

SANDIA REPORT

SAND2021-3972
Printed March 2021



Sandia
National
Laboratories

Analysis and Optimization of Seismo-Acoustic Monitoring Networks with Bayesian Optimal Experimental Design

Thomas A. Catanach and Kevin Monogue

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185
Livermore, California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: <http://www.osti.gov/scitech>

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: <https://classic.ntis.gov/help/order-methods>



Analysis and Optimization of Seismo-Acoustic Monitoring Networks with Bayesian Optimal Experimental Design

Thomas A. Catanach
Extreme Scale Data Science & Analytics
Sandia National Laboratories
P.O. Box 969
Livermore, CA 94551-0969
tacatan@sandia.gov

Kevin Monogue
Institute for Computational & Mathematical Engineering
Stanford University
Stanford, CA 94306
kcmونogue@gmail.com

SAND2021-3972

ABSTRACT

The Bayesian optimal experimental design (OED) problem seeks to identify data, sensor configurations, or experiments which can optimally reduce uncertainty. The goal of OED is to find an experiment that maximizes the expected information gain (EIG) about quantities of interest given prior knowledge about expected data. Therefore, within the context of seismic monitoring, we can use Bayesian OED to configure sensor networks by choosing sensor locations, types, and fidelity in order to improve our ability to identify and locate seismic sources. In this work, we develop the framework necessary to use Bayesian OED to optimize the ability to locate seismic events from arrival time data of detected seismic phases. In order to do utilize Bayesian OED we must develop four elements:

1. A likelihood function that describes the uncertainty of detection and travel times,
2. A Bayesian solver that takes a prior and likelihood to identify the posterior,
3. An algorithm to compute EIG,
4. An optimizer that finds a sensor network which maximizes EIG.

Once we have developed this framework, we can explore many relevant questions to monitoring such as: how and what multiphenomenology data can be used to optimally reduce uncertainty, how to trade off sensor fidelity and earth model uncertainty, and how sensor types, number, and locations influence uncertainty.

ACKNOWLEDGMENT

This research was funded by the U.S. Department of Energy. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003535. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government. This research was funded by the National Nuclear Security Administration, Defense Nuclear Nonproliferation Research and Development (NNSA DNN R&D). The authors acknowledge the important interdisciplinary collaboration with scientists and engineers from LANL, LLNL, MSTS, PNNL, and SNL. We would like to thank Lisa Linville, Joshua Carmichael, and Chris Young for their help and advice in developing this research and feedback while preparing this manuscript. We would also like to thank Ariel Lellouch and Biondo Biondi for their help on the Stanford ICME Xplore project that seeded this work.

CONTENTS

| | |
|---|----|
| Nomenclature | 11 |
| 1. Overview | 13 |
| 2. Bayesian Methods | 14 |
| 2.1. Bayesian Inference | 14 |
| 2.2. Bayesian Optimal Experimental Design | 16 |
| 2.3. Optimization for Bayesian OED | 17 |
| 3. Bayesian Seismic Monitoring | 19 |
| 3.1. General Approach | 19 |
| 3.2. Detection Model | 20 |
| 3.3. Arrival Time Model | 22 |
| 3.4. Integrating Earth model Uncertainty | 24 |
| 3.5. Travel Time Correlation | 25 |
| 4. Computational Approach | 27 |
| 4.1. Estimating Information Gain | 27 |
| 4.2. Optimization | 28 |
| 5. Results | 29 |
| 5.1. Effect of Measurement Noise | 29 |
| 5.2. Effect of Correlation | 31 |
| 6. Conclusion | 35 |
| References | 37 |

LIST OF FIGURES

| | |
|--|----|
| Figure 2-1. Illustration of the Bayesian inference process for seismic source location. Bayesian inference begins with a prior distribution for different earthquake locations θ (left). As an observer collects data, they use a likelihood function model to quantify the probability of observing that data, given that an earthquake occurs at a specific location. The observer constructs the likelihood model from physical models of seismic wave propagation, models of the sensors that detect seismic signals, and models of uncertainty (e.g. background noise modeling errors, etc.). The observer then applies Bayes' theorem to update the prior to assimilate this new information. The posterior distribution (right) then quantifies the probability that the seismic source has location θ , given the data. | 15 |
| Figure 2-2. Sensor locations that optimize a sensor network configuration, according to Equation 4. In this example, we place the first sensor at the center of the domain (top, left) and compute optimal locations for the next four subsequent sensors to maximize expected information gain (EIG). White circular markers define the initial sensor configuration, and magenta circular markers define new sensor locations to optimally augment the network. The top row shows the optimization surface that describes the utility of adding a sensor at a given location to the initial network; contours show the EIG, averaged over all event locations. The warmer the color indicates that adding a sensor at that location is better. The bottom row shows the EIG for the augmented sensor network, where the color contours show the EIG about the location of a shallow seismic source at that specific latitude and longitude. | 18 |
| Figure 3-1. Map of Transportable Array stations from December 2007. The pink region indicates the region (Latitude $\in [39^{\circ}N, 43^{\circ}N]$, Longitude $\in [113^{\circ}W, 107.36^{\circ}W]$) that we use gathered sensor data that populates the parameters of the likelihood models. The orange region (Latitude $\in [40^{\circ}N, 42^{\circ}N]$, Longitude $\in [112^{\circ}W, 109.36^{\circ}W]$), corresponds to the monitoring region that we build a sensor network over..... | 20 |
| Figure 3-2. The estimated detection probability from the Transportable Array dataset. We binned these data by distance and magnitude and then estimated the detection probability as the number of detections versus the number of potential detections for stations and events within a given distance and magnitude bin..... | 22 |
| Figure 3-3. Arrival time residual as a function of distance in kilometers. Note that there is a bias towards positive residuals, particularly at longer distances. This bias is particularly evident in Figure 3-4. | 23 |
| Figure 3-4. Arrival time residual histogram compared to the fit of different statistical models. The magenta line indicates the non-centered t-distribution fit to the data. The orange line shows the marginal residual distribution under the more tractable distance dependent Gaussian model discussed in Section 3.4. Note that the non-data driven model does not a priori know the bias so it is centered at zero. | 23 |
| Figure 3-5. Illustration of the 121 1D velocity models for V_p sampled from around the monitoring region. These representative earth models are used to estimate travel time uncertainty from earth model uncertainty..... | 24 |

| | |
|--|----|
| Figure 3-6. <i>Left:</i> A scatter plot of the estimated travel time mean $\mu(\Delta, z)$, and estimated travel time standard deviation, $\sigma(\Delta, z)$ for various distance and depth pairs, superimposed with a fit polynomial model (orange curve). <i>Right:</i> A composite scatter plot of the travel time residuals for all sampled distance, depth, and earth models compared to the 2σ range defined by the fitted polynomial model. The simple polynomial model adequately captures the bulk trend, despite some variability; a few very high residual points inflate the standard deviation away from typical residual values. | 25 |
| Figure 3-7. Visualization of the correlation matrix. The left figure is the correlation matrix, Γ , induced by the earth model uncertainty computed using Equation 12. The axes correspond to the distance along the surface from the source epicenter in km. Stations are approximately spaced 33km apart. The right figure is a simplified model of the correlation, Γ_{GP} , found using a square-exponential kernel, which approximates Γ . The square-exponential kernel assumes that correlation is only a function of the distance between stations. This simplified model reasonably captures the correlation length scale for Γ but is unable to capture the complex, non-translation invariant, block structure of Γ | 26 |
| Figure 5-1. Illustration of the effect of changing the measurement uncertainty standard deviation over several orders of magnitude. The color plots illustrate the EIG of a shallow seismic source with the different stated measurement errors. The line plot illustrates the degradation of EIG for all events as the measurement error increases. For noise levels below about 0.2 seconds the model uncertainty dominates over measurement uncertainty so EIG is fairly stable. | 29 |
| Figure 5-2. The left panel illustrates the change in EIG as additional sensors are placed using greedy optimization for two different networks with different sensor fidelities. One network has sensors with a measurement error standard deviation of 0.1 seconds and the other has sensors with error of 0.4642 seconds. For this the 20 station noisy network is about equivalent to the 11 station high fidelity network. The right panel describes the geometry of the networks based upon how close the stations are to each other. We can see that, particularly in the beginning, stations of the noisy network are being added further apart than stations of the less noisy network. | 30 |
| Figure 5-3. Depiction of the evolution of the sensor placement and expected information gain about surface events given the two different noise models and different numbers of sensors. | 32 |
| Figure 5-4. The left panel illustrates the evolution of the expected information gain as the number of sensors increases for the three different correlation models where the correlation changes over two orders of magnitude. We see that initially the disparity in EIG is small then increases with the number of sensors. However, after about 10 sensors, the disparity begins to decrease. The right panel investigates the geometry of the network by looking at the distances between stations. Few patterns can be identified in this plot although it may be the case that stations are initially further apart for the low correlation model. | 33 |

Figure 5-5. Depiction of the evolution of the sensor placement and expected information gain about surface events given the three different correlation models and different numbers of sensors. The spacing of sensors in all cases appears to be fairly regular..... 34

LIST OF TABLES

| | |
|--------------------------------------|----|
| Table 3-1. Detection Weighting | 22 |
|--------------------------------------|----|

NOMENCLATURE

EIG Expected Information Gain

GP Gaussian Process

KL divergence Kullback-Leibler divergence

OED Optimal Experimental Design

QMC Quasi Monte Carlo

QoI Quantity of Interest

STD Standard Deviation

1. OVERVIEW

Seismo-acoustic monitoring networks provide an important tool for detecting nuclear tests. In order to improve monitoring capabilities and improve decision making, network designers may incorporate new sensors or data types into the network to reduce detection thresholds or improve estimate uncertainties for quantities of interest (QoIs). These QoIs include location, magnitude, or discriminants to separate earthquakes from explosions. To estimate a QoI, modern processing algorithms often employ Bayesian inference because it provides rigorous uncertainty quantification to support decision making. Therefore, when designing or analyzing a monitoring network, we may take a Bayesian approach called Bayesian Optimal Experimental Design (OED) [1, 2]. Within this framework, we optimize sensors within a monitoring network to reduce uncertainty about QoIs under different conditions. Therefore, with Bayesian OED we not only design an effective monitoring network, but also get an understanding of the expected performance of that network under the studied conditions. While the target application of this research is nuclear explosion monitoring, the experimental design framework we have developed applies to arbitrary seismic sources. Therefore, this framework may support other applications of seismic networks such as earthquake seismology or exploration geophysics.

This work quantifies the sensitivity of a seismo-acoustic monitoring network for inferring the location of seismic sources that include shallow earthquakes and explosions either on the surface or underground. We then present a Bayesian OED algorithm to improve the monitoring network sensitivity by optimizing the location of ground motion sensors (seismometers). Our computational approach combines information and Bayesian probability theory to quantify and optimize the sensitivity of our sensor network by estimating the information gain Bayesian inference provides about QoIs. This approach includes four distinct analysis stages:

1. Build the likelihood function to estimate the probability of data, given a seismic event.
2. Solve the Bayesian inference problem for locating events given data (e.g., solve for the posterior).
3. Estimate seismic source location sensitivity through measuring the expected information gain, with a sensor network.
4. Optimize the seismic monitoring network to improve the expected information gain over seismic source events.

We use observational data from the U.S. Transportable Array and physics-based models to build the Bayesian likelihood functions. These likelihood functions incorporate many sources of uncertainty and model the behavior of the seismic sensor network. This model defines how well Bayesian inference can assimilate sensor data to locate seismic sources. For our sensor data, we consider spatially correlated travel-times for seismic phase arrivals detected at our sensor network. We make generally justifiable assumptions on our noise models that match normalized histograms from the data that we consider. While we present our method using seismic P-wave arrivals, our approach is flexible to any event or signature data because it only requires that we can construct likelihood functions.

We study how the optimized sensor configuration and network sensitivity change under different design conditions and uncertainty models. We present the dependence of our solutions over sensor distribution, sensor number, and sensor fidelity. This analysis provides a framework that we will later extend to optimize sensor networks that measure other natural and explosion signatures (e.g., radionuclide or acoustic signals), which supports a more comprehensive need for multi-phenomenology explosion monitoring (e.g., [3, 4]). This framework will improve upon existing tools, like Sandia’s NetMOD [5], to provide a highly flexible and rigorous framework for analyzing and optimizing monitoring networks. This rigor is justified through our usage of Bayesian probability theory.

In our overview, we first describe the Bayesian inference and optimal experimental design problems in general. We then describe the specifics of a Bayesian inference problem to identify the location of sources, using records of their seismic-phase arrivals at distributed receivers, and demonstrate how to build the likelihood models from these data. Next, we describe the algorithms used for solving the Bayesian OED problem. Finally, we will describe several experiments that demonstrate the utility of this approach and identify areas for further exploration.

2. BAYESIAN METHODS

Bayesian probability theory provides a rigorous methodology for quantifying uncertainty [6, 7]. Within this framework, uncertainty is represented using probability distributions. Therefore, within the Bayesian view, probability represents uncertainty and not intrinsic stochasticity and thus is not tied directly to randomness. Uncertainty can come from both epistemic uncertainty, which represents a lack of knowledge about learnable phenomena (ignorance), or aleatory uncertainty, when it represents uncertainty about inherently unknowable randomness (limitations to observer knowledge). The Bayesian perspective treats these sources of uncertainty identically using a probability distribution.

As data or other sources of information become available, an observer can integrate such information into the probability distributions to update observer uncertainty. When this observer then makes predictions, they include the uncertainty represented by these probability distribution in these predictions. The rules of Bayesian probability provide the logic for updating and propagating uncertainty, just as binary logic provides rules for working with statements that are true or false. In this way, Bayesian theory extends propositional logic to situations where there is uncertainty and statements are described as having a certain probability of being true or false.

2.1. Bayesian Inference

The process of updating beliefs using data is known as Bayesian inference. Figure 2-1 illustrates Bayesian inference applied to a hypothetical seismic location problem (see [8, 9] for detailed applications of Bayesian inference to locating seismic sources). Bayesian inference begins by expressing prior beliefs about parameters of interest θ . For example, within the context of identifying characteristics of a seismic event, these beliefs may represent prior knowledge about

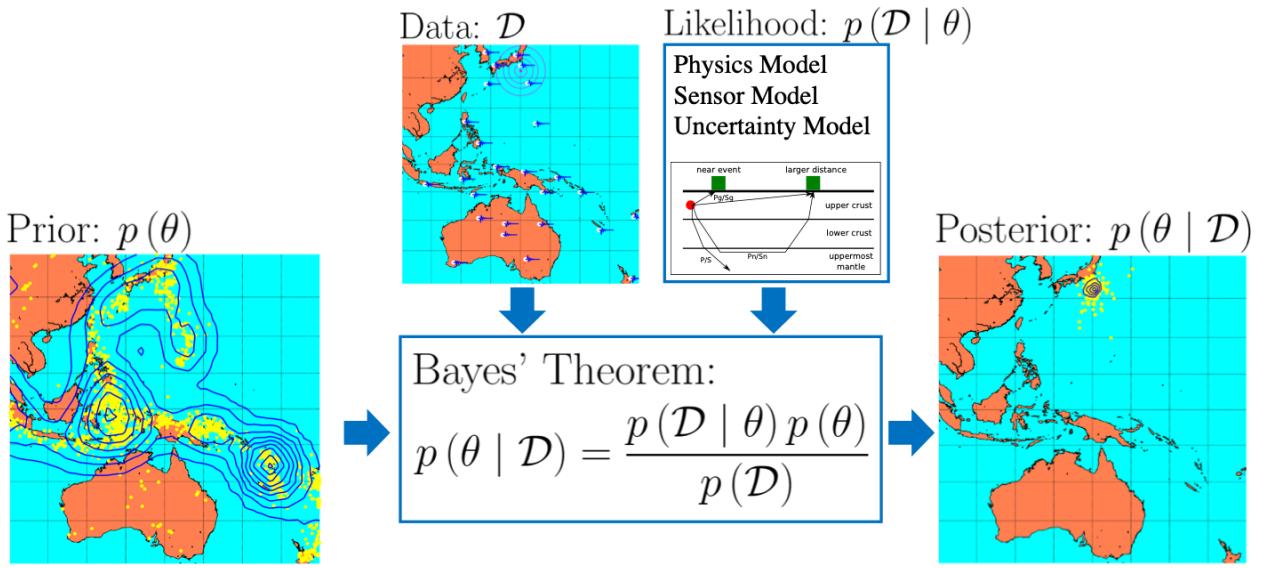


Figure 2-1 Illustration of the Bayesian inference process for seismic source location. Bayesian inference begins with a prior distribution for different earthquake locations θ (left). As an observer collects data, they use a likelihood function model to quantify the probability of observing that data, given that an earthquake occurs at a specific location. The observer constructs the likelihood model from physical models of seismic wave propagation, models of the sensors that detect seismic signals, and models of uncertainty (e.g. background noise modeling errors, etc.). The observer then applies Bayes' theorem to update the prior to assimilate this new information. The posterior distribution (right) then quantifies the probability that the seismic source has location θ , given the data.

the distribution of earthquake magnitudes or their locations e.g. source proximity to lithospheric faults. As an observer gathers data \mathcal{D} and other information, they update the prior distribution using the rules of probability. This updated, or posterior distribution $p(\theta | \mathcal{D})$, now quantifies the likelihood of the source location, given the data. The observer performs this updating from a likelihood function to describe the probability of the data, given an event hypothesis, i.e. $p(\mathcal{D} | \theta)$. An observer can construct such a likelihood function as a probabilistic forward model of the data from a set of source parameters. The likelihood function includes source parameters that describe the seismic source and then uses physical models, sensor models, and models of background signals or noise processes to map this source description to plausible sensor data. As an example, if the arrival time of a seismic phase at a seismometer constitutes observed data, and event parameters describe the location and origin time of an earthquake, then the likelihood uses an earth structure model to predict uncertainty in the travel time of a seismic phase from the source to any receivers. The model of the measured travel time then includes (predicted) earth model uncertainty and measurement uncertainty on the sensor.

Once an observer constructs such a likelihood function, they can easily construct the posterior

distribution on events given data. This construction is an application of Bayes' Theorem:

$$p(\theta | \mathcal{D}) = \frac{p(\mathcal{D} | \theta) p(\theta)}{p(\mathcal{D})} \quad (1)$$

We emphasize that the probability terms in Equation 1 can be either probabilities when θ is a discrete random variable or a probability density when θ is continuous. Equation 1 provides the foundational statement of belief, and the machinery to update these beliefs as new information becomes available. In practice, solving for the updated Bayesian posterior requires approximate computational methods since the posterior may not have an analytical expression. Common approaches generate samples that represent the posterior distribution and can estimate QoIs. Examples of these methods include importance sampling using Monte Carlo, Quasi Monte Carlo, or meshing and Markov Chain Monte Carlo [10].

To quantify sensor performance, we require a measure of how much belief changes due to inference on observed data. This is a measure of the sensor data's utility. To perform this measure, we apply the Kullback-Leibler (KL) divergence from information theory:

$$\text{KL}[p(\theta | \mathcal{D}) || p(\theta)] = \int p(\theta | \mathcal{D}) \log \frac{p(\theta | \mathcal{D})}{p(\theta)} d\theta \quad (2)$$

The KL divergence in Equation 2 measures how many units of information (bits for \log_2 or nats for \ln) lead to a change in the distribution from $p(\theta)$ to $p(\theta | \mathcal{D})$. A KL divergence of 0 means that the distributions are the same, side-stepping certain technical details. The KL divergence is always non-negative and as it increases from zero, Equation 2 implies that the distributions increasingly differ. A relatively large KL divergence therefore indicates that the data was very informative, and the prior and posterior are measurably distinct.

2.2. Bayesian Optimal Experimental Design

We now consider the Bayesian optimal experimental design (OED) problem [2, 11, 12] built upon the concepts of Bayesian probability and information theory. Our goal requires an observer that applies Bayes' Theorem (a Bayesian agent) to select a sensor configuration \mathcal{S} that maximizes utility; we term \mathcal{S} as the “experiment”. For Bayesian OED, the Bayesian agent uses a utility function that maximizes the expected information gain (EIG) from the prior to the posterior, in the view of the posterior. The expectation $E_{\mathcal{D}|\mathcal{S}}$ operator that computes this EIG notationally indicates that the observer computes the expectation with respect to the prior distribution of hypothetical data from the experiment given by $p(\mathcal{D} | \mathcal{S})$. The expected information gain for a specific experimental configuration is (from Equation 2):

$$\begin{aligned} I(\mathcal{S}) &= E_{\mathcal{D}|\mathcal{S}}[\text{KL}[p(\theta | \mathcal{D}, \mathcal{S}) || p(\theta)] | \mathcal{D}] \\ &= \int p(\mathcal{D} | \mathcal{S}) \int p(\theta | \mathcal{D}, \mathcal{S}) \log \frac{p(\theta | \mathcal{D}, \mathcal{S})}{p(\theta)} d\theta d\mathcal{D} \end{aligned} \quad (3)$$

The first integral in Equation 3 is the expectation over the hypothetical data from the experiment, while the second integral computes the KL divergence, given a realization of the hypothetical data. To compute the EIG in practice, we express $p(\mathcal{D} | \mathcal{S})$ as the marginal distribution $p(\mathcal{D} | \mathcal{S}) = \int p(\mathcal{D} | \theta', \mathcal{S}) p(\theta') d\theta'$, because it is often only implicitly known by definition of the likelihood and prior over parameters θ' . We then draw samples from the marginal distribution by first sampling the prior, and then sampling the data according to the likelihood.

We now maximize $I(\mathcal{S})$ to estimate the best experimental design \mathcal{S}^* from $\mathcal{S} \in \mathbb{S}$, where \mathbb{S} is the set of possible designs under consideration:

$$\mathcal{S}^* = \operatorname{argmax}_{\mathcal{S} \in \mathbb{S}} I(\mathcal{S}) \quad (4)$$

This optimization is generalizable to include constraints that include, for example, a sensor budget through methods like Lagrange multipliers or nonlinear programming. Further, while we have formulated this problem through maximizing the EIG for the posterior, we could more specifically optimize EIG about a specific quantity of interest derived from its parameters.

Solving this optimization problem is challenging because it requires solving many Bayesian inference problems for many hypothetical realizations of data from many hypothetical sensor configurations. This nested complexity means that significant care must be taken to make this approach computationally tractable.

2.3. Optimization for Bayesian OED

Greedy optimization algorithms provide an effective computational solution to sequentially place sensors in OED and other network optimization problems [1, 13]. Such greedy optimization involves sequentially adding sensors one at a time so that the optimization problem at a particular iteration is low dimensional, and therefore only requires updating the location of that particular sensor. During iteration, the algorithm computes the EIG from the an average over all possible source locations specified by the prior, and then computes an optimal location. Figure 2-2 illustrates the process of adding sensor one-by-one. The optimization surfaces, shown in the top row of the figure, start out fairly symmetric with multiple optima when there are few sensors. However, as more sensors are added these symmetries are broken so there is a unique optimal location of the next sensor. The bottom row illustrates how the EIG increases with sensor density, particularly about sources that are near several sensors.

We concede that the true, optimal sensor network configuration requires that we compute sensor location solutions all at once. However, greedy optimization often does reasonably well with a significantly reduced computational cost. Suboptimal bounds, in fact, exist for certain classes of optimization problems. One such class is that of submodular functions. At a high level, submodular utility functions exhibit diminishing returns with each iteration, that is, adding a sensor to a smaller network yields higher gains than adding a sensor to a larger network. The EIG objective for Bayesian OED is submodular when the sensors are independent (or more precisely, conditionally independent given the event description). If our utility function is submodular, then we can show that the greedy optimum will be near-optimal, meaning that there is a multiplicative

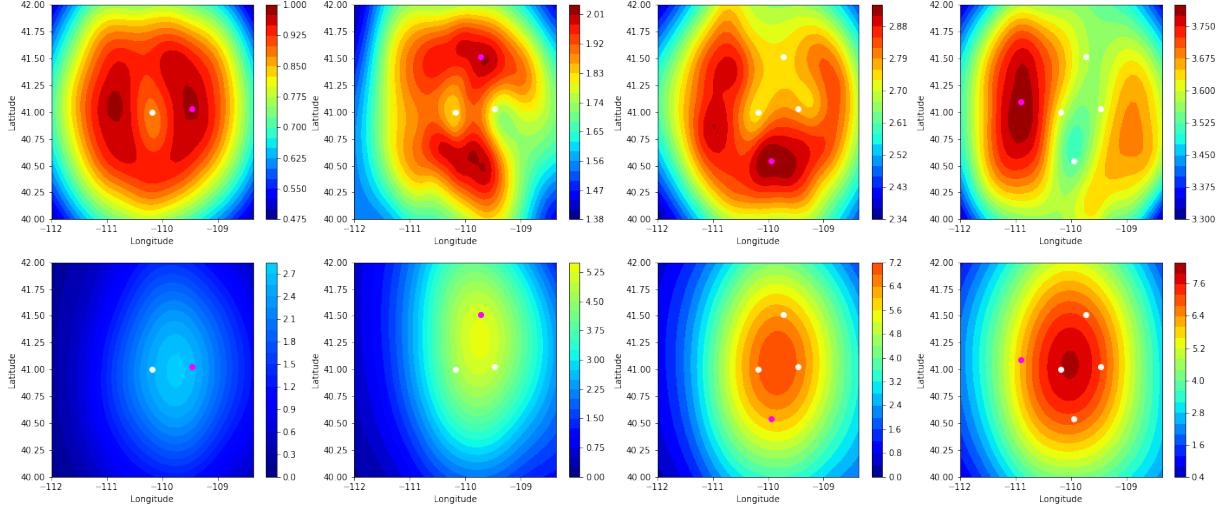


Figure 2-2 Sensor locations that optimize a sensor network configuration, according to Equation 4. In this example, we place the first sensor at the center of the domain (top, left) and compute optimal locations for the next four subsequent sensors to maximize expected information gain (EIG). White circular markers define the initial sensor configuration, and magenta circular markers define new sensor locations to optimally augment the network. The top row shows the optimization surface that describes the utility of adding a sensor at a given location to the initial network; contours show the EIG, averaged over all event locations. The warmer the color indicates that adding a sensor at that location is better. The bottom row shows the EIG for the augmented sensor network, where the color contours show the EIG about the location of a shallow seismic source at that specific latitude and longitude.

approximation guarantee of $(1 - 1/e)$. Therefore, the greedy optimum is guaranteed to be within approximately 63% of the global optimum. This bound is loose in practice, and stronger assumptions on the problem structure may yield smaller optimality guarantees. We refer to [1] for details on submodular functions and greedy optimization.

We use Bayesian optimization [14] to optimize sensor placement locations. This requires sampling the utility function to build a surrogate model of the optimization surface from the samples, such as a Gaussian Process model. Using this surrogate model, we choose new points at which to evaluate the utility function according to an acquisition function. The choice of acquisition function determines how we balance the exploration of high uncertainty regions of the parameter space, improving our surrogate model, with optimizing the existing surrogate to sample new points that will be close to the predicted optimum. The residual between the optimal solution and the sampled solution improves with iteration. Details of Bayesian optimization and descriptions of acquisition functions can be found in [15, 16].

3. BAYESIAN SEISMIC MONITORING

3.1. General Approach

Bayesian inference requires constructing a likelihood model $p(\mathcal{D} | \theta, \mathcal{S})$ that quantifies the likelihood of the data given an event θ , and given a seismic sensor network configuration \mathcal{S} . We assume that a sufficient source description is defined by a vector that stores its origin time, location, and magnitude $\theta = \{\text{Time}, \text{Lat}, \text{Long}, \text{Depth}, \text{Mag}\}$. Notationally, these parameters are epicentral location $\mathcal{L} = \{\text{Lat}, \text{Long}\}$, source depth x , event magnitude m , and origin time t_o . The network \mathcal{S} consists of individual stations \mathcal{S}_i . Such stations may have heterogenous response or sampling features, but our formulation assumes they are homogenous. Therefore, station description is sufficiently described by $\mathcal{S}_i = \{\mathcal{S}_i^{\text{Loc}}\}$ where $\mathcal{S}_i^{\text{Loc}}$ is the station's location.

We limit our analysis to modeling arrivals of seismic phases from their sources, and leave inclusion of waveform features to future research. Our data then takes the form of $\mathcal{D} = \{\mathbb{D}, \mathbb{A}\}$, where \mathbb{D} stores data about which stations detected different seismic phases and \mathbb{A} stores information about the arrival times, t_{ij} , of the detected phases.

$$\mathbb{D}_{ij} = \begin{cases} 1 & \text{if station } i \text{ detects phase } j \\ 0 & \text{if station } i \text{ does not detect phase } j \end{cases} \quad (5)$$

$$\mathbb{A}_{ij} = \begin{cases} t_{ij} & \text{if station } i \text{ detects phase } j \\ \emptyset & \text{if station } i \text{ does not detect phase } j \end{cases} \quad (6)$$

Note that $\mathbb{A}_{ij} = \emptyset$ when no phase is detected since there is no arrival time to capture. We make the simplifying assumptions that the likelihood of detection is independent of the origin time t_o , and that phase arrival time uncertainty is independent of the event magnitude m . This first assumption seems reasonable, but may not hold in cases that background noise is diurnally variable. The second assumption is most appropriate when the dominate source of travel time uncertainty is Earth model uncertainty, rather than measurement error. A more general treatment would model the effect of event magnitude on measurement uncertainty, but avoiding this complexity makes this initial study significantly more tractable. Finally, we assume that the priors are independent. The posterior then is:

$$\begin{aligned} p(\mathcal{L}, x, m, t_o | \mathbb{A}, \mathbb{D}, \mathcal{S}) &= \frac{p(\mathbb{A}, \mathbb{D} | \mathcal{L}, x, m, t_o, \mathcal{S}) p(\mathcal{L}, x, m, t_o)}{p(\mathbb{A}, \mathbb{D} | \mathcal{S})} \\ &\propto p(\mathbb{A} | \mathcal{L}, x, t_o, \mathbb{D}, \mathcal{S}) p(\mathbb{D} | \mathcal{L}, x, m, \mathcal{S}) p(\mathcal{L}) p(x) p(m) p(t_o) \end{aligned} \quad (7)$$

To estimate the likelihood $p(\mathbb{D} | \mathcal{L}, x, m, \mathcal{S})$ we must estimate the detection probability for a given arrival. When historic data is available, we can build a model for the detection of a phase at a station given an event at a specific location and with a specified magnitude.

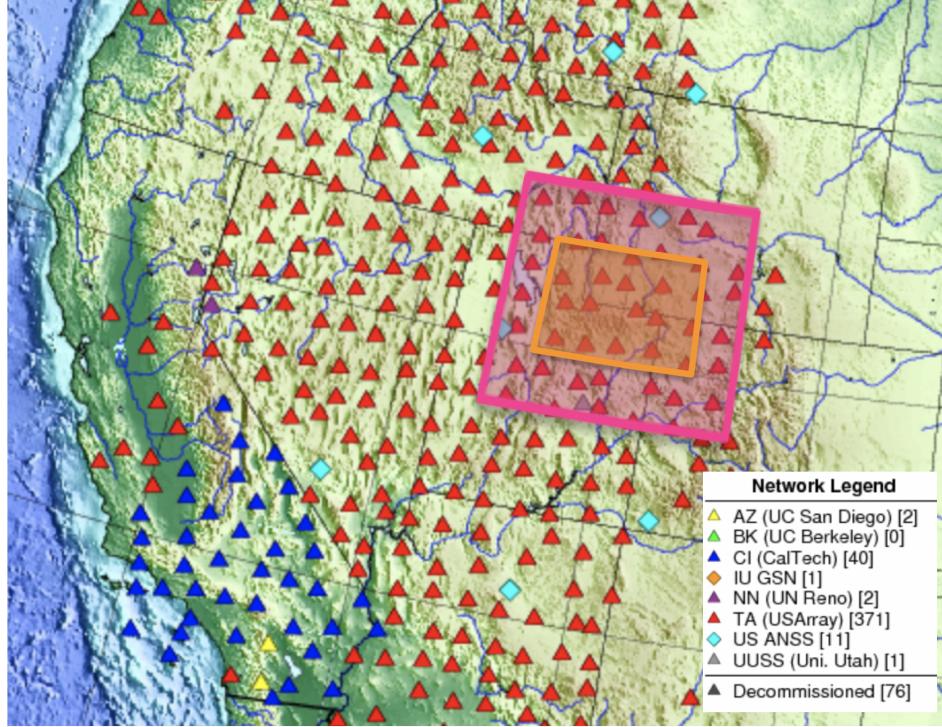


Figure 3-1 Map of Transportable Array stations from December 2007. The pink region indicates the region ($\text{Latitude} \in [39^\circ\text{N}, 43^\circ\text{N}]$, $\text{Longitude} \in [113^\circ\text{W}, 107.36^\circ\text{W}]$) that we use gathered sensor data that populates the parameters of the likelihood models. The orange region ($\text{Latitude} \in [40^\circ\text{N}, 42^\circ\text{N}]$, $\text{Longitude} \in [112^\circ\text{W}, 109.36^\circ\text{W}]$), corresponds to the monitoring region that we build a sensor network over.

We now separate two sources of uncertainty in our arrival time likelihood $p(\mathbb{A} | \mathcal{L}, x, t_o, \mathbb{D}, \mathcal{S})$: measurement noise and model prediction uncertainty. We assume that the measurement noise distribution for each station is independent and identically distributed (which is not true, but provides a tractable simplifying assumption). We can approach estimating measurement uncertainty in several ways. First, we can explicitly specify the station's measurement uncertainty based upon the known sensitivity of that sensor and its processing methods. Alternatively, we can use either historic or simulated data to estimate measurement error by comparing model predictions to data. In this latter case, our model prediction uncertainty will likely include correlated travel time errors across different stations. This correlation reflects that real Earth structure will likely be different than the modeled Earth structure, and this discrepancy will induce correlated errors. We therefore model this uncertainty, and the correlation induced in the sensor network, by sampling a distribution of Earth models to estimate the distribution in arrival times.

3.2. Detection Model

We use a catalog from the USAArray Transportable Array experiment [17] to build a simple detection model for seismic phases, specifically the first P arrival, i.e. arrivals labeled as P, Pg, and Pn in the catalog. Details of the modeling region can be seen in Figure 3-1. This model is

based upon the logistic regression model used by NET-VISA [9]. The USArray dataset was chosen because of the homogeneity of the sensor network and its good coverage of the region. Sensors were deployed in this region from August 2007 - August 2008 and during that time, 1089 events were registered on 45 stations. 11487 P arrivals were detected out of the 49005 potential detections, which corresponds to 23% of potential P detections. Note we assume that every station has the potential to detect each event so the number of potential detections is the just the number of events multiplied by the number of stations. Of the 1089 events, 833 had estimated magnitudes. The minimum magnitude of the dataset was 0.51, maximum was 4.37, and median was 2.03.

Figure 3-2 shows the mean detection probability of seismic sources in our catalog, binned over magnitude and distance. We construct a logistic regression model using input features that include the distance between the event and the sensor (degrees) the depth of the event, and the magnitude of the event. Our catalog data also included events with missing magnitude estimates. We therefore used an additional indicator feature to reflect the absence of magnitude information in our source vector θ . This feature is 1 when the magnitude is absent and 0 when magnitude data is present. This feature helps us train using the data with missing magnitudes. When this model is used as part of the Bayesian OED framework, the magnitudes of hypothetical events will all be known so this indicator feature is always ignored after training. As we described previously, we assume that the detection probability for each station is conditionally independent, so the likelihood model becomes:

$$p(\mathbb{D} | \mathcal{L}, x, m, \mathcal{S}) = \prod_i p(\mathbb{D}_i | \mathcal{L}, x, m, \mathcal{S}_i) \quad (8)$$

$$p(\mathbb{D}_i | \mathcal{L}, x, m, \mathcal{S}_i) = \begin{cases} \frac{\exp(\alpha \text{Dist}[\mathcal{L}, \mathcal{S}_i] + \beta x + \gamma m + \delta)}{1 + \exp(\alpha \text{Dist}[\mathcal{L}, \mathcal{S}_i] + \beta x + \gamma m + \delta)}, & \text{if station } i \text{ detects a phase} \\ \frac{1}{1 + \exp(\alpha \text{Dist}[\mathcal{L}, \mathcal{S}_i] + \beta x + \gamma m + \delta)}, & \text{if station } i \text{ does not detect a phase} \end{cases} \quad (9)$$

The coefficients $\alpha, \beta, \gamma, \delta$ in Equation 8 correspond to the regression coefficients that fit the data. $\text{Dist}[\mathcal{L}, \mathcal{S}_i]$ is the distance in degrees from \mathcal{L} to \mathcal{S}_i , x is the depth, and m is the magnitude. Since we only consider one phase, we remove the phase index in \mathbb{D} hereon. We re-emphasize that only 23% of the event-sensor pairs contained a detection. We therefore tuned our data to balance the performance of the model. We fit the logistic regression model by minimizing binary cross entropy loss across the dataset. This loss comes from the KL divergence between the predicted detection probability and the realized detection. Placing higher weight in the loss function on detections biases the model to predicting detection and balances the dataset composition. We experimented with different weights, Table 3-1 summarizes these results.

A weighting of 2 was chosen to maintain accuracy while providing a significant recall boost (catching the actual positive detections, which contribute more uniquely to the information gain). With this choice, we find the distance coefficient, $\alpha = -2.82$, the depth coefficient, $\beta = -0.03$, the magnitude coefficient, $\gamma = 1.14$, and the intercept, $\delta = 1.95$. From this we see that the distance and magnitude have a much higher influence than depth on the detection probability of the first P arrival.

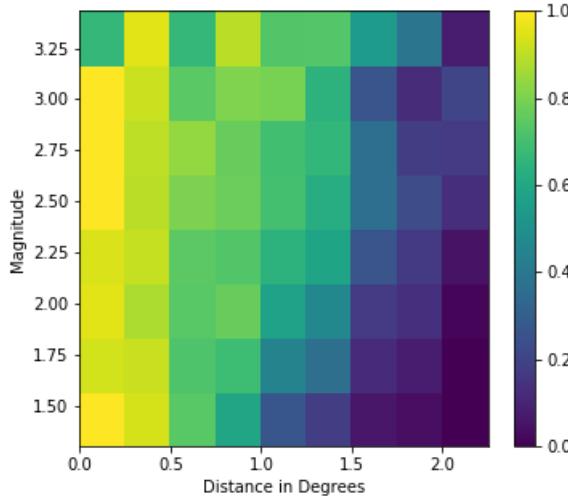


Figure 3-2 The estimated detection probability from the Transportable Array dataset.
We binned these data by distance and magnitude and then estimated the detection probability as the number of detections versus the number of potential detections for stations and events within a given distance and magnitude bin.

| Detection Weight | Accuracy | Precision | Recall | AUC |
|------------------|----------|-----------|--------|------|
| 1 | 0.870 | 0.728 | 0.686 | 0.92 |
| 2 | 0.865 | 0.665 | 0.820 | 0.92 |
| 3 | 0.850 | 0.617 | 0.874 | 0.92 |
| 4 | 0.835 | 0.583 | 0.905 | 0.92 |
| 5 | 0.819 | 0.553 | 0.920 | 0.92 |

Table 3-1 Over weighting of detections in the loss function was considered to correct for the dataset imbalance towards non-detections. A weight of 2 was chosen to balance the different performance metrics.

3.3. Arrival Time Model

While the analysis in the rest of this document does not rely on a data driven arrival time likelihood model, it is helpful to consider the complexities of real arrival time data to understand some of the modeling choices. Again the same Transportable Array dataset was used. For each P arrival, we predict the arrival time for a phase given the event and sensor locations and origin time in the catalog using the IASP91 velocity model. Then, we calculated the residuals observed in the data. Figure 3-3 shows a scatter plot of the residual data as a function of distance.

We see little obvious relationship between distance and the residual. This residual is probably a combination of many factors: measurement noise, travel time model errors, phase categorization errors, and location errors in the catalog. In the histogram, Figure 3-4, we see that the residuals follow a heavy tailed distribution. We choose to model it with a non-centered t-distribution.

For the non-centered t-distribution we fit the data and found that the degrees of freedom parameter was 2.198, location parameter was 0.214, and scale parameter was 0.293. While this distribution fits the data reasonably well, it does not give us the ability to tune the various sources

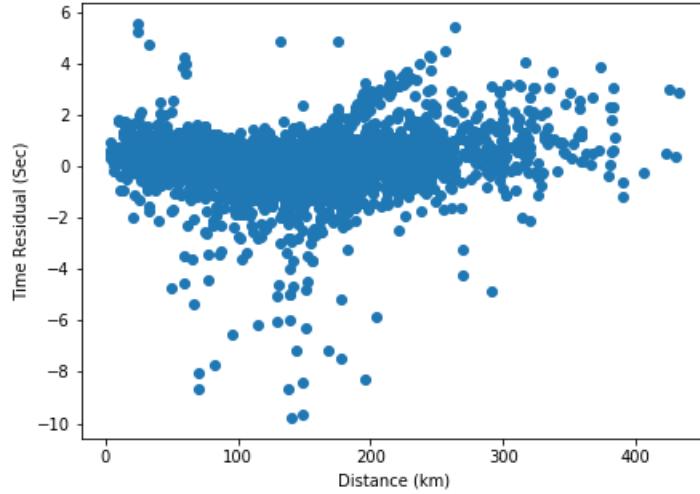


Figure 3-3 Arrival time residual as a function of distance in kilometers. Note that there is a bias towards positive residuals, particularly at longer distances. This bias is particularly evident in Figure 3-4.

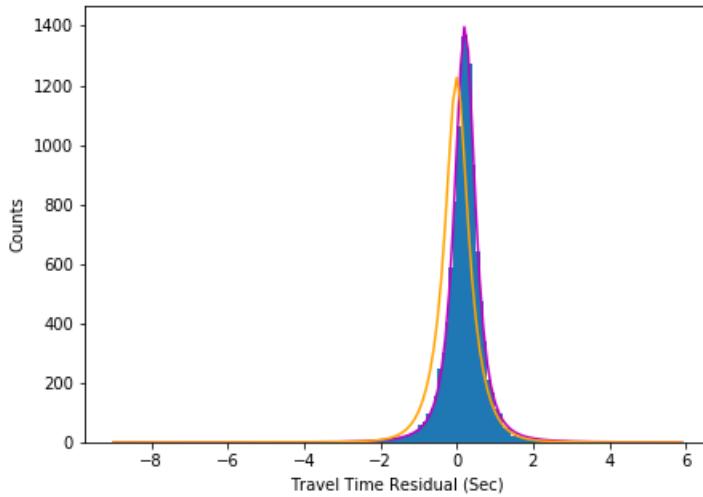


Figure 3-4 Arrival time residual histogram compared to the fit of different statistical models. The magenta line indicates the non-centered t-distribution fit to the data. The orange line shows the marginal residual distribution under the more tractable distance dependent Gaussian model discussed in Section 3.4. Note that the non-data driven model does not a priori know the bias so it is centered at zero.

of uncertainty when analyzing and optimizing the seismic network. Further, considering station correlation and marginalizing out origin time uncertainty is very hard for this distribution. Therefore, we instead turn to a simple Gaussian distribution because the Gaussian distribution allows us to easily model correlation and marginalize out origin time uncertainty analytically. Finally, for the purpose of Bayesian OED, ignoring the bias and choosing to use a zero mean Gaussian is justified because adding a constant, known, bias to all travel times would only affect the time of the arrivals but not their uncertainty and therefore the likelihood would be the same. Obviously, for inference with real data including the bias is necessary.

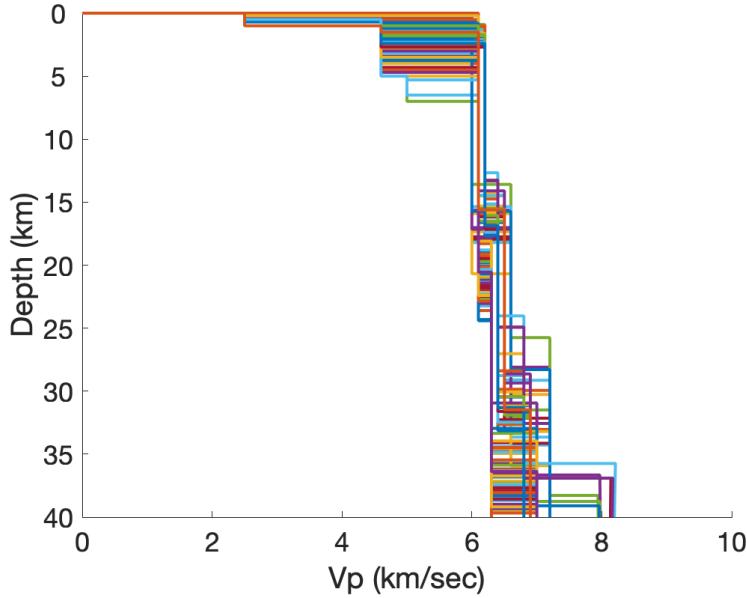


Figure 3-5 Illustration of the 121 1D velocity models for Vp sampled from around the monitoring region. These representative earth models are used to estimate travel time uncertainty from earth model uncertainty.

3.4. Integrating Earth model Uncertainty

For our travel time uncertainty model, first we will build an uncertainty model that captures the uncertainty due to the earth model. Then we will include a simple independent and additive measurement uncertainty model. To capture earth model uncertainty, we selected 121 vertical cross-sections from Crust 1.0 [18] from the area around the monitoring region to get 121 different 1D earth models with different Vp velocity profiles. These models can be seen in Figure 3-5. For each of these models we used TauP [19] to compute the travel times for different distances, Δ , and depth, z , pairs. For a given distance and depth pair we compute the mean and variance of the travel times, t_i , given the travel times computed by TauP for the $N = 121$ models:

$$\mu(\Delta, z) = \frac{1}{N} \sum_{i=1}^N t_i(\Delta, z) \quad (10)$$

$$\sigma(\Delta, z) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N [t_i(\Delta, z) - \mu(\Delta, z)]^2} \quad (11)$$

Given the estimated mean and standard deviation pairs (Figure 3-6), we can derive a model for the standard deviation of the travel time as a cubic function of the predicted travel time:
 $\sigma_p(\mu_p) = a\mu_p + b\mu_p^2 + c\mu_p^3$. Therefore, given any prediction of the travel time we can construct an appropriate estimate of the uncertainty to be used in the likelihood.

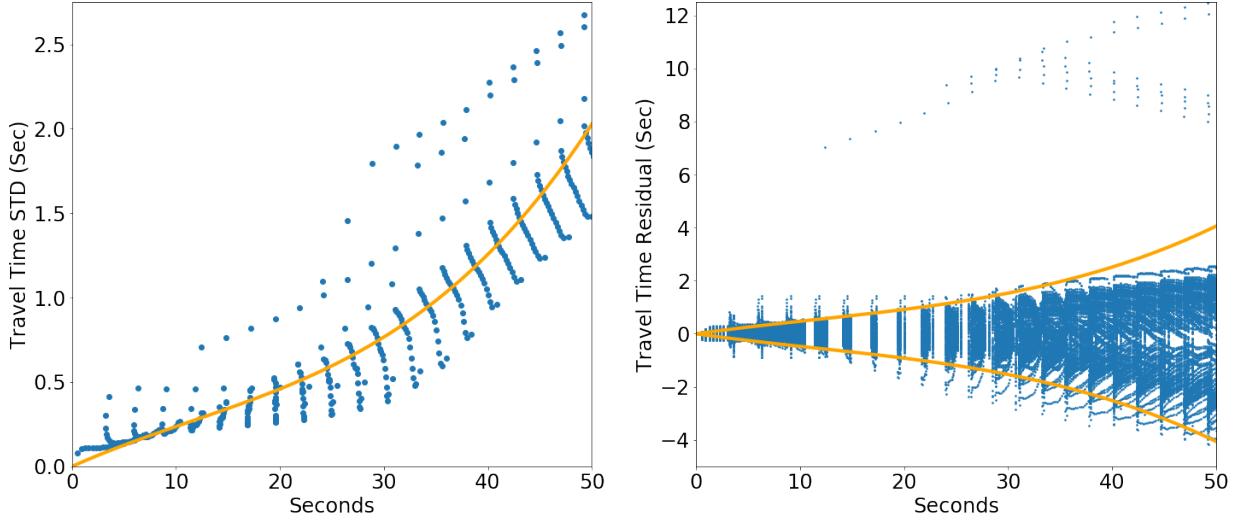


Figure 3-6 Left: A scatter plot of the estimated travel time mean $\mu(\Delta, z)$, and estimated travel time standard deviation, $\sigma(\Delta, z)$ for various distance and depth pairs, superimposed with a fit polynomial model (orange curve). **Right:** A composite scatter plot of the travel time residuals for all sampled distance, depth, and earth models compared to the 2σ range defined by the fitted polynomial model. The simple polynomial model adequately captures the bulk trend, despite some variability; a few very high residual points inflate the standard deviation away from typical residual values.

3.5. Travel Time Correlation

We can also compute the correlation between the travel times observed at two different stations at locations Δ_j and Δ_k for an event at depth z . This correlation is induced by the earth model uncertainty as:

$$\rho(\Delta_j, \Delta_k, z) = \frac{\sum_{i=1}^N [t_i(\Delta_j, z) - \mu(\Delta_j, z)][t_i(\Delta_k, z) - \mu(\Delta_k, z)]}{(N-1)\sigma(\Delta_j, z)\sigma(\Delta_k, z)} \quad (12)$$

Here μ and σ are computed from earth models in different locations from Crust 1.0 as in Equations 10 and 11. For simplicity we will remove the depth dependence of the correlation by averaging the correlation over all, L , depths. Therefore we estimate the correlation between two sensors as $\rho(\Delta_j, \Delta_k) = \frac{1}{L} \sum_{l=1}^L \rho(\Delta_j, \Delta_k, z_l)$.

The full correlation matrix, Γ , between the stations at distances Δ_i from the source has elements $\Gamma_{jk} = \rho(\Delta_j, \Delta_k)$. We want to fit a Gaussian process model with a square exponential kernel to this data so we can estimate the correlation between arbitrary sensor pairs when designing the network i.e. we want $\Gamma \approx \Gamma_{GP}$. Therefore we want to find the a correlation length, l , such that $\{\Gamma_{GP}\}_{jk} = \exp\left[-\frac{1}{2l^2}(\Delta_j - \Delta_k)^2\right]$ and Γ_{GP} minimizes the discrepancy with Γ . We find the correlation length scale as $l = 147.5$ km. The comparison of Γ and the resulting Γ_{GP} can be seen in Figure 3-7. We observe that the square exponential kernel is able to capture the general length scale of the induced correlation, meaning that stations that are close together are more correlated, but does not capture its complexity. The induced correlation has a block-like structure where

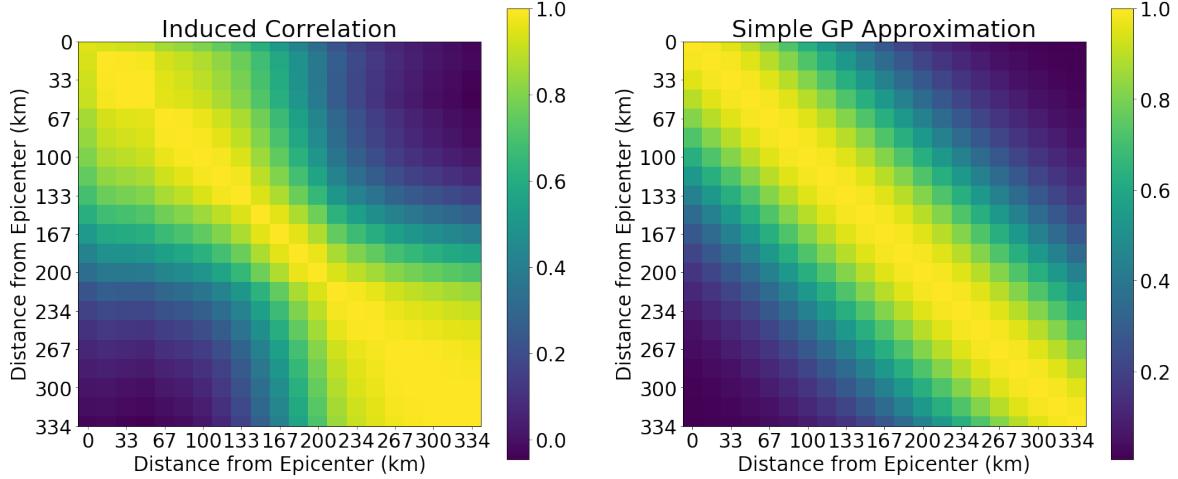


Figure 3-7 Visualization of the correlation matrix. The left figure is the correlation matrix, Γ , induced by the earth model uncertainty computed using Equation 12. The axes correspond to the distance along the surface from the source epicenter in km. Stations are approximately spaced 33km apart. The right figure is a simplified model of the correlation, Γ_{GP} , found using a square-exponential kernel, which approximates Γ . The square-exponential kernel assumes that correlation is only a function of the distance between stations. This simplified model reasonably captures the correlation length scale for Γ but is unable to capture the complex, non-translation invariant, block structure of Γ .

stations that are near to the source are highly correlated, stations far from the source are highly correlated, and stations in the transition region exhibit less strong correlation with nearby stations. This likely corresponds to the type of first arrival that is being observed at each station, where close stations observe a Pg while far stations observe a Pn. Our choice of GP kernel is translation invariant meaning that the sensor correlation is only a function of the distance between the two sensors and does not depend on the source parameters. More generally, a different GP kernel would need to be constructed for each seismic source, which is computationally challenging. Considering only a translation invariant GP kernel is obviously a simplification but provides a first step towards modeling station correlation which is typically very difficult and often ignored.

We can now construct the P arrival time likelihood $p(\mathbb{A} | \mathcal{L}, x, t_o, \mathbb{D}, \mathcal{S})$ by combining our model of the travel time prediction, μ_p ; standard deviation, σ_p ; correlation matrix, Γ_{GP} ; and independent and identically distributed measurement noise with covariance $\epsilon^2 \mathbb{I}$:

$$p(\mathbb{A} | \mathcal{L}, x, t_o, \mathbb{D}, \mathcal{S}) = \frac{1}{(2\pi)^{|\mathbb{D}|/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} [\mathbb{A} - \mu_p(\mathcal{L}, x, \mathbb{D}, \mathcal{S}) - t_o]^T \Sigma^{-1} [\mathbb{A} - \mu_p(\mathcal{L}, x, \mathbb{D}, \mathcal{S}) - t_o] \right) \quad (13)$$

$$\Sigma = \text{diag}[\sigma_p(\mu_p(\mathcal{L}, x, \mathbb{D}, \mathcal{S}))] \Gamma_{GP}(\mathbb{D}, \mathcal{S}) \text{diag}[\sigma_p(\mu_p(\mathcal{L}, x, \mathbb{D}, \mathcal{S}))] + \epsilon^2 \mathbb{I} \quad (14)$$

Here $|\mathbb{D}|$ is the number of detections, $|\Sigma|$ is the determinant, $\mu_p(\mathcal{L}, x, \mathbb{D}, \mathcal{S})$ is a vector of the predicted travel times for stations that had a detection, and $\text{diag}[\sigma_p(\mu_p(\mathcal{L}, x, \mathbb{D}, \mathcal{S}))]$ is a diagonal

matrix of the predicted standard deviations of the travel time to each station as a function of the mean. Finally, $\Gamma_{GP}(\mathbb{D}, \mathcal{S})$ is the estimated correlation between stations using the GP model.

We further note that we marginalized our source origin time prior over t_o , assuming a uniform improper prior, and therefore omit it from the model. This reduces the dimension of our seismic source parameterization space and leads to our final model of the arrival time likelihood:

$$p(\mathbb{A} | \mathcal{L}, x, \mathbb{D}, \mathcal{S}) = \int_{t_o} p(\mathbb{A} | \mathcal{L}, x, t_o, \mathbb{D}, \mathcal{S}) p(t_o) dt_o \quad (15)$$

$$= \frac{1}{(2\pi)^{(1-|\mathbb{D}|)/2} |\Sigma|^{1/2} \beta^{1/2}} \exp\left(-\frac{1}{2} [\mathbb{A} - \mu_p(\mathcal{L}, x, \mathbb{D}, \mathcal{S})]^T \Sigma^{-1} [\mathbb{A} - \mu_p(\mathcal{L}, x, \mathbb{D}, \mathcal{S})]\right) \exp\left(\frac{\alpha^2}{\beta}\right) \quad (16)$$

$$\alpha = \mathbb{1}^T \Sigma^{-1} [\mathbb{A} - \mu_p(\mathcal{L}, x, \mathbb{D}, \mathcal{S})] \quad (17)$$

$$\beta = \mathbb{1}^T \Sigma^{-1} \mathbb{1} \quad (18)$$

4. COMPUTATIONAL APPROACH

4.1. Estimating Information Gain

Given this approach we are now able to estimate the expected information gain, $I(\mathcal{S})$, of the sensor network \mathcal{S} . Recall that we can express EIG as:

$$I(\mathcal{S}) = \int p(\theta') \int p(\mathcal{D} | \theta', \mathcal{S}) \int p(\theta | \mathcal{D}, \mathcal{S}) \log \frac{p(\theta | \mathcal{D}, \mathcal{S})}{p(\theta)} d\theta d\mathcal{D} d\theta' \quad (19)$$

We can further define $I(\mathcal{S} | \theta')$ as the expected information gained about a specific event θ' where

$$I(\mathcal{S} | \theta') = \int p(\mathcal{D} | \theta', \mathcal{S}) \int p(\theta | \mathcal{D}, \mathcal{S}) \log \frac{p(\theta | \mathcal{D}, \mathcal{S})}{p(\theta)} d\theta d\mathcal{D} \quad (20)$$

and thus express $I(\mathcal{S})$ as:

$$I(\mathcal{S}) = \int I(\mathcal{S} | \theta') p(\theta') d\theta' \quad (21)$$

$I(\mathcal{S} | \theta')$ is an important quantity on its own as it can be used to tell how sensitive the network is to a specific event.

We will use the approach of estimating $I(\mathcal{S} | \theta')$ to estimate EIG. First, we draw samples from our prior to construct a large set of candidate seismic events using a method like importance sampling with a Quasi Monte Carlo mesh. Then, for each element in our event space, we will estimate $I(\mathcal{S} | \theta')$ and average them to estimate $I(\mathcal{S})$. To estimate $I(\mathcal{S} | \theta')$ we will construct hypothetical datasets by sampling $p(\mathcal{D} | \theta', \mathcal{S})$. Then we will solve the Bayesian inference problem given the datasets to estimate the information gain measured via the KL divergence. The hypothetical data is constructed by sampling $p(\mathbb{A} | \mathcal{L}, x, \mathbb{D}, \mathcal{S})$ and $p(\mathbb{D} | \mathcal{L}, x, m, \mathcal{S})$. This is summarized in Algorithm 1.

Algorithm 1: Expected Information Gain (EIG) Calculation

Result: Information gain, $I(\mathcal{S} | \theta)$, for individual events, θ , and total EIG, $I(\mathcal{S})$, given sensor configuration, \mathcal{S}

Construct the set of plausible events $\theta' \in \Theta$ with parameters: locations \mathcal{L}' , depth x' , and magnitudes m' such that $\theta' \sim p(\theta')$;

for each event hypothesis, θ' do

Assuming the seismic event hypothesis, θ' , simulate hypothetical datasets of which stations detect an arrival according to the distribution $\mathbb{D} \sim p(\mathbb{D} | \mathcal{L}', x', m', \mathcal{S})$ (Equation 8);

For each arrival dataset, simulate the arrival time according to the distribution

$\mathbb{A} \sim p(\mathbb{A} | \mathcal{L}', x', \mathbb{D}, \mathcal{S})$ (Equation 13);

for each simulated data set, $\mathcal{D} = \{\mathbb{A}, \mathbb{D}\}$ do

Compute the likelihood $p(\mathcal{D} | \theta, \mathcal{S})$ of the observation data $\mathcal{D} = \{\mathbb{A}, \mathbb{D}\}$ given each event in the event space $\theta \in \Theta$ using the detection and arrival time likelihood functions, (Equations 8 and 15, respectively);

Compute the posterior probability of each event from the likelihood,
 $p(\theta | \mathcal{D}, \mathcal{S}) \propto p(\mathcal{D} | \theta, \mathcal{S}) p(\theta)$;

Compute the KL divergence i.e. information gain for this realization

$$I(\mathcal{S} | \theta', \mathcal{D}) = \int p(\theta | \mathcal{D}, \mathcal{S}) \log \frac{p(\theta | \mathcal{D}, \mathcal{S})}{p(\theta)} d\theta;$$

end

Compute the EIG for the event hypothesis, $I(\mathcal{S} | \theta')$, as the average over KL divergences of the simulated data realizations, $I(\mathcal{S} | \theta', \mathcal{D})$, (Equation 20);

end

Compute the total EIG, $I(\mathcal{S})$, as average across all event hypotheses and simulated data (Equation 21);

4.2. Optimization

Once we have the algorithms to estimate $I(\theta' | \mathcal{S})$ and $I(\mathcal{S})$, we can formulate the optimal experimental design problem to choose the the location and type of different seismic stations. We can use the greedy Bayesian optimization method described in 2.3. We use the Python library skopt [20] to implement Bayesian optimization with a GP surrogate.

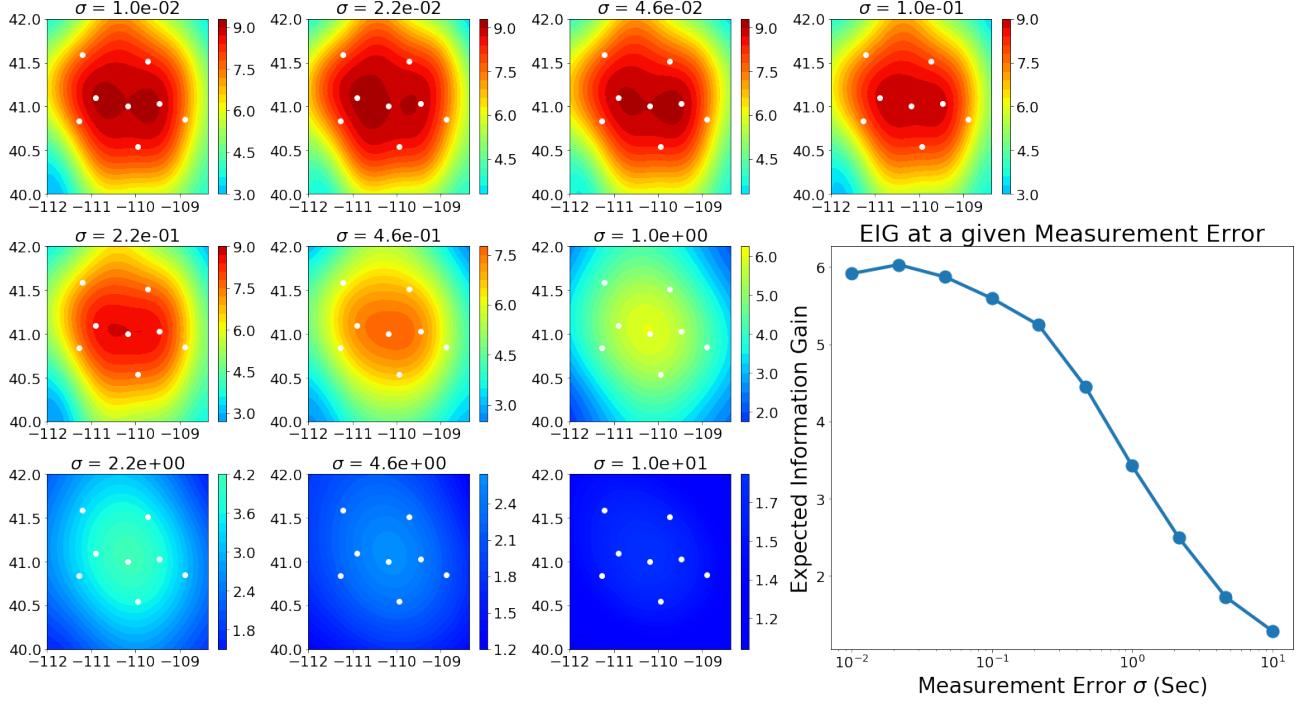


Figure 5-1 Illustration of the effect of changing the measurement uncertainty standard deviation over several orders of magnitude. The color plots illustrate the EIG of a shallow seismic source with the different stated measurement errors. The line plot illustrates the degradation of EIG for all events as the measurement error increases. For noise levels below about 0.2 seconds the model uncertainty dominates over measurement uncertainty so EIG is fairly stable.

5. RESULTS

We explore a simple model for placing sensors to monitor a square domain for latitudes between 40°N and 42°N , longitudes between 112°W and 108.36°W , and depth between 0km and 40km. Seismic sources are assumed to have a uniform prior probability in this domain. We also assume that the magnitude prior is an exponential distribution with rate parameter $\lambda = \log(10)$ and a minimum magnitude of 0.5. This prior means that the likelihood of an event of a given magnitude falls off exponentially as the magnitude increases. We assume that the origin time is a uniform improper prior meaning that all times are equally likely. The sensors are also limited to be placed in this domain. For computing the EIG, 10k events were chosen using a Quasi Monte Carlo (QMC) mesh defined by the Sobol sequence. For each candidate event 20 hypothetical data realizations were used. 100 steps of Bayesian optimization were used to optimize the sensor configuration.

5.1. Effect of Measurement Noise

As described in Section 3.4, we can use simulations of different travel time models to get a notion of modeling error and the induced modeling correlation, which we capture with a Gaussian

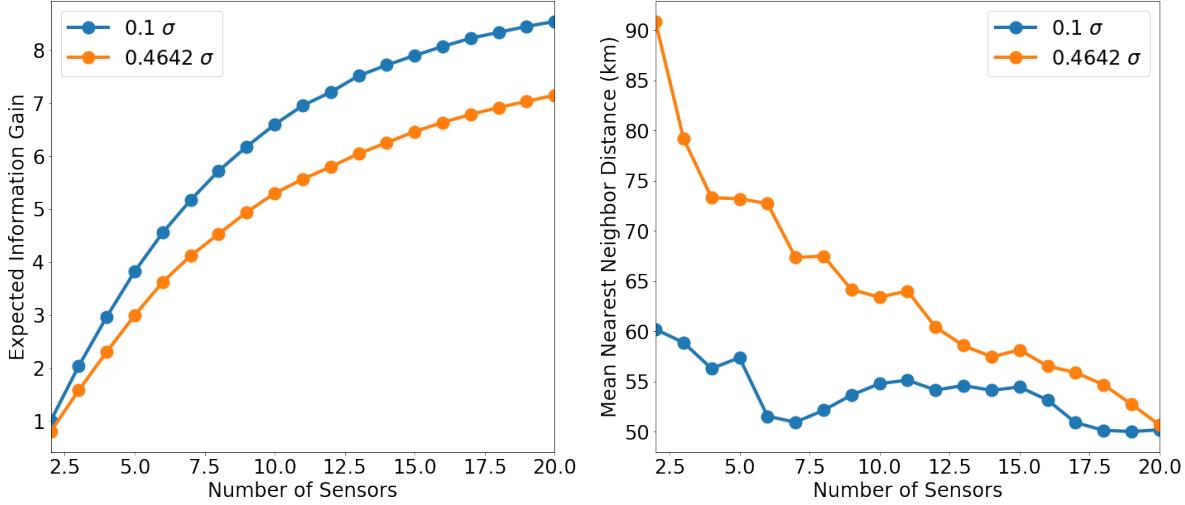


Figure 5-2 The left panel illustrates the change in EIG as additional sensors are placed using greedy optimization for two different networks with different sensor fidelities. One network has sensors with a measurement error standard deviation of 0.1 seconds and the other has sensors with error of 0.4642 seconds. For this the 20 station noisy network is about equivalent to the 11 station high fidelity network. The right panel describes the geometry of the networks based upon how close the stations are to each other. We can see that, particularly in the beginning, stations of the noisy network are being added further apart than stations of the less noisy network.

process. Then we additionally add measurement noise with a defined standard deviation to replicate phase arrival pick uncertainty. This gives the GP model for the arrival times from Equation 13. For a given network we can see how this added measurement noise affects the expected information gain. These results are seen in Figure 5-1. When the measurement noise is small enough, the model uncertainty dominates. Therefore, the expected information gain remains relatively unchanged as the measurement noise increases until it exceeds the model uncertainty. In this eight-sensor network, that occurs when the measurement noise standard deviation is around 0.2 seconds. Given this analysis we can see whether measurement noise or model uncertainty is the dominant source of error in a network and assess the impact of either reducing modeling uncertainty or measurement uncertainty.

Next we can investigate how the sensor network evolves as we use greedy optimization to place new sensors under two different noise conditions, see Figures 5-2 and 5-3. We find that sensors are initially placed further apart from each other for the noisier network compared to the less noisy network. One possible explanation is that if stations are close together timings must be estimated very accurately to be able to triangulate a distant event. So, if the stations are noisier, they need to be further apart, so they do not have to be as accurate. This analysis can be useful to determine the trade off between using few high fidelity sensors versus numerous low fidelity sensors.

5.2. Effect of Correlation

We can also investigate the effect of station correlation on the placement of sensors. We look at three different correlation length scales, l : 14.75km, 147.5km, and 1475km. The measurement noise standard deviation was fixed at 0.1 seconds. Twenty stations were then placed using greedy optimization. The results can be seen in the Figures 5-4 and 5-5. We observe that correlation increases expected information gain for these experiments, which is expected, as overall correlation means less uncertainty. However, if there is a discrepancy between the correlation model used and the true correlation, the bias introduced may be very significant. This should be investigated more in the future. Further, in Figure 5-4 we see that correlation has relatively modest effect on the EIG of the optimized network particularly as the number of sensors grow large and EIG begins to saturate for all correlation length scales. This behavior could mean that correlation does not influence sensor placement too much, but it also could mean that the greedy optimization approach is poor for large numbers of sensors. In future studies we will disambiguate these possibilities. We see in Figure 5-5 that the distance between stations is fairly regular and therefore the mean nearest neighbor distance in Figure 5-4 reasonably represents the network. The mean nearest neighbor distance is very similar for all correlation models once the number of sensors grows large and the domain is fully covered, and again this could be due to the use of greedy optimization.

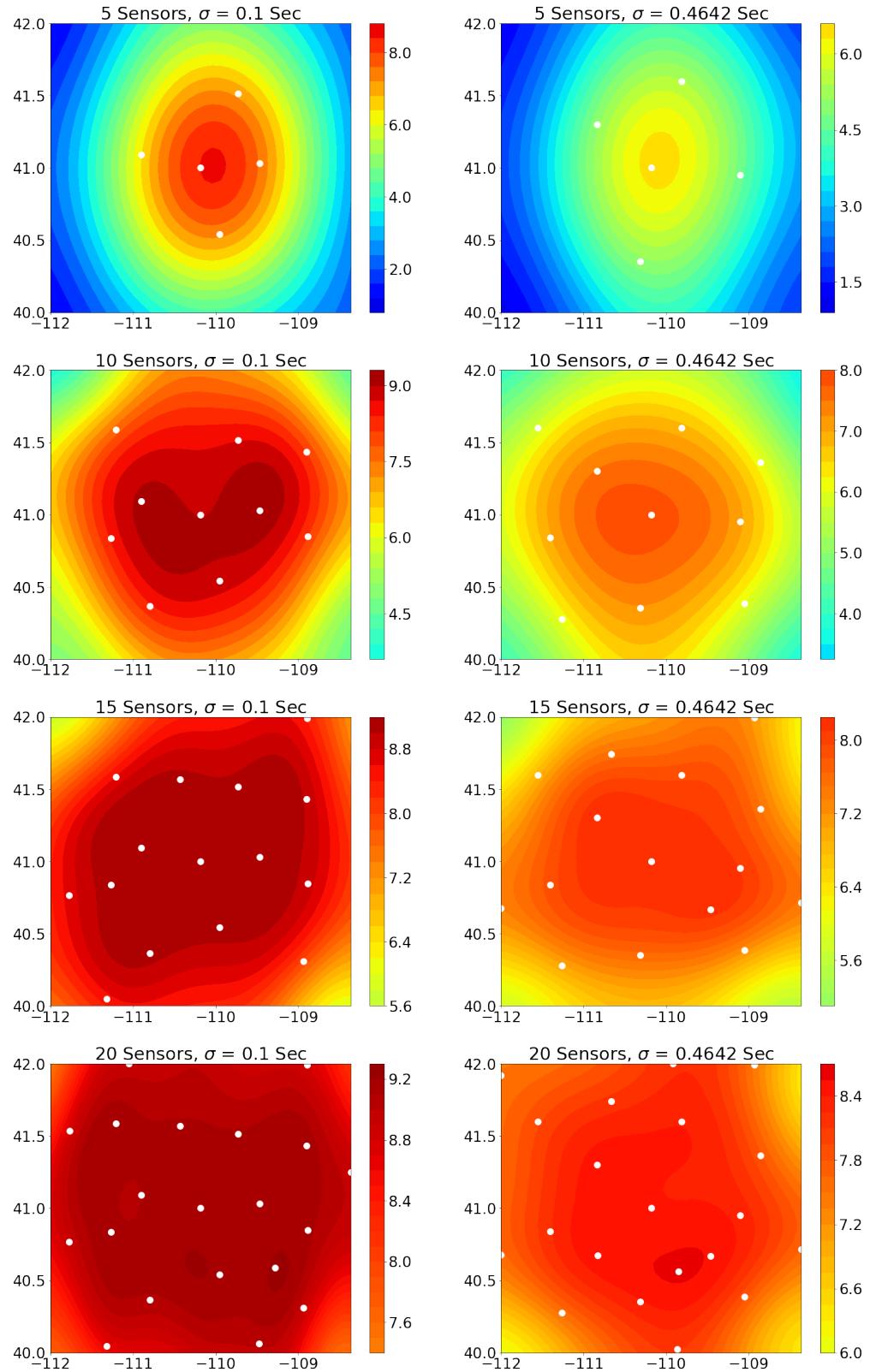


Figure 5-3 Depiction of the evolution of the sensor placement and expected information gain about surface events given the two different noise models and different numbers of sensors.

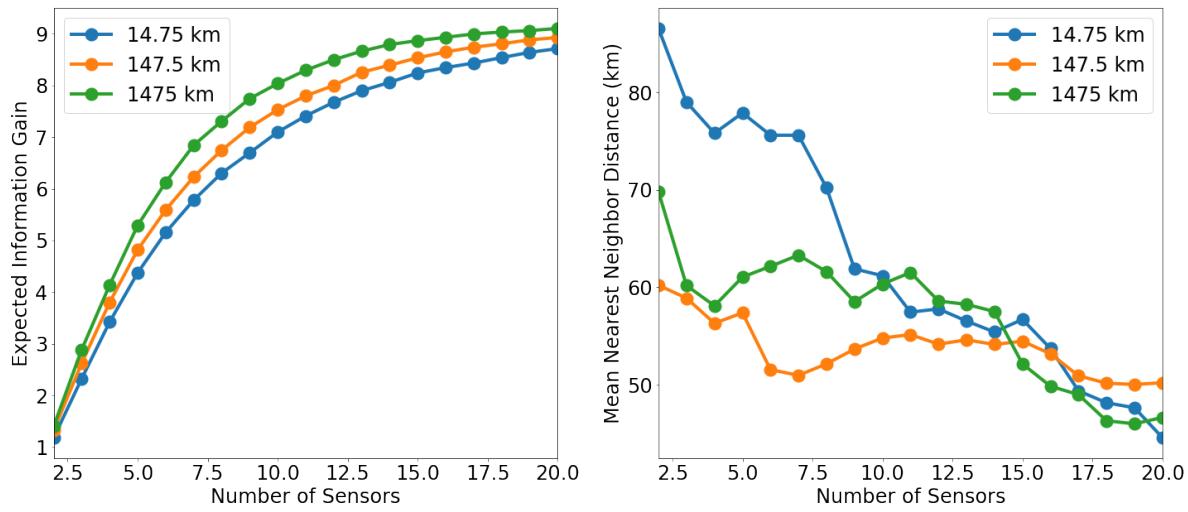


Figure 5-4 The left panel illustrates the evolution of the expected information gain as the number of sensors increases for the three different correlation models where the correlation changes over two orders of magnitude. We see that initially the disparity in EIG is small then increases with the number of sensors. However, after about 10 sensors, the disparity begins to decrease. The right panel investigates the geometry of the network by looking at the distances between stations. Few patterns can be identified in this plot although it may be the case that stations are initially further apart for the low correlation model.

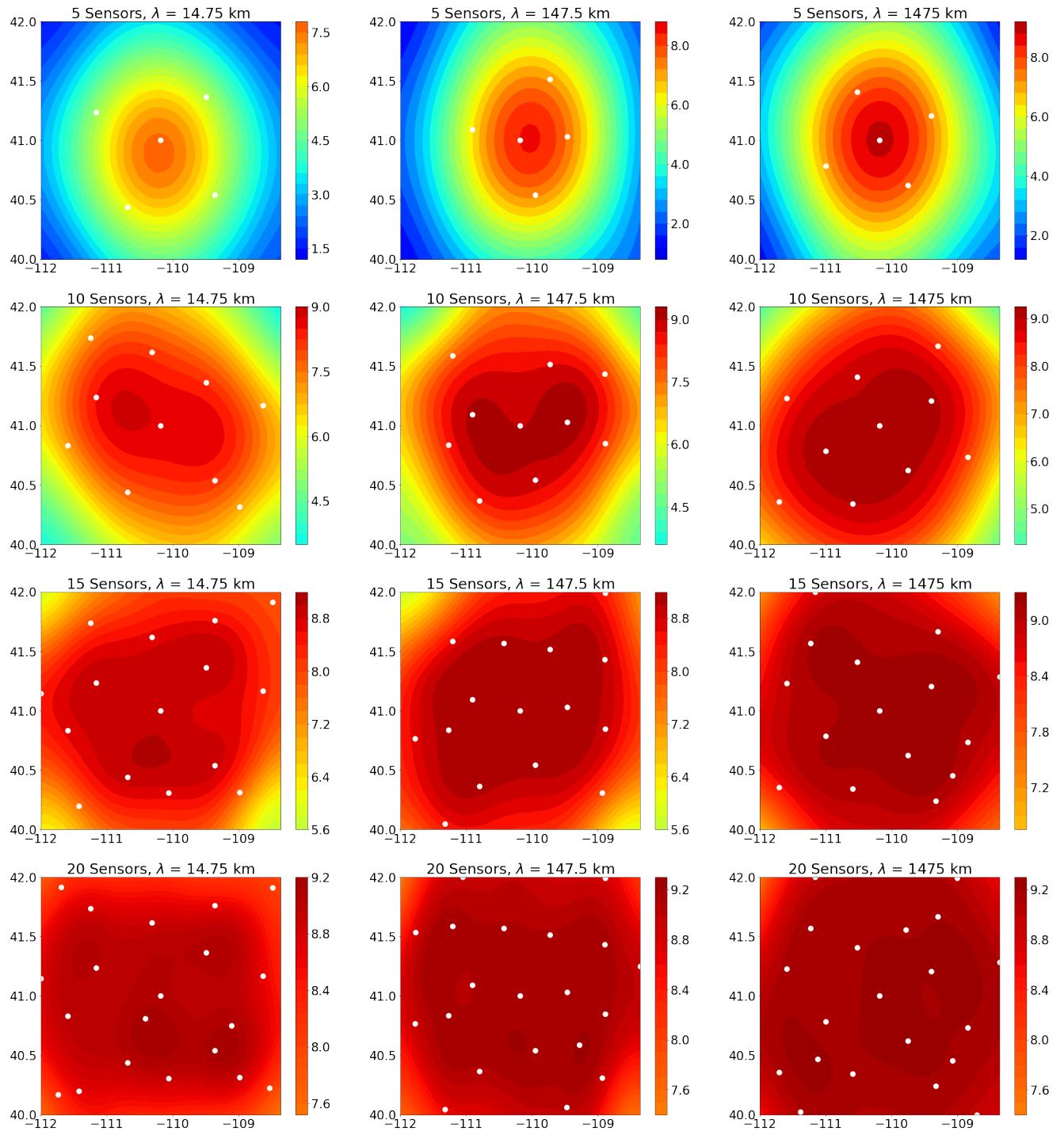


Figure 5-5 Depiction of the evolution of the sensor placement and expected information gain about surface events given the three different correlation models and different numbers of sensors. The spacing of sensors in all cases appears to be fairly regular.

6. CONCLUSION

In this work, we have demonstrated and implemented a modern framework for Bayesian optimal experimental design for analyzing and optimizing a seismic monitoring network. We used this framework on a seismic source location problem with uncertainty in both the detection of seismic phases and uncertainty in the arrival time. We selected these models using data from the U.S. Transportable Array and physics-based travel time modeling with earth model uncertainty. Using these models, we capture the often-ignored influence of earth model uncertainty and station correlation on travel times. We further investigate the influence of station correlation, earth model uncertainty, and phase-arrival pick uncertainty on the sensor placement and sensitivity of the monitoring network.

Our Bayesian OED approach will enable rigorous and flexible analysis and design of monitoring networks for applications like nuclear explosion monitoring. When evaluating a monitoring network, decision makers in high-consequence domains can trust the rigor of the Bayesian approach to provide coherent uncertainty quantification. Further, decision makers may employ Bayesian OED to assess the monitoring network's sensitivity to different types of seismic sources and locations and therefore can certify the capabilities of the network to meet design requirements. Bayesian OED may answer other questions critical to seismic monitoring such as: how may multiphenomenology data be used to reduce uncertainty; what is the appropriate sensor fidelity or earth model resolution for estimating a QoI; and how do sensor types, number, and locations influence estimates of QoIs?

While this work provides a meaningful first step towards analyzing and optimizing monitoring networks, many simplifications were made during this exploratory study. Based on these results we have identified several follow-on directions to increase its applicability to real monitoring problems:

1. In this work, we used a very simple Gaussian travel time model because it enabled marginalization of origin time and handling correlation between stations. As we saw, real data is much more complex and so more complex travel time models should be explored. Further, the measurement error model should incorporate the event magnitude as that will affect the signal-to-noise ratio.
2. Further, we may extend the correlation model to include more event characteristics. We ultimately assumed that the station correlation was independent of the event and was only a function of how far apart the stations were. Real data exhibits more complex correlation structures, such as depth dependence. Further, we assumed that the detections of each station were independent. Again, we would expect this not to be true.
3. We also assumed that the stations were identical. Studying a heterogenous sensor network is much more realistic. Stations are heterogenous both because of the use of different sensors but also based upon how the stations are installed, which could introduce different uncertainties and background noise environments. Modeling this heterogeneity also would enable us to better assess the tradeoff between different sensor types and installation methods.

4. Finally, we may incorporate many other sources of data into this analysis. We only considered P arrivals so other seismic phases should be studied using the same workflow and incorporated into the likelihood function. Also, infrasound sensors and seismic arrays could be included to make the analysis multi-modal by providing directional information. This would then give us the ability to explore the utility of different sensor types as we could see how the expected information gain changes as we add sensors with these different modalities. We could also then deduce the types of seismic sources different data modalities most benefit.

REFERENCES

- [1] Andreas Krause, Ajit Singh, and Carlos Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(Feb):235–284, 2008.
- [2] Xun Huan and Youssef M Marzouk. Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317, 2013.
- [3] Joshua Carmichael, Robert Nemzek, Neill Symons, and Mike Begnaud. A method to fuse multiphysics waveforms and improve predictive explosion detection: theory, experiment and performance. *Geophysical Journal International*, 222(2):1195–1212, 2020.
- [4] Stephen Arrowsmith, Junghyun Park, Il-Young Che, Brian Stump, and Gil Averbuch. Event location with sparse data: When probabilistic global search is important. *Seismological Research Letters*, 2020.
- [5] Bion John Merchant. Netmod, version 00, 9 2013.
- [6] James L Beck. Bayesian system identification based on probability logic. *Structural Control and Health Monitoring*, 17(7):825–847, 2010.
- [7] Edwin T Jaynes. *Probability theory: The logic of science*. Cambridge university press, 2003.
- [8] Stephen C Myers, Gardar Johannesson, and William Hanley. A bayesian hierarchical method for multiple-event seismic location. *Geophysical Journal International*, 171(3):1049–1063, 2007.
- [9] Nimar S Arora, Stuart Russell, and Erik Sudderth. Net-visa: Network processing vertically integrated seismic analysis. *Bulletin of the Seismological Society of America*, 103(2A):709–729, 2013.
- [10] Art B Owen. Monte carlo theory, methods and examples. 2013.
- [11] Josep Ginebra et al. On the measure of the information in a statistical experiment. *Bayesian Analysis*, 2(1):167–211, 2007.
- [12] Dennis V Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pages 986–1005, 1956.
- [13] Joshua Daniel Carmichael. Hypothesis tests on rayleigh wave radiation pattern shapes: A theoretical assessment for source screening; independent review: Lanl source physics lcp. 2020.
- [14] Jonas Močkus. On bayesian methods for seeking the extremum. In *Optimization techniques IFIP technical conference*, pages 400–404. Springer, 1975.
- [15] Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.
- [16] Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

- [17] Usarray transportable array monthly catalogs picked by anf.
<https://anf.ucsd.edu/tools/events/>. Accessed: 2020-04-30.
- [18] Gabi Laske, Guy Masters, Zhitu Ma, and Mike Pasquano. Update on crust1.0—a 1-degree global model of earth’s crust. In *Geophys. Res. Abstr.*, volume 15, page 2658, 2013.
- [19] H Philip Crotwell, Thomas J Owens, and Jeroen Ritsema. The taup toolkit: Flexible seismic travel-time and ray-path utilities. *Seismological Research Letters*, 70(2):154–160, 1999.
- [20] Tim Head, Manoj Kumar, Holger Nahrstaedt, Gilles Louppe, and Iaroslav Shcherbatyi. scikit-optimize/scikit-optimize, September 2020.



Sandia
National
Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.