is the neutron tracklength through the subpixel, and a_k is the subpixel area perpendicular to the ray. The first term is the probability of penetrating to reach the subpixel. The second term is the probability of interacting in the subpixel. The third term is the arcangle fraction the subpixel subtends. For the P24 imager, a typical $\Sigma_m \approx 0.19\,/\text{cm}$.

It is difficult to rapidly calculate a_k and t_k unless the ray travels perpendicular to a subpixel face. However, if the subpixels are small then we can reasonably approximate

$$(1-\exp(-\Sigma_s t_k))a_k = \Sigma_s v_k + \mathcal{O}\left(t_k\right) \quad , \tag{3.4}$$

where the subpixel volume $v_k \equiv a_k t_k$ is trivial to calculate. Therefore the subpixel interaction likelihood is approximately proportional to

$$\frac{\Sigma_s v_k \exp(-\Sigma_s t_s)}{4\pi|x_k-x_j|^2} \quad . \tag{3.5}$$

Light collection efficiency varies across the face of the photomultiplier tubes and varies as a function of the source spectrum. Therefore pixel i has an effective intrinsic efficiency $[\varepsilon]_i \in [0,1]$, which should be determined by calibration. This calibrated parameter subsumes the constant Σ_s outside the exponential of Eq. (3.5).

Although the scintillator pixels are optically separated, the Anger logic can misidentify the interaction pixel. Indeed, the scintillator pixels represent a lower bound on the detector's spatial resolution; the resolution may be much poorer due to misplacement. We model this effect by blurring the expectation with a five-point stencil. If the unblurred expectation is $[A_u]_{i,j}$, then the blurred response is

$$[A]_{i,j} = (1-4b)\,[A_u]_{i,j} + b\sum_{i'\in N_i}[A_u]_{i',j} \quad , \tag{3.6}$$

with blur factor $b \in [0, 1]$ and N_i indexing the four immediate neighbors of pixel i. If i is at the edge of the scintillator block, $[A_u]_{i',j} = 0$ for the "absent" neighbor (since misplacement across a block boundary is unlikely).

The expectation is proportional to the measurement livetime. Preliminary studies indicate that the P24 deadtime is not significant for the source strengths available in our lab. We therefore approximate the livetime with the walltime.

The walltime is generally identical for all pixels in a single measurement. However, it is convenient to aggregate data from a series of measurements by stacking the response matrix. In this case, the index i refers to a specific pixel in a specific measurement. For example, the top-left pixel may be indexed $i = 0$ in one measurement, $i = 576$ in a second measurement, $i = 1152$ in a third measurement, etc. Because the livetime varies among measurements, we specify the livetime for pixel i as T_i.

Finally, we choose to model the difference between a mask–antimask pair of measurements instead of modeling each separately. The pair difference is much less sensitive to sources and scatter outside of the fully coded field of view. (Refer to the mask–antimask explanation in Section 2.3.) Using the pair difference reduces calibration error. Denote the mask response matrix A_0 and the antimask response matrix A_1. If multiple mask–antimask pairs are measured, each new pair adds rows to A_0 and A_1 as described in the previous paragraph.

In summary, our model predicts that in the normal mask orientation, neglecting blur, pixel i will record an average of

$$[A_{0,u}]_{i,j} = [\varepsilon]_i T_i \sum_{k \in K_i} \left(\frac{v_k \exp\left(-\Sigma_m t_m - \Sigma_s t_s\right)}{4\pi |x_k - x_j|^2} \right) \tag{3.7}$$

from a unit-strength point source at x_j. (The calculation is identical in the antimask configuration, except that t_m has changed.) The expectation is then blurred according to Eq. (3.6). Response matrices for the mask (A_0) and antimask (A_1) configurations are intended to be used together to predict the antimask-subtracted hit pattern. To use the model, the analyst must specify the source distribution or discretization; the mask size, thickness, position, orientation, and pattern; and the detector position, orientation, thickness, and pixel geometry. Calibration parameters ε, Σ_m, Σ_s, and b will be experimentally determined in Section 3.3.

3.2.2 Covariance Model I

The covariance models predict the typical distribution of prediction–measurement residuals. Covariance model I assumes that the model is perfect, and the detector counts are Poisson distributed. We use the usual Gaussian approximation: Given a vector s of point-source strengths, the counts would be distributed with mean

$$\text{mean}[\mathbf{d}] = As \tag{3.8}$$

and covariance

$$\text{covar}[\mathbf{d}] = \text{diag}(As) \quad . \tag{3.9}$$

We define prediction–measurement residuals

$$\hat{r}_0 \equiv A_0 s - \hat{d}_0 \quad \text{and} \tag{3.10}$$

$$\hat{r}_1 \equiv A_1 s - \hat{d}_1 \quad . \tag{3.11}$$

Since we are using mask–antimask subtracted data, the residuals of interest are

$$\hat{r} \equiv (A_0 s - A_1 s) - \left(\hat{d}_0 - \hat{d}_1\right) = r_0 - r_1 \quad . \tag{3.12}$$

Accounting for only counting statistics, $\hat{r}$ will be distributed with mean

$$\text{mean}[\mathbf{r}] = 0 \tag{3.13}$$

and covariance

$$\text{covar}[\mathbf{r}] = \text{diag}(A_0 s + A_1 s) \quad . \tag{3.14}$$

Unfortunately this model typically underestimates the counting-statistics uncertainty. Scatter around the mask makes the count rates higher than the model prediction. These counts are typically cancelled in the mean (Eq. (3.13)) but still contribute to the variance. To account for this additional noise we approximate

$$\text{covar}[\mathbf{r}] \approx \text{diag}(\hat{d}_0 + \hat{d}_1) \quad . \tag{3.15}$$

Although this variance model can bias quantitative estimates, the effect appears to be reasonably small in pratice.

This approximation also assumes that the model error is insignificant compared to the data noise. Later results will show that the model error is signficant. This motivates us to create a second, calibrated covariance model that captures the uncertainty introduced by model error.

3.2.3 Covariance Model II

Here we add three calibrated terms to the noise model of Eq. (3.15): relative uncertainty, same-measurement correlation, and mask–antimask correlation. These terms effectively enhance the covariance matrix to include uncertainty from modeling error, calibration uncertainty, and aleatoric uncertainty (discussed in Chapter IV).

The relative uncertainty term increases the variance in the pixel-i mask residual to

$$[\text{covar}[\mathbf{r_0}]]_{i,i} = \left[\hat{d}_0\right]_i + \rho\,[A_0 s]_i^2 \quad , \tag{3.16}$$

with $\rho \in [0, 1]$ determined by calibration. Similarly for the antimask,

$$[\text{covar}[\mathbf{r_1}]]_{i,i} = \left[\hat{d}_1\right]_i + \rho\,[A_1 s]_i^2 \quad , \tag{3.17}$$

with the same value of ρ.

We also observe some correlation among the pixels within a mask or antimask measurement. This could be explained by gain shifts in the electronics, for example. We estimate the covariance between pixels i and i' in a mask measurement

$$[\text{covar}[\mathbf{r_0}]]_{i,i'} = \rho_{00}\sqrt{[\text{covar}[\mathbf{r_0}]]_{i,i}\,[\text{covar}[\mathbf{r_0}]]_{i',i'}}\ \forall\ i \neq i' \quad , \tag{3.18}$$

with $\rho_{00} \in [0, 1)$ determined by calibration. The antimask correlation takes the same form using the same coefficient.

We also allow for some correlation between same-pixel measurements of a mask–antimask measurement pair. We denote the covariance between pixel i of the mask measurement and pixel i' of the antimask measurement as $[\text{covar}[\mathbf{r_0}, \mathbf{r_1}]]_{i,i'}$. Similar

to Eq. (3.18),

$$[\text{covar}[\mathbf{r_0}, \mathbf{r_1}]]_{i,i'} = \begin{cases} \rho_{01}\sqrt{[\text{covar}[\mathbf{r_0}]]_{i,i}\,[\text{covar}[\mathbf{r_1}]]_{i,i}} & i = i' \\ 0 & \text{else} \end{cases}, \tag{3.19}$$

with mask–antimask correlation constant $\rho_{01} \in [0, 1)$ determined by calibration.

Propagating the uncertainty gives the covariance of the mask–antimask difference

$$[\text{covar}[\mathbf{r}]]_{i,i'} = [\text{covar}[\mathbf{r_0}]]_{i,i'} + [\text{covar}[\mathbf{r_1}]]_{i,i'} - [\text{covar}[\mathbf{r_0}, \mathbf{r_1}]]_{i,i'} - [\text{covar}[\mathbf{r_0}, \mathbf{r_1}]]_{i',i} \tag{3.20}$$

$$... = \begin{cases} \left[\hat{d}_0\right]_i + r\,[A_0 s]_i^2 + \left[\hat{d}_1\right]_i + r\,[A_1 s]_i^2 + ... \\ \quad ... - 2\rho_{01}\sqrt{[\text{covar}[\mathbf{r_0}]]_{i,i}\,[\text{covar}[\mathbf{r_1}]]_{i,i}} & i = i' \\ \rho_{00}\sqrt{\left(\left[\hat{d}_0\right]_i + r\,[A_0 s]_i^2\right)\left(\left[\hat{d}_0\right]_{i'} + r\,[A_0 s]_{i'}^2\right)} + ... & \\ \quad ... + \rho_{00}\sqrt{\left(\left[\hat{d}_1\right]_i + r\,[A_1 s]_i^2\right)\left(\left[\hat{d}_1\right]_{i'} + r\,[A_1 s]_{i'}^2\right)} & i \neq i' \end{cases}. \tag{3.21}$$

3.2.4 Covariance Model III

We also propose a third covariance model that is similar to covariance model II. However, this model uses only the observed counts, not the predicted counts:

$$[\text{covar}[\mathbf{r}]]_{i,i'} = \begin{cases} \left[\hat{d}_0\right]_i + r\left[\hat{d}_0\right]_i^2 + \left[\hat{d}_1\right]_i + r\left[\hat{d}_1\right]_i^2 + ... \\ \quad ... - 2\rho_{01}\sqrt{[\text{covar}[\mathbf{r_0}]]_{i,i}\,[\text{covar}[\mathbf{r_1}]]_{i,i}} & i = i' \\ \rho_{00}\sqrt{\left(\left[\hat{d}_0\right]_i + r\left[\hat{d}_0\right]_i^2\right)\left(\left[\hat{d}_0\right]_{i'} + r\left[\hat{d}_0\right]_{i'}^2\right)} + ... & \\ \quad ... + \rho_{00}\sqrt{\left(\left[\hat{d}_1\right]_i + r\left[\hat{d}_1\right]_i^2\right)\left(\left[\hat{d}_1\right]_{i'} + r\left[\hat{d}_1\right]_{i'}^2\right)} & i \neq i' \end{cases}. \tag{3.22}$$

This difference means that the noise is constant with respect to s.

We determine parameters for the forward model and variance models II and III using calibration.

3.3 Calibration

We calibrate the model by taking a series of measurements and calculating the best-fit calibration parameters.

3.3.1 Measurements

We took a series of $P = 15$ measurement pairs. A measurement pair consists of twin mask and antimask measurements with identical walltime, detector position, etc. For all measurements, the mask is centered at $(190.5, 0, 116.5)$ cm. The imager faces the $-x$ direction. See Fig. 3.4.

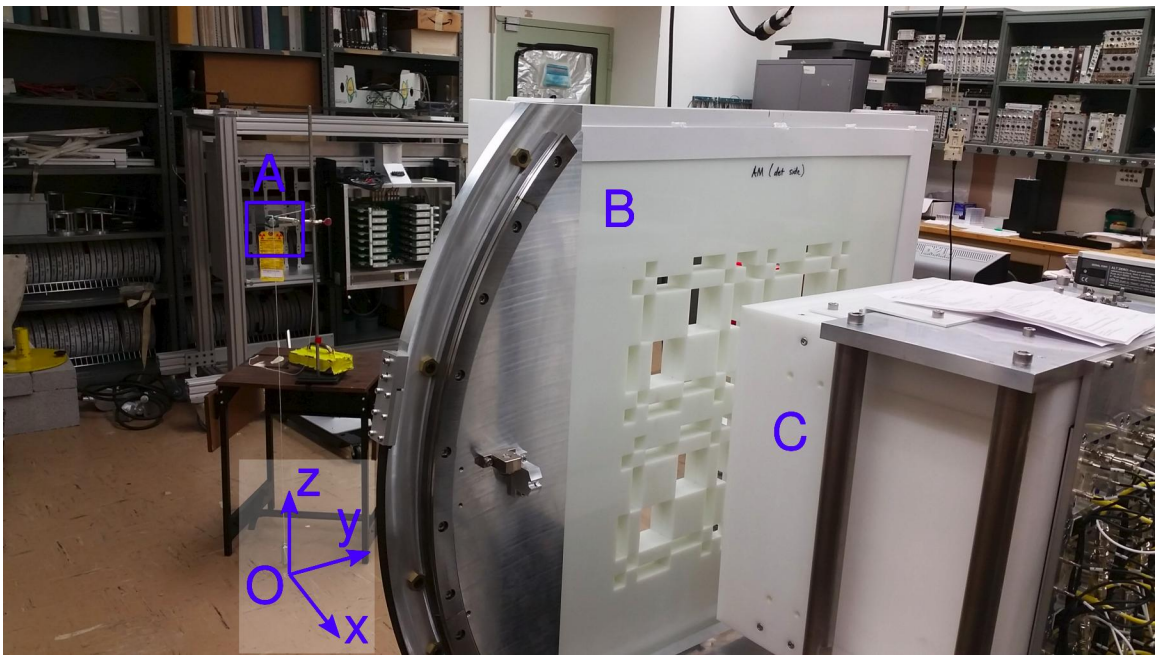


Figure 3.4: Calibration setup with origin (O) and coordinate system (xyz) marked.

For pair p, we measured the source position as $\hat{x}_p$. See Table 3.2. The measured source positions are denoted as a set $\hat{X}_P \equiv \{\hat{x}_p \mid p = 1...P\}$. The source locations were chosen to span the detector's fully and partially coded field of view. (A source position is "fully encoded" if the mask outline overshadows the entire detector face; the position is "partially encoded" if the mask outline overshadows a portion of the detector face.) We anticipate that application measurements may include nuisance sources outside the fully encoded region. We did not include calibration positions outside the partially encoded region, such as the region opposite the FNCA field of view.

Table 3.2: Source locations (centimeters), walltimes (seconds), and total counts for the calibration measurement pairs.

	position	walltime (s)		total counts	
p	(cm)	mask	antimask	mask	antimask
1	$(-5.1, -0.6, 121.0)$	21600	21600	579712	585937
2	$(-3.2, 108.1, 121.0)$	5400	5400	105960	98447
3	$(61.1, -82.5, 121.0)$	5400	5400	195510	177780
4	$(-136.7, -23.2, 121.0)$	27000	27000	319898	319840
5	$(113.0, -0.3, 121.0)$	1800	1800	189115	189079
6	$(89.5, 92.1, 121.0)$	3600	3600	131484	116840
7	$(89.9, 92.7, 155.9)$	3600	3600	118016	102812
8	$(121.9, -2.2, 155.6)$	1800	1800	174984	188594
9	$(-69.5, 153.2, 155.6)$	27000	27000	298122	283653
10	$(-1.0, 0.0, 155.9)$	5400	5400	137636	144154
11	$(0.3, -83.8, 156.2)$	5400	5400	118418	119137
12	$(-134.3, -41.0, 155.9)$	27000	27000	312607	309235
13	$(34.8, -46.2, 155.9)$	1800	1800	58580	58231
14	$(96.7, -81.1, 155.9)$	1800	1800	70613	66026
15	$(-51.8, -84.0, 156.2)$	27000	27000	431456	436547

We stored the measured mask data in a matrix $\hat{D}_0 \in \mathbb{R}^{\mathrm{I}\times\mathrm{P}}$, where $\left[\hat{D}_0\right]_{i,p}$ is the number of counts in pixel i in measurement p. For the P24 imager, $I = 576$ pixels. Similarly, the antimask data are stored in $\hat{D}_1 \in \mathbb{R}^{\mathrm{I}\times\mathrm{P}}$.

3.3.2 Forward Model

We used nonlinear optimization to determine calibration parameters that best fit the subtracted data $\hat{D}_0 - \hat{D}_1$. For convenience, let us summarize the scalar calibration parameters as a vector

$$c \equiv \begin{bmatrix} \Sigma_m & \Sigma_s & b \end{bmatrix}^\top \in \mathbb{R}^3 \quad . \tag{3.23}$$

Table 3.3 lists constraints for the calibration parameters.

We also allow for small adjustments of the source positions. Our measurements $\hat{X}_P$ are precise within a few centimeters, but our P24 measurements are somewhat sensitive to perturbations of this order of magnitude—especially parallel to the detector face. For this reason we allow some in-plane adjustment of source position x_p from the as-measured value $\hat{x}_p$. Adjustments of the source distance ($[x_p]_x$) are not considered. Again, see Table 3.3.

Table 3.3: Bounds on the calibration parameters c and adjusted source positions x_p.

Pixel intrinsic efficiency	$0 <$	$[\varepsilon]_i$	≤ 1	$\forall i \in [1, I]$
Mask cross section	$0.02 \leq$	$c_1 \equiv \Sigma_m$	≤ 0.2	
Scintillator cross section	$0.02 \leq$	$c_2 \equiv \Sigma_s$	≤ 0.2	
Blur factor	$0 \leq$	$c_3 \equiv b$	≤ 0.15	
Source distance		$[x_p]_x = [\hat{x}_p]_x$		$\forall p \in [1, P]$
Source lateral	$[\hat{x}_p]_y - 15 \leq$	$[x_p]_y$	$\leq [\hat{x}_p]_y + 15$	$\forall p \in [1, P]$
Source height	$[\hat{x}_p]_z - 15 \leq$	$[x_p]_z$	$\leq [\hat{x}_p]_z + 15$	$\forall p \in [1, P]$

We model the measurements to determine best-fit estimates of ε, c and X_P. Our forward model predicts counts

$$[D_0(\varepsilon, c, X_P)]_{i,p} \equiv [A_0(\varepsilon, c, X_P)]_{i,p} \; s^{(p)} \quad , \tag{3.24}$$

with source intensity $s^{(p)} = 3.65 \times 10^5$ neutrons/s for all p. The response matrix (as defined in Eqs. (3.6) and (3.7)) now depends on c and X_P:

$$[A_0(\varepsilon, c, X_P)]_{i,p} = (1 - 4c_3) \, [A_{0,u}(\varepsilon, c, X_P)]_{i,p} + c_3 \sum_{i' \in N_i} [A_{0,u}(\varepsilon, c, X_P)]_{i',p} \tag{3.25}$$

with

$$[A_{0,u}(\varepsilon, c, X_P)]_{i,p} \equiv [\varepsilon]_i \, T_{p,0} \sum_{k \in K_i} \left(\frac{v_k \exp\left(- [c]_1 \, t_m - [c]_2 \, t_s\right)}{4\pi |x_k - x_p|^2} \right) \quad . \tag{3.26}$$

The walltime for the mask measurement of pair p is $T_{p,0}$. In these expressions (D_0, A_0) are for the mask; use analogous expressions (D_1, A_1) for the antimask. This defines a matrix of subtracted residuals

$$\hat{R}(\varepsilon, c, X_P) \equiv (D_0(\varepsilon, c, X_P) - D_1(\varepsilon, c, X_P)) - \left(\hat{D}_0 - \hat{D}_1 \right) \quad , \tag{3.27}$$

with $\hat{R}(\varepsilon, c, X_P) \in \mathbb{R}^{\mathrm{I} \times \mathrm{P}}$.

Note that $\hat{R}(\varepsilon, c, X_P)$ is sampled (since $\hat{D}$ is sampled). For simplicity, use covariance model I: assume $\left[\hat{R}(\varepsilon, c, X_P) \right]_{i,p}$ is sampled with variance

$$\mathrm{var}[[\mathbf{R}]_{i,p}] = \left[\hat{D}_0 \right]_{i,p} + \left[\hat{D}_1 \right]_{i,p} \tag{3.28}$$

and with no modeling error and no covariance among pixels, measurements, or pairs.

Based on this model, a reasonable goodness-of-fit metric is

$$\zeta(\varepsilon, c, X_P) \equiv \sum_{i,p} \frac{\left[\hat{R}(\varepsilon, c, X_P)^2\right]_{i,p}}{\left[\hat{D}_0\right]_{i,p} + \left[\hat{D}_1\right]_{i,p}} \quad . \tag{3.29}$$

A small value of $\zeta(\varepsilon, c, X_P)$ indicates that c and X_P fit the subtracted data well. The weighted least-squares calibration estimator is then

$$\varepsilon^*, c^*, X_P^* \equiv \underset{\varepsilon, c, X_P}{\operatorname{argmin}}\, \zeta(\varepsilon, c, X_P) \quad . \tag{3.30}$$

We evaluate ε^*, c^*, and X_P^* using nonlinear optimization. The intrinsic efficiencies ε can be estimated in one step. Given estimates of c^* and X_P^*, the optimal ε is approximately

$$[\varepsilon^*]_i \approx [\varepsilon]_i \frac{\sum_p w_{0,p} \left[\hat{D}_0\right]_{i,p} + w_{1,p} \left[\hat{D}_1\right]_{i,p}}{\sum_p w_{0,p} \left[D_0(\varepsilon, c^*, X_P^*)\right]_{i,p} + w_{1,p} \left[D_1(\varepsilon, c^*, X_P^*)\right]_{i,p}} \quad \text{, with} \tag{3.31}$$

$$w_{0,p} \equiv \frac{1}{\sqrt{\left[\hat{D}_0\right]_{i,p}}} \quad . \tag{3.32}$$

This is a weighted average ratio between measured and predicted counts. Although this may not satisfy Eq. (3.30) exactly, it is guaranteed to be non-negative.

We use this prescription for ε to optimize the components of c and X_P. Using a bisection line search, we can minimize each component. We update ε at each evaluation in the line search. Each line search terminates when the solution ζ is estimated to be within within 0.1% of the minimum ζ. Because ζ is not a separable function, we cycle through the line searches serially until every component's change is smaller than 10^{-6}.

Using this analysis procedure, the effective intrinsic efficiencies range from 0.1539% to 3.048%. See Fig. 3.5.

The best-fit calibration parameters are $\Sigma_m = 0.1392\,\mathrm{cm}^{-1}$, $\Sigma_s = 0.178\,\mathrm{cm}^{-1}$, and $b = 0.0949$. The mask cross section is reasonable for the californium-252 spectrum in polyethylene. It indicates that the mask is 0.71 mean-free-paths thick, and a neutron beam normally incident on a closed element is attenuated to 49% of its original strength. The scintillator cross section is much higher than expected. This

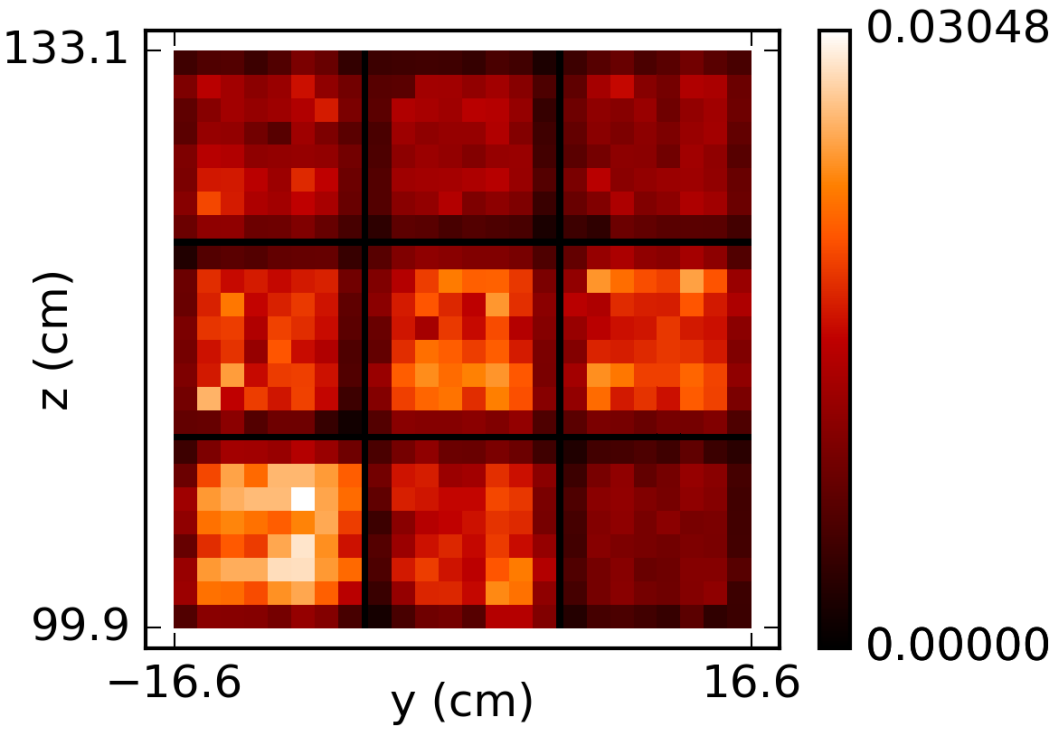


Figure 3.5: Best-fit intrinsic efficiencies from the calibration measurements.

suggests that the typical interaction depth in the scintillator is quite shallow, perhaps for reasons other than scintillator attenuation. The blur value indicates that roughly half of the fast neutron events are misplaced in a neighboring pixel. This finding is consistent with previous findings personally communicated with Paul Hausladen.

The X_P adjustments show a distinct trend shifting the sources in the $+y$ direction. See Table 3.4. Follow-up measurements confirm that the imager was oriented slightly to the left of the $-x$-axis, so the $+y$ correction is reasonable.

Table 3.4: Calibration adjustments to X_p.

p	$X_p - \hat{X}_p$
1	(0.0, 2.44629, -0.75439)
2	(0.0, 3.04688, -0.82031)
3	(0.0, 2.46094, -0.74707)
4	(0.0, 4.21875, -1.40625)
5	(0.0, 0.82031, 0.0293)
6	(0.0, 1.875, -0.23438)
7	(0.0, 2.10938, -0.05859)
8	(0.0, 0.9668, -0.0293)
9	(0.0, 3.75, -1.75781)
10	(0.0, 2.92969, -0.58594)
11	(0.0, 3.28125, -0.82031)
12	(0.0, 4.10156, -1.64062)
13	(0.0, 2.46094, -0.70312)
14	(0.0, 1.64062, -0.70312)
15	(0.0, 3.92578, -1.23047)

It is possible to propagate measurement uncertainty from the calibration data to the calibrated parameters. However, one expects that the model's approximations limit its predictive power far more than calibration data uncertainty does. We therefore forgo uncertainty analysis on the calibration parameters. Instead we will train a covariance model based on the observed prediction–measurement residuals (Section 3.3.4), then use cross-validation to determine the validity of the calibrated model (Section 3.4).

3.3.3 Covariance Model I

Before we train our generalized covariance model, we should assess the accuracy of the Poisson covariance model. Covariance model I is based on counting statistics and has no calibration parameters. (Refer to Eq. (3.15).)

We can tabulate a goodness-of-fit parameter

$$\hat{\chi}_p^2 \equiv \left[\hat{R}\right]_{:,p}^{\top} \left(\text{covar}[[\mathbf{R}]_{:,p}]\right)^{-1} \left[\hat{R}\right]_{:,p} \tag{3.33}$$

to evaluate the fit between the modeled and measured data of pair p. The residuals are calculated using the calibrated forward model, and the covariance matrix is calculated using covariance model I. If $[\mathbf{R}]_{:,p}$ is a normal distribution with mean zero and covariance covar$[[\mathbf{R}]_{:,p}]$, we expect $\hat{\chi}_p^2$ to be sampled from a chi-squared distribution with I degrees of freedom. Based on the chi-squared distribution, we expect $\hat{\chi}_p^2 \approx I$ and the right-tail probability density function (PDF) integral $P(\chi^2 > \hat{\chi}_p^2)$ to distribute uniformly between 0% and 100%. Calculated values of $\hat{\chi}_p^2$, etc. are listed in Table 3.5.

The values of $\hat{\chi}_p^2$ are unreasonably large. This indicates that the model significantly under-predicts the magnitude of prediction–measurement residuals. This motivates us to use of a more generalized covariance model.

Table 3.5: Goodness-of-fit parameter for each calibration pair using covariance model I. (Pair p is included in the calibration data.)

p	$\hat{\chi}_p^2$	I	$\hat{\chi}_p^2/I$	$P(\chi^2 > \hat{\chi}_p^2)$
1	2256.5	576	3.92	0.0%
2	691.0	576	1.20	0.1%
3	1034.4	576	1.80	0.0%
4	1166.4	576	2.03	0.0%
5	1286.2	576	2.23	0.0%
6	1009.7	576	1.75	0.0%
7	1131.8	576	1.96	0.0%
8	1017.7	576	1.77	0.0%
9	1202.2	576	2.09	0.0%
10	880.3	576	1.53	0.0%
11	739.6	576	1.28	0.0%
12	1141.5	576	1.98	0.0%
13	679.9	576	1.18	0.2%
14	845.9	576	1.47	0.0%
15	1401.1	576	2.43	0.0%

3.3.4 Covariance Model II

The analysis to calibrate covariance model II is similar to the analysis to calibrate the forward model. The calibration parameters are the relative error ρ, the

same-measurement correlation ρ_{00}, and the mask–antimask correlation ρ_{01}. (Refer to Eq. (3.21).)

In later sections we perform non-Bayesian inference using the χ^2 goodness-of-fit metric. There we take advantage of the fact that χ^2 follows a chi-squared distribution. As the results in Section 3.3.3 show, this property is only true if the covariance is predicted correctly.

Using covariance model II, the goodness-of-fit metric for pair p is

$$\hat{\chi}^2_p(\rho, \rho_{00}, \rho_{01}) \equiv \left[\hat{R}\right]^\top_{:,p} \Big(\text{covar}[[\mathbf{R}]_{:,p}](\rho, \rho_{00}, \rho_{01})\Big)^{-1} \left[\hat{R}\right]_{:,p} \quad , \tag{3.34}$$

where $\left[\hat{R}\right]_{:,p}$ is column p of the calibrated-model residuals $\hat{R}(\varepsilon^*, c^*, X^*_P)$ and $\text{covar}[[\mathbf{R}]_{:,p}]$ is the estimated covariance of $\left[\hat{R}\right]_{:,p}$. Since $\text{covar}[[\mathbf{R}]_{:,p}]$ depends on the calibration parameters $(\rho, \rho_{00}, \rho_{01})$, $\hat{\chi}^2_p$ does too.

Since $\left[\hat{R}\right]_{:,p}$ has I elements, we expect $\hat{\chi}^2_p \approx I$. Therefore a reasonable objective function for fitting is

$$\zeta(\rho, \rho_{00}, \rho_{01}) \equiv \sum_p \left(\hat{\chi}^2_p(\rho, \rho_{00}, \rho_{01}) - I\right)^2 \quad . \tag{3.35}$$

Our best-fit calibration parameters are

$$\rho^*, \rho^*_{00}, \rho^*_{01} \equiv \underset{\rho, \rho_{00}, \rho_{01}}{\text{argmin}}\, \zeta(\rho, \rho_{00}, \rho_{01}) \quad . \tag{3.36}$$

We use the Nelder–Mead algorithm [58, 59, 60] to solve this relatively small nonlinear optimization problem.

The best-fit covariance parameters are $\rho = 3.522 \times 10^{-3}$, $\rho_{00} = 7.77 \times 10^{-1}$, and $\rho_{01} = -5.63 \times 10^{-1}$. Roughly speaking, we expect a prediction–measurement error on the order of $\sqrt{\rho} \approx 5\%$ in addition to counting statistics. Based on the correlation coefficients, we note that roughly $\rho^2_{00} \approx 60\%$ of $\text{var}[\mathbf{R}]$ can be explained by same-measurement covariance and $\rho^2_{01} \approx 32\%$ of $\text{var}[\mathbf{R}]$ can be explained by same-pixel mask–antimask covariance.

The magnitude of these parameters is surprisingly high. Model improvements may yield less error and covariance, and further effort could improve the calibraiton measurements by minimizing counts from scattered neutrons. However, if the covariance model is predictive, then we can proceed with quantitative inference.

Table 3.6 shows the goodness-of-fit metric using covariance model II. The values

of $\hat{\chi}_p^2$ calculated using covariance mdoel II are much more reasonable than the values calculated using model I. These results indicate that the forward model and covariance model II fit the calibration data well.

Table 3.6: Goodness-of-fit parameter for each calibration pair using covariance model II. (Pair p is included in the calibration data.)

p	$\hat{\chi}_p^2$	I	$\hat{\chi}_p^2/I$	$P(\chi^2 > \hat{\chi}_p^2)$
1	586.9	576	1.02	36.7%
2	532.5	576	0.92	90.2%
3	579.2	576	1.01	45.5%
4	579.5	576	1.01	45.1%
5	649.2	576	1.13	1.8%
6	559.8	576	0.97	67.8%
7	595.4	576	1.03	27.9%
8	563.4	576	0.98	63.8%
9	520.7	576	0.90	95.2%
10	600.5	576	1.04	23.2%
11	555.3	576	0.96	72.5%
12	580.3	576	1.01	44.2%
13	604.8	576	1.05	19.7%
14	634.9	576	1.10	4.5%
15	508.2	576	0.88	98.0%

3.3.5 Covariance Model III

We repeat the calibration procedure of covariance model II with covariance model III. (Refer to the model definition in Eq. (3.22).) The best-fit covariance parameters are $\rho = 1.612 \times 10^{-3}$, $\rho_{00} = 9.01 \times 10^{-2}$, and $\rho_{01} = -1.94 \times 10^{-1}$. Table 3.7 shows the goodness-of-fit metric using covariance model III.

Covariance model III predicts far less correlation among the datapoints than model II. This is useful because correlation can make the inference problem poorly conditioned. Unfortunately, the fit metric in Table 3.7 shows significant trends.

Tables 3.5–3.7 show reasonably good fit between the predictions, covariance models II and III, and the data. However, showing that the model fits the calibration data is insufficient to show that the model predicts future measurements [61]. A model with many degrees of freedom could perfectly fit an arbitrary dataset without providing any capability to predict future data. Since we need a predictive capability to perform quantitative inference, we will use cross-validation to assess this capability.

Table 3.7: Goodness-of-fit parameter for each calibration pair using covariance model III. (Pair p is included in the calibration data.)

p	$\hat{\chi}^2_p$	I	$\hat{\chi}^2_p/I$	$P(\chi^2 > \hat{\chi}^2_p)$
1	619.8	576	1.08	10.1%
2	453.9	576	0.79	100.0%
3	607.2	576	1.05	17.8%
4	502.0	576	0.87	98.8%
5	642.4	576	1.12	2.8%
6	636.1	576	1.10	4.2%
7	712.7	576	1.24	0.0%
8	572.6	576	0.99	53.2%
9	570.2	576	0.99	56.0%
10	528.0	576	0.92	92.5%
11	493.1	576	0.86	99.5%
12	505.5	576	0.88	98.4%
13	492.6	576	0.86	99.5%
14	628.2	576	1.09	6.5%
15	547.3	576	0.95	80.0%

3.4 Cross-Validation

We hypothesize that predictions made using the calibrated model will match experimental data and that the prediction–measurement residuals will be distributed according to covariance model I, II, or III. To show this we use leave-one-out cross-validation: recalibrate using all pairs except a validation pair p_{val}, then compare the p_{val} measurement data to the model prediction. See Fig. 3.6. The resulting goodness-of-fit metric $\chi^2_{p_{\text{val}}}$ ideally follows a chi-squared distribution with I degrees of freedom. Even when it does not follow a chi-squared distribution, we can use the cross-validation result to set a threshold on what values of χ^2 are reasonable for a 95% confidence region.

Table 3.8 lists the recalibrated parameters for the forward model. The source positions X_P were not adjusted. Tables 3.9 and 3.10 list the recalibrated parameters for covariance models II and III. We use these calibration parameters to evaluate the model fit for individual pixels and for the overall chi-squared distribution.

We would like to show that the prediction–measurement residuals $\hat{r}_{p_{\text{val}}}$ follow the Gaussian distribution defined by the covariance model. We can de-correlate the prediction–measurement residuals of a measurement pair using the covariance ma-

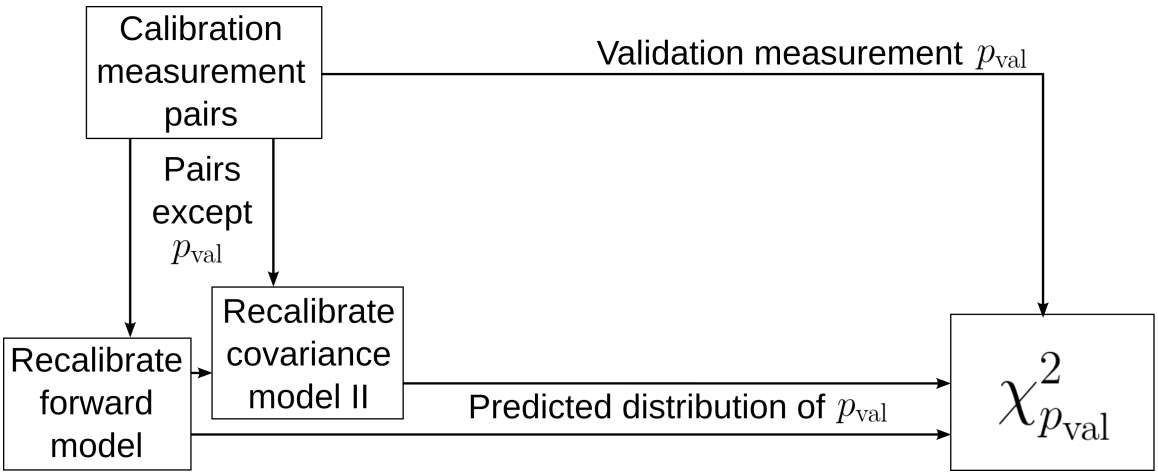


Figure 3.6: Diagram of leave-one-out cross-validation.

Table 3.8: Forward model calibration parameters calculated using all pairs except p_{val}.

p_{val}	Σ_m	Σ_s	b
1	0.1388	0.178	0.0984
2	0.1392	0.178	0.0949
3	0.1385	0.178	0.0938
4	0.1409	0.178	0.0967
5	0.1378	0.178	0.0938
6	0.1385	0.178	0.0949
7	0.1385	0.183	0.0949
8	0.1378	0.172	0.0938
9	0.1402	0.178	0.0932
10	0.1390	0.178	0.0949
11	0.1392	0.178	0.0938
12	0.1406	0.178	0.0949
13	0.1388	0.178	0.0938
14	0.1385	0.178	0.0949
15	0.1402	0.178	0.0938

Table 3.9: Calibration parameters for covariance model II calculated using all pairs except p_{val}.

p_{val}	ρ	ρ_{00}	ρ_{01}
1	2.775×10^{-3}	0.515	-0.361
2	3.231×10^{-3}	0.661	-0.477
3	3.447×10^{-3}	0.606	-0.387
4	3.563×10^{-3}	0.639	-0.414
5	3.534×10^{-3}	0.815	-0.591
6	3.403×10^{-3}	0.669	-0.456
7	3.313×10^{-3}	0.539	-0.324
8	3.486×10^{-3}	0.548	-0.324
9	3.515×10^{-3}	0.715	-0.499
10	3.571×10^{-3}	0.600	-0.365
11	3.367×10^{-3}	0.585	-0.377
12	3.51×10^{-3}	0.607	-0.387
13	3.767×10^{-3}	0.614	-0.355
14	3.566×10^{-3}	0.516	-0.270
15	3.826×10^{-3}	0.582	-0.333

Table 3.10: Calibration parameters for covariance model III calculated using all pairs except p_{val}.

p_{val}	ρ	ρ_{00}	ρ_{01}
1	1.348×10^{-3}	0.0354	-0.161
2	1.583×10^{-3}	0.136	-0.225
3	1.791×10^{-3}	0.0982	-0.136
4	1.933×10^{-3}	0.0884	-0.0939
5	1.891×10^{-3}	0.0672	-0.0792
6	1.81×10^{-3}	0.100	-0.128
7	1.927×10^{-3}	0.106	-0.0896
8	1.87×10^{-3}	0.0816	-0.110
9	1.783×10^{-3}	0.0765	-0.122
10	1.763×10^{-3}	0.118	-0.159
11	1.746×10^{-3}	0.0813	-0.143
12	1.928×10^{-3}	0.0841	-0.0937
13	1.613×10^{-3}	0.135	-0.216
14	2.036×10^{-3}	0.0946	-0.0715
15	2.027×10^{-3}	0.0514	-0.0639

trix:

$$\hat{r}'_{p_{\text{val}}} \equiv \Sigma^{-1/2} \hat{r}_{p_{\text{val}}} \quad , \tag{3.37}$$

where $\Sigma^{-1/2}$ is the inverse of the Cholesky decomposition of the covariance matrix (calculated using the covariance model). We expect that the elements of $\hat{r}'_{p_{\text{val}}}$ are uncorrelated and follow a standard normal distribution. Figure 3.7 compares $\hat{r}'_{p_{\text{val}}}$ to the quantiles of a standard normal distribution. For example, we plot the lowest element of $\hat{r}'_{p_{\text{val}}}$ against the $\frac{1-1/2}{I}$ quantile of a standard normal distribution, plot the second-lowest element against the $\frac{2-1/2}{I}$ quantile, etc. Figure 3.7 is a Quantile–Quantile (QQ) plot using covariance model I; the procedure is repeated for models II and III in Figs. 3.8 and 3.9.

If there is covariance among the pixels, a scaled residual does not correspond to any particular detector pixel. However, covariance model I assumes no covariance, so the residuals correspond to pixels. Figure 3.10 plots $\hat{r}'_{p_{\text{val}}}$ as calculated using covariance model I. This plot shows significant residuals for $p_{\text{val}} = 1$ and in the lower-left $(-y, -z)$ scintillator block.

These QQ plots show the best agreement between the residuals and the distribution predicted by covariance model II. In a few pairs—especially $p_{\text{val}} = 1$—the observed lower quantiles are more extreme than the expected quantiles. This "long tail" of the distribution indicates that the residual magnitude is underpredicted in some pixels. This warrants further investigation.

We also would like to show that $\hat{\chi}^2_{p_{\text{val}}}$ follows a chi-squared distribution with I degrees of freedom. Observed values of $\hat{\chi}^2_{p_{\text{val}}}$ are listed in Tables 3.11–3.13. Figures 3.11–3.13 are QQ plots of $\hat{\chi}^2_{p_{\text{val}}}$.

For covariance model I, the quantiles of $\hat{\chi}^2_{p_{\text{val}}}$ far exceed a chi-squared distribution. However we can use the observed distribution to assign an empirical threshold $\chi^2_{I;0.95} = 1500$ for $I = 576$ degrees of freedom.

For covariance model II, the observed distribution of $\hat{\chi}^2_{p_{\text{val}}}$ appears to have the same mean—but a somewhat broader dispersion—compared to the expected distribution. The clear outlier $p_{\text{val}} = 1$ is the measurement with the most counts. This could indicate a limit on the model's predictive capabiility: the calibrated forward model and covariance model are valid for measurements with fewer than 5×10^5 counts. We assign an empirical threshold $\chi^2_{I;0.95} = 750$ for $I = 576$ degrees of freedom for model II.

Results for covariance model III are similar to model II, but with even broader

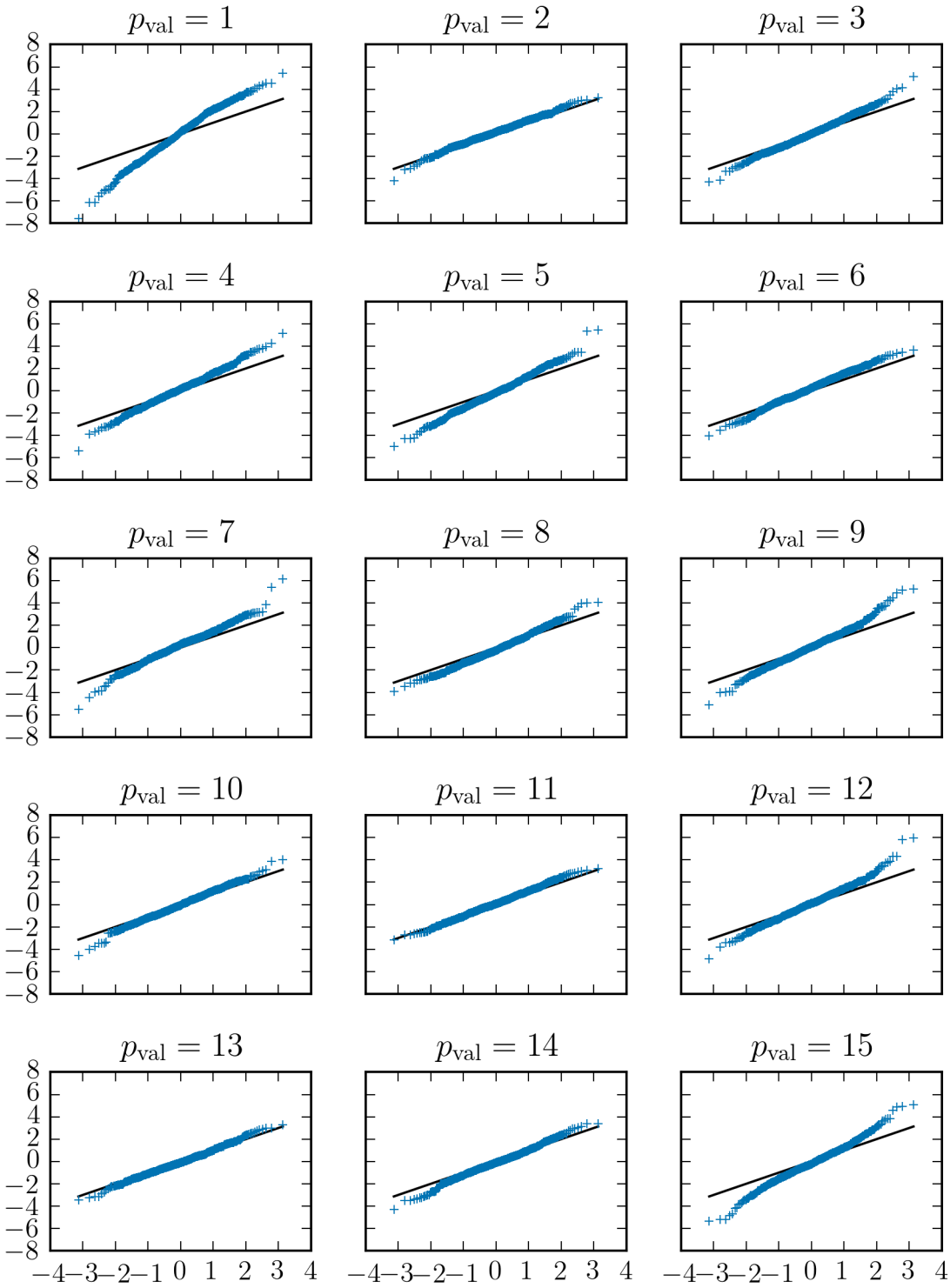


Figure 3.7: Quantiles of the cross-validation residuals (vertical axis) vs a standard normal distribution (horizontal axis) using covariance model I. Expected trend marked in black.

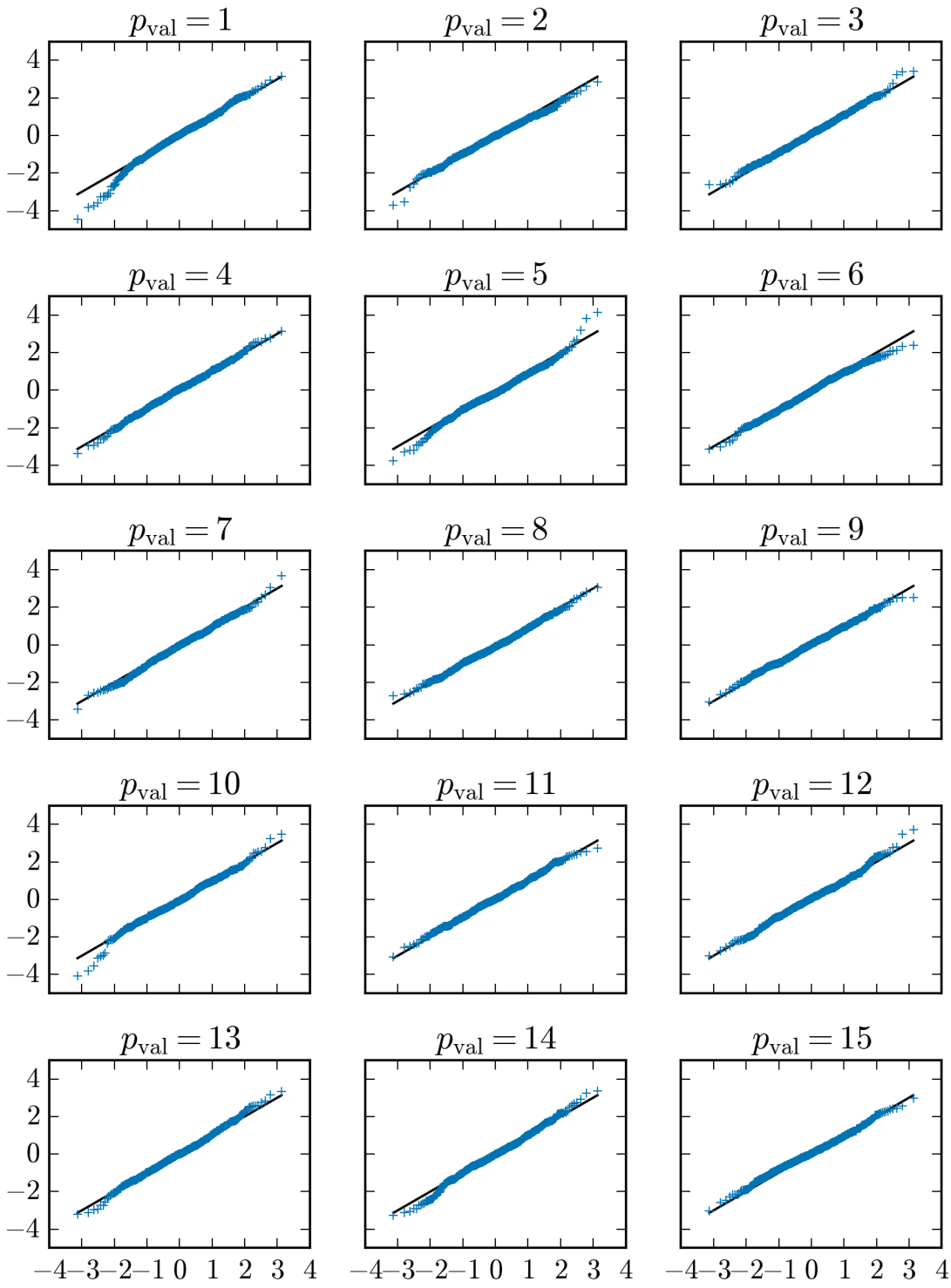


Figure 3.8: Quantiles of the cross-validation residuals (vertical axis) vs a standard normal distribution (horizontal axis) using covariance model II. Expected trend marked in black.

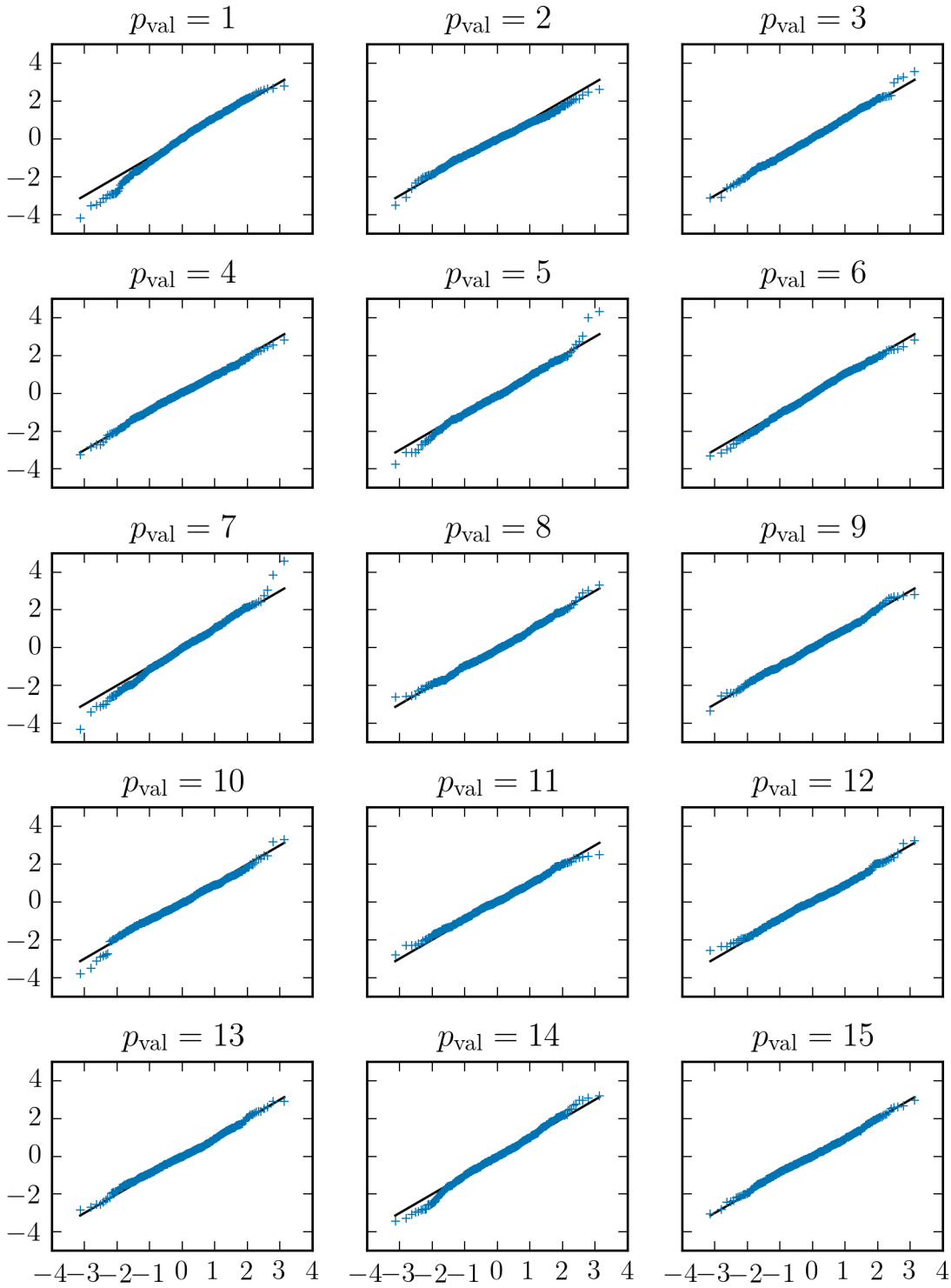


Figure 3.9: Quantiles of the cross-validation residuals (vertical axis) vs a standard normal distribution (horizontal axis) using covariance model III. Expected trend marked in black.

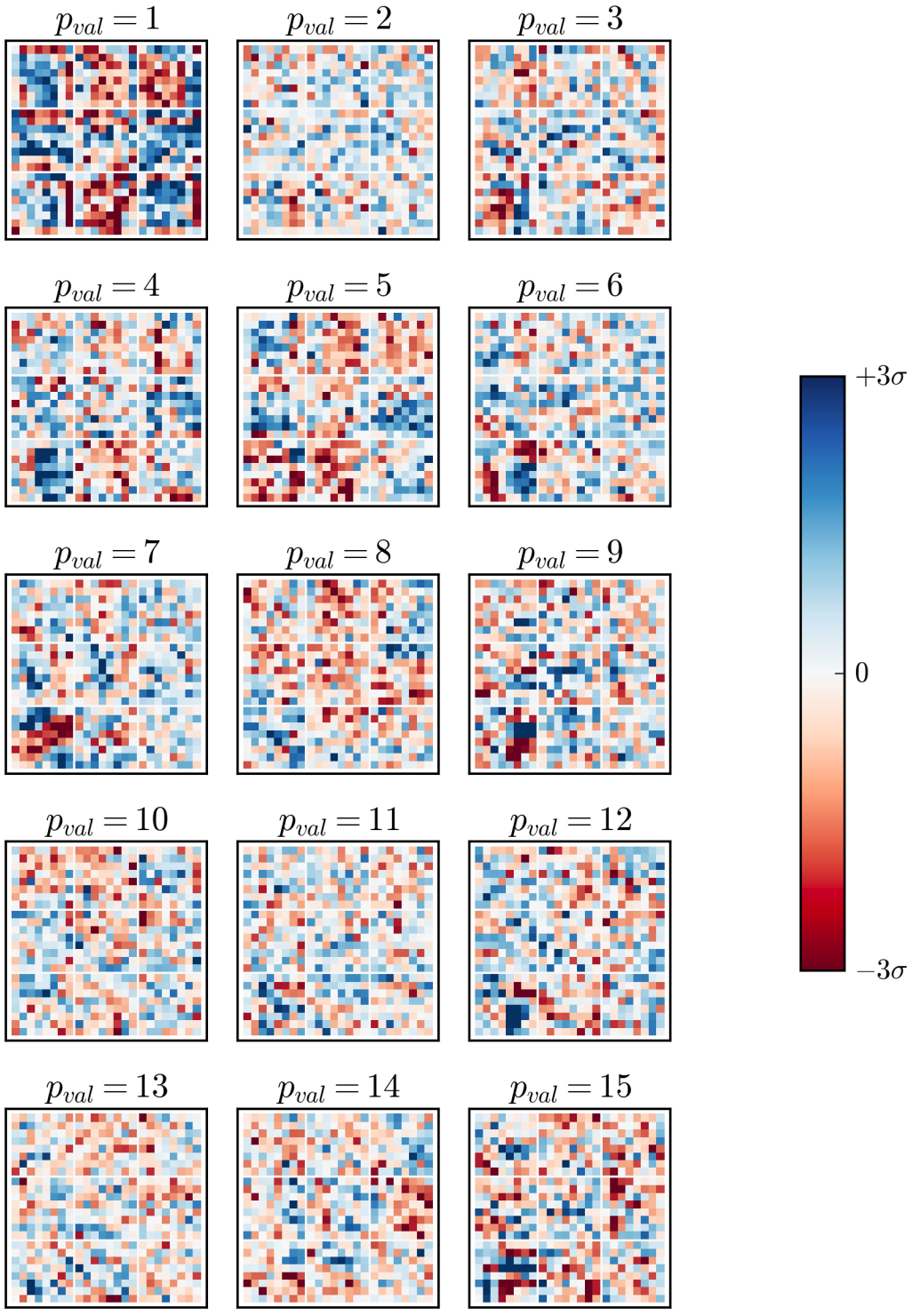


Figure 3.10: Cross-validation scaled residuals $\hat{r}'_{p_{\text{val}}}$ calculated using covariance model I. Horizontal and vertical axes correspond to y and z, respectively.

Table 3.11: Goodness-of-fit parameter for each calibration pair using covariance model I. (Pair p_{val} is excluded from the calibration data.)

p_{val}	I	$\hat{\chi}^2_{p_{\text{val}}}$	$\hat{\chi}^2_{p_{\text{val}}}/I$	$P(\chi^2 > \hat{\chi}^2_{p_{\text{val}}})$
1	576	2348.5	4.08	0.0%
2	576	673.1	1.17	0.3%
3	576	999.2	1.73	0.0%
4	576	1181.0	2.05	0.0%
5	576	1280.6	2.22	0.0%
6	576	961.6	1.67	0.0%
7	576	1108.9	1.93	0.0%
8	576	986.3	1.71	0.0%
9	576	1201.5	2.09	0.0%
10	576	861.1	1.49	0.0%
11	576	733.6	1.27	0.0%
12	576	1119.0	1.94	0.0%
13	576	685.3	1.19	0.1%
14	576	834.2	1.45	0.0%
15	576	1391.6	2.42	0.0%
overall	8640	16365.4	1.89	0.0%

Table 3.12: Goodness-of-fit parameter for each calibration pair using covariance model II. (Pair p_{val} is excluded from the calibration data.)

p_{val}	I	$\hat{\chi}^2_{p_{\text{val}}}$	$\hat{\chi}^2_{p_{\text{val}}}/I$	$P(\chi^2 > \hat{\chi}^2_{p_{\text{val}}})$
1	576	743.8	1.29	0.0%
2	576	504.7	0.88	98.5%
3	576	575.2	1.00	50.2%
4	576	596.7	1.04	26.7%
5	576	613.9	1.07	13.3%
6	576	546.7	0.95	80.5%
7	576	604.3	1.05	20.0%
8	576	535.0	0.93	88.8%
9	576	512.9	0.89	97.2%
10	576	593.2	1.03	30.1%
11	576	547.2	0.95	80.0%
12	576	570.8	0.99	55.3%
13	576	623.2	1.08	8.5%
14	576	644.1	1.12	2.6%
15	576	482.3	0.84	99.8%
overall	8640	8694.1	1.01	33.9%

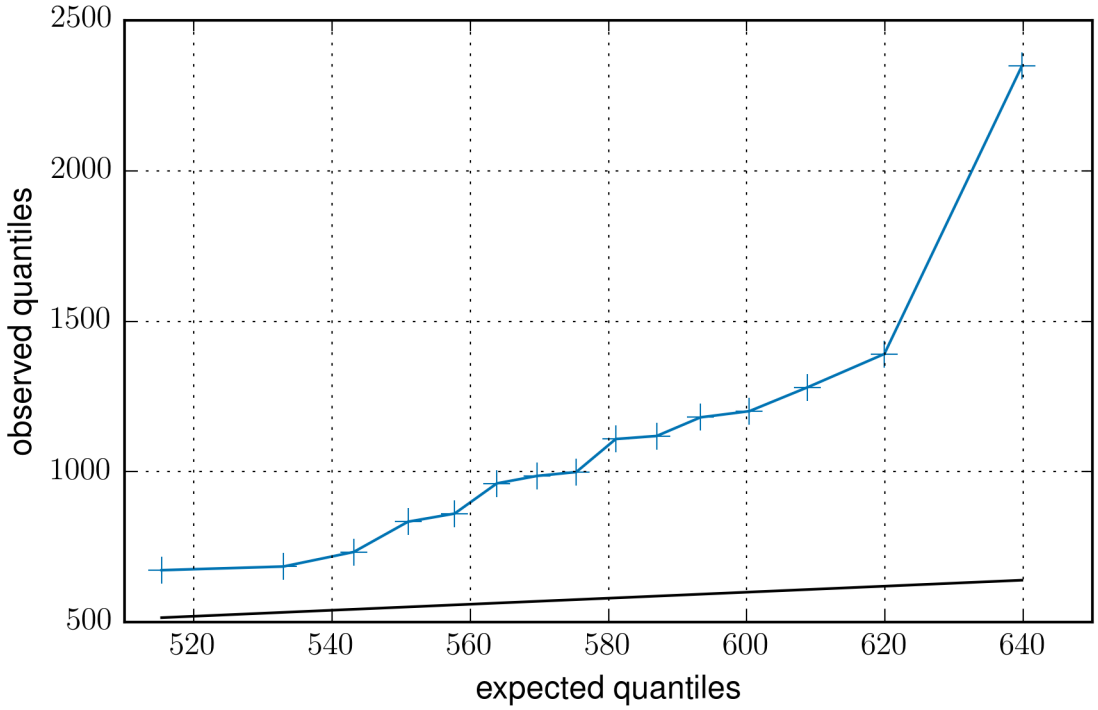


Figure 3.11: Quantiles of $\hat{\chi}^2_{p_{\text{val}}}$ (vertical axis) from covariance model I vs a chi-squared distribution with $I = 576$ degrees of freedom (horizontal axis). Expected trend marked in black.

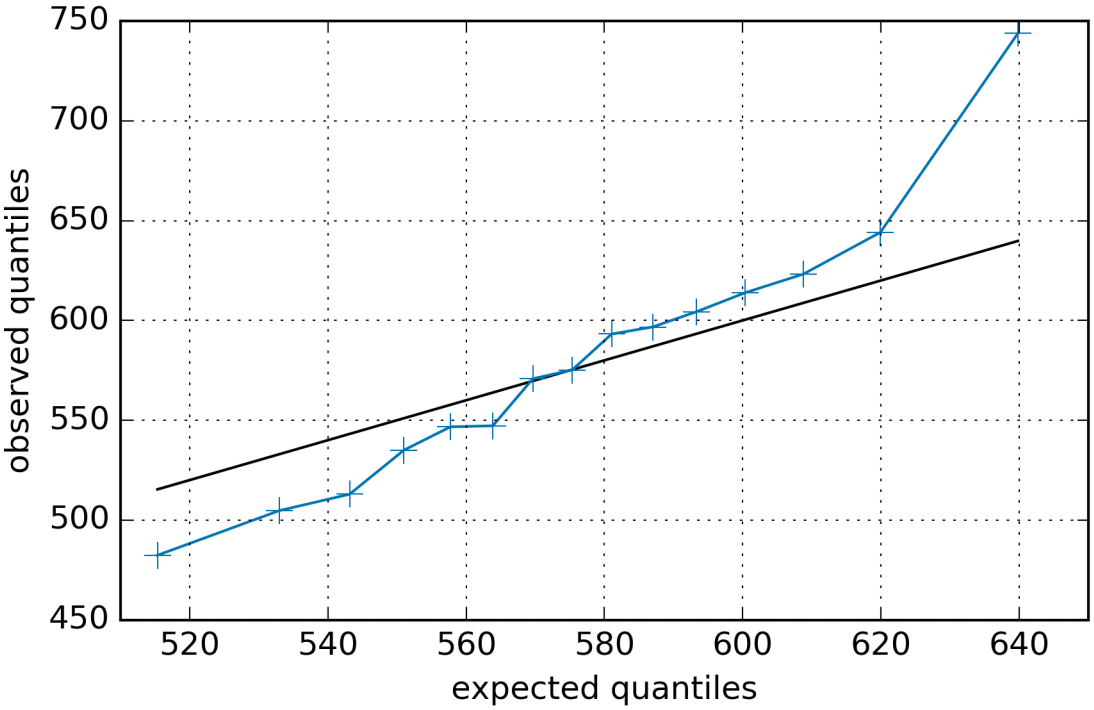


Figure 3.12: Quantiles of $\hat{\chi}^2_{p_{\text{val}}}$ (vertical axis) from covariance model II vs a chi-squared distribution with $I = 576$ degrees of freedom (horizontal axis). Expected trend marked in black.

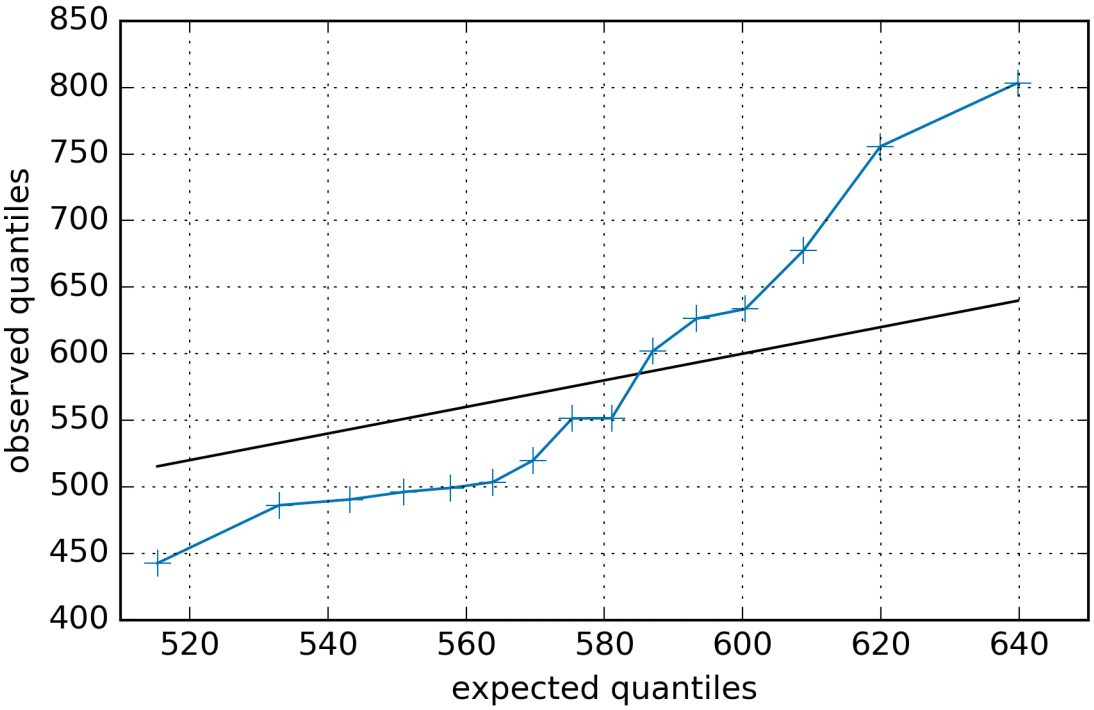


Figure 3.13: Quantiles of $\hat{\chi}^2_{p_{\text{val}}}$ (vertical axis) from covariance model III vs a chi-squared distribution with $I = 576$ degrees of freedom (horizontal axis). Expected trend marked in black.

Table 3.13: Goodness-of-fit parameter for each calibration pair using covariance model III. (Pair p_{val} is excluded from the calibration data.)

p_{val}	I	$\hat{\chi}^2_{p_{\text{val}}}$	$\hat{\chi}^2_{p_{\text{val}}}/I$	$P(\chi^2 > \hat{\chi}^2_{p_{\text{val}}})$
1	576	803.5	1.39	0.0%
2	576	442.6	0.77	100.0%
3	576	601.8	1.04	22.1%
4	576	503.5	0.87	98.7%
5	576	626.3	1.09	7.2%
6	576	633.6	1.10	4.8%
7	576	755.7	1.31	0.0%
8	576	551.5	0.96	76.2%
9	576	551.4	0.96	76.3%
10	576	519.7	0.90	95.5%
11	576	490.4	0.85	99.6%
12	576	486.2	0.84	99.7%
13	576	496.1	0.86	99.3%
14	576	677.4	1.18	0.2%
15	576	499.1	0.87	99.1%
overall	8640	8638.9	1.00	50.1%

dispersion. We assign an empirical threshold $\chi^2_{I;0.95} = 800$ for $I = 576$ degrees of freedom for model III.

3.5 Region-of-Interest Phenomena

The model specified in Section 3.2 ignores scatter and attenuation near the ROI. In this section we model a simplified glove box holdup scenario using Monte Carlo N-Particle Transport Code version 5 [11] (MCNP5) to evaluate the impact of these phenomena.

The model is shown in Fig. 3.14. An example of the MCNP5 input is included in Appendix A. The model includes 1 cm-thick lead glass representing the glove box wall and 0.8 cm-thick stainless steel 304 representing the wall of a pipe. It also includes plutonium nitrate on the sides and in the corners of the glove box and along the sides of the pipe. Cross-section libraries are specified in Table A.6.

We calculate the angular flux at the point $(225, 0, 0)$ cm using a pinhole tally. The image is recorded by a grid of pixels on the $x = 250$ cm plane. The simulations use 10^6 histories and terminates neutrons below 0.1 MeV. In this work we use MCNP5 version 1.60.

If we compare the images formed neglecting materials (vacuum) to the images

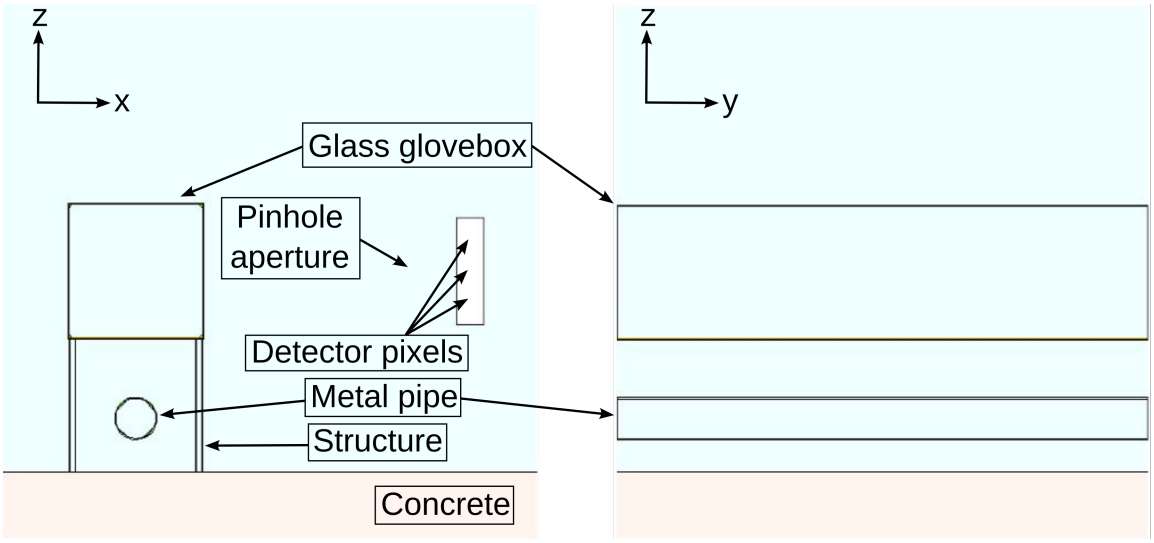


Figure 3.14: Slices through an MCNP5 model of a glove box holdup scenario at $y = 0$ (left) and $x = 0$ (right).

formed with materials, the differences are significant. For example, images tallied with holdup along the sides of the glove box are plotted in Fig. 3.15.

The overall number of neutrons reaching the detector is increased by 33%. See Table 3.14. MCNP5 predicts similar increases when the source is distributed in the corners of the glove box and lining the pipe. It is somewhat surprising that the flux at the detector increases with the addition of materials. This could be explained by the particular model geometry—more neutrons are scattered toward the detector than away from the detector. From these result we conclude that the vacuum approximation causes unacceptably large errors when modeling detector response.

Table 3.14: Neutron flux through a pinhole camera modeled using MCNP5.

Holdup location	Glove box sides	Glove box corners	Pipe lining
Holdup thickness (cm)	0.25 cm	3 cm	1 cm
Flux (neutrons/cm^2 per source neutron)			
Vacuum	5.8544×10^{-7}	5.9131×10^{-7}	4.7014×10^{-7}
Uncollided	4.1198×10^{-7}	3.8639×10^{-7}	2.6912×10^{-7}
Diff w/rt vac.	-29.6%	-34.7%	-42.8%
Full physics	7.7706×10^{-7}	7.9568×10^{-7}	7.7488×10^{-7}
Diff w/rt vac.	+32.7%	+34.6%	+64.8%
No self-shielding	7.6759×10^{-7}	7.7148×10^{-7}	7.4331×10^{-7}
Diff w/rt full phys.	-1.2%	-3.0%	-4.1%

We also need to assess whether self-shielding (interaction of neutrons within the

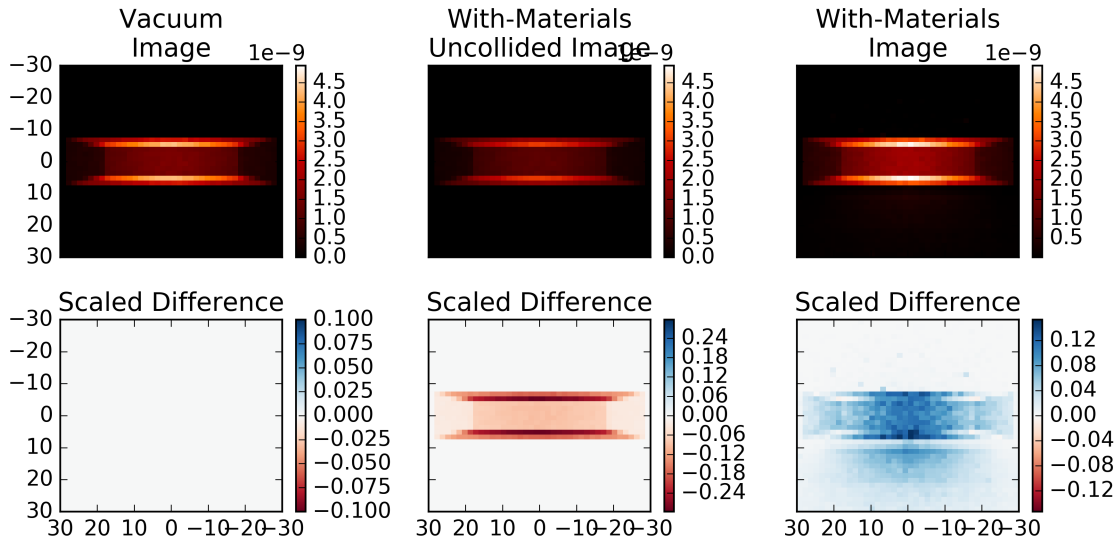


Figure 3.15: MCNP5 simulated images for 0.25 cm of holdup along the sides of a glove box. Images are in neutrons/cm^2 per source neutron. Differences scaled with respect to the brightest pixel in the vacuum image.

deposit) significantly affects detector response. To do this we compare the full-fidelity model to a model with the holdup material replaced with vacuum. See the last rows of Table 3.14. The self-shielding increases the detector response by a few percent. Since real-world deposits would likely be smaller than the modeled deposits, the effect is probably much smaller than other sources of error and uncertainty. This finding is important because it means that the linear model (neglecting self-shielding) is often sufficient.

3.6 Discretization

Source and scintillator discretizations are major sources of modeling error. In this section we assess the affect of discretization on the model estimates.

The unknown source is discretized as a 3D grid of point sources. The grid is ideally fine enough that the goodness-of-fit parameter is insensitive to the approximation. We use the calibration models to assess the resolution of measurements similar to the calibration measurements.

Figure 3.16 plots the goodness-of-fit metric $\hat{\chi}^2_p$ as a function of source position for pair $p = 1$. The fit metric rises rapidly away from a small minimum value. We can define a width of the minimum by setting a threshold on $\hat{\chi}^2_p$ at 1.05× the minimum.

This resolution width is listed for all pairs in Table 3.15. These results suggest that a discretization on the order of 10 cm is appropriate in the x direction and on the order of 3 cm in the y and z directions.

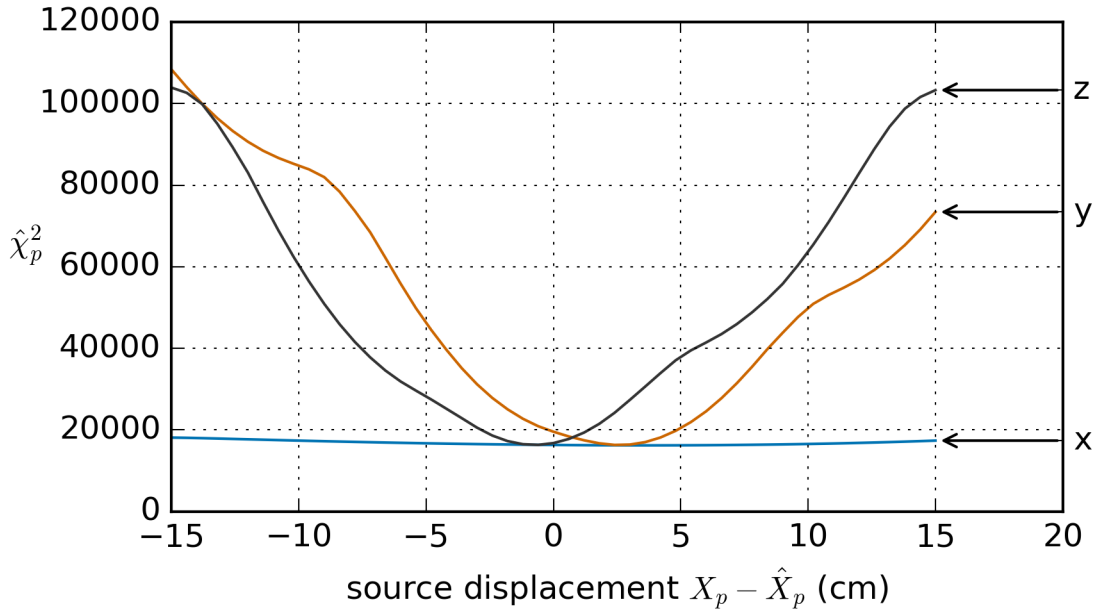


Figure 3.16: Goodness of fit as a function of source repositioning in each dimension (x, y, z) for calibration pair $p = 1$.

Scintillator pixel i is discretized into $|K_i|$ subpixels. This discretization introduces errors in the forward model. We perform a mesh refinement study on the calibration models to determine a sufficiently large $|K_i|$.

The first part of the mesh refinement study compares pairs' predictions $(D_0 - D_1)$ to reference predictions with $|K_i| = 30^3 = 27000$. The comparison metric is the 2-norm of the predictions' difference, i.e.,

$$\left\| (D_0 - D_1)^{(|K_i|)} - (D_0 - D_1)^{(27000)} \right\|_2 \quad . \tag{3.38}$$

The results are plotted in Fig. 3.17. These results indicate limited benefit above $|K_i| = 14^3 = 2744$.

The second part of the mesh refinement study compares pairs' evaluation of $\hat{\chi}^2$ across all of the pairs. This study uses covariance model I. The results plotted in Fig. 3.17 indicated limited affect on $\hat{\chi}^2$ above $|K_i| = 10^3 = 1000$.

Based on these results, we make recommendations for source and scintillator dis-

Table 3.15: Width of the source domain in x, y, or z for which $\hat{\chi}^2_p$ is less than 1.05× its minimum.

p	Width		
	x	y	z
1	20.4	1.8	1.2
2	16.2	7.8	4.2
3	7.2	3.6	2.4
4	30.0	4.8	4.2
5	10.8	1.2	1.2
6	6.6	4.2	1.8
7	7.2	4.2	2.4
8	4.2	0.6	1.2
9	13.8	6.6	4.8
10	20.4	3.6	3.6
11	15.0	5.4	4.2
12	30.0	4.8	4.8
13	19.2	6.0	5.4
14	7.8	5.4	3.6
15	11.4	3.0	2.4

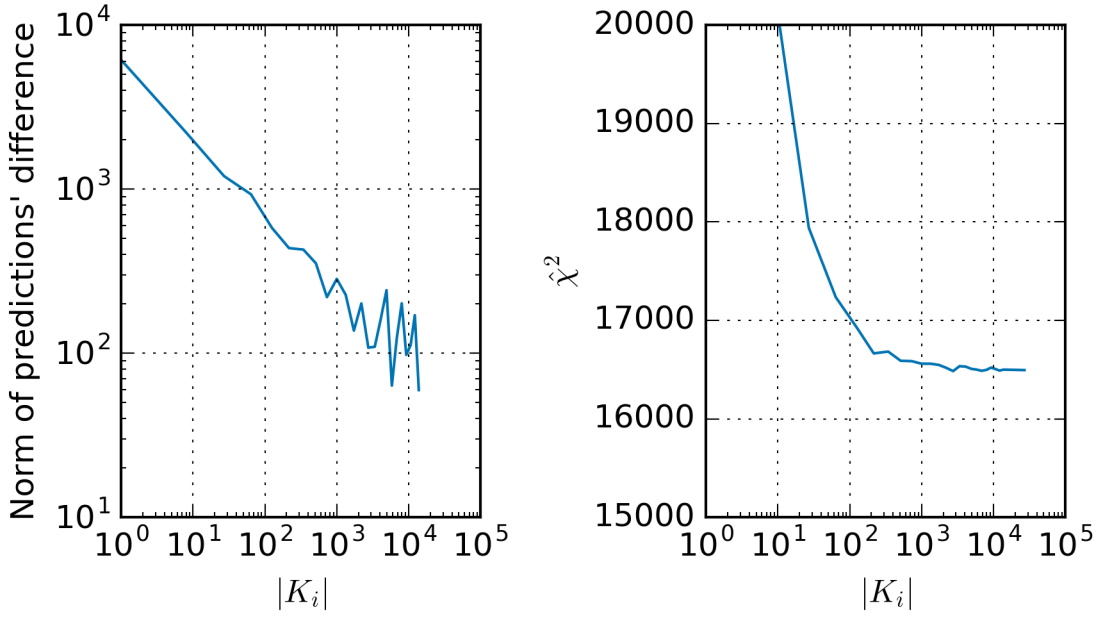


Figure 3.17: Mesh refinement study results for the detector quadrature (per pixel) $|K_i|$ based on the predictions' 2-norm (left) and the goodness-of-fit metric (right).

cretization. Source voxels should be 10 cm perpendicular to the detector plane and 3 cm parallel to the detector plane. Scintillator pixels should be discretized into 14^3 subpixels.

3.7 Conclusions

In this chapter we described and modeled Oak Ridge National Laboratory's P24 FNCA imager. We used point-source measurements to calibrate the model, infer the distribution of prediction–measurement residuals, and cross-validate this distribution. The observed values of $\hat{\chi}^2_{p_{\text{val}}}$ are distributed more broadly than the chi-squared distribution would predict, but the approximation is reasonable.

We also determined minimum requirements for FNCA modeling. Models should account for scatter and attenuation near the ROI, but self-shielding can be neglected. Source voxels should be 10 cm perpendicular to the detector plane and 3 cm parallel to the detector plane. Scintillator pixels should be discretized into 14^3 subpixels.

CHAPTER IV

Aleatoric Uncertainty

Aleatoric uncertainty refers to the limited precision of a measurement. If we repeat a detection measurement using the same setup, we expect slightly different results due to counting statistics. In particular, we expect the counts to follow a nearly Poisson distribution. Although our model attempts to predict the mean of this distribution, it cannot possibly predict the exact value that will be sampled. In this sense, even a "perfect" model will have non-zero prediction–measurement residuals.

If our estimate of the parameters is an inversion of the model, then inverting noisy data will result in a noisy estimate. The magnitude of the noise is typically amplified by the inverse problem, so aleatoric uncertainty can contribute significantly to the overall uncertainty in the estimate.

In our QI UQ problem, noise in the data $\hat{d}$ propagates through the inverse problem to uncertainty in $\hat{s}$ and $\hat{S}$. For simplicity of illustration we will use covariance model I (no modeling error, counting statistics only). In this chapter we derive delta-method uncertainty estimates using MLEM reconstruction. We also show that this approach is similar to GLLS. The results show that the delta method describes the aleatoric uncertainty well in problems that are well-posed ($I > J$) and when the solution is away from inequality constraints. These approaches are less reliable when the solution approaches inequality constraints like the nonnegativity constraints on s.

4.1 Delta-Method Approach

Reconstructed-voxel covariance estimates similar to [41] are possible using MLEM. See also [62, 63, 64]. The covariance matrix could be combined with a total-strength sensitivity matrix to calculate a first-order estimate of the strength uncertainty. The usefulness of this approach depends on the computational cost of estimating the

covariance matrix and the validity of a first-order uncertainty estimate. This may also be useful for creating a 2D benchmark.

We can use a first-order uncertainty approximation to estimate the variance in $\hat{S}$ among repeated realizations of the data $\hat{d}$. See background on the delta method in Section 2.1.1.2.

We first propagate the variance in $\mathbf{d}$ to a covariance matrix for $\mathbf{s}$. The detector data is Poisson in nature, so it has covariance $\text{covar}[\mathbf{d}] = \text{diag}(d) = \text{diag}(As)$. Using the delta method, the reconstructions are approximately Gaussian-distributed with covariance

$$\text{covar}[\mathbf{s}] = B^\top \text{covar}[\mathbf{d}] B \quad , \tag{4.1}$$

where the $I \times J$ Jacobian matrix B has elements $[B]_{i,j} \equiv \frac{\partial [\hat{s}]_j}{\partial [\hat{d}]_i}\Big|_{\hat{s}=s, \hat{d}=As}$. (This definition of B transposes the mathematical convention for the Jacobian so that the shapes of A and B match.) Equation (4.1) is sometimes called the "sandwich equation." The Jacobian matrix is also called the sensitivity matrix. When s and d are not known, we can approximate

$$\text{covar}[\mathbf{d}] \approx \text{diag}(A\hat{s}) \quad \text{and} \tag{4.2}$$

$$B \approx \hat{B} \quad , \tag{4.3}$$

where $[\hat{B}]_{i,j} \equiv \frac{\partial [\hat{s}]_j}{\partial [\hat{d}]_i}\Big|_{\hat{s}=\hat{s}, \hat{d}=A\hat{s}}$. (The approximation $\text{covar}[\mathbf{d}] \approx \text{diag}(A\hat{s})$ is more accurate than $\text{covar}[\mathbf{d}] \approx \text{diag}(\hat{d})$.)

We then propagate $\text{covar}[\mathbf{s}]$ to $\text{var}[\mathbf{S}]$. Since $\hat{S} = \sum_j [\hat{s}]_j$, $\frac{\partial \hat{S}}{\partial [\hat{s}]_j} = 1$. The delta method then approximates

$$\text{var}[\mathbf{S}] \approx \vec{1}^\top \text{covar}[\mathbf{s}] \vec{1} = \sum_j \sum_{j'} [\text{covar}[\mathbf{s}]]_{j,j'} \quad . \tag{4.4}$$

The balance of this section is dedicated to the non-trivial step of calculating the Jacobian matrix B. For convenience, we refer to row i of B as $b_i \equiv \frac{\partial \hat{s}}{\partial [\hat{d}]_i}\Big|_{\hat{s}=s, \hat{d}=As}$. We can derive an estimator of b_i by differentiating the ML equation to form I linear systems of J equations.

Differentiating Eq. (2.42) with respect to the counts in detector pixel i gives

$$\frac{\partial}{\partial[\hat{d}]_i}\left(A^\top \vec{1}\right) = \frac{\partial}{\partial[\hat{d}]_i}\left(A^\top \mathrm{diag}(A\hat{s})^{-1}\hat{d}\right) \tag{4.5}$$

$$\vec{0} = A^\top \frac{\partial}{\partial[\hat{d}]_i}\left(\mathrm{diag}(A\hat{s})^{-1}\right)\hat{d} + A^\top \mathrm{diag}(A\hat{s})^{-1}\vec{e}_i \quad , \tag{4.6}$$

where $\vec{0}$ is a J-vector of zeros and $\vec{e}_i$ is an I-vector such that

$$[\vec{e}_i]_{i'} \equiv \begin{cases} 1 & i' = i \\ 0 & else \end{cases} \quad . \tag{4.7}$$

Chain rule gives

$$\frac{\partial}{\partial[\hat{d}]_i}\left(\mathrm{diag}(A\hat{s})^{-1}\right) = -\mathrm{diag}(A\hat{s})^{-2}\ \mathrm{diag}\left(A\frac{\partial\hat{s}}{\partial[\hat{d}]_i}\right) \quad . \tag{4.8}$$

We can substitute Eq. (4.8) into Eq. (4.6) and re-arrange to show

$$A^\top \mathrm{diag}(A\hat{s})^{-2}\mathrm{diag}\left(A\frac{\partial\hat{s}}{\partial[\hat{d}]_i}\right)\hat{d} = A^\top \mathrm{diag}(A\hat{s})^{-1}\vec{e}_i \quad . \tag{4.9}$$

Since $\mathrm{diag}(x)y = \mathrm{diag}(y)x$,

$$A^\top \mathrm{diag}(A\hat{s})^{-2}\mathrm{diag}(\hat{d})A\frac{\partial\hat{s}}{\partial[\hat{d}]_i} = A^\top \mathrm{diag}(A\hat{s})^{-1}\vec{e}_i \quad . \tag{4.10}$$

We evaluate this expression at $\hat{s} = s, \hat{d} = As$ to show

$$A^\top \mathrm{diag}(As)^{-1}A\ b_i = A^\top \mathrm{diag}(As)^{-1}\vec{e}_i \tag{4.11}$$

or evaluate at $\hat{s} = \hat{s}, \hat{d} = A\hat{s}$ to show

$$A^\top \mathrm{diag}(A\hat{s})^{-1}A\ \hat{b}_i = A^\top \mathrm{diag}(A\hat{s})^{-1}\vec{e}_i \quad \text{or} \tag{4.12}$$

$$A^\top \mathrm{diag}(A\hat{s})^{-1}A\ \hat{B}^\top = A^\top \mathrm{diag}(A\hat{s})^{-1} \quad . \tag{4.13}$$

We must use linear algebra techinques to solve for $\hat{B}$. Conveniently the left-side

coefficients in Eqs. (4.11) and (4.12) are independent of i. Therefore a single LDL$^\top$-decomposition of the left side can accelerate the Gaussian-elimination solution of all I systems. The system can be ill-conditioned for large J, which precludes accurate calculation of $\hat{B}$.

Note that this result of combining MLEM with the delta method is very similar to GLLS. Substituting $\hat{B}$ in Eq. (4.13) as B in Eq. (4.1) and undoing the approximation $\text{covar}[\mathbf{d}] \approx \text{diag}(A\hat{s})$ gives

$$\text{covar}[\mathbf{s}] = \left(\left(A^\top \text{covar}[\mathbf{d}]^{-1} A\right)^{-1} A^\top \text{covar}[\mathbf{d}]^{-1}\right) \ ...$$

$$... \ \text{covar}[\mathbf{d}] \left(\left(A^\top \text{covar}[\mathbf{d}]^{-1} A\right)^{-1} A^\top \text{covar}[\mathbf{d}]^{-1}\right)^\top \tag{4.14}$$

$$... = \left(A^\top \text{covar}[\mathbf{d}]^{-1} A\right)^{-1} A^\top \left(\text{covar}[\mathbf{d}]^{-1\top} A \left(A^\top \text{covar}[\mathbf{d}]^{-1} A\right)^{-1\top}\right) \quad . \tag{4.15}$$

Because $\text{covar}[\mathbf{d}]$ and $A^\top \text{covar}[\mathbf{d}]^{-1} A$ are symmetric,

$$... = \left(A^\top \text{covar}[\mathbf{d}]^{-1} A\right)^{-1} A^\top \left(\text{covar}[\mathbf{d}]^{-1} A \left(A^\top \text{covar}[\mathbf{d}]^{-1} A\right)^{-1}\right) \tag{4.16}$$

$$... = \left(A^\top \text{covar}[\mathbf{d}]^{-1} A\right)^{-1} \quad , \tag{4.17}$$

which is similar to the GLLS covariance estimate given in Eq. (2.32).

4.2 Application

We propose three test problems here. The scalar-source problem is a small conceptual model to demonstrate that the delta-method approach is consistent with non-imaging counting statistics. Two five-points models demonstrate the UQ approaches for small, well-posed systems of varying condition number. Future test problems should analyze the scalability of the UQ approaches.

4.2.1 Scalar Source

This trivial reconstruction problem is used to show consistency with non-imaging problems. Consider a point source at a known location; hence $J = 1$. Since A is an

$I \times 1$ matrix and s is a scalar, the ML equation simplifies as

$$A^\top \mathrm{diag}(A)^{-1} A\hat{s} = A^\top \mathrm{diag}(A)^{-1}\hat{d} \tag{4.18}$$

$$\vec{1}^\top A\hat{s} = \vec{1}^\top \hat{d} \tag{4.19}$$

$$\hat{s} = \frac{\sum_i [\hat{d}]_i}{\sum_i [A]_i} \tag{4.20}$$

with $I \times 1$ sensitivity matrix elements

$$b_i \equiv \left.\frac{\partial \hat{s}}{\partial [\hat{d}]_i}\right|_{\hat{s}=s, \hat{d}=As} = \frac{1}{\sum_{i'} [A]_{i'}} \quad . \tag{4.21}$$

The variance of the reconstructed source strength is

$$\mathrm{var}[\mathbf{S}] = B^\top \mathrm{diag}(As) B = \frac{\vec{1}^\top \mathrm{diag}(As)\vec{1}}{\left(\sum_{i'} [A]_{i'}\right)^2} = \frac{s}{\sum_{i'} [A]_{i'}} = \frac{\sum_{i'} [d]_{i'}}{\left(\sum_{i'} [A]_{i'}\right)^2} \quad , \tag{4.22}$$

which we expect since $\mathbf{d}$ is Poisson. If s is unknown, $\hat{b}_i = b_i$, so we can approximate

$$\mathrm{var}[\mathbf{S}] \approx \frac{\hat{s}}{\sum_{i'} [A]_{i'}} \quad . \tag{4.23}$$

Note that this approximation under-estimates var[(]$\mathbf{S}$) when $\hat{s} < s$ and over-estimates var[$\mathbf{S}$] when $\hat{s} > s$.

4.2.2 Well-Conditioned

Here we make our first attempt to derive our system coefficients A from a physical system. Five source points are arranged perpendicular to a FNCA imager. See Fig. 4.1. Each source is assigned a random strength. Detector data $\hat{d}$ is sampled from $\mathbf{d} = \mathrm{Poisson}(As)$. Our goal is to calculate $\hat{S}$ and its associated standard deviation var[$\mathbf{S}$].

The sources are spaced 12 cm apart along the z axis and centered at the origin. A one-dimensional base-19 MURA-patterned mask is centered at $x = 120$ cm; it is 60 cm wide. The mask is oriented as shown in Fig. 4.1 to enhance discrimination in the z direction. A 100-pixel detector is 60 cm behind the mask; it is 42.8 cm wide.

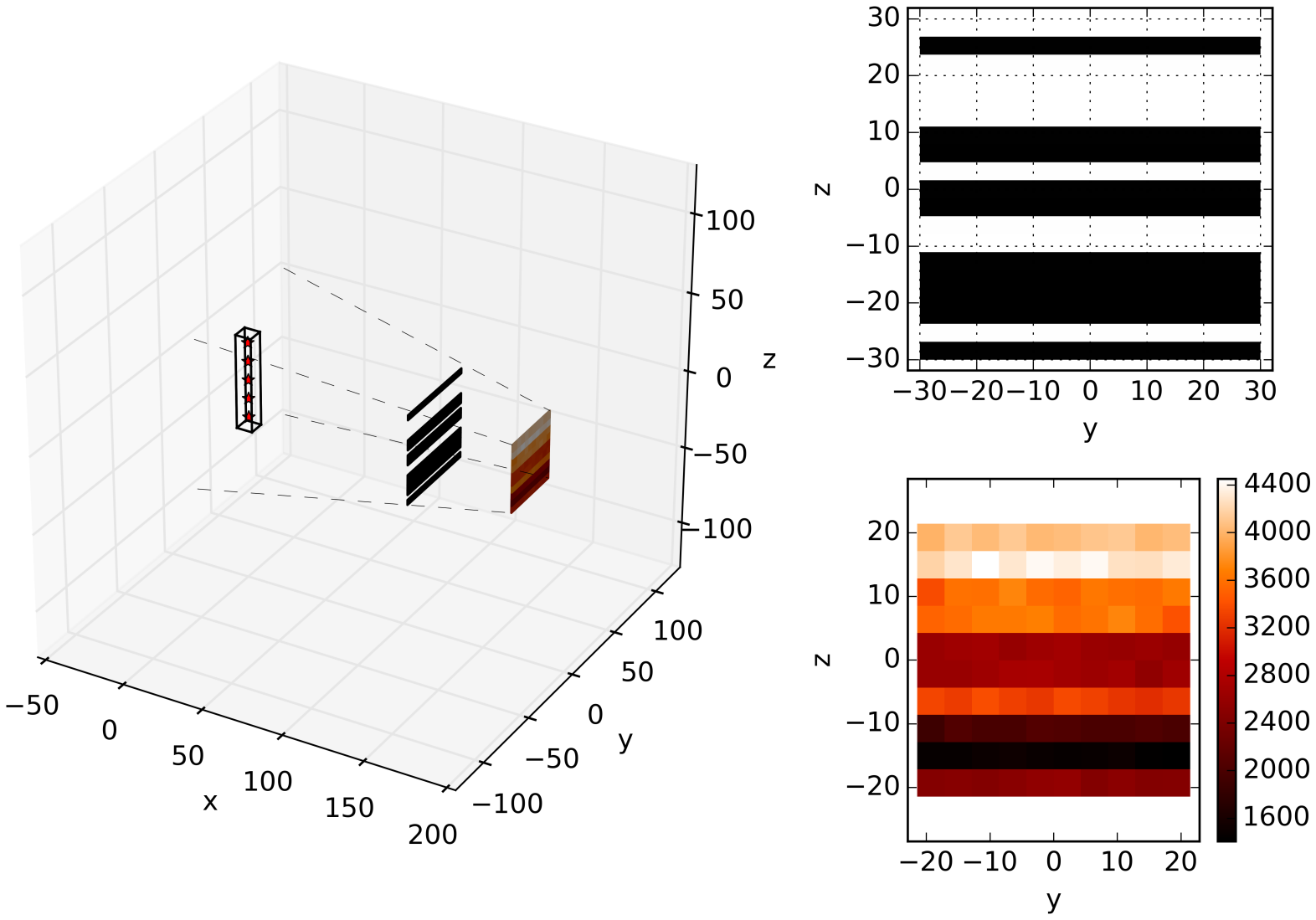


Figure 4.1: The well-conditioned test problem, including the aperture mask pattern (top right) and hit pattern (bottom right). Source positions are marked with red stars. Spatial units are centimeters; black aperture elements are opaque; hit pattern color is number of counts.

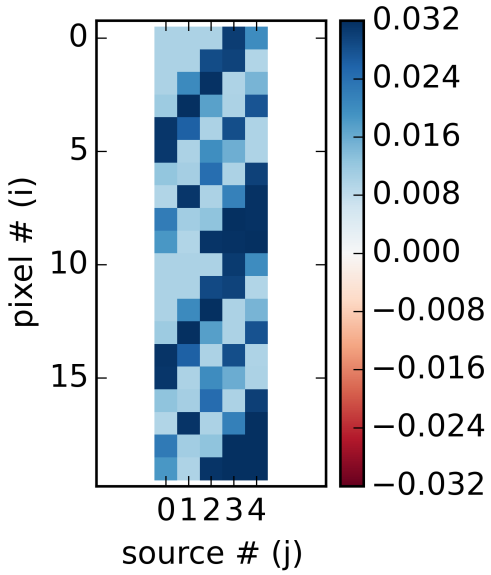


Figure 4.2: The first 20 rows of A in the well-conditioned test problem. Units: count-seconds per emission.

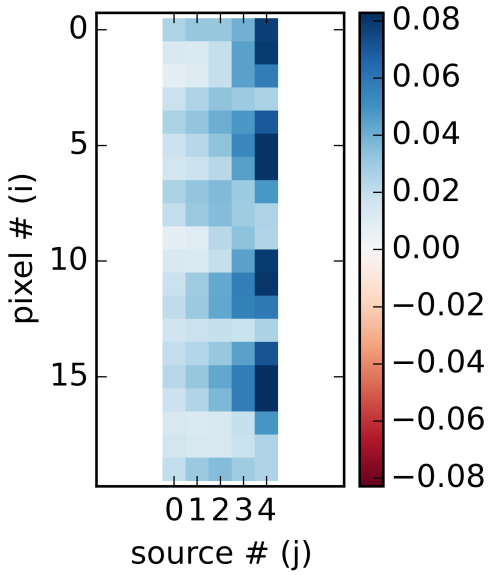


Figure 4.3: The first 20 rows of A in the ill-conditioned test problem. Units: count-seconds per emission.

The forward physics model accounts for mask attenuation, the distance and orientation of the detector pixels, and the intrinsic efficiency of the detector. Opaque mask elements attenuate 68% of incident neutrons; transparent elements do not attenuate. The detector intrinsic efficiency is 20%. Counts are recorded for 3600 seconds.

Each source strength is chosen randomly for each realization of the problem. Strength s_j in neutrons per second is sampled uniformly in the interval $[10^6, 2 \times 10^6]$. This "meta-distribution" of source strengths allows us to evaluate the estimators' behavior for many sources.

The first 20 rows of A for this problem are plotted in Fig. 4.2. The row-sums of A range from 1.695 to 2.214; thus an incorrect $\hat{s}$ is unlikely to accidentally yield a correct $\hat{S}$. Values of κ_{ML} and ρ_{MLEM} depend on s; typically $\kappa_{ML} \approx 45$ and $\rho_{MLEM} \approx 0.975$. Based on ρ_{MLEM}, 2000 MLEM iterations are used to achieve precision on the order of machine ϵ.

We first show that the only error in $\hat{s}$ is from statistical noise in $\hat{d}$. For a source s sampled as described, we can calculate an exact value of d. We introduce error into $\hat{d}$ by rounding each element to the nearest 10, 100, etc. The linear convergence in Fig. 4.4 indicates that other sources of error are trivial.

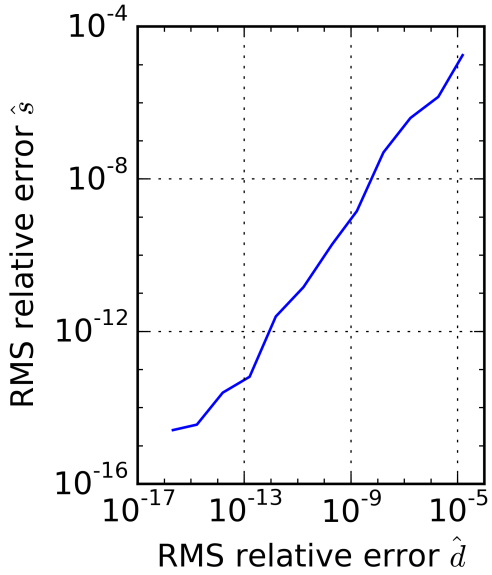


Figure 4.4: Root-mean-square error $\hat{d}$ vs $\hat{s}$ for the well-posed five-points problem.

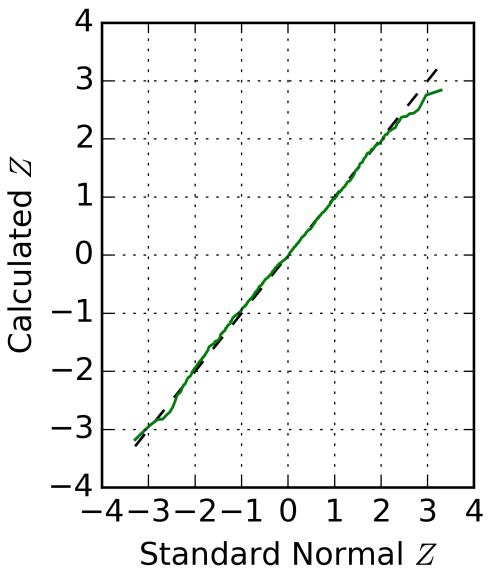


Figure 4.5: A quantile–quantile plot for the well-conditioned five-points problem. Variances estimated using the delta-method approach with $\hat{B}$ (blue) and B (green) overlap.

We then test the delta-method approach using K trials. Consider trial k of K. A true source s_k is sampled from the meta-distribution. From this true source we sample a dataset $\hat{d}_k$ from $\mathbf{d_k}$. We then reconstruct $\hat{s}_k$ and $\hat{S}_k$ and estimates var$[\mathbf{S}_k]$. Estimates of var$[\mathbf{S}_k]$ are calculated using both B (supposing s_k is known) and $\hat{B}$ (supposing s_k is hidden). The delta method is implemented using LDL$^\top$-decomposition.

We then calculate the standard score

$$Z_k \equiv \frac{\hat{S}_k - S_k}{\sqrt{\text{var}[\mathbf{S}_k]}} \quad . \tag{4.24}$$

Figure 4.5 plots the quantiles of Z_k in a QQ plot. We also list

$$\chi^2 \equiv \sum_k Z_k^2 \tag{4.25}$$

in Table 4.1 as a quantitative assessment of the variance estimator.

Table 4.1: Quantitative analysis of the delta-method approach applied to the well-conditioned and ill-conditioned problems.

Problem	Well-posed	Ill-posed	Near constraints
χ^2 degrees of freedom $= K$	1000	1000	1000
χ^2 from B	934.3	961.0	934.1
$p(\chi'^2 > \chi^2 \vert K)$	93.1%	80.8%	93.2%
χ^2 from $\hat{B}$	934.3	960.9	933.9
$p(\chi'^2 > \chi^2 \vert K)$	93.1%	80.8%	93.3%
xyz			

Based on the QQ plots and χ^2, we draw two conclusions: The Poisson error in $\hat{d}$ propagates to an approximately Gaussian error in $\hat{S}$. The delta-method estimates of var$[\mathbf{S}]$ are reasonably accurate and unbiased.

4.2.3 Ill-conditioned

This test problem modifies the well-conditioned test problem of Section 4.2.2 to be ill conditioned. We use this problem to confirm that our UQ methods work well even when the Poisson uncertainty is greatly amplified by reconstruction.

The differences between this problem and the well-conditioned problem are as follows: The five source points are re-arranged perpendicular to the detector. The sources are each 22 cm apart and are centered 145 cm from the mask. The mask pattern is changed to a base-19 2-dimensional MURA.

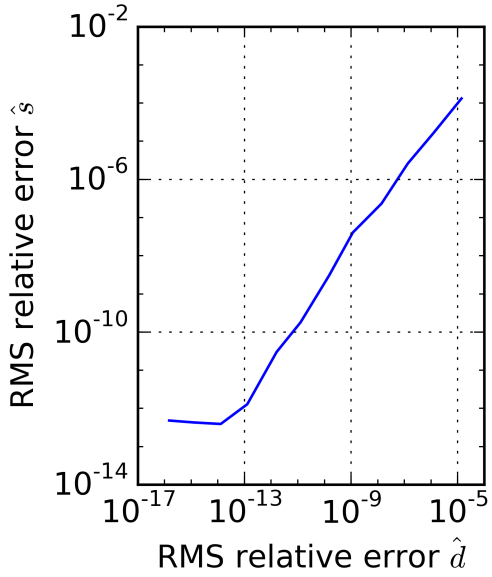


Figure 4.6: Root-mean-square error $\hat{d}$ vs $\hat{s}$ for the ill-posed five-points problem.

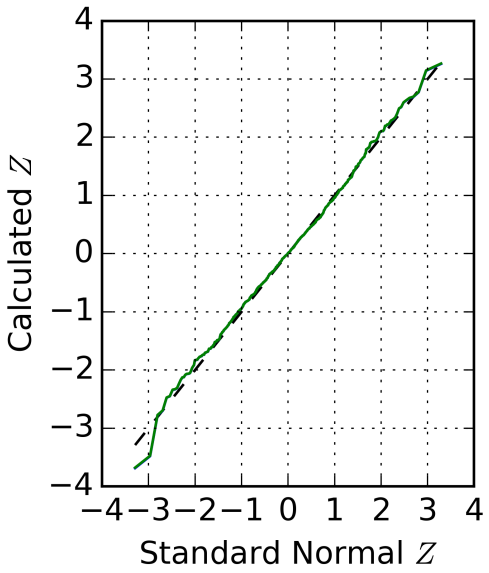


Figure 4.7: A quantile–quantile plot for the ill-conditioned five-points problem. Variances estimated using the delta-method approach with $\hat{B}$ (blue) and B (green) overlap.

The first 20 rows of A for this problem are plotted in Fig. 4.3. Note that the columns of A are nearly multiples of one another—a demonstration of a high condition number. The row-sums of A range from 1.719 to 5.348; thus an incorrect $\hat{s}$ is unlikely to accidentally yield a correct $\hat{S}$. Values of κ_{ML} and ρ_{MLEM} depend on s; typically $\kappa_{ML} \approx 10^4$ and $\rho_{MLEM} \approx 0.99986$. Based on ρ_{MLEM}, 200000 MLEM iterations are used to achieve precision on the order of machine ϵ.

Again, the error in $\hat{s}$ decreases as $\hat{d}$ converges on d. See Fig. 4.6. The QQ plot of Z_k and the overall χ^2 agree with a standard normal distribution. See Fig. 4.7 and Table 4.1. From this we conclude that the distribution of $\hat{S}$ is insensitive to κ_{ML}.

4.2.4 Near Constraints

The well- and ill-conditioned problems both sampled the elements of s_k from Uniform$(10^6, 2 \times 10^6)$. Since the true source strength is far from the nonnegativity constraints, it is unlikely that $\hat{s}$ will be near any of the constraints. This problem is like the well-conditioned problem, but with $\hat{s}$ close to the nonnegativity constraints.

The source strengths are now sampled from a delta function. With 50% probability a point source's intensity is sampled from Uniform$(0, 1e4)$; otherwise it is zero. If all sources are zero the entire source is resampled.

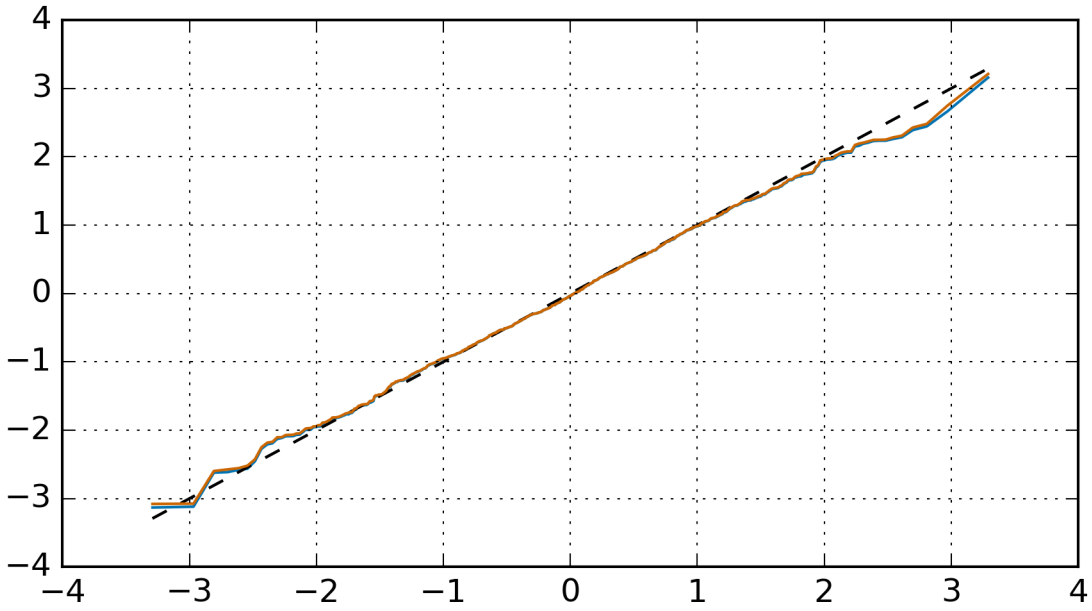


Figure 4.8: A quantile–quantile plot for the five-points problem near the constraints. Variances estimated using the delta-method approach with $\hat{B}$ (blue) and B (orange) overlap.

The QQ plot in Fig. 4.8 show that the distribution of $\hat{S}$ is predicted well. The goodness-of-fit metric is also reasonable; see Table 4.1. However, if we evaluate a goodness of fit metric based on $\hat{s}$ and covar[$\mathbf{s}$], the fit is improbably good: $P(\chi^2 > 3323.6|\text{dof} = 5000) > 99.9\%$. This indicates that the propagation approach does not account for the additional information provided by the inequality constraints. However, this has only a small impact on $\hat{S}$ in this example because the elements of $\hat{s}$ are negatively correlated.

4.3 Conclusions

The delta-method UQ approach accurately estimates var[$\mathbf{S}$] for small problems with varying κ_{ML}. However, it tends to over-estimate the uncertainty when many elements in s are near zero.

These test problems are all fully determined, i.e., $I \geq J$; the approaches in this

chapter will not work when $I < J$. When $I < J$ there may be many equally valid reconstructions, and these approaches do not consider this possibility. Because underdetermined problems are common in FNCA imaging, we must also evaluate incertitude, i.e., epistemic uncertainty.

CHAPTER V

Epistemic Uncertainty

The ML equation is often underdetermined. Rather than having a unique maximum, the solution space may have a region that maximizes likelihood. Any s in that region will satisfy the ML equation. Several authors have provided notes on propagating epistemic incertitude in the absence of aleatoric uncertainty, e.g., [65, 66, 67].

Qualitative imaging addresses solution underdeterminancy in several ways. The reconstruction domain can be coarsened in such a way that the ML solution is unique. Qualitative imaging objectives and algorithms typically converge on a solution $\hat{s}$ that smoothes the source, i.e., regularization. These solutions are the most reasonable for clinical interpretation.

Quantitative imaging must address underdeterminancy in the opposite way. Smoothing regularization is not appropriate because the $\hat{S}$ extrema have many zero elements of s. The relevant solutions will have compact non-zero regions, not broad non-zero regions created by smoothing.

In this chapter we analyze underdetermined reconstruction using subspace methods. We then show that linear programming approaches can quantify uncertainty in $\hat{S}$ when statistical noise is neglegible. First, we demonstrate underdeterminancy problem in QI using a simple example.

5.1 Illustration

The single-pixel problem is a simple conceptual model. Suppose we have two sources of strength s_0 and s_1. We observe these sources simultaneously with a non-imaging detector ($I = 1$). The reponse matrix

$$A = \begin{bmatrix} A_0 & A_1 \end{bmatrix} \tag{5.1}$$

is strictly positive. We observe $\hat{d} > 0$ counts in the detector with trivial statistical error.

Arbitrarily set $A_0 = 100$, $A_1 = 200$, and $s_0 = s_1 = 1$. Figure 5.1 plots the Poisson likelihood of $\hat{d} = 300$ for hypothetical values of s_0 and s_1. The surface has no unique maximum, but a stationary curve connecting the intercepts $s_0 = \hat{d}/A_0$ and $s_1 = \hat{d}/A_1$. All first-quadrant points on this curve satisfy the ML equation. It is impossible to map $\hat{d}$ to a unique $\hat{s}$ without some form of regularization. Because statistical error is trivial, our true s lies on that curve.

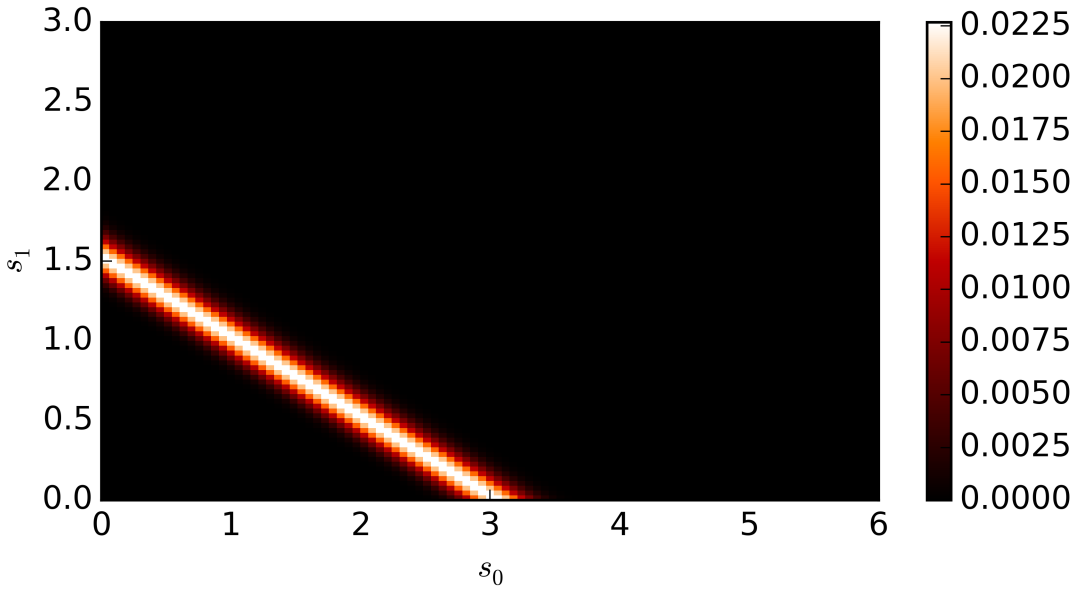


Figure 5.1: Likelihood of $\hat{d} = 300$ as a function of s_0 and s_1 in the single-pixel test problem.

We can decompose s as the weighted sum of an influential vector and a noninfluential vector:

$$s = \begin{bmatrix} \hat{d}/A_1 \\ \hat{d}/A_0 \end{bmatrix} c_{\mathcal{I}} + \begin{bmatrix} \hat{d}/A_0 \\ -\hat{d}/A_1 \end{bmatrix} c_{\mathcal{N}} \quad , \tag{5.2}$$

where $c_{\mathcal{I}}$ and $c_{\mathcal{N}}$ are scalar. The influential vector is perpendicular to the ML solution curve. Both $c_{\mathcal{I}}$ and $c_{\mathcal{N}}$ affect s, but only changes in $c_{\mathcal{I}}$ affect d. When considering $c_{\mathcal{N}}$, the reconstruction can only rule out values violate the nonnegativity constraint on s.

We conclude that $\hat{S} \equiv \hat{s}_0 + \hat{s}_1$ is bounded

$$\min(\frac{d}{A_0}, \frac{d}{A_1}) \leq \hat{S} \leq \max(\frac{d}{A_0}, \frac{d}{A_1}) \quad , \tag{5.3}$$

but not uniquely determined—despite the absence of statistical uncertainty.

5.2 Noninfluential Subspaces

We begin by defining the influential and noninfluential subspaces of s [68]. Again consider a system with response matrix A with $I < J$. In most physical systems A is full rank (but perhaps ill-conditioned) because of small differences in response even among neighboring voxels—so $R \equiv \mathrm{rk}(A) = I$. We can then divide $\mathbb{R}^{\mathrm{J}}$ into an influential subspace $\mathcal{I}$ and a noninfluential subspace $\mathcal{N}$.

The influential subspace $\mathcal{I}$ corresponds to the rowspace of A. Equivalently, $\mathcal{I}$ is the columnspace or range of $A^\top$. If we represent $\mathcal{I}$ as a $J \times R$ matrix of orthogonal basis vectors, we can define define a vector of coefficients $c_\mathcal{I} \in \mathbb{R}^{\mathrm{R}}$ such that $A\mathcal{I}c_\mathcal{I} = As$ for any $s \in \mathbb{R}^{\mathrm{J}}$.

The remaining noninfluential subspace $\mathcal{N}$ corresponds to the nullspace of A. Represent $\mathcal{N}$ as a $J \times (J-R)$ matrix of orthogonal basis vectors. By definition, $A\mathcal{N}c_\mathcal{N} = \vec{0}$ for all $c_\mathcal{N} \in \mathbb{R}^{\mathrm{J-R}}$. To summarize in block notation,

$$s = \begin{bmatrix} \mathcal{I} & \mathcal{N} \end{bmatrix} \begin{bmatrix} c_\mathcal{I} \\ c_\mathcal{N} \end{bmatrix} \quad \text{, but} \tag{5.4}$$

$$As = A\mathcal{I}c_\mathcal{I} \quad . \tag{5.5}$$

Smith recommends rank-revealing QR decomposition for determining $\mathcal{I}$ and singular value decomposition (SVD) for determining $\mathcal{N}$ [68]. We rely on SciPy (version 0.15.1) implementations of these decompositions.

We make two notes. First, the positivity constraint on s effectively constrains $c_\mathcal{N}$. In physical systems with penetration and scatter, all elements of A are (at least slightly) positive. The basis vectors in $\mathcal{N}$ then must have mixed sign, i.e.,

$$\min_k([\mathcal{N}]_{k,j}) \ \max_k([\mathcal{N}]_{k,j}) < 0 \ \forall \ j \in (J-R) \quad , \tag{5.6}$$

and any linear combination of the basis vectors must have mixed sign (except $c_\mathcal{N} = \vec{0}$).

Therefore $\|c_{\mathcal{N}}\| \to \infty$ would preclude $[s]_j \geq 0 \forall j$, and our solution is bounded.

Second, in the underdetermined system, ML solutions also satisfy $A\hat{s} = \hat{d}$. This is rarely true in the overdetermined system; the ML equation represents a compromise among the data. As the subspace analysis of s shows, $A\hat{s} = \hat{d}$ defines a $\hat{s}$ domain with $J - R$ degrees of freedom. Any $\hat{s}$ satisfying $A\hat{s} = \hat{d}$ also solves the ML equation, and linear techniques can be used to evaluate $\hat{s}$.

Two exceptions exist: First, some members of $\hat{s}$ can be overdetermined while others are underdetermined (in which case A is not full rank). This is unlikely, but it reminds us that the fully determined problem $A\mathcal{I}\hat{c}_{\mathcal{I}} = \hat{d}$ may be ill-conditioned. Second, statistical or numerical error may yield a $\hat{d}$ that has no corresponding non-negative $\hat{s}$ solutions. Except for these cases, linear programming methods to explore $\mathcal{N}$ can bound $\hat{S}$.

5.3 The Simplex Method

Given noise-free data $\hat{d}$ and a full rank matrix A with $I < J$, we can bound $\hat{S}$:

$$\hat{S} \leq \max_{\hat{s}} \sum_j [\hat{s}]_j \tag{5.7}$$

with $\hat{s}$ subject to constraints

$$A\hat{s} = \hat{d} \quad \text{and} \tag{5.8}$$

$$[s]_j \geq 0 \ \forall \ j \quad . \tag{5.9}$$

We can also bound $\hat{S}$ from below:

$$\hat{S} \geq \min_{\hat{s}} \sum_j [\hat{s}]_j \tag{5.10}$$

subject to the same constraints. These are linear programming problems that can be solved using the simplex method [69].

The simplex method is based on Gaussian elimination-like manipulations of an augmented system matrix. In its first phase, simplex uses "artificial" variables to satisfy equality constraints (Eq. (5.8)). In its second phase, simplex determines which inequality constraints (Eq. (5.9)) limit $\hat{S}$. We rely on the SciPy (version 0.15.1) implementation of simplex for now.

5.4 Application

We propose two underdetermined reconstruction problems to test our UQ approaches. First, a single-pixel problem illustrates the ML region and provides an analytical benchmark. Second, an underdetermined problem of arbitrary size is given to demonstrate the scalability of the UQ approaches and the interactions between the $A\hat{s} = \hat{d}$ equations and the nonnegative-source constraints. These problems omit statistical uncertainty on $\hat{d}$ to isolate the underdetermined-reconstruction uncertainty.

5.4.1 Single Pixel

We can first consider the single-pixel illustration of Section 5.1. If we calculate $\mathcal{I}$ using QR, we get

$$\mathcal{I} = \begin{bmatrix} -0.447 \\ -0.894 \end{bmatrix} \propto \begin{bmatrix} \hat{d}/A_1 \\ \hat{d}/A_0 \end{bmatrix} \quad . \tag{5.11}$$

If we calculate $\mathcal{N}$ using SVD, we get

$$\mathcal{N} = \begin{bmatrix} -0.894 \\ 0.447 \end{bmatrix} \propto \begin{bmatrix} \hat{d}/A_0 \\ -\hat{d}/A_1 \end{bmatrix} \quad . \tag{5.12}$$

The simplex method returns the expected bounds $\hat{S} \in [1.5, 3]$.

5.4.2 Half-Shade

The half-shade test problem mimics an idealized one-dimensional aperture system with I pixels. We prescribe $\sim 2\times$ as many source locations as detector pixels: $J = 2I - 1$. Elements of A are prescribed

$$[A]_{i,j} = \begin{cases} 2(j+1) & 2i < j+1 \\ 1(j+1) & 2i = j+1 \\ 10^{-6}(j+1) & 2i > j+1 \end{cases} \tag{5.13}$$

with elements indexed from zero. (Each row is multiplied by $j + 1$ so different arrangements of $\hat{s}$ are less likely to accidentally yield the same $\hat{S}$.) For example, if

$I = 3$,

$$A = \begin{bmatrix} 2 & 4 & 6 & 8 & 10 \\ 10^{-6} & 2 & 6 & 8 & 10 \\ 10^{-6} & 2 \times 10^{-6} & 3 \times 10^{-6} & 4 & 10 \end{bmatrix} \tag{5.14}$$

The response matrix A is full rank for any $I > 0$. When $[s]_j$ are sampled uniformly in $[1, 2]$, $\hat{S}$ is not unique but is bounded. In this case we can observe the accuracy and central processing unit (CPU) cost as $I \to \infty$. When $[s]_j = \delta_{j=k}$ for some k in $[0, J)$, the $s \geq 0$ constraints force a unique solution. In this case we can observe the accuracy and robustness of the simplex implementation when many $s \geq 0$ constraints are active.

First consider $I = 15$ with $s_j = \vec{1}$. The true $S = 29$ falls within the simplex bounds $\hat{S} \in [28.52, 29.48]$. With $I = 20$, the simplex implementation is unable to find the extrema. The code reports that the simplex algorithm is unable to satisfy the equality constraints in phase 1. This does not match our expectations and warrants further study.

Second consider $I = 10$ with $s_j = \delta_{j=4}$. The true $S = 1$ is at the edge of the simplex bounds $\hat{S} \in [1, 5]$. This matches our understanding that the extrema occur in solutions in which many s_j are bounded. To evaluate the impact of statistical and numerical errors, we perturb $[\hat{d}]_3$ from 5×10^{-6} to 4.9999×10^{-6}. The code again reports that the simplex algorithm is unable to satisfy the equality constraints in phase 1. This matches our expectation that data errors in $\hat{d}$ may cause $A\hat{s} = \hat{d}$ to have no nonnegative solutions.

5.5 Conclusions

The single-pixel test problem demonstrates that underdetermined reconstruction problems can yield uncertain estimates, even in the absence of statistical noise. Straightforward application of the simplex algorithm can quantify the assicated uncertainty in $\hat{S}$. In many cases, however, numerical or statistical errors will make all linear programming solutions infeasible. For robustness we must identify a method that estimates overall uncertainty in underdetermined problems with statistical errors.

CHAPTER VI

Frequentist Approach

Solving for the frequentist confidence interval is a convex optimization problem. This is true when the parameters are unknown source strengths and the QOI is their sum; it is not true for all inverse problems.

We begin with background on confidence regions and intervals and on convex optimization. We then introduce our confidence interval formulation, describe improvements that make the convex optimization solver converge reliably, and calculate the derivatives necessary to implement the solver. We conclude with demonstration and analyses of the solver, including a scaling study.

6.1 Background

A frequentist analysis accepts or rejects a hypothesized model by scoring its fit with the data—without regard for alternative models. Here "model" is defined broadly, and in QI it refers to a specific forward model and parameters. One contrasts frequentist analysis with Bayesian analysis, in which the posterior normalization effects a competition among the models.

Because a frequentist scores models independent of one another, calculating $\mathbf{CI}_f$ is a search problem, rather than a sampling problem. This attribute is computationally convenient. However, the frequentist approach should be used with caution. The solution may be rather conservative in the sense that the inequality of Eq. (6.5) may be far from equality.

In Section 6.1.1 we define a χ^2 residual to score the model–data fit. Determining $\mathbf{CI}_f$ is then an optimization problem constrained by a prescribed threshold on χ^2. Sections 6.1.2 and 6.1.3 describe an applicable constrained-optimization solver. Section 6.1.4 shows that the convex optimization solver is appropriate for our FNCA emission QI problem.

6.1.1 Confidence Intervals Based on a Chi-squared Threshold

A frequentist analysis of the forward model could construct a "confidence region" in the parameter space. A 95% confidence region is defined to capture the true parameter for at least 95% of sampled datasets. When data are repeatedly sampled from some true, hidden Gaussian distribution $\mathbf{d}$, we expect the sampled χ^2 statistic

$$\hat{\chi}^2 \equiv (\hat{d} - \text{mean}[\mathbf{d}])^\top \text{covar}[\mathbf{d}]^{-1} (\hat{d} - \text{mean}[\mathbf{d}]) \tag{6.1}$$

to follow a chi-squared distribution with I degrees of freedom. (This is a general definition of the chi-squared distribution, not specific to QI.) In 95% of samples, $\hat{\chi}^2 \leq \chi^2_{I;0.95}$, where $\chi^2_{I;0.95}$ can be calculated from the chi-squared distribution's percent point function. Therefore a 95% confidence region

$$\mathbf{CR}_f \equiv \left\{ x \;\middle|\; \chi^2(x) \leq \chi^2_{I;0.95} \bigcap x \geq 0 \right\} \quad \text{with} \tag{6.2}$$

$$\chi^2(x) \equiv \left(\hat{d} - \text{mean}[\mathbf{d}; x] \right)^\top \text{covar}[\mathbf{d}; x]^{-1} \left(\hat{d} - \text{mean}[\mathbf{d}; x] \right) \tag{6.3}$$

will capture x_{true} for 95% of samples of $\hat{d}$. (Here we use x to represent the source distribution s, not the spatial variable. The nonnegativity constraint is specific to the emission QI problem.)

Hastie and colleagues mention a similar confidence-interval approach (using an upper threshold on χ^2); see Eq. 8.19 of [61].

The QOI confidence interval is then a constrained optimization problem:

$$\mathbf{CI}_f \equiv \left\{ S \in \left[\min_{x \in \mathbf{CR}_f} S(x), \max_{x \in \mathbf{CR}_f} S(x) \right] \right\} \quad \text{since} \tag{6.4}$$

$$p \left(\min_{x \in \mathbf{CR}_f} S(x) \leq S(x_{\text{true}}) \leq \max_{x \in \mathbf{CR}_f} S(x) \right) \geq 0.95 \tag{6.5}$$

in repeated realizations of $\mathbf{CR}_f$ for every possible x_{true}. Here the QOI function $S(x)$ refers to the total source intensity; see Eq. (2.4).

This constrained optimization approach is more robust than a "generalized confidence interval" [36]. With a generalized confidence interval the inequality of Eq. (6.5) is satisfied only under repeated sampling of $\hat{d}$ *among many situations with various* x_{true}. The generalized confidence interval is less conservative and would likely be sim-

ilar to a Bayesian credible interval. However, the differences shrink when ζ is small, since situations in which x_{true} has many zero elements are the situations in which a generalized confidence interval would violate Eq. (6.5).

6.1.2 Unconstrained Convex Optimization: Newton's Method

Before we consider the constrained optimization problem, let us consider the unconstrained optimization problem:

$$\text{seek } F(x^*) = \min_{x \in \mathbb{R}^J} F(x) \quad . \tag{6.6}$$

This optimization problem is convex if F is a convex function on $\mathbb{R}^J$. A function f_j is convex if

$$F(\alpha x + \beta x') \leq \alpha F(x) + \beta F(x') \; \forall \; x, x' \in \mathcal{D} \quad , \tag{6.7}$$

where the weights $0 \leq \alpha, \beta \leq 1$ sum to unity. Equivalently, F is convex on a domain if its Hessian $\nabla^2 F$ is positive semi-definite [70] everywhere in the domain:

$$y^\top \; \nabla^2 F\big|_x \; y \geq 0 \; \forall \; y \in \mathbb{R}^J \bigcap x \in \mathcal{D} \quad . \tag{6.8}$$

Convexity is useful because it implies that any local minimum $F(x^*)$ is the global minimum, so we can use local minimization techniques to find the global minimum. (The minimum may exist along a continuum, so x^* may not be unique.)

Newton's method solves the unconstrained convex optimization problem using a Taylor series expansion of the objective [9]. Approximate

$$F(x^{(k)} + s) = F(x^{(k)}) + s^\top \nabla F\big|_{x^{(k)}} + s^\top \nabla^2 F\big|_{x^{(k)}} s + \mathcal{O}\left(\|s\|_3^3\right) \quad . \tag{6.9}$$

The step minimizing the second-order expansion satisfies

$$-\nabla^2 F\big|_{x^{(k)}} \; s^{(k)} = \nabla F\big|_{x^{(k)}} \quad . \tag{6.10}$$

This defines a sequence of steps $x^{(k+1)} = x^{(k)} + s^{(k)}$, $k = 0, 1, 2, ..., K$. Each step requires solving a $J \times J$ system of linear equations. The steps converge rapidly on x^* if the third derivative of F at $x^{(k)}$ is small (with respect to the second derivative and step size). This implies that the Newton steps converge rapidly when $x^{(k)}$ is in some Rapid Convergence Region (RCR) around x^*.

If the $J \times J$ linear solve is much more expensive than evaluating F, it is reasonable to rescale each step using a line search. By rescaling we can guarantee that each step is decreasing, even if it is not converging rapidly. To do this we use bisection method to estimate the real scalar

$$\alpha \equiv \underset{\alpha' \in (0,\infty)}{\operatorname{argmin}} F(x^{(k)} + \alpha' s^{(k)}) \quad . \tag{6.11}$$

If F is convex, then α is the unique root of

$$s^{(k)\top} \nabla F\big|_{x^{(k)}+\alpha s} = 0 \quad . \tag{6.12}$$

We will use α in Section 6.2.2 to indicate the Newton step quality: If $\alpha \approx 1$ then the second-order expansion appears to be a reasonable approximation of F, so $x^{(k)}$ appears to be in RCR.

6.1.3 Constrained Convex Optimization: Logarithmic Barriers

A constrained global minimization problem

$$\text{seek } f_0(x^*) \equiv \min_{x \in \mathbf{CR}_f} f_0(x) \quad \text{with} \tag{6.13}$$

$$\mathbf{CR}_f \equiv \{x \mid f_j(x) \leq 0,\ j = 1, 2, ...\} \tag{6.14}$$

is a convex optimization problem if all f_j (including f_0) are convex functions on $\mathcal{D}$, the domain of x [9]. Again convexity is a useful property, since it implies that local minimization techniques are guaranteed to find the global minimum.

The optimization problems used to determine a confidence interval in Section 6.1.1 are convex. The nonnegativity constraints can be expressed as

$$f_j(x) \ \equiv - [x]_j \leq 0 \ \ \forall\ j \in [1, J] \quad . \tag{6.15}$$

The functions are linear and therefore convex. The minimization objective functions will be

$$f_0 \equiv \left\{ \begin{matrix} S, & \text{to seek } \min_{x \in \mathbf{CR}_f} S \\ -S, & \text{to seek } \max_{x \in \mathbf{CR}_f} S \end{matrix} \right\} \quad . \tag{6.16}$$

In both cases f_0 is linear and therefore convex. In Section 6.1.4 we will show that the

"functional" constraint

$$f_{J+1}(x) \equiv \chi^2(x) - \chi^2_{I;0.95} < 0 \tag{6.17}$$

is convex for emission QI (although it is not convex for other inverse problems). In total we have $J+1$ inequality constraints.

Boyd gives an algorithm based on logarithmic barrier functions and Newton's method to solve the convex optimization problem [9]. Similar path-following prescriptions are given elsewhere for convex optimization [71, 72] and linear programming [73]. The barrier function

$$\phi(x) \equiv \sum_{j=1}^{J+1} -\ln\left(-f_j(x)\right) \tag{6.18}$$

approaches $+\infty$ as x approaches any of the constraints and is undefined where the constraints are violated. We can therefore rewrite our constrained optimization problem as a family of *un*constrained optimization problems seeking

$$x^*(\mu) \equiv \underset{x}{\operatorname{argmin}}\, F(x;\mu) \quad \text{, with} \tag{6.19}$$

$$F(x;\mu) \equiv f_0(x) + \mu\phi(x) \quad . \tag{6.20}$$

Given an initial feasible solution $x^{(0)} \in \mathbf{CR}_f$, any problem in this family could be solved using Newton's method.

The scalar $\mu \in (0,1]$ adjusts the smoothness of the objective. When $\mu = 1$, F is very smooth, so the RCR is large and the optimization problem is easy. As $\mu \to 0$ the RCR shifts and shrinks, but optimization problem becomes more accurate:

$$\lim_{\mu\to 0} F(x;\mu) = f_0(x) \quad . \tag{6.21}$$

We therefore need to gradually decrease μ in successive steps to converge as rapidly as possible. A geometric series

$$\mu^{(k+1)} = \gamma\mu^{(k)} \quad , \tag{6.22}$$

with $\mu^{(0)} = 1$ is typically used; $\gamma \approx 0.95$ appears effective for our application. The

Newton step is then

$$s^{(k)} = -\left(\nabla_x^2 F(x;\mu^{(k)})\big|_{x^{(k)}}\right)^{-1} \nabla_x F(x;\mu^{(k)})\big|_{x^{(k)}} \tag{6.23}$$

for steps $k = 1, 2, ..., K$.

We conclude this section with a brief on the Karush–Kuhn–Tucker [7, 8, 9] (KKT) conditions. The KKT conditions are necessary and sufficient conditions for the solution of a convex optimization problem. We will use them in Section 6.2.1 to create a convergence criterion for our constrained optimization problem.

At the optimum x^*, constraint j may be active ($f_j(x^*) = 0$) or inactive ($f_j(x^*) < 0$). The KKT conditions tells us that the optimal x^* must satisfy stationarity

$$\nabla f_0|_{x^*} + \sum_{j=1}^{J+1} \lambda_j \nabla f_j|_{x^*} = \vec{0} \quad , \tag{6.24}$$

with "KKT multipliers" constrained to

$$\lambda_j \geq 0 \tag{6.25}$$

for dual feasibility. "Complementary slackness" indicates that

$$\lambda_j f_j(x^*) = 0 \tag{6.26}$$

for every constraint. In other words, if objective gradient at the solution is always zero—unless the solution cannot move further down the gradient because it is "blocked" squarely by active constraints. The f_0 gradient at x^* can therefore be expressed as a linear combination of the active constraint gradients. We will use these properties of the solution to solve for a bounding estimator on the extrema and to show it converges.

6.1.4 Convexity of Chi-squared

Here we show that the Hessian matrix $\nabla^2\chi^2$—the second derivatives of χ^2 with respect to x—is always positive semi-definite when $Ax > 0$. In other words, we show

$$x^\top \ \nabla^2\chi^2\big|_x \ x \geq 0 \ \forall \ x \in \mathbb{R}^{\mathrm{J}} \bigcap x \in \{x|\,[Ax]_i > 0 \forall i\} \quad . \tag{6.27}$$

This indicates that χ^2 is a convex function of x in that domain. This agrees with previous findings [74] and matches similar analysis that the log-likelihood function is

concave [42].

Using covariance model I, our χ^2 statistic is

$$\chi^2(x) \equiv (Ax-d)^\top \mathrm{diag}(Ax)^{-1}(Ax-d) = \sum_i \frac{[Ax-d]_i^2}{[Ax]_i} \quad . \tag{6.28}$$

For convenience define I-vectors

$$e \equiv Ax \quad \text{and} \tag{6.29}$$

$$r \equiv Ax - d \quad , \tag{6.30}$$

the expectation and residual, respectively. Hence

$$\chi^2(x) \equiv r^\top \mathrm{diag}(e)^{-1} r = \sum_i \frac{[r]_i^2}{[e]_i} \quad . \tag{6.31}$$

The expectation and residual have derivatives

$$\frac{\partial e}{\partial [x]_j} = \frac{\partial r}{\partial [x]_j} = [A]_{:,j} \quad , \tag{6.32}$$

where $[A]_{:,j}$ is column j of A. For emission QI, all elements of A are at least slightly positive, so all elements of e are positive when any element of x is positive. We use this to restrict the domain of x: $x \in \{x | [e]_i > 0 \forall i\}$.

Differentiating χ^2 with respect to $[x]_j$ gives

$$\frac{\partial \chi^2}{\partial [x]_j} = \sum_i \frac{2[r]_i [A]_{i,j}}{[e]_i} - \sum_i \frac{[r]_i^2 [A]_{i,j}}{[e]_i^2} \tag{6.33}$$

$$\ldots = r^\top \left(2\mathrm{diag}(e)^{-1} - \mathrm{diag}(r)\mathrm{diag}(e)^{-2}\right) [A]_{:,j} \quad , \tag{6.34}$$

so the gradient with respect to x is a J-vector

$$\nabla \chi^2 = A^\top \left(2\mathrm{diag}(e)^{-1} - \mathrm{diag}(r)\mathrm{diag}(e)^{-2}\right) r \quad . \tag{6.35}$$

Differentiating again gives

$$\frac{\partial}{\partial [x]_{j'}} \frac{\partial \chi^2}{\partial [x]_j} = \sum_i \frac{2 [A]_{i,j'} [A]_{i,j}}{[e]_i} - \sum_i \frac{2 [r]_i [A]_{i,j} [A]_{i,j'}}{[e]_i^2} \dots \tag{6.36}$$

$$\dots - \sum_i \frac{2 [r]_i [A]_{i,j} [A]_{i,j'}}{[e]_i^2} + \sum_i \frac{2 [r]_i^2 [A]_{i,j} [A]_{i,j'}}{[e]_i^3} \tag{6.37}$$

$$\dots = \left[A^\top\right]_{j',:} \left(2\text{diag}(e)^{-1} - 4\text{diag}(r)\text{diag}(e)^{-2} + 2\text{diag}(r)^2\text{diag}(e)^{-3}\right) [A]_{:,j} \quad , \tag{6.38}$$

so the Hessian is

$$\nabla^2 \chi^2 = 2A^\top \left(\text{diag}(e)^{-1} - 2\text{diag}(r)\text{diag}(e)^{-2} + \text{diag}(r)^2\text{diag}(e)^{-3}\right) A \tag{6.39}$$

$$... = 2A^\top \text{diag}(e)^{-1} \left(\mathbf{I} - \text{diag}(r)\text{diag}(e)^{-1}\right)^2 A \tag{6.40}$$

$$... = \Upsilon^\top \Upsilon \quad \text{, with} \tag{6.41}$$

$$\Upsilon \equiv \sqrt{2}\text{diag}(e)^{-1/2} \left(\mathbf{I} - \text{diag}(r)\text{diag}(e)^{-1}\right) A \quad . \tag{6.42}$$

The inverse-square-root can be applied elementally to e. Because all elements of e are positive and all elements of A and r are real, $\Upsilon \in \mathbb{R}^{\mathrm{I} \times \mathrm{J}}$.

Showing that the Hessian can be written in this form implies that it must be positive semi-definite: For any $x \in \mathbb{R}^{\mathrm{J}}$,

$$x^\top (\nabla^2 \chi^2) x = (\Upsilon x)^\top \Upsilon x = \sum_i [\Upsilon x]_i^2 \geq 0 \quad . \tag{6.43}$$

Therefore determining the confidence interval is a convex optimization problem.

6.2 Theory

Here we derive an estimator based on the KKT conditions that inexpensively bounds S. The estimator can function as a convergence metric, stopping criterion, and conservative estimate of the confidence interval. We also improve the reliability

of the barrier method by modifying the barrier scale prescription and initialization procedure. These measures make the solver reliable enough for non-expert use. We then derive the appropriate derivatives of the objective functions and constraints to implement the solver for the QI problem.

6.2.1 Convergence Criterion

For optimization problems, an ideal convergence metric bounds the error on the solver estimate of $F(x^*)$. In this section we use linear programming to derive the relative error bound

$$w^- \left(x^{(k)}\right) \leq \left| \frac{S(x^{(k)}) - \min_{x \in \mathbf{CR}_f} S(x)}{\min_{x \in \mathbf{CR}_f} S(x)} \right| \tag{6.44}$$

for the minimization search and

$$w^+ \left(x^{(k)}\right) \leq \left| \frac{S(x^{(k)}) - \max_{x \in \mathbf{CR}_f} S(x)}{\max_{x \in \mathbf{CR}_f} S(x)} \right| \tag{6.45}$$

for the maximization search.

To do this we first bound $S(x^*)$. At later iterations the convex optimization solver has a solution $x^{(k)}$ with

$$x^{(k)} \in \mathbf{CR}_f \quad \text{and} \tag{6.46}$$

$$\lim_{k \to \infty} S(x^{(k)}) = S(x^*) \quad . \tag{6.47}$$

We would like an estimator that uses $x^{(k)}$ to bound the extremum, e.g.,

$$\tilde{S}^- \left(x^{(k)}\right) \leq \min_{x \in \mathbf{CR}_f} S(x) \tag{6.48}$$

that also converges

$$\lim_{x \to x^*} \tilde{S}^-(x) = S(x^*) \tag{6.49}$$

as we minimize $f_0 = S$. An analogous

$$\tilde{S}^+ \left(x^{(k)}\right) \geq \max_{x \in \mathbf{CR}_f} S(x) \tag{6.50}$$

is useful when maximizing S. We can create these estimators by approximating the convex optimization problem as a linear optimization problem.

We begin by taking a Taylor series expansion of χ^2 about $x^{(k)}$. Equation (6.7) implies that, since $\chi^2(s)$ is a convex function,

$$\chi^2(hx + (1-h)x^{(k)}) \leq h\chi^2(x) + (1-h)\chi^2(x^{(k)}) \quad , \tag{6.51}$$

with $0 \leq h \leq 1$. (For brevity, the domain restriction $[Ax]_i \geq 0 \ \forall \ i$ is implied.) Algebra gives

$$\frac{\chi^2(x^{(k)} + h(x - x^{(k)})) - \chi^2(x^{(k)})}{h} \leq \chi^2(x) - \chi^2(x^{(k)}) \quad , \tag{6.52}$$

and taking the limit $h \to 0$ leaves

$$(x - x^{(k)})^\top \ \nabla\chi^2\big|_{x^{(k)}} \leq \chi^2(x) - \chi^2(x^{(k)}) \quad \text{or} \tag{6.53}$$

$$\chi^2(x) \geq \chi^2(x^{(k)}) + (x - x^{(k)})^\top \ \nabla\chi^2\big|_{x^{(k)}} \quad . \tag{6.54}$$

Therefore the order-1 Taylor series expansion of χ^2 at $x^{(k)}$ bounds χ^2 from below.

A lower bound on χ^2 defines an outer bound on the confidence region:

$$\mathbf{CR}_f \subset \widetilde{\mathbf{CR}_f} \equiv \left\{ x \ \middle| \ \chi^2(x^{(k)}) + (x - x^{(k)})^\top \ \nabla\chi^2\big|_{x^{(k)}} \leq \chi^2_{I;0.95} \bigcap x \geq 0 \right\} \quad , \tag{6.55}$$

which in turn bounds the minimum

$$\tilde{S}^- \equiv \min_{x \in \widetilde{\mathbf{CR}_f}} S(x) \leq \min_{x \in \mathbf{CR}_f} S(x) \tag{6.56}$$

located at

$$\tilde{x}^- \equiv \operatorname*{argmin}_{x \in \widetilde{\mathbf{CR}_f}} S(x) \quad . \tag{6.57}$$

The omit the dependence of $\widetilde{\mathbf{CR}_f}$, $\tilde{S}^-$, and $\tilde{x}^-$ on $x^{(k)}$ for brevity.

Here we use linear programming to solve for $\tilde{S}^-$. Evaluating $\tilde{S}^-$ is a linear programming problem with nonnegativity constraints. Its sole functional constraint is

re-arranged from Eq. (6.55):

$$\tilde{f}_{J+1}(x) \equiv \nabla\chi^2\big|^{\top}_{x^{(k)}}\, x - \left(\chi^2_{I;0.95} - \chi^2(x^{(k)}) + \nabla\chi^2\big|^{\top}_{x^{(k)}}\, x^{(k)}\right) \leq 0 \quad , \tag{6.58}$$

and the minimization objective function is $S(x) = \sum_j [x]_j$. The gradients used in Eq. (6.24) depend only on $x^{(k)}$, not on $\tilde{x}^-$:

$$\nabla S|_{\tilde{x}^-} = \vec{1} \quad , \tag{6.59}$$

$$\nabla f_j|_{\tilde{x}^-} = -\hat{e}_j \;\forall\; j \leq J \quad , \text{ and} \tag{6.60}$$

$$\nabla\tilde{f}_{J+1}\Big|_{\tilde{x}^-} = \nabla\chi^2\big|_{x^{(k)}} \quad . \tag{6.61}$$

Consider element j of Eq. (6.24):

$$(1) + \lambda_j(-1) + \lambda_{J+1}\left[\nabla\chi^2\big|_{x^{(k)}}\right]_j = 0 \quad \text{or} \tag{6.62}$$

$$\lambda_j = 1 + \lambda_{J+1}\left[\nabla\chi^2\big|_{x^{(k)}}\right]_j \quad . \tag{6.63}$$

Assume that $\nabla\chi^2|_{x^{(k)}}$ has at least one negative element at j' that bounds $[\tilde{x}^-]_{j'}$ above zero. The corresponding constraint is inactive: $f_{j'} < 0$, so Eq. (6.26) requires $\lambda_{j'} = 0$, so

$$\lambda_{J+1} = \frac{-1}{[\nabla\chi^2|_{x^{(k)}}]_{j'}} \quad . \tag{6.64}$$

Equation (6.25) requires $\lambda_{J+1} \geq 0$, so $[\nabla\chi^2|_{x^{(k)}}]_{j'} < 0$—but we have already assumed this. Substituting Eq. (6.64) into Eq. (6.63) for all other $j \neq j'$ gives

$$\lambda_j = 1 - \frac{[\nabla\chi^2|_{x^{(k)}}]_j}{[\nabla\chi^2|_{x^{(k)}}]_{j'}} \quad , \tag{6.65}$$

so $\lambda_j \geq 0$ and $[\nabla\chi^2|_{x^{(k)}}]_{j'} < 0$ together imply

$$\left[\nabla\chi^2\big|_{x^{(k)}}\right]_j \geq \left[\nabla\chi^2\big|_{x^{(k)}}\right]_{j'} \quad \text{and} \tag{6.66}$$

$$j' = \underset{j}{\operatorname{argmin}} \left[\nabla\chi^2 \big|_{x^{(k)}} \right]_j \quad . \tag{6.67}$$

This implies that $[x]_j > 0$ only if $[\nabla\chi^2|_{x^{(k)}}]_j = [\nabla\chi^2|_{x^{(k)}}]_{j'}$—in which case the linear programming problem is "degenerate" with many $\tilde{x}^-$ but a unique $\tilde{S}^-$. We assume that $[\tilde{x}^-]_{j'}$ is as small as permitted by the functional constraint:

$$\tilde{x}^- = \left\{ \begin{array}{cc} \frac{\chi^2_{I;0.95} - \chi^2(x^{(k)}) + \nabla\chi^2\big|^\top_{x^{(k)}} x^{(k)}}{\left[\nabla\chi^2|_{x^{(k)}}\right]_j} & j = j' \\ 0 & \text{else} \end{array} \right\} \quad \text{and} \tag{6.68}$$

$$\tilde{S}^- = \frac{\chi^2_{I;0.95} - \chi^2(x^{(k)}) + \nabla\chi^2\big|^\top_{x^{(k)}} x^{(k)}}{\left[\nabla\chi^2|_{x^{(k)}}\right]_{j'}} \quad . \tag{6.69}$$

This bound can be calculated using vector arithmetic; evaluating the gradient is arithmetic as shown in Section 6.1.4.

We can show that the bound converges as $x^{(k)}$ converges, i.e.,

$$\lim_{x^{(k)} \to x^*} \tilde{S}^-(x^{(k)}) = S(x^*) \quad . \tag{6.70}$$

As $x^{(k)}$ approaches x^*, the KKT conditions apply:

$$\lim_{k\to\infty} \lambda_j^{(k)} f_j(x^{(k)}) = 0 \quad \text{and} \tag{6.71}$$

$$\lim_{k\to\infty} \lambda_j^{(k)} \geq 0 \quad , \tag{6.72}$$

where $\lambda_j^{(k)}$ is a pseudo-KKT-multiplier that satisfies

$$\lambda_j^{(k)} = 1 + \lambda_{J+1}^{(k)} \left[\nabla\chi^2 \big|_{x^{(k)}} \right]_j \quad . \tag{6.73}$$

Together these equations and the assumption $\chi^2(x^*) = \chi^2_{I;0.95}$ imply that

$$\lim_{k\to\infty} \left[\nabla\chi^2 \big|_{x^{(k)}} \right]_j \begin{cases} = & \frac{-1}{\lambda_{J+1}} \quad \forall\ j \mid [x^*]_j > 0 \\ \geq & \frac{-1}{\lambda_{J+1}} \quad \forall\ j \mid [x^*]_j = 0 \end{cases} \quad \text{with} \tag{6.74}$$

$$\lambda_{J+1} = \min_j \frac{-1}{\left[\nabla\chi^2|_{x^*}\right]_j} = \frac{-1}{\min_j \left[\nabla\chi^2|_{x^*}\right]_j} \quad . \tag{6.75}$$

Taking the limit $k \to \infty$ of Eq. (6.69) gives

$$\lim_{k\to\infty} \tilde{S}^- = \lim_{k\to\infty} \frac{\chi^2_{I;0.95} - \chi^2(x^{(k)}) + \nabla\chi^2|^\top_{x^{(k)}}\, x^{(k)}}{\min_j \left[\nabla\chi^2|_{x^{(k)}}\right]_j} \tag{6.76}$$

$$... = \frac{(0) - \frac{1}{\lambda_{J+1}} \sum_j [x^*]_j}{\min_j \left[\nabla\chi^2|_{x^*}\right]_j} \tag{6.77}$$

$$... = \sum_j [x^*]_j = S(x^*) \quad . \tag{6.78}$$

In other words, $\tilde{S}^-(x^{(k)}) \to S(x^*)$ as $x^{(k)} \to x^*$.

In summary, we have defined an estimator $\tilde{S}^-$ that bounds the lower limit of $\mathbf{CI}_f$ from below. As $S(x^{(k)})$ descends, $\tilde{S}^-(x^{(k)})$ rises to meet it. Since

$$\tilde{S}^-(x^{(k)}) \leq S(x^*) \leq S(x^{(k)}) \quad \text{and} \tag{6.79}$$

$$S(x^{(k)}) - S(x^*) \leq S(x^{(k)}) - \tilde{S}^-(x^{(k)}) \quad , \tag{6.80}$$

the relative error is bounded

$$\frac{S(x^{(k)}) - S(x^*)}{S(x^*)} \leq \frac{S(x^{(k)}) - \tilde{S}^-(x^{(k)})}{\tilde{S}^-(x^{(k)})} \equiv w^-(x^{(k)}) \quad . \tag{6.81}$$

We can define a stopping criterion

$$\text{if } w^-(x^{(k)}) < \epsilon_w : \text{ halt} \quad , \tag{6.82}$$

for some ϵ_w several orders larger than machine precision. In this work we choose $\epsilon_w = 10^{-4}$ unless otherwise specified, since the confidence interval is probably several orders wider. The esimator S^- can also be used as a conservative lower limit for $\mathbf{CI}_f$, since $\tilde{S}^- < \min_{x\in\mathbf{CR}_f} S(x)$.

An analogous upper bound can be derived for the upper limit of $\mathbf{CI}_f$: When minimizing $-S$, the objective gradient is $-\vec{1}$ and the functional-constraint gradient

is $\nabla\chi^2|_{x^{(k)}} > 0$. Because $\nabla f_0|_{x^{(k)}}$ and $\nabla\chi^2|_{x^{(k)}}$ have both reversed sign, the upper bound

$$\tilde{S}^+ = \frac{\chi^2_{I;0.95} - \chi^2(x^{(k)}) + \nabla\chi^2|^\top_{x^{(k)}}\, x^{(k)}}{\min_j \left[\nabla\chi^2|_{x^{(k)}}\right]_j} \tag{6.83}$$

is apparently identical to $\tilde{S}^-$. However, the sign of $\nabla\chi^2$ has reversed. Since

$$S(x^{(k)}) \leq S(x^*) \leq \tilde{S}^+(x^{(k)}) \quad \text{and} \tag{6.84}$$

$$S(x^*) - S(x^{(k)}) \leq \tilde{S}^+(x^{(k)}) - S(x^{(k)}) \quad , \tag{6.85}$$

the relative error is bounded

$$\frac{S(x^*) - S(x^{(k)})}{S(x^*)} \leq \frac{\tilde{S}^+(x^{(k)}) - S(x^{(k)})}{S(x^{(k)})} \equiv w^+(x^{(k)}) \tag{6.86}$$

and create a stopping criterion

$$\text{if } w^+(x^{(k)}) < \epsilon : \text{ halt} \quad . \tag{6.87}$$

6.2.2 Rapid Convergence Region Monitoring

We solve the optimization problems in Section 6.1.1 using Newton's method with logarithmic barriers. For unconstrained optimization, the solver almost never steps out of the RCR. For constrained optimization the RCR will shift and shrink as the solver decreases μ, so the RCR may shift away from $x^{(k)}$. If the solver continues to decrease μ, the steps may be unable to locate the shrinking RCR. Stranded outside the RCR, the solver will not converge in reasonable time.

For example, consider the convergence plot in Fig. 6.1. In this problem we are seeking $\text{argmin}_{x\in\mathbb{R}_+}\, \chi^2(x)$ for an imaging problem with $I = 576$ datapoints and $J = 1000$ parameters. The barrier scale μ decreases by a factor $\gamma = 0.88$ at every iteration. However, the error bound w^- does not converge. After 116 iterations, $x^{(k)}$ moves outside $\mathbf{CR}_f$ because of discretization error.

The analyst faces a dilemma: aggressively decrease μ (set γ small) and risk convergence failure or timidly decrease μ and converge slowly. The ideal γ can only be determined by trial. To reduce this dilemma, the solver can freeze the RCR when $x^{(k)}$ is outside the RCR. This "RCR monitoring" feature helps the solver find or return

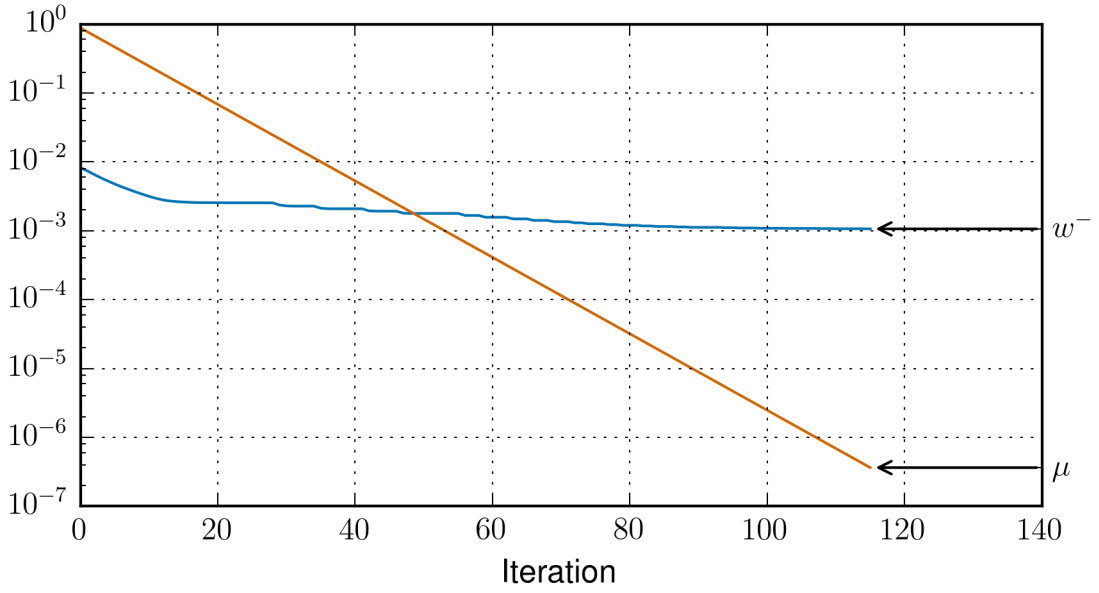


Figure 6.1: Convergence fails without Rapid Convergence Region monitoring.

to the RCR.

One indication that $x^{(k)}$ is outside the RCR is that the line-search scalar α is far from unity. See Eq. (6.11). Using this indicator, we can update Eq. (6.22) to

$$\mu^{(k+1)} = \begin{cases} \gamma\mu^{(k)}, & |\alpha - 1| \leq \epsilon_\alpha \\ \mu^{(k)}, & \text{else} \end{cases} . \tag{6.88}$$

Tolerance $\epsilon_\alpha = 0.9$ appears to work well. This prescription for the barrier coefficient monitors whether $x^{(k)}$ is in the RCR.

Using RCR monitoring, the solver typically converges dispite aggressive choices of γ. If we repeat the example of Fig. 6.1, the solver converges to $w^- \approx 10^{-6}$ in 200 iterations. See Fig. 6.2.

In this plot we see that the solver decreases μ in fewer than half of the iterations. This indicates that $\gamma = 0.88$ is rather aggressive, and a larger γ may converge more quickly. Near iteration 60 the solver requires dozens of iterations to find the RCR. However, the solver still converges beyond sufficient precision in finitely many iterations.

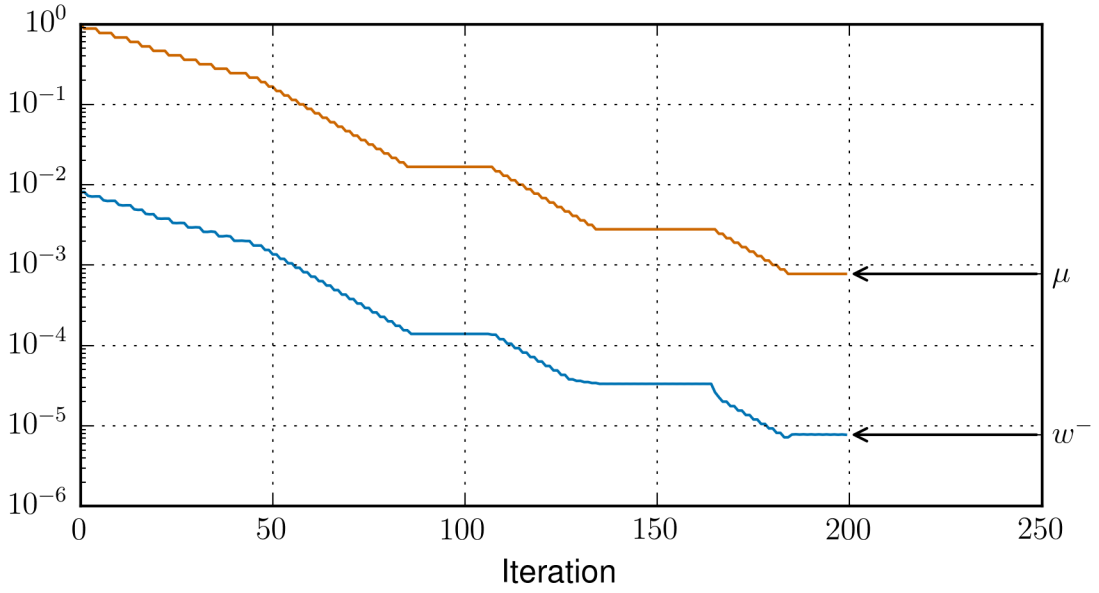


Figure 6.2: The solver converges with Rapid Convergence Region monitoring. Compare to Fig. 6.1

6.2.3 Initialization

The barrier method assumes that the analyst can provide an initial $x^{(0)} \in \mathbf{CR}_f$. Since $F(x;\mu)$ is undefined outside of $\mathbf{CR}_f$, logarithmic barriers cannot be applied until $x^{(k)}$ is in $\mathbf{CR}_f$. Various iterative schemes, e.g., expectation maximization, can minimize χ^2 and find $\mathbf{CR}_f$. We prefer to apply the barrier method to the minimum-χ^2 problem,

$$\text{seek } x^* \equiv \underset{x \in \mathbb{R}_+}{\text{argmin}}\, \chi^2(x) \quad \text{until} \tag{6.89}$$

$$\chi^2(x^{(k)}) < \chi^2_t \quad , \tag{6.90}$$

with the transition threshold $\chi^2_t = 0.999\ \chi^2_{I;0.95}$. We begin iteration with $x^{(0)} = \vec{1}$. Our optimization problem therefore has two stages: first locate $\mathbf{CR}_f$, then search $\mathbf{CR}_f$ for the extrema of S.

One challenge with this approach is the transition from stage 1 to stage 2. In Section 6.2.2 we noted that large changes of F can shift the RCR away from $x^{(k)}$. In that section we focused on changes due to μ, but altering the constrained objective

f_0 can lead to the same problem. An immediate transition from minimizing $f_0(x) = \chi^2(x)$ to minimizing $f_0(x) = S(x)$ significantly shifts the RCR.

For example, the analyst took special steps to find a suitable $x^{(0)}$) for the example in Fig. 6.1. If we naïvely apply the two-stage scheme to that problem, the solver is unable to find the RCR after the transition. See Fig. 6.3. The plot uses $\chi^2_t = 0.999\ \chi^2_{I;0.95}$, but results are similar with $0.9\ \chi^2_{I;0.95}$ and $0.99\ \chi^2_{I;0.95}$, and $0.9999\ \chi^2_{I;0.95}$.

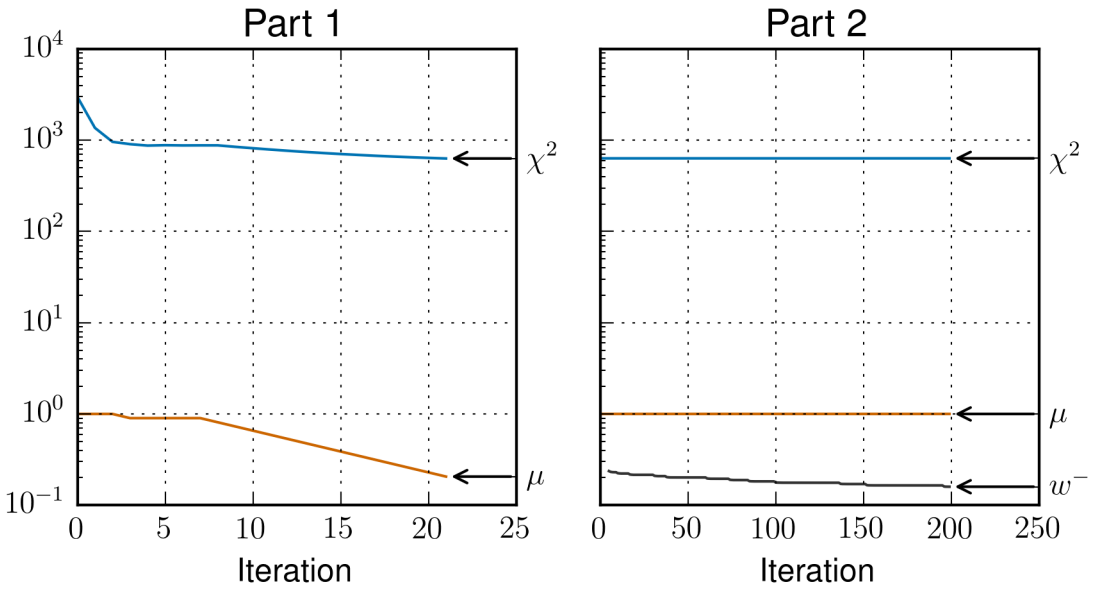


Figure 6.3: Two-stage initialization scheme using an abrupt transition fails to converge. $\gamma = 0.9$ for stage 1 and 0.95 for stage 2.

A more sophisticated approach could gradually adjust F. For example, we can augment F with additional parameters to emphasize different objectives:

$$F_1\left(x;\eta_0,\eta_p,\mu_l,\mu_f\right) \equiv \eta_0 f_0(x) + \eta_p \chi^2(x) + \mu_l \sum_{m=1}^{M-1} -\ln\left[s\right]_j \tag{6.91}$$

for stage 1 and

$$F_2\left(x;\eta_0,\eta_p,\mu_l,\mu_f\right) \equiv \eta_0 f_0(x) + \eta_p \chi^2(x) + \mu_l \sum_{m=1}^{M-1} -\ln\left[s\right]_j + \ldots$$

$$\ldots + -\mu_f \ln\left(\chi^2_{I;0.95} - \chi^2(x)\right) \tag{6.92}$$

for stage 2. The parameters $\eta_0 \in [0,1]$ and $\eta_p \in [0,1]$ scale the true objective function ($f_0 = S$ or $-S$) and a χ^2 penalty. The parameters $\mu_l \in (0,1]$ and $\mu_f \in (0,1]$ reshape the linear- and functional-constraint barriers.

These augmented functions F_1 and F_2 are useful because they can smoothly transition between stage 1 and stage 2. For stage 1 we adjust the parameters toward the limit

$$\begin{aligned} \lim \quad & F_1(x;\eta_0,\eta_p,\mu_l,\mu_f) = \chi^2(x) \ \forall x \in \mathbb{R}_+ \quad . \\ \eta_0 &\to 0 \\ \eta_p &= 1 \\ \mu_l &\to 0 \\ \mu_f &= 1 \end{aligned} \tag{6.93}$$

For stage 2 we adjust the parameters toward the limit

$$\begin{aligned} \lim \quad & F_2(x;\eta_0,\eta_p,\mu_l,\mu_f) = f_0(x) \ \forall x \in \mathbf{CR}_f \quad . \\ \eta_0 &\to 1 \\ \eta_p &\to 0 \\ \mu_l &\to 0 \\ \mu_f &\to 0 \end{aligned} \tag{6.94}$$

To make incremental changes in F, we change the parameters by no more than a factor γ at each step. One prescription is

$$\eta_0^{(k+1)} = \begin{cases} \eta_0^{(k)}, & \text{if } |\alpha - 1| > \epsilon_\alpha \\ \gamma_1 \eta_0^{(k)}, & \text{else if } x_k \notin \mathbf{CR}_f \\ \min\left(\eta_0^{(k)}/\gamma_2, 1\right), & \text{else} \end{cases} , \tag{6.95}$$

$$\eta_p^{(k+1)} = \begin{cases} \eta_p^{(k)}, & \text{if } |\alpha - 1| > \epsilon_\alpha \\ \eta_p^{(k)}, & \text{else if } x_k \notin \mathbf{CR}_f \\ \gamma_2 \eta_p^{(k)}, & \text{else} \end{cases} , \tag{6.96}$$

$$\mu_l^{(k+1)} = \begin{cases} \mu_l^{(k)}, & \text{if } |\alpha - 1| > \epsilon_\alpha \\ \gamma_1 \mu_l^{(k)}, & \text{else if } x_k \notin \mathbf{CR}_f \\ \gamma_2 \mu_l^{(k)}, & \text{else} \end{cases} \text{, and} \tag{6.97}$$

$$\mu_f^{(k+1)} = \eta_p^{(k+1)} \quad , \tag{6.98}$$

with $\eta_0^{(0)} = \eta_p^{(0)} = \mu_l^{(0)} = \mu_f^{(0)} = 1$. The ϵ_α condition implements RCR monitoring; see Section 6.2.2. The scale ratio γ is split into γ_1 for stage 1 and γ_2 for stage 2. In this work we use $\gamma_1 = 0.9$ and $\gamma_2 = 0.95$ unless otherwise specified. An example of this prescription is plotted in Fig. 6.4.

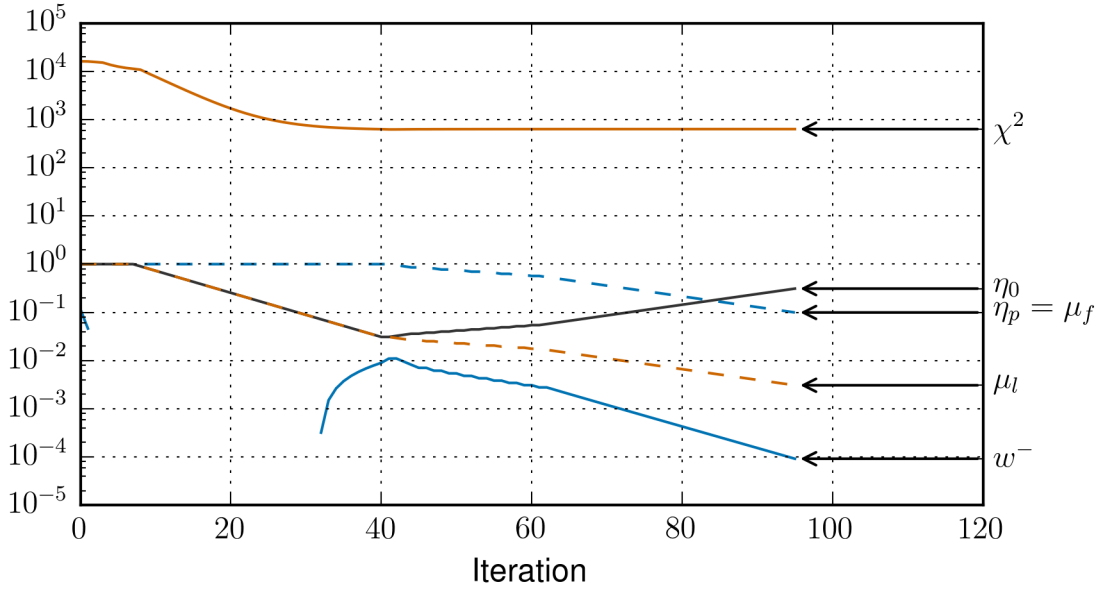


Figure 6.4: Two-stage initialization scheme using a smooth transition at iteration 41.

The naïve transition abruptly sets $\eta_0 = 1$, $\eta_p = 0$, and $\mu_l = 1$. The smooth transition avoids these abrupt changes, so the RCR shifts less. The only abrupt change is the addition of the functional-constraint barrier. This approach is much more likely to keep $x^{(k)}$ near the RCR and converge in reasonably few iterations.

6.2.4 Derivatives of the Objective Function

Section 6.1.3 describes a constrained optimization solver that requires the first and second derivatives of the objective function F. Using the augmented definition of F (defined in Sec. 6.2.3), our solver needs the first and second derivatives of F_2

with respect to x, i.e.,

$$\nabla_x F_2\left(x;\eta_0,\eta_p,\mu_l,\mu_f\right) \equiv \eta_0\nabla_x f_0(x) + \eta_p\nabla_x\chi^2(x) + \mu_l\sum_{j=1}^{J} -\nabla_x \ln\left[x\right]_j + ...$$

$$... + -\mu_f\nabla_x \ln\left(\chi^2_{I;0.95} - \chi^2(x)\right) \quad \text{and} \tag{6.99}$$

$$\nabla^2_x F_2\left(x;\eta_0,\eta_p,\mu_l,\mu_f\right) \equiv \eta_0\nabla^2_x f_0(x) + \eta_p\nabla^2_x\chi^2(x) + \mu_l\sum_{j=1}^{J} -\nabla^2_x \ln\left[x\right]_j + ...$$

$$... + -\mu_f\nabla^2_x \ln\left(\chi^2_{I;0.95} - \chi^2(x)\right) \quad . \tag{6.100}$$

(The terms in F_1 are a subset of the terms in F_2.) This section derives these terms or estimators for these terms.

For convenience, note here that

$$-\nabla_x \ln\left(f(x)\right) = -\frac{\nabla_x f}{f(x)} \quad \text{and} \tag{6.101}$$

$$-\nabla^2_x \ln\left(f(x)\right) = \frac{\left(\nabla_x f\right)^2 - f(x)\nabla^2_x f}{f(x)^2} \quad . \tag{6.102}$$

The derivatives of the original objective function have elements

$$\left[\nabla_x f_0\right]_j = 1 \quad \text{and} \tag{6.103}$$

$$\left[\nabla^2_x f_0\right]_{j,j'} = 0 \tag{6.104}$$

since f_0 is the sum of x.

The derivatives of the χ^2 depend on the choice of covariance model. For covariance model I we have derived

$$\nabla_x\chi^2 = A^\top\left(2\text{diag}(Ax)^{-1} - \text{diag}(Ax-d)\text{diag}(Ax)^{-2}\right)(Ax-d) \quad \text{and} \tag{6.105}$$

$$\nabla^2_x\chi^2 = 2A^\top\text{diag}(Ax)^{-1}\left(\mathbf{I} - \text{diag}(Ax-d)\text{diag}(Ax)^{-1}\right)^2 A \tag{6.106}$$

in Sec. 6.1.4. For covariance model II we assume that the covariance matrix covar$[Ax-d]$ depends only weakly on x. Using this assumption,

$$\chi^2(x) = (Ax-d)^\top \,\mathrm{covar}[Ax-d]^{-1}\,(Ax-d) \quad , \tag{6.107}$$

$$\nabla_x \chi^2 \approx 2A^\top \mathrm{covar}[Ax-d]^{-1}\,(Ax-d) \quad \text{, and} \tag{6.108}$$

$$\nabla_x^2 \chi^2 \approx 2A^\top \mathrm{covar}[Ax-d]^{-1} A \quad . \tag{6.109}$$

For nonnegativity constraint j, the derivatives have elements

$$\left[-\nabla_x \ln [x]_j\right]_{j'} = \begin{cases} -1/x & j = j' \\ 0 & \text{else} \end{cases} \quad \text{and} \tag{6.110}$$

$$\left[-\nabla_x^2 \ln [x]_j\right]_{j',j''} = \begin{cases} 1/x^2 & j = j' = j'' \\ 0 & \text{else} \end{cases} \quad . \tag{6.111}$$

The sums in Eqs. (6.99) and (6.100) are then

$$\sum_{j=1}^{J} -\nabla_x \ln [x]_j = -\mathrm{diag}(x)^{-1} \quad \text{and} \tag{6.112}$$

$$\sum_{j=1}^{J} -\nabla_x^2 \ln [x]_j = \mathrm{diag}(x)^{-2} \quad . \tag{6.113}$$

Like the χ^2 penalty, the χ^2 threshold constraint depends on the choice of noise model. The derivatives are

$$-\nabla_x \ln\left(\chi^2_{I;0.95} - \chi^2(x)\right) = -\frac{\nabla_x\left(\chi^2_{I;0.95} - \chi^2(x)\right)}{\left(\chi^2_{I;0.95} - \chi^2(x)\right)} = \frac{\nabla_x \chi^2(x)}{\left(\chi^2_{I;0.95} - \chi^2(x)\right)} \quad \text{and} \tag{6.114}$$

$$-\nabla_x^2 \ln\left(\chi^2_{I;0.95} - \chi^2(x)\right) = ...$$

$$... = \frac{\left(\nabla_x\left(\chi^2_{I;0.95} - \chi^2(x)\right)\right)^2 - \left(\chi^2_{I;0.95} - \chi^2(x)\right)\nabla_x^2\left(\chi^2_{I;0.95} - \chi^2(x)\right)}{\left(\chi^2_{I;0.95} - \chi^2(x)\right)^2} = ...$$

$$... = \frac{\left(\nabla_x \chi^2\right)^2 + \left(\chi^2_{I;0.95} - \chi^2(x)\right) \nabla^2_x \left(\chi^2(x)\right)}{\left(\chi^2_{I;0.95} - \chi^2(x)\right)^2} \quad , \tag{6.115}$$

with χ^2 and its derivatives given in Eqs. (6.105)–(6.109).

6.3 Application

In this section we use small problems to illustrate the frequentist UQ approach. We also use a scaling study to show the computational tractability of complex QI problems (with large I and/or J).

6.3.1 Small Problems

Let us first demonstrate the convex optimization algorithm using two small problems. The system matrices, data, and algorithm parameters are listed in Table 6.1. Because the problems have only 2 parameters, we can create a pseudocolor plot of $\chi^2(s)$; see Figs. 6.5 and 6.6.

Table 6.1: Defining parameters for two small problems.

Problem	6.3.1a	6.3.1b
A	$\begin{bmatrix} 40 & 60 \end{bmatrix}$	$\begin{bmatrix} 39.6 & 21.6 \\ 40 & 60 \end{bmatrix}$
d	$\begin{bmatrix} 158 \end{bmatrix}$	$\begin{bmatrix} 141 & 174 \end{bmatrix}^\top$

The χ^2-minimization arrives at a feasible solution within 2 iterations. The extrema searches proceed from there.

6.3.2 Scaling

We expect the $\mathbf{CI}_f$ to approach some steady value as we refine the source discretization. However, the computation time grows rapidly as $J \to \infty$. To study the effect of J on compute time and $\mathbf{CI}_f$, we used a pair of calibration measurements to perform a scaling study.

For pair $p = 10$, source Cf-252-5557 is located near $xyz = (-1.0, 0.0, 155.9)$ cm. We selected this pair because the position is central among the claibration measurements and has a median number of total counts. Here we treat the source distribution as an unknown distribution in a $60 \times 60 \times 60$ cm domain around the measured posi-

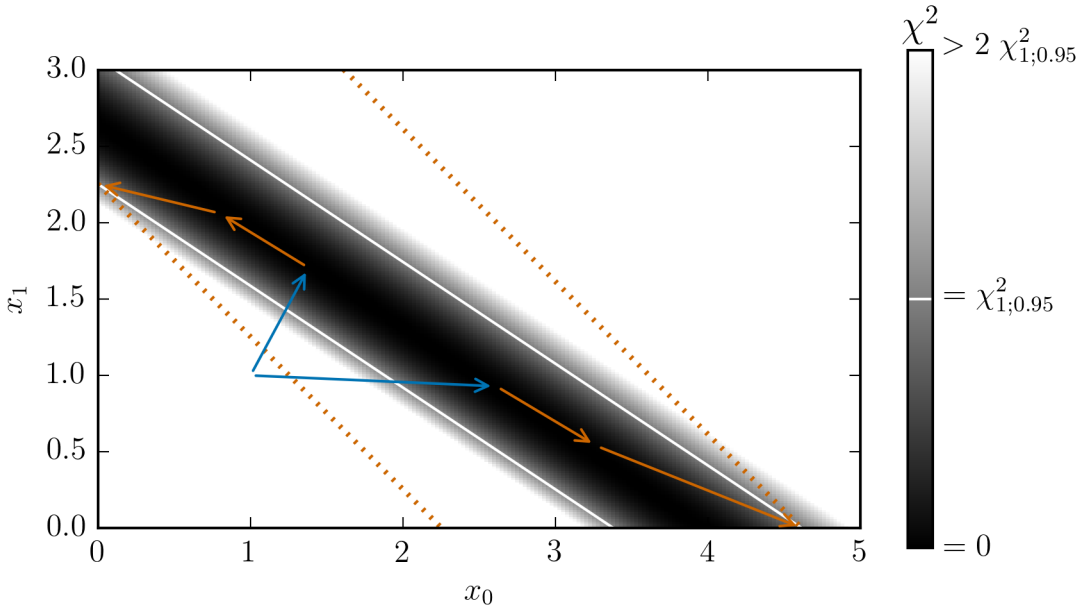


Figure 6.5: The convex optimization sequence for problem 6.3.1a: first seek $\min_{s\geq 0} \chi^2$ (blue arrows), then $\min_{s\in\mathbf{CR}_f} S$ and $\max_{s\in\mathbf{CR}_f} S$ (orange arrows). For clarity, only selected steps are drawn. Extrema of $S = x_0 + x_1$ are marked with orange dotted lines.

tion. In other words, the source is distributed in the domain $(-31.0, -30.0, 125.9) < (x, y, z) < (29., 30., 185.9)$ cm.

To better understand the information provided by the data, we performed an MLEM reconstruction of the pair data. Views of the reconstruction are plotted in Fig. 6.7. The reconstruction indicates that the data poorly identifies the source–detector distance.

We then performed the frequentist analysis using varying discretizations of the source domain. The analyses used covariance model I. The calculations are performed on a laptop with an Intel i7-3630QM processor. Key results for this analysis are summarized in Table 6.2.

We note several interesting trends. First, the computational walltime increases roughly proportionally to $J^{2.4}$. This scaling is reasonable based on the LAPACK routines we use to solve the $J \times J$ linear system to calculate the Newton step. This is true for each stage of the minimization searches and maximization searches as well; see Fig. 6.8. In this plot the walltime is divided between stage 1 (seeking $\mathbf{CR}_f$) and stage 2 (seeking the extrema of S in $\mathbf{CR}_f$). The extrema searches are marked $(-)$ for the minimum search and $(+)$ for the maximimum search. The increasing

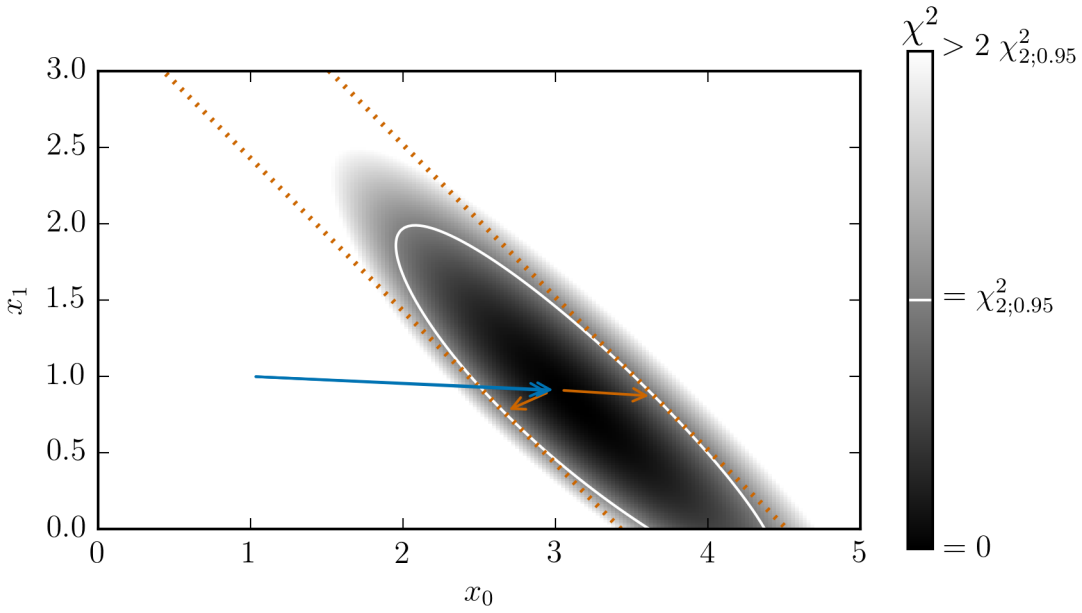


Figure 6.6: The convex optimization sequence for problem 6.3.1b: first seek $\min_{s\geq 0} \chi^2$ (blue arrows), then $\min_{s\in\mathbf{CR}_f} S$ and $\max_{s\in\mathbf{CR}_f} S$ (orange arrows). For clarity, only selected steps are drawn. Extrema of $S = x_0 + x_1$ are marked with orange dotted lines.

Table 6.2: Frequentist analysis results for calibration pair $p = 10$ as a function of source domain discretization.

J		walltime (s)	$\mathbf{CI}_f$ (neutrons / s)
$1 \times 1 \times 1$	$= 1$	5.53×10^1	a
$2 \times 2 \times 2$	$= 8$	5.88×10^1	a
$5 \times 5 \times 5$	$= 125$	5.67×10^1	$[3.390 \times 10^5, 7.581 \times 10^5]$
$5 \times 5 \times 10$	$= 250$	7.44×10^1	$[3.143 \times 10^5, 7.864 \times 10^5]$
$5 \times 10 \times 10$	$= 500$	9.70×10^1	$[2.729 \times 10^5, 8.313 \times 10^5]$
$10 \times 10 \times 10$	$= 1000$	2.29×10^2	$[2.703 \times 10^5, 8.442 \times 10^5]$
$10 \times 10 \times 20$	$= 2000$	6.89×10^2	$[2.669 \times 10^5, 8.541 \times 10^5]$
$10 \times 20 \times 20$	$= 4000$	2.85×10^3	$[2.622 \times 10^5, 8.756 \times 10^5]$
$20 \times 20 \times 20$	$= 8000$	1.45×10^4	$[2.619 \times 10^5, 8.829 \times 10^5]$
S_{true}			3.65×10^5

a Unable to find $\mathbf{CR}_b$

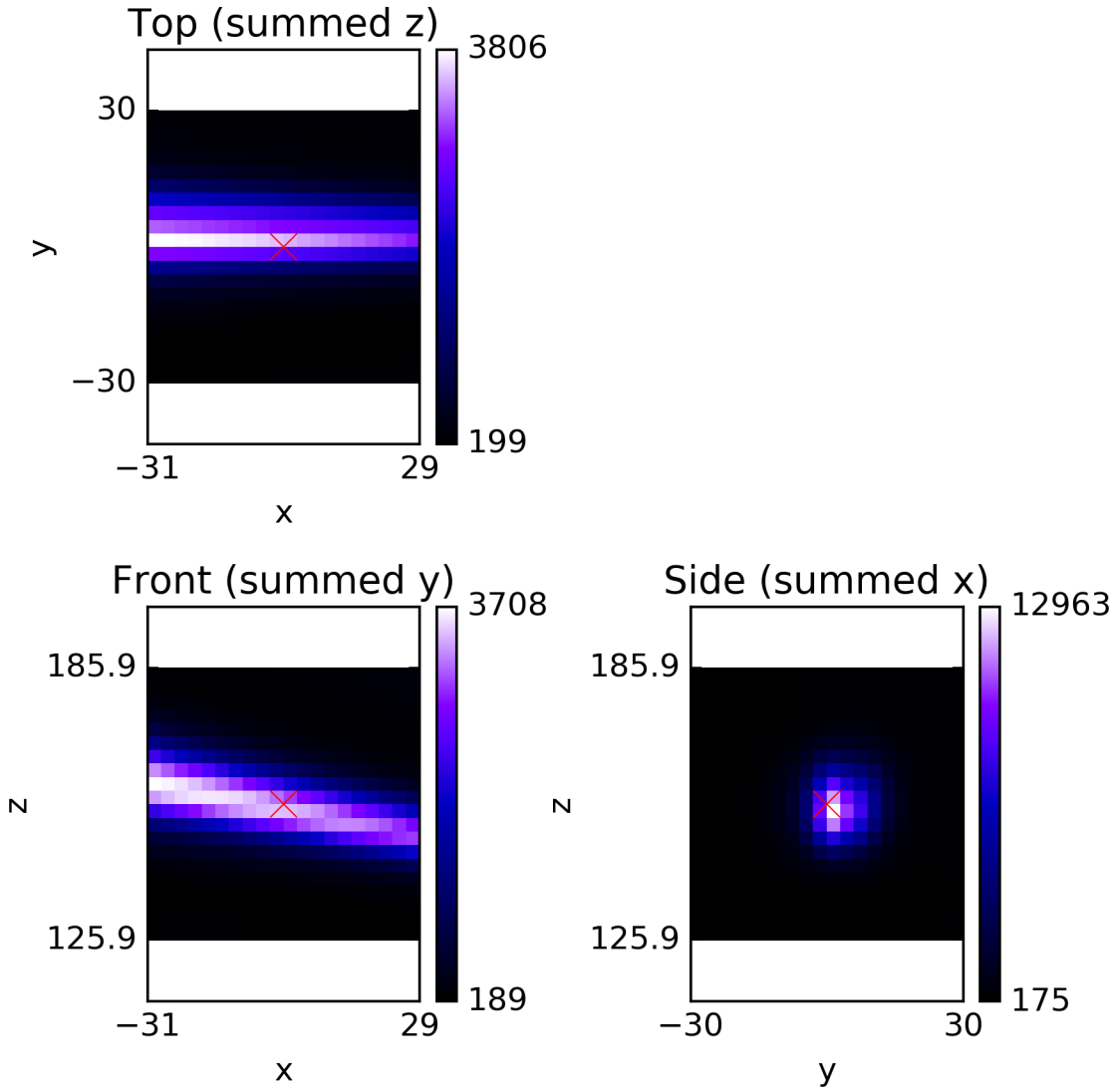


Figure 6.7: Views of the MLEM reconstruction of calibration pair $p = 10$. The measured source position is marked with a red $\times$.

computational cost is driven by increasing cost per step, since the number of steps is insensitive to J; see Fig. 6.9.

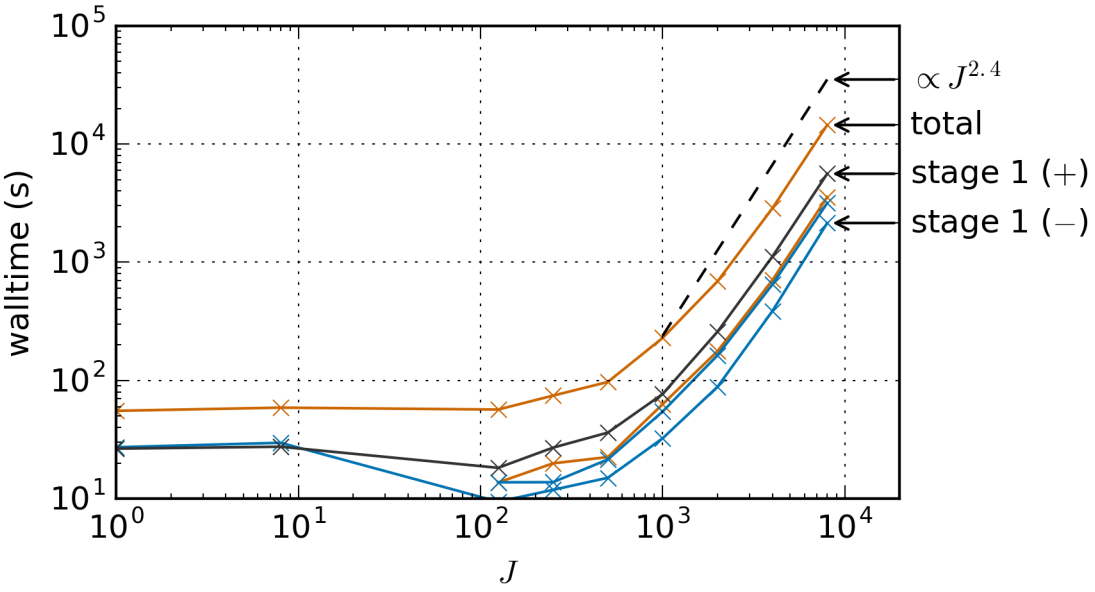


Figure 6.8: Computational walltime for frequentist analysis of calibration pair $p = 10$ as a function of source discretization. Stage 2 times (unlabeled) fall between the curves stage 1 times in the minimization (−) and maximization (+) problems.

Second, seeking $\mathbf{CR}_f$ requires many more steps for the maximization problem than the minimization problem. This is because the RCR is harder to locate in stage 1 of the maximization problem. For example, observe the large number of iterations before the barrier scaling parameter μ_l changes; compare Figs. 6.10 and 6.11. This behavior is not observed in stage 1 of the minimization problem, suggesting that it is caused by the η_0 term of F_1. In future work we may be able to reduce the computational cost by beginning the maximization problem with smaller values of η_0.

Third, the confidence intervals widen as J increases. This is expected, since adding parameters will improve the fit between the model and the data and increase the size of $\mathbf{CR}_f$. We even observe cases where $\mathbf{CR}_f$ is an empty set when $J < 125$. (In these cases, the solver performs 200 stage-1 steps and halts.) However as $J \to \infty$, $\mathbf{CI}_f$ approaches a constant interval. In Section 3.6 we suggest using voxels that are 10 cm perpendicular to the detector plane and 3 cm parallel to the plane. The results in this section support suggest a similar discretization: 6 cm perpendicular to the detector plane and 3 cm parallel to the plane.

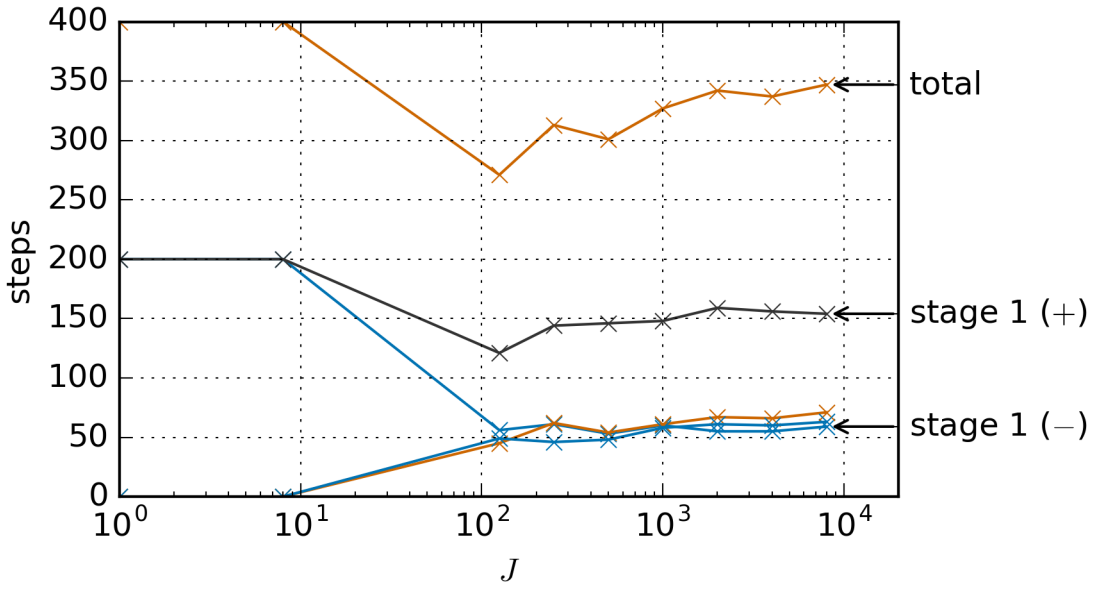


Figure 6.9: Number of computational steps for frequentist analysis of calibration pair $p = 10$ as a function of source discretization. Stage 2 step counts (unlabeled) lie near the stage 1 step counts for the minimization problem $(-)$.

Finally, each step has a fixed minimum cost on the order of 0.2 seconds. This cost is driven in part by the need to invert the $I \times I$ covariance matrix. We can assess the scaling of this cost by repeating the $J = 1000$ trial with the data and response matrix repeated once, twice, etc. See Table 6.3. The cost per step scales as $\mathcal{O}(I^{2.2})$.

Table 6.3: Frequentist analysis walltimes for calibration pair $p = 10$ with $J = 1000$ as a function of number of datapoints.

I	walltime (s)
576×1	0.53
576×2	1.81
576×3	4.42
576×4	8.29

From these measurements, our overall time prediction is

$$1.16 \times 10^{-4} I^{2.2} + 6.22 \times 10^{-6} J^{2.4} \quad \text{seconds.} \tag{6.116}$$

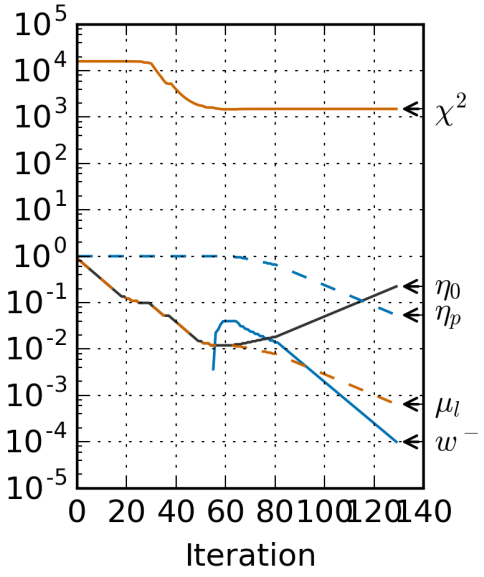


Figure 6.10: Convergence of the minimization problem for calibration pair $p = 10$ with $J = 8000$ shows an immediate decrease in μ_l.

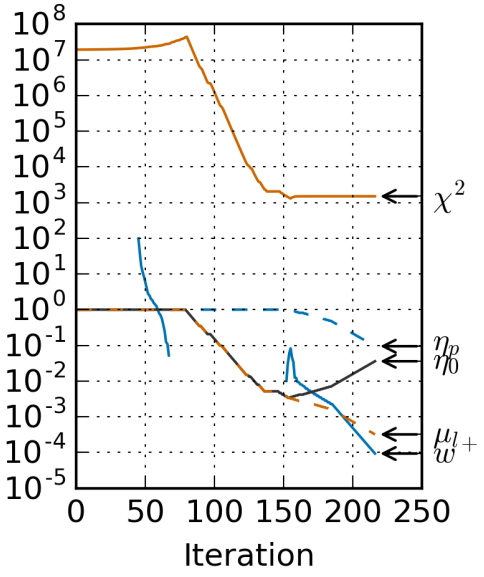


Figure 6.11: Convergence of the maximization problem for calibration pair $p = 10$ with $J = 8000$ shows constant μ_l for the first 80 iterations.

6.4 Conclusions

This section builds on existing constrained optimization techniques to create a robust UQ approach. The convex optimization methods in Section 6.2 may be useful for other optimization applications. For a problem with $I = 576 \times 2$ datapoints and $J = 10^4$ parameters, the estimate can be computed on a laptop in roughly 7 hours.

CHAPTER VII

Demonstration

The ultimate demonstration of the UQ approach of Chapter VI is a demonstration with an unknown source distribution. This exercise also highlights pratical challenges of QI using an FNCA imager.

7.1 Equipment and Setup

We used the P24 imager (described in Section 3.1) and a set of small californium-252 sources. Paul Hausladen selected the sources and arranged them behind a curtain. See Fig. 7.1. This is the "black box" ROI with an unknown neutron source distribution.

The number, strength, and arrangement of the sources hidden from the inspector (Bevill). Hausladen placed a total of seven sources in three groups in the ROI. The sources are listed in Table 7.1. (The calculation of source intensity is identical to Sec. 3.1. Refer to the coordinate system in Fig. 7.1.)

Table 7.1: Sources used in the hidden-source demonstration.

name	cert. date	cert. strength (Ci)	intensity on 2016-08-19 (neutrons/s)	approx. position (cm)
Cf-252-5214	2001-09-18	1.23×10^{-3}	1.059×10^{5}	(-228.6, -182.9, 121.9)
Cf-252-4863	2001-03-14	132×10^{-6}	9.926×10^{3}	(-320.0, -182.9, 132.1)
Cf-252-4864	2001-03-14	132×10^{-6}	9.926×10^{3}	(-320.0, -182.9, 132.1)
Cf-252-4865	2001-03-14	132×10^{-6}	9.926×10^{3}	(-320.0, -182.9, 132.1)
Cf-252-4866	2001-03-14	132×10^{-6}	9.926×10^{3}	(-320.0, -114.3, 132.1)
Cf-252-4867	2001-03-14	132×10^{-6}	9.926×10^{3}	(-320.0, -114.3, 132.1)
Cf-252-4868	2001-03-14	132×10^{-6}	9.926×10^{3}	(-320.0, -114.3, 132.1)
total			165410.	

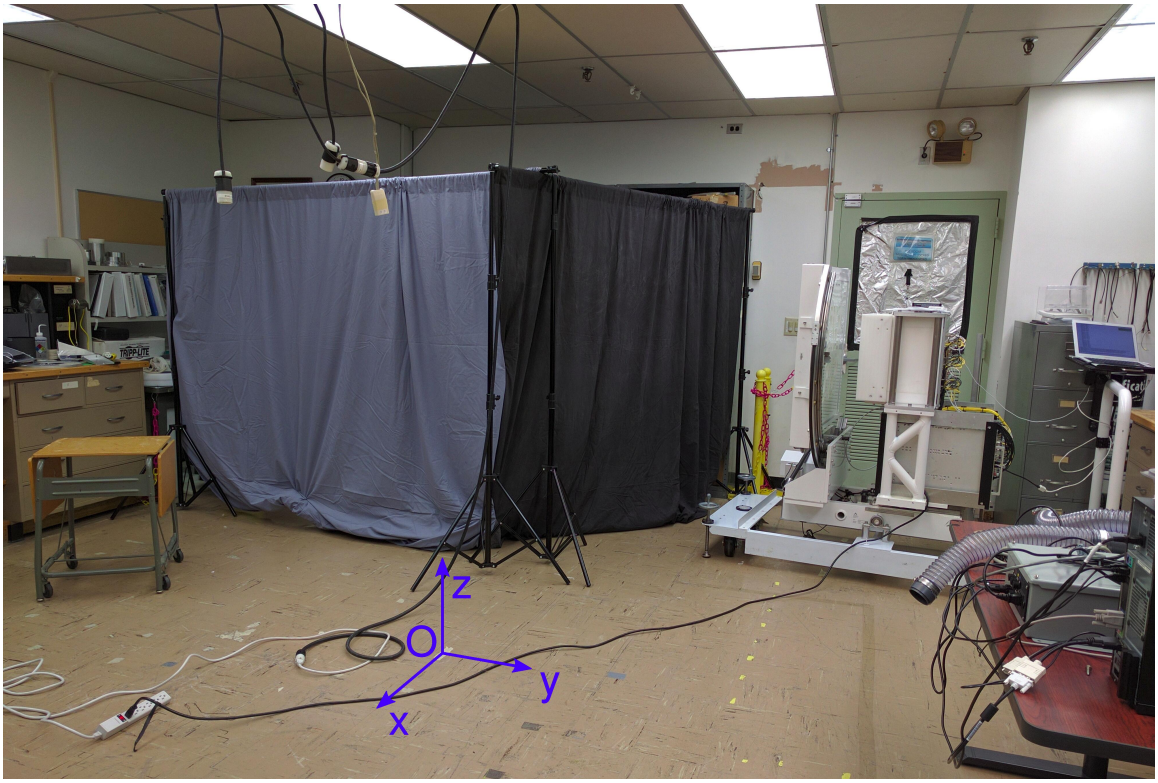


Figure 7.1: A view of the curtain concealing an unknown distribution of californium-252 sources.

We note that this total source strength is equivalent to approximately 254 to 1479 grams of plutonium. The upper estimate is based on 1 gram of "fuel grade plutonium" [75] emitting $\sim$ 111.8 neutrons per second via spontaneous-fission [76]. Refer to caclulations in Table 7.2. More neutrons could be created by (α, n) knockout reactions and americium-241 spontaneous fission, depending on the chemical composition and age of the plutonium. A different reference estimates 650 neutrons per second per gram of plutonium-oxide [23]. A source intensity of around a kilogram of plutonium is very significant in the context of holdup measurement.

Table 7.2: Calculation of spontaneous-fission neutron intensity from fuel grade plutonium.

Isotope	weight fraction [75]	s.f. neutron intensity [76]	intensity per gram Pu (neutrons / s)
Pu-238	0.001	2600.0	2.6
Pu-239	0.861	0.022	0.018 942
Pu-240	0.120	910.0	109.2
Pu-241	0.016	0.0491	7.856×10^{-4}
Pu-242	0.002	0.0017	3.4×10^{-6}
total			111.8

7.2 Measurements

The inspector first noted the geometry of the ROI. The curtained region is bounded (in centimeters) $-297.2 < x < -137.3$, $-251.5 < y < 22.5$, and $0 < z < 200.7$.

The inspector also noted that metal can be a significant contributor of fast-neutron scatter toward the detector. The curtain is suspended using a metal frame, and metal shelves were located along the $-x$ and $-y$ sides of the ROI.

The inspector then acquired multiple measurements of the ROI using the P24 imager. The measurements occured at various positions (A, B, C, ...) and for various walltimes (1 hour, 2 hours, ...). The inspector selected measurements A8 (from the $+x$ side of the ROI) and C8 (from the $+y$ side of the ROI) for further analysis. See details listed in Table 7.3.

Table 7.3: Geometry and walltime of measurements used in the demonstration analysis.

Name	A8	C8
Aperture center xyz (cm)	$(121.6, -160.4, 116.2)$	$(-207.7, 130.7, 116.2)$
Imager orientation uvw	$(-0.9999, -0.01416, 0.)$	$(-0.1123, -0.99367, 0.)$
Walltime (seconds)		
mask	14400	14400
antimask	14400	14400
Counts recorded		
mask	102 458	101 627
antimask	130 583	129 680

7.3 Preliminary Reconstruction

Using measurements A8 and C8, the inspector created an MLEM reconstruction of the black box source distribution. The mask and antimask measurements are considered separately, so there are a total of $I = 4 \times 576$ datapoints. The ROI is discretized into a uniform 3D grid of $J = 20 \times 20 \times 20$ voxels. The response matrix was then calculated using the calibrated model of Section 3.3. The MLEM reconstruction used covariance model I (i.e., standard MLEM reconstruction).

Views of the 3D reconstruction are plotted in Fig. 7.2. It appears that most of the fast neutrons reaching the detector originate at one of three regions. See Table 7.4. When we infer the total source strength, we will limit our unknown source distribution to these three regions.

Table 7.4: Regions of the demonstration ROI from which significant numbers of fast neutrons reach the detector.

color	domain (cm)			discretization
cyan	$(-250, -210, 100)$	$< (x, y, z) <$	$(-200, -155, 140)$	coarse: $8 \times 9 \times 7$ fine: $17 \times 18 \times 13$
orange	$(-180, -209, 101)$	$< (x, y, z) <$	$(-137, -156, 139)$	coarse: $7 \times 9 \times 6$ fine: $14 \times 18 \times 13$
gray	$(-180, -250, 0)$	$< (x, y, z) <$	$(-137, -220, 90)$	coarse: $4 \times 3 \times 8$ fine: $7 \times 5 \times 15$

The first region (cyan in Fig. 7.2) appears to be a true source or set of sources indistinguishably close to one another. Fast neutrons may scatter on metal objects in or near the ROI, leading to nuisance "scatter sources" in the ROI. We assume that this is the case, so we will model the cyan region as our true ROI and treat the other

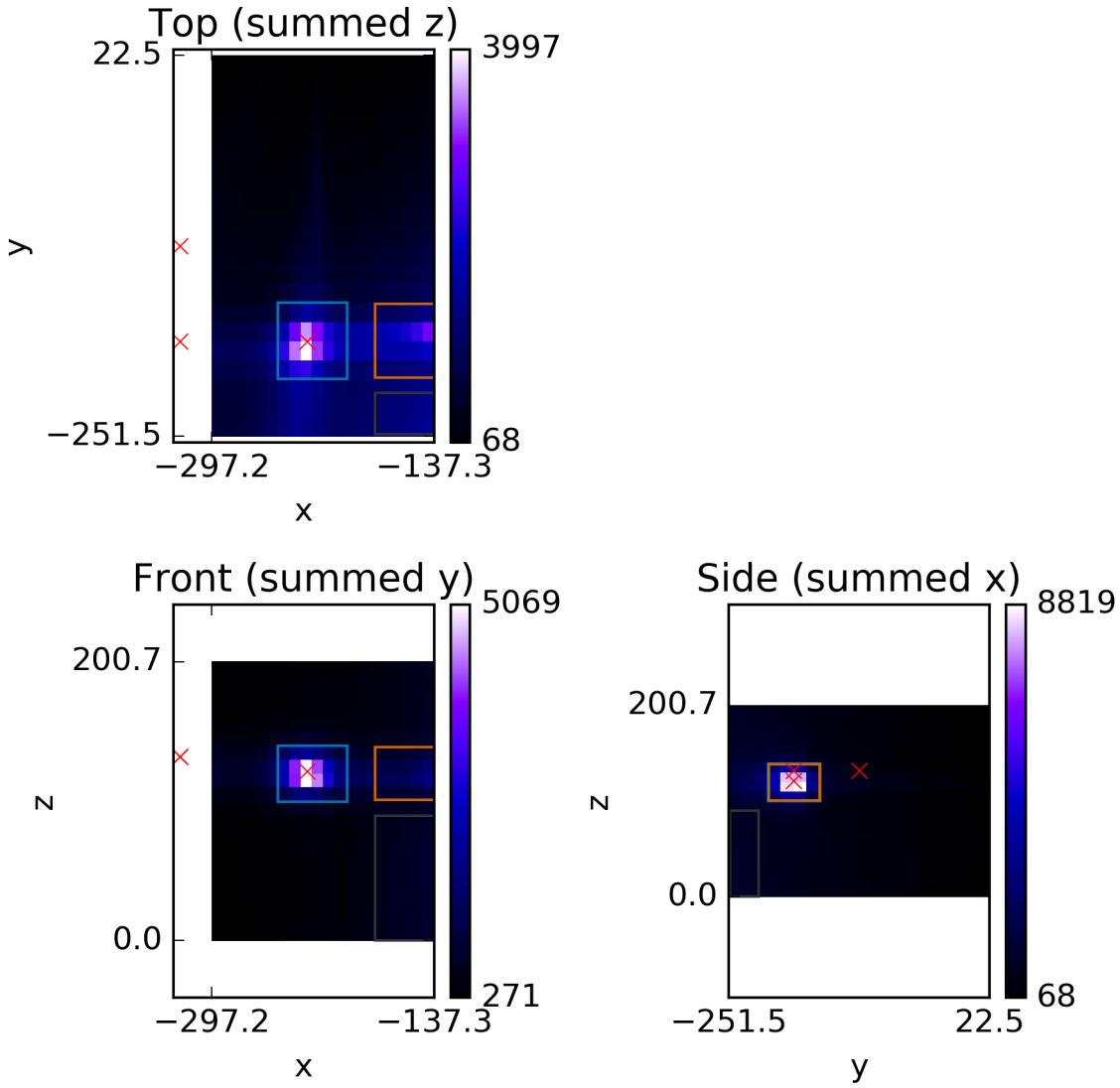


Figure 7.2: The 3D MLEM reconstruction of the ROI suggests three significant origin areas (cyan, orange, and gray boxes). The true (hidden) source positions are marked with red ×; two of the positions are outside the MLEM domain.

regions as nuisance unknowns (excluded from the optimization objective).

We discretize these regions as the source domain to infer the source strength. The sensitvity study in Section 3.6 suggests inference with voxels no wider than 3 cm. However for the sake of numerical tractability we limit our inference calculation to $J < 10^4$ unknown source voxels. Table 7.4 includes voxel counts for a coarse discretization (with $J = 978$) and a fine discretization (with $J = 7779$). The fine discretization achieves ∼3 cm resolution for the cyan and orange regions and ∼6 cm resolution for the gray region.

7.4 Frequentist Analysis

We can now use the frequentist analysis approach of Chapter VI to infer the total source strength. The inference results vary depending on which datasets are analyzed, which ROI discretization is used, and which covariance model is used. As recommended in Section 3.4, threshold values of $\chi^2_{I;0.95} = 1500, 750,$, and 800 are used for covariance models, I, II, and III (respectively) when one measurement is analyzed ($I = 576$). When two measurements are used, $\chi^2_{I;0.95}$ is doubled. The results are listed in Tables 7.5 and 7.6.

Table 7.5: Confidence interval of the total source intensity (neutrons/s) of hidden sources using the coarse mesh ($J = 978$).

dataset(s)	cov. model I	cov. model II	cov. model III
A8	$[0.0185, 3.981 \times 10^5]$	$[0.0094, a]$	$[0.0240, 3.603 \times 10^5]$
C8	b	b	b
A8 + C8	$[1.313 \times 10^5, 2.634 \times 10^5]$	b	$[1.371 \times 10^5, 2.027 \times 10^5]$

a Interval appears unconstrained
b Unable to find $\mathbf{CR}_b$

Table 7.6: Confidence interval of the total source intensity (neutrons/s) of hidden sources using the fine mesh ($J = 7779$).

dataset(s)	cov. model I	cov. model II	cov. model III
A8	$[0.0539, 4.117 \times 10^5]$	$[3.977 \times 10^3, a]$	$[0.0209, 3.722 \times 10^5]$
C8	b	b	b
A8 + C8	$[1.283 \times 10^5, 2.705 \times 10^5]$	b	$[1.339 \times 10^5, 2.092 \times 10^5]$

a Interval appears unconstrained
b Unable to find $\mathbf{CR}_b$

Covariance models I and III give fairly consistent values of $\mathbf{CI}_f$. The intervals

calculated using model III tend to be slightly smaller than using model I, but all of the intervals contain the true $S = 1.65 \times 10^5$ neutrons per second. However, the solver seldom finds values of χ^2 smaller than the threshold using covariance model II. This indicates that $\mathbf{CR}_f$ is an empty set. When $\mathbf{CR}_f$ is not an empty set, it appears to be unbounded as s increases. This is possible because the covariance increases roughly linearly as the predicted count rate increases.

Using two measurements yields a much narrower $\mathbf{CI}_f$. In fact, the algorithm cannot distinguish neutrons from the true source region (cyan) from neutrons from the scatter-source regions (orange and gray). This explains why the lower bounds created from measurement A8 approach zero at the lower bound. When using only measurement C8, the model cannot find a reasonable s. This is because the model excludes the sources hidden at the $-x$ edge of the ROI, which significantly affect the C8 data.

The intervals calculated using the fine mesh ($J = 7779$) are about 10% wider than the intervals created using the coarse mesh. This suggests that the voxel discretization error is reasonable.

The confidence intervals are quite wide. Even the narrowest interval has a relative width of $\frac{2.027 \times 10^5 - 1.371 \times 10^5}{1.716 \times 10^5} = 44\%$. This is similar to a relative standard deviation of 22%.

7.5 Reanalysis

Once the true source positons were disclosed, the inspector recognized that the modeled ROI geometry did not include two of the source positions. Refer to the source positions in Table 7.1 and the source locations marked in Fig. 7.2. This highlights one possibility for human error, but also raises the question of how the method would have performed had the inspector correctly bounded the ROI. In this section we repeat the analysis procedures of Sections 7.3 and 7.4 using a larger MLEM domain.

The new domain is bounded (in centimeters) $-350 < x < -137.3$, $-251.5 < y < 22.5$, and $0 < z < 200.7$. (This is the original domain with the x lower bound decreased to -350 cm.) The MLEM reconstruction using the new domain is plotted in Fig. 7.3.

The updated reconstruction only suggests one new source domain. See the cyan region marked at the $-x$ edge of Fig. 7.3. The new domain spans $(-350, -225, 100) < (x, y, z) < (-300, -170, 140)$. We subdivide it into $8 \times 9 \times 7 = 504$ uniform voxels to achieve ~ 6 cm resolution. Adding these voxels to the "coarse" discretization of

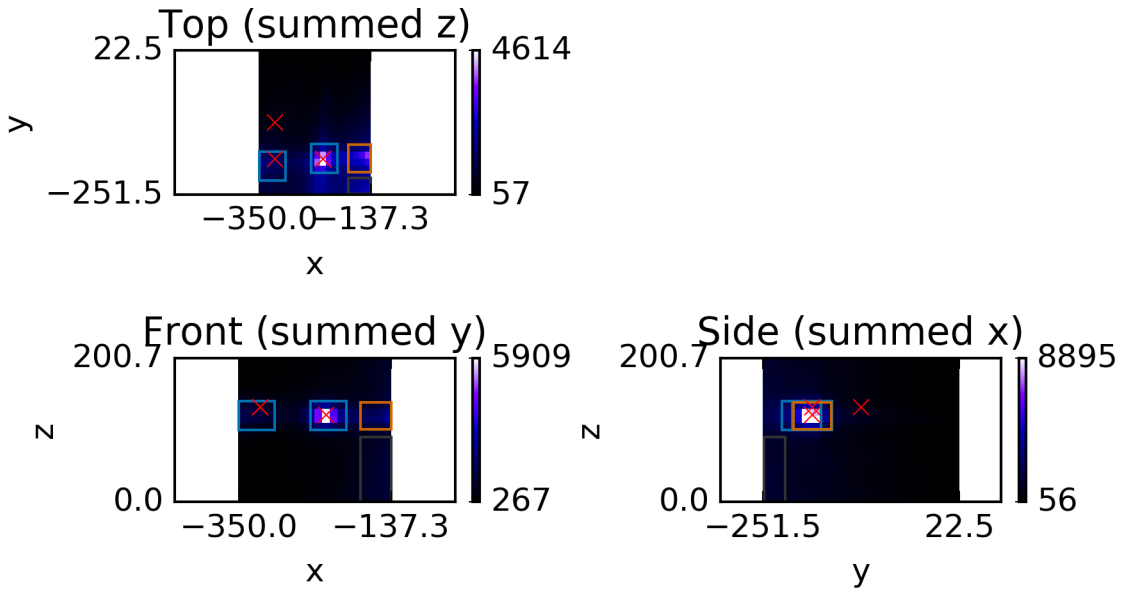


Figure 7.3: The updated 3D MLEM reconstruction of the ROI suggests four significant origin areas (cyan, orange, and gray boxes). The true (hidden) source positions are marked with red ×.

Section 7.3 gives a total of $J = 1482$ voxels to infer the total source intensity.

For the frequentist analysis, the inspector assumed that the two cyan regions were true sources and that the orange and gray regions were nuisance scatter sources. Results of the frequentist analysis are listed in Table 7.7.

Table 7.7: Confidence interval of the total source intensity (neutrons/s) of hidden sources using the reanalyzed coarse mesh ($J = 1482$).

dataset(s)	cov. model I	cov. model II	cov. model III
A8	$[0.0433, 5.553 \times 10^5]$	$[0.0187, a]$	$[0.0433, 4.945 \times 10^5]$
C8	$[1.372 \times 10^5, 5.890 \times 10^5]$	$[1.224 \times 10^5, a]$	$[1.299 \times 10^5, 5.497 \times 10^5]$
A8 + C8	$[1.293 \times 10^5, 4.868 \times 10^5]$	$[1.264 \times 10^5, a]$	$[1.248 \times 10^5, 4.432 \times 10^5]$

a Interval appears unconstrained

The resulting intervals are wider than in the original analysis. This is not surprising, since the original $\mathbf{CR}_f$ is subset of the reanalyzed $\mathbf{CR}_f$. The extrema of the reanalyzed $\mathbf{CR}_f$ are at least as extreme as the extrema of the original $\mathbf{CR}_f$. The narrowest $\mathbf{CI}_f$ is 112% of the mean.

Interestingly, the $\mathbf{CR}_f$ from measurement C8 alone is no longer an empty set. The new region provided additional flexibility to the model that explained the observed

data.

The interval calculated using both measurements is only slightly narrower than the interval calculated using C8 alone. This is expected because the source–detector distance is crucial for inferring source strength. Because the source regions are narrow along the y-axis, the distance was narrowly constrained for measurement C8. However, the regions would not be narrow if they were determined using an MLEM reconstruction of only measurement C8. The distance is broadly constrained along the x-axis, so measurement A8 needs the second measurement to infer the distance.

7.6 Conclusions

The results in Chapter VII demonstrate that FNCA imaging can infer total source intensity within a region of interest. This inference overcomes small amounts of scatter that the inspector identified as a nuisance source. However, the 112% measurement uncertainty is quite large, given that the equivalent of ~ 1 kilogram of plutonium was measured for 16 hours. Since the data includes a large ($> 4 \times 10^5$) number of counts, the uncertainty appears to be mostly due to modeling error. Models should be improved in future work.

The results also identify the method's reliance on subjective decisions made by inspector. Because the inspector omitted a portion of the ROI, several sources were absent from the initial analysis. Follow-up analysis shows that the confidence intervals change significantly when these sources are considered. These subjective decisions could be avoided entirely if analyses with $J \gg 10^4$ were computationally tractable.

CHAPTER VIII

Conclusions

This work lays the foundation for emission QI using FNCA imagers.

We created and calibrated a model to predict the difference between mask–antimask measurement pairs. We used cross-validation to test the predictive power of a calibrated forward model. Although the prediction–measurement residuals are statistically significant (vis-à-vis measurement noise), the residuals are distributed in a predictable way. Based on this validation work we can set bounds on reasonable values of a chi-squared goodness-of-fit parameter. Having a validated model of prediction–measurement residuals enables meaningful UQ.

Future work should include incorporation of objects near the ROI (e.g., glove boxes) into the model. This could be a merger of the modeling here with the MCNPX models of [1, 2].

We have also demonstrated several concepts relevant to the UQ problem. The UQ problem is similar to other UQ problems: a (nearly) Gaussian-distributed dataset undergoes a (nearly) linear operation to infer a quantity of interest. For problems like this, traditional UQ approaches (e.g., GLLS and the delta method) are appropriate. However, the holdup QI problem also has inequality constraints and is usually underdetermined. These complications motivated us to pursue a non-traditional UQ approach.

We account for the multiple sources of uncertainty in this problem—noisy data, limited model accuracy, physical constraints, and epistemic uncertainty—using a frequentist stastical approach. In this approach we search within a confidence region of reasonable solutions to determine bounds on the QOI. This approach is computationally tractable up to $J = 10^4$ unknowns, which is roughly a 1 m^3 grid at typical FNCA resolution. Larger problems will likely arise in practice, so alternative approaches should be considered in future work.

These alternative approaches could be high-performance implementations of our frequentist approaches or perhaps Bayesian approaches. As Section 2.2 demonstrates, the Bayesian posterior for this problem can be multimodal. This calls for advanced Bayesian techniques such as DREAM or RJMCMC; see Section 2.1.1.1.

We conclude with a blind-inspector demonstration of the frequentist approach. The inspector correctly inferred the total strength of multiple sources hidden in a large volume. The resulting uncertainty is very large however, due mostly to modeling error. Future work could improve the models to reduce the uncertainty due to modeling error.

Two additional topics may warrant significant further research.

First, a reconstruction approach that incorporates data updates mid-calculation would decrease the overall time-to-solution. With this system, the analysis software could begin processing during data acquisition. This approach could also allow analysts to adjust the imager based on preliminary reconstructions. Imager adjustments improve the robustness of the measurement by providing complementary information, e.g., parallax.

Second, a new aperture design could provide that parallax information in a single view. Existing FNCA imagers are rather insensitive to source distance. Since source distance affects detector efficiency more than in-plane motion, a distance-sensitive system would acquire more information relevant to holdup QI. For example, a mask with apertures cut non-perpendicular to the scintillator could provide parallax in a single-view measurement.

APPENDIX

APPENDIX A

MCNP5 Model

This appendix details the MCNP5 input used to model a glove box in Section 3.5. See Tables A.1–A.6. Variations of this template were used to generate the results in Section 3.5.

Table A.1: MCNP5 cell cards used in Section 3.5.

```
 Templated problem for fast-neutron imaging phenomena
c
c glovebox
100 $ fill 5 -0.001205 -130 -141 -142 -143 -144 imp:n=1
130 $ holdup lining sides 4 -2.477 130 -10 imp:n=1
141 $ holdup in corner s020 4 -2.477 -130 141 imp:n=1
142 $ holdup in corner s022 4 -2.477 -130 142 imp:n=1
143 $ holdup in corner s220 4 -2.477 -130 143 imp:n=1
144 $ holdup in corner s222 4 -2.477 -130 144 imp:n=1
101 $ glass 21 -6.22 10 -11 imp:n=1
110 $ front leg 1 -8.0 -12 13 imp:n=1
111 $ front leg hollow 5 -0.001205 -13 imp:n=1
121 $ back leg 1 -8.0 -14 15 imp:n=1
122 $ back leg hollow 5 -0.001205 -15 imp:n=1
c
c pipe
200 $ pipe wall 1 -8.0 -20 21 imp:n=1
210 $ pipe fill 5 -0.001205 -21 -221 -222 -223 -224 imp:n=1
221 $ pipe holdup r020 4 -2.477 -21 221 imp:n=1
222 $ pipe holdup r022 4 -2.477 -21 222 -221 imp:n=1
223 $ pipe holdup r220 4 -2.477 -21 223 -221 imp:n=1
224 $ pipe holdup r222 4 -2.477 -21 224 -222 -223 imp:n=1
c
c detector
500 $ void around detector to preclude nearby scatter 0 -500 imp:n=1
c
c environment
300 $ concrete 3 -2.18 -30 -99 imp:n=1
400 $ air 5 -0.001205 30 -99 11 12 14 20 500 imp:n=1
999 $ beyond 0 99 imp:n=0
```

Table A.2: MCNP5 surface cards used in Section 3.5.

```
 c glovebox
10 rpp -50.0 50.0 -200.0 200.0 -50.0 50.0 $ glovebox glass--holdup
11 rpp -51.0 51.0 -201.0 201.0 -51.0 51.0 $ glovebox glass outer
130 rpp -49.75 49.75 -200.0 200.0 -49.75 49.75 $ glovebox holdup--fill
141 p -1 0 -1 97.0 $ glovebox corner s020 plane
142 p -1 0 1 97.0 $ glovebox corner s022 plane
143 p 1 0 -1 97.0 $ glovebox corner s220 plane
144 p 1 0 1 97.0 $ glovebox corner s222 plane
12 rpp 45.0 50.0 -2.5 2.5 -150 -51.0 $ front leg outer
13 rpp 45.5 49.5 -2.0 2.0 -150 -51.0 $ front leg inner
14 rpp -50.0 -45.0 -2.5 2.5 -150 -51.0 $ rear leg outer
15 rpp -49.5 -45.5 -2.0 2.0 -150 -51.0 $ rear leg inner
c
c pipe
20 rcc 0 -200.0 -110 0 400 0 15.8 $ wall outer
21 rcc 0 -200.0 -110 0 400 0 15 $ wall--holdup interface
221 p -1 0 -1 129.79898987322332 $ pipe corner s020 plane
222 p -1 0 1 -90.20101012677667 $ pipe corner r022 plane
223 p 1 0 -1 129.79898987322332 $ pipe corner r220 plane
224 p 1 0 1 -90.20101012677667 $ pipe corner r222 plane
c
c detector
500 rpp 240 260 -40 40 -40 40
c
c environment
30 pz -150 $ ground level
99 so 10000.0 $ edge of universe
```

Table A.3: MCNP5 data cards used in Section 3.5.

```
 print 10 30 35 100 110 126 140 170
phys:n
nps 1000000 1000000
c notrn $ notrn == source contributions only
cut:n j 0.1 $ cut below 0.1 MeV
c
c tally spec
c f5:n 250 0 0 0 $ x=250
c e5 0.1 20ilog 10 1000
fip15:n
........225 0 0 $ pinhole center
........0 $ place-holder
........224 0 0 $ grid axis alignment
........0 $ no collimation
........0 $ perfect pinhole
........25 $ grid--pinhole distance
fs15 -30 39i 30 $ fir grid s-axis (y-axis)
c15 -30 39i 30 $ fir grid t-axis (z-axis)
c fir25:n $
c 250 0 0 $ grid center
c 0 $ place-holder
c -1 0 0 $ grid plane normal
c 0 $ F1 include source and scatter contributions
c 0 $ F2 no field-of-view restriction
c 0 $ F3 no spatial jitter
c fs25 -30 2i 30 $ fir grid s-axis
c c25 -30 2i 30 $ fir grid t-axis
c
```

Table A.4: MCNP5 source cards used in Section 3.5.

```
 sdef par=n tme=d1 cel=ftme d2
........x=ftme d31 y=ftme d32 z=ftme d33
........erg=d4 eff=0.012464650561431558
si1 h 0 1 2 3 4 5 6 7 8 9 10 $ t is used to choose a source (sub)cell
sp1 0
........0.3524170050576154 0.35418794478152305
........0.02213674654884519 0.02213674654884519
........0.02213674654884519 0.02213674654884519
........0.0512120159913702 0.0512120159913702
........0.0512120159913702 0.0512120159913702
ds2 l
........130 130
........141 142 143 144
........221 222 223 224
ds31 s 310 311 312 313 314 315 316 317 318 319
ds32 s 321 321 322 322 322 322 326 326 326 326
ds33 s 331 330 332 333 334 335 336 337 338 339
c si4 50e-3 10
sp4 -3
c glovebox sides---no y-sides, overlap included in z-sides
si310 h -50.0 -49.75 49.75 50.0
sp310 0 0.25 0 0.25
si311 h -50.0 50.0
sp311 0 1
si321 h -200.0 200.0
sp321 0 1
si330 h -50.0 -49.75 49.75 50.0
sp330 0 0.25 0 0.25
si331 h -49.75 49.75
sp331 0 1
```

Table A.5: MCNP5 source cards used in Section 3.5 (continued).

```
 c glovebox corners
si312 -49.75 -47.25
sp312 0 1
si313 -49.75 -47.25
sp313 0 1
si314 47.25 49.75
sp314 0 1
si315 47.25 49.75
sp315 0 1
si322 -200.0 200.0
sp322 0 1
si332 -49.75 -47.25
sp332 0 1
si333 47.25 49.75
sp333 0 1
si334 -49.75 -47.25
sp334 0 1
si335 47.25 49.75
sp335 0 1
c pipe corners
si316 -13.70738148954362 -6.091608383679711
sp316 0 1
si317 -13.70738148954362 -6.091608383679711
sp317 0 1
si318 6.091608383679711 13.70738148954362
sp318 0 1
si319 6.091608383679711 13.70738148954362
sp319 0 1
si326 -200.0 200.0
sp326 0 1
si336 -123.70738148954361 -116.09160838367971
sp336 0 1
si337 -103.90839161632029 -96.29261851045639
sp337 0 1
si338 -123.70738148954361 -116.09160838367971
sp338 0 1
si339 -103.90839161632029 -96.29261851045639
sp339 0 1
c
```

Table A.6: MCNP5 material cards used in Section 3.5.

```
 c material spec per PNNL-15870Rev1.pdf, using default XS libraries
m1 $ stainless steel 304 at -8 g/cc
........6000.80c 0.001830
........14000.60c 0.009781
........15031.80c 0.000408
........16000.62c 0.000257
........24000.50c 0.200762
........25055.80c 0.010001
........26000.55c 0.690375
........28000.50c 0.086587
m21 $ lead glass at -6.22 g/cc
........8016.80c 0.592955
........14000.60c 0.174592
........22000.62c 0.010251
........33075.80c 0.002146
........82000.50c 0.220056
m3 $ Hanford dry concrete at -2.18 g/cc
........1001.80c 0.078440
........8016.80c 0.595591
........11023.80c 0.001864
........12000.62c 0.011462
........13027.80c 0.050831
........14000.60c 0.195330
........19000.62c 0.006577
........20000.62c 0.039567
........26000.55c 0.020338
m4 $ plutonium nitrate at -2.447 g/cc
........7014.80c 0.235294
........8016.80c 0.705882
........94238.80c 0.000030
........94239.80c 0.055016
........94240.80c 0.003516
........94241.80c 0.000233
........94242.80c 0.000029
m5 $ dry sea-level air at -0.001205 g/cc
........6000.80c 0.000150
........7014.80c 0.784431
........8016.80c 0.210748
........18000.35c 0.004671
```

BIBLIOGRAPHY

[1] H. Nakamura, Y. Mukai, and T. Kurita, "Development of advanced MOX holdup measurement technology for improvement of MC&A and safeguards," Report IAEA-CN–220, International Atomic Energy Agency, Oct. 2014.

[2] H. Nakamura, Y. Mukai, and T. Kurita, "Development of advanced MOX holdup measurement technology for improvement of MC&A and safeguards." presentation slides, Oct. 2014.

[3] K. C. Bledsoe and M. A. Jessee, "Final report for ORNL12-RS-012L: quantitative uncertainty analysis for inverse transport methods," tech. rep., Oak Ridge National Lab, Sep 2013.

[4] K. C. Bledsoe, J. P. Lefebvre, and R. A. Lefebvre, "Implementation of parallel evolutionary algorithms in a prototype of scale for ORNL12-RS-012: quantitative uncertainty analysis for inverse transport methods," Tech. Rep. ORNL/SPR-2015/178, Oak Ridge National Lab, 2015.

[5] T. D. Reilly, "Nondestructive assay of holdup," Tech. Rep. LA-UR-07-5149, Los Alamos National Lab, 2007.

[6] R. Venkataraman *et al.*, "Validation of in situ object counting system (ISOCS) mathematical efficiency calibration software," *Nuclear Instruments and Methods in Physics Research A*, vol. 422, pp. 450–454, 1999.

[7] W. Karush, "Minima of functions of several variables with inequalities as side constraints," m.s. dissertation, Univ. of Chicago, Chicago, Illinois, 1939.

[8] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Proceedings of 2nd Berkeley Symposium*, pp. 481–492, University of California Press, 1951.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[10] R. P. Gardner and L. Xu, "Status of the Monte Carlo library least-squares (mclls) approach for non-linear radiation analyzer problems," *Radiation Phys. and Chemistry*, vol. 78, pp. 843–851, 2009.

[11] X-5 Monte Carlo Team, *MCNP - Version 5, Vol. I: Overview and Theory*, 2003. LA-UR-03-1987.

[12] *MCNPX Users Manual Version 2.7.0*, 2011. LA-CP-11-00438.

[13] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *Medical Imaging, IEEE Transactions on*, vol. 1, no. 2, pp. 113–122, 1982.

[14] S. R. Gottesman and E. E. Fenimore, "New family of binary arrays for coded aperture imaging," *Applied Optics*, vol. 28, pp. 4344–4352, Oct. 1989.

[15] M. S. Wyatt, T. Uckan, and T. F. Hannon, "Physics of enriched uranyl fluoride deposit characterizations using active neutron and gamma interrogation techniques with 252Cf," Tech. Rep. ORNL/CP-98771, Oak Ridge National Lab, Long Island, NY, Oct 1998. presented at ICPNST International Conference.

[16] P. Hausladen, J. Newby, F. Liang, and M. Blackston, "The deployable fast-neutron coded-aperture imager: demonstration of locating one or more sources in three dimensions," Tech. Rep. ORNL/TM-2013/248, Oak Ridge National Laboratory, 2013.

[17] P. J. Green, "Reversible jump markov chain monte carlo computation and bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[18] T. Iwamoto, "Hold-up measurement in a reprocessing facility." presentation slides, 2006.

[19] R. Hagenauer and R. Mayer, "Lessons learned from the holdup measurement program at the K-25 site." presentation slides, 2006.

[20] B. Rasmussen, M. Clapham, and R. E. James, "In situ measurement in low enrichment uranium holdup process gas piping at K-25," Tech. Rep. BJC/OR-3357/R1, Bechtel Jacobs, Oak Ridge, TN, 2010. paper for Waste Management Symposia.

[21] D. L. Mozhayev, G. M. Mapili, and A. V. Mozhayev, "Plutonium hold-up measurements at Pacific Northwest National Laboratory." presentation slides, 2006.

[22] J. Sprinkle, "Holdup measurement uncertainty." presentation slides, Oct 2006. presented at the INMM Workshop on Best Practices for Material Holdup Monitoring.

[23] P. Hausladen, M. Blackston, F. Liang, and J. Newby, "Measurements and analysis of mock holdup configurations using fast-neutron imaging," technical memo ORNL/TM02012/400, ORNL, 2012.

[24] T. D. Jackson, *Quantification of fast-neutron sources with coded aperture imaging*. PhD thesis, University of Tennessee, Knoxville, Dec 2015.

[25] A. M. Bevill and K. C. Bledsoe, "Uncertainty quantification for quantitative imaging holdup measurements," in *Proceedings of the INMM 57th Annual Meeting*, Institute of Nuclear Materials Management, Jul 2016.

[26] A. M. Bevill and W. R. Martin, "Model validation for fast-neutron coded aperture quantitative imaging," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2016 IEEE*, IEEE, 2016.

[27] D. F. Yu and J. A. Fessler, "Mean and variance of single photon counting with deadtime," *Phys. Med. Biol.*, vol. 45, pp. 2043–2056, 2000.

[28] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machinesequation," *J. Chem. Phys.*, vol. 21, no. 1087, 1953.

[29] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, p. 97, 1970.

[30] S. Chib and E. Greenberg, "Understanding the metropolis–hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.

[31] A. Tarantola, *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.

[32] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions*, vol. 1. Wiley, 1994.

[33] B. Efron, *The jackknife, the bootstrap, and other resampling plans*. Society for Industrial and Applied Mathematics, 1982.

[34] M. C. Fleenor, K. P. Ziock, and M. A. Blackston, "Bootstrapped uncertainties in coded-aperture images." poster, Oct 2016. poster presented at Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC).

[35] K. P. Ziock and M. A. Blackston, "Fun with blocks, gamma-ray images of different source configurations." presentation, Nov 2016. presentation at Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC).

[36] S. Weerahandi, "Generalized confidence intervals," *Journal of the American Statistical Association*, vol. 88, pp. 899–905, Sep 1993.

[37] P. Hausladen, P. Bingham, J. Neal, J. Mullens, and J. Mihalczo, "Portable fast-neutron radiography with the nuclear materials identification system for fissile material transfers," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 261, no. 1, pp. 387–390, 2007.

[38] B. Perot, C. Carasco, C. Deyglun, C. Eleon, W. El Kanawati, A. Mariani, and J.-L. Ma, "Materials characterisation with the associated particle technique," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, pp. 1702–1711, IEEE, 2012.

[39] R. Accorsi, *Design of near-field coded aperture cameras for high-resolution medical and industrial gamma-ray imaging*. PhD thesis, MIT, Jun 2001.

[40] R. Accorsi, F. Gasparini, and R. C. Lanza, "Optimal coded aperture patterns for improved SNR in nuclear medicine imaging," *Nuclear Instruments and Methods in Physics Research A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 474, no. 3, pp. 273–284, 2001.

[41] M. C. Fleenor, M. A. Blackston, and K. P. Ziock, "Correlated statistical uncertainties in coded-aperture imaging," *Nuclear instruments and methods in physics research A*, vol. 784, pp. 370–376, 2015.

[42] J. Fessler, "Emission ML image reconstruction." Chap. 16 of an unpublished textbook draft, Jun 2015.

[43] A. O. Hero and J. A. Fessler, "Convergence in norm for alternating expectation-maximization (EM) type algorithms," *Statistica Sinica*, vol. 5, pp. 41–54, 1995.

[44] A. Sitek, "Representation of photon limited data in emission tomography using origin ensemblesrepresentation," *Phys. Med. Biol*, vol. 53, pp. 3201–3216, 2008.

[45] "CBRNE training/exercise event:GeGI imaging measurements." Unpublished slides accessed 2016-12-05, 2016.

[46] E. Hull and M. Kiser, "Radioactive threat vision via quantitative gamma-ray imaging," tech. rep., PHDs Corporation, 3011 Amherst Road, Knoxville, TN 37921, USA, 2016. accessed 2016-12-05.

[47] E. Hull, "Nuclear visualization via quantitative imaging and detection." presentation for the American Nuclear Society Oak Ridge/Knoxville Section, Nov 2016.

[48] D. I. Goodman and Z. He, "Compton and coded aperture radiation image reconstruction using stochastic origin ensembles in 3-D position sensitive CdZnTe detectors." presentation at IEEE RTSD conference in Strasbourg, France, Nov 2016.

[49] K. C. Bledsoe, J. A. Favorite, and T. Aldemir, "Application of the Differential Evolution method to solving inverse transport problems," *Nuclear Science and Engineering*, vol. 169, pp. 208–221, 2011.

[50] K. C. Bledsoe, J. A. Favorite, J. P. Lefebvre, R. A. Lefebvre, and M. A. Jessee, "Application of the DiffeRential Evolution Adaptive Metropolis (DREAM) method for uncertainty quantification in inverse transport problems," in *Trans. of the Am. Nucl. Soc.*, vol. 111, pp. 743–746, 2014.

[51] J. Mattingly and D. J. Mitchell, "A framework for the solution of inverse radiation transport problems," *IEEE Transactions on Nuclear Science*, vol. 57, no. 6, pp. 3734–3743, 2010.

[52] M. Streicher, S. Brown, Y. Zhu, D. Goodman, and Z. He, "A method to estimate the atomic number and mass thickness of intervening materials in uranium and plutonium gamma-ray spectroscopy measurements," *IEEE Transactions on Nuclear Science*, vol. 63, no. 5, pp. 2639–2648, 2016.

[53] J. T. Mihalczo, "Holdup characterization for a 24-in.-OD pipe at the K-29 plant using Cf-source-driven transmission measurements." presentation slides, 2006. presented at the International Workshop on Best Practices For Material Holdup Monitoring.

[54] D. H. Beddingfield, T. L. Burr, and C. E. Longo, "Error evaluation in distributed source-term analysis," Tech. Rep. LA-UR-10-01947, Los Alamos National Laboratory, 2010.

[55] P. Hausladen, J. Newby, F. Liang, and M. Blackston, "A deployable fast-neutron coded-aperture imager for quantifying nuclear material," Tech. Rep. ORNL/TM-2013/446, Oak Ridge National Laboratory, 2013.

[56] H. O. Anger, "Scintillation camera," *Review of Scientific Instruments*, vol. 29, p. 27, Jan 1958.

[57] R. A. Winyard, J. E. Lutkin, and G. W. McBeth, "Pulse shape discrimination in inorga nic and organic scintillators," *Nuclear Instruments and Methods*, vol. 95, pp. 141–153, 1971.

[58] SciPy.org, *scipy.optimize.minimize*, 0.18.1 ed., 2016. https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html accessed 2016-11-15.

[59] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, pp. 308–313, 1965.

[60] W. M. H, "Direct search methods: Once scorned, now respectable, in numerical analysis," in *Numerical Analysis 1995: Proceedings of the 1995 Dundee Biennial Conference in Numerical Analysis* (D. F. Griffiths and G. A. Watson, eds.), (Harlow, UK), pp. 155–162, Addison Wesley Longman, 1995.

[61] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2 (10th printing) ed., 2009.

[62] J. A. Fessler, "Mean and variance of implicitly defined bias estimators (such as penalized maximum likelihood): applications to tomography," *IEEE Trans on Image Processing*, vol. 5, p. 493, Mar 1996.

[63] J. W. Stayman and J. A. Fessler, "Fast methods for approximation of resolution and covariance for spect," in *Nuclear Science Symposium Conference Record*, vol. 2, pp. 786–788, IEEE, Nov 2002.

[64] J. W. Stayman and J. A. Fessler, "Efficient calculation of resolution and covariance for penalized-likelihood reconstruction in fully 3-d spectefficient," *IEEE Trans on Medical Imaging*, vol. 23, p. 1543, Dec 2004.

[65] S. Ferson, L. Ginzburg, V. Kreinovich, and M. Aviles, "Exact bounds on sample variance of interval data," in *Extended Abstracts of the 2002 SIAM Workshop on Validated Computing*, pp. 67–69, May 2002.

[66] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, and L. Ginzburg, "Experimental uncertainty estimation and statistics for data having interval uncertainty," Tech. Rep. SAND2007-0939, Sandia National Laboratories, May 2007.

[67] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*. Society for Industrial and Applied Mathematics, 2009.

[68] R. C. Smith, *Uncertainty quantification: theory, implementation, and applications*. Society for Industrial and Applied Mathematics, 2013.

[69] F. S. Hillier and G. J. Lieberman, *Introduction to operations research*. New York, NY: McGraw-Hill, tenth ed., 2015.

[70] R. Bhatia, *Positive definite matrices*. Princeton University Press, 2007.

[71] R. J. Vanderbei, *Linear Programming Foundations and Extensions*. Springer, 2014.

[72] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*. Springer, 2016.

[73] P.-Q. Pan, *Linear Programming Computation*. Springer, 2014.

[74] B. A. Mair, M. Rao, and J. M. M. Anderson, "A minimum chi-squared method for indirect parameter estimation from poisson data," *Statistica Neerlandica*, vol. 56, no. 2, pp. 165–178, 2002.

[75] R. McConn Jr, C. J. Gesh, R. T. Pagh, R. A. Rucker, and R. Williams III, "Compendium of material composition data for radiation transport modeling," Tech. Rep. PNNL-15870, Pacific Northwest National Laboratory, 2011.

[76] J. C. Mark, F. von Hippel, and E. Lyman, "Explosive properties of reactor-grade plutonium," *Science & Global Security*, vol. 17, Dec 2009.