

A Bayesian sampling optimisation strategy for finite element model updating

*Original*

A Bayesian sampling optimisation strategy for finite element model updating / Raviolo, D., Civera, M., Zanotti Fragonara, L.. - In: JOURNAL OF CIVIL STRUCTURAL HEALTH MONITORING. - ISSN 2190-5452. - 15:6(2025), pp. 1585-1607. [10.1007/s13349-023-00759-5]

*Availability:*

This version is available at: 11583/2989470 since: 2024-06-13T01:57:42Z

*Publisher:*

Springer

*Published*

DOI:10.1007/s13349-023-00759-5

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)



# A Bayesian sampling optimisation strategy for finite element model updating

Davide Raviolo<sup>1</sup> · Marco Civera<sup>2</sup> · Luca Zanotti Fragonara<sup>3</sup>

Received: 28 March 2023 / Accepted: 27 December 2023  
© The Author(s) 2024

## Abstract

Model Updating (MU) aims to estimate the unknown properties of a physical system of interest from experimental observations. In Finite Element (FE) models, these unknowns are the elements' parameters. Typically, besides model calibration purposes, MU and FEMU procedures are employed for the Non-Destructive Evaluation (NDE) and damage assessment of structures. In this framework, damage can be located and quantified by updating the parameters related to stiffness. However, these procedures require the minimisation of a cost function, defined according to the difference between the model and the experimental data. Sophisticated FE models can generate expensive and non-convex cost functions, which minimization is a non-trivial task. To deal with this challenging optimization problem, this work makes use of a Bayesian sampling optimisation technique. This approach consists of generating a statistical surrogate model of the underlying cost function (in this case, a Gaussian Process is used) and applying an acquisition function that drives the intelligent selection of the next sampling point, considering both exploitation and exploration needs. This results in a very efficient yet very powerful optimization technique, necessitating of minimal sampling volume. The performance of this proposed scheme is then compared to three well-established global optimisation algorithms. This investigation is performed on numerical and experimental case studies based on the famous Mirandola bell tower.

**Keywords** Bayesian optimisation · Bayesian expected improvement · Finite Element Model Updating · Bell Tower · Masonry structures

## 1 Introduction

Finite Element Model Updating (FEMU) refers to all strategies and algorithms intended for the calibration of an existing FE model based on experimental evidence, especially vibration data [1]. Data used for such calibration or updating purposes can be acquired from occasional in situ surveys or by an embedded, permanent monitoring system. For the specific case considered here—and as very commonly used in real-life circumstances—modal parameters, extracted from

acceleration time histories, are utilised. This is a well-known example of the indirect FEMU method [2], where the input parameters are varied to match the output results (natural frequencies and mode shapes). At its core, this represents an optimisation problem.

Therefore, the aim is to estimate the mechanical properties to be assigned to the numerical model, given its geometry. Once calibrated, the FE model can be used in several ways. The most obvious application is for predictive analysis, e.g., to estimate the remaining resilience of the structure in case of strong motions or other potentially dangerous events. If a predictive model was already available before a specific damaging event (e.g., a major earthquake), updating that FE model and comparing the estimated stiffnesses before and after the seism can be used for model-based Structural Health Monitoring (SHM) and damage assessment [3, 4]. That allows not only basic damage detection but also specific advanced tasks such as damage localisation and severity assessment. A third use is for hybrid simulations [5, 6]; in these applications, a target structure is

✉ Marco Civera  
marco.civera@polito.it

<sup>1</sup> Department of Structural Engineering, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway

<sup>2</sup> Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, 10129 Turin, Italy

<sup>3</sup> School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, Bedfordshire MK43, UK

divided into experimental and numerical substructures, due to practical limitations or to save costs. FEMU allows one to match the response of the numerical components to their experimental counterparts. Finally, in the case of continuously monitored structures and infrastructures, constantly re-updating a detailed Finite Element Model represents an enabling technology required for Digital Twins. This can serve the decision maker to evaluate the current and future structural situation of the assets under management. More details about the basic and general concepts of FEMU can be found in the works of Friswell and Mottershead—e.g., [7] and [8].

### 1.1 Efficient Bayesian sampling for finite element model updating

Arguably, one of the major issues about the FEMU procedure described so far is that the optimisation problem can become computationally expensive and very time-consuming, especially when dealing with complex FE models.

On the one hand, numerical models are becoming more and more complex, and so very computationally demanding. The need for very efficient optimisation techniques suitable for potentially highly demanding tasks is therefore clear. Nonetheless, an optimisation algorithm should discern the global minimum across the function domain, thereby circumventing the risk of encountering local minima. Unfortunately, sampling efficiency and global search capabilities are somewhat conflicting goals. Consequently, global optimisation techniques that require high sampling volumes to search the space for the global optimum are frequently employed.

For these reasons, the approach proposed here employs Bayesian Sampling Optimisation [9] in the framework of FEMU. As will be described in detail in the Methodology section, Bayesian Sampling Optimisation (or simply Bayesian Optimisation, BO, for short) uses the basics of Bayes' Theorem to infer the best sampling strategy in the search domain. This greatly increases the computational efficiency of the procedure, vastly reducing the sampling volume required to attain a solution, especially when a larger number of parameters needs to be estimated at once, as in the case of damaged structures and infrastructures, where multiple areas can be affected by different levels of damage.

In common practice, also according to the visible crack pattern, the target system is divided into macro-areas, under the assumption that these substructures will have different mechanical properties [10]. In the most common case, the local parameters of these macro-areas (Young's moduli, etc.) must be jointly estimated to match the global dynamic response of the structure. Hence, the dimensionality of the search space of the optimisation function, defined by these numerous parameters, easily ten or more [11]—can become very high. As a note, it is important to remark that

this will be the intended use of the term 'Bayesian' in this work; other research works, for example [12–15], and Ref. [16] among many others, use the same adjective to refer to the estimated output. Instead, this paper focuses solely on Bayesian sampling and its effectiveness for the optimisation of expensive functions in search spaces characterized by high dimensionality.

### 1.2 Applications of FEMU to historical architectural heritage and earthquake engineering.

The proposed Bayesian Optimization-based FEMU strategy is validated on a case study of interest for Structural Dynamics purposes, the bell tower of the Santa Maria Maggiore Cathedral in Mirandola. This historical high-rise masonry building suffered extensive damage after the 2012 Emilia Earthquake and has been the subject of several research studies throughout the years—see e.g., [1]. Both numerically simulated and experimental data were employed, thereby allowing to benchmark the proposed approach with a known ground truth and in a controlled fashion.

Indeed, regarding this specific application, FEMU is especially important for Earthquake Engineering. After major seismic events, reliable and predictive FE models are required as soon as possible to design and evaluate temporary interventions that should be deployed in the immediate aftermath to secure the damaged structures. However, the calibration of such FE models is not trivial. This situation is relatively worse for masonry structures, where even the properties of the original (pristine) structure are more difficult to estimate than homogeneous materials such as structural steel. In architectural and cultural heritage (CH) sites, even the pre-earthquake material properties are often unknown due to the lack of historical records; yet they have notoriously low mechanical resistance, due to their centuries-old ageing. These aspects make these unique and irreplaceable structures strongly vulnerable. Among them, historical bell towers are at particular risk during seismic events, due to various factors such as their relative slenderness, several potential failure mechanisms, and building material (bricks and mortar) [17]. These further underscore the importance of implementing robust monitoring strategies to detect and track damage development in such structures [18].

Thus, FEMU represents a precious tool for CH. Moreover, even after the first phase of a post-earthquake emergency, model updating is an important tool for vibration-based continuous monitoring and/or periodic dynamic investigations [19]. In the short to medium term, seismic aftershocks can cause more damage than the main shock, as strong motions insist on accumulated damage; in the long run, the initial cracks can expose structural vulnerabilities to external environmental factors.

Some noteworthy examples of FEMU applications can be found in [17, 20–23] and [24]. A broader, up-to-date review of Structural Health Monitoring (SHM) techniques successfully applied to CH structures is given by [25], while [26] specifically delves into the historical and contemporary advancements in SHM concerning the Garisenda tower in Bologna, Italy, a heritage structure similar in many aspects to the case study under examination. Similarly, Refs. [27] and [28] thoroughly analyses the Civic Tower of Ostra, Italy, and the Civic Clock tower of Rotella, respectively, using detailed numerical models and experimental data to assess the structural condition and establish standards for ongoing maintenance, posing the accent on the use of Genetic Algorithms. The remainder of this paper is organised as follows. In Sect. 2, the theoretical background of Finite Element Model Updating and Bayesian Sampling Optimisation are discussed in detail. In Sect. 3, the specific methodology of the algorithm implemented for this research work is reported. The three optimisation algorithms used for the comparison of the results are also briefly recalled. Section 4 describes the case study. Section 5 comments on the results, comparing the BO estimates with the three benchmark algorithms and the findings retrieved from the published scientific literature on the same case study. Finally, Sect. 6 concludes this paper.

## 2 Theoretical background

Parametric models (such as finite elements models) are described by a vector of model parameters  $\theta$ . Thus, being  $M$  the model operator,  $y = M(x, \theta)$  returns the output vector  $y$  for a given input vector  $x$ . For obvious reasons, in model updating, it is preferable to adopt outputs that are independent of the input and dependent on the model parameters only (such as modal features). According to this assumption, the  $x$  vector can be dropped, and the input–output relationship is simply represented by  $y = M(\theta)$ .

Finite elements model updating methods fall into two categories, direct methods and iterative methods (the latter also called deterministic). *Direct methods* try to improve observed data and computed data agreement by directly changing the mass and stiffness matrices; this leads to little physical meaning (no correlation with physical model parameters), problems with elements connectivity, and fully populated stiffness matrices. For these reasons, they are seldom used in common structural engineering applications. The *iterative methods* attempt to obtain results that fit the observations by iteratively changing the model parameters: this enables retaining good physical understanding of the model and doesn't present the above-mentioned problems. The degree of correlation is determined by a *penalty function* (or *cost function*): optimising this function requires the

problem to be solved iteratively, which means computing the output (i.e., performing a FE analysis) of the numerical model at each iteration. Hence, a higher computational cost is the major drawback of iterative methods.

Many FEMU methods have been proposed and successfully used: sensitivity-based methods, [29, 30], and [31]; eigenstructure-assignment methods, [32] and [33]; uncertainty quantification methods [34]; sensitivity-independent iterative methods, [35]; and many more [36].

As described, model updating is an inverse problem, as it aims at inverting the relationship between model parameters and the model output to find the *optimal* set of parameters  $\theta$  that minimises the difference between computed data and measured data.

In this sense, model updating can be simply considered as the following constrained optimisation problem:

$$\theta^* = \underset{\theta \in D}{\operatorname{argmin}} F(M(\theta), f), \quad (1)$$

where  $\theta^*$  is the set of optimal parameters,  $D$  is the parameter space,  $F$  is the cost function and  $f$  is the measured data.

The whole process of solving  $F(M(\theta), f)$  – the output of the numerical model “post-processed” in some way by a cost function – may be conceived as computing an unknown (non-linear) objective function of the model parameters  $\theta$ , which constitutes the sole input of the numerical model to be updated. Typically, this objective function is non-convex and expensive to evaluate. The output surface of the objective function lies in a  $d$ -dimensional space, where  $d$  is the number of parameters to be optimised. The sampling volume is exponential to  $d$ , thus posing an implicit restriction to the number of parameters that can be optimised.

Many optimisation algorithms have been developed in the last decades, each of them with its peculiar strengths and weaknesses. Among them, three of the better-known and most extensively used are Generalized pattern search (GPS) algorithms, Genetic Algorithms (GA), and simulated annealing (SA) algorithms.

In recent years, BO has proven itself to be a powerful strategy for finding the global minimum of non-linear functions that are expensive to evaluate, non-convex and whose access to the derivatives is burdensome. Furthermore, Bayesian sampling optimisation techniques distinguish themselves as being among the most efficient approaches in terms of a number of objective evaluations [37–41].

The essence of the Bayesian approach lies in the reading of the optimisation problem given by the ‘Bayes’ Theorem’:

$$P(M|E) \propto P(E|M)P(M), \quad (2)$$

which mathematically states that the conditional probability of event  $M$  occurring given the event  $E$  is true is proportional to the conditional probability of event  $E$  occurring if event  $M$  is true multiplied by the probability of  $M$ . Here,

$P(M|E)$  is seen as the *posterior probability* of the model  $M$  given the evidence (or observations)  $E$ ,  $P(E|M)$  as the *likelihood* of  $E$  given  $M$  and  $P(M)$  as the *prior probability* of the model  $M$ . Essentially, the prior,  $P(M)$ , represents the extant beliefs about the type of possible objective functions, eventually based on the observations already at disposal. The posterior  $P(M|E)$ , on the other hand, represents the updated beliefs about the objective function, given the new observations. The process basically aims at estimating the objective function by means of a statistical surrogate function, or *surrogate model*.

Many stochastic regression models can be used as a surrogate: the model must be able to describe a predictive distribution that represents the uncertainty in the reconstruction of the objective function, in practice by providing a mean and a variance.

To efficiently select the next sampling point, the proposed approach makes use of an *acquisition function* defined over the statistical moments of the posterior distribution given by the surrogate. The role of the acquisition function is crucial since it governs the trade-off between exploration (aptitude for a global search of the minimum) and exploitation (aptitude for sampling regions where the function is expected to be low) of the optimisation process. Probability of Improvement (PI), Expected Improvement (EI) and Upper Confidence Bound (UCB) are among the most used and most popular acquisition functions in Bayesian optimisation applications.

## 2.1 Finite element model updating

As mentioned, when iterative model updating methods are involved, the solution to the problem described by Eq. (1) entails the optimisation of a highly non-convex, high-dimensional cost function. In this case, modal features have been chosen to evaluate the degree of correlation between experimental and theoretical results, by employing both natural frequencies and associated mode shapes.

The selection of the parameters to be updated is a crucial step to reduce optimisation complexity, retain good physical understanding and ensure the well-posedness of the problem [42]. Generally, good practices to avoid ill-conditioning or ill-posedness are (1) choosing updating parameters that adequately affect the model output and (2) reducing the number of parameters to limit the occurrence of under-determinacy issues in the updating problem [43]. The first task can be accomplished by using sensitivity-based methods to discard non-sensitive parameters, and the second by dividing the structure into sub-parts with the same material properties. Additionally, the richness and the nature of the measured data, in contrast to the degree of discretization of the finite element model, places a limit on the type and number of

parameters that can be updated while retaining physical meaningfulness.

Various issues of ill-conditioning or rank-deficiency can arise in relation to the specific optimisation technique used. In the case of the BO approach, the rank of the covariance matrix of the Gaussian Process (i.e., kernel matrix) may be source of some concern. The matrix can become nearly singular if (i) the original function that is being optimised is so smooth and predictable that leads to a high correlation between sampling points, thereby generating columns of near-one values, and/or if (ii) the sampled points are very close one to another (which typically happens towards the end of the optimisation process), thereby generating several columns that are almost identical [44].

## 2.2 Bayesian sampling optimisation algorithm

For highly non-convex cost functions and problems denoted by high dimensionality, traditional optimization algorithms may encounter difficulty in identifying the global optimum or fail to converge, even within the framework of well-posed problem sets. In this study, we undertake a comparison between the performance of the proposed Bayesian sampling optimization approach and the outcomes derived from the aforementioned three classical alternatives. These will be discussed later in a dedicated paragraph.

When dealing with expensive and non-convex functions to optimise, both efficiency (in terms of sampling) and global search capabilities are paramount. Indeed, several global optimisation techniques have been developed over the years, but very few perform well when the number of function evaluations is kept to a minimum. One way to deal with expensive functions is by using surrogate optimisation techniques. This approach consists in substituting the objective function with a fast surrogate model, which is then used to search for the optimum and speed up the optimisation process. Of course, the validity of the surrogate model, that is to say, its capability to represent the behaviour of the underlying objective function, is of uttermost importance to obtain good and reliable results. Unfortunately, when a linear regression of the form

$$y(\mathbf{x}^{(i)}) = \sum_h \beta_h f_h(\mathbf{x}^{(i)}) + \epsilon^{(i)} (i = 1, \dots, n), \quad (3)$$

is used to fit the data (where  $\mathbf{x}^{(i)}$  is the  $i$ -th sampled point out of a total of  $n$ ,  $y(\mathbf{x}^{(i)})$  is the associated objective value,  $f_h(\mathbf{x})$  is a function of  $\mathbf{x}$ ,  $\beta_h$  are coefficients to be estimated, and  $\epsilon^{(i)}$  are the independent errors, normally distributed), it is arduous to determine which functional form should be employed if none or scanty a priori information about the function of interest is available. As such, these strategies are often impracticable for model updating optimisation problems.

The approach of Bayesian sampling optimisation consists of a change of paradigm. Instead of trying to minimise the error  $\epsilon^{(i)}$  by selecting some functional form that aligns with the data, the focus is placed on modelling the error by means of a stochastic process, so that the surrogate model is of the form:

$$y(\mathbf{x}^{(i)}) = \mu + \epsilon(\mathbf{x}^{(i)})(i = 1, \dots, n), \tag{4}$$

where  $\mu$  is the regression term (the functional form is just a constant), and the error term  $\epsilon(\mathbf{x}^{(i)})$  is a stochastic process with mean zero (in other words, a set of correlated random variables indexed by space). This change of perspective about the surrogate function is comprehensively described in one of the most interesting papers on modern Bayesian optimisation, [38], where the proposed method is called Efficient Global Optimisation, EGO. Besides modelling the surrogate as a stochastic process, the Bayesian sampling optimisation method makes use of an acquisition function to perform a utility-based selection of the points to be sampled. These (a stochastic predictive/surrogate model combined with the acquisition function) are in fact the two key elements of Bayesian optimisation.

BO has gained much attention in the last decades only. However, the first works on the topic have been published in the early 60s by [45]. After some developments by [46], who used Wiener processes, the concept of Bayesian optimisation using Gaussian Processes as the surrogate model was first used in the EGO formulation, combined with the expected improvement (EI) concept [47].

In the last years, several research works have proven the advantages of using Bayesian optimisation with expensive non-convex functions [48], making it a popular and well-known global optimisation technique.

Fitting a surrogate model to the data requires carrying out an additional optimisation process for the determination of hyperparameters. Furthermore, the next point to be sampled is found by searching for the maximum of the acquisition function. Hence, the BO approach entails two secondary (arguably fast-computing) optimisation problems, to be solved at each iteration: this results in a somewhat fancy and potentially heavy algorithm, which is suitable only if the objective function is considerably expensive.

In the following, this notation is often used:

$$\mathcal{D}_{1:t} = \{\mathbf{x}_{1:t}, f(\mathbf{x}_{1:t})\}, \tag{5}$$

where  $\mathcal{D}_{1:t}$  denotes the observations set, or sample, made of  $t$  observations in total.  $\mathbf{x}_i$  is the input point vector of the  $i$ -th observation. This vector, in other words, contains the updating parameters (in the input domain). The length of  $\mathbf{x}_i$  equals  $d$ , the dimensionality of the updating problem, i.e., the number of updating parameters. Finally,  $f(\mathbf{x}_{1:t})$ , also abbreviated in  $\mathbf{f}_t$ , are the observed values of the objective

function at  $\mathbf{x}_{1:t}$ , i.e., the outputs of the cost function at each set of updating parameters  $\mathbf{x}_i$ .

While any probabilistic model can be adopted to describe the prior and the posterior, it should be (i) relatively light and fast, to provide quick access to predictions and related uncertainties, (ii) able to adequately fit the objective function with a small number of observations, since sampling efficiency is pursued, and (iii) the conditional variance must cancel if and only if the distance between an observation and the prediction point is zero, as this is one condition to ensure the convergence of the BO method [49].

Given these requisites, Gaussian Process priors are the chosen probabilistic model in the majority of modern Bayesian optimisation implementations. To mention some popular alternatives, [50] worked with random forests, [51] with deep neural networks, [52] made use of Bayesian neural networks, while [53] used Mondrian trees. GPs are well-suited for model updating problems where the penalty function to be minimised is continuous.

Given a Gaussian Process (seen as a continuous collection of random variables, any finite number of which have consistent joint Gaussian distributions [54]), of the form:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \tag{6}$$

where  $m(\mathbf{x})$  is the mean function, and  $k(\mathbf{x}, \mathbf{x}')$  is the covariance function (which models the level of correlation between two observations,  $f_i$  and  $f_j$ , relatively to the distance between the points  $x_i$  and  $x_j$ ), the covariance can be computed for each pair of sampled points and conveniently arranged in matrix form:

$$\mathbf{K} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \dots & k(\mathbf{x}_1, \mathbf{x}_t) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_t, \mathbf{x}_1) & \dots & k(\mathbf{x}_t, \mathbf{x}_t) \end{bmatrix}. \tag{7}$$

Many covariance functions  $k(\mathbf{x}, \mathbf{x}')$  (or *kernel functions*) can be chosen, as decreasing functions of the distance between points  $x_i$  and  $x_j$  in the input space.

Considering the joint Gaussian distribution:

$$\begin{bmatrix} \mathbf{f}_{1:t} \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k}^T & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix}\right), \tag{8}$$

where  $f_*$  is the objective output at  $\mathbf{x}_*$ , that is  $f_* = f(\mathbf{x}_*)$ , and

$$\mathbf{k} = [k(\mathbf{x}_*, \mathbf{x}_1) k(\mathbf{x}_*, \mathbf{x}_2) \dots k(\mathbf{x}_*, \mathbf{x}_t)], \tag{9}$$

the following *predictive distribution* can be derived—for a full analytical derivation, see [55]:

$$P(f_* | \mathcal{D}_{1:t}, \mathbf{x}_*) = \mathcal{N}(\mu_t(\mathbf{x}_*), \sigma_t^2(\mathbf{x}_*)), \tag{10}$$

where:

$$\mu_t(\mathbf{x}_*) = \mathbf{k}^T \mathbf{K}^{-1} \mathbf{f}_{1:t} \quad (11)$$

$$\sigma_t^2(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k} \sigma_t^2(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k} \quad (12)$$

In the above set of equations,  $\mu_t(\mathbf{x}_*)$  is the prediction over the objective function value at any chosen point  $\mathbf{x}_*$ , and  $\sigma_t^2(\mathbf{x}_*)$  is the variance of the prediction at  $\mathbf{x}_*$  (the subscripts here denote that the prediction and its variance come from a GP trained with the  $\mathcal{D}_{1:t} = \{\mathbf{x}_{1:t}, f(\mathbf{x}_{1:t})\}$  data sample).

To compute the prediction and the variance at  $\mathbf{x}_*$  from Eqs. (11) and (12) (by means of which exact inference is computed), it is necessary to invert the kernel matrix  $\mathbf{K}$ . This operation has a computational complexity of  $\mathcal{O}(N^3)$ , where  $N$  is the size of the (square) kernel matrix (which equals the number of observations,  $t$ ). While this operation is relatively fast on its own, it can lead to computationally burdensome workflows as (i) the BO approach entails the maximisation of the acquisition function, a task that may require computing thousands of predictions (especially in high dimensional problems), and (ii) the number of observations keeps increasing (and so the size of  $\mathbf{K}$ ) as the optimisation advances and new points are added to the observations set.

Therefore, when using Gaussian Processes, BO badly scales with the number of observations. One way to mitigate such a problem consists of limiting the number of observations used to fit the GP to a certain amount (e.g., defining an “active set” size of a few hundreds), by randomly choosing a new set of training points among the sample at each iteration of the algorithm. Indeed, this practice is applied in the implementation used within this work, with an active set size of 300 being used.

The choice of the kernel function deeply affects the smoothness properties of a GP. This must be coherent with the features of the underlying objective function to obtain a quality surrogate model. Moreover, as each problem has its own specifics, the kernel function must be properly scaled. To this extent, the kernel functions are generalized by introducing *hyperparameters*. In the case of a squared exponential kernel function, this results in the equation:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f \exp\left(-\frac{1}{2\theta^2} \|\mathbf{x}_i - \mathbf{x}_j\|^2\right), \quad (13)$$

where  $\sigma_f$  is the *vertical scale*, which is the GP’s standard deviation (i.e., describes the vertical scaling of the GP’s variance), and the hyperparameter  $\theta$  is the *characteristic length scale*, which defines how far apart the input points  $\mathbf{x}_i$  can be for the output to become uncorrelated. When dealing with anisotropic problems (as it is often the case with model updating), it is much more convenient to use separate length scales, one for each parameter. This is typically done with

automatic relevance determination kernels (ARD), that consists in using a vector of hyperparameters  $\theta$ , which size equals  $d$ .

In practical terms, when a specific length scale  $\theta_l$  assumes a significantly higher value compared to other length scales, the kernel matrix becomes independent of the  $l$ -th parameter.

The optimal set of hyperparameters  $\theta$  is computed by maximisation of the marginal log-likelihood of the evidence  $\mathcal{D}_{1:t} = \{\mathbf{x}_{1:t}, f(\mathbf{x}_{1:t})\}$  given  $\theta$ :

$$\log(p(\mathbf{f}_{1:t} | \mathbf{x}_{1:t}, \theta^+)) = -\frac{1}{2} \mathbf{f}_{1:t}^T \mathbf{K}^{-1} \mathbf{f}_{1:t} - \frac{1}{2} \log|\mathbf{K}| - \frac{t}{2} \log(2\pi), \quad (14)$$

where the  $\theta^+$  vector contains the length scales  $\theta_{1:l}$  plus the vertical scale, and the mean  $\mu_0$  (i.e., the constant regression term) of the GP (and therefore all the  $d + 2$  hyperparameters), so that  $\theta^+ := (\theta_{1:l}, \mu_0, \sigma_f)$ . In the previous equation, the dependency on  $\theta^+$  is obviously found in the kernel matrix  $\mathbf{K}$ .

By employing this approach, a sort of sensitivity analysis of the parameters over the sampled points is performed. This built-in feature of the Bayesian optimisation technique is certainly useful for what concerns structural model updating problems, where the system sensitivity to the updating parameters is often dissimilar and usually unknown.

Four different ARD kernels will be employed, two of which are in the form of Matérn functions [56]. These functions are defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{2^{\zeta-1} \Gamma(\zeta)} \left(2\sqrt{\zeta} \mathbf{x}_i - \mathbf{x}_j\right)^{\zeta} H_{\zeta} \left(2\sqrt{\zeta} \mathbf{x}_i - \mathbf{x}_j\right), \quad (15)$$

where  $\zeta$  is a smoothness coefficient, while  $\Gamma(\bullet)$  and  $H_{\zeta}(\bullet)$  are the Gamma function and the Bessel function of order  $\zeta$ , respectively. As the smoothness coefficient  $\zeta$  tends towards infinite, the Matérn function reduces to the squared exponential function; when  $\zeta$  tends towards zero, the Matérn function reduces to the unsquared exponential function. The four employed kernels are:

- an ARD unsquared exponential kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j | \theta^+) = \sigma_f^2 \exp(-D) \quad (16.a)$$

$$D = \sqrt{\sum_{l=1}^d \frac{(x_{i,l} - x_{j,l})^2}{\theta_l^2}}$$

- an ARD squared exponential kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j | \theta^+) = \sigma_f^2 \exp\left[\sum_{l=1}^d \frac{(x_{i,l} - x_{j,l})^2}{\theta_l^2}\right] \quad (16.b)$$

- an ARD Matérn 3/2 kernel ( $\zeta = \frac{3}{2}$ ):

$$k(\mathbf{x}_i, \mathbf{x}_j | \theta^+) = \sigma_f^2 \left( 1 + \sqrt{3}D \right) \exp \left( -\sqrt{3}D \right)$$

$$D = \sqrt{\sum_{l=1}^d \frac{(x_{i,l} - x_{j,l})^2}{\theta_l^2}} \tag{16.c}$$

- an ARD Matérn 5/2 kernel ( $\zeta = \frac{5}{2}$ ):

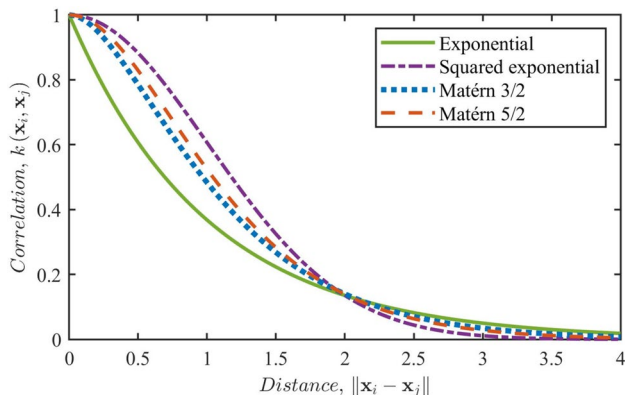
$$k(\mathbf{x}_i, \mathbf{x}_j | \theta^+) = \sigma_f^2 \left( 1 + \sqrt{5}D + \frac{5}{3}D^2 \right) \exp \left( -\sqrt{5}D \right)$$

$$D = \sqrt{\sum_{l=1}^d \frac{(x_{i,l} - x_{j,l})^2}{\theta_l^2}} \tag{16.d}$$

The difference between the four kernel functions is visible in Fig. 1. The exponential kernel stands out due to its rapid decline in correlation as distance increases. Consequently, function samples drawn from a GP constructed with the exponential kernel exhibit notably rugged features, whereas those generated using Matérn 3/2, Matérn 5/2, and squared exponential kernel functions display increasingly smoother characteristics.

As mentioned already, two different approaches can be followed to search for the optimum: the exploitative approach and the explorative approach. The automatic trade-off between exploitation and exploration is taken care of by the acquisition function.

Typically, Bayesian Optimization has found its historical roots within the scientific literature as a methodology primarily designed for the maximization of objective functions. Consequently, acquisition functions are conventionally formulated to yield high values in regions where the objective is deemed to exhibit a high value. Thus, when seeking



**Fig. 1** Correlation between observations  $\mathbf{x}_i$  and  $\mathbf{x}_j$  according to the four different kernel functions, plotted against the distance ( $\theta = 0.25$ )

the minimum of a function  $f(\mathbf{x})$ , such in the case of model updating, it is sufficient to consider the equivalent problem:

$$\operatorname{argmax}_{\mathbf{x}} g(\mathbf{x})$$

$$g(\mathbf{x}) = -f(\mathbf{x}) \tag{17}$$

The next point  $\mathbf{x}_{t+1}$  that will be chosen for sampling is found by maximising the acquisition function  $a(\mathbf{x})$  according to the optimisation problem:

$$\mathbf{x}_{t+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x} | \mathcal{D}_{1:t}) \tag{18}$$

Four acquisition functions were tested on a preliminary numerical case study, consisting of the updating of a simple 2D shear-frame (see Fig. 2): Probability of Improvement (PI), Expected Improvement (EI), a modified version of Expected Improvement [57] and Upper Confidence Bound (UCB). Among these, UCB was selected for implementation in the Bayesian sampling algorithm used in this study, as it was found to strike the best balance between exploitation and exploration. In particular, UCB was found to be about 40%, 20%, and 70% better than PI, EI, and EI (Bull) in terms of accuracy, respectively, given the same initial seed and the same total sampling volume.

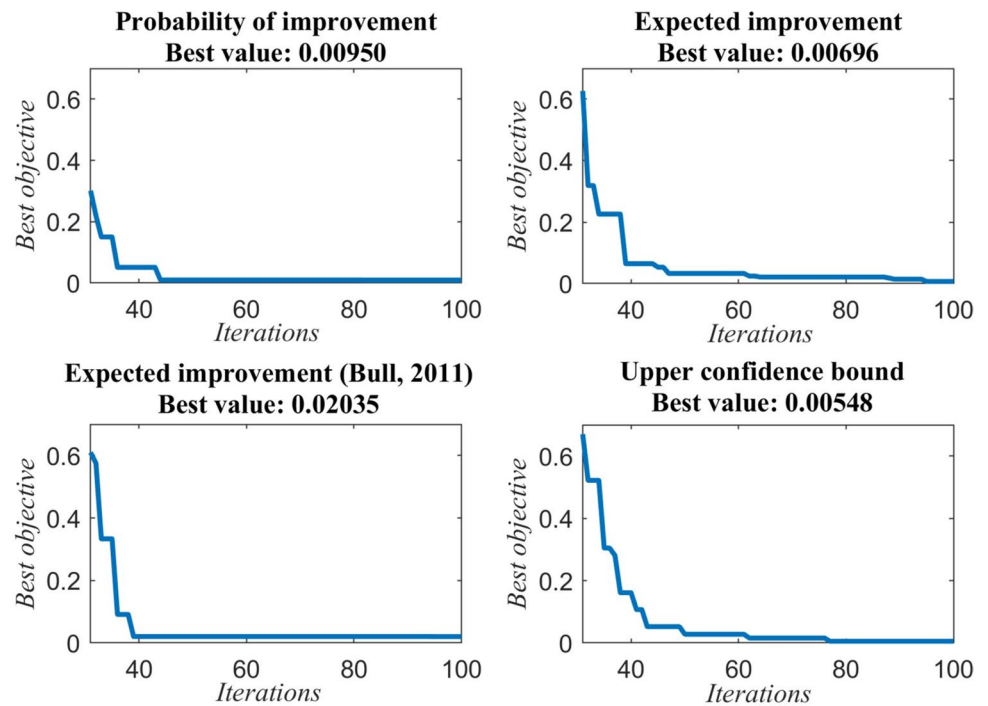
Upper Confidence Bound (or Lower Confidence Bound, LCB, if minimisation is involved), first proposed by [58] in the “Sequential Design for Optimisation” algorithm (SDO), consists of a very simple yet very effective approach. The UCB function is defined as:

$$\operatorname{UCB}(\mathbf{x}) = \mu(\mathbf{x}) + \kappa \sigma(\mathbf{x}) \tag{19}$$

where  $\kappa$  is typically a positive integer number, which controls the bound width identified by the standard deviation  $\sigma(\mathbf{x})$  and therefore the propensity of the algorithm to explore the search space. Often,  $\kappa$  is taken equal to 2, so that the confidence bound is about 95% (indeed, this is the value used in the following case-study).

The “intelligent” sampling performed by the BO approach when using UCB is displayed in Fig. 3. Here, a simple numerically-simulated updating problem consisting of a 3-DOF shear-type system is considered. Levels are all denoted by the same stiffness  $k$ , while the lumped masses are  $m_1, m_2$  and  $m_3$ . The parameters being updated consist of the stiffness  $k$  and the mass  $m_2$ , which target values are known in advance (as well as the target response of the system). The associated 2D penalty function is sampled at 9 randomly chosen points, a Gaussian Process is fitted to the observations, and the UCB function is computed knowing  $\mu(\mathbf{x})$  and  $\sigma(\mathbf{x})$  by means of Eqns. (11) and (12). The minimum according to the surrogate model and the acquisition function maximum (i.e., the succeeding sampling point) are visible. When using UCB (with  $\kappa = 2$ ), the choice of the next sampling point is

**Fig. 2** Comparative analysis of four distinct acquisition functions. The plots illustrate the trajectory of the optimal objective function value throughout the optimization iterations for the updating of a basic 2D shear-frame



heavily influenced by the high level of uncertainty in the predictions. The acquisition maximum is found in an area far from other observations, where uncertainty is very high, while the predicted objective is still reasonably low. Furthermore, a tendency to explore the optimization space at the early stages of the optimization procedure, when the number of observations is low and the uncertainty is high, followed by a gradually more exploitative behaviour as the overall uncertainty decreases, is a remarkable aspect of UCB functions for locating the global optimum, as both exploration and exploitation needs are upheld.

### 3 Methodology

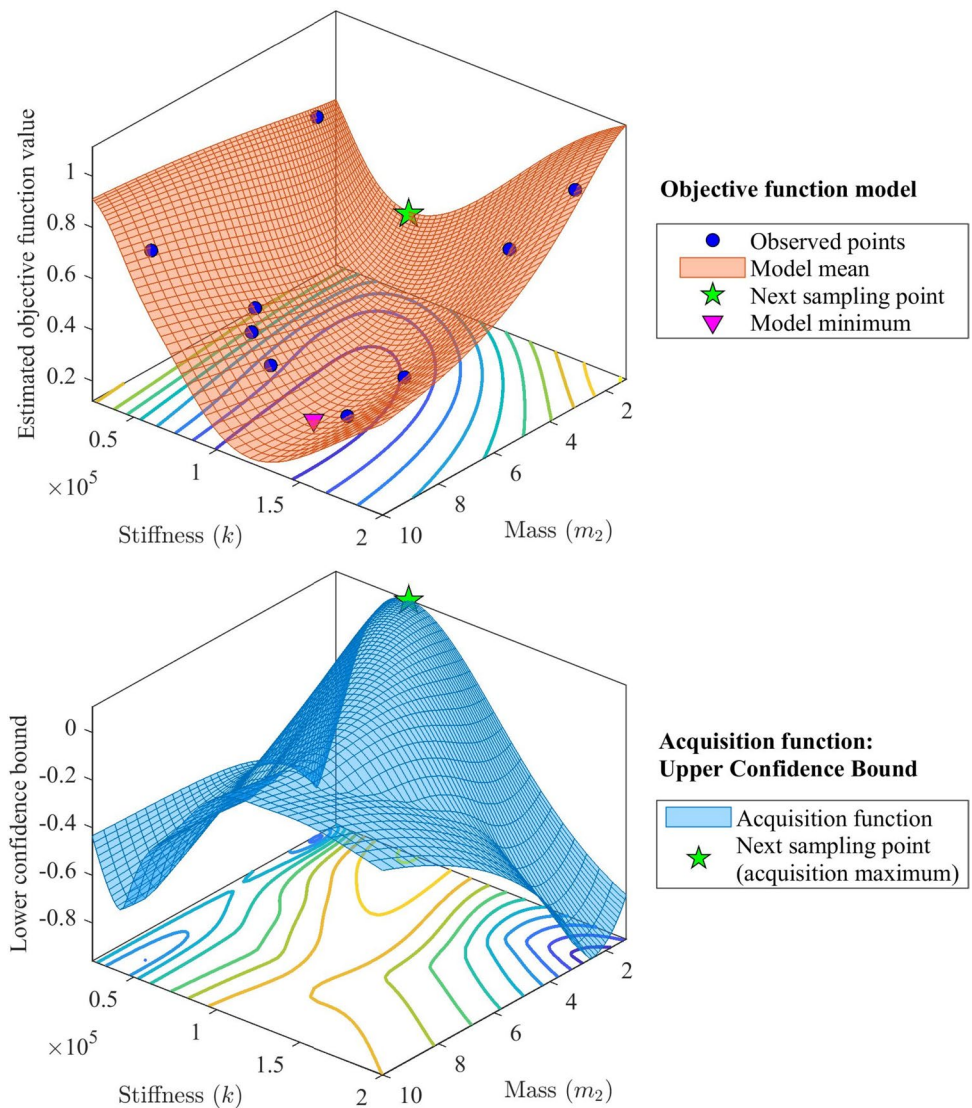
In this section, the implementation of the Bayesian optimisation algorithm strategy used in this study is described. For completeness, the main technical details about the optimisation techniques used for comparison are discussed as well. These algorithms are in fact very susceptible to specific implementation choices and initial parameters' values, as the optimisation outcome is affected both in terms of sampling efficiency and accuracy. This is particularly true for Simulated Annealing and Genetic Algorithm.

#### 3.1 Bayesian optimisation: the proposed algorithm

Technical details about the implementation of the employed Bayesian sampling optimisation procedure are summarized as follows, according to the flowchart represented in Fig. 4.

- (1) The optimisation procedure is initialized by computing the objective function at the seed points, which are randomly chosen within the optimisation domain, defined by the search bounds of each input parameter. The seed size should be sufficiently large to avoid overfitting when selecting the optimal set of kernel hyperparameters through log-likelihood maximisation. As a rule of thumb, [38] suggest setting the initial seed size at  $10 \cdot d$  at least, where  $d$  is the number of dimensions of the optimisation problem (i.e., updating parameters). Indeed, this criterion will be followed for the presented case study.
- (2) The fitting of the Gaussian Process (i.e., the surrogate model at iteration  $i$ ) occurs by maximising the marginal log-likelihood, which enables to select the optimal set of hyperparameters  $\theta^+$ . Moreover, a small amount of Gaussian noise  $\sigma^2$  is added to the observations (such that the prior distribution has covariance  $\mathbf{K}(\mathbf{x}, \mathbf{x}'; \theta) + \sigma^2 \mathbf{I}$ ).
- (3) To maximise the acquisition function, several thousands of predictions  $\mu_t(\mathbf{x}_*)$  are computed at points  $\mathbf{x}_*$  randomly chosen within the optimisation space. Then,

**Fig. 3** The upper visualization shows the GP mean (surface in red), the nine observations  $f_{1,9}$  used for training, the model minimum (i.e., the *believed* lowest function value according to the GP), and the next sampling point selected by the UCB acquisition function. The lower depicts the acquisition function surface  $UCB(x)$  and the next chosen sampling point, which corresponds to the acquisition function maximum



some of the best points are further improved with local search (for this application, the “*fmincon*” MatLab® function is used), among which the best point is finally selected.

- (4) The objective function is computed at the point corresponding to the acquisition function maximum.

By following this workflow, a newly fitted GP is used at each algorithm iteration. In fact, the objective function value computed at each  $i$ -th iteration is added to the set of observations at iteration  $i + 1$ , which is then employed to train the GP used to model the objective function, by determining a new set of hyperparameters via log-likelihood maximisation.

Before starting the actual procedure (step (2)), the code is implemented to perform cross-validation tests to determine which configuration of the GP is most suitable

for the specific updating problem, according to the following procedure. First, non-exhaustive cross-validation tests are performed to choose whether to log transform the input variables (i.e., updating parameters), as this is often found to improve the GP regression quality. To this extent, validation loss is computed for two GPs, one fitted using non-transformed variables, and the other fitted using log-transformed variables. Secondly, after choosing whether to transform the input variables or not, the GP is fitted four different times using the four kernel functions previously introduced. Once more, cross-validation loss is computed to establish which kernel is the fittest for modelling the objective function. Once the input-variables transformation is established and the GP kernel chosen, the algorithm is actually initialized.

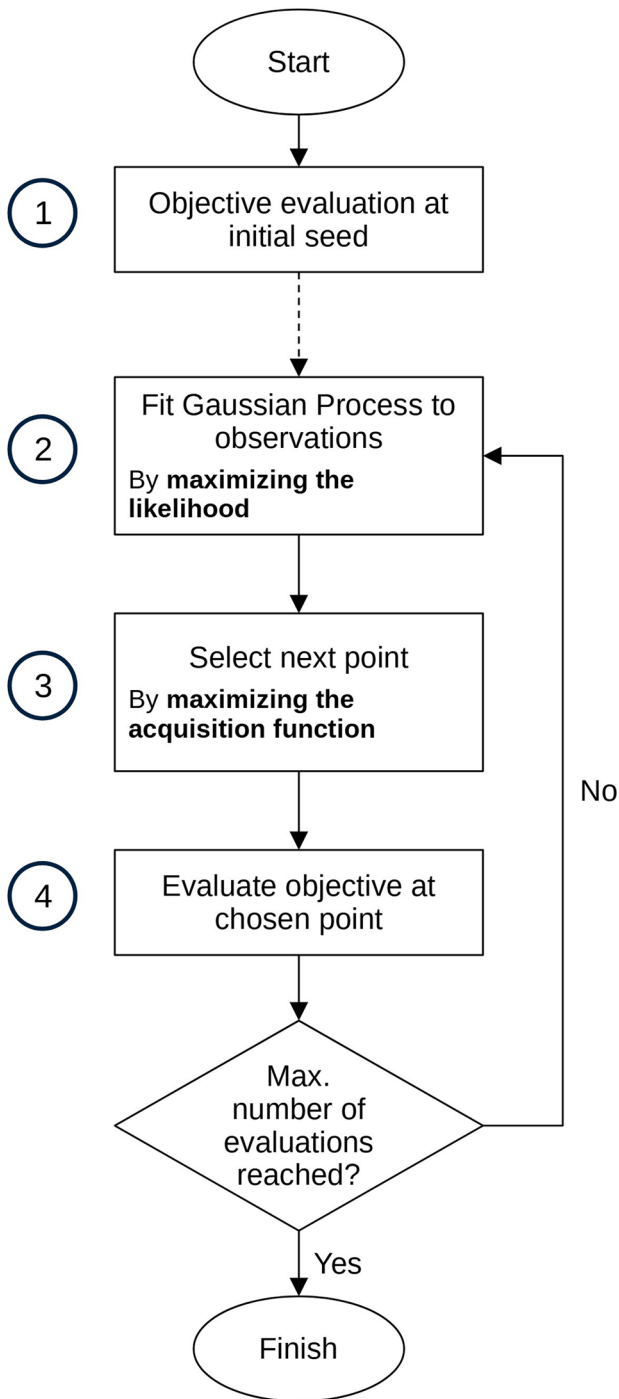


Fig. 4 Bayesian sampling optimisation algorithm flowchart

### 3.2 Benchmark algorithms: GPS, SA, and GA

Generalized pattern search is a relatively simple traditional optimisation algorithm, while Simulated Annealing, the Genetic Algorithm and BO are generally considered to be “computational intelligence” optimisation techniques. All four algorithms have in common that no use of derivatives

is made, hence the function is not required to be differentiable. Besides, despite the different approaches and backgrounds, Simulated annealing, GA and Bayesian sampling optimisation techniques share many elements: all algorithms are designed to carry out a global search of the minimum, avoiding local minima; they are appropriate for non-linear and non-smooth functions; finally, the algorithms are especially suitable for black-box models, where establishing in advance a functional form that effectively aligns with the data is often impossible. The key difference between the BO approach and the other techniques is that the former requires a much smaller sampling volume to achieve comparable optimization performance, greatly enhancing the computational efficiency when expensive cost functions are involved. Nonetheless, a greater sampling efficiency comes at the expense of a more sophisticated algorithm, that requires computationally intensive operations at each iteration. One shared drawback of GA, simulated annealing and Bayesian optimisation techniques is that these algorithms tend to give results close to the global minimum, although not very accurate. On the other hand, GPS can achieve high accuracy of solution.

The selection of parameters proves to be a critical aspect across all techniques. However, parameter selection for GPS, GA, and SA, as found by the authors, presents greater challenges compared to BO. The optimization performance exhibits heightened sensitivity to factors such as initial temperature, cooling schedule, and acceptance probability function for SA, crossover and mutation operators for GA and mesh shape and size in the case of GPS. On the contrary, while BO also demands user-defined parameters as described in Sect. 2.2—most notably, the kernel function—these do not pose significant difficulty and demonstrate robust generalization within the optimization framework of model updating cost functions. Furthermore, kernel hyperparameters are automatically optimized through maximization of the marginal log-likelihood, as already mentioned.

Reference can be made to [59] for a first implementation of a GPS algorithm, to [60] who initially proposed the GA algorithm, and to [61] for a first application of the SA concept to optimisation problems.

Given the diverse nature of the optimisation techniques employed, it is essential to exercise particular care in selecting each algorithm’s implementation strategy to ensure a fair comparison, necessary to evaluate the performance of the Bayesian sampling optimisation approach in model updating applications. In particular, GPS, SA and GA are allowed to sample the objective function 1000 times, while BO is stopped at 500 function evaluations. This is necessary since the former techniques typically need a much greater sampling volume to achieve sufficient levels of accuracy.

The specific technical details of each alternative are briefly described as follows. The GPS algorithm used in

the case studies adheres to the standard procedure firstly introduced by Hooke & Jeeves and is set according to the following details:

- (1) Input parameters are linearly scaled to the interval [0, 100], according to the optimisation bounds: this is needed since the employed mesh size is equal in all dimensions.
- (2) The mesh size is multiplied by a factor of 2 at every successful poll, and it is divided by the same factor after any unsuccessful poll.
- (3) The algorithm stops if the maximum allowed number of objective function evaluations is reached.

Simulated annealing is implemented in its most common formulation as proposed by Kirkpatrick et al. Specifically, it is set according to the following strategy:

- (1) Input parameters are linearly scaled to the interval [0, 1]
- (2) The initial temperature  $T_0$  is set at 50.
- (3) The temperature gradually decreases at each iteration according to the cooling schedule  $T = T_0/k$ , where  $k$  is a parameter equal to the iteration number.
- (4) Each new sampled point, if its objective is higher than the current one, is accepted according to the acceptance function  $\frac{1}{1 + \exp(\frac{\Delta}{\max(T)})}$ , where  $\Delta$  is the difference between the objective values (at the incumbent point and the newly sampled one).

Finally, the Genetic Algorithm implementation used is based on the following principles:

- (1) Compliance with optimisation bounds is enforced by ensuring each individual is generated within the constraints at each generation though proper crossover and mutation operators.
- (2) The initial population, necessary to initialize the algorithm, consists of points randomly chosen within the space defined by the optimisation bounds of each parameter. As the population size should increase accordingly to the number of dimensions, the initial population size was set to 200.
- (3) The choice of parents is made according to their fitness value. In particular, the chances of breeding are higher for higher fitness values.
- (4) The crossover fraction is set to 0.8.
- (5) The elite size is set at 5% of the population size.
- (6) The mutation fraction varies dynamically, according to the genetic diversity of each generation.

### 3.3 The objective function

A well-known objective function is employed, based on the difference between the estimated and the actual values of natural frequencies and on the Modal Assurance Criterion—MAC [62] between target and computed modes. This can be defined as follows:

$$P = \sum_{i=1}^N \left| \frac{\omega_i^{targ/id} - \omega_i^{calc}}{\omega_i^{targ/id}} \right| + \sum_{i=1}^N \left( 1 - \text{diag} \left( \text{MAC} \left( \phi_i^{calc}, \phi_i^{targ/id} \right) \right) \right) \tag{20}$$

where  $\omega_i^{targ/id}$  and  $\omega_i^{calc}$  are respectively the  $i$ -th target (or identified) natural angular frequency and the  $i$ -th computed natural angular frequency out of the  $N$  modes used for updating, and  $\text{MAC}(\phi_i^{calc}, \phi_i^{targ/id})$  is the MAC value relative to the  $i$ -th computed mode shape  $\phi_i^{calc}$  and the  $i$ -th target (or identified) mode shape  $\phi_i^{targ/id}$ . This objective function includes both natural frequencies and mode shapes, with equal weights.

### 3.4 Performance metrics

The root mean square relative error (RMSRE) is considered as a global metric for the accuracy of the optimisation procedure, both in the input domain (i.e., for each updating parameter) and in the output domain (i.e., for the natural frequencies and mode shapes). In this latter case, this is computed as:

$$\text{RMSRE}_{\text{output}} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n \Delta f_{\text{rel},i}^2 + \frac{1}{n} \cdot \sum_{i=1}^n (1 - \text{MAC}_i)^2} \tag{21}$$

where  $f_{\text{rel},i}$  is the relative error of the  $i$ -th natural frequency,  $\text{MAC}_i$  is the MAC value of the  $i$ -th mode shape and  $n$  is the number of modes considered for updating. In the former case, for the parameters estimated in the input space, the RMSRE is instead given by:

$$\text{RMSRE}_{\text{input}} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n \Delta X_{\text{rel},i}^2} \tag{22}$$

where in this case  $X_{\text{rel},i}$  is the relative error between the  $i$ -th updating parameter and its target value, and  $n$  is the number of updating parameters considered. Obviously, this calculation was only possible for the numerical dataset, were the inputs of the numerically-generated results are user-defined and thus known and comparable.

Finally, the total computational time is reported as well, specifically as a way to compare BO to the other algorithms—GPS, SA, and GA. This allows to prove the

time-saving advantage of using a very efficient optimisation technique such as Bayesian sampling optimisation.

#### 4 The case study: the Mirandola Bell Tower

The proposed approach has been validated on a well-known CH case study, the Mirandola bell tower, resorting to data collected from on-site surveys [63, 64]. Indeed, historical CH structures require very specific monitoring and assessment strategies; a complete overview can be found in [65].

Both experimental data and numerically-simulated data have been employed. This latter case is intended for assessing the algorithm capabilities in a more controlled fashion. The experimental validation, on the other hand, proves its feasibility for data collected from operational (output-only) Ambient Vibration (AV) tests.

The masonry-made bell tower of the Santa Maria Maggiore Cathedral in Mirandola (Emilia-Romagna, Italy) is pictured in Fig. 5.a. For this application, model updating is used as a means of structural damage assessment, as mentioned in the Introduction. The results from this procedure, carried out through the use of a Bayesian sampling optimisation approach, will be compared to the damage analysis of the bell tower structure conducted in [63]. The data analysed here originate from the same dataset, which will be discussed later.

The target structure represents an important piece of historical and cultural heritage. Built during the late fourteenth century, it underwent several structural modifications

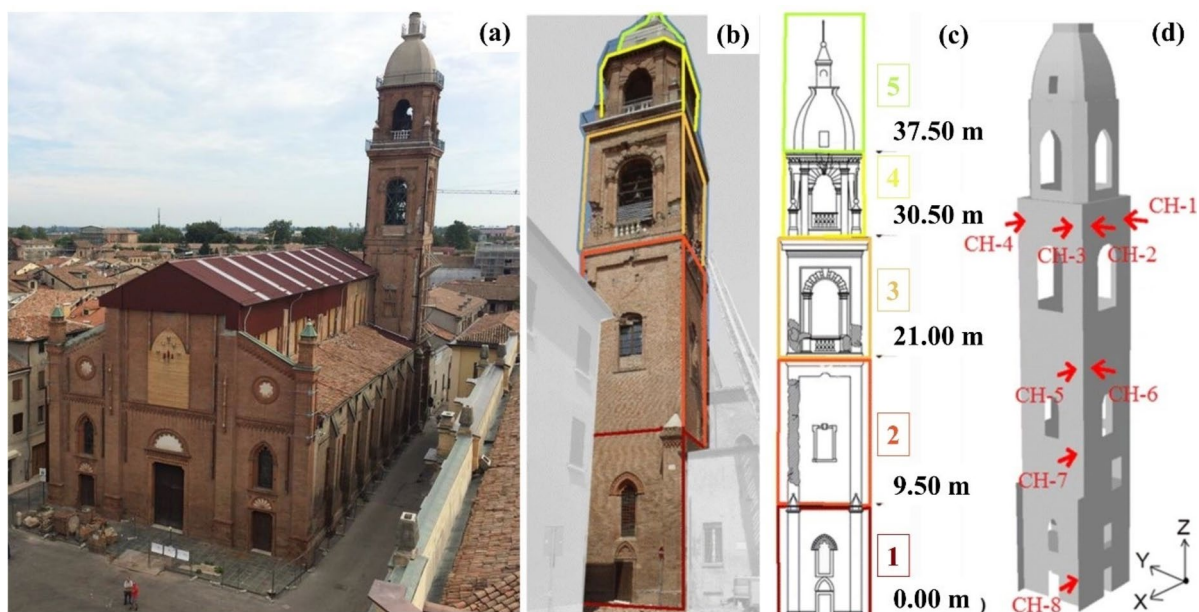
throughout the centuries, especially in the seventeenth century when the height of the original tower was tripled [66]; the existing portions were reinforced as well to withstand the additional weights. The octagonal stone roofing was finally added in the eighteenth century. From a geometric perspective, the tower has a square plan ( $5.90 \times 5.90$  m) for a total height of 48 m with four levels of openings (Fig. 5b).

Historically, the area of interest was only considered to be at modest seismic risk. However, two events linked to the 2012 Emilia Earthquake (specifically on the 20th and 29th of May) caused significant damage to the tower, as detailed in [64]. The damage pattern influenced the FE modelling as well, as will be discussed in the next subsections. More details from the visual inspection and in-situ surveys can be found in [64] and [63].

##### 4.1 Acquisition setup

The dynamic testing discussed here was performed by the laboratory of Strength of Materials (LabSCo) of the IUAV (Istituto Universitario di Venezia) in August 2012 [64], considering the post-earthquake situation before the installation of the provisional safety interventions. An Operational Modal Analysis (OMA) was conducted, relying on the structural response from ambient vibrations to identify the modal properties of the target structure [67].

Eight uniaxial piezoelectric accelerometers (PCB Piezotronics type 393C) were deployed as portrayed in Fig. 5.d, using metal bases to attach them to the masonry walls [1]. All recorded signals had a length of  $\sim 300$  s and a sampling



**Fig. 5** Pictures of the Santa Maria Maggiore Cathedral (a), the bell tower located on its south-east corner (b), the discretisation in five substructures (c), and the sensor layout with the location and orientation of the recording channels (d). Adapted from [63]

frequency  $f_s = 192$  Hz: this was then reduced to 48 Hz in post-processing, focusing on the first eight modes, to expedite the output-only identification procedure. This was done by employing the Stochastic Subspace Identification (SSI) algorithm [68] on the detrended and filtered data. However, only the first six modes (reported in Table 1) were employed as done in [63] since the 7th and the 8th ones were deemed as less reliable.

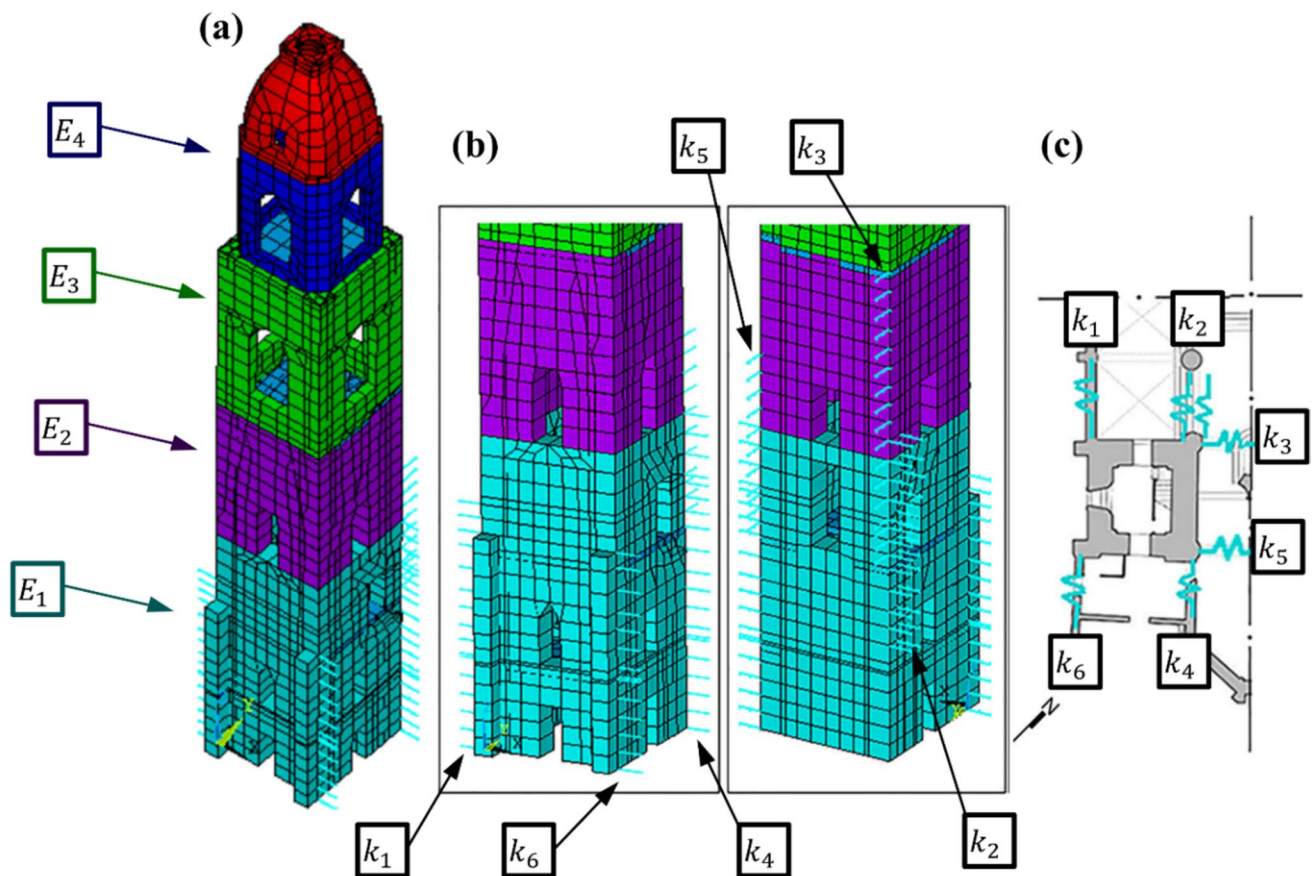
**Table 1** Identified natural frequencies (OMA)

Mode n	Mode description	$f_n$ [Hz]
1	First bending mode along the East–West direction	0.68
2	First bending mode along the North–South direction	0.72
3	First torsional mode	1.41
4	Second bending mode along the East–West direction	2.29
5	Second bending along the North–South direction	2.70
6	Second torsional mode	3.68

## 4.2 Crack pattern and finite element model

The FE model (Fig. 6) was organised in substructures (i.e. macro-areas), following the structural partition of the tower structure and reflecting the strong localisation of damage. This macro-zoning procedure reflects what is generally done for similar structures after seismic damage—see the similar case of the Fossano bell tower in [69]. The rationale is to encompass areas that are expected to show similar and homogeneous mechanical properties.

In this case, the tower base (subsection 1, highlighted in red in Fig. 5.c), i.e., the portion from ground level (0.00 m) to +9.50 m, suffered minimal to no damage, as well as the tower top (subsection 4) and the belfry (subsection 5). These are highlighted respectively in yellow, between +30.5 m and +37.5 m, and green, from +37.5 m to +48.0 m. The most severe and large damage portions were found in subsection 2 (dark orange, +9.50 m to +21.0 m) and subsection 3 (light orange, +21.0 m to +30.5 m). This latter floor represents a structural peculiarity of this case study, as it includes very large window openings on all façades. The reason is that the portion corresponds to the first belfry,



**Fig. 6** Finite Element Model of the Mirandola bell tower. **a** The whole structure **b** detail of the linear springs elements **(c)** the idealized connections with the nearby buildings, shown in plan-view. Adapted from [63]

built before the addition of the second one on top of it. In any case, this layout is quite different from the most common cases where the opening size decreases moving from the tower bottom. This affects the local and global dynamic response of the whole structure [70]. In fact, during the seismic events of interest, this locally more flexible portion underwent a twisting rotation, which caused deep diagonal cracks to arise right below this order of openings, extending down to the (less wide) openings of the underlying level on all four sides. Importantly, the second and third portions were also the ones covered by the eight metal tie rods (four per portion) installed as provisional safety interventions.

As for the previous case study, the model was realized in ANSYS Mechanical APDL. SHELL181 elements were applied for all façades in all macro-areas, as well as for the stone roof and the masonry vaults at the basement level, for a total of 1897 elements. The interactions with the Cathedral of Santa Maria Maggiore and the rectory were modelled by 104 COMBIN14 linear springs, distributed along the whole contact surface in correspondence to the apse arches, the nave walls, and the rectory wall on the East, West, and North side (Fig. 6b). That is intended to simulate the in-plane stiffness of the attached masonry walls, thus removing these external elements and replacing their reaction forces with springs' elastic forces, acting as boundary conditions as also suggested in [63].

The complete FE model consisted of 2052 nodes.

### 4.3 Model updating setup

The following eleven parameters were considered for updating (see Table 2):

- $E_1, E_2, E_3, E_4$ : the Young's modulus of the damaged masonry in the four sub-structures (according to Fig. 6).

- $\nu_{mas}$ : the Poisson's ratio of the masonry, assumed as constant everywhere.
- $k_1, k_2, k_3, k_4, k_5, k_6$ : the linear stiffness of the six distributed springs (used to model the connections with the nearby structures, see Fig. 6).

These parameters are exactly the same ones considered in [63], thus enabling a direct comparability of the results. No sensors were available on the belfry roof, therefore, for lack of reliable data, the 5th macro-area was not considered in the FE updating.

Table 2 shows these input-parameters and the assumed optimisation bounds. Notice that the optimisation range of the link-element parameters spans through several orders of magnitude, thus generating an extremely wide optimisation space. This reflects the high uncertainty about the boundary conditions, which might significantly affect the dynamic response of the structure. The optimisation bounds of the elasticity moduli consider the values suggested by literature and Italian regulations for brick masonry, while allowing to capture the level of damage suffered by the structure.

The (arbitrarily chosen) target input parameters and the related system-output parameters (in terms of frequency only) used for the numerically simulated data setup are reported in Table 3. System-output parameters used for updating the experimental data setup are the identified natural frequencies shown in Table 1 (and the related mode shapes).

**Table 2** Parameters selected for updating and associated optimisation bounds

Updating parameters	Lower bound	Upper bound
$E_1$ (GPa)	0.375	6
$E_2$ (GPa)	0.375	6
$E_3$ (GPa)	0.375	6
$E_4$ (GPa)	0.375	6
$k_1$ (kN/m)	1.00	1.00E+07
$k_2$ (kN/m)	1.00	1.00E+07
$k_3$ (kN/m)	1.00	1.00E+07
$k_4$ (kN/m)	1.00	1.00E+07
$k_5$ (kN/m)	1.00	1.00E+07
$k_6$ (kN/m)	1.00	1.00E+07
$\nu$ (–)	0.40	0.50

**Table 3** Target natural frequencies (generated by the set of target parameters)

Updating parameter	Target value	Mode $n$	Freq. <sup>TARG</sup> (Hz)
$E_1$ (GPa)	0.5	1	0.72
$E_2$ (GPa)	1.5	2	0.74
$E_3$ (GPa)	1	3	1.96
$E_4$ (GPa)	4.5	4	2.77
$k_1$ (kN/m)	4.10E+03	5	2.96
$k_2$ (kN/m)	2.00	6	4.57
$k_3$ (kN/m)	1.60E+03		
$k_4$ (kN/m)	6.30E+02		
$k_5$ (kN/m)	1.50E+06		
$k_6$ (kN/m)	1.00E+02		
$\nu$ (–)	0.45		

## 5 Results

The results for BO and each benchmark optimisation algorithm are firstly presented for the numerical dataset, drawing considerations about accuracy and computational time. These are followed by the results of the experimental datasets, which allow to assess BO performance in a real application. Furthermore, concerning BO, the use of several kernels is investigated, and the benefits of accessing parameters length scales when using ARD kernel functions in the framework of model updating are highlighted.

### 5.1 Numerically-simulated data results

For the numerical case, the initial seed size is set to 220 points (20 times the number of parameters being updated). A logarithmic transformation is employed on the input parameters, as it was observed to augment the quality of the GP regression. The choice of the kernel function is driven by a cross-validation test, since some kernels may happen to be more suitable at modelling the underlying objective function specific to this updating problem, resulting in surrogate models with enhanced validity. The outcome of a tenfold cross-validation test of four different Gaussian Processes, fitted using the ARD exponential, the ARD Matérn 3/2, the ARD Matérn 5/2 and the ARD squared exponential kernels are shown in Table 4. Whilst all kernel functions are seen to return reliable regression models, the ARD Matérn 5/2 kernel is found to be the most suitable, returning excellent validation results. Fitting the GP to the initial seed also allows retrieving information on system sensitivity through the optimized hyperparameters of the selected ARD kernel (Table 5).

As expected, the eleven length scales differ by several orders of magnitude due to the anisotropy of the problem. The parameters which mostly affect the system response are the elasticity moduli: such behaviour is foreseeable, as material elasticity is usually blamed for significantly impact modal properties. However, kernel hyperparameters suggest that the elasticity modulus of the fourth sub-structure has instead low sensitivity: this is likely due to the scarcity of sensors at that level of the building, which actually disqualifies from capturing the necessary vibration information. Therefore, reliable estimations of  $E_4$  should not be expected,

**Table 4** The root mean square error (RMSE) is reported for each kernel, as the average of the validation losses obtained with tenfold cross-validation tests

	ARD Exp	ARD Matérn 3/2	ARD Matérn 5/2	ARD squared Exp
<b>RMSE (–)</b>	0.104	0.090	0.066	0.101

**Table 5** Parameters' length-scales (hyperparameters) obtained by maximisation of the marginal log-likelihood

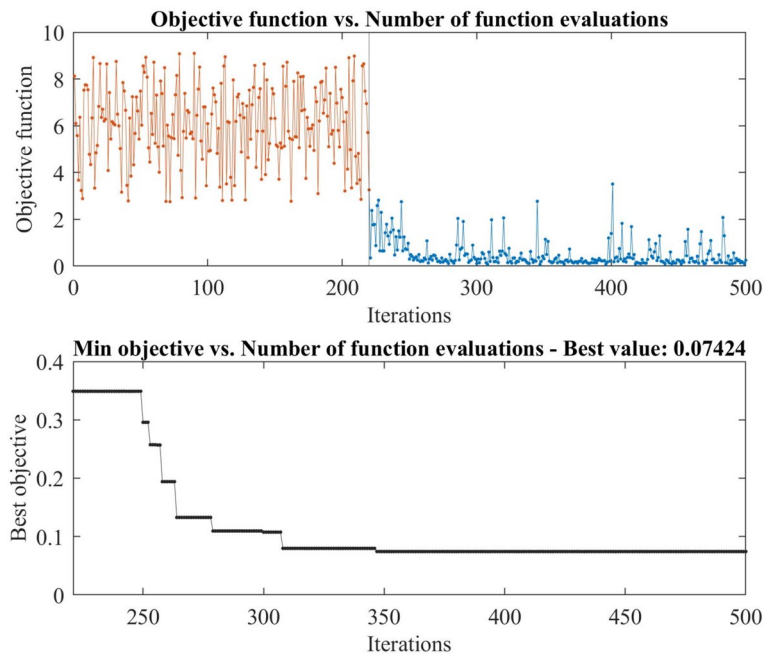
	Hyperparameter (–)
$\vartheta(E_1)$	3.59
$\vartheta(E_2)$	3.58
$\vartheta(E_3)$	5.76
$\vartheta(E_4)$	4.57E+05
$\vartheta(k_1)$	61.91
$\vartheta(k_2)$	216.27
$\vartheta(k_3)$	45.47
$\vartheta(k_4)$	27.64
$\vartheta(k_5)$	4.35E+05
$\vartheta(k_6)$	3.32E+05
$\vartheta(\nu)$	48.55

neither in the simulated-data setup, nor in the experimental one. The problem appears to be scarcely sensitive to changes in the Poisson's ratio as well: this makes perfect sense as damping has only marginal effect on modal properties. For what concerns the springs, these are generally found to have lesser effects on the modal response compared to the elasticity moduli of the first three sub-structures. In particular, as hyperparameters suggest,  $k_5$  and  $k_6$  are found to have lower impact on the system modal response, while  $k_1$ ,  $k_3$  and  $k_4$  feature higher sensitivity.

The Bayesian sampling optimisation is carried out using the upper confidence bound (UCB) acquisition function, which was generally found to be the most effective, as it provided a good balance between exploitation and exploration. In Fig. 7 (at the top) it is clear how the first selected sampling point already represents a massive improvement over the best cost function value, suggesting that the Gaussian Process is able to model the objective function impressively well. As the GP is updated with newly sampled points, UCB steadily converges towards a minimum, gradually improving the accuracy of the optimisation solution. For additional clarity, the best-computed objective against the iteration number is shown at the bottom. The best-obtained objective is 0.0742, which is fairly close to zero, the (known) global optimum in the output space. Also, it is noticeable how BO struggles at further improving the result after 350 iterations in this case.

The results relative to the output, that are the raw (best) cost function value and the estimated modal data, are shown in Table 6 for each optimisation technique. For all modes, the updated value and the target value are reported, as well as the relative error and the MAC value. Also, the RMSRE and the best achieved objective function value are reported for each algorithm. These results highlight how GPS, SA and GA fail at minimising the cost function under the proposed conditions: all algorithms, and especially SA, return objective values that are far from

**Fig. 7** Top: Bayesian optimisation progress over iterations for the numerical dataset. The objective value at the randomly sampled seed points is displayed in red, while the objective at points selected by UCB is displayed in blue. Bottom: best objective function value over iterations



**Table 6** Optimisation results in the output space, obtained with GPS, Simulated annealing, Genetic Algorithm and Bayesian optimisation

Mode $n$	Freq. $^{\text{TARG}}$ (Hz)	Generalized pattern search 1000 fun. eval			Simulated annealing 1000 fun. eval		
		Freq. $^{\text{UPD}}$ (Hz)	Freq. err (%)	MAC (-)	Freq. $^{\text{UPD}}$ (Hz)	Freq. err (%)	MAC (-)
1	0.72	0.74	3.17	0.86	0.70	-3.06	0.25
2	0.74	0.90	21.26	0.84	1.01	35.63	0.20
3	1.96	2.09	6.36	0.98	1.89	-3.54	0.99
4	2.77	2.47	-10.58	0.93	2.52	-8.75	0.02
5	2.96	2.96	0.26	0.35	3.01	1.76	0.41
6	4.57	4.24	-7.19	0.70	3.84	-15.86	0.23
RMSRE (%)		22.89			52.25		
Best cost function value		0.9472			2.4817		
Mode $n$	Freq. $^{\text{TARG}}$ (Hz)	Genetic Algorithm 1000 fun. eval			Bayesian optimisation 500 fun. eval		
		Freq. $^{\text{UPD}}$ (Hz)	Freq. err (%)	MAC (-)	Freq. $^{\text{UPD}}$ (Hz)	Freq. err (%)	MAC (-)
1	0.72	0.56	-22.35	0.86	0.73	0.87	0.99
2	0.74	0.75	0.23	0.82	0.78	4.28	0.99
3	1.96	1.90	-2.99	0.98	1.96	-0.41	1.00
4	2.77	2.24	-18.92	0.99	2.82	1.89	0.99
5	2.96	2.90	-2.03	0.27	2.98	0.68	0.98
6	4.57	3.92	-14.15	0.80	4.62	1.26	0.99
RMSRE (%)		29.69			1.63		
Best cost function value		0.9436			0.0742		

zero. Bayesian optimisation, on the contrary, at only 500 evaluations, achieves quite impressive results. The error about the frequencies is kept to a minimum (here, only

the second mode is showing a higher divergence), and so is the error about the mode shapes. The relative errors are computed, for each  $n$ -th mode, as  $(f_n^{\text{UPD}} - f_n^{\text{TARG}})/f_n^{\text{TARG}}$ ,

**Table 7** Optimisation results in the input space, obtained with GPS, SA, GA and BO

Parameters	Target value	Updating values			
		GPS 1000 obs	SA 1000 obs	GA 1000 obs	BO 500 obs
$E_1$ (GPa)	0.50	0.83	0.38	0.37	0.63
$E_2$ (GPa)	1.50	0.83	0.58	0.85	1.53
$E_3$ (GPa)	1.00	2.63	5.54	2.01	0.82
$E_4$ (GPa)	4.50	1.50	5.79	3.76	3.97
$k_1$ (kN/m)	4.10E+03	4.00E+05	1.00E+07	9.21E+06	1.89E+01
$k_2$ (kN/m)	2.00	4.00E+05	8.45E+05	1.00	5.62E+03
$k_3$ (kN/m)	1.60E+03	1.00	2.07E+03	1.00	1.48E+03
$k_4$ (kN/m)	6.30E+02	1.00	9.98E+06	1.00	1.01E+01
$k_5$ (kN/m)	1.50E+06	1.20E+06	9.98E+06	5.53E+06	1.75E+06
$k_6$ (kN/m)	1.00E+02	1.00E+07	3.41E+05	1.00	8.18E+04
$\nu$ (-)	0.45	0.421	0.477	0.430	0.465
RMSRE (%)		96.31	229.62	72.67	16.79

where  $f_n^{\text{UPD}}$  is the updated value and  $f_n^{\text{TAR}}$  is the corresponding target.

The resulting updated parameters (that generate the updated modal features just discussed), are displayed in Table 7. The updated value, the target value and the RMSRE value (about the four elasticity moduli) are reported. Up to 1000 observations are used for the three former algorithms, while BO employs 500 observations only, leading to comparable total optimization times. Among all parameters, the most interesting are the elasticity moduli of the four sub-structures as these parameters respond to the main goal of the updating problem, that is assessing the level of damage to the structure. The stiffnesses of the six springs are of secondary interest since these are introduced in the updating procedure only due to unawareness of the degree of support provided by the adjacent architectural elements as well as of their impact on the dynamic response of the building. Moreover, given their extremely wide optimisation range, identifying the right order of magnitude can already be considered a satisfactory result. In light of the above, GPS, SA and GA show quite poor results, failing to attain a good estimation of the first four parameters (and providing even worse estimations for the rest). On the contrary, Bayesian optimisation returns acceptable errors over the estimated values, showing good agreement with the first four in particular. Although a ~25% error in the elasticity moduli may appear significant, these results are in fact remarkable given the intricate nature of the optimization problem under consideration. Indeed, this level of precision allows for informed assessments regarding which portions of the structure likely experienced the most damage and which sections remain structurally sound. The scale of the springs is in some cases recognized as well, except for

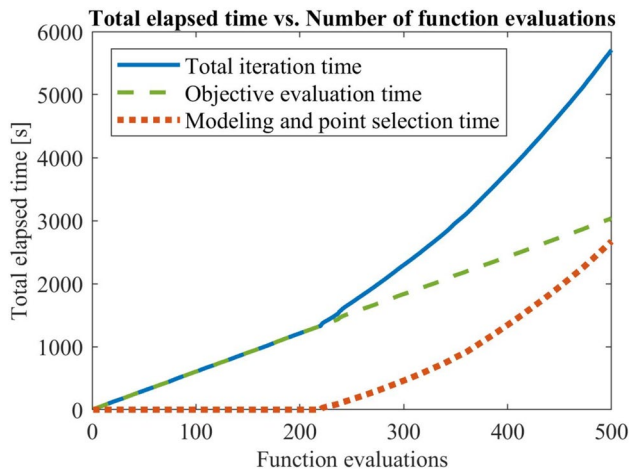
$k_2$  and  $k_6$  (and to a lower extent  $k_1$  and  $k_4$ ), suggesting once more that the algorithm could have run into a local minimum. Potentially, had Bayesian optimisation been able to properly sort out the right scale of  $k_2$  and  $k_6$ , it would have then returned even better accuracy about the parameters of interest (i.e.,  $E_1, E_2, E_3$  and  $E_4$ ).

In the damage assessment study of 2017, the updating procedure was carried out in batches: at first, the springs were calibrated while holding the elasticity moduli constant, afterwards,  $E_1, E_2, E_3$  and  $E_4$  were optimised using the link-element stiffness values previously estimated. As traditional optimisation techniques were employed, this approach aimed at facilitating the optimisation procedure by lowering the dimensionality of the problem. Given the outcome of this numerical test, such an approach can be avoided when using BO, as this technique is powerful enough to allow considering all parameters at once, cutting off computational time and enhancing chances of ending up close enough to the global minimum.

The optimisation time employed by each algorithm is reported in Table 8. As the number of observations is relatively high, the secondary optimisation problems (i.e., the maximisation of the marginal log-likelihood and the

**Table 8** The table reports the elapsed optimisation time for each technique. The sampling volume comprises 500 observations for BO and 1000 observations for GPS, SA, and GA. Minor variations in computational time required to calculate the objective across algorithms stem from differences in computer-related computational performance among different runs

	GPS	SA	GA	BO
Time for modelling and point selection (s)	–	–	–	2672
Time for computation of the objective (s)	6097	6116	6083	3036
Total optimisation time (s)	6118	6213	6237	5708



**Fig. 8** Total optimisation time of BO, as the sum of the objective evaluation time (the cumulative time employed for evaluating the objective function) and the modelling and point selection time (the cumulative time employed for maximising the marginal log-likelihood and the acquisition function)

maximisation of the acquisition function) are relatively burdensome tasks, significantly extending the total optimisation time (Fig. 8). As computing a prediction has a computational complexity of  $\mathcal{O}(N^3)$ , the modelling and point selection time gradually increases as the number of observations grows. These two tasks are crucial to obtain high sampling efficiency, which in turn allows for keeping the total number of observations to a minimum. In fact, with 500 iterations only (actually, in this case, about 350 would have been sufficient to obtain the same final results), BO still enables saving some computational time when compared to other techniques, even if negatively affected by increasingly burdensome secondary optimization problems, while retaining far superior accuracy.

## 5.2 Experimental data results

Differently from numerically simulated data, the experimental case study is affected by both the implicit limitations of the FE model and the measurement noise of the acquisitions. Thus, the minimum of the cost function is never found at zero when using real data, since significant misfit between experimental and computed modal properties will endure.

The same kernel and acquisition function of the numerical case are used, while the seed size is set to 250 points. For a given number of total observations (in this case 500), using larger seeds leads to shorter computational time (as the actual number of algorithm iterations is reduced, meaning less time is spent on secondary optimisation problems), potentially improved surrogate quality, at the expense of reduced exploitation of the algorithm's "intelligent" sampling capabilities. In a way, we rely more on the surrogate

**Table 9** Parameters' length scales obtained by maximising the marginal log-likelihood for the experimental dataset. GP fitted on 250 seed points

Parameter	Length scale (–)
$E_1$	2.69
$E_2$	1.96
$E_3$	16.19
$E_4$	1.74E+05
$k_1$	76.69
$k_2$	1.44E+03
$k_3$	40.14
$k_4$	33.44
$k_5$	3.73E+05
$k_6$	1.26E+06
$\nu$	23.31

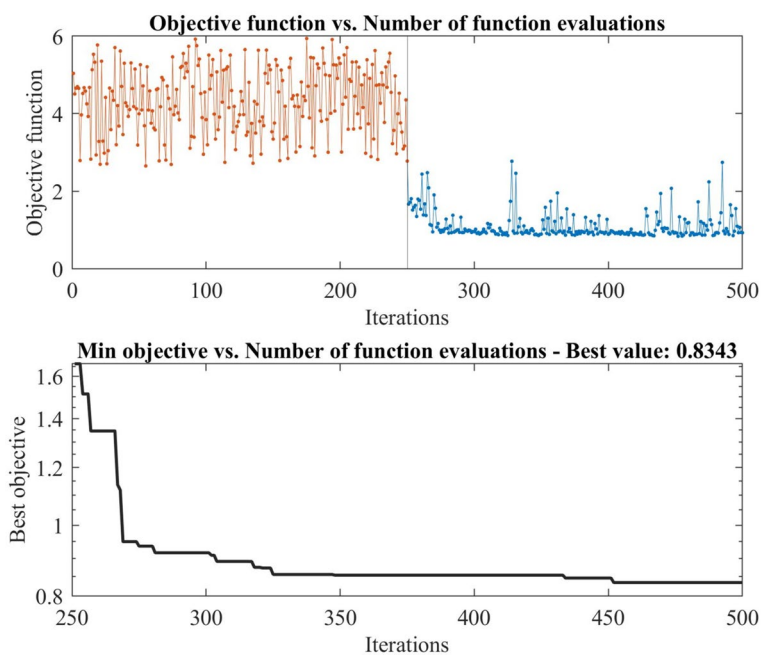
represented by the GP and less on the point-selection process operated by the acquisition function.

Table 9 provides the length-scales obtained when fitting the GP to the new 250-point seed. As different observations are here employed, hyperparameters (only) marginally differ from the ones obtained for the numerically simulated dataset. It is noticeable a substantial correlation between the two. As such, previous considerations about length scales apply in this case as well.

The results obtained through the optimisation process (shown in Fig. 9) reveal how in this case the optimiser cannot be expected to converge at zero, since FE modelling deficiencies coupled with identified modal data inaccuracies lead to the ineluctable misfit between computed and measured modal response. For additional clarity, the best-computed objective against the iteration number is shown as well. The value associated with the function minimum is 0.8343.

The results relative to the output, that are the raw (best) cost function value and the estimated modal data, are shown in Table 10, along with the modal parameters attained in the paper of reference. Once again, for each considered mode, the updated and the identified values are reported, as well as the relative error and the MAC value. Looking at the updated modal parameters, Bayesian optimisation gives results that are consistent with what was obtained in the 2017 study. Generally, natural frequencies obtained through the Bayesian sampling optimisation show a good correlation with the identified ones. The modes that exhibit the highest amount of error are the third and the fifth: this was already the case in the former study. The MAC values of the first three modes suggest a good correlation with the identified mode shapes of the tower, while the last three denote some degree of incoherence (especially for the fifth and the sixth). This issue is in common with the study made in 2017, indicating some problems probably due to the quality of the measurements at the highest modes or the inadequacy of the FE model to capture the dynamic behaviour of the bell tower at higher

**Fig. 9** Top: Bayesian optimisation progress over iterations for the experimental dataset. Once again, the objective value at the randomly sampled seed points is displayed in red, while the objective value at points selected by UCB is displayed in blue. Bottom: in best objective function value over iterations



**Table 10** Optimisation results in the output space, obtained with Bayesian optimisation. Results from 2017 are reported for reference. Also, the RMSRE is computed for both BO and the former paper

Mode n	Freq. <sup>ID</sup> (Hz)	Bayesian optimisation (UCB)			Original paper –Ref. (2017)		
		Freq. <sup>UPD</sup> (Hz)	Freq. error (%)	MAC (–)	Freq. <sup>UPD</sup> (Hz)	Freq. error (%)	MAC (–)
1	0.68	0.67	1.75	0.98	0.68	0.20	0.98
2	0.72	0.71	1.17	0.96	0.69	3.80	0.98
3	1.41	1.64	16.22	0.87	1.73	18.30	0.91
4	2.29	2.28	0.46	0.78	2.33	1.60	0.68
5	2.70	2.43	9.97	0.48	2.42	11.80	0.66
6	3.68	3.79	2.99	0.59	3.80	3.30	0.35
RMSRE (%)		21.29			24.15		

frequencies. Indeed, due to these reasons, fitting the system response with modal features beyond the third mode is often unpractical in many FEMU applications [71].

The results relative to the input space (i.e., the estimated parameters) are reported in Table 11, as the estimated value obtained through BO and the estimations of the former damage assessment study. Focusing on the elasticity moduli of the four sub-structures, results stemming from the Bayesian sampling optimisation approach are mostly compliant with the former study, except for the elasticity modulus of the third sub-structure.

These findings can be used to assess the damage condition of the bell tower. The low values obtained for the two lowest levels suggest that the bell tower has probably endured high levels of damage in these areas, which mostly affect the lower modes of the building. Concerning these two subparts, a slightly higher estimate of the second elasticity modulus is the only marginal difference between the

**Table 11** Optimisation results in the input space, obtained with Bayesian optimisation

Parameter	Bayesian optimisation	Original paper [63]
	Estimated value	Estimated value
$E_1$ (GPa)	0.726	0.508
$E_2$ (GPa)	0.792	0.806
$E_3$ (GPa)	1.908	0.850
$E_4$ (GPa)	3.098	4.500
$k_1$ (kN/m)	5.34	4.13E+03
$k_2$ (kN/m)	4.60E+01	0.51
$k_3$ (kN/m)	1.11E+01	1.60E+03
$k_4$ (kN/m)	5.81	6.34E+02
$k_5$ (kN/m)	6.25E+04	1.50E+06
$k_6$ (kN/m)	8.60	1.06E+02
$\nu$ (–)	0.46	N.A

two results. On the contrary, a substantial difference can be seen for the elasticity of the walls at the third level of the structure: the former study highlighted a much higher level of damage. Judging by the value estimated through BO, this specific part of the tower could have been either less damaged by the seismic event, or originally characterized by walls built with stiffer and more qualitative material. This hypothesis is made more plausible by considering that the building, which construction started in the late fourteenth century, was severally altered in the seventeenth century, when the height of the bell tower was tripled and the original structure reinforced [63]. Finally, the walls of the fourth level were found to be significantly stiffer than the rest of the structure. However, estimates of parameters concerning this level of the structure should not be considered as reliable for the reasons stated before (scarcity of sensors).

Regarding the springs elements which model the degree of constrain enforced by the adjacent structures, the results of the BO approach agree with the estimations of the former analysis, particularly for what concerns the value of  $k_3$ , which stands out in both cases with respect to the other link-element parameters by a factor of  $10^3$ . The greater stiffness of the fifth spring element suggests that the architectural element having the greatest impact on the dynamic response of the bell tower is the easternmost apse arch. All the other elements (particularly the ones modelled by  $k_2$ ,  $k_4$  and  $k_6$ , hence the nave walls and the rectory wall) seem to provide little contrast to the motion of the bell tower when considering small-amplitude environmental vibrations.

## 6 Conclusions

When expensive FE models are involved, the optimization algorithm represents one of the most critical aspects of FEMU applications that make use of iterative methods. The optimisation technique is a key element of the updating procedure, as it should feature good sampling efficiency, global search attitudes and adequate accuracy to cope with non-convex and complex cost functions.

This research presented and validated a Bayesian sampling optimisation (BO) approach for such a task, with an application to a real case study—the Mirandola bell tower—that represents an interesting example of post-seismic assessment of a historical building of cultural and architectural relevance. Overall, the proposed procedure proved itself to be well-suited for this challenging task. Especially, BO outperformed the other well-established global optimisation techniques selected for the benchmark (namely Generalized Pattern Search, Simulated annealing and Genetic Algorithm), featuring far superior sampling efficiency, greater accuracy, and better capabilities

of finding the global function minimum, particularly as the dimensionality of the problem increases. Results were achieved with half of the objective function evaluations allowed for GPS, SA, and GA. This, practically, translates into shorter computational times and costs.

One major drawback of SA and GA is that both techniques rely on large sampling volumes. Furthermore, these algorithms tend to provide results denoted by poor accuracy. The GPS algorithm, on the contrary, may exhibit scarce global search aptitudes, as it was either found converging too quickly or failing at finding the global function optimum. The proposed BO algorithm is not affected by any of these drawbacks.

Regarding the implementation of the described technique, a logarithmic transformation of the input variables was found to improve the quality of the Gaussian Process regression, albeit marginally. Furthermore, this research revealed that all four investigated kernels can be successfully used, although the ARD Matérn 5/2 kernel provided the best results in terms of validation of the surrogate model for this specific case-study. When using automatic relevance determination (ARD) kernels, it is possible to retrieve hyperparameters (i.e., length scales) of the GP, gathering useful information about the problem sensitivity to each parameter. Additionally, ARD kernels automatically discard irrelevant dimensions from the optimization procedure, making this approach well-suited for highly anisotropic problems. This sort of “built-in” sensitivity analysis is particularly advantageous in structural model updating applications, where information about the relevance of parameters (often scarcely known in advance) can be considerably useful for a better understanding of the structural behaviour.

**Acknowledgements** The authors would like to thank the Laboratory of Strength of Materials (LabSCo) of IUAV University of Venice for the acquisition, processing, and dynamic identification of the case study.

**Author contributions** Conceptualization: M.C. and L.Z.F.; Methodology: D.R., M.C., and L.Z.F.; Software: D.R., M.C., and L.Z.F.; Validation: D.R.; Formal Analysis: D.R. and M.C.; Investigation: D.R., M.C., and L.Z.F.; Resources: L.Z.F.; Data Curation: D.R., M.C., and L.Z.F.; Writing—Original Draft: D.R.; Writing—Review & Editing: M.C. and L.Z.F.; Visualization: D.R. and M.C.; Supervision: M.C. and L.Z.F.; Project administration: L.Z.F.

**Funding** Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement. This research did not receive external fundings.

**Availability of data and materials** Data available upon reasonable request from the authors.

## Declarations

**Conflict of interest** There are no conflicts of interest.

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

## References

- Boscatto G, Dal Cin A, Ientile S, Russo S (2016) Optimized procedures and strategies for the dynamic monitoring of historical structures. *J Civ Struct Health Monit* 6(2):265–289. <https://doi.org/10.1007/S13349-016-0164-9>
- Friswell MI, Mottershead JE (1995) Finite element model updating in structural dynamics, vol 38. Springer, Netherlands, Dordrecht. <https://doi.org/10.1007/978-94-015-8508-8>
- Friswell MI (2007) Damage identification using inverse methods. *Philos Trans R Soc A Math Phys Eng Sci* 365(1851):393–410. <https://doi.org/10.1098/rsta.2006.1930>
- Durmazgezer E, Yucel U, Ozelik O (2019) Damage identification of a reinforced concrete frame at increasing damage levels by sensitivity-based finite element model updating. *Bull Earthq Eng* 17(11):6041–6060. <https://doi.org/10.1007/s10518-019-00690-5>
- Mohagheghian K, Karami Mohammadi R (2017) Comparison of online model updating methods in pseudo-dynamic hybrid simulations of TADAS frames. *Bull Earthq Eng* 15(10):4453–4474. <https://doi.org/10.1007/s10518-017-0147-1>
- Miraglia G, Petrovic M, Abbiati G, Mojsilovic N, Stojadinovic B (2020) A model-order reduction framework for hybrid simulation based on component-mode synthesis. *Earthq Eng Struct Dyn* 49(8):737–753. <https://doi.org/10.1002/EQE.3262>
- Mottershead JE, Friswell MI (1993) Model updating in structural dynamics: a survey. *J Sound Vib* 167(2):347–375. <https://doi.org/10.1006/JSVI.1993.1340>
- Mottershead JE, Link M, Friswell MI (2011) The sensitivity method in finite element model updating: a tutorial. *Mech Syst Signal Process* 25(7):2275–2296. <https://doi.org/10.1016/j.ymssp.2010.10.012>
- Brochu E, Cora VM, de Freitas N (2010) A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. <https://doi.org/10.48550/arXiv.1012.2599>
- Ceravolo R, Pistone G, Fragonara LZ, Massetto S, Abbiati G (2016) Vibration-based monitoring and diagnosis of cultural heritage: a methodological discussion in three examples. *Int J Arch Heritage* 10(4):375–395. <https://doi.org/10.1080/15583058.2013.850554>
- Boscatto G, Russo S, Ceravolo R, Fragonara LZ (2015) Global sensitivity-based model updating for heritage structures. *Comput-Aided Civ Infrastruct Eng* 30(8):620–635. <https://doi.org/10.1111/mice.12138>
- Bracchi S, Rota M, Magenes G, Penna A (2016) Seismic assessment of masonry buildings accounting for limited knowledge on materials by Bayesian updating. *Bull Earthq Eng* 14(8):2273–2297. <https://doi.org/10.1007/s10518-016-9905-8>
- Secanell R, Martin C, Viallet E, Senfaute G (2018) A Bayesian methodology to update the probabilistic seismic hazard assessment. *Bull Earthq Eng* 16(6):2513–2527. <https://doi.org/10.1007/s10518-017-0137-3>
- Monchetti S, Viscardi C, Betti M, Bartoli G (2022) Bayesian-based model updating using natural frequency data for historic masonry towers. *Probab Eng Mech* 70:103337. <https://doi.org/10.1016/J.PROBENGMECH.2022.103337>
- Ponsi F, Bassoli E, Vincenzi L (2022) Bayesian and deterministic surrogate-assisted approaches for model updating of historical masonry towers. *J Civ Struct Health Monit* 12(6):1469–1492. <https://doi.org/10.1007/s13349-022-00594-0>
- Monchetti S, Viscardi C, Betti M, Clementi F (2023) Comparison between Bayesian updating and approximate Bayesian computation for model identification of masonry towers through dynamic data. *Bull Earthq Eng*. <https://doi.org/10.1007/S10518-023-01670-6>
- Ferraioli M, Miccoli L, Abruzzese D (2018) Dynamic characterisation of a historic bell-tower using a sensitivity-based technique for model tuning. *J Civ Struct Health Monit* 8(2):253–269. <https://doi.org/10.1007/S13349-018-0272-9>
- Civera M, Ferraris M, Ceravolo R, Surace C, Betti R (2019) The Teager-Kaiser energy cepstral coefficients as an effective structural health monitoring tool. *Appl Sci* 9(23):5064. <https://doi.org/10.3390/app9235064>
- Ceravolo R, de Lucia G, Lenticchia E, Miraglia G (2019) Seismic structural health monitoring of cultural heritage structures. *Springer Tracts Civ Eng* 11:51–85. [https://doi.org/10.1007/978-3-030-13976-6\\_3](https://doi.org/10.1007/978-3-030-13976-6_3)
- Gentile C, Saisi A (2007) Ambient vibration testing of historic masonry towers for structural identification and damage assessment. *Constr Build Mater* 21(6):1311–1321. <https://doi.org/10.1016/J.CONBUILDMAT.2006.01.007>
- De Stefano A, Ceravolo R (2007) Assessing the health state of ancient structures: the role of vibrational tests. *J Intell Mater Syst Struct* 18(8):793–807. <https://doi.org/10.1177/1045389X06074610>
- Standoli G, Giordano E, Milani G, Clementi F (2021) Model updating of historical bellfries based on oma identification techniques. *Int J Arch Heritage* 15(1):132–156. <https://doi.org/10.1080/15583058.2020.1723735>
- Ceravolo R, de Lucia G, Miraglia G, Pecorelli ML (2020) Thermoelastic finite element model updating with application to monumental buildings. *Comput-Aided Civ Infrastruct Eng* 35(6):628–642. <https://doi.org/10.1111/mice.12516>
- Ozelik O, Misir IS, Yucel U, Durmazgezer E, Yucel G, Amadeo C (2022) Model updating of Masonry courtyard walls of the historical Isabey mosque using ambient vibration measurements. *J Civ Struct Health Monit* 12(5):1157–1172. <https://doi.org/10.1007/s13349-022-00610-3>
- Rossi M, Bournas D (2023) Structural health monitoring and management of cultural heritage structures: a state-of-the-art review. *Appl Sci* 13(11):6450. <https://doi.org/10.3390/APP13116450>
- Cartiaux F, Olivetti G, Fort V, Pelletier P (2022, Preserving the heritage of world's monuments through Structural Health Monitoring. A case study: the Garisenda Tower. In: 5th Joint International Symposium on Deformation Monitoring (JISDM 2022)
- Standoli G, Salachoris GP, Masciotta MG, Clementi F (2021) Modal-based FE model updating via genetic algorithms: Exploiting artificial intelligence to build realistic numerical models of historical structures. *Constr Build Mater* 303:124393. <https://doi.org/10.1016/J.CONBUILDMAT.2021.124393>
- Salachoris GP, Standoli G, Betti M, Milani G, Clementi F (2023) Evolutionary numerical model for cultural heritage structures via genetic algorithms: a case study in central Italy. *Bull Earthq Engg*. <https://doi.org/10.1007/S10518-023-01615-Z>

29. Fox RL, Kapoor MP (1968) Rates of change of eigenvalues and eigenvectors. *AIAA J* 6(12):2426–2429. <https://doi.org/10.2514/3.5008>
30. Chen JC, Garba JA (1980) Analytical model improvement using modal test results. *AIAA J* 18(6):684–690. <https://doi.org/10.2514/3.50805>
31. Alvin KF (1997) Finite element model update via Bayesian estimation and minimization of dynamic residuals. *AIAA J* 35(5):879–886. <https://doi.org/10.2514/2.7462>
32. Zimmerman DC, Kaouk M (1992) Eigenstructure assignment approach for structural damage detection. *AIAA J* 30(7):1848–1855. <https://doi.org/10.2514/3.11146>
33. Datta BN (2002) Finite-element model updating, eigenstructure assignment and eigenvalue embedding techniques for vibrating systems. *Mech Syst Signal Process* 16(1):83–96. <https://doi.org/10.1006/mssp.2001.1443>
34. Simoen E, de Roeck G, Lombaert G (2015) Dealing with uncertainty in model updating for damage assessment: a review. *Mech Syst Signal Process* 56–57:123–149. <https://doi.org/10.1016/j.ymssp.2014.11.001>
35. Levin RI, Lieven NAJ (1998) Dynamic finite element model updating using simulated annealing and genetic algorithms. *Mech Syst Signal Process* 12(1):91–120. <https://doi.org/10.1006/mssp.1996.0136>
36. Wang DJ, Tan ZC, Li Y, Liu Y (2013) Review of the application of finite element model updating to civil structures. *Key Eng Mater* 574:107–115. <https://doi.org/10.4028/www.scientific.net/KEM.574.107>
37. Mockus J (1994) Application of Bayesian approach to numerical methods of global and stochastic optimization. *J Global Optim* 4(4):347–365. <https://doi.org/10.1007/BF01099263>
38. Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J Global Optim* 13(4):455–492. <https://doi.org/10.1023/A:1008306431147>
39. Streltsov S, Vakili P (1999) A non-myopic utility function for statistical global optimization algorithms. *J Global Optim* 14(3):283–298. <https://doi.org/10.1023/A:1008284229931>
40. Jones DR (2001) A taxonomy of global optimization methods based on response surfaces. *J Global Optim* 21(4):345–383. <https://doi.org/10.1023/A:1012771025575>
41. Sasena MJ (2002) Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations [Doctoral dissertation, University of Michigan]. <https://hdl.handle.net/2027.42/132844>
42. Friswell MI, Penny JE (1992) Assessing model quality in parameter updating procedures. In: 10th International Modal Analysis Conference, San Diego, California, USA, 1992, pp 188–194
43. Ahmadian H, Gladwell GM, Ismail F (1997) Parameter selection strategies in finite element model updating. *J Vib Acoust* 119(1):37–45
44. Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J Glob Optim* 13(4):455–492. <https://doi.org/10.1023/A:1008306431147>
45. Kushner HJ (1964) A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *J Basic Eng* 86(1):97–106. <https://doi.org/10.1115/1.3653121>
46. Mockus J, Tiesis V, Zilinskas A (1978) The application of Bayesian methods for seeking the extremum. *Towards Glob Optim* 2:117–129
47. Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. *Stat Sci*. <https://doi.org/10.1214/ss/1177012413>
48. Shahriari B, Swersky K, Wang Z, Adams RP, de Freitas N (2016) Taking the human out of the loop: a review of Bayesian optimization. *Proc IEEE* 104(1):148–175. <https://doi.org/10.1109/JPROC.2015.2494218>
49. Mockus J (1982) The Bayesian approach to global optimization. In: Drenick RF, Kozin F (eds) *System modeling and optimization*, vol 38. Springer Berlin Heidelberg, Berlin, pp 473–481. <https://doi.org/10.1007/BFb0006119>
50. Hutter F, Hoos HH, Leyton-Brown K (2011) Sequential model-based optimization for general algorithm configuration, pp 507–523. [https://doi.org/10.1007/978-3-642-25566-3\\_40](https://doi.org/10.1007/978-3-642-25566-3_40)
51. Snoek J et al (2015) Scalable Bayesian optimization using deep neural networks. *arXiv*. <https://doi.org/10.48550/ARXIV.1502.05700>
52. Springenberg JT, Klein A, S, Hutter F (2016) Bayesian optimization with robust Bayesian neural networks. In: Lee D, Sugiyama M, Luxburg U, Guyon I and Garnett R (eds) *Advances in neural information processing systems*, pp 4134–4142
53. Wang Z, Gehring C, Kohli P, Jegelka S (2018) Batched large-scale Bayesian optimization in high-dimensional spaces. In: *International Conference on artificial intelligence and statistics (AISTATS)*, pp 745–754
54. Rasmussen C (2004) Gaussian processes in machine learning. In: Bousquet O, Luxburg U, Rätsch G (eds) *Advanced lectures on machine learning*. Springer Berlin, Heidelberg, pp 63–71. <https://doi.org/10.1007/b100712>
55. Rasmussen CE, Williams CKI (2005) *Gaussian processes for machine learning*. The MIT Press. <https://doi.org/10.7551/mitpress/3206.001.0001>
56. Matérn B (1986) *Spatial variation*, vol 36. Springer, New York, New York. <https://doi.org/10.1007/978-1-4615-7892-5>
57. Bull AD (2011) Convergence rates of efficient global optimization algorithms. *J Mach Learn Res* 12:2879–2904
58. Cox DD, John S A statistical method for global optimization. In: *Proc. IEEE International Conference on systems, man, and cybernetics, IEEE*, pp 1241–1246. <https://doi.org/10.1109/ICSMC.1992.271617>
59. Hooke R, Jeeves TA (1961) ‘Direct Search’ solution of numerical and statistical problems. *J ACM* 8(2):212–219
60. Holland J (1992) *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor
61. Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. *Science* (1979) 220(4598):671–680. <https://doi.org/10.1126/science.220.4598.671>
62. Allemang RJ, Brown DL (1982) A correlation coefficient for modal vector analysis. In: *Proceedings of the 1st International Modal Analysis Conference, Orlando, 8-10 November 1982*, pp 110–116
63. Zanotti Fragonara L et al (2017) Dynamic investigation on the Mirandola bell tower in post-earthquake scenarios. *Bull Earthq Eng* 15(1):313–337. <https://doi.org/10.1007/s10518-016-9970-z>
64. Boscatto G, Dal Cin A, Russo S (2015) Collapse mechanisms due to earthquake in the structural typologies of historic constructions: the case of Mirandola. *Key Eng Mater* 624:59–65. <https://doi.org/10.4028/WWW.SCIENTIFIC.NET/KEM.624.59>
65. De Stefano A, Matta E, Clemente P (2016) Structural health monitoring of historical heritage in Italy: some relevant experiences. *J Civ Struct Health Monit* 6(1):83–106. <https://doi.org/10.1007/S13349-016-0154-Y>
66. Ceretti F (1889) *Delle chiese dei conventi e delle confraternite della Mirandola*. In: *Commissione Municipale di Storia Patria e di Arti Belle della Mirandola (Eds.) Memorie Storiche della Città e dell'antico Ducato della Mirandola*
67. Mugnaini V, Zanotti ragonara L, Civera M (2022) A machine learning approach for automatic operational modal analysis. *Mech Syst Signal Process*. <https://doi.org/10.1016/j.ymssp.2022.108813>

68. Van Overschee P, De Moor B (1996) Subspace identification for linear systems: theory and implementation—applications. Kluwer Academic Press, Dordrecht
69. Civera M, Pecorelli ML, Ceravolo R, Surace C, Zanotti Fragonara L (2021) A multi-objective genetic algorithm strategy for robust optimal sensor placement. *Comput-Aided Civ Infrastruct Eng* 36(9):1185–1202. <https://doi.org/10.1111/mice.12646>
70. Bartoli G, Betti M, Marra AM, Monchetti S (2020) On the role played by the openings on the first frequency of historic masonry towers. *Bull Earthq Eng* 18(2):427–451. <https://doi.org/10.1007/s10518-019-00662-9>
71. Friswell MI, Mottershead JE (1995), Finite element model updating in structural dynamics, vol. 38. In: *Solid mechanics and its applications*, vol. 38. Springer Netherlands, Dordrecht. <https://doi.org/10.1007/978-94-015-8508-8>.

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