# Geoacoustic inversion using Bayesian optimization with a Gaussian process surrogate model

William F. Jenkins,[a] Peter Gerstoft, and Yongsung Park
*Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, USA*

**ABSTRACT:**

Geoacoustic inversion can be a computationally expensive task in high-dimensional parameter spaces, typically requiring thousands of forward model evaluations to estimate the geoacoustic environment. We demonstrate Bayesian optimization (BO), an efficient global optimization method capable of estimating geoacoustic parameters in seven-dimensional space within 100 evaluations instead of thousands. BO iteratively searches parameter space for the global optimum of an objective function, defined in this study as the Bartlett power. Each step consists of fitting a Gaussian process surrogate model to observed data and then choosing a new point to evaluate using a heuristic acquisition function. The ideal acquisition function balances exploration of the parameter space in regions with high uncertainty with exploitation of high-performing regions. Three acquisition functions are evaluated: upper confidence bound, expected improvement (EI), and logarithmically transformed EI. BO is demonstrated for both simulated and experimental data from a shallow-water environment and rapidly estimates optimal parameters while yielding results comparable to differential evolution optimization. © 2024 Acoustical Society of America.

https://doi.org/10.1121/10.0028177

## I. INTRODUCTION

Geoacoustic inversion can be a computationally expensive task, requiring thousands of evaluations of a forward model to estimate the geoacoustic parameters that explain measured acoustic data.[1] While computational costs have been mitigated through advances in computing capabilities, the complexity of some models has grown in an attempt to accurately predict more complicated environments. For example, underwater acoustic propagation models that use the parabolic equation or finite element/finite difference modeling are time consuming and computationally expensive, particularly for higher frequencies and when applied to range-dependent or three-dimensional environments.[2–5] This study proposes Bayesian optimization (BO) as a sample-efficient method for rapidly obtaining optimal estimates of geoacoustic parameters in high-dimensional parameter spaces. Though the forward model used in this work is computationally inexpensive, the method generalizes to more complex models.

In the forward problem, acoustic propagation models provide predictions of sound fields in underwater environments. The inverse problem, i.e., estimating model parameters given a propagation model and observed data, can be directly solved through matched field processing (MFP), which evaluates samples from the parameter space with the objective of finding parameters that yield a predicted acoustic field that matches the observed data.[6,7] These matches, or correlations, comprise an ambiguity surface that is computed over the parameter space. For two- or three-dimensional parameter space, such as with source localization, MFP is typically implemented as a grid search,[6] where parameters like source range and depth are discretized into equally spaced resolution cells according to physical characteristics of the waveguide.[8] For geoacoustic inversion, grid search—even with coarse resolution—is computationally unfeasible due to the high dimensionality of the parameter space, as computational cost scales exponentially with dimensionality.

To overcome this challenge, we demonstrate BO, a sample-efficient, global optimization strategy that adaptively chooses which samples to evaluate according to previous evaluations of the ambiguity surface.[9–11] Treating the ambiguity surface as an objective function, the goal of BO is to find the global optimum of the objective function with as few function evaluations as possible. At every step of the optimization, BO fits a Gaussian process (GP)[12] surrogate model to the ambiguity surface. GP regression results in a probabilistic interpolator and has been used in a variety of acoustic signal processing tasks.[13–15] In BO, the uncertainty encoded in the GP surrogate model is used by a heuristic acquisition function to suggest which point in parameter space to evaluate next.[16] BO requires no information about the gradient of the ambiguity surface, and though not invulnerable to converging on local optima, its adaptive nature makes it more robust for multimodal objective function optimization than gradient-based methods. BO provides an optimal estimate of the parameters that minimize the objective function but does not inherently provide a full posterior distribution of the parameters. BO was previously

[a]Email: wfjenkins@ucsd.edu

demonstrated for acoustic source localization and array tilt estimation;[17,18] this study expands the parameter space to encompass geoacoustic inversion of seabed parameters.

Previous global optimization methods applied to geoacoustic inversion include simulated annealing (SA),[7] which mimics the annealing process of metals to find the global optimum of an objective function; and genetic algorithms (GA)[19] and differential evolution (DE),[20] which obtain estimates of the optimal parameters by mimicking natural selection and evolution. A drawback of SA, GA, and DE is their computational cost: thousands of evaluations of the forward model are required, and the methods have tunable hyperparameters that require optimization.[1,19,21,22] A detailed history and review of these and other geoacoustic inversion methods, such as Bayesian inference,[23–28] are provided in Ref. 1.

The paper is organized as follows: Sec. II presents the geoacoustic inversion problem, Sec. III describes the BO framework, Sec. IV summarizes the geoacoustic environment and acoustic data, Sec. V presents an analysis of BO results, and Sec. VI provides a discussion of the BO results.

## II. INVERSION FRAMEWORK

### A. Parameterization

The acoustic pressure field $\mathbf{q}(\omega_l) = [q_1 \cdots q_J]^\mathsf{T} \in \mathbb{C}^J$ is sampled by a vertical line array (VLA) with $J$ elements processed at frequencies $\boldsymbol{\Omega} = (\omega_1 \cdots \omega_L)$; $\mathsf{T}$ denotes the transpose operator. Observed data are the sum of predictions of the data $\mathbf{p}(\mathbf{m}, \omega_l) = [p_1 \cdots p_J]^\mathsf{T} \in \mathbb{C}^J$ given a model $\mathbf{m} \in \mathcal{M}^D$, with $D$ parameters, and an error term $\mathbf{e}(\omega_l) = [e_1 \cdots e_J]^\mathsf{T} \in \mathbb{C}^J$:

$$\mathbf{q}(\omega_l) = \mathbf{p}(\mathbf{m}, \omega_l) + \mathbf{e}(\omega_l). \tag{1}$$

Predicted data are modeled by

$$\mathbf{p}(\mathbf{m}, \omega_l) = \mathbf{w}(\mathbf{m}, \omega_l)S(\omega_l), \tag{2}$$

where $\mathbf{w}(\mathbf{m}, \omega_l) \in \mathbb{C}^J$ is a replica pressure field produced by an acoustic propagation model and $S(\omega_l) \in \mathbb{C}$ is an unknown deterministic source term. For brevity, we denote frequency dependence by $\mathbf{q}_l = \mathbf{q}(\omega_l)$, etc. Replica pressure fields $\mathbf{w}_l(\mathbf{m})$ are generated using the KRAKEN normal mode propagation code.[29]

### B. Objective function

The estimated model $\widehat{\mathbf{m}}$ is obtained by matching $\mathbf{q}_l$ and $\mathbf{p}_l$ through minimization of an objective function $\phi$:

$$\widehat{\mathbf{m}} = \arg \min_{\mathcal{M}}[\phi(\mathbf{m})], \tag{3}$$

$$\widehat{\phi} = \phi(\widehat{\mathbf{m}}). \tag{4}$$

$\phi$ is constructed from the Bartlett objective function,[23,30]

$$\phi(\mathbf{m}) = \prod_{l=1}^{L} \left( \text{tr} \, \widehat{\mathbf{R}}_l - \frac{\mathbf{w}_l^\mathsf{H}(\mathbf{m})\widehat{\mathbf{R}}_l\mathbf{w}_l(\mathbf{m})}{\mathbf{w}_l^\mathsf{H}(\mathbf{m})\mathbf{w}_l(\mathbf{m})} \right), \tag{5}$$

where H denotes the Hermitian transpose and $\widehat{\mathbf{R}}_l \in \mathbb{C}^{J \times J}$ is the sample covariance matrix (SCM) from $K$ snapshots:

$$\widehat{\mathbf{R}}_l = \frac{1}{K} \sum_{k=1}^{K} \widehat{\mathbf{q}}_{k,l}\widehat{\mathbf{q}}_{k,l}^\mathsf{H}. \tag{6}$$

The normalization $\widehat{\mathbf{q}} = \mathbf{q}/||\mathbf{q}||$ is introduced to restrict the output of Eq. (5) to the interval $[0, 1]$, as the maximum possible eigenvalue of $\widehat{\mathbf{R}}$ is 1.

## III. BAYESIAN OPTIMIZATION

In BO, a GP surrogate model $\mathbf{f} \sim \mathcal{GP}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ approximates the objective function $\phi(\mathbf{m})$ [Eq. (5)] by performing GP regression over a set of observed data $\mathcal{D}$. Next, an acquisition function takes the GP surrogate model as its input and suggests a new point in $\mathcal{M}$ to evaluate. The process repeats as data are added by exploration of $\mathcal{M}$ until a fixed budget of $N_{\text{total}}$ evaluations, or trials, is exhausted. One iteration of BO is shown in Fig. 1 for a one-dimensional ambiguity surface over source range.

### A. Gaussian process surrogate model

Here we follow the derivations of Ref. 31. Given a set of $N$ points,

$$\mathbf{M} = [\mathbf{m}_1 \cdots \mathbf{m}_N] \in \mathcal{M}^{D \times N}, \tag{7}$$

we model the deterministic objective function $\boldsymbol{\phi} = [\phi(\mathbf{m}_1) \cdots \phi(\mathbf{m}_N)]^\mathsf{T}$ [Eq. (5)] using a probabilistic surrogate model, $\mathbf{f}$, such that

$$\mathbf{f} \approx \boldsymbol{\phi}, \quad \mathbf{f} \sim \mathcal{GP}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{8}$$

where $\mathbf{f} = [f(\mathbf{m}_1) \cdots f(\mathbf{m}_N)]^\mathsf{T}$ is a GP, a finite collection of random variables assumed to have a joint Gaussian distribution. The GP is fully described by a mean function,

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{f}] = [\mu(\mathbf{m}_1) \cdots \mu(\mathbf{m}_N)]^\mathsf{T} \in \mathbb{R}^N, \tag{9}$$

where $\mu(\mathbf{m}_n)$ is the mean at $\mathbf{m}_n$, and a covariance function,

$$\boldsymbol{\Sigma}_{ij} = \mathbb{E}\big[(\phi(\mathbf{m}_i) - \mu(\mathbf{m}_i))(\phi(\mathbf{m}_j) - \mu(\mathbf{m}_j))\big] \tag{10}$$

$$= \mathcal{K}(\mathbf{m}_i, \mathbf{m}_j) \in \mathbb{R}^{N \times N}, \tag{11}$$

where $\mathcal{K}(\mathbf{m}_i, \mathbf{m}_j)$ is a kernel function measuring the similarity between points $\mathbf{m}_i$ and $\mathbf{m}_j$. Any random variable $f(\mathbf{m}_i)$ from the GP is normally distributed with mean $\mu(\mathbf{m}_i)$ and variance $\sigma^2(\mathbf{m}_i) = \boldsymbol{\Sigma}_{ii}$ (Ref. 12, Sec. 2.2):

$$f(\mathbf{m}_n) \sim \mathcal{N}(\mu(\mathbf{m}_n), \sigma^2(\mathbf{m}_n)). \tag{12}$$

Observations of the GP surrogate model are obtained by evaluating Eq. (5) at points $\mathbf{M}$. Equation (5) is deterministic, leading to

J. Acoust. Soc. Am. **156** (2), August 2024
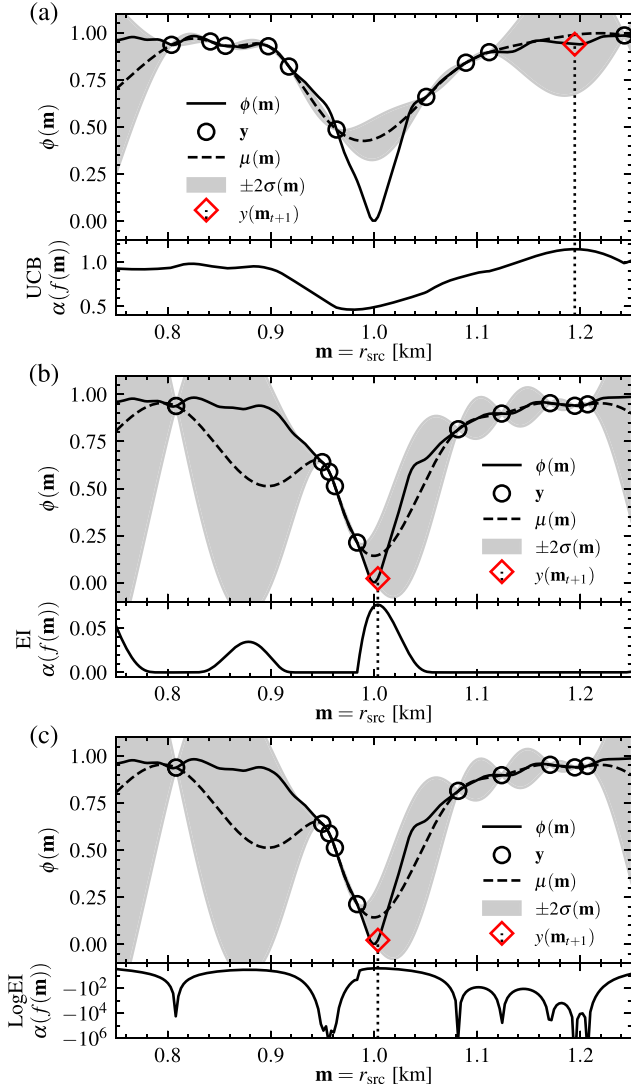
Jenkins *et al.*     813

FIG. 1. (Color online) (Upper panels) Gaussian process (GP) regression on the objective function $\phi(\mathbf{m})$ (solid) for one-dimensional ambiguity surface over source range $r_{src}$. $\phi(\mathbf{m})$ is approximated by the GP mean function $\mu(\mathbf{m})$ (dashed) and uncertainty $\sigma(\mathbf{m})$ (shaded) conditioned on observed data $\mathbf{y}$ (circles). (Lower panels) The next point, $\mathbf{m}_{t+1}$, is obtained from the maximum of the (a) UCB with $\kappa = 1$, (b) EI, and (c) LogEI acquisition functions $\alpha(f(\mathbf{m}))$.

$$f(\mathbf{m}_n) = \phi(\mathbf{m}_n). \tag{13}$$

For numerical stability in subsequent matrix inversions, Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma_y^2)$ with variance $\sigma_y^2 = 10^{-4}$ is added to the observations:

$$y_n = f(\mathbf{m}_n) + \epsilon_n. \tag{14}$$

The observed data $\mathcal{D}$ are then

$$\mathcal{D} = \{(\mathbf{m}_n, y_n), n = 1 : N\} = \{\mathbf{M}, \mathbf{y}\}, \quad \mathbf{y} \in \mathbb{R}^N. \tag{15}$$

The GP predicts $N^*$ unobserved outputs $\mathbf{f}_*$ at test points $\mathbf{M}_{*,D \times N^*} = [\mathbf{m}_1^* \cdots \mathbf{m}_{N^*}^*]$. The joint distribution of the observed process $\mathbf{y}$ and unobserved outputs $\mathbf{f}_*$ at the test points is [Ref. 31, Eq. (17.33)]

$$p(\mathbf{y}, \mathbf{f}_* | \mathbf{M}, \mathbf{M}_*) = \begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix}$$

$$= \mathcal{N}\left( \begin{bmatrix} \boldsymbol{\mu}_M \\ \boldsymbol{\mu}_* \end{bmatrix}, \begin{bmatrix} \widehat{\mathbf{K}}_{M,M} & \mathbf{K}_{M,*} \\ \mathbf{K}_{M,*}^{\mathsf{T}} & \mathbf{K}_{*,*} \end{bmatrix} \right), \tag{16}$$

where $\boldsymbol{\mu}_M$ and $\boldsymbol{\mu}_*$ are the mean functions at $\mathbf{M}$ and $\mathbf{M}_*$, respectively, and

$$\widehat{\mathbf{K}}_{M,M} = \mathbf{K}_{M,M} + \sigma_y^2 \mathbf{I} = \mathcal{K}(\mathbf{M}, \mathbf{M})^{N \times N} + \sigma_y^2 \mathbf{I}, \tag{17}$$

$$\mathbf{K}_{M,*} = \mathcal{K}(\mathbf{M}, \mathbf{M}_*)^{N \times N_*}, \tag{18}$$

$$\mathbf{K}_{*,*} = \mathcal{K}(\mathbf{M}_*, \mathbf{M}_*)^{N_* \times N_*}. \tag{19}$$

The posterior predictive density at the test points $\mathbf{M}_*$ is [Ref. 31, Eq. (17.34)]

$$p(\mathbf{f}_* | \mathcal{D}, \mathbf{M}_*) = \mathcal{N}(\mathbf{f}_* | \boldsymbol{\mu}_{*|M}, \boldsymbol{\Sigma}_{*|M}), \tag{20}$$

$$\boldsymbol{\mu}_{*|M} = \boldsymbol{\mu}_* + \mathbf{K}_{M,*}^{\mathsf{T}} \widehat{\mathbf{K}}_{M,M}^{-1} (\boldsymbol{\phi} - \boldsymbol{\mu}_M), \tag{21}$$

$$\boldsymbol{\Sigma}_{*|M} = \mathbf{K}_{*,*} - \mathbf{K}_{M,*}^{T} \widehat{\mathbf{K}}_{M,M}^{-1} \mathbf{K}_{M,*}. \tag{22}$$

The kernel function of Eq. (11) measures similarity between two points in $\mathcal{M}$. We use a positive-definite and stationary kernel with real-valued inputs, i.e.,

$$\mathcal{K}(\mathbf{m}_i, \mathbf{m}_j) = \mathcal{K}(\mathbf{r}), \quad \mathbf{r} = \mathbf{m}_i - \mathbf{m}_j, \tag{23}$$

and specifically adopt the Matern kernel with smoothness parameter $\nu = 5/2$,

$$\mathcal{K}\left(\mathbf{r}; \frac{5}{2}, \mathbf{l}\right) = \sigma_y^2 \prod_{d=1}^{D} \left( 1 + \frac{\sqrt{5}r_d}{l_d} + \frac{5r_d^2}{3l_d^2} \right) \exp\left( -\frac{\sqrt{5}r_d}{l_d} \right), \tag{24}$$

where $\sigma_y \in \mathbb{R}$ and $l_d \in \mathbb{R}$ are hyperparameters, with $\sigma_y^2$ estimating the noise variance of the GP and $\mathbf{l} = [l_1 \ldots l_D]$ controlling the length scale in dimension $d$ [Ref. 31, Eq. (17.13)].

Kernel hyperparameters $\boldsymbol{\theta} = [\sigma_y^2, \mathbf{l}]$ must be optimized for the GP surrogate model to appropriately reflect the data. Hyperparameter optimization is performed with an empirical Bayes approach. Noting $\boldsymbol{\theta}$ is implicit in $\widehat{\mathbf{K}}_{M,M}$, the product of the following Gaussians,

$$p(\mathbf{f} | \mathbf{M}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{f} | \boldsymbol{\mu}_M, \widehat{\mathbf{K}}_{M,M}) \tag{25}$$

and

$$p(\mathbf{y} | \mathbf{f}, \mathbf{M}) = \prod_{n=1}^{N} \mathcal{N}(y_n | f_n, \sigma_y^2), \tag{26}$$

is marginalized with respect to $\mathbf{f}$, resulting in the marginal likelihood, itself a Gaussian [Ref. 31, Eq. (17.51)]:

$$p(\mathbf{y} | \mathbf{M}, \boldsymbol{\theta}) = \int p(\mathbf{y} | \mathbf{f}, \mathbf{M}, \boldsymbol{\theta}) p(\mathbf{f} | \mathbf{M}, \boldsymbol{\theta}) d\mathbf{f} \tag{27}$$

$$= \mathcal{N}(\mathbf{y} | \boldsymbol{\mu}_M, \widehat{\mathbf{K}}_{M,M}). \tag{28}$$

To find the optimal $\theta$, the log marginal likelihood is maximized [Ref. 31, Eq. (17.52)]:

$$\mathcal{L} = \log p(\mathbf{y}|\mathbf{M}, \theta) = \log \mathcal{N}(\mathbf{y}|\boldsymbol{\mu}_M, \widehat{\mathbf{K}}_{M,M})$$
$$= -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_M)^\mathsf{T}\widehat{\mathbf{K}}_{M,M}^{-1}(\mathbf{y} - \boldsymbol{\mu}_M)$$
$$- \frac{1}{2}\log|\widehat{\mathbf{K}}_{M,M}| - \frac{N}{2}\log(2\pi), \tag{29}$$

$$\frac{\partial \mathcal{L}}{\partial \theta_j} = \frac{1}{2}\mathrm{tr}\left[(\gamma\gamma^\mathsf{T} - \widehat{\mathbf{K}}_{M,M})\frac{\partial \widehat{\mathbf{K}}_{M,M}}{\partial \theta_j}\right], \tag{30}$$

where $\gamma = \widehat{\mathbf{K}}_{M,M}^{-1}(\mathbf{y} - \boldsymbol{\mu}_M)$. The empirical Bayes hyperparameter optimization of Eq. (24) is a form of automatic relevancy determination (ARD) (Ref. 31, Sec. 17.1.2.1). During optimization of an ARD kernel, if one of the parameters in $\mathbf{m}$ does not affect the value of the objective function, that parameter's associated length scale hyperparameter will approach $l_d = \infty$.

## B. Acquisition function

BO is performed sequentially over a fixed budget of $N_{\text{total}}$ trials. At trial $t$, a point $\mathbf{m}_t$ is evaluated by the objective function $\phi(\mathbf{m}_t)$ [Eq. (5)]. The next point $\mathbf{m}_{t+1}$ is suggested through an auxiliary optimization of a heuristic acquisition function $\alpha(f(\mathbf{m}))$:

$$\mathbf{m}_{t+1} = \arg\max_{\mathbf{m}\in\mathcal{M}}[\alpha(f(\mathbf{m}))]. \tag{31}$$

This auxiliary optimization is typically performed using gradient-based methods; further implementation details are provided in Sec. III C.

The ideal acquisition function balances the exploration of regions of high uncertainty with the exploitation of well-performing regions. There are many acquisition functions in the literature, and their development remains a vigorous area of research. We examine the upper confidence bound (UCB), expected improvement (EI), and logarithmically transformed expected improvement (LogEI) acquisition functions for their robust performance and simplicity.

### 1. Upper confidence bound

The UCB acquisition function[16,32] is the weighted sum of the GP mean function $\mu(\mathbf{m})$ and uncertainty $\sigma(\mathbf{m})$,

$$\alpha_{\text{UCB}}(f(\mathbf{m})) = \mu(\mathbf{m}) + \kappa\sigma(\mathbf{m}), \tag{32}$$

where $\kappa$ is a tunable parameter that controls the trade-off between exploration and exploitation. UCB is sensitive to the choice of $\kappa$ and requires tuning to achieve optimal performance.[16] Initially, UCB is dominated by the uncertainty term $\sigma(\mathbf{m})$ and favors exploration of the parameter space [see Fig. 1(a), lower panel]. As more data are added, the GP mean function $\mu(\mathbf{m})$ dominates and UCB exploits the best-performing regions.

### 2. EI

The EI acquisition function adaptively balances exploitation with exploration, requiring no hyperparameter tuning.[11,16] EI quantifies the expected amount of improvement a point is expected to yield over the best previously observed value of the objective function $\widehat{\phi}$ [Eq. (4)] and is defined as

$$\alpha_{\text{EI}}(f(\mathbf{m})) = \mathbb{E}\left[\max(f(\mathbf{m}) - \widehat{\phi}, 0)\right]. \tag{33}$$

Defining the standardized improvement $x$ as

$$x = \frac{\widehat{\phi} - \mu(\mathbf{m})}{\sigma(\mathbf{m})}, \tag{34}$$

Eq. (33) is expressed through integration by parts as a function of the cumulative distribution function $\Phi$ and probability density function $\varphi$:[16]

$$\alpha_{\text{EI}}(f(\mathbf{m})) = \sigma(\mathbf{m})(x\Phi(x) + \varphi(x)). \tag{35}$$

Of note, $\alpha_{\text{EI}} > 0$ but approaches zero in regions where $x \approx 0$ [see Fig. 1(b), lower panel]. This causes the gradient of Eq. (35) to vanish, making gradient-based optimization of Eq. (31) difficult. The vanishing gradient problem worsens as optimization progresses and more data are used for the GP regression, since $\mathbf{f}$ increasingly resembles $\phi$ [Eq. (8)] and the likelihood of finding improved values of $\widehat{\phi}$ decreases.[33] Random restarts of the auxiliary optimization can help mitigate this issue,[34] but if restarts only occur in regions where the gradient has vanished, this approach is reduced to a random search.

### 3. LogEI

To overcome vanishing gradients in Eq. (35), we evaluate LogEI, a log-transformed version of EI:[33]

$$\alpha_{\text{LogEI}}(f(\mathbf{m})) = \log(\sigma(\mathbf{m})) + \log(x\Phi(x) + \varphi(x)). \tag{36}$$

Analytically, LogEI provides the same exploration-exploitation trade-off and suggested trials as EI. A numerically stable implementation of LogEI handles regions where $x \approx 0$ using asymptotic approximations [Ref. 33, Eq. (8)], making LogEI robust to vanishing gradients [see Fig. 1(c), lower panel].

## C. Implementation

BO is implemented using the BoTorch Python library,[34] with pseudocode given in Table I and algorithm parameters in Table II. To establish a prior over the objective function, the GP surrogate model is initialized with $N_{\text{init}}$ warmup trials generated by a Sobol sequence,[35] a quasi-random sequence that generates points in $\mathcal{M}$ such that the points are spread more evenly than when drawn from a uniform distribution.[35] GP regression is performed on the observed data $\mathcal{D}$, and the next trial is suggested by Eq. (31); this repeats until a fixed budget of $N_{\text{total}}$ trials is expended. The ratio between $N_{\text{init}}$ and $N_{\text{total}}$ is a hyperparameter analyzed in Sec. VI.

J. Acoust. Soc. Am. **156** (2), August 2024

Jenkins *et al.*    815

TABLE I. Pseudocode for Bayesian optimization.

---

**Input:** Parameter domain $\mathcal{M}$, objective function $\phi$, kernel function $\mathcal{K}$, acquisition function $\alpha$, total trials $N_{total}$.

**Output:** Best estimate of parameters $\widehat{\mathbf{m}}$

1. $\widehat{\mathbf{m}}, \widehat{\phi} \leftarrow \text{Sobol}[\mathcal{M}, N_{init}]$
2. **for** $t = N_{init} + 1$ **to** $N_{total}$ **do**
3. $\quad \boldsymbol{\mu} \leftarrow \mathbb{E}[f(\mathbf{M})]$ [Eq. (9)]
4. $\quad \boldsymbol{\theta} \leftarrow \text{AdamW}[\mathcal{K}(\mathbf{M}, \mathbf{M}; \boldsymbol{\theta})]$ [Eqs. (29) and (30)]
5. $\quad \boldsymbol{\Sigma} \leftarrow \mathcal{K}(\mathbf{M}, \mathbf{M}; \boldsymbol{\theta})$ [Eq. (11)]
6. $\quad \mathbf{m}_t \leftarrow \arg\max_{\mathbf{m}} \alpha(f(\mathbf{m}))$ [Eq. (31)]
7. $\quad y_t \leftarrow \phi(\mathbf{m}_t)$ [Eq. (14)]
8. $\quad$ **if** $y_t < \widehat{\phi}$ **then**
9. $\quad\quad \widehat{\mathbf{m}} \leftarrow \mathbf{m}_t, \widehat{\phi} \leftarrow y_t$

---

Since the acquisition function can be non-convex but is inexpensive to evaluate, Eq. (31) is optimized through quasi-random Monte Carlo sampling:[34] for $N_{restart}$ random restarts, the acquisition function is evaluated at $N_{acq}$ points drawn from a Sobol sequence, and an auxiliary optimization algorithm is used to find the maximum of the acquisition function. The auxiliary optimization is performed using the bounded limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS-B) algorithm, a quasi-Newtonian method that approximates the Hessian matrix with bounds handling to solve constrained optimization problems.[36,37] The maximum of the $N_{restart}$ auxiliary optimizations is used to suggest the next trial, $\mathbf{m}_{t+1}$. The values for $N_{acq}$ and $N_{restart}$ in Table II are default values implemented in BoTorch. Generally, the greater their values, the more accurate but more computationally expensive the auxiliary optimization becomes. As the dimensionality of $\mathcal{M}$ increases, higher values of $N_{acq}$ and $N_{restart}$ should be considered to ensure the auxiliary optimization is robust.

To improve numerical stability while fitting the GP surrogate model, data are transformed by normalizing model parameters $\mathbf{m}$ to $[0, 1]$ and standardizing $\mathbf{y}$ to zero mean and unit variance. GP regression is aided by assigning prior distributions to the kernel hyperparameters: kernel length scale $\mathbf{l}$ values are drawn from a Gamma distribution with shape $a = 3$ and rate $b = 6$, giving a mean of 0.5; and noise variance $\sigma_y^2$ values are drawn from a Gamma distribution with shape $a = 2$ and rate $b = 0.15$. Kernel hyperparameter optimization is performed using 50 steps of AdamW, an adaptive gradient-based stochastic optimizer.[38]

## IV. DATA AND ENVIRONMENT

### A. Experimental data

BO is demonstrated using data from the SWellEx-96 experiment, conducted off the coast of Southern California

TABLE II. Bayesian optimization algorithm parameters.

| Parameter | Description | Value |
|---|---|---|
| $N_{total}$ | Total trials | 100 |
| $N_{init}$ | Warm-up trials | 64 |
| $N_{acq}$ | Samples for acquisition function auxiliary optimization | 1024 |
| $N_{restart}$ | Restarts of acquisition function auxiliary optimization | 40 |

in shallow water.[39] Data were recorded on a 64-element VLA deployed in 217 m of water with elements spaced evenly between 94.125 m and 212.25 m. The sampling rate of the VLA was 1.5 kHz. During event S5, R/V *Sproul* towed an acoustic source from south to north at a speed of 5 knots and depth of 60 m, with the closest point of approach (CPA) to the VLA occurring 1 km to the east of the array. The source transmitted a comb signal comprising 13 tones between 49 and 388 Hz. Hann-windowed time series data are processed from 21 of the 64 channels using the discrete Fourier transform with $2^{13}$ samples (2.7 s) at $\boldsymbol{\Omega} = [148, 235, 388]$ Hz. $K = 8$ segments with 50% overlap are used to form the SCM [Eq. (6)]. An SCM corresponding to the source's CPA is used for the inversion. At CPA, the estimated signal-to-noise ratios for $\boldsymbol{\Omega}$ are 11.1, 15.7, and 16.0 dB, respectively.

### B. Environment model

The SWellEx-96 geoacoustic and oceanographic environments are well characterized[39,40] and serve as a useful testbed for BO. A range-independent environment model is adopted for simplicity (see Fig. 2). The primary discrepancy in the range-independent assumption is that water depth at the source is 40 m shallower than at the VLA. However, the equivalent range-independent environment is obtained by applying corrections to source range and depth as proposed by Ref. 41.

The water column consists of a shallow mixed layer atop a downward refracting sound speed profile (SSP). The geoacoustic environment is parameterized by two layers of sediment (subscript $s$) and mudrock (subscript $m$) atop a bedrock half-space (subscript $b$). Array tilt $\tau$ is defined as an effective tilt in the source-receiver plane, with positive values indicating the array is tilted toward the source. Model parameters $\mathbf{m}$ and parameter space bounds are listed in
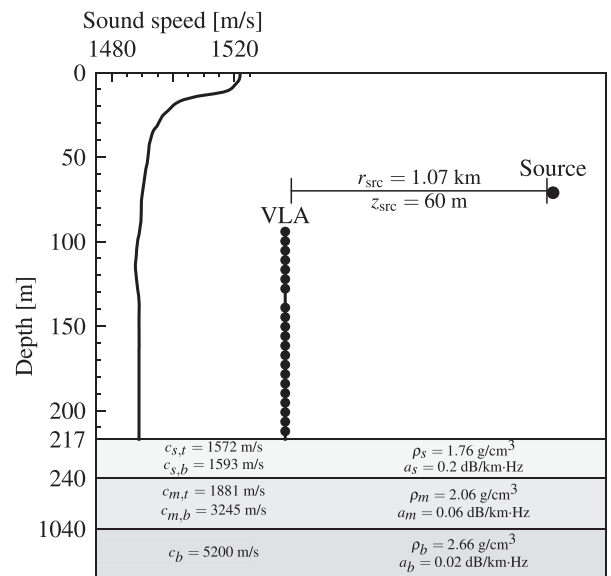


FIG. 2. Range-independent environmental model used for simulations and geoacoustic inversion.

816    J. Acoust. Soc. Am. **156** (2), August 2024

Jenkins *et al.*

TABLE III. Model **m** parameterization and bounds.

| Parameter | Definition | Bounds |
|---|---|---|
| $r_{src}$ (km) | Source range | $[0.8, 1.3]$ |
| $z_{src}$ (m) | Source depth | $[60, 80]$ |
| $\tau$ (°) | Array tilt | $[-3, 3]$ |
| $h_w$ (m) | Water depth | $[212, 222]$ |
| $h_s$ (m) | Sediment thickness | $[15, 25]$ |
| $c_{s,t}$ (m/s) | Sediment top sound speed | $[1560, 1580]$ |
| $\delta c_s$ (m/s) | Sediment SSP change $c_{s,b} = c_{s,t} + \delta c_s$ | $[10, 30]$ |

Table III. At CPA, any discrepancy between *Sproul*'s Global Positioning System range and the source's true range as a result of tow-cable scope is negligible since the VLA is perpendicular to the ship-source axis.

## C. Simulations

Acoustic data are simulated to examine the performance of BO under ideal conditions, i.e., with no environmental mismatch or added noise. Synthetic observations of Eq. (1) are generated with the range-independent parameters in Fig. 2 and with $\mathbf{e}_l = 0$ using the KRAKEN normal mode propagation model.[29] The SCM of the noise-free data is formed by Eq. (6) with $K = 1$.

## D. Sensitivity analysis

Underwater acoustic propagation depends on many parameters, not all of which contribute equally to the acoustic field. To determine approximately which parameters are most important, we perform a sensitivity analysis by sweeping through one parameter in **m** at a time while fixing all others at their anticipated values (see Fig. 3). Though a one-dimensional sensitivity analysis may fail to fully convey higher-dimensional structure in the data, it is a useful tool for coarsely identifying which parameters affect the acoustic field most strongly. From Fig. 3, the acoustic field is most sensitive to source range $r_{src}$, source depth $z_{src}$, and array tilt $\tau$. Simulations indicate the remaining parameters show far lower sensitivity, but experimental data reveal moderate sensitivity to water depth $h_w$, sediment thickness $h_s$, and sound speeds $c_{s,t}$, $c_{s,b}$ at the top and bottom of the sediment layer, respectively.

## V. EXAMPLE

## A. Evaluation method

BO with the upper confidence bound (BO-UCB), expected improvement (BO-EI), and log-expected improvement (BO-LogEI) acquisition functions is demonstrated for both simulated and experimental data using a budget of $N_{total} = 100$ trials and $N_{init} = 64$ warmup trials (36 trials of BO). All computations were performed on an Nvidia DGX workstation (Nvidia, Santa Clara, CA) using a 20-core Intel Xeon E5-2698v4 central processing unit (Intel, Santa Clara, CA) and a Tesla V100 graphics processing unit. One evaluation of the objective function took $\mathcal{O}(0.05)$ s, and GP
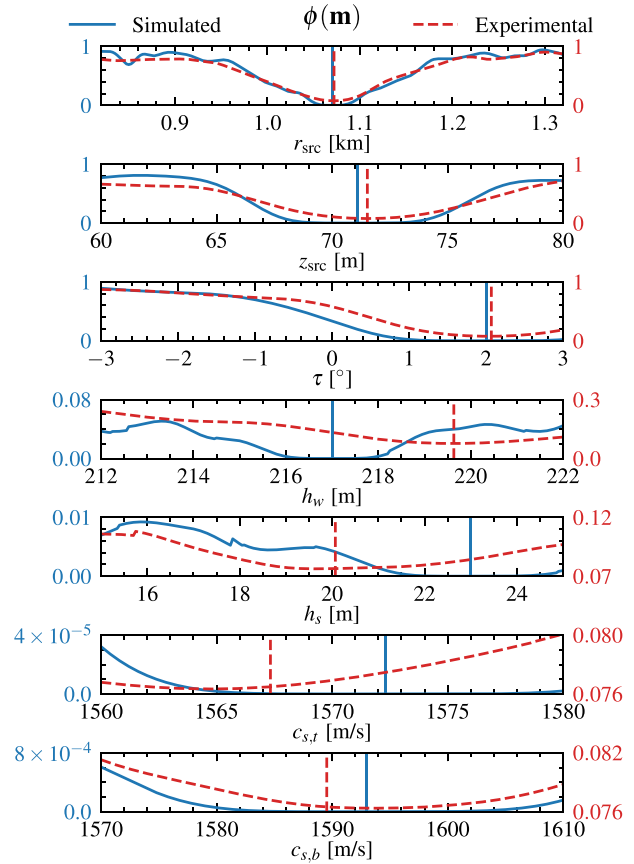


FIG. 3. (Color online) Sensitivity analyses for simulated (blue solid) and experimental (red dashed) data. True simulated and estimated experimental parameter values are indicated by vertical lines.

regression and acquisition function optimization took $\mathcal{O}(1)$ s, $\mathcal{O}(5)$ s, and $\mathcal{O}(10)$ s for the UCB, LogEI, and EI acquisition functions, respectively.

The BO strategies are compared to random search and DE optimization, a stochastic global optimization algorithm related to evolutionary and genetic algorithms.[20,42] Random search performance is examined at the 100th trial to compare against BO on the basis of a limited budget of trials. In practice, a random or grid search would not be restricted to $N_{total} = 100$ trials; thus, the random search is also examined at the 10 000th trial to compare against BO on the basis of elapsed wall time.

DE, which is included as a benchmark for global optimization, is controlled by three hyperparameters set to values per Ref. 42: the population size is set to $10D = 70$, where $D = 7$ is the dimensionality of the parameter space $\mathcal{M}$; the crossover rate is set to 0.7; and the differential weight factor is set to 0.9. DE is initialized with 70 random points in $\mathcal{M}$ and run for 200 iterations. Within each iteration, the objective function is evaluated 70 times to evaluate the fitness of the current population and select the next population. Each DE optimization consists of 14 070 evaluations of $\phi$ and takes approximately 650 s to run (see Fig. 4). Unlike the simulated data, true values of **m** are nearly impossible to obtain from the experimental data directly due

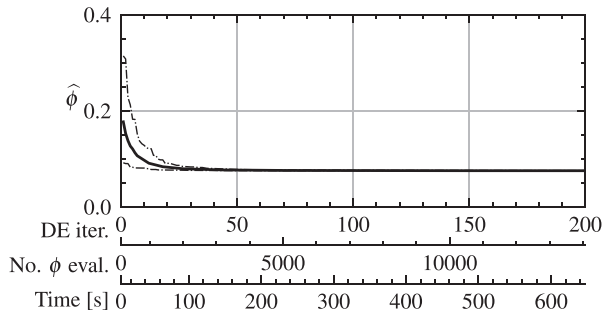J. Acoust. Soc. Am. **156** (2), August 2024

Jenkins *et al.*     817

FIG. 4. Results from differential evolution (DE) optimization on experimental data with the lowest observed values of $\widehat{\phi}$ vs DE iteration, number of evaluations of $\phi$, and wall time. The solid line indicates the mean value at that iteration, and dashed lines indicate minimum and maximum values.

to variability in the ocean and assumption of a range-independent environment for generating replica pressure fields. Thus, we consider DE results as a benchmark against which the BO and random search strategies are evaluated.

Since BO, random search, and DE are sensitive to initialization, 100 random starts are performed for each strategy to broadly characterize performance trends.

## B. Results

Table IV contains a summary of the mean values of $\widehat{\phi}$ and the mean absolute errors (MAE) of the parameter estimates. With the GP surrogate model conditioned on the warmup trials, BO rapidly locates optimal regions to evaluate: given the same trial budget, the BO strategies yield superior optimization performance compared to random search [see Fig. 5(a) and 5(d)]. For simulated data, the 10 000 trials of random search achieve the best performance but over the longest duration [see Fig. 5(b)]. For experimental data, BO achieves better values for $\widehat{\phi}$ than random search

even when the latter is permitted to evaluate two orders of magnitude more trials [see Fig. 5(e)]. With the exception of a small number of outliers, the distribution of final optimization results for BO is centered near optimal values [see Fig. 5(c) and 5(f)]; outliers indicate cases where BO was unable to navigate beyond local optima within the trial budget.

While BO strategies exhibited similar performance to each other in terms of lowest observed $\widehat{\phi}$, BO-UCB achieved this performance faster than BO-EI and BO-LogEI due to the computationally simpler optimization of the UCB acquisition function. From Fig. 5, BO-LogEI was faster than BO-EI due to its robustness to vanishing gradients. However, from Table IV, BO-LogEI exhibited higher MAE than BO-EI, likely due to the log-shape of the acquisition function, as discussed by Ref. 33; if allowed to run for more trials, BO-LogEI may outperform BO-EI in both wall time and MAE. In all cases, BO achieves comparable performance to DE but with two orders of magnitude fewer evaluations of $\phi$.

From Fig. 6, parameter estimates from random search are widely distributed around true simulated values and DE-estimated experimental values of the most sensitive parameters, indicating poor estimation performance of source range $r_{src}$, depth $z_{src}$, and VLA tilt $\tau$ despite low values of $\widehat{\phi}$ indicated in Fig. 5 and Table IV; estimates of water depth $h_w$, sediment thickness $h_s$, and sediment sound speeds $c_{s,t}$, $c_{s,b}$ are uniformly distributed. From Fig. 7, parameter estimates from BO-UCB, the best-performing strategy in this study, are more tightly distributed around the true simulated values and DE-estimated experimental values of $r_{src}$, $z_{src}$, and $\tau$. Estimates of $h_w$, $h_s$, $c_{s,t}$, and $c_{s,b}$ are more widely distributed but exhibit a distribution centered around their true and DE-estimated values. Of note, the shapes of the histograms in Fig. 7 are consistent with the sensitivities shown in Fig. 3, indicating that the optimization can estimate the most

TABLE IV. Mean $\widehat{\phi}$ and MAE of parameter estimates.[a]

| Strategy | Mean $\widehat{\phi}$ | $r_{src}$ (km) | $z_{src}$ (m) | $\tau$ (°) | $h_w$ (m) | $h_s$ (m) | $c_{s,t}$ (m/s) | $c_{s,b}$ (m/s) |
|---|---|---|---|---|---|---|---|---|
| Simulated data | | | | | | | | |
| True | 0.0000 | 1.07 | 71.1 | 2.0 | 217.0 | 23.0 | 1572.3 | 1593.0 |
| DE | $1.07 \times 10^{-11}$ | 0.00 | 0.0 | 0.0 | 0.1 | 0.1 | 1.1 | 1.1 |
| Random (100) | 0.1023 | 0.12 | 5.3 | 2.5 | 2.4 | 3.5 | 5.7 | 6.8 |
| Random (10 000) | **0.0022** | 0.10 | 5.3 | 2.3 | 2.4 | 3.3 | 5.1 | 7.3 |
| BO-UCB | 0.0044 | **0.01** | **0.6** | **0.1** | **1.5** | 2.0 | **4.4** | **5.8** |
| BO-EI | 0.0031 | 0.01 | 0.7 | 0.2 | 1.6 | 2.0 | 5.0 | 7.0 |
| BO-LogEI | 0.0029 | 0.03 | 1.4 | 0.4 | 1.9 | **1.8** | 4.7 | 5.9 |
| Experimental data | | | | | | | | |
| DE | 0.0765 | 1.07 | 71.5 | 2.1 | 219.6 | 20.1 | 1567.3 | 1589.5 |
| Random (100) | 0.2074 | 0.12 | 4.9 | 2.4 | 3.4 | 2.4 | 5.9 | 6.7 |
| Random (10 000) | 0.0920 | 0.12 | 5.1 | 2.3 | 3.3 | 2.5 | 5.3 | 7.4 |
| BO-UCB | **0.0789** | **0.01** | **0.2** | **0.0** | **1.1** | **1.1** | 3.6 | **4.0** |
| BO-EI | 0.0793 | 0.01 | 0.4 | 0.1 | 1.4 | 1.3 | **3.4** | 4.4 |
| BO-LogEI | 0.0790 | 0.02 | 0.8 | 0.2 | 1.5 | 1.4 | 4.0 | 4.4 |

[a]Bold values indicate the lowest mean $\widehat{\phi}$ and mean absolute error (MAE) for each parameter. Differential evolution (DE) for simulated data is for reference only; DE for experimental data is the benchmark.
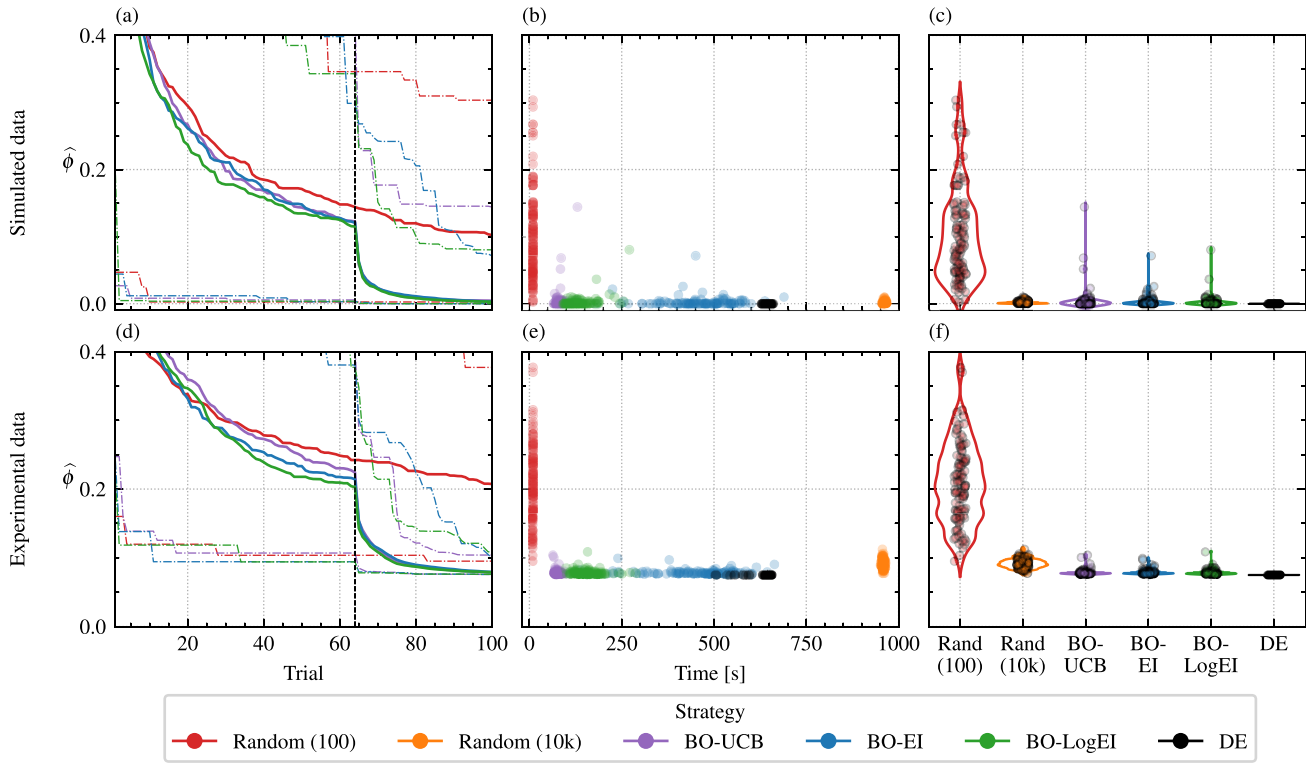
FIG. 5. (Color online) Lowest observed values of $\widehat{\phi}$ for (a)–(c) simulated data and (d)–(f) experimental data. (a) and (d) $\widehat{\phi}$ vs trial. Vertical dashed black lines indicate when Bayesian optimization (BO) commenced after $N_{init} = 64$ warmup trials. Solid colored lines indicate the mean value at that trial, and dashed colored lines indicate minimum and maximum values. (b) and (e) $\widehat{\phi}$ vs wall time at the final trial. (c) and (f) Violin plots of the distribution of final values of $\widehat{\phi}$.

sensitive parameters ($r_{src}$, $z_{src}$, and $\tau$) but is vulnerable to terminating in local optima in the least sensitive parameters ($h_w$, $h_s$, $c_{s,t}$, and $c_{s,b}$).

## VI. DISCUSSION

The previous analyses use $N_{init} = 64$ warmup trials to initialize the GP surrogate model in the BO strategies. Previous work[17] used a high proportion of warmup trials to ensure the GP surrogate model was well-fit to the data and then used BO to rapidly obtain an optimal estimate of source localization. Ultimately this choice is a hyperparameter that can be tuned according to the requirements of the optimization problem. For example, if more rapid optimization is desired, $N_{init}$ can be set to higher values; this enables a better

fit of the GP to the data, but reduces the number of Bayesian trials within a fixed budget. Setting $N_{init}$ too high could preclude BO from exploring the parameter space; conversely, setting $N_{init}$ too low could result in a poorly fit GP surrogate model. $N_{init} = 64$ yields the desirable behavior of rapid optimization with a well-fit GP surrogate model (see Fig. 8).

A key consideration in selecting an optimization strategy is the computational cost and duration of the objective function evaluation. For this study, the objective function evaluation took $\mathcal{O}(0.05)$ s. For such a fast computation, direct-search methods like random search, DE, and Markov chain Monte Carlo (MCMC) sampling are feasible and practical; however, these methods may require a large number of evaluations to converge to the global optimum, especially as
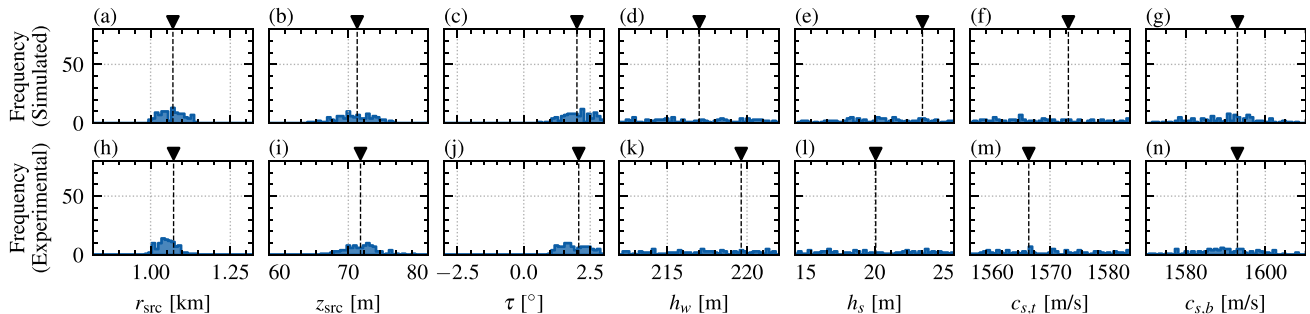


FIG. 6. (Color online) Histograms of parameter estimates of random search with 10 000 trials for (a)–(g) simulated and (h)–(n) experimental data. True and DE-estimated values are indicated by the vertical dashed line and arrowhead for simulated and experimental data.
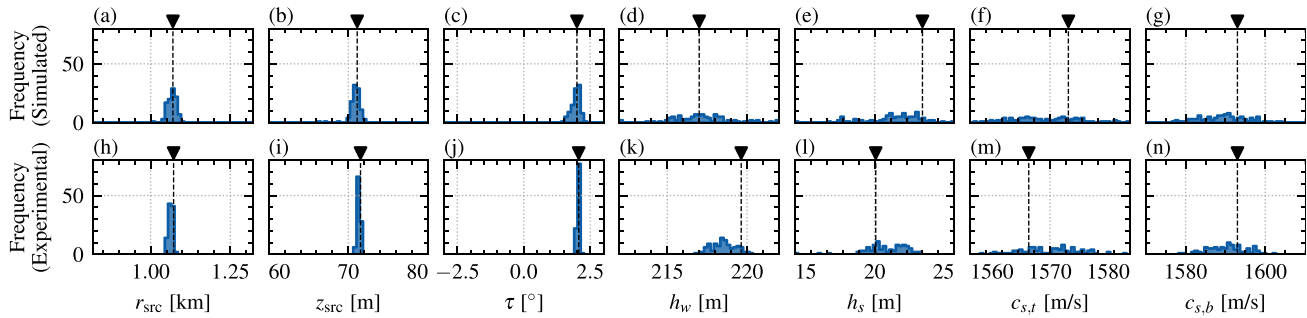
FIG. 7. (Color online) As in Fig. 6, but from Bayesian optimization (BO) with the BO-UCB acquisition function, $N_{total} = 100$, and $N_{init} = 64$.

the dimensionality of the parameter space $\mathcal{M}$ increases. In reality, state-of-the-art inverse problems require objective function evaluations that are computationally expensive, such as those that involve finite element, range-dependent, and three-dimensional propagation models. As the duration of the objective function evaluation increases, the computational cost of direct-search methods increases (see Fig. 9). The advantages of BO therefore become more pronounced as the cost of the objective function evaluation increases.

Because BO operates on the ambiguity surface, its performance relative to the true model parameters is dependent on the quality of the ambiguity surface. For example, in cases with model mismatch or noisy data, sidelobes in the ambiguity surface will become more prominent, making the surface more multimodal and increasing the likelihood that BO might converge on a local optimum. With enough noise or mismatch, sidelobes in the ambiguity surface may obfuscate the peak associated with the true parameters, and BO will estimate the wrong parameters. This dependence is not unique to BO, as any sampling algorithm that makes use of the Bartlett objective in Eq. (5) is susceptible. Like any optimization algorithm, BO can converge on local optima if terminated too early, though various techniques have been implemented to make BO more robust in the optimization of multimodal objective functions, including the development of quasi-Monte Carlo acquisition functions, incorporation of random restart heuristics, and use of trust regions that adaptively update the parameter search space.[34,43]

A flexible and intuitive aspect of BO is the ability to incorporate prior knowledge of the geoacoustic environment by placing prior distributions over the GP surrogate model

hyperparameters and parameter space boundaries. For example, length scales associated with source range and depth can be constrained according to the size of the expected resolution cell,[8] and parameters with little sensitivity can be assigned kernel length scale priors that favor large values to enforce little curvature in that dimension. Though not required, prior knowledge and expertise about the physics of propagation and the geoacoustic environment—e.g., from geologic surveys or computational models—can directly inform the shape of the GP surrogate model and thus inform the search for the global optimum via the acquisition function. The ability to specify kernel functions and hyperparameter priors in multiple dimensions enables the construction of more sophisticated GP surrogate models that can encompass this prior knowledge and has been demonstrated in acoustic sound field reconstruction[13] and for high-dimensional BO problems.[44] Nevertheless, thanks to automatic relevancy determination (Sec. III A), BO remains a viable optimization strategy even in the absence of prior knowledge.

## VII. CONCLUSION

BO efficiently finds regions of optimal performance due to its ability to incorporate observations of the objective function into its decision-making about where to sample next. With a suitably designed kernel function, the GP surrogate model provides a flexible and tractable method to incorporate prior knowledge of the geoacoustic environment into the optimization problem. The heuristic acquisition function relies on the GP to balance exploration of regions
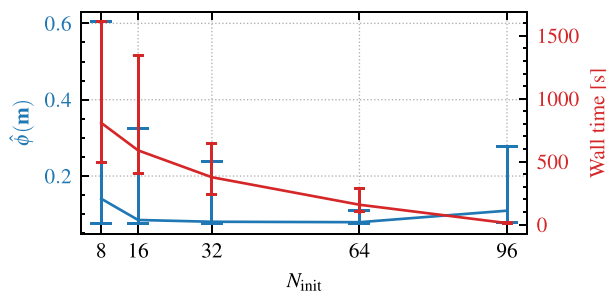


FIG. 8. (Color online) Lowest observed values of $\widehat{\phi}$ (blue) and wall time (red) using BO-LogEI with $N_{total} = 100$ trials and varying $N_{init}$. Mean, minimum, and maximum values are indicated.
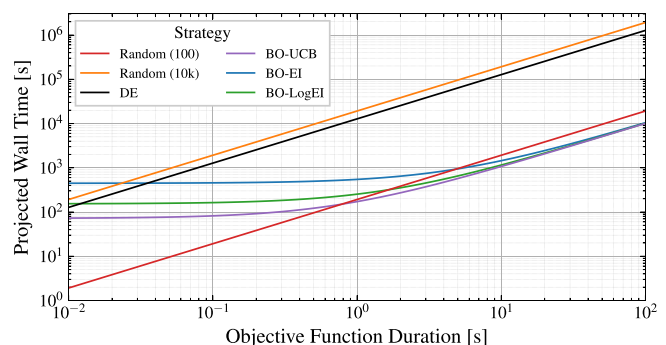


FIG. 9. (Color online) Projected optimization wall time vs duration of each objective function evaluation.

820    J. Acoust. Soc. Am. **156** (2), August 2024

Jenkins *et al.*

of high uncertainty with exploitation of regions where the best observed objective function values reside. The BO framework enables optimization of non-convex objective functions and requires no information about the gradient of the objective function.

Using simulated and real data from a shallow-water experiment, we demonstrate geoacoustic inversion in a seven-dimensional parameter space using BO with just 100 evaluations of the objective function. We show that BO strategies provide more accurate estimates of the geoacoustic parameters than random search and do so in less time. Furthermore, we achieve comparable results to DE, a robust global optimization algorithm that requires two orders of magnitude more evaluations of the objective function than BO. The results of this study suggest that BO is a promising optimization strategy for geoacoustic inversion problems, particularly when the objective function evaluation is computationally expensive or when prior knowledge of the geoacoustic environment can be incorporated into the optimization problem, though the latter consideration is not a strict requirement for BO to be effective.

## ACKNOWLEDGMENTS

## AUTHOR DECLARATIONS
### Conflict of Interest

The authors have no conflicts to disclose.

## DATA AVAILABILITY

The data that support the findings of this study are openly available in Ref. 39 and the code is available at https://github.com/NeptuneProjects/BOGP.

[1] N. R. Chapman and E. C. Shang, "Review of geoacoustic inversion in underwater acoustics," J. Theor. Comput. Acoust. **29**(03), 2130004 (2021).

[2] F. B. Jensen and C. M. Ferla, "Numerical solutions of range-dependent benchmark problems in ocean acoustics," J. Acoust. Soc. Am. **87**(4), 1499–1510 (1990).

[3] A. Tolstoy, "3-D propagation issues and models," J. Comput. Acoust. **04**(3), 243–271 (1996).

[4] Y.-T. Lin, T. F. Duda, and A. E. Newhall, "Three-dimensional sound propagation models using the parabolic-equation approximation and the split-step Fourier method," J. Comput. Acoust. **21**(01), 1250018 (2013).

[5] B. J. DeCourcy and T. F. Duda, "A coupled mode model for omnidirectional three-dimensional underwater sound propagation," J. Acoust. Soc. Am. **148**(1), 51–62 (2020).

[6] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**(4), 401–424 (1993).

[7] M. D. Collins and W. A. Kuperman, "Focalization: Environmental focusing and source localization," J. Acoust. Soc. Am. **90**(3), 1410–1422 (1991).

[8] S. Kim, G. F. Edelmann, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, and T. Akal, "Spatial resolution of time-reversal arrays in shallow water," J. Acoust. Soc. Am. **110**(2), 820–829 (2001).

[9] J. Mockus, "Application of Bayesian approach to numerical methods of global and stochastic optimization," J. Glob. Optim. **4**(4), 347–365 (1994).

[10] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of Bayesian optimization," Proc. IEEE **104**(1), 148–175 (2016).

[11] S. Greenhill, S. Rana, S. Gupta, P. Vellanki, and S. Venkatesh, "Bayesian optimization for adaptive experimental design: A review," IEEE Access **8**, 13937–13948 (2020).

[12] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006).

[13] D. Caviedes-Nozal, N. A. B. Riis, F. M. Heuchel, J. Brunskog, P. Gerstoft, and E. Fernandez-Grande, "Gaussian processes for sound field reconstruction," J. Acoust. Soc. Am. **149**(2), 1107–1119 (2021).

[14] Z.-H. Michalopoulou, P. Gerstoft, and D. Caviedes-Nozal, "Matched field source localization with Gaussian processes," JASA Express Lett. **1**(6), 064801 (2021).

[15] Z.-H. Michalopoulou and P. Gerstoft, "Inversion in an uncertain ocean using Gaussian processes," J. Acoust. Soc. Am. **153**(3), 1600–1611 (2023).

[16] D. R. Jones, "A taxonomy of global optimization methods based on response surfaces," J. Glob. Optim **21**(4), 345–383 (2001).

[17] W. F. Jenkins II, P. Gerstoft, and Y. Park, "Bayesian optimization with Gaussian process surrogate model for source localization," J. Acoust. Soc. Am. **154**(3), 1459–1470 (2023).

[18] W. F. Jenkins and P. Gerstoft, "Bayesian optimization with Gaussian processes for robust localization," in *Proceedings of IEEE ICASSP*, Seoul, Republic of Korea (IEEE, New York, 2024), pp. 6010–6014.

[19] P. Gerstoft, "Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions," J. Acoust. Soc. Am. **95**(2), 770–782 (1994).

[20] Y.-M. Jiang, N. R. Chapman, and P. Gerstoft, "Short range travel time geoacoustic inversion with vertical line array," J. Acoust. Soc. Am. **124**(3), EL135–EL140 (2008).

[21] M. D. Collins, W. A. Kuperman, and W. L. Siegmann, "Propagation and inversion in complex ocean environments," in *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer Academic, Dordrecht, The Netherlands, 1995), pp. 15–20.

[22] M. D. Collins and L. Fishman, "Efficient navigation of parameter landscapes," J. Acoust. Soc. Am. **98**(3), 1637–1644 (1995).

[23] P. Gerstoft and C. F. Mecklenbräuker, "Ocean acoustic inversion with estimation of a posteriori probability distributions," J. Acoust. Soc. Am. **104**(2), 808–819 (1998).

[24] S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**(1), 129–142 (2002).

[25] J. Dettmer, S. E. Dosso, and C. W. Holland, "Model selection and Bayesian inference for high-resolution seabed reflection inversion," J. Acoust. Soc. Am. **125**(2), 706–716 (2009).

[26] J. Dettmer, S. E. Dosso, and C. W. Holland, "Trans-dimensional geoacoustic inversion," J. Acoust. Soc. Am. **128**(6), 3393–3405 (2010).

[27] C. Yardim, P. Gerstoft, and W. S. Hodgkiss, "Geoacoustic and source tracking using particle filtering: Experimental results," J. Acoust. Soc. Am **128**(1), 75–87 (2010).

[28] S. E. Dosso and J. Bonnel, "Joint trans-dimensional inversion for water-column sound speed and seabed geoacoustic models," JASA Express Lett. **3**(6), 060801 (2023).

[29] M. B. Porter, "The KRAKEN normal mode program," SACLANT Undersea Research Centre Memorandum (SM-245)/Naval Research Laboratory Memorandum Report 6920 (1991).

[30] C. F. Mecklenbräuker and P. Gerstoft, "Objective functions for ocean acoustic inversion derived by likelihood methods," J. Comput. Acoust. **08**(2), 259–270 (2000).

[31] K. P. Murphy, *Probabilistic Machine Learning: An Introduction* (MIT Press, Cambridge, MA (2022).

[32] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Information-theoretic regret bounds for Gaussian process optimization in the bandit setting," IEEE Trans. Inf. Theory **58**(5), 3250–3265 (2012).

[33] S. Ament, S. Daulton, D. Eriksson, M. Balandat, and E. Bakshy, "Unexpected improvements to expected improvement for Bayesian optimization," Adv. Neural Inf. Process. Syst. **36**, 20577–20612 (2023).

J. Acoust. Soc. Am. **156** (2), August 2024

Jenkins *et al.*     821

[34]M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, "BoTorch: A framework for efficient Monte-Carlo Bayesian optimization," Adv. Neural Inf. Process. Syst. **33**, 21524–21538 (2020).

[35]I. M. Sobol, "On the distribution of points in a cube and the approximate evaluation of integrals," USSR Comput. Math. Math. Phys **7**(4), 86–112 (1967).

[36]R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," SIAM J. Sci. Comput. **16**(5), 1190–1208 (1995).

[37]C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," ACM Trans. Math. Softw. **23**(4), 550–560 (1997).

[38]I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, New Orleans, LA (2019), available at https://openreview.net/forum?id=Bkg6RiCqY7.

[39]Marine Physical Laboratory, "SWellEx-96 experiment," http://swellex96.ucsd.edu/ (Last viewed January 14, 2024).

[40]R. T. Bachman, P. W. Schey, N. O. Booth, and F. J. Ryan, "Geoacoustic databases for matched-field processing: Preliminary results in shallow water off San Diego, California," J. Acoust. Soc. Am. **99**(4), 2077–2085 (1996).

[41]G. L. D'Spain, J. J. Murray, and W. S. Hodgkiss, "Mirages in shallow water matched field processing," J. Acoust. Soc. Am. **105**(6), 3245–3265 (1999).

[42]R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," J. Glob. Optim. **11**(4), 341–359 (1997).

[43]D. Eriksson and M. Jankowiak, "High-dimensional Bayesian optimization with sparse axis-aligned subspaces," Proc. Mach. Learn. Res. **161**, 493–501 (2021).

[44]L. Papenmeier, L. Nardi, and M. Poloczek, "Increasing the scope as you learn: Adaptive Bayesian optimization in nested subspaces," Adv. Neural Inf. Process. Syst. **35**, 11586–11601 (2022).