



## PAPER

## OPEN ACCESS

RECEIVED  
24 January 2025

REVISED  
28 April 2025

ACCEPTED FOR PUBLICATION  
5 June 2025

PUBLISHED  
13 June 2025

Original content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).

Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.



# Failure modes and mitigations for Bayesian optimization of neuromodulation parameters

Evan M Dastin-van Rijn and Alik S Widge\*

Department of Psychiatry and Behavioral Sciences, University of Minnesota, Minnesota, MN, United States of America

\* Author to whom any correspondence should be addressed.

E-mail: [awidge@umn.edu](mailto:awidge@umn.edu)

**Keywords:** neuromodulation, bayesian optimization, simulation

Supplementary material for this article is available [online](#)

## Abstract

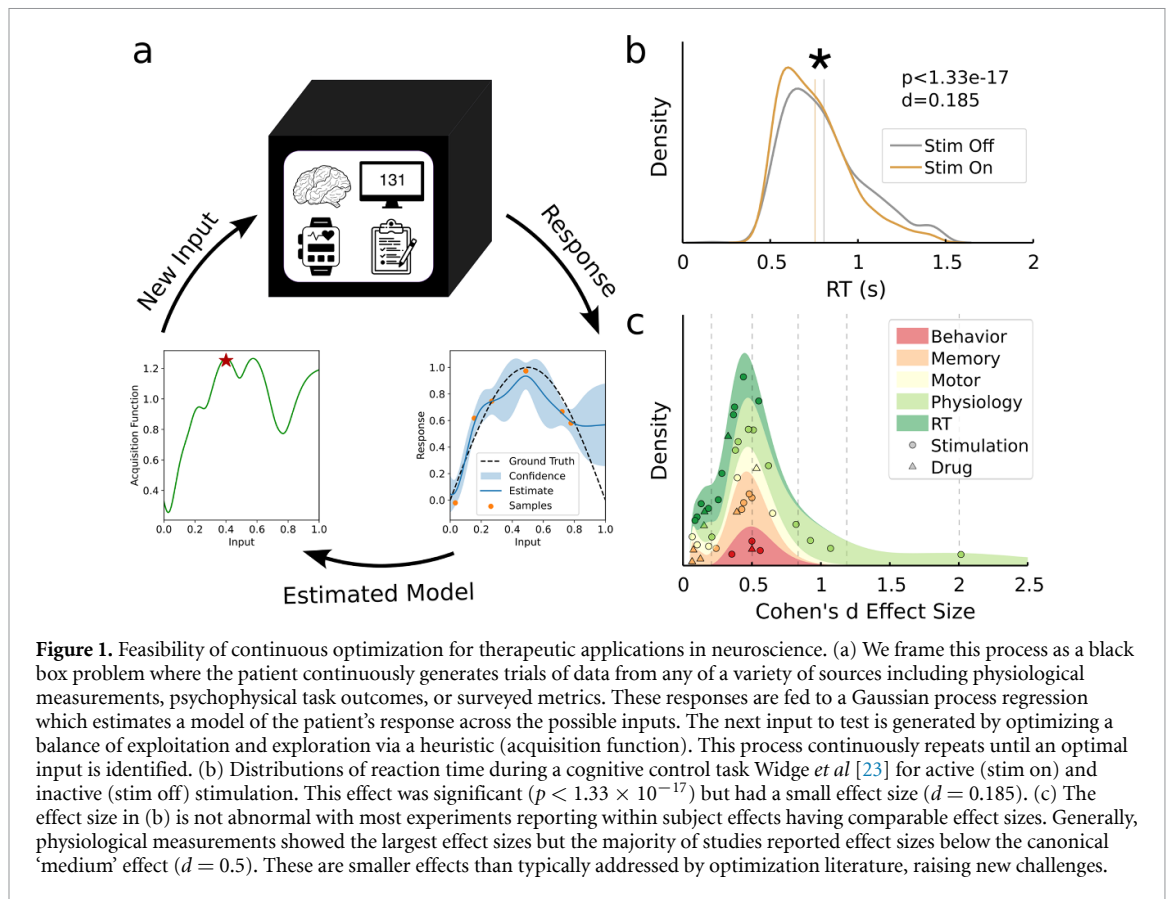
**Objective.** Precision medicine holds substantial promise for tailoring neuromodulation techniques to the symptomatology of individual patients. Precise selection of stimulation parameters for individual patients requires the development of robust optimization techniques. However, standard optimization approaches, like Bayesian optimization, have historically been assessed and developed for applications with far less noise than is typical in neuro-psychiatric outcome measures and with minimal focus on parameter safety. **Approach.** We conducted a literature review of individual effects in neurological and psychiatric applications to build a series of simulated patient responses of varying signal to noise ratio. Using these simulations, we assessed whether existing standards in Bayesian optimization are sufficient for robustly optimizing such effects. **Main results.** For effect sizes below a Cohen's d of 0.3, standard Bayesian optimization methods failed to consistently identify optimal parameters. This failure primarily results from over-sampling of the boundaries of the space as the number of samples increases, because the variance on the edges becomes disproportionately greater than in the remainder of parameter space. Using a combination of an input warp and a boundary avoiding Iterated Brownian-bridge kernel we demonstrated robust Bayesian optimization performance for problems with a Cohen's d effect size as low as 0.1. **Significance.** Our results demonstrate that caution should be taken when applying standard Bayesian optimization in neuromodulation applications with potentially low effect sizes, as standard algorithms are at high risk of converging to local rather than global optima. Mitigating techniques, like boundary avoidance, are effective and should be used to improve robustness.

## 1. Introduction

The human brain is a complex, interconnected system with considerable variability between individuals. This is especially true for diseases of the brain; neurological and psychiatric diseases have substantial variation in symptoms, timecourse, and treatment efficacy that arises from a complex interplay of risk genes, dynamic biological determinants, and environmental factors [1]. Addressing this individual variability is the focus of precision medicine and will be integral to improving current ineffective treatments and reducing disease burden [2, 3]. Neuromodulation is an especially promising technique for applying precision medicine due to the wide array of tunable parameters

including current dose, frequency, pattern, and location. However, for precision medicine to be effective in neuromodulation applications, it will be necessary to develop methods to rapidly identify the treatment parameters that would be most effective for an individual patient [4].

The primary difficulty with developing precision neuromodulation is that the connection between an individual's unique background and their specific disease symptomatology is often a black box problem [5]. Extensive screens of physiological, behavioral, and symptom measures have enabled 'big data' and artificial intelligence methods to make extensive strides towards relating individual patient data to specific treatments [6, 7]. However, these approaches



focus on inferring the treatment for a new patient based on experience with prior patients. While larger, more multimodal datasets will increasingly help to address patient heterogeneity [8], the exact parameters or dosage for an individual patient will inevitably differ somewhat from any predictive model [9]. Truly optimizing a treatment for an individual patient requires an algorithm that can test treatment parameters, obtain a response, and select future parameters based on a balance of exploration and exploitation specific to the patient.

In essence, this is a global function optimization problem where the minimum or maximum of a function (the patient response) is queried based on selected locations in the parameter space (treatment parameters). In the case of neuromodulation, the function being optimized could be a physiological measurement, performance on a behavioral task, or symptom self-report. The parameters, correspondingly, might be intensity, temporal pattern, or location of stimulation. Bayesian optimization is a particularly effective approach for global optimization that has already shown promise in neurological and psychiatric applications [10–16]. Bayesian optimization relies on a surrogate model of the response surface (typically a Gaussian process) (GP) which is constructed from available observations of the function [17, 18]. The parameters at which the function is next evaluated are chosen as the location that maximizes

an acquisition function—a heuristic balancing the predicted value of the function at each location with the uncertainty in the prediction [19–22]. This three step process of evaluation, model fitting, and parameter selection would repeat until the optimal treatment parameters are identified with sufficient certainty (figure 1(a)).

However, the existing literature applying and assessing Bayesian optimization is predominantly focused on applications with relatively high signal to noise ratio [24–26]. While this is an accurate assumption in fields like materials science, robotics, and machine learning, most measurements from the brain and related neural systems are defined by substantial noise [27]. As a result, the within-subject effects that would be leveraged by Bayesian optimization are characterized by small changes relative to the magnitude of the noise. While such effects may seem modest, changes of 5%–10% can be significant and behaviorally meaningful. As an example of such an effect, consider the DBS study conducted by Widge *et al* (figure 1(b)) [23]. In this study, reaction time was used as a readout of cognitive control and was shown to decrease with stimulation. Bayesian optimization could feasibly be used to identify the DBS parameters that would optimize this outcome [12]. However, the effect, while highly significant ( $p < 1.33 \times 10^{-17}$ ), is visibly small relative to the noise resulting in a small Cohen's  $d$  effect size of 0.185. This

effect size is substantially smaller than what is typically addressed by Bayesian optimization and it has not been conclusively demonstrated that the technique is effective at these noise levels.

Additionally, much of the existing Bayesian optimization literature focuses on optimization independent of potential safety concerns associated with the choice of parameters [28, 29]. In patient facing applications like neuromodulation, safety cannot be ignored during the optimization process. Slightly too much current in the wrong location is sufficient to initiate seizures or off-target motor and cognitive effects [30–32]. Sensitivity to treatment parameters is also patient-specific, necessitating an optimization approach that is compatible with complex safety boundaries rather than the simple hypercubes that are typically employed in Bayesian optimization.

To investigate the potential of using Bayesian optimization for neuromodulation in neurological and psychiatric applications, we conducted a set of simulations over a variety of potential response surfaces, problem effect sizes, and safety boundaries informed by a meta-analysis of typical within-subject effects of neuro-stimulation and pharmaceuticals. We found that standard methods for Bayesian optimization perform poorly for problems with low effect sizes, even though low effect size problems are the majority of the applications in neurology and psychiatry. We identified that this poor performance resulted from excessive variance on the boundaries of parameter space and a corresponding over-sampling of these boundaries. Methods that addressed this variance and boundary over-sampling substantially improved optimization performance suggesting that Bayesian optimization, with appropriate adjustments, can be an effective technique for optimizing treatments for nervous system disorders.

## 2. Methods

### 2.1. Effect size meta-analysis

To ensure our simulations accurately represented typical problems in the field, we conducted a meta-analysis of the within-subject effects of neuro-stimulation and pharmaceuticals on a variety of clinically relevant outcomes.

Our review focused on two primary questions:

- 1) What are the typical within-subject effect sizes of stimulation and drug interventions for psychiatric and neurological applications?
- 2) Do these effect sizes depend on the outcome type?

We searched using Google Scholar for articles reporting within-subject measures of mean change

and standard deviation (or directly reporting within-subject effect size) for a neural-related outcome. Effect size was computed using Cohen's  $d$  [33]. We considered articles published in any indexed year. We did not specify which outcome measures to consider in order to broadly assess reported effects on a variety of relevant brain and behavioral metrics. We only included effects that were reported as statistically significant ( $p < 0.05$ ). For articles that reported multiple measurements for the same outcome, we only included the one with the largest effect size [34–46]. Articles with more than one outcome type or intervention were treated as separate studies [35, 37, 39, 41, 44, 47–50]. After reviewing the articles, we divided the outcomes into five categories to identify any potential differences: general behavior (accuracy, performance, etc.) [40, 43, 51, 52], memory [35, 38, 41, 42, 47, 53], motor [44, 49, 50, 54], physiology (brain activity, pupillometry, skin conductance response, etc.) [35, 37, 39, 44–46, 48, 55, 56], and reaction time [23, 34–37, 49, 50, 57–61].

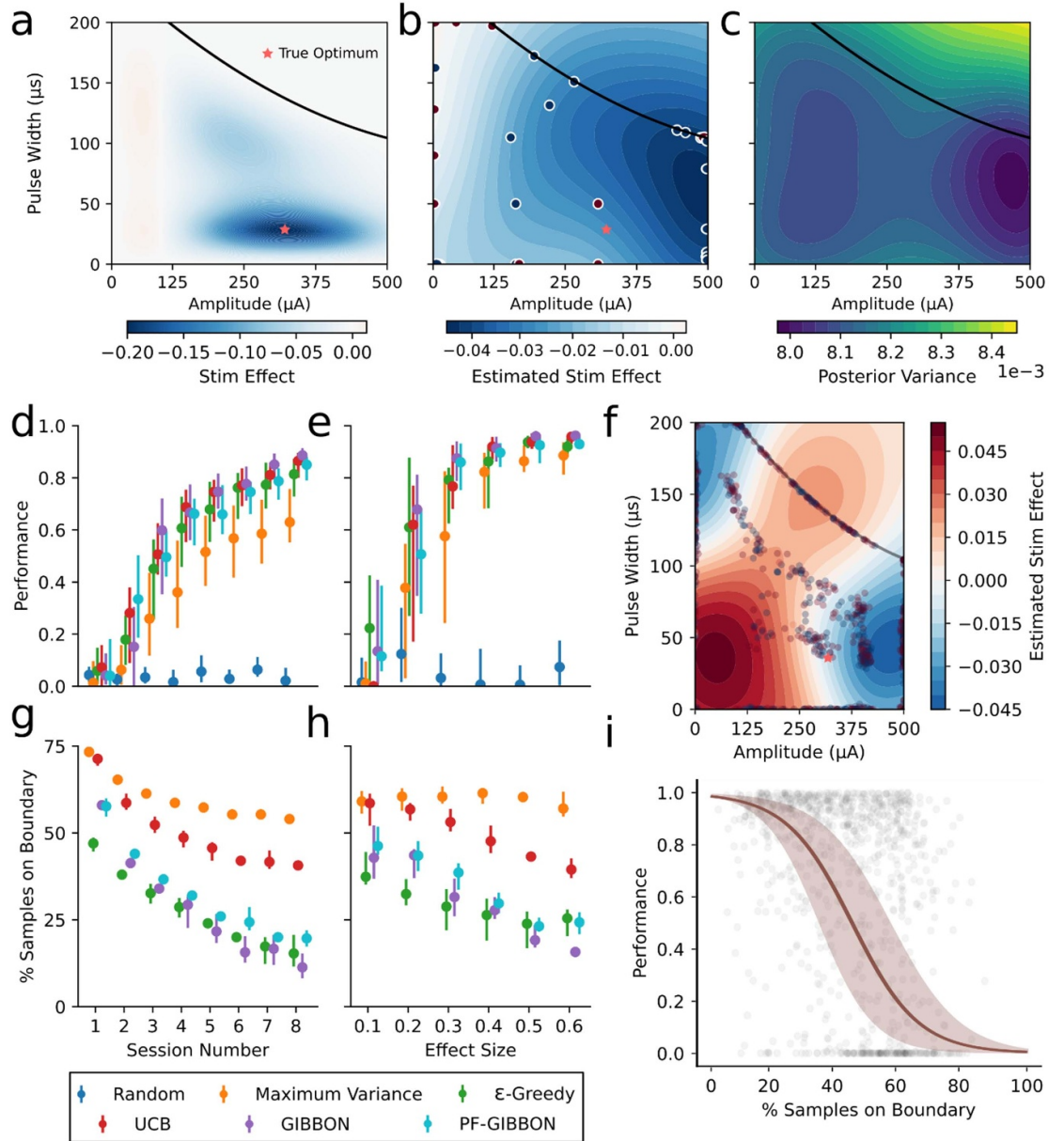
### 2.2. Simulation approach

To assess the impact of effect size on the efficacy of Bayesian optimization, we generated simulated response surfaces for a two-parameter optimization problem based on our prior work using DBS to modulate cognitive control [23, 34, 57] (example surfaces are shown in figure 2(a) and supplemental figure 1). Our two parameters represent the amplitude and pulse width of the stimulation and respectively ranged from 0 to 500  $\mu A$  and 0–200  $\mu s$  corresponding to a reasonable range from prior experiments in rodents [57]. The range of values is arbitrary for the optimization methods but is useful for assessing if the curvature and shape of the surface would be reasonable in practice. To produce the surfaces, 10 sinusoids were generated for each parameter ( $x$  or  $y$ ) with amplitudes ( $A$ ) ranging from  $-0.5$  to  $0.5$  and frequencies ( $F$ ) ranging from 0 to 1 cycle over the length of the parameter space. All sinusoids for both parameters were summed together to produce the full 2D surface maps ( $f^o$ ). The full equation for generating the surfaces is described below:

$$f^o(x, y) = \sum_{i=1}^{10} A_{i,0,0} \sin(2\pi F_{i,0}x) + A_{i,0,1} \cos(2\pi F_{i,0}x) + \sum_{i=1}^{10} A_{i,1,0} \sin(2\pi F_{i,1}y) + A_{i,1,1} \cos(2\pi F_{i,1}y) \quad (1)$$

$$A \sim \text{Uniform}(-0.5, 0.5) \quad (2)$$

$$F \sim \text{Uniform}(0, 1) \quad (3)$$



**Figure 2.** Assessing Bayesian optimization performance for a typical simulated problem. (a) Example simulated ground truth surface used to generate samples as a function of input parameters. (b), (c) Gaussian process posterior mean (b) and variance (c) after 30 example samples for the problem shown in (a). Variance is high in regions with few samples (for example, beyond the safety boundary) and low where there are many samples. (d), (e) Performance of Bayesian optimization for five different acquisition functions, compared against random selection. Performance was calculated as the ratio between the effect of the learned optimal parameters to the ground truth optimal parameters. (d) Performance increased with the number of completed sessions, with all acquisition functions outdoing random selection after a single session. (e) Performance was substantially reduced for low effect size problems with UCB and variance maximization failing to outdo random sampling for the lowest effect size. (f) Example completed optimization for the GIBBON acquisition function showing excessive sampling on the boundaries of the parameter space. (g), (h) Percentage of samples on the boundary of the space for the four acquisition functions. (g) The number of samples on the boundary decreased with each session for all acquisition functions. (h) The boundary was most oversampled for problems with a low effect size. (i) Logit transformed linear regression between the percentage of samples on the boundary and final optimizer performance across the five acquisition functions. 95% confidence interval on the fit is indicated by the shaded region. Problems with a high degree of boundary sampling were significantly more likely ( $p < 7.08 \times 10^{-50}$ ) to have poor optimization performance.

A quadratic safety boundary was also generated for each simulation according to the following equation:

$$y = \frac{(x-h)^2}{a} + k. \quad (4)$$

A quadratic boundary was chosen for this particular simulation as an approximation of a charge delivery curve. The tolerability of stimulation parameters is typically a function of the charge delivered which would be proportional to the product of amplitude and pulse width (an inverse relationship) [62].



We found that a simulation using an inverse equation as a boundary required numerical rather than analytic solutions which would have substantially slowed the analysis. To address this bottleneck, we used a quadratic approximation which we found mimicked an inverse boundary over a reasonable range of values (supplemental figure 2) while allowing for an analytic solution.

The parameters for the boundary were randomly sampled according to the following equations:

$$h \sim \text{Uniform}(500, 1500) \quad (5)$$

$$k \sim \text{Uniform}(0, 200) \quad (6)$$

$$a \sim \text{Uniform}\left(\frac{(500-h)^2}{200-k}, \frac{h^2}{200-k}\right). \quad (7)$$

The portions of the simulated surface near the left, bottom, and safety boundaries were sigmoidally scaled according to the distance ( $d$ ) from the corresponding boundary to mimic negligible effects for zero output and expected decreases in response efficacy along the safety bound. The exact equation for this scaling function ( $s$ ) is shown below:

$$s(d) = \begin{cases} 0 & d < 0 \\ \frac{d^{1.5}}{(0.2-d)^{1.5}} & 0 < d < 0.2 \\ 1 & d > 0.2 \end{cases} \quad (8)$$

This produces a new, modified response surface ( $f^*$ ) according to the following equation, where  $D$  is the set of boundary distances:

$$f^*(x, y) = f^o(x, y) \prod_d^D s(d). \quad (9)$$

The value and location of the maximum and minimum of each surface was identified using the *scipy* function *fmin\_l\_bfgs\_b* [63]. Surfaces were generated to always have the absolute value of the minimum exceed the maximum.

Using the values for the maximum and minimum, each surface was rescaled to have a minimum value corresponding to the effect size (ES) for the problem. This surface ( $f^R$ ) is described by the following equation:

$$f^R(x, y) = \frac{f^*(x, y)}{|\min(f^*(x, y))|} \text{ES}. \quad (10)$$

We modeled effect sizes of 0.1–0.6 in steps of 0.1 by appropriately rescaling the response surfaces and setting the sampling noise to have a standard deviation of 1. Therefore, an individual sample ( $f$ ) from the response surface would be generated according to the following equation:

$$f(x, y) = N(f^R(x, y), 1) \quad (11)$$

20 simulated surfaces were generated per effect size resulting in a total of 240 simulated optimization problems. Examples of surfaces with different optimal locations, presence of multiple local minima, and various safety bounds are shown in figure 2(a) and supplementary figure 1.

### 2.3. Standard Bayesian optimization

Bayesian optimization is a popular method for global optimization that is particularly effective for problems where each evaluation of the function is expensive to compute [17]. This global optimization problem can generally be expressed as follows:

$$\max_{x \in A} f(x) \quad (12)$$

where  $A \subset \mathbb{R}^d$  is the feasible space. The overall approach of Bayesian optimization is to sequentially sample the parameter space at likely or informative locations of the global optimum by taking into account not only the predictions of some surrogate model but also the uncertainty in the predictions. We began each of our optimization protocols by sampling 6 points in the left corner of the parameter space ( $[0, 0, 150, 0, 300, 0, 300, 50, 1, 50, 0]$ ). We then fit a GP as our choice of surrogate model to these initial points. At each iteration, this GP was updated based on the acquired data and an acquisition function was computed from the mean and variance of the GP and maximized to choose the next sampled point according to a heuristic balance of exploration and exploitation. Every 150 trials, the hyperparameters of the GP were re-optimized as this is a more time consuming problem that might be done intermittently after a patient has finished a treatment session, behavioral task, or sequence of measurements. A total of 8 blocks of 150 trials were run for each optimization protocol.

For our GP, we chose a Matern 5/2 kernel with zero mean and fixed output noise of 1 standard deviation for a standard multivariate Gaussian likelihood using the *gpytorch* package [64]. Remaining hyperparameters (lengthscale and output scale) were optimized using the *fit\_gpytorch\_mll* function which is part of the *botorch* package [65]. Acquisition functions were optimized using the *optimize\_acqf* in *botorch* with a nonlinear constraint corresponding to the true safety bound from the simulated response surface. For some applications, this bound would be known in advance (e.g. charge density limits) however many robust methods have been described in the literature to identify the safe regions of parameter space in cases where bounds are unknown [28, 29, 66]. We ran our optimization simulations using five different acquisition functions: maximum variance, upper confidence bound (UCB), general-purpose information-based Bayesian optimization (GIBBON),  $\epsilon$ -greedy, and pareto front-GIBBON. These acquisition functions were chosen based on

their robustness to noise and prior successes in similar applications [12]. Maximum variance is a purely exploratory acquisition function where at each iteration the point in parameter space with the greatest variance is selected. This acquisition function serves as a useful benchmark for determining if the actual optimization process yields greater performance than naively exploring unseen parameters [20]. UCB is an acquisition function that explicitly balances exploration and exploitation by computing a weighted sum of the GP posterior mean and standard deviation:

$$UCB(x, y) = \mu(x, y) + \sqrt{\beta_t} \sigma(x, y) \quad (13)$$

where  $\beta_t$  is scaled as the optimization progresses to ensure a theoretical zero-regret bound [19]:

$$\beta_t = 2 \log \left( t^3 \frac{\pi^2}{3p} \right) \\ p = 0.01. \quad (14)$$

GIBBON is an extension of the max-value entropy search acquisition function which identifies points in the parameter space that would maximally reduce the uncertainty regarding the optimal value of the outcome [21, 22]. GIBBON employs an approximation scheme that allows it to be more computationally-lightweight than standard max-value entropy search while having comparable performance and noise tolerance.  $\varepsilon$ -greedy is an acquisition function that randomly selects points in parameter space that have a pareto optimal tradeoff between exploration (posterior variance) and exploitation (posterior mean) whereas standard acquisition functions like UCB and expected improvement (EI) determine a fixed heuristic tradeoff [20]. At each iteration,  $\varepsilon$ -greedy will either select the predicted, optimal location in parameter space with probability  $p$  or a random point on the exploration-exploitation pareto front with probability  $1 - p$ . For the purposes of our simulations, we fixed  $p$  to a constant value of 0.1. At each iteration, points on the exploration-exploitation pareto front were identified using the NSGA-II algorithm implemented in the *pymoo* package [67, 68]. Based on the observation that GIBBON would occasionally converge to a solution too quickly without adequately exploring the parameter space, we implemented a combination of  $\varepsilon$ -greedy and GIBBON that would use the parameters suggested by GIBBON with probability  $p$  or a random point on the exploration-exploitation pareto front with probability  $1 - p$ . For the purposes of our simulations, we fixed  $p$  to a constant value of 0.3.

The effectiveness of each optimization was evaluated by computing a performance metric at the end of each session. Performance was determined by evaluating the ground truth response surface at the optimal parameters according to the GP posterior mean and normalizing by the effect size for the problem. The optimal parameters from the GP

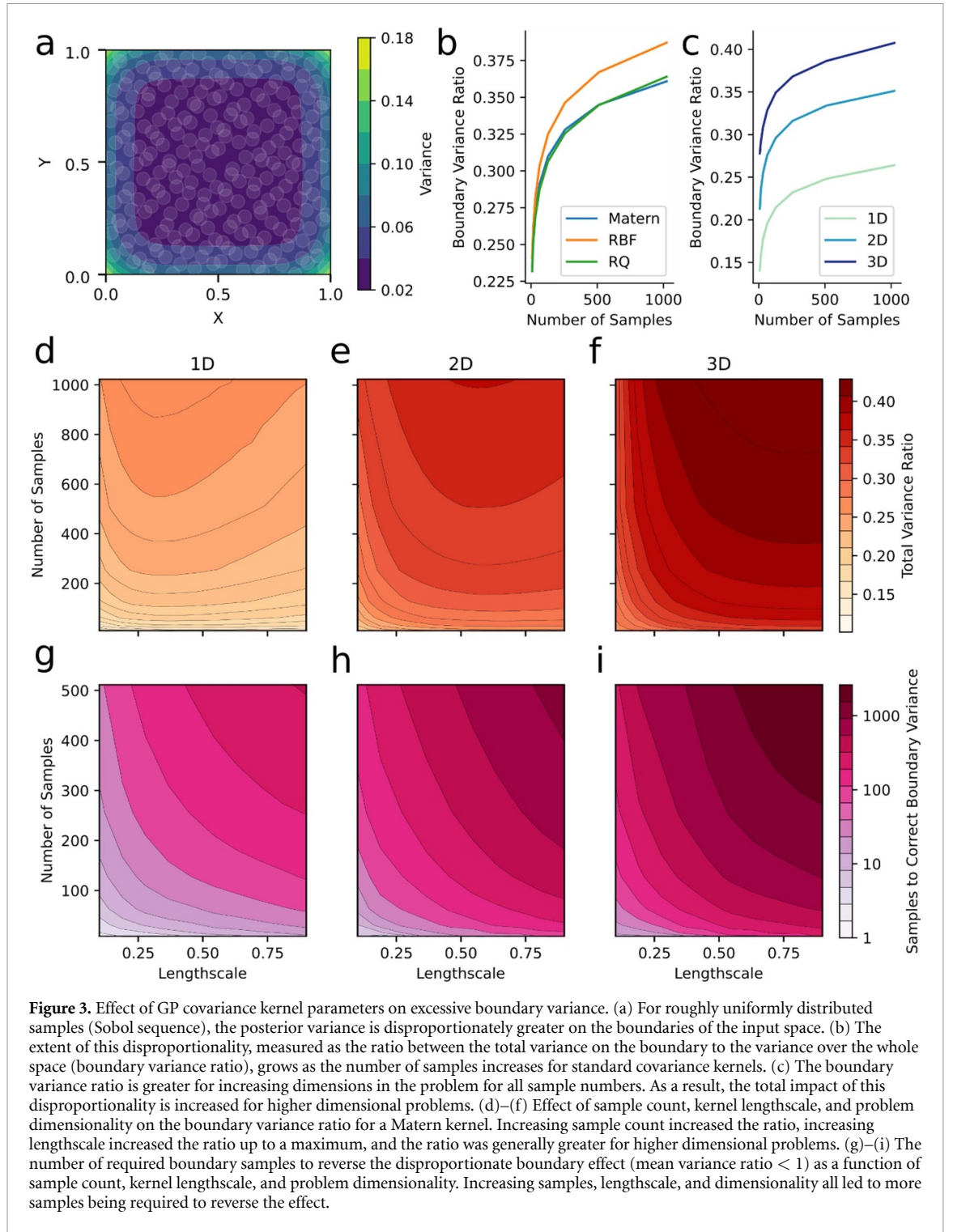
were identified by maximizing the *posterior mean* acquisition function in *Botorch*. This metric is analogous to a normalized version of inverse regret which is commonly employed to evaluate the efficacy of acquisition functions in Bayesian optimization. We also evaluated boundary over-sampling by computing the percentage of samples that were on the boundary of the space or the safety boundary. High values for this percentage would indicate an optimization run where an excessive number of samples were in uninformative areas of parameter space. Performance and boundary over exploration were evaluated across sessions as a function of the session number or effect size. Confidence intervals on median performance were computed using bootstrapping with 1000 resamples via the *bootstrap* method in *scipy* [63]. The relationship between performance and boundary over-exploration at the end of the optimization process was assessed by means of an ordinary least squares regression between the degree of boundary over-exploration and logistically scaled performance using the OLS method of the *statsmodels* package (see figure 2(i) in the results for a visualization of the model) [69].

#### 2.4. Assessment of boundary over-sampling

The source of boundary over-sampling was assessed by a series of simulations using GPs with variable amounts of training data, kernel types, lengthscales, and dimensionality. Training data (number of samples) was sampled to fill the parameter space using a Sobol sequence in steps of powers of 2 ranging from 2 to 1024 using the *SobolEngine* class in *pytorch* (see figure 3(a) in the results for a visualization of typical training data) [70–72]. For kernel comparison, we considered the Matern 5/2, radial basis functions, and rational quadratic. Lengthscales for these kernels were generated by selecting 20 values linearly spaced between 0.1 and 0.9. We considered 1D, 2D, and 3D problems. The primary outcome for these simulations was the discrepancy between the variance on the boundaries and in the remainder of the parameter space. We determined this discrepancy by computing the ratio between the sum of the variance evaluated at points along the boundary of the space and the sum at regularly spaced points in the full parameter space. We term this the boundary variance ratio. We performed a second set of simulations where we computed how many samples were required to correct this discrepancy by sequentially adding samples at the highest variance locations on the boundary and recomputing the GP until the boundary variance ratio was less than one.

#### 2.5. Boundary-avoiding Bayesian optimization

To prevent boundary over exploration during Bayesian optimization, we modified the GP by using an iterated Brownian bridge kernel (IBBK) and an input warp. IBBKs are two parameter kernels



modeled after the Matern family with the key distinction that they are non-stationary [73]. While standard GP kernels have covariance functions that only depend on the distance between the points, the covariance of the IBBK is also a function of the location in parameter space. Specifically, the IBBK forces a covariance of 0 at the boundaries of the parameter space. Therefore, unlike standard kernels that have progressively more variance on the boundary of parameter space over the course of optimization,

the lowest variance for the IBBK is always along the boundary. This has the added effect of making the GP's posterior mean equal to a constant value along the boundary. Since most acquisition functions select samples based on a balance of mean and variance, these two effects will dissuade acquisition functions from sampling along the boundary. The two parameters of the IBBK,  $\beta$  and  $\varepsilon$  are balanced to localize the kernel and address the rate at which the kernel is warped towards the boundaries. For the purposes of

this paper, we fixed these parameters to values of 20 and 50 respectively. These values were chosen because they represent a relatively tight length-scale which can describe a broad range of response surfaces. However, a smaller length-scale does run a greater risk of fitting to local noise. We show some comparative examples of GPs with various IBBK kernel parameters fit to example simulated data in supplemental figure 3.

While the IBBK prevents excessive variance along the *actual* boundaries of the parameter space, it cannot do the same for a boundary *within* the parameter space, like a safety boundary. To counteract this type of over-exploration, we applied a warping function to the input data of the GP that shifted the safety boundary to the actual boundary of the parameter space (example shown in figure 4(e) and a labelled illustration of the method is included in supplemental figure 4). Warping functions have previously been used in other Bayesian optimization applications to address non-stationarity [74]. To begin, we determined the intersection of the safety boundary with the upper ( $x_1$ ) and right boundaries ( $y_1$ ). Using these points, we found the location ( $x_0$ ) along the safety boundary where the path lengths on either side of the point were equal:

$$\frac{\int_{x_1}^{x_0} \sqrt{1 + (dy/dx)^2} dx}{\int_{x_0}^{500} \sqrt{1 + (dy/dx)^2} dx} = \frac{500 - x_1}{200 - y_1} \quad (15)$$

$$dy/dx = \frac{2(x - h)}{a} \quad (16)$$

The warp direction ( $m$ ) is then the slope of the line connecting  $x_0$  to the upper right corner (500, 200) of the parameter space. A similar procedure could be used in higher dimensional spaces by instead using a surface area integral (3D) or volume integral (4D) for the corresponding higher dimensional safety bounds in those spaces.

Subsequently, sampled points could be warped into an adjusted parameter space. For each parameter pair ( $x_i, y_i$ ), we calculated the intersection of the line containing the point and having slope  $m$  with the safety boundary. Parameters that had a point of intersection outside the bounds of the parameter space were not warped. Parameters with intersection points to the left of  $x_0$  were warped towards the upper boundary while those to the right were warped towards the right boundary. The warping distance ( $d_w$ ) was then determined with the following equation:

$$d_w = D_s \frac{D_p - D_o}{D_p} \quad (17)$$

where  $D_s$  is the length of the segment passing through ( $x_i, y_i$ ) and meeting the opposing boundaries,  $D_o$  is the distance from the lower boundaries to ( $x_i, y_i$ ), and

$D_p$  is the length of the segment passing through ( $x_i, y_i$ ) between the lower boundaries and the safety boundary. The coordinates ( $x_*, y_*$ ) of the parameters in the warped space can then be computed as follows:

$$x_* = x_i + \frac{d_w}{\sqrt{m^2 + 1}} \quad (18)$$

$$y_* = y_i + \frac{d_w m}{\sqrt{m^2 + 1}}. \quad (19)$$

This process forces points along the safety boundary to be warped to the upper boundaries of parameter space while points on the lower boundary are not warped at all. Those in between are warped more or less depending on how close they are to the safety boundary. Together with the IBBK, this leads to 0 variance along the safety boundary as well as the true boundaries of parameter space.

Bayesian optimization was performed similarly as in the standard simulations. However, there was no longer a nonlinear constraint since the warped parameter space covered the full range of input values. Additionally, since we kept the hyperparameters of the IBBK fixed, there were no other hyperparameters to optimize with these simulations

### 3. Results

#### 3.1. Effect size meta-analysis

To assess the kinds of problems that could be addressed using Bayesian optimization in neurological and psychiatric applications, we conducted a literature search of typical within-subject effect sizes in the field. Within-subject effect sizes would indicate the size of the change from the intervention relative to the variability in the outcome, to which any optimizer would need to be robust. Our search revealed many studies that reported within-subject changes due to stimulation or drug-based interventions, however, very few reported the necessary information (trial-wise standard deviations) to actually compute a within-subject effect size. Our search identified 42 studies reporting sufficient information to calculate a within-subject effect size (figure 1(c)). Effect-sizes ranged from 0.087 to 2.016 with an average of  $0.418 \pm 0.342$  ( $\pm$ standard deviation), which constitutes a small effect size according to standard descriptors [33]. The largest effect sizes were reported for physiological outcomes ( $0.689 \pm 0.525$ ) while the smallest were reported for reaction time ( $0.271 \pm 0.145$ ). Across the board, these studies indicate that any within-subject optimization would have to contend with a substantial amount of noise.

#### 3.2. Simulation of standard Bayesian optimization

Using the range of effect sizes from our literature search, we conducted a set of simulations to



assess the ability of standard Bayesian optimization to identify the optimal parameters in a standard problem. We modeled a two parameter (amplitude and pulse width) stimulation optimization problem for effect sizes ranging from 0.1 to 0.6 with the goal of identifying the minimum value in the response surface. We generated smooth, simulated surfaces with substantial variability in peak locations and shape including local maxima and minima (supplemental figure 2). Surfaces also included a safety boundary as any optimization of neuromodulation would have to avoid excessive dosage [62]. The response outcome was fixed to 0 at the left, bottom, and safety boundaries to model no dosage or aversive response (figure 2(a)). Bayesian optimization would sequentially sample the space based on a combination of the estimated mean response and variance (figures 2(b) and (c)). Over the course of 8 simulated sessions, all tested acquisition functions showed increasing performance with a maximum performance of 0.886 for GIBBON (figure 2(d)). All acquisition functions outperformed random selection by the eighth session. Maximum variance, a purely exploratory acquisition function, performed the worst, demonstrating that the exploitation–exploration trade-offs of the other acquisition functions were effective. All acquisition functions achieved performances above 80% for effect sizes greater than 0.3 (figure 2(e)). However, for an effect size of 0.1, only GIBBON and  $\varepsilon$ -greedy performed better than random selection after 8 sessions of optimization. Performance for an effect size of 0.2 was mediocre with the best acquisition function, GIBBON, only having a performance of 0.7. For reference, roughly a third of the studies in our literature search had effect sizes below 0.2 (figure 1(c)). After looking at the sampling behavior of our optimization, we noticed that there was substantial over-exploration of the boundaries of parameter space even for the acquisition functions that performed well (figure 2(f)). This over-exploration resulted in poor estimation of the response surface and relative under-sampling of potentially informative regions of parameter space. Additionally, in the context of an optimization problem for a neurological or psychiatric treatment, these samples would have been predominantly ineffective (zero dose) or actively harmful (provoking aversive effects when sampled). We investigated whether this over-sampling depended on the choice of acquisition function or optimization course and found that the worse performing acquisition functions (UCB and maximum variance) had a substantially greater proportion of samples on the boundaries (figure 2(g)). The degree of this over-sampling dropped over the course of the 8 sessions for all acquisition functions. Over-sampling also depended on effect size, with the largest degree of over-sampling occurring for the lower effect sizes for

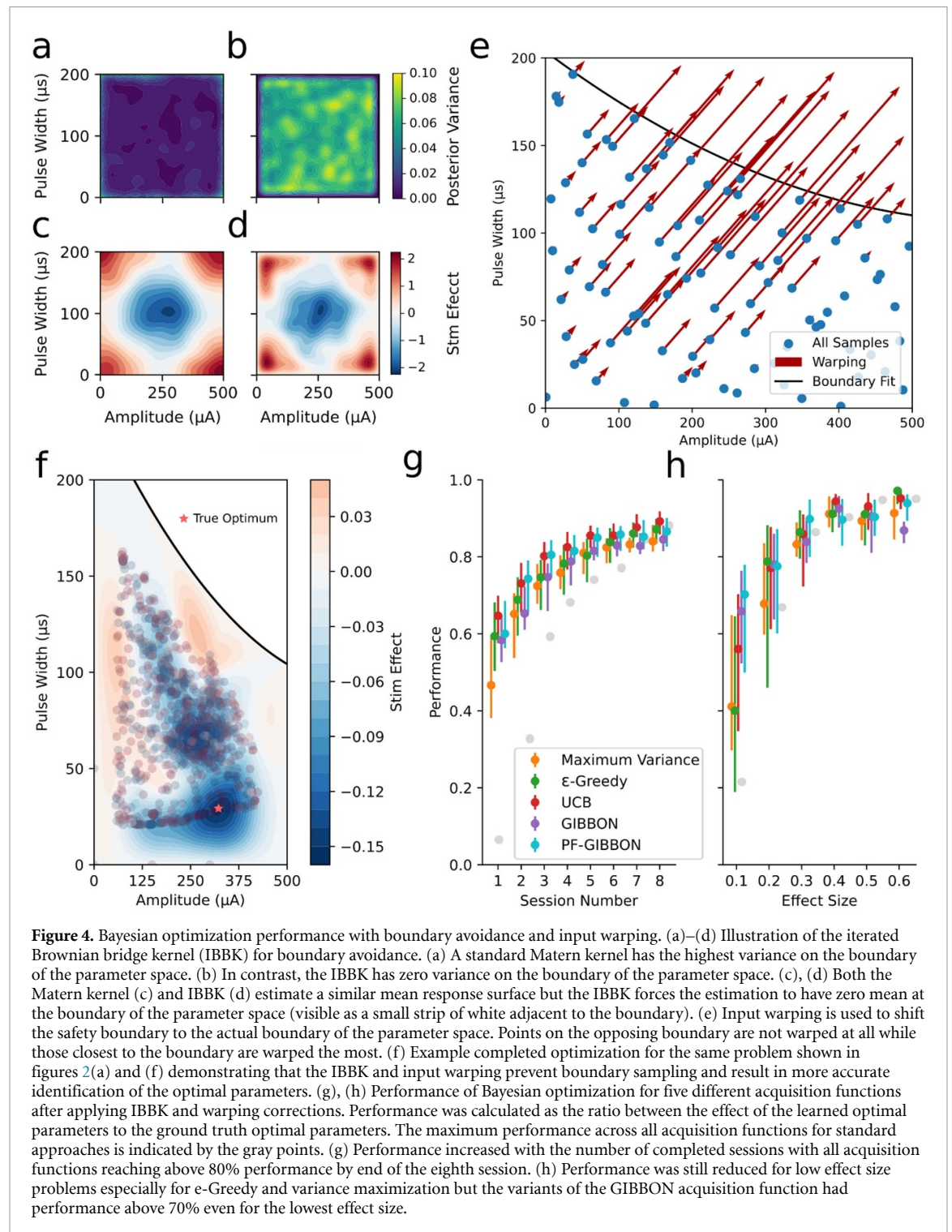
all acquisition functions except maximum variance (figure 2(h)). To determine if there was a direct relationship between over-sampling and optimization performance, we ran a regression between the two quantities for the final session of optimization across all acquisition functions. Over-sampling was significantly associated with poor optimization performance ( $\beta = -0.105$ ,  $p < 7.08 \times 10^{-50}$ ,  $R^2 = 0.168$ ) with performance rapidly falling off as the degree of over-sampling increased.

### 3.3. Assessing boundary over-sampling

To determine why over-sampling occurred with Bayesian optimization, we assessed the behavior of the underlying GP surrogate as the parameter space was filled with samples. We observed that for a uniformly filled parameter space, posterior variance was orders of magnitude greater on the boundaries of the space (figure 3(a)). This discrepancy in variance would result in most common acquisition functions almost exclusively selecting points on the boundaries when choosing exploratory samples. We investigated this variance effect by a series of simulations where we varied the GP kernel family, number of training samples, problem dimensionality, and kernel lengthscale. The severity of the effect was evaluated by computing the ratio between the total variance on the boundary to the total variance in the remainder of the space (boundary variance ratio). For comparable lengthscale in a 2D problem, three common kernel families (Matern, RBF, and RQ) all showed increasing boundary variance ratio as the number of training samples increased (figure 3(b)). The boundary variance ratio also depended on dimensionality, with the 3D simulations having the greatest values across variable amounts of training data (figure 3(c)). These effects were also present across all lengthscales, with larger lengthscales generally showing greater boundary variance ratios regardless of the problem dimensionality or the amount of training data (figures 3(d)–(f)). Together, these results indicate that the excessive variance on the boundary is an innate feature of using GPs with smoother surfaces, and that it increases with problem dimensionality. Lastly, we assessed how many samples on the boundary would be required to balance the variance on the boundary and the remainder of parameter space. As the amount of training data grew, and for larger lengthscales, an exponentially larger number of boundary samples were required for balancing (figures 3(g)–(i)).

### 3.4. Simulation of boundary-avoiding Bayesian optimization

To account for the excessive variance on the boundary of parameter space, we adjusted the GP to use an IBBK and an input warp. Whereas a standard



stationary kernel like the Matern family has the greatest variance on the boundary of parameter space (figure 4(a)), the IBBK has 0 variance on the boundary (figure 4(b)). This leads to the optimizer finding no exploratory value for samples on the boundary. Additionally, the Matern kernel and IBBK estimate the response surface comparably (figures 4(c) and (d)) but the IBBK forces a constant mean along the boundary of the parameter space visible as a white line along the border in figure 4(d). For a GP with

0 mean, this also leads to no exploitative value for points on the boundary. While the IBBK prevents the excessive variance on the boundaries of parameter space, it would not prevent similar variance inflation along a safety boundary. To address this, we warped the inputs to the GP to shift points along the safety boundary to the upper-right edges of the parameter space (figure 4(e)). Together, these measures dissuade the optimizer from selecting points near the boundaries leading to a more effective balance of exploration

and exploitation in informative regions of parameter space and more accurate estimation of the response surface (figure 4(f)). The boundary avoiding optimization substantially improved performance across all 8 sessions for all acquisition functions (figure 4(g)). This improvement was particularly apparent early in the optimization process, where the best performance with boundary avoidance was 0.647 compared to 0.073 without. Boundary avoidance also substantially improved performance for problems with smaller effect sizes with PF-GIBBON having a performance of 0.702 using boundary avoidance compared to a maximum of 0.223 without for an effect size of 0.1.

#### 4. Discussion

Identifying the optimal treatment parameters on an individual basis is the central focus of precision medicine and a promising avenue for improving therapeutic efficacy. However, especially in neuromodulation, substantial noise processes complicate efforts to identify robust relationships between treatment parameters and efficacy. We have demonstrated that standard approaches using Bayesian optimization for identifying this individualized parameter-response relationship fail for problems with relatively low signal to noise ratio. From our literature review, these noise levels are not uncommon and resolving this failure is necessary to ensure optimization methods are sufficiently robust for use in a clinical setting. Through a series of simulations, we identified that excessive variance on the boundaries of parameter space is the primary source of this poor performance. That excessive variance is an inherent feature of the standard kernels used in GP surrogates. We showed that this excessive variance can be addressed on the boundaries of parameter space and safety bounds via the combination of a non-stationary IBBK and input warp. These measures widen the space of problems that can be addressed by Bayesian optimization, improving the future potential for precision medicine in neurology and psychiatry.

In our boundary avoidance simulations, UCB was the overall best performer. However, PF-GIBBON was more effective at lower effect sizes. Srinivas *et al* extensively described regret bounds in their original paper proposing UCB which we leveraged in our implementation [19]. The same rigorous analysis has not been performed for GIBBON (partially for standard max-value entropy search) nor Pareto-front sampling [20–22]. The improvements of PF-GIBBON over standard  $\epsilon$ -greedy and GIBBON suggest that there may be problems or points in the optimization progression where the two approaches complement each other. While we fixed the probability of using either method to a constant value, it is

likely that this weighting could be optimally adjusted over the course of optimization as is done for the  $\beta_t$  term in UCB. This kind of hyperparameter optimization is beyond the scope of this paper but would be important for maximizing performance in future applications.

We identified that the boundary over-sampling hindering the performance of standard Bayesian optimization was primarily caused by excessive variance on the boundary. This largely results from points in the center of parameter space having many nearby prior samples whereas those on the boundary lack neighbors beyond the boundary. As a result, the posterior variance drops more rapidly for samples within parameter space than those on the boundary. This effect is amplified by the increasing surface to volume ratio that happens in higher dimensions, which is a well-documented concern when applying Bayesian optimization in many dimensions [75]. Increasing lengthscales will also worsen this variance discrepancy since an even greater number of points will share variance within parameter space compared to those on the boundary. As many outcomes in neurology and psychiatry have smooth responses to parameter changes [16, 76], this lengthscale effect is particularly problematic without corrective measures.

Aside from the IBBK, other methods have previously been tested for reducing over-sampling of the boundaries in Bayesian optimization. One approach that is often suggested involves augmenting the training data with reflected copies of samples near the boundaries. This technique is commonly used in kernel density estimation for correcting the edges of bounded kernels [77]. Unfortunately, with GPs, this technique incurs a large computational cost by adding a substantial number of samples which is especially problematic as the problem dimensionality and lengthscale increase. Another similar approach involves augmenting the GP with virtual samples from the derivative when the acquisition function would suggest a sample on the boundary [78]. These virtual derivatives are directed to disincentivize sampling of the boundary. While powerful, this technique still increases the amount of training data in the problem. Further, this derivative based sampling approach is not well supported in the majority of existing packages for Bayesian optimization and GPs. The cylindrical kernel, another non-stationary kernel, has also been used to focus sampling during Bayesian optimization on the central portion of parameter space rather than the boundaries [75]. However, this technique still can have issues with excess variance on a hyper-spherical region within parameter space rather than on the boundary itself.

While the IBBK was effective in our simulations, it is not without issues. In cases where the boundaries of parameter space have been selected to

include the full range of ineffective to aversive parameters, the IBBK's 0 variance and a constant mean on the boundary is desirable as it will disincentivize boundary sampling. However, should an optimum exist on the boundary of parameter space, the IBBK would be unable to identify it. This could of course be addressed by widening the bounds but nevertheless presents a point of concern when applying the technique. Additionally, the hyperparameters of the IBBK are discrete making optimization more difficult compared to other kernel families [73]. It is possible that other more flexible kernels could be developed with properties similar to the IBBK, with kernels inspired by the beta distribution being a particularly promising option [79].

All the alternative boundary avoidance methods are unable to handle boundaries within parameter space. These boundaries are especially important for clinical applications because they can be used to model safety cut-offs, dosage limits, or hard ranges on combinations of device parameters [62]. By warping our safety boundary to the actual bounds of parameter space, we were able to take advantage of the IBBK's boundary avoiding features to disincentivize sampling on this boundary as well. Our chosen strategy for input warping resembled a 'shearing' effect which seemed to cause minimal distortion in the response surface. This choice is still somewhat arbitrary and other warps that would concentrate the bound at a single point or edge could be more appropriate in other settings. It is likely that the warp could have the added benefit of improving the stationarity of the GP as the effects of typical interventions are often more sensitive to parameters near safety boundaries [62].

The performance of Bayesian optimization tends to suffer as the number of parameters grows. Especially in the neurostimulation space, this poses a problem as clinicians need to select a number of parameters (site, frequency, amplitude, pulse width, waveform, etc.) which quickly becomes a combinatorial explosion. For the purposes of our simulations, we focused on a 2D case with fixed output variance because we had substantial prior literature to inform the shape of our response surfaces. 3D problems addressing parameters beyond amplitude and pulse width have been underexplored for psychiatric and neurological applications, because the effects of frequency are generally poorly understood and may be non-linear. Similarly, combining the optimization here with optimization of stimulation site or waveform would require a mixed continuous-discrete optimizer. In less established test cases, output noise would likely have to also be estimated during optimization, requiring additional sessions or trials for a robust estimate of an effect. In practice, a hierarchical

optimization process would likely be a promising solution where a range of reasonable parameters could be identified by 'big data' approaches which would subsequently be refined via Bayesian optimization. Strong prior probabilities from past patients would assist in regularization and enable personalization of parameters without the impossibility of searching for a needle in a therapeutic haystack.

## 5. Conclusion

In summary, we have demonstrated that with the proper considerations, Bayesian optimization can be effectively used for the high noise problems typically observed in neurological and psychiatric applications. This will hopefully assist in the development of optimization approaches for personalizing therapies in this field and improving patient outcomes. Future work will focus on refining these algorithms and determining the unique modifications that may be required in individual treatment contexts.

## Data availability statement

The data and code that support the findings of this study are openly available at the following URL/DOI: <https://github.com/tne-lab/NeuroOptimizationPaper>.

## Acknowledgments

This work was supported by the US National Institutes of Health (R01NS120851 to ASW). EDvR was supported by a National Science Foundation Graduate Research Fellowship under Award Number 2237827 and by a MnDRIVE Data Science Initiative Fellowship.


## Conflict of interest

ASW has consulted with Abbott on DBS and anonymously to investors interested in psychiatric indications through expert networks that prohibit him from revealing specific clients. None of those clients involves any ongoing relationship, financial, or otherwise. ASW has received nonfinancial research support from Medtronic and Boston Scientific, companies that manufacture deep brain stimulators. This work is indirectly related to Patent US11241188B2, 'System and methods for monitoring and improving cognitive flexibility' and patent application US20240017069A1, 'Systems and methods for measuring and altering brain activity related to flexible behavior,' both of which name ASW as an inventor.



## ORCID iDs

Evan M Dastin-van Rijn  <https://orcid.org/0000-0002-1428-0723>

Alik S Widge  <https://orcid.org/0000-0001-8510-341X>

## References

- [1] Hampel H, Gao P, Cummings J, Toschi N, Thompson P M, Hu Y, Cho M and Vergallo A 2023 The foundation and architecture of precision medicine in neurology and psychiatry *Trends Neurosci.* **46** 176–98
- [2] Vigo D, Thornicroft G and Atun R 2016 Estimating the true global burden of mental illness *Lancet Psychiatry* **3** 171–8
- [3] Steinmetz J D *et al* 2024 Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the global burden of disease study 2021 *Lancet Neurol.* **23** 344–81
- [4] Widge A S 2024 Closing the loop in psychiatric deep brain stimulation: physiology, psychometrics, and plasticity *Neuropsychopharmacology* **1**–12
- [5] MacEachern S J and Forkert N D 2021 Machine learning for precision medicine *Genome* **64** 416–25
- [6] Koppe G, Meyer-Lindenberg A and Durstewitz D 2021 Deep learning for small and big data in psychiatry *Neuropsychopharmacology* **46** 176–90
- [7] Myszczyńska M A, Ojames P N, Lacoste A M B, Neil D, Saffari A, Mead R, Hautbergue G M, Holbrook J D and Ferraiuolo L 2020 Applications of machine learning to diagnosis and treatment of neurodegenerative diseases *Nat. Rev. Neurol.* **16** 440–56
- [8] Desai R J, Wang S V, Vaduganathan M, Evers T and Schneeweiss S 2020 Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes *JAMA Netw. Open* **3** e1918962
- [9] Afrose S, Song W, Nemeroff C B, Lu C and Yao D 2022 Subpopulation-specific machine learning prognosis for underrepresented patients with double prioritized bias correction *Commun. Med.* **2** 1–14
- [10] Connolly M J, Park S-E, Laxpati N G, Zaidi S A, Ghetiya M, Fernandez A M, Gutekunst C-A and Gross R E 2021 A framework for designing data-driven optimization systems for neural modulation *J. Neural Eng.* **18** 016025
- [11] Cole E R, Connolly M J, Ghetiya M, Sendi M E S, Kashlan A, Eggers T E and Gross R E 2024 SAFE-OPT: a Bayesian optimization algorithm for learning optimal deep brain stimulation parameters with safety constraints *J. Neural Eng.* **21** 046054
- [12] Nagrale S S, Yousefi A, Netoff T I and Widge A S 2023 In silico development and validation of Bayesian methods for optimizing deep brain stimulation to enhance cognitive control *J. Neural Eng.* **20** 036015
- [13] Bonizzato M *et al* 2023 Autonomous optimization of neuroprosthetic stimulation parameters that drive the motor cortex and spinal cord outputs in rats and monkeys *Cell Rep. Med.* **4** 101008
- [14] Dastin-van Rijn E M *et al* 2022 Personalizing dual-target cortical stimulation with Bayesian parameter optimization successfully treats central post-stroke pain: a case report *Brain Sci.* **12** 25
- [15] Louie K H, Petrucci M N, Grado L L, Lu C, Tuite P J, Lamperski A G, MacKinnon C D, Cooper S E and Netoff T I 2021 Semi-automated approaches to optimize deep brain stimulation parameters in Parkinson's disease *J. Neuroeng. Rehabil.* **18** 83
- [16] Stieve B J, Richner T J, Krook-Magnuson C, Netoff T I and Krook-Magnuson E 2022 Optimization of closed-loop electrical stimulation enables robust cerebellar-directed seizure control *Brain* **146** 91–108
- [17] Shahriari B, Swersky K, Wang Z, Adams R P and De Freitas N 2016 Taking the human out of the loop: a review of Bayesian optimization *Proc. IEEE* **104** 148–75
- [18] Frazier P I 2018 A Tutorial on Bayesian Optimization (arXiv:1807.02811v1)
- [19] Srinivas N, Krause A, Kakade S and Seeger M 2010 Gaussian process optimization in the bandit setting: no regret and experimental design *The 27th Int. Conf. on Machine Learning (ICML 2010)*
- [20] De Ath G, Everson R M, Rahat A A M and Fieldsend J E 2021 Greed is good: exploration and exploitation trade-offs in Bayesian optimisation *ACM Trans. Evol. Learn. Optim.* **1** 1–22
- [21] Moss H B, Leslie D S, Gonzalez J and Rayson P 2021 GIBBON: general-purpose information-based Bayesian optimisation *J. Mach. Learn. Res.* **22** 235 (available at: <http://jmlr.org/papers/v22/21-0120.html>)
- [22] Wang Z and Jegelka S 2017 Max-value entropy search for efficient Bayesian optimization *Proc. 34th Int. Conf. on Machine Learning (PMLR)* (available at: <https://proceedings.mlr.press/v70/wang17e.html>) (Accessed 12 June 2024) pp 3627–35
- [23] Widge A S, Zorowitz S, Basu I, Paulk A C, Cash S S, Eskandar E N, Deckersbach T, Miller E K and Dougherty D D 2019 Deep brain stimulation of the internal capsule enhances human cognitive control and prefrontal cortex function *Nat. Commun.* **10** 1536
- [24] Zhang Y, Apley D W and Chen W 2020 Bayesian optimization for materials design with mixed quantitative and qualitative variables *Sci. Rep.* **10** 4924
- [25] Calandra R, Seyfarth A, Peters J and Deisenroth M P 2016 Bayesian optimization for learning gaits under uncertainty *Ann. Math. Artif. Intell.* **76** 5–23
- [26] Nguyen V 2019 Bayesian optimization for accelerating hyper-parameter tuning 2019 *IEEE 2nd Int. Conf. on Artificial Intelligence and Knowledge Engineering (AIKE)* (available at: <https://ieeexplore.ieee.org/document/8791696>) (Accessed 10 December 2024) pp 302–5
- [27] Faisal A A, Selen L P J and Wolpert D M 2008 Noise in the nervous system *Nat. Rev. Neurosci.* **9** 292–303
- [28] Berkenkamp F, Krause A and Schoellig A P 2023 Bayesian optimization with safety constraints: safe and automatic parameter tuning in robotics *Mach. Learn.* **112** 3713–47
- [29] Bottero A G, Luis C E, Vinogradskaya J, Berkenkamp F and Peters J 2022 Information-theoretic safe exploration with Gaussian processes 36th *Conf. on Neural Information Processing Systems (NeurIPS 2022)*
- [30] Kisely S, Hall K, Siskind D, Frater J, Olson S and Crompton D 2014 Deep brain stimulation for obsessive-compulsive disorder: a systematic review and meta-analysis *Psychol. Med.* **44** 3533–42
- [31] Coley E, Farhadi R, Lewis S, Whittle I R, Coley E, Farhadi R, Lewis S and Whittle I R 2009 The incidence of seizures following deep brain stimulating electrode implantation for movement disorders, pain and psychiatric conditions *Br. J. Neurosurg.* **23** 179–83
- [32] Massano J and Garrett C 2012 Deep brain stimulation and cognitive decline in Parkinson's disease: a clinical review *Front. Neurol.* **3** 66
- [33] Cohen J 1988 *Statistical Power Analysis for the Behavioral Sciences* (Routledge) p 579
- [34] Basu I *et al* 2021 Closed-loop enhancement and neural decoding of cognitive control in humans *Nat. Biomed. Eng.* **7** 576–88
- [35] Nikolin S, Martin D, Loo C K and Boonstra T W 2018 Effects of TDCS dosage on working memory in healthy participants *Brain Stimul.* **11** 518–27
- [36] Biel A L, Sterner E, Röhl L and Sauseng P 2022 Modulating verbal working memory with fronto-parietal transcranial electric stimulation at theta frequency: does it work? *Eur. J. Neurosci.* **55** 405–25
- [37] Saville C W N, de Morree H M, Dundon N M, Marcora S M and Klein C 2018 Effects of caffeine on reaction time are

- mediated by attentional rather than motor processes *Psychopharmacology* **235** 749–59
- [38] Alekseiuk I, Turi Z, Amador de Lara G, Antal A and Paulus W 2016 Spatial working memory in humans depends on theta and high gamma synchronization in the prefrontal cortex *Curr. Biol.* **26** 1513–21
- [39] Wienke C, Grueschow M, Haghighi A and Zaehle T 2023 Phasic, event-related transcutaneous auricular vagus nerve stimulation modifies behavioral, pupillary, and low-frequency oscillatory power responses *J. Neurosci.* **43** 6306–19
- [40] Chou T *et al* 2024 A low-intensity transcranial focused ultrasound parameter exploration study of the ventral capsule/ventral striatum *Neuromodulation Technol. Neural Interface* **1** 146–54
- [41] Schuster R M, Mermelstein R J and Hedeker D 2016 Ecological momentary assessment of working memory under conditions of simultaneous marijuana and tobacco use *Addiction* **111** 1466–76
- [42] Balderston N L, Beydler E M, Roberts C, Deng Z D, Radman T, Lago T, Luber B, Lisanby S H, Ernst M and Grillon C 2020 Mechanistic link between right prefrontal cortical activity and anxious arousal revealed using transcranial magnetic stimulation in healthy subjects *Neuropsychopharmacology* **45** 694–702
- [43] Dantas A M, Sack A T, Bruggen E, Jiao P and Schuhmann T 2021 Reduced risk-taking behavior during frontal oscillatory theta band neurostimulation *Brain Res.* **1759** 147365
- [44] Bange M *et al* 2024 Subthalamic stimulation modulates context-dependent effects of beta bursts during fine motor control *Nat. Commun.* **15** 3166
- [45] Keller C J, Huang Y, Herrero J L, Fini M E, Du V, Lado F A, Honey C J and Mehta A D 2018 Induction and quantification of excitability changes in human cortical networks *J. Neurosci.* **38** 5384–98
- [46] Desideri D, Zrenner C, Ziemann U and Belardinelli P 2019 Phase of sensorimotor  $\mu$ -oscillation modulates cortical responses to transcranial magnetic stimulation of the human motor cortex *J. Physiol.* **597** 5671–86
- [47] Prehn K, Stengl H, Grittner U, Kosiolek R, Ölschläger A, Weidemann A and Flöel A 2017 Effects of anodal transcranial direct current stimulation and serotonergic enhancement on memory performance in young and older adults *Neuropsychopharmacology* **42** 551–61
- [48] Mayeli A *et al* 2023 Parieto-occipital ERP indicators of gut mechanosensation in humans *Nat. Commun.* **14** 3398
- [49] Verheij J R, Vyas S and Shenoy K V 2023 Methylphenidate modulates motor cortical dynamics and behavior *bioRxiv* (<https://doi.org/10.1101/2023.10.15.562405>)
- [50] Hu Z, Xu S, Hao M, Xiao Q and Lan N 2019 The impact of evoked cutaneous afferents on voluntary reaching movement in patients with Parkinson's disease *J. Neural Eng.* **16** 036029
- [51] Lustenberger C, Boyle M R, Foulser A A, Mellin J M and Fröhlich F 2015 Functional role of frontal alpha oscillations in creativity *Cortex* **67** 74–82
- [52] Becker S, Gandhi W, Elfassy N M and Schweinhardt P 2013 The role of dopamine in the perceptual modulation of nociceptive stimuli by monetary wins or losses *Eur. J. Neurosci.* **38** 3080–8
- [53] Kahana M J, Ezzyat Y, Wanda P A, Solomon E A, Adamovich-Zeitlin R, Lega B C, Gross R E, Ding K and Diaz-Arrastia R R 2023 Biomarker-guided neuromodulation aids memory in traumatic brain injury *Brain Stimul.* **16** 1086–93
- [54] Leunissen I, Van Steenkiste M, Heise K F, Monteiro T S, Dunovan K, Mantini D, Coxon J P and Swinnen S P 2022 Effects of beta-band and gamma-band rhythmic stimulation on motor inhibition *iScience* **25** 104338
- [55] Van't Wout M, Longo S M, Reddy M K, Philip N S, Bowker M T and Greenberg B D 2017 Transcranial direct current stimulation may modulate extinction memory in posttraumatic stress disorder *Brain Behav.* **7** e00681
- [56] Jung P and Ziemann U 2006 Differences of the ipsilateral silent period in small hand muscles *Muscle Nerve* **34** 431–6
- [57] Reimer A E *et al* 2024 Striatal stimulation enhances cognitive control and evidence processing in rodents and humans *Sci. Transl. Med.* **16** eadp1723
- [58] Weichart E R, Sederberg P B, Sammartino F, Krishna V, Corrigan J D and Rezai A R 2020 Cognitive task performance during titration predicts deep brain stimulation treatment efficacy: evidence from a case study *Front. Psychiatry* **11** 30
- [59] Meteyard L and Holmes N P 2018 TMS SMART—scalp mapping of annoyance ratings and twitches caused by transcranial magnetic stimulation *J. Neurosci. Methods* **299** 34–44
- [60] Waldthaler J, Sperlich A, König A, Stüssel C, Bremmer F, Timmermann L and Pedrosa D 2023 High (130 Hz)- and mid (60 Hz)-frequency deep brain stimulation in the subthalamic nucleus differentially modulate response inhibition: a preliminary combined EEG and eye tracking study *Neuroimage Clin.* **37** 103314
- [61] Ni H C, Hung J, Wu C T, Wu Y Y, Chang C J, Chen R S and Huang Y-Z 2017 The impact of single session intermittent theta-burst stimulation over the dorsolateral prefrontal cortex and posterior superior temporal sulcus on adults with autism spectrum disorder *Front. Neurosci.* **11** 255
- [62] Cogan S F, Hara S and Ludwig K A 2018 Chapter 7—The safe delivery of electrical currents and neuromodulation *Neuromodulation* ed E S Krames, P H Peckham and A R Rezai 2nd edn (Academic) pp 83–94 (available at: [www.sciencedirect.com/science/article/pii/B9780128053539000073](http://www.sciencedirect.com/science/article/pii/B9780128053539000073))
- [63] Virtanen P *et al* 2020 SciPy 1.0: fundamental algorithms for scientific computing in Python *Nat. Methods* **17** 261–72
- [64] Gardner J R, Pleiss G, Bindel D, Weinberger K Q and Wilson A G 2021 GPyTorch: blackbox matrix-matrix Gaussian process inference with GPU acceleration (arxiv:1809.11165)
- [65] Balandat M, Karrer B, Jiang D, Daulton S, Letham B, Wilson A G, Bakshy E 2020 BoTorch: a framework for efficient Monte-Carlo Bayesian optimization *Advances in Neural Information Processing Systems* (Curran Associates, Inc.) (available at: [https://proceedings.neurips.cc/paper\\_files/paper/2020/hash/f5b1b89d98b7286673128a5fb112cb9a-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2020/hash/f5b1b89d98b7286673128a5fb112cb9a-Abstract.html)) (Accessed 13 June 2024) pp 21524–38
- [66] Sui Y, Gotovos A, Burdick J W and Krause A 2015 Safe exploration for optimization with Gaussian processes (extended version with supplementary material) *Proc. of the 32nd Int. Conf. on Machine Learning*
- [67] Blank J and Deb K 2020 Pymoo: multi-objective optimization in Python *IEEE Access* **8** 89497–509
- [68] Deb K, Pratap A, Agarwal S and Meyarivan T 2002 A fast and elitist multiobjective genetic algorithm: NSGA-II *IEEE Trans. Evol. Comput.* **6** 182–97
- [69] Seabold S and Perktold J 2010 *Statsmodels: Econometric and Statistical Modeling with Python* (Austin, Texas) (available at: <https://conference.scipy.org/proceedings/scipy2010/seabold.html>) (Accessed 14 June 2024) pp 92–96
- [70] Sobol' I M 1967 On the distribution of points in a cube and the approximate evaluation of integrals *USSR Comput. Math. Math. Phys.* **7** 86–112
- [71] Owen A B 1998 Scrambling sobol' and niederreiter-xing points *J. Complex* **14** 466–89
- [72] Paszke A *et al* 2019 PyTorch: an imperative style, high-performance deep learning library (arxiv:1912.01703)
- [73] Cavoretto R, Fasshauer G E and McCourt M 2015 An introduction to the Hilbert-Schmidt SVD using iterated brownian bridge kernels *Numer. Algorithms* **68** 393–422
- [74] Snoek J, Swersky K, Zemel R and Adams R 2014 Input warping for bayesian optimization of non-stationary functions *Proc. 31st Int. Conf. on Machine Learning (PMLR)* (available at: <https://proceedings.mlr.press/v32/snoek14.html>) (Accessed 14 June 2024) pp 1674–82

- [75] Oh C, Gavves E and Welling M 2018 BOCK : bayesian optimization with cylindrical kernels *Proc. 35th Int. Conf. on Machine Learning* (PMLR) (available at: <https://proceedings.mlr.press/v80/oh18a.html>) (Accessed 19 June 2024) pp 3868–77
- [76] Grado L L, Johnson M D and Netoff T I 2018 Bayesian adaptive dual control of deep brain stimulation in a computational model of Parkinson's disease *PLoS Comput. Biol.* **14** e1006606
- [77] Jones M C 1993 Simple boundary correction for kernel density estimation *Stat. Comput.* **3** 135–46
- [78] Siivola E, Vehtari A, Vanhatalo J, González J and Andersen M R 2018 Correcting boundary over-exploration deficiencies in Bayesian optimization with virtual derivative sign observations (arxiv:1704.00963)
- [79] Chen S X 1999 Beta kernel estimators for density functions *Comput. Stat. Data Anal.* **31** 131–45