Efficient and adaptive Bayesian data-driven design of reliable solder joints for micro-electronic devices

Guo, Leo; Inamdar, Adwait; van Driel, Willem D.; Zhang, Guoqi

DOI
10.1016/j.apm.2025.116645

Publication date
2026

Document Version
Final published version

Published in
Applied Mathematical Modelling

Applied Mathematical Modelling 154 (2026) 116645

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Efficient and adaptive Bayesian data-driven design of reliable solder joints for micro-electronic devices

Leo Guo, Adwait Inamdar, Willem D. van Driel, Guoqi Zhang *

Department of Electronic Components, Technology and Materials, Delft University of Technology, Mekelweg 4, Delft, 2628CM, Netherlands

ARTICLE INFO

Keywords:
Data-driven design
Adaptive hyperparameters
Bayesian optimization
Solder joint reliability
Micro-electronics

ABSTRACT

Solder joint reliability related to failures due to thermomechanical loading is a critically important yet physically complex engineering problem. As a result, simulated behavior is oftentimes computationally expensive. In an increasingly data-driven world, it is popular to use efficient data-driven design schemes. Among the family of efficient optimization methods, Bayesian optimization with Gaussian process regression is a key representative. The authors argue that additional computational savings can be obtained from exploiting thorough surrogate modeling and selecting a design candidate based on multiple acquisition functions. This is feasible due to the relatively low computational cost, compared to the expensive simulation objective. This paper presents a novel heuristic framework for performing Bayesian optimization with adaptive hyperparameters across multiple optimization iterations. A comparative study shows the ability of adaptive Bayesian optimization to save on expensive objective evaluations with respect to the worst-performing regular Bayesian optimization scheme. As an engineering use case, the solder joint reliability problem is tackled by minimizing the accumulated non-linear creep strain under a cyclic thermal load. Results show that adaptive Bayesian optimization can at least match the performance of regular Bayesian optimization in terms of raw objective performance, but achieves this with half of the computational expense budget. This practical result underlines the methodological potential of the novel adaptive Bayesian data-driven methodology to achieve more efficient results and significantly cut optimization-related expenses. Lastly, to promote the reproducibility of the results, the data-driven implementations are made available on an open-source basis.

1. Introduction

Computer simulations of integrated circuit packages, particularly finite element model (FEM) simulations, have become an effective tool in improving their designs. The extensive use of data-driven design in addition to the traditional physics-based design process has been a central theme in engineering for the past years [1]. However, the simulations that underpin the design process can be computationally expensive. One way to handle this is by leveraging data to construct related sub-objectives that are much faster and simpler to evaluate and optimize. One of the most popular methods that satisfies these properties is Bayesian optimization (BO) with Gaussian process regression (GPR) [2]. BO is a proxy-optimization method that employs knowledge from a surrogate model to perform iterative design space sampling to optimize an expensive objective. Although many types of surrogate models can be selected, the surrogate is commonly chosen to be a Gaussian process (GP) regression model in engineering applications [3–5]. The

* Corresponding author.
E-mail addresses: l.l.guo@tudelft.nl (L. Guo), g.q.zhang@tudelft.nl (G. Zhang).

https://doi.org/10.1016/j.apm.2025.116645
Received 2 July 2025; Received in revised form 21 November 2025; Accepted 24 November 2025
Available online 5 December 2025

proxy-optimization component of BO hinges on the selection of an acquisition function, which maps the design space onto a belief landscape that serves to guide the adaptive sampling process.

About 70 % of the failures in electronic components occur during the packaging and assembly processes, and the predominant failure mode is associated with the solder joints [6]. Temperature, humidity, mechanical vibrations, and dust are the four key environmental factors that are responsible for component degradation, in which the temperature factor is the most dominant one [7] and is responsible for about 55 % of the failures, whereas mechanical vibrations contribute to about 20 % of the failures [8,9]. Failures related to semiconductors, connectors, and solder joints together account for over one-third of the share for power electronics [10,11]. A solder joint failure is primarily governed by the variation of temperature and mechanical loads [12,13], and thus, it is one of the key aspects of reliability engineering for integrated circuit packages and electronics-enabled systems. Chief among the indicators of solder joint failure is accumulated non-linear creep strain [14], which can be calculated by means of FEM simulations.

Accurately modeling the thermomechanical behavior of solder joints is often a computationally expensive endeavor. To compensate for this expense, response surface modeling of accumulated (plastic) strain has previously been achieved by means of long short-term memory neural network models [15]. This work served as a foundation to utilize GPR in modeling nonlinear creep and was previously adapted with BO to solve for reliability in solder joints [16]. BO with GP regression models has similarly proven to be a fruitful methodology in efficiently solving other high-cost electronic design problems [17–19], with authors aiming to implement even higher-fidelity – and thus more expensive – numerical models in future work. This is exacerbated by the fact that BO techniques with multiple objectives [20–22] and multiple levels of data fidelity [23,24] have previously been successfully applied to electronic and mechanical reliability [25].

Keeping in mind the potential increase in computational expense to procure the necessary data to perform Bayesian data-driven design, it would be prudent to critically and thoroughly discuss the hyperparameter or model parameter assumptions on which BO with GPR hinges. Practitioners of supervised machine learning, such as regression for surrogate modeling, are familiar with the importance of obtaining a model with a model parameter vector $\boldsymbol{\theta}$ that generalizes the regression model across the entire design space as accurately as possible [26,27]. Crucially, to the authors' knowledge, the handling of supervised model hyperparameters in the context of Bayesian data-driven design lacks representation in the literature: it is noted that the type of covariance kernel and acquisition function is commonly kept the same throughout all steps of BO. This is despite the lack of precedent in believing that other available acquisition function types would perform worse. Due to the high stakes that the high-cost objective evaluations in solder joint reliability modeling carry over to the optimization problem, it is a critically important research gap to address.

This work aims to show that it is possible to create statistically robust heuristics towards the selection of hyperparameters when performing BO, to the benefit of efficient usage of a given computational budget. First, a brief introduction of BO with GPR is given, after which the novel ideas of surrogate model selection and acquisition candidate selection are introduced. Regarding the micro-electronic case study, variance-based sensitivity analysis is performed to identify the relevant design parameters that should be considered for an optimization problem. Finally, the adaptive BO framework is applied to optimize a solder joint design for minimal accumulated creep strain and compared to the non-adaptive, standard variant of BO. Because the authors value the accessibility and reproducibility of the results, the code implementations and data-related resources are open-source as mentioned in the Data availability statement.

In summary, this work provides the following novel contributions:

- We introduce a heuristic to perform hyperparameter selection, specifically in the context of Bayesian data-driven design.
- A family of novel BO methods, collectively called adaptive BO, is described, implemented, and compared to standard BO.
- As a proof of concept, adaptive BO is applied to optimize a selection of synthetic objective functions as well as the minimization of accumulated creep strain on the solder joint of an electronic package.

2. Bayesian data-driven methods

The mathematical description of a single-objective design problem typically comprises a design space $\mathcal{X} \subset \mathbb{R}^D$ consisting of designs or design vectors $\mathbf{x} = (x_1, x_2, \dots, x_D)^\top$ of length D, and the relevant objective function $f : \mathcal{X} \to \mathbb{R}$. It is often the case that the permissible domain of each design parameter x_j is a closed interval. In this scenario, it is possible to scale $\mathcal{X}$ into the unit cube $[0,1]^D$, which can be interpreted as a generic design space.

2.1. Bayesian optimization with Gaussian process regression

GP modeling of an objective function $f : [0,1]^D \to \mathbb{R}$ is a Bayesian method, in that it assumes all dependent variables to be stochastic [28]. To be precise, for any $\mathbf{x} \in [0,1]^D$, the value $f(\mathbf{x})$ is modeled as a stochastic variable. For any $\mathbf{u}, \mathbf{v} \in [0,1]^D$, the covariance between $f(\mathbf{u})$ and $f(\mathbf{v})$ is stipulated by a covariance function or kernel κ, such that $\kappa(\mathbf{u}, \mathbf{v}) := \mathrm{Cov}(f(\mathbf{u}), f(\mathbf{v}))$. Various choices for κ exist, each of them relying on a vector of learnable model parameters $\boldsymbol{\theta}$. The most popular ones in literature are:

$$\kappa_{\mathrm{RBF},\boldsymbol{\theta}_{\mathrm{RBF}}}(\mathbf{u},\mathbf{v}) := c \cdot \exp\left(-\frac{r^2(\mathbf{u},\mathbf{v})}{2\lambda^2}\right) + s^2\delta(\mathbf{u}-\mathbf{v}),$$

$$\kappa_{\mathrm{Mat},\boldsymbol{\theta}_{\mathrm{Mat}}}(\mathbf{u},\mathbf{v}) := c \cdot \left(1 + \frac{\sqrt{3}r(\mathbf{u},\mathbf{v})}{\lambda}\right)\exp\left(-\frac{\sqrt{3}r(\mathbf{u},\mathbf{v})}{\lambda}\right) + s^2\delta(\mathbf{u}-\mathbf{v}),$$

$$\kappa_{\mathrm{RQ},\boldsymbol{\theta}_{\mathrm{RQ}}}(\mathbf{u},\mathbf{v}) := c \cdot \left(1 + \frac{r^2(\mathbf{u},\mathbf{v})}{2\alpha\lambda^2}\right)^{-\alpha} + s^2\delta(\mathbf{u}-\mathbf{v}),$$

where $r(\mathbf{u},\mathbf{v}) := \|\mathbf{u}-\mathbf{v}\|$ and δ is the Dirac-δ function. Furthermore, the scalars c, λ, s^2, α represent covariance kernel parameters. These are summarized into a model parameter vector per kernel type, generically denoted by $\boldsymbol{\theta}$. For example, the RBF kernel has $\boldsymbol{\theta}_{\mathrm{RBF}} := (c, \lambda, s^2)^\top$ as its model parameter vector.

Furthermore, assume that $\mathbf{X} := (\mathbf{x}_1^\top, \dots, \mathbf{x}_N^\top)^\top$ is a matrix of N design parameter vectors, then $\mathbf{y} := (f(\mathbf{x}_1), \dots, f(\mathbf{x}_N))^\top$ is the realization of a multivariate random variable. If one now assumes that $\boldsymbol{\theta}$, as a dependent parameter, is a realization of a random variable $\boldsymbol{\Theta}$, and

$$\mathbf{Y}|(\boldsymbol{\Theta}=\boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{\boldsymbol{\theta}})$$

with covariance matrix $\mathbf{K}_{\boldsymbol{\theta}} = \mathbf{K}_{\boldsymbol{\theta}}(\mathbf{X}) := (\kappa_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1,\dots,N}$, then

$$f(\mathbf{x})|(\mathbf{Y}=\mathbf{y}, \boldsymbol{\Theta}=\boldsymbol{\theta}) \sim \mathcal{N}(\mu_{\boldsymbol{\theta}}(\mathbf{x}), \sigma^2_{\boldsymbol{\theta}}(\mathbf{x})),$$

where

$$\mu_{\boldsymbol{\theta}}(\mathbf{x}) := \kappa_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X})^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1} \mathbf{y},$$
$$\sigma^2_{\boldsymbol{\theta}}(\mathbf{x}) := \kappa_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}) - \kappa_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X})^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1} \kappa_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X}).$$

Selecting or finding a fitting value for $\boldsymbol{\theta}$ is called GP regression (GPR). One common method of doing so is by numerically solving for the maximum (log) likelihood estimate (MLE):

$$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} := \operatorname*{argmin}_{\boldsymbol{\theta}\in\mathcal{T}_\kappa} \ln(\det(\mathbf{K}_{\boldsymbol{\theta}}(\mathbf{X}))) + \mathbf{y}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1}(\mathbf{X})\mathbf{y}. \tag{1}$$

Here, $\mathcal{T}_\kappa$ stands for the space of all permissible model parameter vectors $\boldsymbol{\theta}$. Throughout this manuscript, it is assumed that any time the optimization problem in Eq. (1) arises, it is solved numerically with an L-BFGS optimizer [29] and 10 random restarts.

The normal distribution that results from inserting $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$, i.e. $\mathcal{N}(\mu_{\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}}(\mathbf{x}), \sigma^2_{\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}}(\mathbf{x}))$, is called the regressive-predictive distribution (RPD). As a normal distribution, an RPD is fully described by its probability density function, which will be denoted by $\hat{\phi}$. Accordingly, the mean and variance of this distribution are denoted as $\hat{\mu} := \mu_{\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}}$ and $\hat{\sigma}^2 := \sigma^2_{\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}}$.

In the assumption that f is to be numerically minimized, an RPD $\hat{\phi}$ carries with it valuable knowledge to suggest new design parameter vectors to sample f at. A common way to extract this knowledge is to build an acquisition function $\alpha : [0,1]^D \to \mathbb{R}$ such that $\alpha(\mathbf{x}; \hat{\phi})$ quantifies a level of belief that $f(\mathbf{x})$ is less than any component of $\mathbf{y}$.

Several acquisition functions have been used widely in BO literature [30]. One such example of α is the expected improvement (EI) acquisition [31]. Other popular examples that have been utilized in data-driven literature include probability of improvement (PI) [32] and lower / upper confidence bound (UCB) [33] with hyperparameter β. See Eq. (2) for the description of these acquisition functions.

$$\begin{aligned}
\alpha_{\mathrm{EI}}(\mathbf{x}; \hat{\phi}) &:= \hat{\sigma}(\mathbf{x})(z(\mathbf{x})\hat{\Phi}(z(\mathbf{x})) + \hat{\phi}(z(\mathbf{x}))); \\
\alpha_{\mathrm{PI}}(\mathbf{x}; \hat{\phi}) &:= \hat{\Phi}(z(\mathbf{x})); \\
\alpha_{\mathrm{UCB}}(\mathbf{x}; \hat{\phi}, \beta) &:= \hat{\mu}(\mathbf{x}) + \beta\hat{\sigma}(\mathbf{x}).
\end{aligned} \tag{2}$$

By numerically optimizing (maximizing) α across the design parameter domain, a promising design parameter vector $\mathbf{x}^*$ can be suggested to evaluate $f(\mathbf{x}^*)$. It should be noted that the formulations in Eq. (2) are all analytical and differentiable, meaning that gradient-based optimizers such as Adam [34] and L-BFGS may be employed. In practice, as reported in this article, the latter is utilized with 10 random restarts. In the case of EI and PI, these gradient-based schemes may not always converge quickly, so the enhanced logarithmic EI (LogEI) [35] and logarithmic PI are often used as practical alternatives.

By appending $\mathbf{X}$ with $\mathbf{x}^*$ and $\mathbf{y}$ with $f(\mathbf{x}^*)$, the process of obtaining a renewed surrogate model can start anew. This process as a whole is called BO with GPR, and is the most popular variant of BO in literature [36]. See Algorithm 1 for an overview of this optimization scheme.

Algorithm 1 Bayesian optimization with Gaussian process regression.

Require: Design of training experiments $\mathbf{D}^{(0)} = (\mathbf{X}^{(0)}, \mathbf{y}^{(0)})$, covariance function κ, acquisition function α, number of iterations I

1: **for** $i = 1, \dots, I$ **do**
2: $\quad\hat{\boldsymbol{\theta}}^{(i)}_{\mathrm{MLE}} \leftarrow \operatorname{argmin}_{\boldsymbol{\theta}\in\mathcal{T}_\kappa} \ln(\det(\mathbf{K}_{\boldsymbol{\theta}}(\mathbf{X}^{(i-1)}))) + \mathbf{y}^{(i-1)\top}\mathbf{K}_{\boldsymbol{\theta}}^{-1}(\mathbf{X}^{(i-1)})\mathbf{y}^{(i-1)}$ ▷ Eq. (1)
3: $\quad\hat{\phi}^{(i)} \leftarrow \hat{\boldsymbol{\theta}}^{(i)}_{\mathrm{MLE}}$
4: $\quad\mathbf{x}^{(i)} \leftarrow \operatorname{argmax}_{\mathbf{x}\in[0,1]^D} \alpha(\mathbf{x}; \hat{\phi}^{(i)})$
5: $\quad y^{(i)} \leftarrow f(\mathbf{x}^{(i)})$
6: $\quad\mathbf{D}^{(i)} \leftarrow (\mathbf{D}^{(i-1)}, (\mathbf{x}^{(i)\top}, y^{(i)}))^\top$
7: **end for**
8: $(\mathbf{x}_{\mathrm{rec}}, y_{\mathrm{rec}}) \leftarrow \mathrm{Rec}(\mathbf{D}^{(I)})$ ▷ Recommends the best-found optimizer and objective
9: **return** $(\mathbf{x}_{\mathrm{rec}}, y_{\mathrm{rec}})$

Steps 1–7 of Algorithm 1 are sometimes referred to as the "outer" optimization loop, to distinguish it from the "inner" optimization loops at step 2 and step 4 performed at every outer loop iteration. The assumption underpinning the motivation to use BO at all is

the fact that step 5, the evaluation of f, is very expensive, e.g., a complete FEM simulation – possibly orders of magnitude costlier than the inner optimization loops. This discrepancy in (computational) cost can be further exploited by expanding step 2 and step 4 appropriately.

2.2. Surrogate model initialization for Bayesian optimization

New expensive data is sampled during the BO process (step 5 of Algorithm 1) during every outer loop iteration. This computationally critical step places substantial importance on selecting the appropriate hyperparameters for BO. In the context of GPR, this latter point equates to the possibility of using any covariance function from a size K tool set $\{\kappa_1, \kappa_2, \dots, \kappa_K\}$. The question of finding the optimal κ has previously been posed in the framework of Bayesian statistics as a Type-II likelihood maximization problem [28]. An approach specifically geared towards discovering structure in time series exists. This is undertaken by means of exploring a search space comprised of algebraic compositions from a set of base kernels [37]. Despite the aforementioned, there exists no practical implementation of GPR model evaluation and comparison when $D > 1$. An informed search methodology is constructed to select a covariance $\hat{\kappa}$ and a corresponding restricted model parameter search space $\hat{\mathcal{T}} \subset \mathcal{T}_{\hat{\kappa}}$ given a set of GPR models arising from optimizing the likelihood in a restricted manner.

Assume that several distinct GPR models are constructed based on $\mathbf{D}_{\text{train}}$ with a portion of N_{train} out of the N design rows that populate $\mathbf{D}$. By then defining a size N_{test} design of test experiments withheld from training $\mathbf{D}_{\text{test}} := (\mathbf{X}_{\text{test}}, \mathbf{y}_{\text{test}})$ where $\mathbf{X}_{\text{test}} := (\mathbf{x}_{\text{test},1}^\top, \dots, \mathbf{x}_{\text{test},N_{\text{test}}}^\top)^\top$ and $\mathbf{y}_{\text{test}} := (y_{\text{test},1}, \dots, y_{\text{test},N_{\text{test}}})^\top$ as a realization of $f(\mathbf{X}_{\text{test}})$. A train-test split commonly used in practice is 20:80, i.e., $N_{\text{test}}/N = 1/5$ and $N_{\text{train}}/N = 4/5$. This split value is used throughout this manuscript.

Moving forward, a fundamental assumption regarding f needs to be made to decide on the quality of an RPD density $\hat{\phi}$. If f is assumed noiseless, i.e. $f(\mathbf{X}_{\text{test}}) = \mathbf{y}_{\text{test}}$ exactly, then any deviation between the RPD mean $\hat{\mu}(\mathbf{x})$ and the objective evaluation $y = f(\mathbf{x})$ for $\mathbf{x} \in \mathbf{X}_{\text{test}}$ can be interpreted as purely resulting from epistemic uncertainty. In this case, the relative mean squared error (RelMSE) is able to sketch a reasonable picture with regard to the prediction quality of the GP surrogate model. Given a non-constant control vector $\mathbf{y} := (y_1, \dots, y_M)$ and a prediction vector $\hat{\mathbf{y}} := (\hat{y}_1, \dots, \hat{y}_M)$, it is defined by

$$\text{RelMSE}(\mathbf{y}, \hat{\mathbf{y}}) := \frac{\text{MSE}(\mathbf{y}, \hat{\mathbf{y}})}{\text{Var}(\mathbf{y})} = \frac{\sum_{j=1}^M (y_j - \hat{y}_j)^2}{\sum_{i=1}^M (y_j - \bar{y})^2} \quad \text{with} \quad \bar{y} := \frac{1}{M} \sum_{j=1}^M y_j. \tag{3}$$

In other words, for a noiseless objective function f, the value $\text{RelMSE}(\mathbf{y}_{\text{test}}, \hat{\mu}(\mathbf{X}_{\text{test}}))$ is a precise indicator of the quality of the GPR's RPD. The RelMSE is sometimes also known as the fraction of variance unexplained (FVU), equal to $1 - R^2$ where R^2 is the coefficient of determination. It is a popular choice of score to measure the (lack of) goodness of fit, especially when comparing regression models applied to different datasets [38].

However, when f is considered noisy, evaluating the quality of $\hat{\phi}$ becomes more complicated. This is because $\hat{\mu}$ no longer carries an exact interpolation role through train and test data. Instead, $\hat{\mu}(\mathbf{x})$ represents the prediction of the mean of $f(\mathbf{x}) = Y$, a normally distributed random variable. One possible solution is to first let $\mathbf{Y}_{\text{test}} := (\mathbf{y}_{\text{test},1}^\top, \dots, \mathbf{y}_{\text{test},N_{\text{test}}}^\top)^\top$ with $\mathbf{y}_{\text{test},n} := (y_{\text{test},n,1}, \dots, y_{\text{test},n,R})^\top$, for each $n \in \{1, \dots, N_{\text{test}}\}$ and $y_{\text{test},n,r}$ being a realization of $f(\mathbf{x}_{\text{test},n})$ for any $r \in \{1, \dots, R\}$. Then, $\bar{\mathbf{y}}_{\text{test}} := (\bar{\mathbf{y}}_{\text{test},1}, \dots, \bar{\mathbf{y}}_{\text{test},N_{\text{test}}})^\top$ is an estimator the true mean of $f(\mathbf{X}_{\text{test}})$, which indicates $\text{RelMSE}(\bar{\mathbf{y}}_{\text{test}}, \hat{\mu}(\mathbf{X}_{\text{test}}))$ as being a possible error measure in the objective space.

There are major drawbacks to this approach: the RelMSE is being measured between two approximations, while R cannot be large because of the expensive cost of evaluating f. In scenarios like these, it will be useful to follow state-of-the-art practice and place RelMSE scoring alongside a probabilistic scoring to judge the quality of $\hat{\phi}$. To this end, define the statistical test log-likelihood (TLL) error score [39] as:

$$\text{TLL}(\mathbf{D}_{\text{test}}, \hat{\phi}) := -\frac{\ln(2\pi)}{2} - \frac{1}{N_{\text{test}}} \sum_{j=1}^{N_{\text{test}}} \frac{1}{2} \left[\ln(\hat{\sigma}^2(\mathbf{x}_j)) + \left(\frac{y_j - \hat{\mu}(\mathbf{x}_j)}{\hat{\sigma}(\mathbf{x}_j)} \right)^2 \right], \tag{4}$$

i.e., a sample mean of logarithmic RPD density values evaluated at test outputs. The closer y_j is located to $\hat{\mu}(\mathbf{x}_j)$, the higher the value of TLL, which implies higher predictive quality on a probabilistic basis. The TLL also has a global maximum in terms of $\hat{\sigma}^2$, which means that the TLL punishes both overconfident and unconfident predictions. Now, consider the RPD densities $\hat{\phi}$ and $\hat{\phi}'$ arising from two different GPR models. Given the interpretation of the TLL score, one would be inclined to prefer $\hat{\phi}$ over $\hat{\phi}'$ if $\text{TLL}(\mathbf{D}_{\text{test}}, \hat{\phi}) > \text{TLL}(\mathbf{D}_{\text{test}}, \hat{\phi}')$.

While some authors conclude from their findings based on the TLL alongside predictive mean squared error measurements [37], this is generally speaking not straightforward. Indeed, there exist practical scenarios in which the TLL as a probabilistic quality measure does not correlate with RelMSE as a physical quality measure [40]. In these circumstances, priority should be placed on predictive RelMSE scores, ahead of TLL scores. This is because building a surrogate model for practical engineering applications requires a correspondingly practical measure of error in terms of the (relative) physical units of the objective. This measurement is readily provided by RelMSE, while TLL is a purely statistical score. Concretely, the following is proposed: assume for two RPD densities $\hat{\phi}$ and $\hat{\phi}'$ that

$$\text{TLL}(\mathbf{D}_{\text{test}}, \hat{\phi}') > \text{TLL}(\mathbf{D}_{\text{test}}, \hat{\phi}),$$

but simultaneously

$$\text{RelMSE}(\mathbf{y}_{\text{test}}, \hat{\mu}'(\mathbf{X}_{\text{test}})) > \text{RelMSE}(\mathbf{y}_{\text{test}}, \hat{\mu}(\mathbf{X}_{\text{test}})).$$

In this case, preference is assumed for $\hat{\phi}$ over $\hat{\phi}'$ if $R > \mathrm{RelMSE}(\mathbf{y}_{\mathrm{test}}, \hat{\mu}'(\mathbf{X}_{\mathrm{test}}))$ for some threshold $R > 0$.

Taking into consideration the high expense of the training data as well as the multimodal nature of many likelihood landscapes, it is prudent to critically investigate solving the likelihood optimization problem for $\hat{\theta}_{\mathrm{MLE}}$ in Eq. (1). For a given covariance function κ from a set of K covariance functions $K_{\mathrm{set}} := \{\kappa_1, \dots, \kappa_K\}$, it might be beneficial for the numerical optimization process to reduce the T_κ-dimensional search space

$$\mathcal{T}_\kappa = P_1 \times P_2 \times \cdots \times P_{T_\kappa}.$$

For example, one could define

$$\mathcal{T}'_\kappa := \{\bar{\theta}_1\} \times P_2 \times \cdots \times P_{T_\kappa}$$

for some value $\bar{\theta}_1 \in P_1$ and subsequently (numerically) solving for

$$\hat{\theta}' := \operatorname*{argmax}_{\theta \in \mathcal{T}'_\kappa} \ell(\theta; \mathbf{D}) = \operatorname*{argmin}_{\theta \in \mathcal{T}'_\kappa} \ln(\det(\mathbf{K}_\theta(\mathbf{X}))) + \mathbf{y}^\top \mathbf{K}_\theta^{-1}(\mathbf{X})\mathbf{y}. \tag{5}$$

In this context, $\mathcal{T}'_\kappa$ is called a restricted likelihood domain (RLD).

Given the nonlinear nature of ℓ, it is possible that the numerical approximation of $\hat{\theta}'$ achieves a higher likelihood than that of $\hat{\theta}_{\mathrm{MLE}}$. If this is the case, there is a quantifiable reason to believe that the training data structure allows for the restriction of the optimization of ℓ to $\mathcal{T}'_\kappa$, potentially reducing the search space dimensionality for future optimization attempts.

Of course, there are many other ways to restrict $\mathcal{T}_\kappa$ apart from $\mathcal{T}'_\kappa$. To describe the set of RLDs systematically, the following is proposed:

- **Limited amount of RLDs**. In practice, it is sufficient to consider a small set of popular covariance kernel types: $K_{\mathrm{set}} \doteq \{\kappa_{\mathrm{RBF}}, \kappa_{\mathrm{Mat}}, \kappa_{\mathrm{RQ}}\}$, i.e. $K = 3$. Recall that for these kernels, the following facts hold:

$$\begin{aligned} \theta_{\mathrm{RBF}} &= (c, \lambda, s^2), & T_{\mathrm{RBF}} &= 3; \\ \theta_{\mathrm{Mat}} &= (c, \lambda, s^2), & T_{\mathrm{Mat}} &= 3; \\ \theta_{\mathrm{RQ}} &= (c, \alpha, \lambda, s^2), & T_{\mathrm{RQ}} &= 4. \end{aligned}$$

  Given the relatively small dimensionalities of each model parameter space, and without being overly restrictive, it is therefore sufficient to consider only values of d such that $d \leq 2$. Finally, to reduce redundancy, it will be sufficient to consider three nominal values, $\bar{\Theta}_t \doteq \{\bar{\theta}_{t,\mathrm{low}}, \bar{\theta}_{t,\mathrm{mid}}, \bar{\theta}_{t,\mathrm{high}}\}$, i.e. $V = 3$ for all t. This readily reduces the size of the search space to $\sum_{k=1}^{3} 3 \cdot T_{\kappa_k} + 9 \cdot \binom{T_{\kappa_k}}{2}$.
- **Structured search**. A grid search policy with breadth-first focus is proposed to define a sequence of RLDs $\mathcal{T}_1, \mathcal{T}_2, \dots$ to solve the restricted likelihood optimization problem.
  - If a previous RLD search has been performed at an earlier BO iteration, with result $\mathcal{T}_\kappa(\bar{\theta}_F)$, then use the mixed RelMSE / TLL performance measure, to assess the quality of RPD $\hat{\phi}$ corresponding to the numerically optimized model parameter vector

$$\hat{\theta} = \operatorname*{argmin}_{\theta \in \mathcal{T}_\kappa(\bar{\theta}_F)} \ln(\det(\mathbf{K}_\theta(\mathbf{X}))) + \mathbf{y}^\top \mathbf{K}_\theta^{-1}(\mathbf{X})\mathbf{y}.$$

    If the RPD quality is sufficient, then the search is terminated. In all other cases, continue the search with the next step.
  - The RLD search starts with regular, state-of-the-art *unrestricted* likelihood optimization as defined in Eq. (1). This means:

$$\mathcal{T}_1 = \mathcal{T}_{\mathrm{RBF}}, \quad \mathcal{T}_2 = \mathcal{T}_{\mathrm{Mat}}, \quad \mathcal{T}_3 = \mathcal{T}_{\mathrm{RQ}}.$$

  - For any subsequent RLD, the search algorithm will similarly cycle through the set of covariance kernels for each d in increasing value. For each d, all possible index sets F are considered, and correspondingly, all possible $\bar{\theta}_F \in \bar{\Theta}_F$. As an explicit example with the setting described previously,

$$\begin{aligned} &\mathcal{T}_4 = \mathcal{T}_{\mathrm{RBF}}(\bar{c}_{\mathrm{low}}), \quad \mathcal{T}_5 = \mathcal{T}_{\mathrm{RBF}}(\bar{c}_{\mathrm{mid}}), \quad \mathcal{T}_6 = \mathcal{T}_{\mathrm{RBF}}(\bar{c}_{\mathrm{high}}), \\ &\mathcal{T}_7 = \mathcal{T}_{\mathrm{RBF}}(\bar{\lambda}_{\mathrm{low}}), \quad \mathcal{T}_8 = \mathcal{T}_{\mathrm{RBF}}(\bar{\lambda}_{\mathrm{mid}}), \quad \mathcal{T}_9 = \mathcal{T}_{\mathrm{RBF}}(\bar{\lambda}_{\mathrm{high}}), \\ &\quad \dots, \\ &\mathcal{T}_{13} = \mathcal{T}_{\mathrm{Mat}}(\bar{c}_{\mathrm{low}}), \quad \mathcal{T}_{14} = \mathcal{T}_{\mathrm{Mat}}(\bar{c}_{\mathrm{mid}}), \quad \mathcal{T}_{14} = \mathcal{T}_{\mathrm{Mat}}(\bar{c}_{\mathrm{mid}}) \\ &\quad \dots, \\ &\mathcal{T}_{31} = \mathcal{T}_{\mathrm{RBF}}(\bar{c}_{\mathrm{low}}, \bar{\lambda}_{\mathrm{low}}), \quad \mathcal{T}_{32} = \mathcal{T}_{\mathrm{RBF}}(\bar{c}_{\mathrm{mid}}, \bar{\lambda}_{\mathrm{low}}), \quad \dots \end{aligned}$$

    Apart from this structured grid search, other (hyper)parameter search methods exist. These include (naïve) grid search, quasi-random search, and tree-structured Parzen estimators [41]. Off-the-shelf packages facilitate a practical implementation of these methods for supervised machine learning models in general, such as `SMAC` [42], `Hyperopt` [43], and `Optuna` [44].

This structured search procedure, as described above, is called Gaussian process initialization (GPi). To illustrate GPi, a schematic overview is provided in Fig. 1.

From Fig. 1, it can be seen that the full GPi procedure performs the search with an ascending number of hyperparameters selected. It begins by probing covariance kernels with an unrestricted likelihood domain (1 total selection). Provided that GPi does not find a hyperparameter configuration that leads to a regression model with a sufficient RelMSE and/or TLL score, the search ends with the evaluation of covariance kernels accompanied by an RLD with two fixed dimensions (3 total selections). A precise algorithmic description of GPi can be found in Algorithm 2.

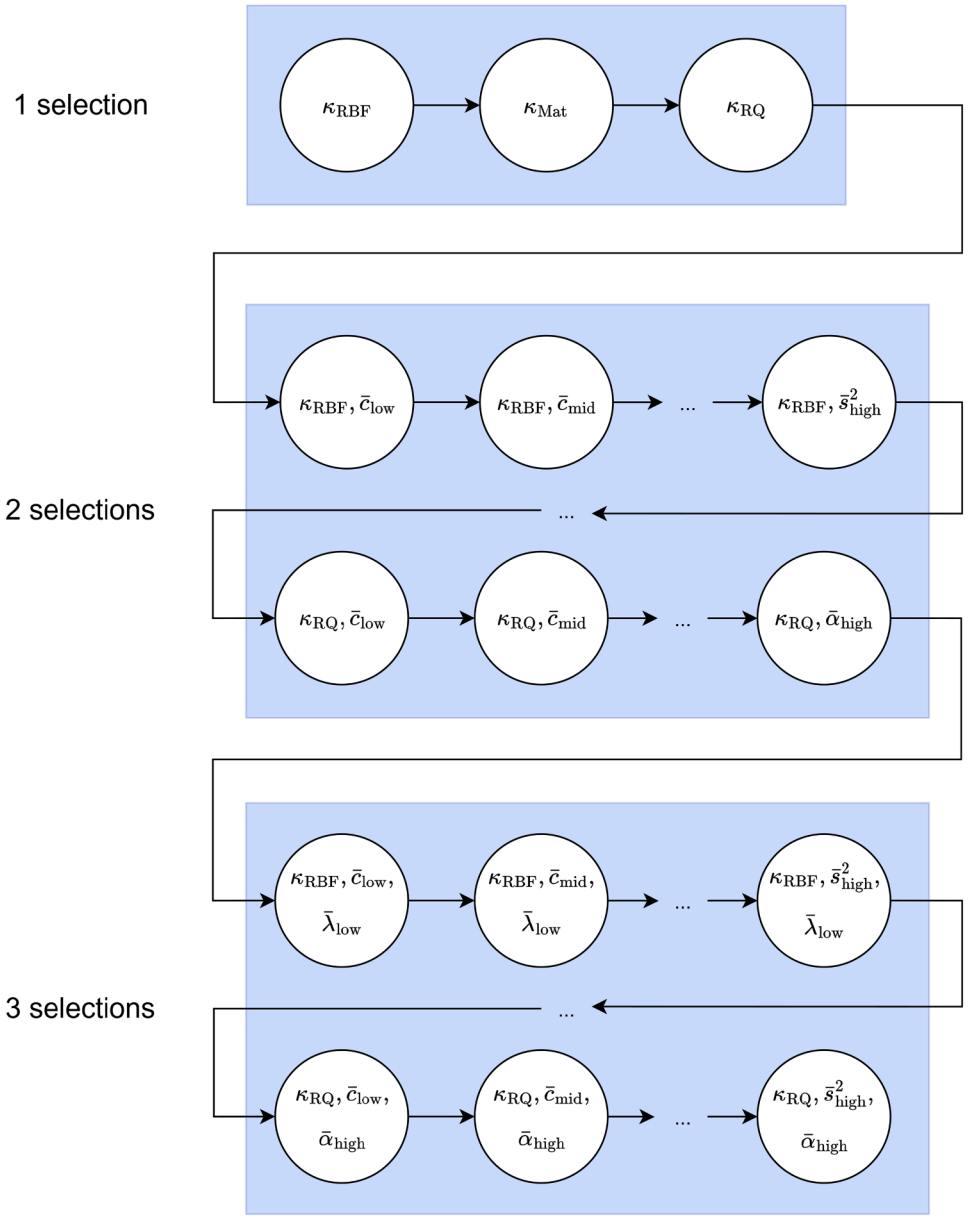


Fig. 1. A visualization of the full breadth-first grid search procedure. The arrow represents the search direction and sequentially traverses the defined space of hyperparameter combinations, characterized by each white disk. Each blue box indicates the subset of configurations with a particular number of selected hyperparameters.

2.3. Adaptive candidate design selection for BO

A reasonable next step in this discussion is exploring the possibilities to automate the selection of α. To do this, it is reasonable to exploit the assumption that the optimization of an acquisition function α is computationally much cheaper than optimizing f. Thus, given a toolbox of acquisition functions $\alpha_1, \alpha_2, \ldots$, it is possible to optimize each of them efficiently and yield design candidates $\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \ldots$, with outer iteration step i. Then, to find the best possible candidate design in terms of optimizing f, each of these designs should ideally be evaluated at every iteration. However, since f is an expensive objective, this is not always feasible.

As a direct comparison between objective evaluations $f(\mathbf{x}_1^{(i)}), f(\mathbf{x}_2^{(i)}), \ldots$ is impractical due to the bottleneck of computational expense, one might turn to comparing acquisition functions instead. However, this method has a number of drawbacks.

Firstly, it should be emphasized that each acquisition landscape represents a belief model given the available data, and might operate on a different scale of magnitude compared to another acquisition. For example, the range of α_{PI} is the unit interval $(0, 1)$, while the range of α_{UCB} is $\mathbb{R}$. This fact rules out any direct comparisons between acquisition function values.

Algorithm 2 Gaussian process model selection/initialization (GPi).

Require: Design of train experiments $\mathbf{D}$, design of test experiments $\mathbf{D}_{\text{test}}$, set of covariance kernels $\{\kappa_{\text{RBF}}, \kappa_{\text{Mat}}, \kappa_{\text{RQ}}\}$, collection of sets of parameter indices to fix $\{\mathcal{F}_0 = \varnothing, \mathcal{F}_1, \mathcal{F}_2\}$, nominal parameter fixture values $\{\{\bar{\theta}_{t,\text{low}}, \bar{\theta}_{t,\text{mid}}, \bar{\theta}_{t,\text{high}}\} : t \in \{1, \dots, D\}\}$, trial threshold Q, RelMSE threshold R

$\text{RelMSE}_* \leftarrow +\infty$
$\text{TLL}_* \leftarrow -\infty$
$\hat{\kappa},\ \hat{\mathcal{T}},\ \hat{\boldsymbol{\theta}} \leftarrow \varnothing$
$q, k \leftarrow 0$
for $d \in \{0, 1, 2\}$ **do**
  for $\kappa \in K_{\text{set}}$ **do**
    for $F \in \mathcal{F}_d$ **do**
      $\bar{\Theta}_F \leftarrow \prod_{t \in F} \{\bar{\theta}_{t,\text{low}}, \bar{\theta}_{t,\text{mid}}, \bar{\theta}_{t,\text{high}}\}$
      for $\bar{\boldsymbol{\theta}}_F \in \bar{\Theta}_F$ **do**
        $\hat{\boldsymbol{\theta}}_{\text{MLE}} \leftarrow \text{argmin}_{\boldsymbol{\theta} \in \mathcal{T}_\kappa(\bar{\boldsymbol{\theta}}_F)} \ln(\det(\mathbf{K}_{\boldsymbol{\theta}}(\mathbf{X}))) + \mathbf{y}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1}(\mathbf{X})\mathbf{y}$ ▷ Eq. (1), (5)
        $q \leftarrow q + 1$
        $\hat{\boldsymbol{\phi}} \leftarrow \hat{\boldsymbol{\theta}}_{\text{MLE}}$
        $\rho := \text{RelMSE}(\mathbf{y}_{\text{test}}, \hat{\mu}(\mathbf{X}_{\text{test}})),\ T := \text{TLL}(\mathbf{D}_{\text{test}}, \hat{\boldsymbol{\phi}}) \leftarrow \hat{\boldsymbol{\phi}}$ ▷ Eq. (3), (4)
        if $\rho < \text{RelMSE}_*$ or $R > \rho > \text{RelMSE}_*$ and $T > \text{TLL}_*$ **then**
          $\text{RelMSE}_* \leftarrow \rho$
          $\text{TLL}_* \leftarrow T$
          $\hat{\kappa},\ \hat{\mathcal{T}},\ \hat{\boldsymbol{\theta}} \leftarrow \kappa,\ \mathcal{T}_\kappa(\bar{\boldsymbol{\theta}}_F),\ \hat{\boldsymbol{\theta}}_{\text{MLE}}$
        end if
        if $\text{RelMSE}_* < 0.05$ or $q \geq Q$ **then return** $\hat{\kappa},\ \hat{\mathcal{T}},\ \hat{\boldsymbol{\theta}}$
        end if
      end for
    end for
  end for
end for
return $\hat{\kappa},\ \hat{\mathcal{T}},\ \hat{\boldsymbol{\theta}}$

The dilemma of choosing between different acquisition recommendations is further exacerbated by the fact that it is difficult to discern the quality of the different belief models and their recommendations. This is because each is based on a different statistical metric, as Eq. (2) shows. In other words, it cannot be generally stated that, or when, one acquisition function yields better design suggestions than another one. This lack of prior knowledge inspires an application of selection methodologies originating from reinforcement learning, in particular k-armed bandits [45]. While the theory of k-armed bandits provides a reward-based strategy towards candidate selection, there is no accompanying measure of reward – stochastic or deterministic – that is both intuitive and easily described or modeled. For this reason, there is not much motivation to employ (partly) deterministic selection schemes such as the (ε-)greedy algorithm. Facing these difficulties, alternative criteria need to be devised in order to select a candidate.

Suppose that $\mathcal{A} := \{\alpha_1, \dots, \alpha_A\}$ is the set of possible acquisition functions in a given toolbox. Each acquisition $\alpha_a : [0,1]^D \to \mathbb{R}$, where $a \in \{1, \dots, A\}$, can be optimized to create a matrix of candidate designs $\mathbf{X}_{\text{cand}}^{(i)} := (\mathbf{x}_1^{(i)\top}, \dots, \mathbf{x}_A^{(i)\top})^\top$ at iteration $i \in \{1, \dots, I\}$, where

$$\mathbf{x}_a^{(i)} := \underset{\mathbf{x} \in [0,1]^D}{\text{argmax}}\ \alpha_a(\mathbf{x}; \hat{\boldsymbol{\phi}}^{(i)}).$$

A selection strategy, generically denoted by "Sel", outputs $\mathbf{x}_a^{(i)}$ for some $a \in \{1, \dots, A\}$ as a response to the candidate design matrix $\mathbf{X}_{\text{cand}}^{(i)}$ and all available data $\mathbf{D}^{(i-1)}$.

One possible strategy consists of selecting $\mathbf{x}_a^{(i)}$ randomly. Out of the possible random strategies, uniform random sampling is the most straightforward:

$$\text{Sel}_{\mathcal{U}}(\mathbf{X}_{\text{cand}}^{(i)}, \mathbf{D}^{(i-1)}) = \text{Sel}_{\mathcal{U}}(\mathbf{X}_{\text{cand}}^{(i)}) := \mathbf{x}_a^{(i)}, \quad a \leftarrow \mathcal{U}_{\{1,\dots,A\}} \tag{6}$$

It should be noted that the selection strategy in Eq. (6) does not actually depend on any of the previously available data $\mathbf{D}^{(i-1)}$. However, inspired by the approach taken by solving the k-armed bandits problem, a (fully stochastic) selection method will be described, which does make use of the available data at optimization iteration i.

Let N be a positive integer and let $\mathbf{p} = (p_1, \dots, p_N)^\top$ be a probability vector, i.e. $0 \leq p_n \leq 1$ for all $n \in \{1, \dots, N\}$ and $\sum_{n=1}^N p_n = 1$. Let $\text{Cat}(N, \mathbf{p})$ denote the categorical probability distribution supported on $\{1, \dots, N\}$, defined by probability mass function $P(C = n) = p_n$ for any $n \in \{1, \dots, N\}$ if $C \sim \text{Cat}(N, \mathbf{p})$.

Define $p_a^{(1)} := 1/A$ for all $a \in \{1, \dots, A\}$. For $i > 1$, let $a^{(i-1)}$ be the selected value for a at iteration $i-1$. Then, define the categorical probability vector $\mathbf{p}^{(i)} = \mathbf{p}^{(i)}(\mathbf{D}^{(i-1)}) := (p_1^{(i)}, \dots, p_A^{(i)})^\top$ recursively as follows:

$$p_a^{(i)} := \frac{n_a^{(i)}}{N^{(i)}} \quad \text{with} \quad n_a^{(i)} := \begin{cases} n_a^{(i-1)} + 1 & \text{if } a = a^{(i-1)} \text{ and } y^{(i-1)} = \min \mathbf{y}^{(i-1)}, \\ n_a^{(i-1)} & \text{otherwise,} \end{cases}$$

$$\text{where} \quad n_a^{(1)} := 1 \quad \text{and} \quad N^{(i)} := \sum_{a=1}^{A} n_a^{(i)}.$$

Then, the categorical (also called multinomial) candidate design selection strategy can be formulated as follows:

$$\mathrm{Sel}_{\mathrm{Cat}}(\mathbf{X}_{\mathrm{cand}}^{(i)}, \mathbf{D}^{(i-1)}) := \mathbf{x}_a^{(i)}, \quad a = a^{(i)} \leftarrow \mathrm{Cat}(A, \mathbf{p}^{(i)}(\mathbf{D}^{(i-1)})). \tag{7}$$

In other words, $\mathrm{Sel}_{\mathrm{Cat}}$ will assign a larger probability to select the acquisition function α_a if it was able to locate the incumbent optimum at the previous iteration, whereas the probability mass distribution over the set of available acquisition functions will remain the same otherwise. This selection strategy is inspired by the `Dragonfly` implementation [46], the authors of which take a similar approach for selecting subsequent candidate designs and regression models as the outer BO steps progress.

The availability of previously sampled data $\mathbf{D}^{(i-1)}$ at iteration i can be exploited further when devising candidate design selection strategies for objective evaluation. While categorical selection defined in Eq. (7) only makes use of $\mathbf{y}^{(i-1)}$, the same selection methodology – and indeed uniform random selection, Eq. (6) – can be expanded based on $\mathbf{X}^{(i-1)}$.

BO needs to employ a careful trade-off between exploration and exploitation of the design space, especially applied to expensive problems. It will precisely be inefficient to sample two very similar designs twice, without exploring the design space first, even if one or both of them have been suggested by optimizing an acquisition function. It is therefore in order to encourage exploration when necessary, but still allow exploitation of promising design candidates as i approaches I. Fortunately, because there is a host of design candidates $\mathbf{X}_{\mathrm{cand}}^{(i)}$ to choose from, those candidates which are overly exploitative can be ruled out: a candidate design can be rejected based on its proximity to $\mathbf{X}^{(i-1)}$.

To make this notion of clustering concrete, the following is proposed: for any N design parameter vectors summarized in a matrix $\mathbf{U} := (\mathbf{u}_1^\top, \dots, \mathbf{u}_N^\top)^\top$, let $\delta_1, \dots, \delta_N$ be the minimum Euclidean distances between $\mathbf{U}$ and itself, defined as follows for any $j \in \{1, \dots, N\}$:

$$\delta_j := \min_{\substack{j' \in \{1,\dots,N\} \\ j' \neq j}} \|\mathbf{u}_{j'} - \mathbf{u}_j\|.$$

Subsequently, define the median minimum distance (MMD) of $\mathbf{U}$ as

$$\mathrm{MMD}(\mathbf{U}) = \mathrm{median}\{\delta_1, \dots, \delta_N\}.$$

An illustration of the MMD on different sample sets is shown in Fig. 2.

Next, define

$$d_{\min}(\mathbf{x}_a^{(i)}, \mathbf{X}^{(i-1)}) := \min_{j \in \{1,\dots,N+(i-1)\}} \|\mathbf{x}_a^{(i)} - \mathbf{x}_j^{(i-1)}\|$$

for $a \in \{1, \dots, A\}$ as the minimal Euclidean distance between $\mathbf{x}_a^{(i)}$ and any design row in $\mathbf{X}^{(i-1)}$.

It is now possible to calculate $\mathrm{MMD}(\mathbf{X}^{(i-1)})$, the median minimum distance of the design matrix at iteration $i-1$, and compare this value to $d_{\min}(\mathbf{x}_a^{(i)}, \mathbf{X}^{(i-1)})$. If $\mathrm{MMD}(\mathbf{X}^{(i-1)}) \gg d_{\min}(\mathbf{x}_a^{(i)}, \mathbf{X}^{(i-1)})$, then $\mathbf{x}_a^{(i)}$ is a relatively exploitative design, while $\mathrm{MMD}(\mathbf{X}^{(i-1)}) \ll d_{\min}(\mathbf{x}_a^{(i)}, \mathbf{X}^{(i-1)})$ indicates exploration by $\mathbf{x}_a^{(i)}$.

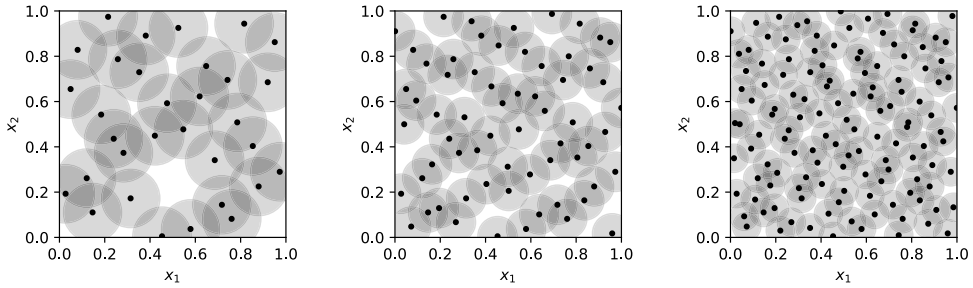
Thus, introducing the exploitation score (ES) of $\mathbf{x}_a^{(i)}$ with respect to $\mathbf{X}^{(i-1)}$ as

$$\mathrm{ES}(\mathbf{x}_a^{(i)}, \mathbf{X}^{(i-1)}) := \ln\left(\frac{\mathrm{MMD}(\mathbf{X}^{(i-1)})}{d_{\min}(\mathbf{x}_a^{(i)}, \mathbf{X}^{(i-1)})}\right),$$

it can be decided to refrain from evaluating the expensive objective f at $\mathbf{x}_a^{(i)}$ if $\mathrm{ES}(\mathbf{x}_a^{(i)}, \mathbf{X}^{(i-1)}) > t^{(i)}$ for some threshold value $t^{(i)} \in \mathbb{R}$. The explicit choice to make this threshold value depend on the outer loop iteration i stems from the desire to encourage exploration when i is small, yet allow exploitation when i is large. In general, $t^{(i)}$ is therefore programmed to decrease with respect to i. It should be noted that the logarithmic nature of ES conforms with the desirable property that a difference in exploitation score is proportional to the difference in *magnitude* of the distance ratio.

Fig. 2 showcases the intuitive notion that the MMD decreases as the number of samples increases. This implies a natural yet adaptive notion of which candidate designs are exploitative or exploratory, given a similar ES threshold parameter $t^{(i)}$: a candidate near a previously evaluated design sample in the 32 samples case is less exploitative in a similar 128 samples scenario.

The samples used in Fig. 2 are Sobol' samples, which are based on a low-discrepancy quasi-random Sobol' sequence [47]. The qualities of these samples are such that they are fully predictable (unlike random samples [48]) and can be easily extended to include an arbitrary number of samples (unlike Latin hypercube samples [49]). Furthermore, Sobol' samples can be used to obtain Saltelli samples [50], which is the core component of performing variance-based sensitivity analysis. A presentation of sensitivity analysis on the case study in Section 3 can be found in Appendix A.



(a) 32 samples (b) 64 samples (c) 128 samples

Fig. 2. Two-dimensional Sobol' samples (black dots) with disks (gray) of radius MMD.

Table 1
Selected synthetic objective functions to benchmark the BO schemes.

Function name	Formula	Unscaled domain	Multimodal	Global minimum
AlpineN2	$-\prod_{d=1}^{D}\sqrt{x_d}\sin(x_d)$	$[0,10]^D$	Yes	$x_d \approx 7.91$
Sphere	$\sum_{d=1}^{D} x_d^2$	$[-5,5]^D$	No	$x_d = 0$

2.4. Adaptive Bayesian optimization

By combining the ideas from surrogate model initialization and adaptive candidate selection with the BO algorithm (Algorithm 1), a set of extensions can be devised.

Algorithm 3 Bayesian optimization with Gaussian process initialization and input-adaptive candidate design selection (BO-GPi-iAda, adaptive BO).

Require: Design of train experiments $\mathbf{D}^{(0)}$, design of test experiments $\mathbf{D}_{\text{test}}$, set of covariance kernels $K_{\text{set}} = \{\kappa_{\text{RBF}}, \kappa_{\text{Mat}}, \kappa_{\text{RQ}}\}$, collection of sets of parameter indices to fix $\mathcal{F} = \{\mathcal{F}_0 = \varnothing, \mathcal{F}_1, \mathcal{F}_2\}$, nominal parameter fixture values $\bar{\Theta} = \{\{\bar{\theta}_{t,\text{low}}, \bar{\theta}_{t,\text{mid}}, \bar{\theta}_{t,\text{high}}\} : t \in \{1,\dots,D\}\}$, trial threshold Q, set of acquisition functions $\mathcal{A}$, number of iterations I, GPR RelMSE threshold R, GP initialization condition C, design candidate selection strategy Sel, exploitation score threshold $\mathbf{t}$

1: **for** $i = 1,\dots,I$ **do**
2: **if** $C(i)$ or $i = 1$ **then**
3: $\hat{\kappa},\ \hat{\mathcal{F}},\ \hat{\theta}^{(i)}_{\text{MLE}} \leftarrow \text{GPi}(\mathbf{D}^{(i-1)}, \mathbf{D}_{\text{test}}, K_{\text{set}}, \mathcal{F}, \bar{\Theta}, Q, R)$ ▷ Algorithm 2
4: **else**
5: $\hat{\theta}^{(i)}_{\text{MLE}} \leftarrow \text{argmin}_{\theta\in\hat{\mathcal{F}}} \ln(\det(\hat{\mathbf{K}}_\theta(\mathbf{X}^{(i-1)}))) + \mathbf{y}^{(i-1)\top}\hat{\mathbf{K}}_\theta^{-1}(\mathbf{X}^{(i-1)})\mathbf{y}^{(i-1)}$ ▷ Eq. (1)
6: **end if**
7: $\hat{\phi}^{(i)} \leftarrow \hat{\theta}^{(i)}_{\text{MLE}}$
8: $\mathbf{X}^{(i)}_{\text{cand}} \leftarrow (\text{argmax}_{\mathbf{x}\in[0,1]^D}\ \alpha_a(\mathbf{x}; \hat{\phi}^{(i)}))_{a=1,\dots,A}$
9: $\mathbf{X}^{(i)}_{\text{cand}} \leftarrow (\mathbf{x}^{(i)}_a \in \mathbf{X}^{(i)}_{\text{cand}} : \text{ES}(\mathbf{x}^{(i)}_a, \mathbf{X}^{(i-1)}) \le t^{(i)})$ ▷ Skip if no candidate satisfies ES threshold.
10: $\mathbf{x}^{(i)} \leftarrow \text{Sel}(\mathbf{X}^{(i)}_{\text{cand}}, \mathbf{D}^{(i-1)})$
11: $y^{(i)} \leftarrow f(\mathbf{x}^{(i)})$
12: $\mathbf{D}^{(i)} \leftarrow (\mathbf{D}^{(i-1)}, (\mathbf{x}^{(i)\top}, y^{(i)}))^\top$
13: **end for**
14: $(\mathbf{x}_{\text{rec}}, y_{\text{rec}}) \leftarrow \text{Rec}(\mathbf{D}^{(I)})$ ▷ Recommends the best-found optimizer and objective
15: **return** $(\mathbf{x}_{\text{rec}}, y_{\text{rec}})$

BO-GPi-Ada is analogously defined by removing the exploitation score threshold step 11 of BO-GPi-iAda. Similarly, BO-(i)Ada opts out on the GPi portion (steps 2–5) and performs the maximum likelihood estimation in the same way as is done in regular BO. Finally, BO-GPi hinges on one single acquisition function — similar to BO — while retaining the covariance kernel selection steps. See Fig. 3 for a schematic overview of BO-GPi-iAda.

Given the newly introduced algorithms, they are now compared to standard BO under a similar optimization budget constraint. To this end, two synthetic objective functions of different landscape qualities are introduced for the optimization algorithms to minimize. See Table 1 for a description of these synthetic objectives.

Fig. 4 shows the performance of BO-GPi-iAda compared to standard BO when optimizing the six-dimensional Sphere function (Sphere-6D) and the three-dimensional AlpineN2 function (AlpineN2-3D) over 64 initial Sobol' samples and 400 outer loop iterations.

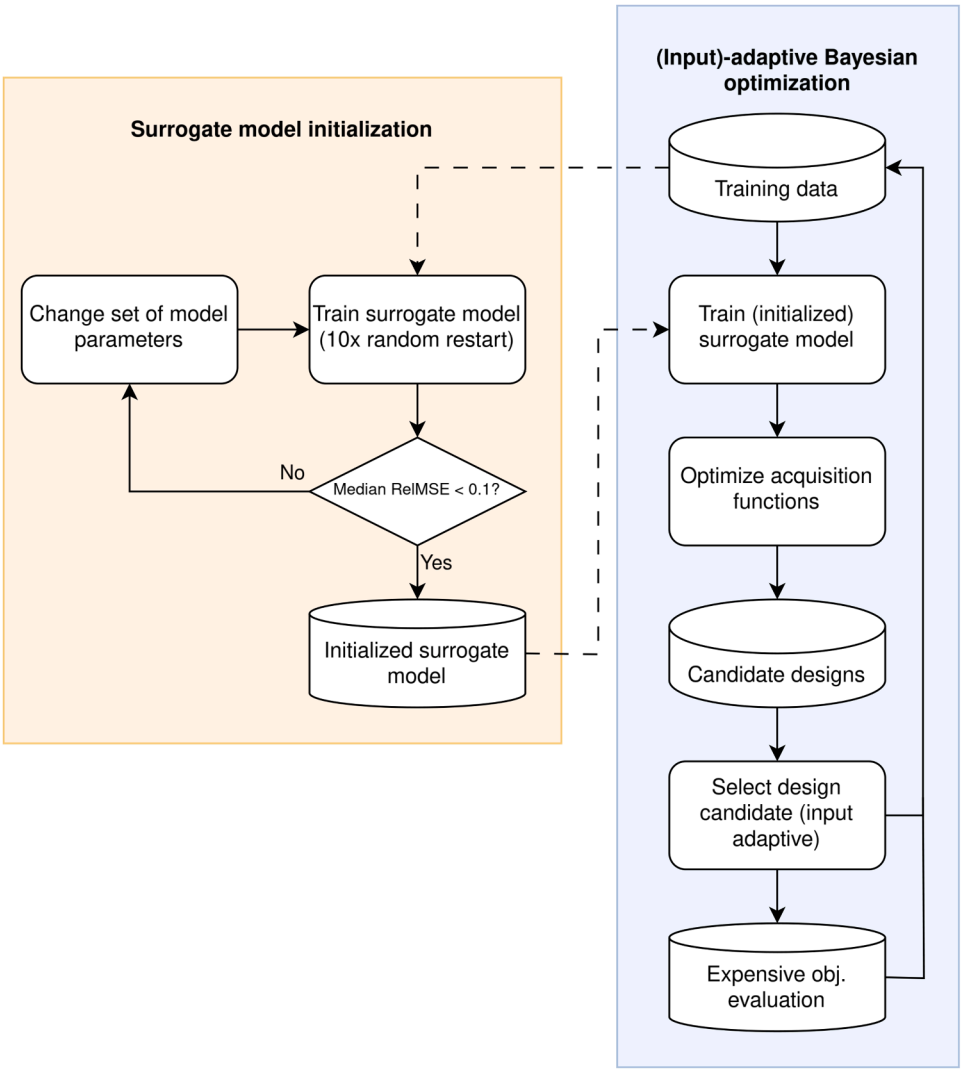


Fig. 3. Flowchart diagram of BO-GPi-(i)Ada, Algorithm 3. The dashed arrows represent connections that are only active when $C(i)$ is true.

In the adaptive BO scenarios, 20 % of the samples, rounded down, are held back for testing during GPi. As an example, this corresponds to 12 design samples when GPi is first performed, at the first adaptive BO iteration.

The core motivation behind using adaptive hyperparameters with BO is to at least perform better than the worst-case standard BO scenario. Let $\Omega := \{O_1, \dots, O_M\}$ be a collection of M (reference) optimizers. Next, let

$$\mathbf{w}_k(\Omega) := (\max\{q_k^{(i)}(O_1), \dots, q_k^{(i)}(O_M)\})_{i=1,\dots,I},$$

be the kth quartile worst-case aggregate across Ω. Then, for a collection of M' optimizers $\Omega' := \{O_1, \dots, O_{M'}\}$ which are to be compared to Ω, define

$$\mathrm{WCRI}_k(\Omega, \Omega') := \mathrm{median}\left(\frac{\mathbf{w}_k(\Omega) - \mathbf{w}_k(\Omega')}{\mathbf{w}_k(\Omega)}\right) = 1 - \mathrm{median}\left(\frac{\mathbf{w}_k(\Omega')}{\mathbf{w}_k(\Omega)}\right)$$

as a worst-case variant of the relative improvement in the kth quartile.

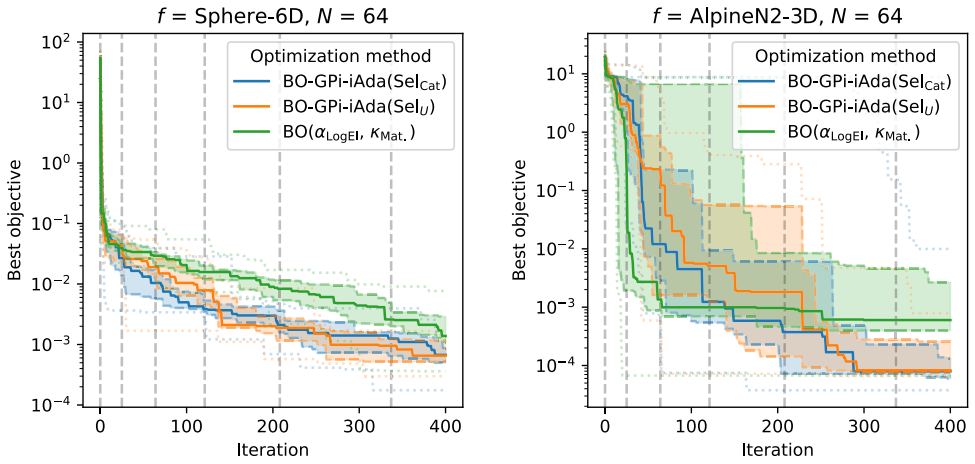
Given this new measure between two sets of optimizers, the optimization runs of the synthetic objectives that gave rise to Fig. 4 are reconsidered and expanded upon. The results are presented in Table 2.

The collection of reference optimizers considered in Table 2 is given by

$$\Omega = \{\mathrm{BO}(\alpha, \kappa) : \alpha \in \{\alpha_{\mathrm{LogEI}}, \alpha_{\mathrm{LogPI}}, \alpha_{\mathrm{UCB}}\}, \kappa \in \{\kappa_{\mathrm{RBF}}, \kappa_{\mathrm{Mat.}}, \kappa_{\mathrm{RQ}}\}\}$$

while the BO-GPi, BO-(i)Ada and BO-GPi-(i)Ada optimizer collections that are being compared to Ω are, respectively:

- $\Omega' = \{\mathrm{BO\text{-}GPi}(\alpha) : \alpha \in \{\alpha_{\mathrm{LogEI}}, \alpha_{\mathrm{LogPI}}, \alpha_{\mathrm{UCB}}\}\}$,



(a) Optimizing the Sphere-6D objective. (b) Optimizing the AlpineN2-3D objective.

Fig. 4. Five-number summary of BO (Algorithm 1) and BO-GPi-iAda optimization histories (Algorithm 3). A comparison between BO, with the Matérn covariance kernel and the logarithmic Expected Improvement acquisition function as chosen hyperparameters, and BO-GPi-iAda with two different candidate selection methods. The minimal incumbent objective value is plotted against the iteration. The vertical gray dashed lines indicate the iterations at which GPi takes place.

Table 2
Table of WCRI values of adaptive Bayesian optimizers compared to a standard BO reference towards the minimization of the indicated synthetic objective functions.

Objective	Model init.	Adaptivity	Sel	Relative improvement per quartile (%) Q_0	Q_1	Q_2	Q_3	Q_4
Sphere-6D	No GPi	No Ada		0.0	0.0	0.0	0.0	0.0
		Ada	$\mathrm{Sel}_{\mathrm{Cat}}$	75.6	72.2	61.5	41.2	34.4
			Sel_{U}	78.8	65.7	55.4	53.8	22.2
		iAda	$\mathrm{Sel}_{\mathrm{Cat}}$	82.4	85.9	71.5	65.9	46.1
			Sel_{U}	84.2	83.4	75.1	67.9	54.2
	GPi	No Ada		88.5	86.8	80.9	79.8	75.5
		Ada	$\mathrm{Sel}_{\mathrm{Cat}}$	93.4	94.5	90.7	88.6	89.0
			Sel_{U}	95.1	89.1	86.6	80.6	82.5
		iAda	$\mathrm{Sel}_{\mathrm{Cat}}$	96.1	94.4	93.9	93.2	92.0
			Sel_{U}	95.5	95.7	95.4	94.1	89.0
AlpineN2-3D	No GPi	No Ada		0.0	0.0	0.0	0.0	0.0
		Ada	$\mathrm{Sel}_{\mathrm{Cat}}$	78.3	59.1	54.5	95.9	0.0
			Sel_{U}	85.5	79.1	52.4	23.9	0.0
		iAda	$\mathrm{Sel}_{\mathrm{Cat}}$	33.2	13.2	-18.8	99.6	0.0
			Sel_{U}	68.9	58.5	43.9	99.7	0.0
	GPi	No Ada		15.0	36.5	-9.9	98.8	0.0
		Ada	$\mathrm{Sel}_{\mathrm{Cat}}$	97.4	96.8	96.3	100.0	0.0
			Sel_{U}	96.3	92.9	94.6	99.2	0.0
		iAda	$\mathrm{Sel}_{\mathrm{Cat}}$	91.9	91.1	90.1	99.9	0.0
			Sel_{U}	85.8	92.8	65.4	99.2	96.8

- $\Omega' = \{\text{BO-(i)Ada}(\kappa, \text{Sel}) : \kappa \in \{\kappa_{\text{RBF}}, \kappa_{\text{Mat.}}, \kappa_{\text{RQ}}\}, \text{Sel} \in \{\text{Sel}_{U'}, \text{Sel}_{\text{Cat}}\}\}$,
- $\Omega' = \{\text{BO-GPi-(i)Ada}(\text{Sel}_{U'}), \text{BO-GPi-(i)Ada}(\text{Sel}_{\text{Cat}})\}$.

From Table 2, it can be seen that there is significant confidence that an adaptive BO methodology will at least perform better than the worst-case standard Bayesian optimizer. These results furthermore show that a fully adaptive implementation (BO-GPi-(i)-Ada) leads to better optimization results than the alternatives (BO-(i)Ada, BO-GPi) when faced with the same computational budget. In other words, it is statistically very likely that adaptive BO will take fewer outer loop iterations to achieve a similar optimization objective value when compared to the worst-performing standard BO method.

Table 3
Table of median run times per iteration in seconds. All results were gathered using conventional CPU cores.

Objective	BO-GPi-iAda		BO
	$\mathrm{Sel}_{\mathrm{Cat}}$	Sel_{U}	–
Sphere-6D	21.20	21.12	9.89
AlpineN2-3D	23.50	20.33	7.05

Lastly, it is important to know about the additional computational expense of adaptive BO over standard BO. This will determine the feasibility of employing it to an engineering problem with an expensive objective. To do this, the run times were recorded that give rise to the results in Fig. 4 and Table 2, and the averages were recorded in Table 3.

From Table 3, it can be seen that the run times of the most extensive adaptive BO scheme, BO-GPi-iAda, are more than double that of standard BO. The relative increase in computational expense is thus quite substantial. However, expensive FEM simulations could take up one or multiple hours by using the same computational resources. An increase of 10 to 15 s per objective evaluation is therefore justified while using adaptive BO.

3. Problem description and data analysis

In order to put the adaptive BO schemes to the test and to confirm the promising findings that Table 2 implies, a case study in the automotive power electronics field was considered. The problem statement involves a commercially available Electronic Control Unit (ECU). The ECU consists of an outer housing shell protecting a printed circuit board (PCB) [51]. The PCB contains a number of electronic packages mounted on it. The aim is to find an optimal location on the PCB for a power package such that it extends the lifetime of the package-to-PCB solder interconnects. Another target is to further minimize the thermomechanical stresses in solder joints by selecting appropriate constituting materials of the layers within the power package. Therefore, the set goal is to optimize the position and constituting materials of the selected package on the available PCB area of approximately 163.4 mm × 163.4 mm for a minimum accumulated creep strain in the most critical solder joint of the package under thermal cyclic load. The accumulated creep strain value is a single dimensionless quantity that is to be interpreted as the single-output objective function f in the context of BO.

A submodeling-based approach was utilized to solve each case for the accumulated creep strain. Fig. 5 shows the relative footprints of the package and the whole PCB, the submodel of the package-on-PCB assembly, the solder joint layout, and two representative accumulated creep strain profiles.

In the submodel, a 13 mm × 13 mm area was defined for the PCB around the package, which has a maximum dimension of 6.45 mm. A commercially available software for FEM simulation was utilized to obtain the accumulated creep strain values after multiple temperature cycles between −40 and 125-C. This approach first solves for the displacements of the PCB-only mode, i.e., a global model of the housing and the PCB without any packages mounted on it. The displacement results are then used as the boundary conditions for the submodel — depending on its location on the PCB — using the cut-boundary interpolation technique along with the same thermal load as the global model. Applying boundary conditions extracted from the PCB-only global model works because the ratio of the relative size between the package and PCB is very small. This multi-scale approach is well-tested and established, the explanation and examples of which can be found in the literature concerning structural mechanics problems [52–54]. The global model utilizes fixed boundary conditions at the screws (indicated in Fig. 5) and linear elastic temperature-dependent material models for the PCB layers for the calculation of deformations.

The sub-model includes more complex material definitions for its sublayers. The outermost layer of the package encapsulation, which is an epoxy-based mold compound, utilizes a linear visco-elastic material model; the plasticity of solder material is defined with a creep law. The rest of the layers are defined using linear elastic temperature-dependent relations. The domains, i.e., the geometries in the global and sub-models, are discretized using three-dimensional finite elements with quadratic shape functions for better accuracy. The FEM was solved using a direct solver and was run on a high-performance computing system with 24 CPU cores and parallel processing enabled. The FEM simulation workflow of the submodel calculates the value of a damage parameter based on the nonlinear accumulated creep strain. This value is a volume-based weighted average of the accumulated creep strain over all the finite elements of a solder joint. The result corresponding to the solder joint 3 is selected as the target objective for the optimization problem. This is due to it being the smallest in dimension and, thus, the most critical one. Fig. 5(a) and (b) show two accumulated creep strain profiles, of which the volume-based average needs to be minimized.

Based on this FEM procedure, several design parameters were considered for the minimization problem regarding the accumulated strain objective f. The resulting design space $\mathcal{X}$ comprises geometric and material parameters. The geometric parameters include the package position, given by two-dimensional x and y coordinates, and the chip rotation angle, which is fixed at either 0° or 90°. The material parameters include the coefficients of thermal expansion (CTEs) of the molding compound before and after its glass transition. This is indicated by the teal-colored top layer of the package submodel in Fig. 5. These are denoted by CTE1 and CTE2, respectively. The selection of material parameters for this study is based on a previous study that shows that among several properties, thermal expansion coefficients of the molding compound affect the stresses in the solder joints the most [55]. The glass transition

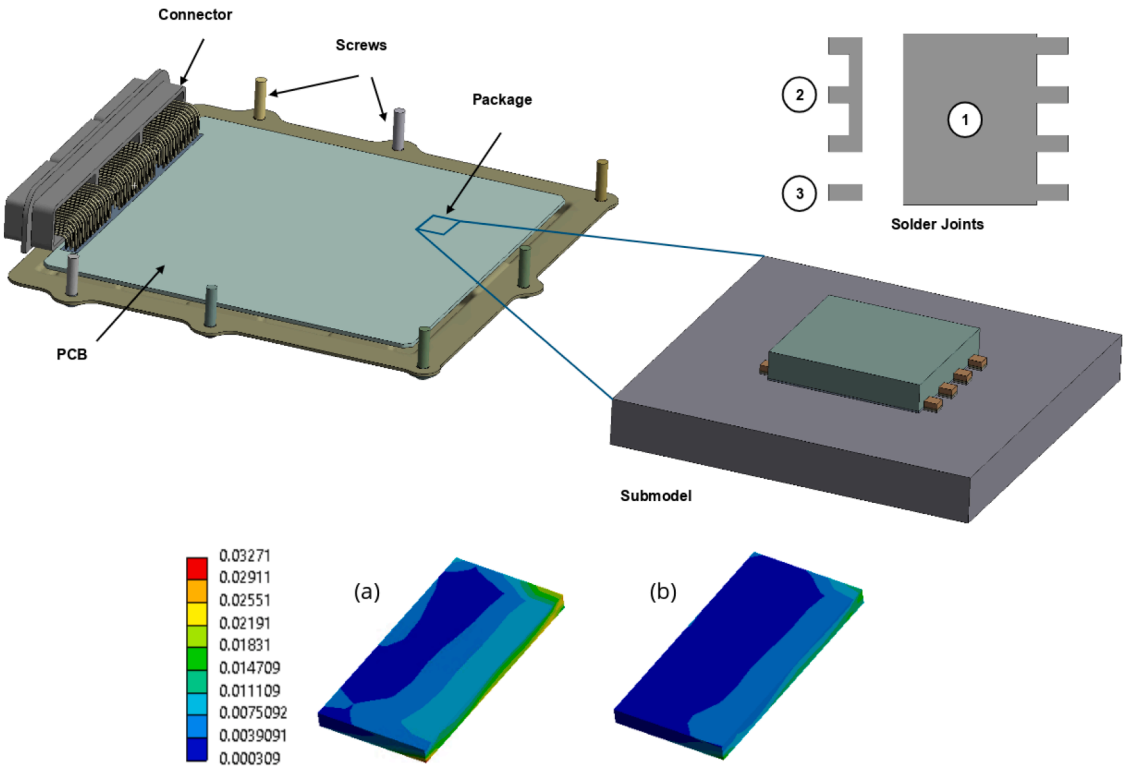


Fig. 5. Schematic overview of the PCB model, the electronic package submodel, and the solder joint interface (top right) subject to the design problem. Some accumulated creep strain profiles across solder joint 3 are displayed in subfigures (a) and (b).

Table 4
Design parameters in $\mathcal{X}$ for the accumulated creep strain optimization problem.

Design parameter	Lower bound	Upper bound
x	15mm	145 mm
y	20 mm	145 mm
rotation	{0°,90°}	
CTE1	5 ppm/°C	12 ppm/°C
CTE2	20 ppm/°C	37 ppm/°C

Table 5
Selected design parameters used for Fig. 5(a) and (b).

Design	CTE1 (ppm/°C)	CTE2 (ppm/°C)	x (mm)	y (mm)	Acc. creep (%)
(a)	6	27.7	58.07	108.35	0.30
(b)	8.5	31.9	58.07	131.28	0.16

temperature (T_g) was defined between 100 and 110° C. Considering the serviceable area available on the PCB, the range for the x and y coordinates of the centroid of the submodel was defined as 15mm–145 mm and 20mm–145 mm, respectively. The rotation was either set to 0° or to 90°. The range for CTE1 and CTE2 was set to 5mm–12ppm/° and 20mm–37ppm/°, respectively. Additional scripting was utilized to automate changing the design parameters, initiating the FEM solver, and extracting the results corresponding to the target objective.

A summary of all considered design parameters in the design space $\mathcal{X}$ can be found in Table 4. These parameters, together with the accumulated creep strain objective function (f), give rise to the designs to perform GP regression and BO, as described in Section 2.

From the design parameter selection as presented in Table 4, it should be remarked that the rotation value is discrete. At first glance, this seems incompatible with the assumed unit cube structure of the design space for applying BO, assumed in Section 2. This issue can be circumvented by separating the full 5D space into two 4D spaces, one for each rotation value. However, after performing data analysis (Appendix A), only one of these reduced design spaces was actually used to generate the results in Section 4.

For completeness, the design parameters leading to the accumulated creep strain profiles in Fig. 5 are given in Table 5.

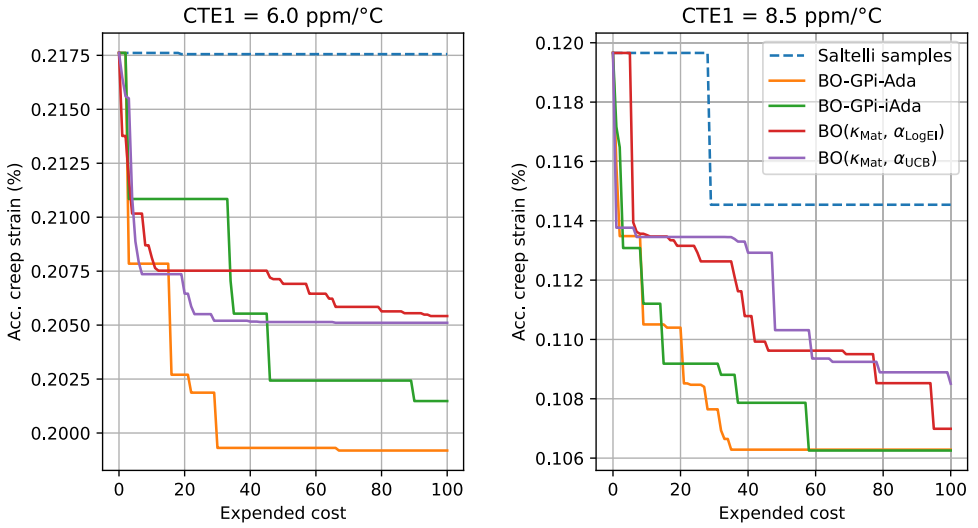


Fig. 6. Accumulated non-linear creep strain optimization histories, with the incumbent minimized strain value plotted against the expended cost incurred by the optimization scheme indicated by the legend. Each optimization method was allotted a computational budget of 100 objective evaluations. The adaptive BO schemes were used with a categorical candidate selection strategy at the acquisition step.

The computational expense of the FEM simulation to yield one single accumulated creep strain value is substantial: between 1.5 and 2 h. This is orders of magnitude above the reported per-iteration run times with the same computational resources reported in Table 3 and justifies the application of BO to optimize the accumulated creep strain objective.

4. Optimization results

In design problems, it is often reasonable to analyze the design space to avoid potential redundant parameters. Effective tools to this end include design axis projection ("pair-plotting") and Sobol' sensitivity analysis [56]. After performing sensitivity analysis, it has been determined that the 5D problem, as indicated in Table 4, can be reduced to a minimization problem with a 3D design space. This takes place by fixing specific values of chip rotation and CTE1, and subsequently optimizing the remaining design parameters. The details of this data analysis workflow can be found in Appendix A.

In line with the selected values of 0° for the rotation value, as well as 6 and 8.5 ppm/°C for CTE1, a variety of BO optimizers have been applied to optimize the accumulated strain value. For this benchmark, BO-GPi-iAda, BO-GPi-Ada, and two standard BO algorithms (Matérn covariance and LogEI / UCB acquisition) are selected. Furthermore, quasi-Monte Carlo sampling based on Saltelli samples has been used as a reference method to compare it with BO methods overall. This baseline methodology has been represented in various engineering studies [4,57] and is therefore important to take into consideration. For an overview of the cumulative optimization histories, see Fig. 6.

The two subfigures of Fig. 6 immediately show that all BO methods outperform quasi-random search by means of Saltelli sampling, both in terms of efficiency and raw objective value. This means that (adaptive) BO are useful optimization algorithms to handle this design problem. Strikingly, in the case where CTE1 = 6.0 ppm/°C, Saltelli samples are not able to provide any meaningful improvement at all.

More importantly, both subfigures of Fig. 6 show that the optimized designs that are located by BO-GPi-(i)Ada are achieved much faster. In the CTE1 = 6.0 ppm/°C case, BO-GPi-iAda achieves an accumulated creep strain level of 0.205 % within 45 expensive objective evaluations, while BO-GPi-Ada achieves it within 15. When CTE1 = 8.5 ppm/°C, BO-GPi-iAda attains 0.108 % creep strain within 40 evaluations, and BO-GPi-Ada within 30. In both cases, the worst-case BO method requires the full budget of 100 expensive objective evaluations to achieve similar levels of design improvement. This implies that more than 50 fewer expensive objective evaluations are needed compared to the (worst-case) BO scheme. This translates to roughly 75 h of computing effort, and is half of the allotted budget of 100 iterations. This shows that the covariance kernel and acquisition candidate selection strategies are effective in guiding the optimization effort, instead of having to rely on a fixed covariance kernel and acquisition type.

Fig. 6 furthermore shows that the kernel and acquisition adaptive BO scheme outperforms both standard BO runs across many of the iterations. It should be noted that the true, global minimum of the accumulated creep strain value across the design space is unknown. Therefore, it cannot be definitively concluded if the synthetic improvement results from Table 2 are reproduced in the solder joint optimization setting. However, as a direct comparison, when CTE1 = 6 ppm/°C, the solder joint design found by BO-GPi-Ada after the optimization budget was expended has a 2.9 % lower accumulated creep than the best-performing vanilla BO scheme. Furthermore, while the improvement resulting from BO-GPi-(i)Ada is only marginal when compared to BO (0.7 %) in the case of CTE1 = 8.5 ppm/°C, the *average* improvement across all iterations is 3.1 %. In other words: adaptive BO will statistically be able to

Table 6
Optimized design results after 100 iterations (CTE1 = 6 ppm/°C). Best values for standard and adaptive BO schemes are expressed in boldface.

Opt. method	CTE2 (ppm/°C)	x (mm)	y (mm)	Acc. creep (%)
BO(κ_{Mat},α_{UCB})	**37.0**	**69.5**	**20.0**	**0.205**
BO(κ_{Mat},α_{LogEI})	37.0	73.0	20.0	0.205
BO-GPi-Ada	**36.8**	**145**	**98.9**	**0.199**
BO-GPi-iAda	35.1	145	99.2	0.202

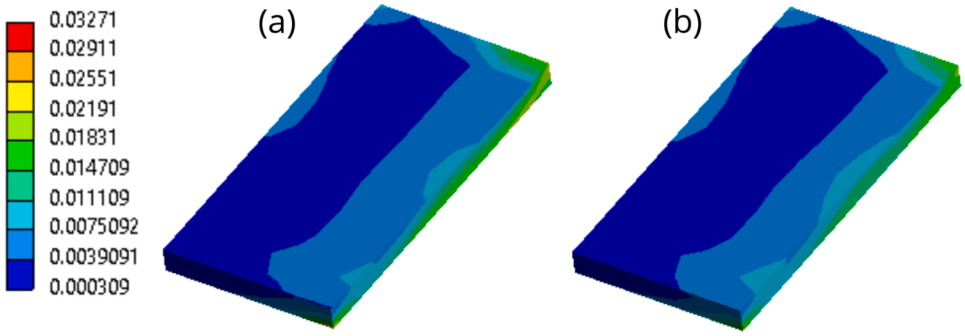


Fig. 7. Accumulated creep profiles across the critical solder joint for selected designs (CTE1 = 6 ppm/°C). The profiles correspond to (a) an optimized design by using BO, and (b) an optimized design by using BO-GPi-Ada. Note that the accumulated creep strain is a dimensionless quantity, which is reflected by the lack of units on the color bar.

Table 7
Optimized results after 100 iterations (CTE1 = 8.5 ppm/°C). Best values for standard and adaptive BO schemes are expressed in boldface.

Opt. method	CTE2 (ppm/°C)	x (mm)	y (mm)	Acc. creep (%)
BO(κ_{Mat},α_{UCB})	32.9	145	10	0.109
BO(κ_{Mat},α_{LogEI})	**35.1**	**145**	**99.8**	**0.107**
BO-GPi-Ada	**37.0**	**145**	**97.8**	**0.106**
BO-GPi-iAda	**37.0**	**145**	**98.2**	**0.106**

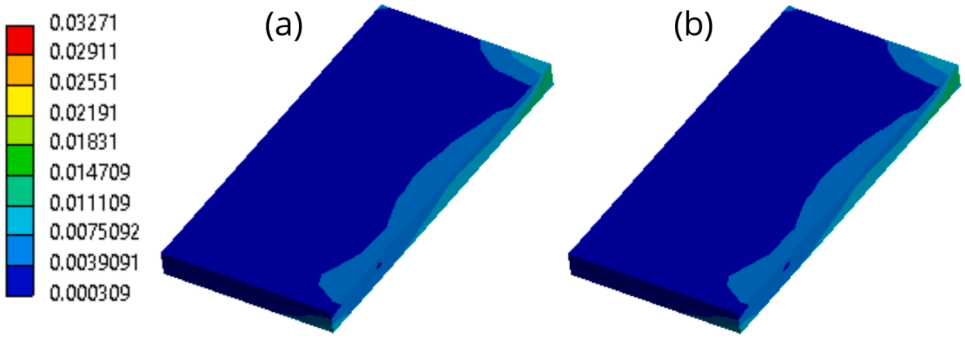


Fig. 8. Accumulated creep profiles across the critical solder joint for selected designs (CTE1 = 8.5 ppm/°C). The format is similar to that of Fig. 7.

achieve similar levels of improvement with a lower computational budget, confirming the positive outlook that the synthetic results present in Table 2.

The optimized design inputs and objective corresponding to each of the four optimization schemes used in Fig. 6 are recorded in Tables 6 and 7. Furthermore, the creep strain profiles of the critical solder joint corresponding to the best performing BO and adaptive BO designs are displayed in Figs. 7 and 8. The design parameters for these strain profiles are given in Table 5.

The rows in Tables 6 and 7 show that the optimizers find (local) minima at various locations in the design space, proving that this design space and objective pose a non-trivial optimization problem. From Figs. 7 and 8, it can be seen that the optimized creep strain profiles are substantially lowered compared to either of the selected non-optimized profiles in Fig. 5(a) and (b). Moreover, by comparing Fig. 7(a) with (b), it can be seen that the adaptive Bayesian scheme achieves lower accumulated creep strain along the edges of the critical solder joint, confirming the observation from the left subfigure of Fig. 6. Finally, it should be noted that the

profiles found by BO and adaptive BO in the case of CTE1 = 8.5 ppm/°C look similar, by comparing Fig. 8(a) to (b). This confirms that the objective performance of adaptive BO is comparable to that of standard BO, while providing a significant cut in computational expense.

The practical implications of these results have great significance in the greater scheme of next-generation micro-electronics reliability. There are more in number, more complex, and ever more capable electronic packages and components being deployed and integrated into products and systems. This makes their reliability even more important than ever before. The industry is moving towards the next wave in reliability engineering [58], where techniques such as prognostics and health management (PHM) play a crucial role. The PHM-based approach requires data-driven and machine learning techniques [59]. A hybrid PHM approach combines physics-based modelling and data-driven techniques. Finite Elements and simulation technology support physics-based methods, facilitating virtual design of experiments, which circumvent the costs of experimental tests. Integrating data-driven techniques, such as BO, into the FEM-based workflow further reduces the required computational efforts significantly. The results presented in this study help establish this and present an example that is key to the hybrid PHM approach.

5. Conclusion

In this work, a novel family of BO schemes is introduced, called adaptive BO. The core novelty behind this methodology is the focus on performing hyperparameter selection on the covariance kernel and acquisition function level between the outer optimization iterations. This happens by exploiting the low computational cost of the inner optimization loops of BO. Synthetic objective optimization results based on worst-case relative improvement show that adaptive BO outperforms BO in almost every quartile. The practical implications are significant: fast convergence of adaptive BO means that it has potential in saving hours of computational time when the objective is powered by expensive numerical simulations.

To investigate this hypothesis, a practical application was selected as a use case to apply adaptive BO to. This objective in this use case comprised the simulated thermomechanical behavior of accumulated creep strain within a critical solder joint on a PCB, where the goal was to minimize this objective. Results show that BO-GPi-(i)Ada achieves similar levels of optimized accumulated creep strain as BO does, while using half of the allotted budget. The implied impact of these results on micro-electronics reliability engineering cannot be understated: achieving a similar optimization result with only half of the allotted objective evaluation budget implies a saving of dozens, if not hundreds, of hours of FEM computations.

The ability of adaptive BO to save days' worth of computational effort showcases impressive potential to perform efficient data-driven design. However, the presented analyses have several limitations. These shortcomings are specified along with possible avenues of further investigation.

- **Longer run times.** The number of FEM simulations allotted for the BO results has been limited to 100. The schemes can find better optima when a larger number of outer loop iterations is considered. This could potentially underline the cost-efficiency of adaptive BO even more clearly.
- **More synthetic results.** The application of the adaptive BO heuristic has delivered promising results. However, the basis on which it was justified, namely the optimization of only two synthetic functions, could be expanded upon. A more extensive set of synthetic objectives should be considered to gain more statistical insights about the performance of adaptive BO. In particular, how do the adaptive schemes perform on classes of objective functions with specific general traits, such as convexity?
- **Different or more complex designs.** The FEM used throughout the design optimization can be expanded in multiple ways. A more complex material model can be used for the molding compound and for the calculation of non-linear strain in solder joints, in addition to the accumulated creep strain. Additional geometrical parameters can also be considered, such as the solder standoff height and the dimensions of the molding compound block. It should be remarked that a lot of these considerations will make the design optimization costlier, and thus more attractive to solve with (adaptive) BO.
- **Adaptive multi-fidelity or multi-objective BO.** Only single-objective, single-fidelity adaptive BO has been considered throughout our work. Multi-objective BO and multi-fidelity BO have previously been successfully applied to improve electronic reliability. It would thus be sensible to follow up with the integration of the novel adaptive BO technique in a multi-objective or multi-fidelity setting.
- **Benchmarking against other global optimizers.** The family of BO algorithms is part of a broader category of global, gradient-free optimizers. These include techniques such as Nelder-Mead or simplex descent, particle swarm optimization, and adaptive-evolutionary algorithms such as CMA-ES. To provide a more complete benchmark of adaptive BO, the performance of these optimization methods should be compared with, regarding both synthetic and engineering objectives.

CRediT authorship contribution statement

Leo Guo: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Data curation, Conceptualization; **Adwait Inamdar:** Writing – review & editing, Writing – original draft, Visualization, Resources, Investigation, Data curation, Conceptualization; **Willem D. van Driel:** Writing – review & editing, Supervision, Funding acquisition; **Guoqi Zhang:** Writing – review & editing, Supervision, Funding acquisition.

Data availability

Data will be made available on request. All presented data and the implemented workflow presented in this manuscript are open-source and accessible via GitHub: https://github.com/llguo95/COMPAS_simulation

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the ECSEL Joint Undertaking (JU) under Grant 826417. The JU receives support from the European Union's Horizon 2020 research and innovation program and Germany, Austria, Spain, Finland, Hungary, Slovakia, the Netherlands, and Switzerland.

A part of this work has been carried out within the project COMPAS, which is supported by ITEA, the Eureka Cluster on software innovation, under project number 19037. COMPAS received funding from Agentschap Innoveren en Ondernemen (Belgium), Bundesministerium für Bildung und Forschung (Germany), and Rijksdienst voor Ondernemend Nederland (Netherlands).

The authors would like to thank Jiaxiang Yi and Dr. Miguel A. Bessa for the valuable discussions and support throughout the implementation and validation of the data-driven methodology. Furthermore, the authors would like to thank Attila Gyarmati and Herbert Güttler from MicroConsult GmbH, Germany, for setting up the submodeling-based workflow and Dr. Martin Niessner (Infineon) and Dr. Michiel van Soestbergen (NXP) for providing the geometrical models of the package and PCB.

Appendix A. Data-driven design space analysis and dimensionality reduction

As a first step to perform analysis on the data, we plot the evaluated creep strain values against the design parameter axes, which can be seen in Fig. A.1.

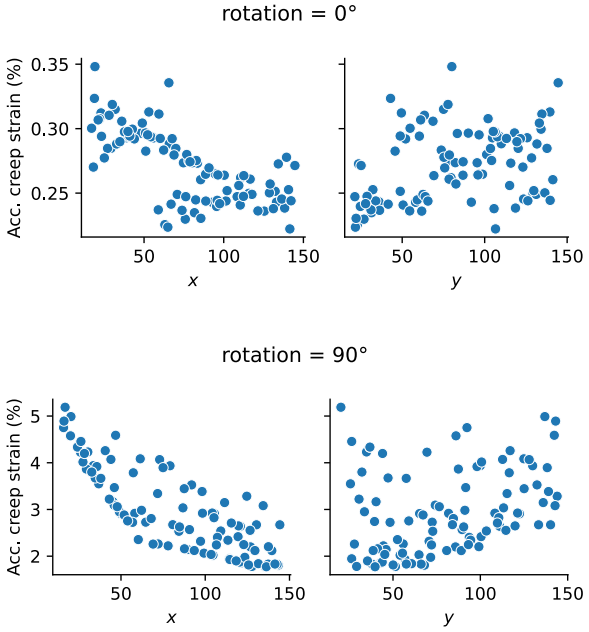


Fig. A.1. The collection of ±100 Saltelli samples (each rotation value) and their accumulated creep strain value (%) projections onto the remaining design parameter axes. For this analysis, fixed values for CTE1 (6 ppm/°C) and CTE2 (30 ppm/°C) were employed.

Note that the set of chip rotation values as determined in Table 4 is discrete, at either 0° or 90°. This yields two projection plots, one for each value. Furthermore, Fig. A.1 shows that the values of the accumulated creep strain in the 0° case are generally an order of magnitude lower when the package is rotated a quarter turn. As the objective is to minimize the strain as much as possible, the original design space can be effectively reduced by fixing the rotation to 0°.

For an objective projection plot corresponding to the remaining four design parameters in Table 4, see Fig. A.2.

From Fig. A.2, it can be seen that CTE1 has a clear correlation with the creep strain on the solder joints. On the other hand, CTE2, x, and y have less effect on the accumulated creep strain. In order to investigate this observation more rigorously, a global

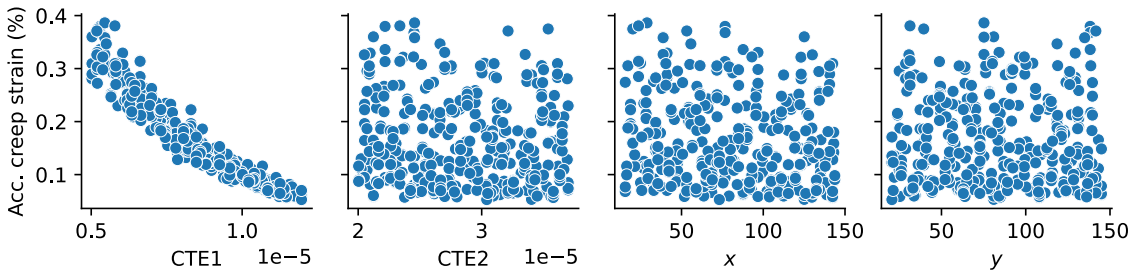


Fig. A.2. The collection of $354 = N_g(D+2)$ Saltelli samples and their accumulated creep strain value (%) projections onto the remaining design parameter axes. Here, $N_g = 59$ and $D = 4$.

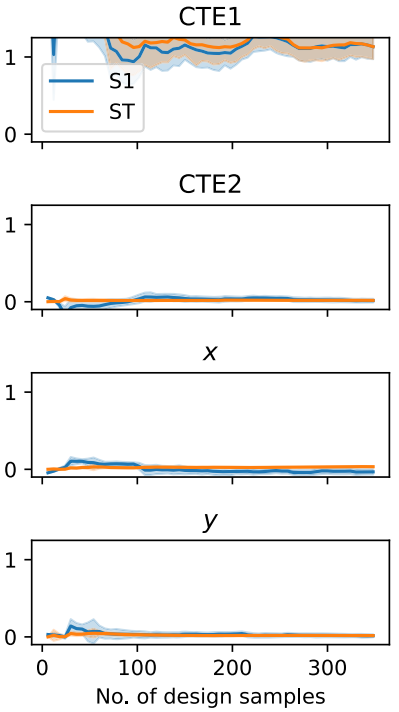


Fig. A.3. First- (S1) and total-order (ST) indices and bootstrapped confidence intervals for various numbers of (Saltelli) design samples, corresponding to the design problem with fixed rotation (0°).

sensitivity analysis is needed. It is commonplace for a more thorough design parameter sensitivity analysis to be performed when micro-electronic design is concerned [60–62]. In particular, Sobol' sensitivity analysis [56] is a common method due to its global and data-efficient nature, and has previously been utilized in e.g., aerospace [63], mechanical [64,65], environmental [66], and clinical [67] studies.

The Sobol' sensitivity analysis approach treats the objective and design parameters $f(X_1, \dots, X_D) = Y$ as fully stochastic, with $X_1, \dots, X_D, Y$ being modeled as random variables. Let $I = \{i_1, \dots, i_d\} \subset \{1, \dots, D\}$ be a subset of size d, and define $\mathbf{X}_I := (X_{i_1}, \dots, X_{i_d})^\top$. By the law of total variance,

$$\mathrm{Var}(Y) = \mathbb{E}(\mathrm{Var}(Y|\mathbf{X}_I)) + \mathrm{Var}(\mathbb{E}(Y|\mathbf{X}_I)),$$

[1] which splits the total, unconditional variance of the dependent variable Y into its unexplained and explained components with respect to the selected design parameters according to the indices in I. Particularly, the first and total-order sensitivity indices are of interest. These are respectively defined as

$$S_i := \frac{\mathrm{Var}(\mathbb{E}(Y|X_i))}{\mathrm{Var}(Y)}, \quad S_{T_i} := \frac{\mathbb{E}(\mathrm{Var}(Y|\mathbf{X}_{\sim i}))}{\mathrm{Var}(Y)}, \tag{A.1}$$

[1] For random variables U and V, it should be noted that $\mathbb{E}(U|V)$ is another random variable as function of V. Consequently, $\mathrm{Var}(\mathbb{E}(U|V))$ measures the variance of said function with respect to V and is therefore a (non-negative) number.

where X_i is an abuse of notation to denote $\mathbf{X}_{\{i\}}$, and $\mathbf{X}_{\sim i} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_D)^\top$. The indices as defined in Eq. (A.1), along with bootstrapped uncertainty bounds, can be efficiently approximated when Saltelli samples are assumed as realizations for the input factors. The details of the implementation will not be described in this manuscript, but can be referred to in the work by Saltelli et al. [50].

The obtained Saltelli samples in Fig. A.2 for the purpose of projection analysis can be subsequently (re)used to perform Sobol' sensitivity analysis. The sensitivity index convergence plots are shown in Fig. A.3. Note that sensitivity indices are dimensionless quantities.

As Fig. A.3 displays, the sensitivity of the strain objective with respect to the remaining parameters is small compared to the CTE1 design parameter. This is an incentive to reduce the 4D problem into a 3D optimization problem by fixing the CTE1 parameter.

In order to analyze the residual problem, a similar data analysis is done on the basis of several nominal values for CTE1. To select these values, note that there is a larger perceived variance of the data when the value of CTE1 is on the lower end of the defined range (leftmost subfigure of Fig. A.2). Hence, two nominal values for CTE1, being 6 ppm/°C and 8.5 ppm/°C, are decided upon. See Fig. A.4 for the accumulated creep strain samples projected on the remaining design parameter axes.

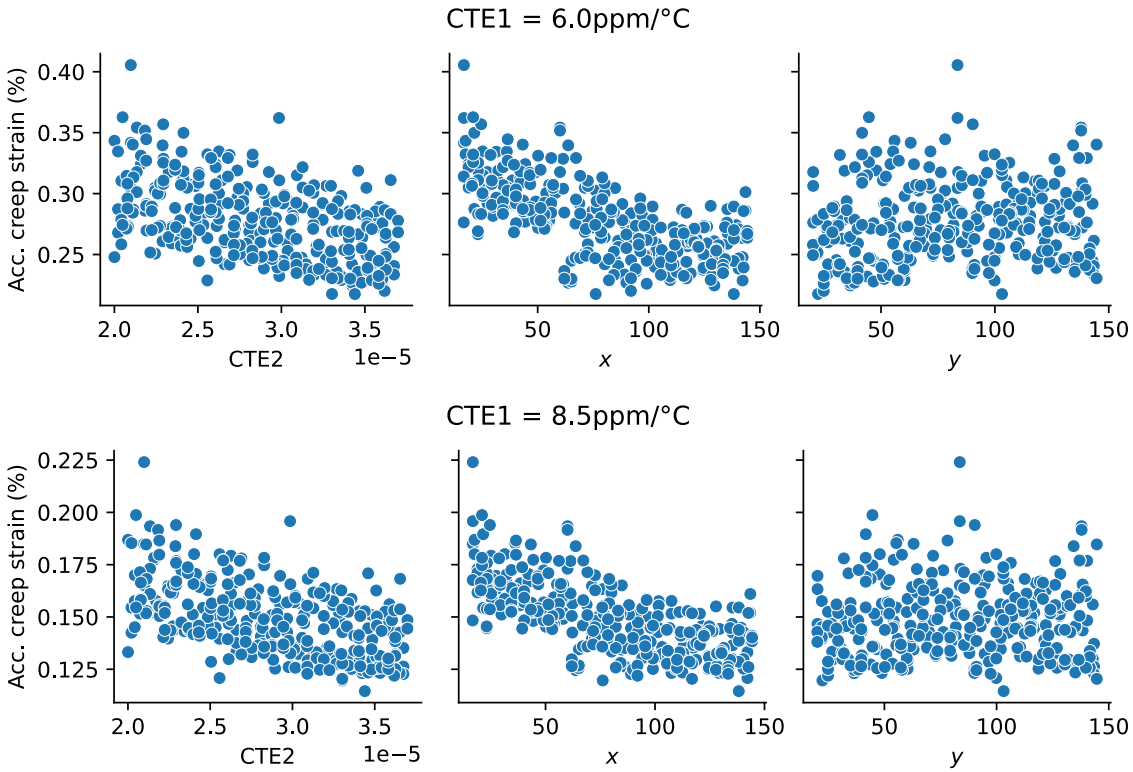
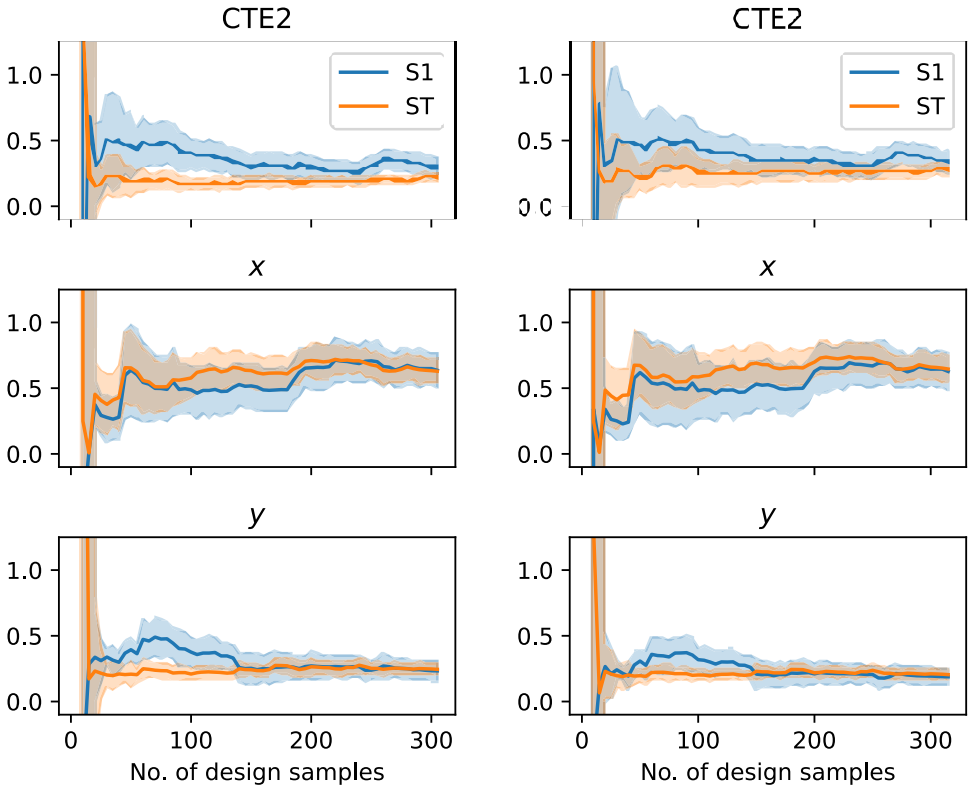


Fig. A.4. Each row shows a collection of $310 = N_g(D+2)$ Saltelli samples and their accumulated creep strain value projections onto the remaining design parameter axes for different fixed values of CTE1. Here, $N_g = 62$ and $D = 3$.

From Fig. A.4, it should be noticed that the behavior of the creep strain as a function of the design parameters CTE2, x, and y is very similar across the various CTE1 values. However, this behavior is exhibited on a different output scale, with higher CTE1 values corresponding to a lower creep strain. This is intuitively clear from the first subfigure in Fig. A.2.

The fact that the sensitivity profile is largely similar on the reduced design space is of importance because this justifies the representation of the entire CTE1 domain by virtue of fixing one or a few CTE1 values. To this end, one should confirm the heuristic that altering CTE1 does not influence the sensitivity of the objective with respect to the remaining parameters. To confirm this, the Saltelli samples have been used to construct a Sobol' sensitivity index convergence plot on the residual design space in Fig. A.5.

Fig. A.5 reveals that there is no significant alteration of the variance-based sensitivity profile when higher values of CTE1 are used. As mentioned previously, this fact allows for the dimensional reduction of the design problem by keeping CTE1 at fixed values when performing the optimization routine.



(a) CTE1 = 6 ppm/°C (b) CTE1 = 8.5 ppm/°C

Fig. A.5. First- (S1) and total-order (ST) indices and bootstrapped confidence intervals for various numbers of (Saltelli) design samples, corresponding to the design problem with fixed rotation (0°) and fixed CTE1.

References

[1] K. Guo, Z. Yang, C.-H. Yu, M.J. Buehler, Artificial intelligence and machine learning in design of mechanical materials, Mater. Horiz. 8 (4) (2021) 1153–1172.
[2] R. Garnett, Bayesian Optimization, Cambridge University Press, 2023.
[3] A. del Val, O.P. Le Maître, T.E. Magin, O. Chazot, P.M. Congedo, A surrogate-based optimal likelihood function for the Bayesian calibration of catalytic recombination in atmospheric entry protection materials, Appl. Math. Model. 101 (2022) 791–810.
[4] G. Li, Z. Chen, Z. Yang, J. He, Novel learning functions design based on the probability of improvement criterion and normalization techniques, Appl. Math. Model. 108 (2022) 376–391.
[5] N. Razaaly, G. Persico, G. Gori, P.M. Congedo, Quantile-based robust optimization of a supersonic nozzle for organic rankine cycle turbines, Appl. Math. Model. 82 (2020) 802–824.
[6] R. Tilgner, Physics of failure for interconnect structures: an essay, Microsyst. Technol. 15 (1) (2009) 129–138.
[7] M. Musadiq, W.D. van Driel, R. Roucou, R. Rongen, G. Zhang, Impact of IP block placement on solder joint reliability in IC packages, in: 2025 26th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), IEEE, 2025, pp. 1–7.
[8] M. Pecht, Handbook of Electronic Package Design, CRC press, 2018.
[9] B. Qiu, J. Xiong, H. Wang, S. Zhou, X. Yang, Z. Lin, M. Liu, N. Cai, Survey on fatigue life prediction of BGA solder joints, Electronics 11 (4) (2022) 542.
[10] E. Wolfgang, Examples for failures in power electronics systems, ECPE Tutorial Reliab. Power Electron. Syst. Nuremberg, Germany (2007) 19–20.
[11] S. Yang, D. Xiang, A. Bryant, P. Mawby, L. Ran, P. Tavner, Condition monitoring for device reliability in power electronic converters: a review, IEEE Trans. Power Electron. 25 (11) (2010) 2734–2752.
[12] T. An, C. Fang, F. Qin, H. Li, T. Tang, P. Chen, Failure study of Sn37Pb PBGA solder joints using temperature cycling, random vibration and combined temperature cycling and random vibration tests, Microelectron. Reliab. 91 (2018) 213–226.
[13] F. Arabi, A. Gracia, J.-Y. Delétage, H. Frémont, Effect of thermal and vibrational combined ageing on QFN terminal pads solder reliability, Microelectron. Reliab. 114 (2020) 113883.
[14] E.-H. Wong, W.D. Van Driel, A. Dasgupta, M. Pecht, Creep fatigue models of solder joints: a critical review, Microelectron. Reliab. 59 (2016) 1–12.
[15] S. de Jong, A.G. Ghezeljehmeidan, W.D. van Driel, Solder joint reliability predictions using physics-informed machine learning, Microelectron. Reliab. 172 (2025) 115797.
[16] A. Wymysłowski, W.D. Van Driel, J. Van De Peer, N. Tzannetakis, G.Q. Zhang, Advanced numerical prototyping methods in modern engineering applications–optimisation for micro-electronic package reliability, Microelectron. Reliab. 47 (2-3) (2007) 280–289.
[17] S. Kinoshita, Y. Inoue, T. Watanabe, K. Ikeda, S. Nishio, A. Teruya, N. Sakai, T. Goda, Space-filling latin hypercube design for efficient Bayesian optimization with application to semiconductor development, IEEE Trans. Semicond. Manuf. 184 (2025) 122263.
[18] D. Otaki, H. Nonaka, N. Yamada, Thermal design optimization of electronic circuit board layout with transient heating chips by using Bayesian optimization and thermal network model, Int. J. Heat Mass Tran. 184 (2022) 122263.

[19] E.T. Surillo, R.A. Sosa, C. Molina, H. Park, P. Nimbalkar, S. Srirangan, L.D. Dahal, Y. Lee, V. Smet, Bayesian optimization of large glass package architecture for system-level reliability in high-performance computing applications, in: 2024 IEEE 74th Electronic Components and Technology Conference (ECTC), IEEE, 2024, pp. 246–253.
[20] T.-T. Nguyen, M.-H. Pong, H.-J. Chiu, Multi-objective design automation in power electronics using Bayesian optimization techniques, in: 2025 IEEE Applied Power Electronics Conference and Exposition (APEC), IEEE, 2025, pp. 389–394.
[21] S. Zhang, S. Li, R.G. Harley, T.G. Habetler, An efficient multi-objective bayesian optimization approach for the automated analytical design of switched reluctance machines, in: 2018 IEEE Energy Conversion Congress and Exposition (ECCE), IEEE, 2018, pp. 4290–4295.
[22] S. Zhang, F. Yang, C. Yan, D. Zhou, X. Zeng, An efficient batch-constrained bayesian optimization approach for analog circuit synthesis via multiobjective acquisition ensemble, IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. 41 (1) (2021) 1–14.
[23] J. Huang, C. Tao, F. Yang, C. Yan, D. Zhou, X. Zeng, Bayesian optimization approach for RF circuit synthesis via multitask neural network enhanced Gaussian process, IEEE Trans. Microwave Theory Tech. 70 (11) (2022) 4787–4795.
[24] S. Zhang, W. Lyu, F. Yang, C. Yan, D. Zhou, X. Zeng, X. Hu, An efficient multi-fidelity bayesian optimization approach for analog circuit synthesis, in: 2019 56th ACM/IEEE Design Automation Conference (DAC), IEEE, 2019, pp. 1–6.
[25] X. Cai, H. Chen, Y. Zheng, L. Gao, Efficient Bayesian optimization for cheap and expensive multi-objective problems, Appl. Math. Model. 156 (2025) 116501.
[26] M. Feurer, J.T. Springenberg, F. Hutter, Initializing bayesian hyperparameter optimization via meta-learning, in: Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.
[27] L. Yang, A. Shami, On hyperparameter optimization of machine learning algorithms: theory and practice, Neurocomputing 415 (2020) 295–316.
[28] C.K.I. Williams, C.E. Rasmussen, Gaussian Processes for Machine Learning, 2, MIT press Cambridge, MA, 2006.
[29] C. Zhu, R.H. Byrd, P. Lu, J. Nocedal, Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization, ACM Trans. Math. Software (TOMS) 23 (4) (1997) 550–560.
[30] P.I. Frazier, A tutorial on Bayesian optimization, arXiv preprint arXiv:1807.02811 (2018).
[31] D.R. Jones, M. Schonlau, W.J. Welch, Efficient global optimization of expensive black-box functions, J. Global Optim. 13 (4) (1998) 455–492.
[32] H.J. Kushner, A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise, J. Basic Eng. 86 (1) (1964) 97–106.
[33] P. Auer, Using confidence bounds for exploitation-exploration trade-offs, J. Mach. Learn. Res. 3 (Nov) (2002) 397–422.
[34] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
[35] S. Ament, S. Daulton, D. Eriksson, M. Balandat, E. Bakshy, Unexpected improvements to expected improvement for Bayesian optimization, Adv. Neural Inf. Process. Syst. 36 (2023) 20577–20612.
[36] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, N. De Freitas, Taking the human out of the loop: a review of Bayesian optimization, Proc. IEEE 104 (1) (2015) 148–175.
[37] D. Duvenaud, J. Lloyd, R. Grosse, J. Tenenbaum, G. Zoubin, Structure discovery in nonparametric regression through compositional kernel search, in: International Conference on Machine Learning, 2013, pp. 1166–1174.
[38] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Comput. Sci. 7 (2021) e623.
[39] A. Gelman, J. Hwang, A. Vehtari, Understanding predictive information criteria for Bayesian models, Stat. Comput. 24 (2014) 997–1016.
[40] S.K. Deshpande, S. Ghosh, T.D. Nguyen, T. Broderick, Are you using test log-likelihood correctly?, arXiv preprint arXiv:2212.00219 (2022).
[41] S. Watanabe, Tree-structured parzen estimator: understanding its algorithm components and their roles for better empirical performance, arXiv preprint arXiv:2304.11127 (2023).
[42] F. Hutter, H.H. Hoos, K. Leyton-Brown, Sequential model-based optimization for general algorithm configuration, in: International Conference on Learning and Intelligent Optimization, Springer, 2011, pp. 507–523.
[43] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, D.D. Cox, Hyperopt: a python library for model selection and hyperparameter optimization, Comput. Sci. Discovery 8 (1) (2015) 014008.
[44] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: a next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 2623–2631.
[45] A.G. Barto, Reinforcement learning, in: Neural Systems for Control, Elsevier, 1997, pp. 7–30.
[46] K. Kandasamy, K.R. Vysyaraju, W. Neiswanger, B. Paria, C.R. Collins, J. Schneider, B. Poczos, E.P. Xing, Tuning hyperparameters without grad students: scalable and robust Bayesian optimisation with dragonfly, J. Mach. Learn. Res. 21 (81) (2020) 1–27.
[47] G. Leobacher, F. Pillichshammer, (T, M, S)-Nets and (T, S)-Sequences, in: Introduction to Quasi-Monte Carlo Integration and Applications, Springer, 2014, pp. 107–142.
[48] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1) (1970) 97–109.
[49] M.D. McKay, R.J. Beckman, W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 42 (1) (2000) 55–61.
[50] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, Comput. Phys. Commun. 181 (2) (2010) 259–270.
[51] M. Niessner, A. Gyarmati, H. Guettler, Impact of mechanical material modeling on the solder joint fatigue analysis of a leadless package mounted at different positions inside a generic aluminum ECU, in: 2023 24th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), IEEE, 2023, pp. 1–13.
[52] J. Zarbakhsh, A. Iravani, Z. Amin-Akhlaghi, Sub-modeling finite element analysis of 3D printed structures, in: 2015 16th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems, IEEE, 2015, pp. 1–4.
[53] S. Chang, K. Liu, M. Yang, L. Yuan, Theory and implementation of sub-model method in finite element analysis, Heliyon 8 (11) (2022) e11427.
[54] L. Kong, Q. Yao, X. Lv, W. Zhao, Submodelling method for modelling and simulation of high-density electronic assemblies, in: Journal of Physics: Conference Series, 1633, IOP Publishing, 2020, p. 012054.
[55] P. Gromala, A. Inamdar, A. Prisacaru, M. Dressler, A. Kabakchiev, Degradation and remaining useful life prediction of automotive electronics, in: Reliability of Organic Compounds in Microelectronics and Optoelectronics, Springer, 2022, pp. 415–447.
[56] I.M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simul. 55 (1–3) (2001) 271–280.
[57] D. Azzimonti, D. Ginsbourger, C. Chevalier, J. Bect, Y. Richet, Adaptive design of experiments for conservative estimation of excursion sets, Technometrics 63 (2020) 1–14.
[58] W.D. van Driel, M.Y. Mehr, X.J. Fan, G. Zhang, Outlook: from physics of failure to physics of degradation, in: Reliability of Organic Compounds in Microelectronics and Optoelectronics: From Physics-of-Failure to Physics-of-Degradation, Springer, 2022, pp. 535–538.
[59] M. Pecht, Prognostics and health management of electronics, Encycl. Struct. Health Monit. 29 (2009) 222–229.
[60] B. Vandevelde, E. Beyne, K.G.Q. Zhang, J.F. Caers, D. Vandepitte, M. Baelmans, Solder parameter sensitivity for CSP life-time prediction using simulation-based optimization method, IEEE Trans. Electron. Packag. Manuf. 25 (4) (2003) 318–325.
[61] D.G. Yang, G.Q. Zhang, W. Van Driel, J. Janssen, H. Bressers, L.J. Ernst, Parameter sensitivity study of cure-dependent underfill properties on flip chip failures, in: 52nd Electronic Components and Technology Conference 2002.(Cat. No. 02CH37345), IEEE, 2002, pp. 865–872.
[62] D.-G. Yang, J.S. Liang, Q.-Y. Li, L.J. Ernst, G.Q. Zhang, Parametric study on flip chip package with lead-free solder joints by using the probabilistic designing approach, Microelectron. Reliab. 44 (12) (2004) 1947–1955.
[63] F. Girault, F.T. Herrador, B. Helber, A. Turchi, T. Magin, P.M. Congedo, Stochastic mesoscale characterization of ablative materials for atmospheric entry, Appl. Math. Model. 135 (2024) 745–758.
[64] S. Ceballes, A. Abdelkefi, Application of sensitivity analysis and uncertainty quantification methods on the dynamic response of general nonlocal beams, Appl. Math. Model. 97 (2021) 322–343.

[65] C. Zha, C. Pan, Z. Sun, Q. Liu, Reliability sensitivity analysis for water hammer-induced stress failure of fluid-conveying pipe, Appl. Math. Model. 130 (2024) 51–65.
[66] J. Nossent, P. Elsen, W. Bauwens, Sobol'sensitivity analysis of a complex environmental model, Environ. Modell. Software 26 (12) (2011) 1515–1525.
[67] X.-Y. Zhang, M.N. Trame, L.J. Lesko, S. Schmidt, Sobol sensitivity analysis: a tool to guide the development and evaluation of systems pharmacology models, CPT: Pharmacometrics Syst. Pharmacol. 4 (2) (2015) 69–79.