



# Bayesian Optimization with Additive Kernels for a Stepwise Calibration of Simulation Models for Cost-Effectiveness Analysis

David Gómez-Guillén<sup>1,2</sup> · Mireia Díaz<sup>2,3</sup> · Josep Lluís Arcos<sup>4</sup> · Jesus Cerquides<sup>4</sup>

Received: 29 February 2024 / Accepted: 23 August 2024  
© The Author(s) 2024

## Abstract

A critical aspect of simulation models used in cost-effectiveness analysis lies in accurately representing the natural history of diseases, requiring parameters such as probabilities and disease burden rates. While most of these parameters can be sourced from scientific literature, they often require calibration to align with the model's expected outcomes. Traditional optimization methods can be time-consuming and computationally expensive, as they often rely on simplistic heuristics that may not ensure feasible solutions. In this study, we explore using Bayesian optimization to enhance the calibration process by leveraging domain-specific knowledge and exploiting structural properties within the solution space. Specifically, we investigate the impact of additive kernel decomposition and a stepwise approach, which capitalizes on the sequential block structure inherent in simulation models. This approach breaks down large optimization problems into smaller ones without compromising solution quality. In some instances, parameters obtained using this methodology may exhibit less error than those derived from naive calibration techniques. We compare this approach with two state-of-the-art high-dimensional Bayesian Optimization techniques: SAASBO and BAXUS. Our findings demonstrate that Bayesian optimization significantly enhances the calibration process, resulting in faster convergence and improved solutions, particularly for larger simulation models. This improvement is most pronounced when combined with a stepwise calibration methodology.

**Keywords** Bayesian optimization · Gaussian processes · Additive kernels · Model calibration · Simulation models · Cancer research

---

This article is an extension of Gómez et al. [1], presented at CCIA 2023.

✉ David Gómez-Guillén  
dgomez@iconcologia.net

Mireia Díaz  
mireia@iconcologia.net

José Lluís Arcos  
arcos@iia.csic.es

Jesús Cerquides  
cerquide@iia.csic.es

<sup>1</sup> Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain

<sup>2</sup> Institut Català d'Oncologia (ICO) - Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Hospitalet de Llobregat, Spain

<sup>3</sup> Consortium for Biomedical Research in Epidemiology and Public Health - CIBERESP, Carlos III Institute of Health, Madrid, Spain

## 1 Introduction

The use of mathematical simulation models in economic evaluation in medicine is a fundamental tool to inform health-related decision-making. Cost-effectiveness analyses (CEA), a common form of economic evaluation, assess the equilibrium between health benefits and the economic viability of various health interventions. This ensures fair allocation of resources while maximizing health benefits for the population. Mathematical simulation models are pivotal in this process, allowing to assess healthcare interventions and identify those offering the best value over the long term [2]. By comparing costs and benefits, policymakers and healthcare providers can prioritize strategies and allocate resources effectively [3].

<sup>4</sup> Institut d'Investigació en Intel·ligència Artificial - Consell Superior d'Investigacions Científiques (IIIA-CSIC), Bellaterra, Spain

Simulation models used in CEA simulate the progression of disease and the impact of various medical strategies on health outcomes over time [4]. These models generate diverse outcomes, such as average cost and life expectancy, often quantified in Quality-Adjusted Life Years (QALYs) [5]. However, the inherent uncertainty in input parameters requires the calibration of the model to align with established scientific literature values, such as disease incidence or mortality rates.

Model calibration, especially for complex models, presents significant challenges due to the multidimensional optimization problems and arbitrary constraints inherent in the medical domain. State-of-the-art calibration methods for simulation models in CEA are typically general-purpose optimization techniques that don't leverage specific structural knowledge about simulation models [6]. These methods include grid search, random search, and gradient descent techniques such as Nelder–Mead or Simulated Annealing. Also, while parameter selection procedures can effectively reduce dimensionality, traditional methods often need a previous parameter selection stage based solely on clinical relevance [7], rather than using a data-driven approach.

Although these methods are adequate for most simulation models of low to moderate complexity, they fall short for more complex simulations like microsimulation models [8]. The large number of simulations required during the optimization process makes these methods impractical, as the computational cost of each simulation causes the total calibration time to increase rapidly.

In this study, we look into the complexities of model calibration and propose Bayesian Optimization (BO) as a promising solution to enhance efficiency and effectiveness. Our research aims to shed light on novel approaches to calibrating simulation models, particularly focusing on the advantages of BO over conventional methods in healthcare decision-making. We also present a stepwise methodology combined with additive kernels for the calibration of high-dimensional simulation models with a sequential-block structure. We show that this stepwise approach manages to drastically reduce the calibration time while maintaining, and sometimes improving, the quality of the calibrated parameters found in a traditional, naive approach. Finally, we compare this stepwise methodology with state-of-the-art BO methods for high dimensional settings.

The paper is organized as follows. Section 2 includes background information on simulation model calibration, Bayesian Optimization and Gaussian processes. Section 3 describes the simulation models and the calibration methodologies used. Section 4 presents the results of the different calibration experiments, with a discussion of their impact in Sect. 5. Finally, future work is outlined in Sect. 6, while Sect. 7 concludes the paper.

## 2 Background

### 2.1 Simulation Model Calibration in CEA

Simulation models used in CEA typically include decision-analytic models, such as decision trees, Markov models, and discrete event simulations. These models are designed to represent the progression of diseases, the impact of interventions, and the associated costs and health outcomes. For example, Markov models are widely used to simulate chronic diseases where patients transition between different health states over time, allowing the incorporation of recurring costs and varying quality-of-life adjustments.

Model calibration is a critical step in the development of accurate and reliable simulation models. Calibration involves adjusting the model parameters so that the outputs of the model align with real-world data. This process ensures that the model can replicate observed clinical outcomes accurately. Calibration enhances the credibility and validity of the model, making the results of the CEA more robust and applicable to decision-making processes.

However, model calibration poses several challenges, primarily due to the potential complexity of the models. Calibration is essentially an optimization task, and the high-dimensional nature of many models can lead to significant computational difficulties. Moreover, some simulation models require extensive simulations, which can be highly time-consuming and resource-intensive. State-of-the-art calibration techniques used in CEA [6], such as grid search or gradient descent methods, often prove ineffective for more complex models.

Despite these challenges, accurate calibration of simulation models is essential for enhancing the reliability of CEA. Well-calibrated models yield more precise estimates of the cost-effectiveness of interventions, guiding resource allocation in healthcare.

### 2.2 Bayesian Optimization

BO is a robust technique utilized for optimizing costly, opaque functions [9]. Its primary advantage lies in its ability to efficiently locate the global optimum of a function, even in high-dimensional spaces, requiring relatively few evaluations. This efficiency stems from BO's proactive approach to target the most promising areas of the input space for evaluation rather than relying on random evaluations or simplistic heuristics. However, the method's computational cost can escalate, particularly for functions with numerous input variables or complex regression models.

BO has found successful applications across various optimization challenges in machine learning, including hyperparameter tuning, experimental design, and automatic algorithm configuration. It facilitates the rapid identification of

optimal hyperparameter values or experimental conditions without requiring exhaustive exploration of the entire parameter space.

### 2.3 Gaussian Processes

Gaussian Processes (GP) serve as non-parametric regression models, where each observation is represented as a random variable drawn from a normal distribution  $f(x) \sim \mathcal{N}(\mu(x), k(x, x))$  [10]. The kernel function, or covariance function, is instrumental in endowing GPs with their expressive capability, and its selection heavily relies on the nature of the function to be modeled [11]. The squared exponential (SE) kernel, expressed as  $k(x, x') = \sigma^2 e^{-\frac{\|x-x'\|^2}{2l^2}}$ , remains a popular choice despite limitations such as locality and susceptibility to the curse of dimensionality [12].

Modeling high-dimensional functions with a single kernel can be computationally expensive, particularly with local kernels, a prominent research area [13, 14]. Additive kernel decomposition simplifies by breaking it into smaller kernels, enhancing interpretability [15].

However, additive kernels are afflicted by the non-identifiability problem, where kernel hyperparameters cannot be uniquely determined from observed data, posing challenges in model selection and interpretation. To address this, Lu et al. [16] proposed an extension of additive kernels by incorporating an additional constant kernel  $\tilde{k}_{add_0}(x, x')$  with an extra variance hyperparameter  $\sigma_0^2$ , along with an orthogonality constraint, yielding Orthogonal Additive Kernels (OAK). Assuming a normal input distribution  $x_i \sim \mathcal{N}(\mu_i, \delta_i^2)$ , dimensionality  $D$  and kernel lengthscales  $l_i$  the constrained base kernel is derived as follows:

$$\begin{aligned}
 k_{add_{OAK}}(x, x') &= \sum_{i=0}^D \sigma_i^2 \tilde{k}_{add_i}(x_i, x'_i) \\
 \tilde{k}_{add_j}(x, x') &= \sum_{1 \leq i_1 < i_2 < \dots < i_j \leq D} \left[ \prod_{d=1}^j \tilde{k}_{i_d}(x_{i_d}, x'_{i_d}) \right] \\
 \tilde{k}_i(x, x') &= e^{-\frac{(x_i-x'_i)^2}{2l_i^2}} - \frac{l_i \sqrt{l_i^2 + 2\delta_i^2}}{l_i^2 + \delta_i^2} e^{-\frac{(x_i-\mu_i)^2 + (x'_i-\mu_i)^2}{2(l_i^2 + \delta_i^2)}}
 \end{aligned} \tag{1}$$

An important advantage of additive kernels lies in their interpretability, as the  $\sigma_i^2$  terms can be interpreted as the contribution of each individual order to the total kernel. Many problems often hinge on a few low-order interactions, so higher orders can be truncated to limit computational costs while retaining most information in the full decomposition. OAKs are particularly valuable as they accurately identify each contribution, precisely representing the function’s composition.

In this document, the BO method using Gaussian processes with a squared exponential (SE) kernel will be denoted as BO-SE, while the same method employing OAK will be referred to as BO-OAK.

## 3 Methodology

In this section, we will describe the simulation model and the optimization methods we will use for its calibration. These include BO and the rest of the techniques that will be compared.

### 3.1 Simulation Model

Our study uses a lung cancer model previously introduced in a published cost-effectiveness analysis as a rapid benchmark for assessing BO on simulation models [17]. This Markov-based microsimulation model simulates the progression of a cohort through six distinct health states from ages 35 to 79, with monthly intervals. The initial cohort starts at a healthy state and as time passes they go through different events such as cancer development, cancer progression through different stages, cancer survival or death, for example (Fig. 1). Transition probabilities within the model are age-specific, with unique values assigned to each 5-year age group (e.g., 35–39, 40–44, ..., 75–79).

Avoiding some model complexities and with the inherent restrictions imposed on the transition matrices, the parameters by age group are reduced from 36 to 11 [1]. The model was intentionally designed to be computationally efficient, with simulation times averaging under 20 ms. By introducing artificial delays, we can examine the correlation between calibration and simulation times, providing insights for more time-intensive models.

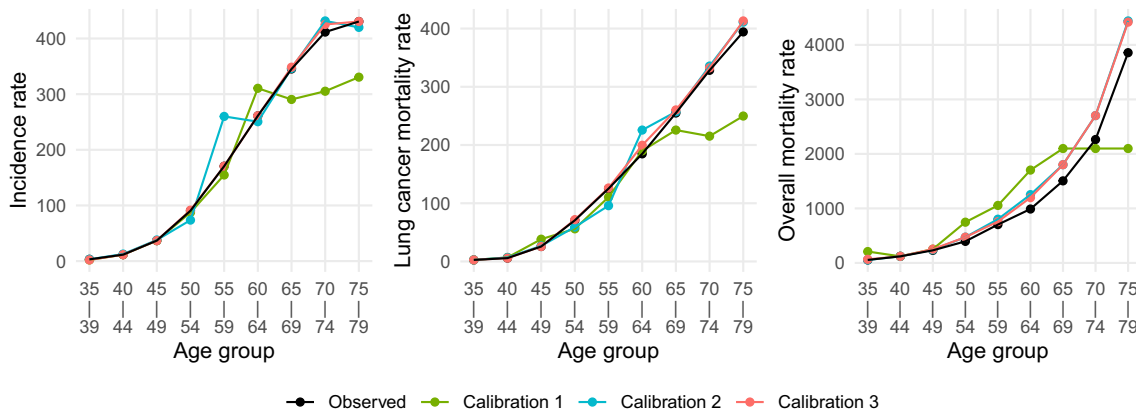
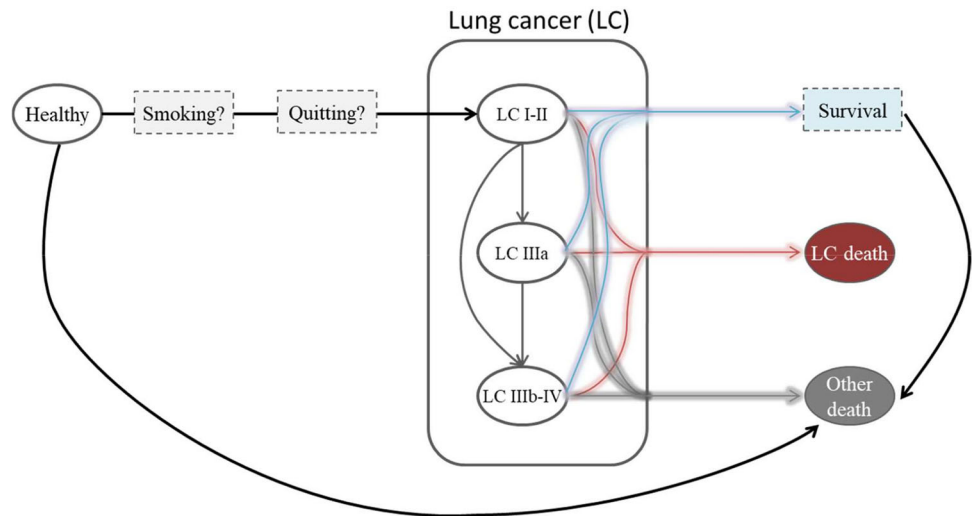
For the calibration process, the weighted sum of Euclidean distances between observed and expected outputs of interest by age groups was used, including lung cancer incidence and mortality (45% each) and mortality from other causes (10%). Figure 2 illustrates these calibration procedures.

### 3.2 Optimization Methods

In health simulation models, it’s typical to start with some estimates derived from scientific literature as starting parameter values. For all optimization experiments conducted, a solution space of  $\pm 50\%$  around these initial values was considered.

BO has been chosen as a promising optimization procedure for simulation models in CEA due to its sample-efficient approach to optimization. This methodology is expected to reduce the number of required simulations, thereby accelerating the calibration process for more time-consuming models.

**Fig. 1** State diagram of the lung cancer simulation model



**Fig. 2** Example of three different calibration results for the lung cancer model, where each attempt tries to match the observed data (black line) as closely as possible

Other traditional optimization methods were tested against BO to compare their performance in error minimization, calibration time and number of simulations required.

Python implementations were employed for the following methods: Nelder–Mead [18], Simulated Annealing (SA) [19], Particle Swarm Optimization (PSO) [20] and BO-SE. Default hyperparameter values were utilized, except for PSO, where the number of particles was 1000 times the number of age groups calibrated.

Additionally, we developed an alternative BO implementation using the R programming language. This was a rapid prototyping environment to assess different optimization process enhancements tailored to our domain. The acquisition function employed was Expected Improvement [21] (Eq. 2, where  $x^+$  represents the best optimum found so far), with PSO utilized to search for its maximum. Implementations of both SE kernels and OAK were incorporated.

$$EI(x) = \mathbb{E}[\max(f(x) - f(x^+), 0)] \tag{2}$$

For more advanced benchmarks, we also include two high-dimensional methods implemented in python. Sparse Axis-Aligned Subspace Bayesian Optimization (SAASBO [22]) employs Bayesian learning to adjust kernel lengthscales through shrinkage, automatically identifying the most critical parameters. Bayesian Optimization with adaptively expanding subspaces (BAXUS [23]) explores subspace embeddings that increase in dimensionality throughout the optimization process and provides theoretical performance guarantees.

To summarize, we will employ a total of seven optimization methods in our study. This includes three traditional methods: Nelder–Mead, Simulated Annealing, and Particle Swarm Optimization (PSO), as well as four Bayesian Optimization (BO) techniques: BO with squared exponential kernels (BO-SE), BO with Orthogonal Additive kernels (BO-OAK), SAASBO, and BAXUS.

### 3.3 Hyperparameter Tuning

Before initiating the BO procedure, we undertake a two-stage process to determine the lengthscales  $l_1, \dots, l_D$  and variances  $\sigma_0^1, \sigma_1^2, \dots, \sigma_n^2$  of the OAK. This strategy allows us to decompose a potentially intricate hyperparameter tuning task for high-dimensional problems into more manageable, lower-dimensional problems.

In the first stage, lengthscales are ascertained by independently maximizing the marginal likelihood for each dimension. This approach is chosen based on the implicit assumption that our simulation models exhibit a robust additive component of order 1 and that a linear combination of one-dimensional kernels can be a reasonable approximation. These optimization subtasks entail  $D$  simple univariate convex problems. For small and large lengthscales, the kernel tends to overfit and underfit, respectively, resulting in models with low likelihood. Each optimum lies within the spectrum between these two extremes and can be swiftly determined using a straightforward binary search.

In the second stage, the marginal likelihood is maximized for the entire set of variances, considering the lengthscales obtained in the previous stage. This task is a  $(D + 1)$ -dimensional optimization subtask, solved using the Nelder–Mead algorithm [18].

### 3.4 Stepwise Calibration

We will also use a stepwise calibration approach as an alternative to the naive calibration. The rationale behind this method is that these simulation models follow a time-dependent sequential block layout. In this structure, the simulation output for a specific age group relies solely on the pertinent parameters for that age group and those preceding it without being influenced by those of later age groups.

Let  $f(p)$  denote the output of the complete simulation model with parameters  $p = \{p_1, p_2, \dots, p_n\}$  and let  $k$  represent the number of age groups. We can make a partition  $p^{(i)}$  of these parameters based on the age group  $i$  they influence, such that  $p = p^{(1)} \cup p^{(2)} \cup \dots \cup p^{(k)}$  and  $\bigcap_{1 \leq i \leq k} p^{(i)} = \emptyset$ . Using this partition and the output for each age group  $s_i$ , we can formulate the final output of the model  $f(p) = s_k$  as a sequence of functions  $f_i$ :

$$\begin{aligned}
 s_1 &= f_1(p^{(1)}) \\
 s_2 &= f_2(p^{(2)} \mid s_1) = f_2(p^{(2)} \mid p^{(1)}) \\
 s_3 &= f_3(p^{(3)} \mid s_2) = f_3(p^{(3)} \mid p^{(1)} \cup p^{(2)}) \\
 &\dots \\
 f(p) = s_k &= f_k(p^{(k)} \mid s_{k-1}) = f_k(p^{(k)} \mid p^{(1)} \cup p^{(2)} \cup \dots \cup p^{(k-1)})
 \end{aligned}
 \tag{3}$$

The last term of each line is also conditional to the set of previous  $f_i$ , but they have been omitted from the notation for clarity purposes.

This decomposition allows a greedy approach to break down the calibration of the full model  $f(p)$  into  $k$  sequential tasks. This involves the iterative calibration of each age group, considering the previously calibrated parameter sets, as outlined in Algorithm 1.

---

#### Algorithm 1 Stepwise calibration with $k$ age groups

---

**Require:** target  $\leftarrow [t_1, \dots, t_k]$

**for**  $i \leftarrow 1$  to  $k$  **do**

$$p_*^{(i)} \leftarrow \operatorname{argmin}_{p^{(i)}} \left\| f_i \left( p^{(i)} \mid \bigcup_{j < i} p_*^{(j)} \right) - t_i \right\|$$

**end for**

**return**  $[p_*^{(1)}, \dots, p_*^{(k)}]$

---

As expected from a greedy method, the optimum for each step does not guarantee to be part of a global optimum, and each solution found by these calibration subtasks might potentially lead to suboptimal solutions in subsequent subtasks. On the other hand, this method reduces the complexity of the full calibration into  $k$  smaller optimization problems, thereby reducing the overall computational effort required. The objective with this approach is to enable the use of simple BO methods in this setting while mitigating the downsides of high dimensionality.

## 4 Results

### 4.1 Naive Calibration

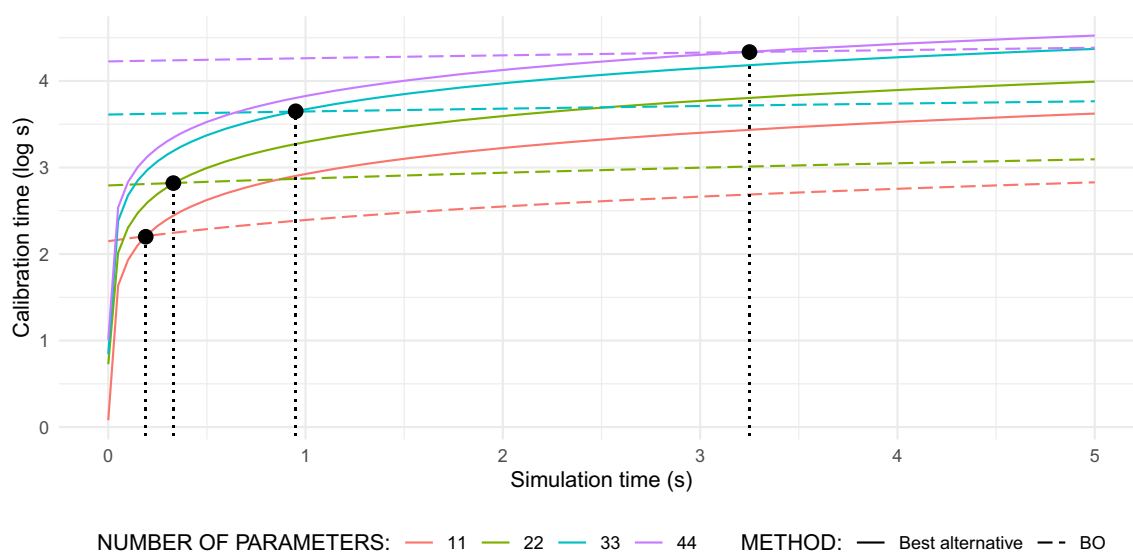
Conceptualizing the calibration problem of the simulation model for all nine age groups (99 parameters) as a single optimization problem, we obtain the results shown in Table 1. Note that the BO methods were truncated at 1000 evaluations due to their exceedingly large calibration time. Despite this, they fail to identify satisfactory solutions after several hours or days of computation.

### 4.2 Partial Calibration

Due to the poor results of a naive approach using BO, additional calibration experiments were made by running partial simulations of the model to see their behavior in smaller settings. The following results comprise a comparative performance analysis of BO and other methods for calibrating simulation models across the initial four age groups. Figure 3 illustrates the relationship between simulation and calibration time. In instances where the models are extremely fast,

**Table 1** Results of calibration of the full model by optimization method

Method	Calibration error	Number of evaluations	Calibration time
Nelder–Mead	78.87	151, 334	6 min
Simulated annealing	78.50	881, 901	35 min
PSO	87.76	243, 450	11 min
BO-SE	1849.56	1000	1 day 3 h
BO-OAK	289.45	1000	7 days 6 h

**Fig. 3** Total calibration time in log scale against model simulation time and their critical points by number of parameters

the inference overhead of BO becomes dominant, and alternative methods achieve faster calibration by simulating the model multiple times. However, as we introduce an artificial delay in the simulation to emulate larger models, the BO approach yields faster calibration times. Specifically, the BO approach outperforms the alternative methods for simulations involving a single age group (with 11 parameters) and a simulation time exceeding 0.2 s. We will refer to this time threshold as the *critical simulation time*.

However, as our problem's dimensionality increases, the BO approach's overhead also increases significantly, as shown in the y-intercepts of Fig. 3. While calibration times for the other methods also increased, the critical simulation time for the BO method experienced an exponential rise with the growing number of parameters. Specifically, it rose from 0.2 s to 0.35, 0.95, and 3.25 s for 11, 22, 33 and 44 parameters, respectively. Figure 4 projects that for all 9 age groups (99 parameters), BO-SE would be the fastest method only when each simulation demands more than 5 min of computation.

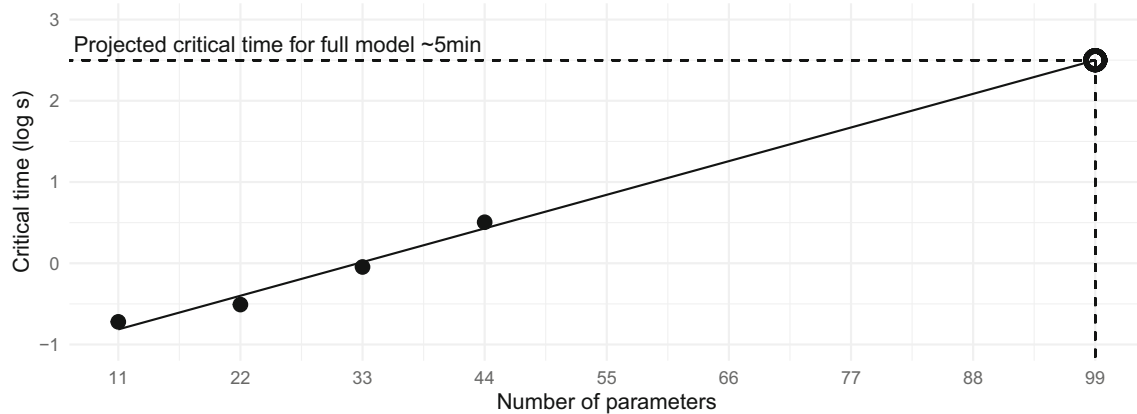
As expected with BO, there is a clear reduction in the number of evaluations needed to achieve a similar level of accuracy compared to other methods, as shown in Fig. 5. Although each iteration requires a significant amount of time due to the Bayesian inference step, this overhead becomes

less relevant as the model simulation time increases. In this scenario, the total execution time per iteration is predominantly determined by the simulation of the model while the inference overhead becomes negligible in comparison.

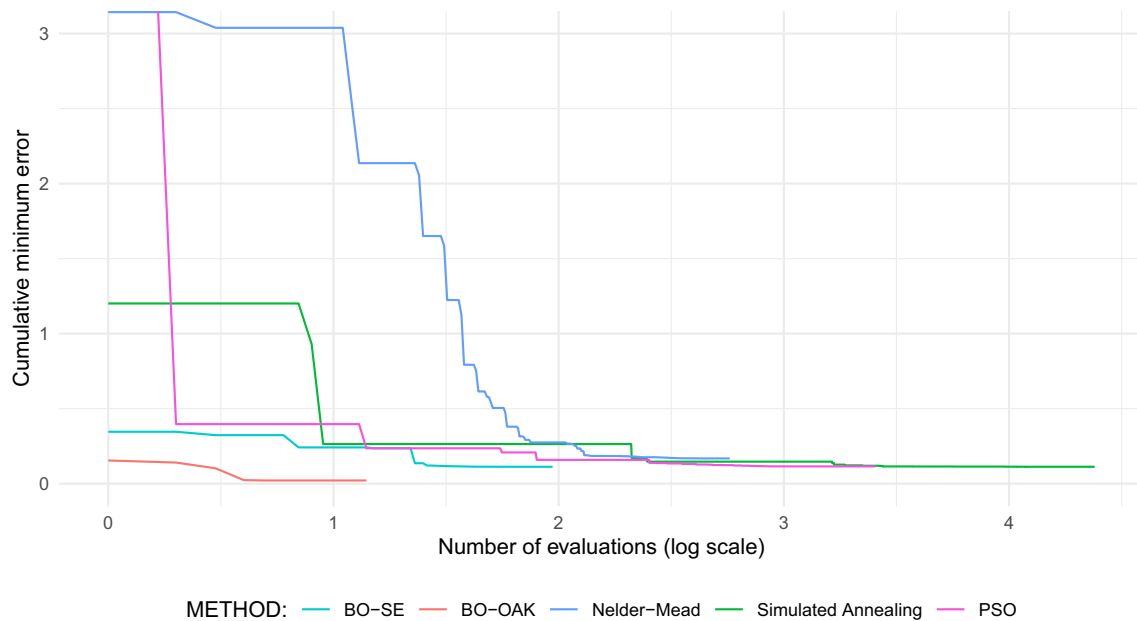
In Fig. 6, we present the error's median progression and interquartile range throughout the optimization process based on a sample of 30 random executions. During the exploration of BO-OAK results, a variable with substantial explanatory power on its own was identified, which could lead to misleading comparisons. To address this, we introduced a third univariate SE kernel that exclusively considers this variable. This kernel demonstrated lower average error and reduced spread compared to the regular SE kernel, simplifying the exploration of a smaller solution space with minimal information loss. However, it is noteworthy that OAK efficiently explores the entire 11-dimensional space, surpassing the average results of the univariate SE kernel while simultaneously reducing dispersion as the optimization progresses.

### 4.3 Stepwise Calibration

Using a stepwise methodology significantly reduces the computational cost dramatically for BO, in contrast with the findings in Sect. 4.1, as seen in Table 3. This enables calibra-



**Fig. 4** Exponential trend of critical simulation times (in log scale) as a function of the number of parameters



**Fig. 5** Lung cancer model calibration error by number of evaluations in logarithmic scale, simulating one age group

tion for all nine age groups within a feasible timeframe while maintaining comparable quality solutions. Furthermore, critical simulation times are drastically reduced, as depicted in Fig. 7. Notably, in a naive calibration with BO-SE, the projected critical simulation time for the full model was approximately 300 s (see Fig. 4). In contrast, with a stepwise calibration, the critical time for the same method is reduced to 0.24 s. To summarize Fig. 7, we can discern three different scenarios according to the time requirements of the simulation model as we can see in Table 2.

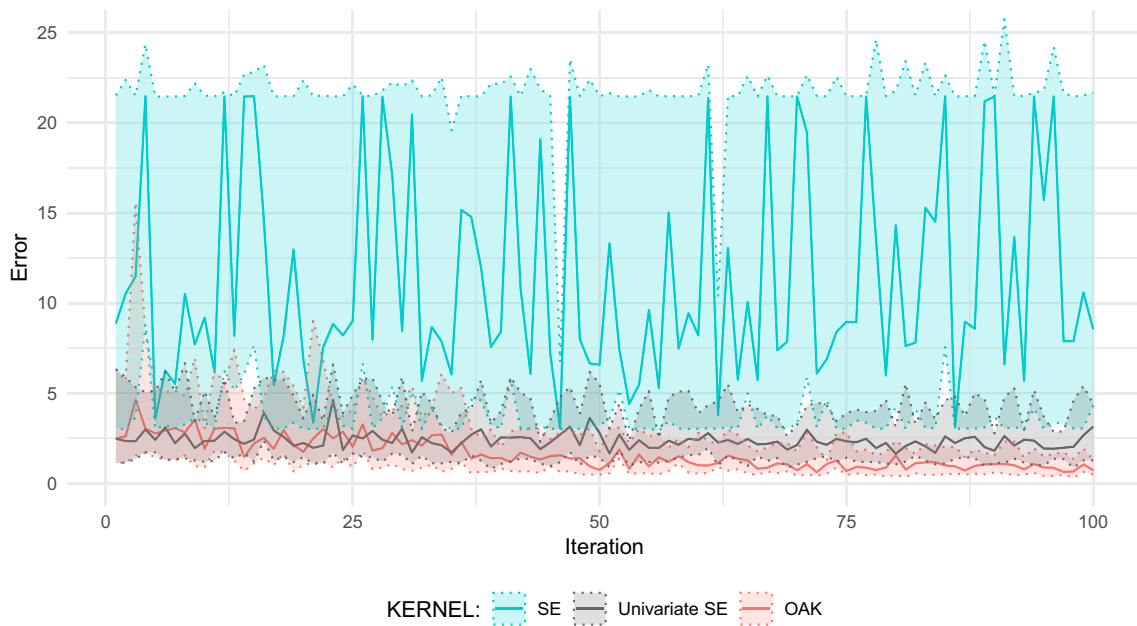
Stepwise calibration not only achieves solutions for the simulation model with comparable error to those obtained by regular calibration but also, in some cases, even lowers the error. In the case of BO-OAK, we observe a significant decrease in error for the first and last few age groups, with a slight increase in the middle ones. This indicates that the

stepwise approach manages to find a better overall solution (see Table 3 and Fig. 8) in a fewer number of simulations.

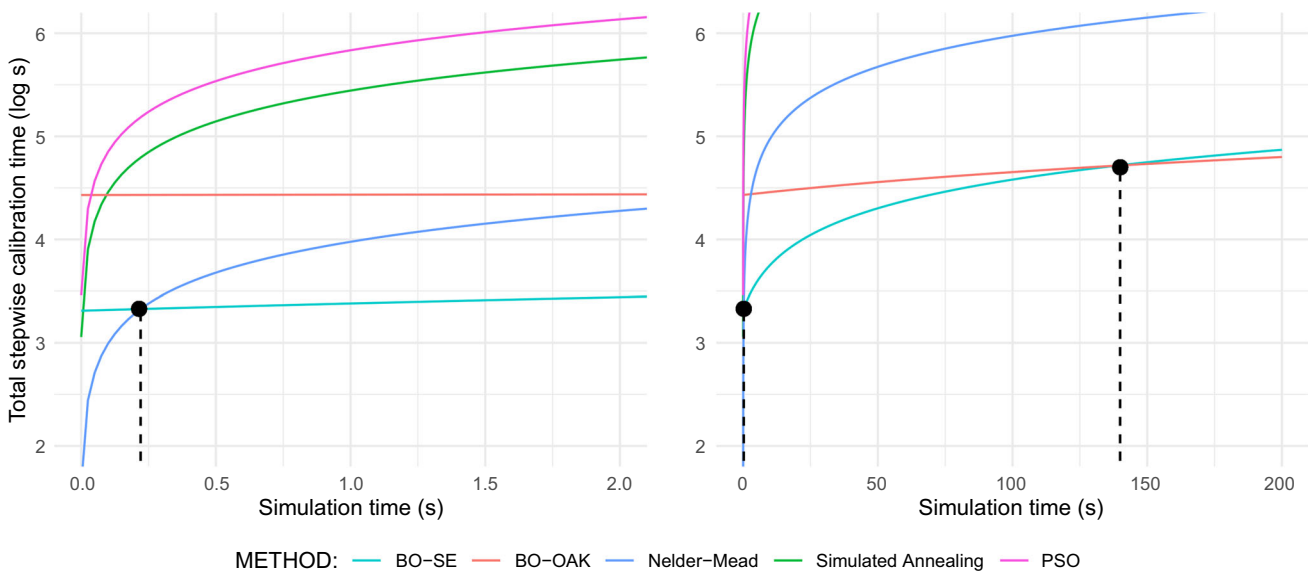
It is important to note a sharp increase in error observed in the seventh age group, a phenomenon observed across all methods. This increase is attributed to the nature of the simplified simulation utilized and how the target has been selected. Consequently, it is crucial to recognize that this effect is a limitation inherent in the simulation model and not a shortcoming of the optimization process.

## 5 Discussion

BO is currently esteemed as a leading optimization method across various domains, particularly those involving costly functions to evaluate. When the computational expense of



**Fig. 6** Median BO error and the interquartile range as the shaded area for the SE kernel (blue), the univariate SE kernel using only the most significant variable (green) and the OAK under a normality assumption for the inputs (red)



**Fig. 7** Total stepwise calibration time for each optimization method in a logarithmic scale against the execution time for one simulation, in two different  $x$ -axis scales to identify the critical times for BO-SE (left) and the BO-OAK (right)

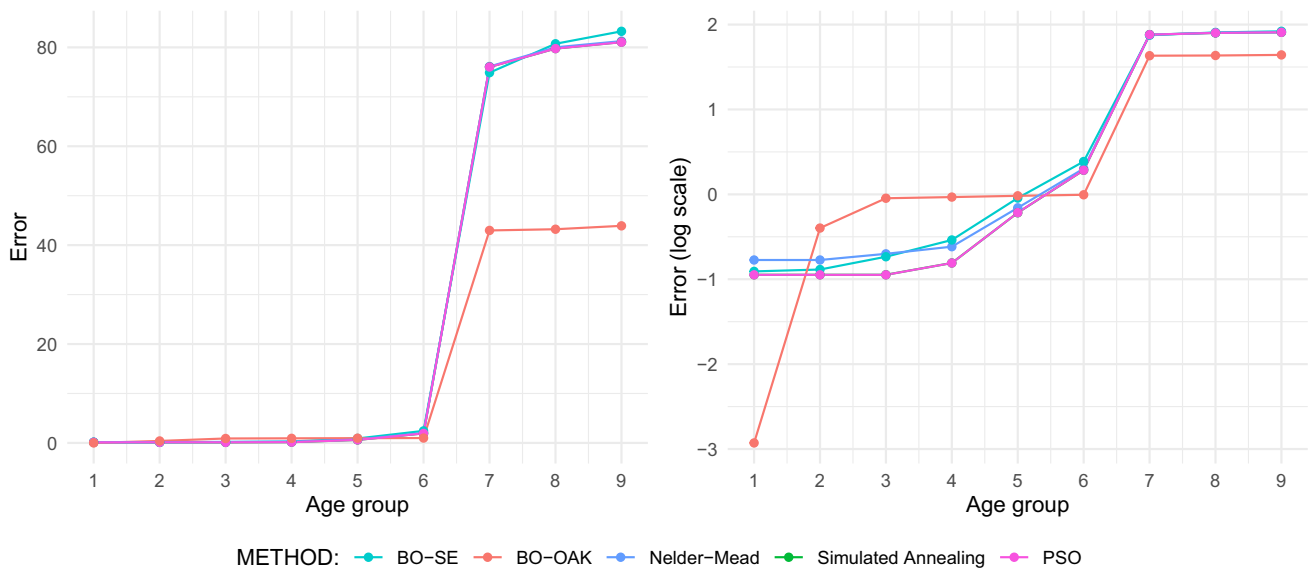
**Table 2** Fastest optimization method according to model simulation time, as seen in Fig. 7

Lung cancer model simulation time	Fastest stepwise calibration method
< 0.24 s	Nelder-Mead
0.24–140 s	BO-SE
> 140 s	BO-OAK



**Table 3** Comparison of naive versus stepwise calibration for the full simulation of all nine age groups, extended with state-of-the-art high-dimensional BO methods

Method	Calibration error		Number of evaluations		Calibration time	
	Naive	Stepwise	Naive	Stepwise	Naive	Stepwise
Nelder–Mead	78.87	81.26	151,334	9447	6 min	18 s
Simulated annealing	78.50	78.95	881,901	272,733	35 min	9 min
PSO	87.76	83.04	243,450	518,656	11 min	36 min
BO-SE	1849.56	83.24	1000	360	1 day 3 h	34 min
BO-OAK	289.45	43.90	1000	180	7 days 6 h	7 h 30 min
SAASBO	117.68	114.82	500	180	1 day 16 h	1 h 4 min
BAxUS	97.5	134.01	1000	225	1 h 9 min	18 s



**Fig. 8** Stepwise calibration error for each age group in linear scale (left) and logarithmic scale (right)

the objective function is substantial, BO often emerges as the natural choice.

The primary bottleneck in the optimization process typically revolves around optimizing the acquisition function, especially when the number of observations is relatively small and model regression is not a significant concern. Since the search space remains constant, the cost of optimizing the acquisition function does not increase significantly with additional data.

However, in model regression, Gaussian processes with SE kernels face considerable challenges, notably due to the curse of dimensionality. In high-dimensional problems, the demand for observations to explore the solution space escalates rapidly, rendering regression impractical, particularly when dealing with large matrices for posterior predictive distribution calculations. To address this issue, OAKs play a pivotal role in reducing the requisite number of observations, mitigating the impact of expanding matrices, and enhancing the efficiency of the search process. The scalability issues evident in Fig. 3, attributed to increasing dimensionality,

underscore the necessity for high-dimensional enhancements such as OAK.

Another complementary approach to circumvent these scaling challenges is through stepwise calibration. This technique enables the calibration of models that would otherwise be infeasible using conventional methods by breaking down many parameters into a sequence of calibration tasks with fewer parameters. While stepwise calibration is particularly suitable for problems with a sequential block structure in the target function, it does not guarantee global optimality despite each step making an optimal choice. Nevertheless, the technique has demonstrated promising results in simulation models used for CEA. The efficacy of stepwise calibration lies in its ability to provide solutions with similar or superior accuracy compared to naive calibration, particularly evident with BO-OAK. Moreover, the tuned OAK hyperparameters suggest a negligible weight on interactions greater than one, implying that the problem’s parameter interactions are not overly complex, making the stepwise approach a reasonable approximation. This result makes intuitive sense, since these simulation models are designed to represent reality in the

simplest way that allows us to accurately mimic the clinical, real-world results we are interested in. By Occam's razor, these models try to avoid unnecessary complication for their intended purpose. Consequently, in practice, most of the interaction between parameters are not excessively convoluted.

The findings strongly support the notion that stepwise calibration represents a substantial enhancement over naive calibration. However, it's crucial to choose the optimization method according to the model simulation time, as highlighted by critical times in Fig. 7. To further illustrate this point, three potential scenarios are outlined in the results section (Table 2).

One notable exception where stepwise calibration doesn't improve upon naive calibration is seen with SAASBO and BxUS. These advanced techniques incorporate built-in dimensionality reduction measures: shrinkage over hyperparameters in SAASBO and lower-dimensionality embeddings in BxUS. However, this dimensionality reduction offers limited advantage when applied to already low-dimensional subtasks. Despite this limitation, both SAASBO and BxUS significantly outperform simpler BO methods when applied in a naive calibration.

Interestingly, stepwise BxUS emerges as an exceptionally fast alternative. It provides a highly attractive option when a certain degree of suboptimal calibration error is acceptable, effectively balancing speed, efficiency and accuracy.

In particular, combining BO-OAK with stepwise calibration yields impressive results. The stepwise approach significantly enhances BO, effectively addressing its inherent poor dimensional scalability. This combination results in faster discovery of better solutions than other methods, especially if the simulation model is moderately time-consuming ( $\sim 2$  min). Notably, these results are achieved despite using a basic BO-OAK implementation, suggesting that further improvements in kernel numerical computation could substantially improve the calibration process, reducing the critical time.

Lu et al. [16] suggest an intriguing avenue for further exploration: extending OAK to a BO workflow by capitalizing on the inferred low-order representation. In our experiments, we demonstrate that even with a straightforward application of OAK on this simple example, a slight improvement over the SE kernel is evident. This enhancement is expected to be more significant for complex models, where additional structure can be leveraged.

It's worth noting that these results hold true despite certain assumptions of the model not being met. Specifically, hyperparameter tuning was conducted using a dataset sampled from a uniform input distribution, while the constrained kernels were computed assuming normality in the input. Even if these distributions were aligned, we would still face the

challenge of determining the input distribution for the optimization process, which would likely be neither normal nor uniform.

## 6 Future Work

An aspect not covered in this article is constraint handling. Constraints in a simulation model provide another layer of the solution space structure that can enhance the calibration process. Extensive literature exists on this topic, exploring the use of additional regression models or novel acquisition functions [24]. While some of these methods assume independent constraints, it is worth noting that simulation models in CEA often involve interdependent constraints. Nevertheless, studies indicate that the independence assumption tends to work well in practice [25]. In cases where this assumption does not work, one related research topic would be adapting these constraint-handling techniques in a stepwise methodology to simplify the dependency structure.

In the discussion, we also highlighted the advantage of harnessing the parallelization potential inherent in various facets of the optimization process. While we utilized the Particle Swarm method for optimizing the acquisition function, more sophisticated avenues for parallelization exist. These include batched optimization [26], parallel acquisition functions [27], and GPU approaches [28], among others. By leveraging parallelization techniques, we can enhance the efficiency and scalability of the optimization process, potentially leading to further improvements in performance and speed.

Finally, other future work involves replicating this work with more simulation models of different nature to gather more evidence on the performance of these Bayesian methods. Furthermore, more experiments could reveal additional patterns to further refine and generalize calibration methods for CEA.

## 7 Conclusion

In this work, we propose BO as a method to calibrate simulation models for CEA and compare them to traditional approaches. In instances where the applicability of BO is compromised by high dimensionality, we propose two strategies to mitigate this limitation: using additive kernels and adopting a stepwise calibration approach. Our findings show that the combined use of both strategies in functions with a sequential block structure, as exemplified in a lung cancer simulation model, outperforms traditional optimization methods in speed and accuracy for moderate computational time requirements.

**Author Contributions** Following the CRediT taxonomy (<https://credit.niso.org>), the authors' contributions are: David Gómez-Guillén: con-

ceptualization, methodology, software, validation, formal analysis, investigation, writing-original draft preparation, visualization. Mireia Diaz: conceptualization, writing-review and editing, visualization, supervision, project administration, funding acquisition. Josep Lluís Arcos: conceptualization, writing-review and editing, visualization, supervision, project administration. Jesus Cerquides: conceptualization, writing-review and editing, visualization, supervision, project administration

**Funding** This work has been funded by the Instituto de Salud Carlos III-ISCIII (Spanish Government) through the Project PI19/01118, as well as CIBERESP CB06/02/0073 (co-funded by European Regional Development Fund, a way to build Europe), and also with the support of the Secretariat for Universities and Research of the Department of Business and Knowledge of the Government of Catalonia. Grants to support the activities of research groups (2021SGR1029). Jesus Cerquides' work was supported by Grant PID2022-136787NB-I00 funded by MCIN/AEI/10.13039/501100011033. We thank the CERCA Programme and Generalitat de Catalunya for institutional support.

**Availability of data and materials** Not applicable.

**Code Availability** The code for the optimization methods used in this work is available in the following links: Nelder–Mead: <https://docs.scipy.org/doc/scipy/reference/optimize.minimize-neldermead.html>. Simulated annealing: [https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.dual\\_annealing.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.dual_annealing.html). PSO: <https://pyswarms.readthedocs.io/en/latest/>. BO (python): <https://secondmind-labs.github.io/trieste/1.1.2/index.html>. BO (R): <https://github.com/david-gomez-guillen/stepwise-calibration>. SAASBO: <https://botorch.org/tutorials/saasbo>. BAXUS: <https://botorch.org/tutorials/baxus>. The lung cancer simulation model used is part of an independent project and its code is not publicly available. If needed, the authors of the original article [17] can be contacted at mireia@iconcologia.net.

## Declarations

**Conflict of interest** All authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Gómez-Guillén, D., Díaz, M., Arcos, J.L., Cerquides, J.: Bayesian optimization with additive kernels for the calibration of simulation models to perform cost-effectiveness analysis. In: Artificial Intelligence Research and Development, pp. 143–152. IOS Press. <https://doi.org/10.3233/FAIA230677>
- Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L., Torrance, G.W.: Methods for the Economic Evaluation of Health Care Programmes, vol. 4, edition Oxford University Press, Oxford (2015)
- Levin, H.M.: Cost-Effectiveness Analysis: Methods and Applications, vol. 2, edition SAGE Publications, Inc, Thousand Oaks (2000)
- Gray, A.M., Clarke, P.M., Wolstenholme, J.L., Wordsworth, S.: Applied Methods of Cost-Effectiveness Analysis in Healthcare. Handbooks in Health Economic Evaluation. Oxford University Press, Oxford (2010)
- Weinstein, M.C., Torrance, G., McGuire, A.: QALYs: the basics. Value in health: the Journal of the International Society for Pharmacoeconomics and Outcomes Research. Value Health **12**(Suppl 1), 5–9 (2009). <https://doi.org/10.1111/j.1524-4733.2009.00515.x>
- Vanni, T., Karnon, J., Madan, J., White, R.G., Edmunds, W.J., Foss, A.M., Legood, R.: Calibrating models in economic evaluation. Pharmacoeconomics **29**, 35–49 (2011). <https://doi.org/10.2165/11584600-000000000-00000>
- Kim, J.J., Kuntz, K.M., Stout, N.K., Mahmud, S., Villa, L.L., Franco, E.L., Goldie, S.J.: Multiparameter calibration of a natural history model of cervical cancer. Am. J. Epidemiol. **166**(2), 137–150 (2007). <https://doi.org/10.1093/aje/kwm086>
- Krijkamp, E.M., Alarid-Escudero, F., Enns, E.A., Jalal, H.J., Hunink, M.G.M., Pechlivanoglou, P.: Microsimulation modeling for health decision sciences using R: a tutorial. Med. Decis. Mak. **38**(3), 400–422 (2018). <https://doi.org/10.1177/0272989X18754513>
- Garnett, R.: Bayesian Optimization. Cambridge University Press, Cambridge (2023). <https://doi.org/10.1017/9781108348973>
- Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning. MIT Press, Cambridge (2006)
- Duvenaud, D., Lloyd, J., Grosse, R., Tenenbaum, J., Zoubin, G.: Structure discovery in nonparametric regression through compositional kernel search. In: Proceedings of the 30th International Conference on Machine Learning, pp. 1166–1174. ISSN: 1938-7228 (2013)
- Bengio, Y.: On the challenge of learning complex functions. Prog. Brain Res. **165**, 521–534 (2007). [https://doi.org/10.1016/S0079-6123\(06\)65033-4](https://doi.org/10.1016/S0079-6123(06)65033-4)
- Durrande, N., Ginsbourger, D., Roustant, O.: Additive covariance kernels for high-dimensional Gaussian process modeling. Annales de la Faculté de Sciences de Toulouse Tome **21**, 481 (2012)
- Bino, M., Wycoff, N.: A survey on high-dimensional Gaussian process modeling with application to Bayesian optimization. ACM Trans. Evol. Learn. Optim. **2**(2), 8–1826 (2022). <https://doi.org/10.1145/3545611>
- Duvenaud, D.K., Nickisch, H., Rasmussen, C.: Additive Gaussian processes. In: Advances in Neural Information Processing Systems, vol. 24, pp. 226–234 (2011)
- Lu, X., Boukouvalas, A., Hensman, J.: Additive Gaussian processes revisited. In: Proceedings of the 39th International Conference on Machine Learning, pp. 14358–14383. ISSN: 2640-3498 (2022)
- Diaz, M., Garcia, M., Vidal, C., Santiago, A., Gnutti, G., Gómez, D., Trapero-Bertran, M., Fu, M.: Lung Cancer Prevention LUCAPREV research group: health and economic impact at a population level of both primary and secondary preventive

- lung cancer interventions: a model-based cost-effectiveness analysis. *Lung Cancer* **159**, 153–161 (2021). <https://doi.org/10.1016/j.lungcan.2021.06.027>
18. Nelder, J.A., Mead, R.: A simplex method for function minimization. *Comput. J.* **7**(4), 308–313 (1965). <https://doi.org/10.1093/comjnl/7.4.308>
  19. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. *Science* **220**(4598), 671–680 (1983). <https://doi.org/10.1126/science.220.4598.671>
  20. Bonyadi, M.R., Michalewicz, Z.: Particle swarm optimization for single objective continuous space problems: a review. *Evol. Comput.* **25**(1), 1–54 (2017). [https://doi.org/10.1162/EVCO\\_r\\_00180](https://doi.org/10.1162/EVCO_r_00180)
  21. Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. *J. Glob. Optim.* **13**(4), 455–492 (1998). <https://doi.org/10.1023/A:1008306431147>
  22. Eriksson, D., Jankowiak, M.: High-dimensional Bayesian optimization with sparse axis-aligned subspaces. In: *Uncertainty in Artificial Intelligence*, pp. 493–503. PMLR (2021)
  23. Papenmeier, L., Nardi, L., Poloczek, M.: Increasing the scope as you learn: adaptive Bayesian optimization in nested subspaces. In: *Advances in Neural Information Processing Systems, NeurIPS 2022* (2022)
  24. Gardner, J.R., Kusner, M.J., Xu, Z., Weinberger, K.Q., Cunningham, J.P.: Bayesian optimization with inequality constraints. In: *Proceedings of the 31st International Conference on Machine Learning*, vol. 32. ICML'14, pp. 937–945. JMLR.org, Beijing, China (2014)
  25. Zhang, S., Lee, R., Shafei, B., Walz, D., Misener, R.: Dependence in constrained Bayesian optimization. *Optim. Lett.* (2023). <https://doi.org/10.1007/s11590-023-02047-z>
  26. González, J., Dai, Z., Hennig, P., Lawrence, N.: Batch Bayesian optimization via local penalization. In: *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS)*. JMLR Workshop and Conference Proceedings, vol. 51, pp. 648–657 (2016)
  27. Wang, J., Clark, S.C., Liu, E., Frazier, P.I.: Parallel Bayesian global optimization of expensive functions. *Oper. Res.* **68**(6), 1850–1865 (2020). <https://doi.org/10.1287/opre.2019.1966>
  28. Wang, K., Pleiss, G., Gardner, J., Tyree, S., Weinberger, K.Q., Wilson, A.G.: Exact Gaussian processes on a million data points. In: *Advances in Neural Information Processing Systems*, vol. 32, pp. 14648–14659 (2019)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.