INTERFACE

Research

CrossMark
click for updates

Model inversion via multi-fidelity Bayesian optimization: a new paradigm for parameter estimation in haemodynamics, and beyond

Paris Perdikaris[1] and George Em Karniadakis[2]

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Division of Applied Mathematics, Brown University, Providence, RI 02912, USA

We present a computational framework for model inversion based on multi-fidelity information fusion and Bayesian optimization. The proposed methodology targets the accurate construction of response surfaces in parameter space, and the efficient pursuit to identify global optima while keeping the number of expensive function evaluations at a minimum. We train families of correlated surrogates on available data using Gaussian processes and auto-regressive stochastic schemes, and exploit the resulting predictive posterior distributions within a Bayesian optimization setting. This enables a smart adaptive sampling procedure that uses the predictive posterior variance to balance the exploration versus exploitation trade-off, and is a key enabler for practical computations under limited budgets. The effectiveness of the proposed framework is tested on three parameter estimation problems. The first two involve the calibration of outflow boundary conditions of blood flow simulations in arterial bifurcations using multi-fidelity realizations of one- and three-dimensional models, whereas the last one aims to identify the forcing term that generated a particular solution to an elliptic partial differential equation.

1. Introduction

Inverse problems are ubiquitous in science. Being inherently ill-posed, they require solution paths that often challenge the limits of our understanding, as reflected by our modelling and computing capabilities. Unlike forward problems, in inverse problems we have to numerically solve the principal equations (i.e. the forward problem) multiple times, often hundreds of times. The complexity in repeatedly solving the forward problem is further amplified in the presence of nonlinearity, high-dimensional parameter spaces and massive datasets; all common features in realistic physical and biological systems. The natural setting for model inversion finds itself within the principles of Bayesian statistics, which provides a formal ground for parameter estimation, i.e. the process of passing from prior belief to a posterior predictive distribution in view of data. Here, we leverage recent advances in machine learning algorithms for high-dimensional input spaces and big data [1] to design a framework for parameter estimation in blood flow models of the human circulation. The proposed paradigm aims at seamlessly using all available information sources, e.g. experimental measurements, computer simulations, empirical laws, etc., through a general multi-fidelity information fusion methodology in which surrogate models are trained on available data, enabling one to explore the interplay between all such sources in a systematic manner.

In general, we model the response of a system as a function $y = g(\mathbf{x})$ of d input parameters $\mathbf{x} \in \mathbb{R}^d$. The goal of model inversion is to identify the parametric configuration in $\mathbf{x}$ that matches a target response y^*. This translates into solving the following optimization problem:

$$\min_{\mathbf{x} \in \mathbb{R}^d} \|g(\mathbf{x}) - y^*\|, \qquad (1.1)$$

THE ROYAL SOCIETY
PUBLISHING

in some suitable norm. In practice, **x** is often a high-dimensional vector and g is a complex, nonlinear and expensive to compute map that represents the system's evolving dynamics. These factors render the solution of the optimization problem very challenging and motivate the use of surrogate models as a remedy for obtaining inexpensive samples of g at unobserved locations. To this end, a surrogate model acts as an intermediate agent that is trained on available realizations of g, and then is able to perform accurate predictions for the response at a new set of inputs. Ever since the seminal work of Sacks *et al.* [2], the use of surrogates in the design and analysis of computer experiments has undergone great growth, establishing a data-driven mindset for design, optimization and, recently, uncertainty quantification problems [1,3]. Of particular importance to our work is the approach of Kennedy & O'Hagan [4], who introduced the use of stochastic auto-regressive maps for building surrogates from a multitude of information sources of variable fidelity. We reckon that this framework enables the meaningful integration of seemingly disparate methodologies and provides a universal platform in which experiments, simulations and expert opinion can coexist in unison. The main challenges here arise from scaling this surrogate-based approach to constructing response surfaces in high-dimensional input spaces and performing tractable machine learning on massive datasets. Here, we highlight the recent work in [1] that aims to address these challenges and enable the construction of multi-fidelity surrogates for realistic high-dimensional cases.

In silico modelling of blood flow in the human vasculature has received great attention over the last 20 years, resulting in the development of computational tools that helped elucidate key biophysical aspects but also aim to provide a customized, patient-specific tool for prediction, intervention and treatment. Despite great growth in computing power and algorithmic sophistication, the applicability of such models is limited to truncated arterial domains, as the complexity introduced from considering circulation in the entire arterial network remains intractable [5]. Addressing this complexity often leads to the introduction of a series of assumptions, parameters and simplified models. Consequently, the physiological relevance of our computations directly relies on the calibration of such parameters and models, which, due to the inherent difficulty of them being determined in the clinical setting, remains very empirical, as the resulting haemodynamic problem could admit an infinite number of solutions [6]. To this end, recent efforts for developing model inversion techniques in haemodynamics have been proposed in [7–11]. A common theme among these efforts is the utilization of reduced order models that can be sampled extensively and with low computational cost, returning a response $f \approx g$ that enables a computationally tractable solution to the optimization problem of equation (1.1). Among possible reduced order model candidates, the most widely used are nonlinear one-dimensional (1D) fluid–structure interaction (FSI) models, linearized 1D-FSI models and zero-dimensional (0D) lumped parameter models [5].

The goal of this work is to incorporate elements of statistical learning towards building a surrogate-based framework for solving inverse problems in haemodynamics, and beyond. Motivated by recent progress in scalable supervised learning algorithms [1], we propose an information fusion framework that can explore cross-correlations between variable fidelity blood flow models (e.g. measurements versus three-dimensional (3D)-FSI, versus 1D-FSI, versus 0D models, etc.), allowing for the efficient construction of high-dimensional

response surfaces that guide the pursuit for a solution to the optimization problem of equation (1.1), while keeping the number of expensive function evaluations at a minimum. Moreover, we aim to demonstrate that this framework is robust with respect to model misspecification, resulting for example from inaccurate low-fidelity models or noisy measurements. Although our main focus here is on parameter estimation for physiologically correct blood flow simulations, the implications of the proposed methodology are far reaching, and practically applicable to a wide class of inverse problems.

In what follows we provide an outline of this paper. In §2.1, we present an overview of Gaussian process (GP) regression that is the basic building block of the proposed multi-fidelity information fusion algorithms. In §2.2, we elaborate on the implementation of efficient multi-source inference schemes using the recursive inference ideas of Le Gratiet & Garnier [12]. In §2.3, we outline how the parameters defining each surrogate model can be calibrated to the data using maximum-likelihood estimation (MLE). In §2.4, we introduce the efficient global optimization (EGO) algorithm, and highlight how maximizing the expected improvement of the surrogate predictor can lead to efficient sampling strategies for global optimization. The framework is put to the test in §3, where we present results for benchmark problems involving the calibration of outflow boundary conditions in blood flow simulations, as well as parameter estimation in elliptic partial differential equations.

2. Material and methods

The basic building block of the proposed framework is GP regression [13] and auto-regressive stochastic schemes [4]. This toolset enables the construction of a flexible platform for multi-fidelity information fusion that can explore cross-correlations between a collection of surrogate models, and perform predictive inference for quantities of interest that is robust with respect to model misspecification. The main advantage of GPs versus other surrogate models, such as neural networks, radial basis functions or polynomial chaos, lies in its Bayesian non-parametric nature. In particular, it results in a flexible prior over functions, and, more importantly, in a fully probabilistic workflow that returns robust posterior variance estimates that are key to Bayesian optimization. Although there have been attempts in the literature to apply Bayesian optimization using Bayesian neural networks (mainly motivated by alleviating the unfavourable cubic scaling of GPs with data, see [14]), GPs provide several favourable properties, such as analytical tractability, robust variance estimates and the natural extension to the multi-fidelity setting, that currently give it the edge over competing approaches to Bayesian optimization. In what follows we provide an overview of the mathematical framework that is used throughout this work. For a more detailed exposition to the subject, the reader is referred to [1,3,4,12,13].

2.1. Gaussian process regression
The main idea here is to model N scattered observations y of a quantity of interest as a realization of a GP, $f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), fK(\mathbf{x}, \mathbf{x}'; \theta) + \sigma_\epsilon^2 I), \mathbf{x} \in \mathbb{R}^d$. The observations could be deterministic or stochastic in nature and may also be corrupted by modelling errors or measurement noise $\epsilon(\mathbf{x})$, which is thereby assumed to be a zero-mean Gaussian random variable, i.e. $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2 I)$. Therefore, we have the following observation model:

$$y = f(\mathbf{x}) + \epsilon. \tag{2.1}$$

The GP prior distribution on f is completely characterized by its mean $\mu(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$ and covariance $K(\mathbf{x}, \mathbf{x}'; \theta) + \sigma_\epsilon^2 I$, where θ is a

vector of hyper-parameters. Typically, the choice of the prior reflects our belief on the structure, regularity and other intrinsic properties of the quantity of interest. Our primary goal here is to compute the conditional posterior distribution of $p(f|y, \mathbf{x}, \theta)$ and explore its predictive capacity at estimating y for unobserved sets of $\mathbf{x}$. This is achieved by calibrating the parametrization of the Gaussian prior, i.e. $\mu(\mathbf{x})$ and $K(\mathbf{x}, \mathbf{x}'; \theta) + \sigma_\epsilon^2 I$, by maximizing the marginal likelihood of the model. Once $f(\mathbf{x})$ has been trained on the observed data y (see §2.3), its inferred mean $\hat{\mu}$, variance $\hat{\sigma}^2$, noise variance $\hat{\sigma}_\epsilon^2$ and kernel hyper-parameters $\hat{\theta}$ are known and can be used to evaluate the predictions $\hat{y}$, as well as to quantify the prediction variance s^2 (see [15] for a derivation), i.e.

$$\hat{y}(\mathbf{x}^*) = \hat{\mu} + r^{\mathrm{T}}(R + \hat{\sigma}_\epsilon^2 I)^{-1}(y - \mathbf{1}\hat{\mu}) \tag{2.2}$$

and

$$s^2(\mathbf{x}^*) = \hat{\sigma}^2 \left[1 - r^{\mathrm{T}}(R + \hat{\sigma}_\epsilon^2 I)^{-1} r + \frac{[1 - r^{\mathrm{T}}(R + \hat{\sigma}_\epsilon^2 I)^{-1} r]^2}{\mathbf{1}^{\mathrm{T}}(R + \hat{\sigma}_\epsilon^2 I)^{-1} \mathbf{1}} \right], \tag{2.3}$$

where $R = \hat{\sigma}^2 K(\mathbf{x}, \mathbf{x}'; \hat{\theta})$ is the $N \times N$ correlation matrix of $f(\mathbf{x})$, $r = \hat{\sigma}^2 \kappa(\mathbf{x}, \mathbf{x}^*; \hat{\theta})$ is a $1 \times N$ vector containing the correlation between the prediction and the N training points, and $\mathbf{1}$ is a $1 \times N$ vector of ones. Note that for $\hat{\sigma}_\epsilon^2$ the predictor exactly interpolates the training data y, returning zero variance at these locations. This is a linear regression scheme that is also known in the literature as Wiener–Kolmogorov filtering (time-series modelling [16]), kriging (geostatistics [17]), and best linear unbiased estimator (statistics [18]).

2.2. Multi-fidelity modelling via recursive Gaussian processes

Here, we provide a brief overview of the recursive formulations recently put forth by Le Gratiet & Garnier [12]—a more efficient version of the well-known auto-regressive inference scheme proposed by Kennedy & O'Hagan [4] in the context of predicting the output from a complex computer code when fast approximations are available. To this end, suppose that we have s levels of variable fidelity information sources producing outputs $y_t(\mathbf{x}_t)$, at locations $\mathbf{x}_t \in D_t \subseteq \mathbb{R}^d$, sorted by increasing order of fidelity and modelled by GPs $f_t(\mathbf{x})$, $t = 1, \dots, s$. Then, the auto-regressive scheme of Kennedy & O'Hagan [4] reads as

$$f_t(\mathbf{x}) = \rho_{t-1}(\mathbf{x}) f_{t-1}(\mathbf{x}) + \delta_t(\mathbf{x}), \quad t = 2, \dots, s, \tag{2.4}$$

where $\delta_t(\mathbf{x})$ is a Gaussian field independent of $\{f_{t-1}, \dots, f_1\}$ and distributed as $\delta_t \sim \mathcal{N}(\mu_{\delta_t}, \sigma_t^2 R_t + \sigma_{\epsilon_t}^2 I)$. Also, $\rho(\mathbf{x})$ is a scaling factor that quantifies the correlation between $\{f_t(\mathbf{x}), f_{t-1}(\mathbf{x})\}$. In the Bayesian setting, $\rho(\mathbf{x})$ is treated as a random field with an assigned prior distribution that is later calibrated to the data. Here, to simplify the presentation we assume that ρ is a deterministic scalar, independent of $\mathbf{x}$, and learned from the data through MLE (see §2.3). In general, one can obtain a more expressive scheme by introducing a spatial dependence on ρ by using a representation in terms of a basis, i.e. $\rho(x) = \sum_{i=1}^M w_i \phi_i(x)$, where the M unknown modal amplitudes w are augmented to the list of parameters that need to be estimated through MLE. For cases that exhibit a smooth distribution of the cross-correlation scaling, only a small number of modal components should suffice for capturing the spatial variability in $\rho(x)$, thus keeping the computational cost of optimization to tractable levels.

The key idea put forth by Le Gratiet & Garnier [12] is to replace the Gaussian field $f_{t-1}(\mathbf{x})$ in equation (2.4) with a Gaussian field $\hat{f}_{t-1}(\mathbf{x})$ that is conditioned on all known models $\{f_{t-1}, f_{t-2}, \dots, f_1\}$ up to level $(t - 1)$, while assuming that the corresponding experimental design sets D_i, $i = 1, \dots, t - 1$ have a nested structure, i.e. $D_1 \subseteq D_2 \subseteq \cdots \subseteq D_{t-1}$. This essentially allows one to decouple the s-level auto-regressive co-kriging

problem to s independent GP estimation problems that can be efficiently computed and are guaranteed to return a predictive mean and variance that are identical to the coupled Kennedy and O'Hagan scheme [4]. To underline the advantages of this approach, note that the scheme of Kennedy and O'Hagan requires inversion of covariance matrices of size $\sum_{t=1}^s N_t \times \sum_{t=1}^s N_t$, where N_t is the number of observed training points at level t. In contrast, the recursive co-kriging approach involves the inversion of s covariance matrices of size $N_t \times N_t$, $t = 1, \dots, s$.

Once $f_t(\mathbf{x})$ has been trained on the observed data $\{y_t, y_{t-1}, \dots, y_1\}$ (see §2.3), the optimal set of hyper-parameters $\{\hat{\mu}_t, \hat{\sigma}_t^2, \hat{\sigma}_{\epsilon_t}^2, \hat{\rho}_{t-1}, \hat{\theta}_t\}$ is known and can be used to evaluate the predictions $\hat{y}_t$, as well as to quantify the prediction variance s_t^2 at all points in $\mathbf{x}_t^*$ (see [12] for a derivation),

$$\hat{y}_t(\mathbf{x}_t^*) = \hat{\mu}_t + \hat{\rho}_{t-1} \hat{y}_{t-1}(\mathbf{x}_t^*) + r_t^{\mathrm{T}}(R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1}[y_t(\mathbf{x}_t) - \mathbf{1}\hat{\mu}_t \\ - \hat{\rho}_{t-1} \hat{y}_{t-1}(\mathbf{x}_t)] \tag{2.5}$$

and

$$s_t^2(\mathbf{x}_t^*) = \hat{\rho}_{t-1}^2 s_{t-1}^2(\mathbf{x}_t^*) \\ + \hat{\sigma}_t^2 \left[1 - r_t^{\mathrm{T}}(R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1} r_t + \frac{[1 - r_t^{\mathrm{T}}(R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1} r_t]^2}{\mathbf{1}_t^{\mathrm{T}}(R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1} \mathbf{1}_t} \right], \tag{2.6}$$

where $R_t = \hat{\sigma}_t^2 K_t(\mathbf{x}_t, \mathbf{x}'_t; \hat{\theta}_t)$ is the $N_t \times N_t$ correlation matrix of $f_t(\mathbf{x})$, $r_t = \hat{\sigma}_t^2 k_t(\mathbf{x}_t, \mathbf{x}_t^*; \hat{\theta}_t)$ is a $1 \times N_t$ vector containing the correlation between the prediction and the N_t training points, and $\mathbf{1}_t$ is a $1 \times N_t$ vector of ones. Note that for $t = 1$ the above scheme reduces to standard GP regression as presented in equations (2.2) and (2.3). Also, $K_t(\mathbf{x}_t, \mathbf{x}'_t; \theta_t) + \sigma_{\epsilon_t}^2 I$ is the auto-correlation kernel that quantifies spatial correlations at level t.

2.3. Maximum-likelihood estimation

Estimating the hyper-parameters requires learning the optimal set of $\{\mu_t, \sigma_t^2, \sigma_{\epsilon_t}^2, \rho_{t-1}, \theta_t\}$ from all known observations $\{y_t, y_{t-1}, \dots, y_1\}$ at each inference level t. In what follows we will confine the presentation to MLE procedures for the sake of clarity. However, in the general Bayesian setting all hyper-parameters are assigned with prior distributions, and inference is performed via more costly marginalization techniques, typically using Markov chain Monte Carlo (MCMC) integration [13,19,20]. Indeed, a fully Bayesian approach to hyper-parameter estimation has been documented to be more robust (hence advantageous) in the context of Bayesian optimization [21]. Although great progress has been made in training GPs using sampling-based techniques, in this work we employed MLE mainly due to its straightforward formulation and reduced computational complexity versus MCMC approaches, as well as to simplify the introduction of the proposed concepts to a wider audience that can readily implement the methods.

Parameter estimation via MLE at each inference level t is achieved by minimizing the negative log-likelihood of the model evidence,

$$\min_{\{\mu_t, \sigma_t^2, \sigma_{\epsilon_t}^2, \rho_{t-1}, \theta_t\}} \frac{n}{2} \log(\sigma_t^2) + \frac{1}{2} \log|R_t(\theta_t) + \sigma_{\epsilon_t}^2 I| \\ + \frac{1}{2\sigma_t^2} [y_t(\mathbf{x}_t) - \mathbf{1}_t \mu_t - \rho_{t-1} \hat{y}_{t-1}(\mathbf{x}_t)]^{\mathrm{T}} [R_t(\theta_t) + \sigma_{\epsilon_t}^2 I]^{-1} \\ \times [y_t(\mathbf{x}_t) - \mathbf{1}_t \mu_t - \rho_{t-1} \hat{y}_{t-1}(\mathbf{x}_t)], \tag{2.7}$$

where we have highlighted the dependence of the correlation matrix R_t on the hyper-parameters θ_t. Setting the derivatives of this expression to zero with respect to μ_t, ρ_{t-1} and σ_t^2, we can express the optimal values of $\hat{\mu}_t$, $\hat{\rho}_{t-1}$ and $\hat{\sigma}_t^2$ as functions of the correlation matrix $(R_t + \sigma_{\epsilon_t}^2 I)$,

$$(\hat{\mu}_t, \hat{\rho}_{t-1}) = [\mathbf{h}_t^{\mathrm{T}}(R_t + \sigma_{\epsilon_t}^2 I)^{-1} \mathbf{h}_t]^{-1} \mathbf{h}_t^{\mathrm{T}}(R_t + \sigma_{\epsilon_t}^2 I)^{-1} y_t(\mathbf{x}_t) \tag{2.8}$$

4

and

$$\hat{\sigma}_t^2 = \frac{1}{c} \Big\{ [y_t(\mathbf{x}_t) - \mathbf{1}_t\hat{\mu}_t - \hat{\rho}_{t-1}\hat{y}_{t-1}(\mathbf{x}_t)]^T [R_t + \sigma_{\epsilon_t}^2 I]^{-1}$$
$$\times [y_t(\mathbf{x}_t) - \mathbf{1}_t\hat{\mu}_t - \hat{\rho}_{t-1}\hat{y}_{t-1}(\mathbf{x}_t)] \Big\}, \quad (2.9)$$

where $\mathbf{h}_t = [\mathbf{1}_t \hat{y}_{t-1}(\mathbf{x}_t)]$ and $c = \begin{cases} N_t - 1, & t = 1 \\ N_t - 2, & t > 1 \end{cases}$. Finally, the optimal $\{\hat{\sigma}_{\epsilon_t}^2, \hat{\theta}_t\}$ can be estimated by minimizing the concentrated restricted log-likelihood

$$\min_{\{\sigma_{\epsilon_t}^2, \theta_t\}} \log |R_t(\theta_t) + \sigma_{\epsilon_t}^2 I| + c \log(\hat{\sigma}_t^2). \quad (2.10)$$

The computational cost of calibrating model hyper-parameters through MLE is dominated by the inversion of correlation matrices $(R_t + \sigma_{\epsilon_t}^2 I)^{-1}$ at each iteration of the minimization procedure in equation (2.10). The inversion is typically performed using the Cholesky decomposition that scales as $\mathcal{O}(N_t^3)$, leading to a severe bottleneck in the presence of moderately big datasets. Alternatively, one may employ iterative methods for solving the linear system and reduce the cubic cost (especially if a good pre-conditioner can be found), but in this case another method for computing the determinant in equation (2.10) should be employed. In general, alleviating the cubing scaling is a pivotal point in the success and practical applicability of GPs, and several studies in the literature have been devoted to this [1,22–25]. This limitation is further pronounced in high-dimensional problems where abundance of data is often required for performing meaningful inference. Moreover, cases where the noise variance $\sigma_{\epsilon_t}^2$ is negligible and/or the observed data points are tightly clustered in space may introduce ill-conditioning that could jeopardize the feasibility of the inversion as well as pollute the numerical solution with errors. Also, if an anisotropic correlation kernel $K_t(\mathbf{x}_t, \mathbf{x}'_t; \theta_t)$ is assumed, then the vector of correlation lengths θ_t is d-dimensional, leading to an increasingly complex optimization problem (see equation (2.10)) as the dimensionality of the input variables $\mathbf{x}_t$ increases. These shortcomings render the learning process intractable for high dimensions and large datasets, and suggest seeking alternative routes to parameter estimation. A possible solution path was recently proposed in [1], where the authors have employed a data-driven additive decomposition of the covariance kernel in conjunction with frequency-domain inference algorithms that bypass the need to invert large covariance matrices.

2.4. Bayesian optimization

Our primary goal here is to use the surrogate models generated by the recursive multi-fidelity formulation towards identifying the global optimum of the optimization problem defined in equation (1.1) by replacing the unknown and expensive to evaluate function g with its corresponding surrogate predictor f_t. The probabilistic structure of the surrogate predictors enables an efficient solution path to this optimization problem by guiding a sampling strategy that balances the trade-off between exploration and exploitation, i.e. the global search to reduce uncertainty versus the local search in regions where the global optimum is likely to reside. One of the most widely used sampling strategies in Bayesian optimization that adopts this mindset is the EGO algorithm proposed by Jones *et al.* [26]. Given a predictive GP posterior, the EGO algorithm selects points in the input space that maximize the *expected improvement* of the predictor. To this end, let $f_{\min}^{(t)} = \min\{y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(N_t)}\}$ be the global minimum of the observed response at the tth inference level. Consequently, the improvement of the Gaussian predictor $f_t(\mathbf{x})$ upon $f_{\min}$ is defined as $I_t(\mathbf{x}) = \max\{f_{\min}^{(t)} - f_t(\mathbf{x}), 0\}$. Then, the infill criterion suggested by the EGO algorithm implies sampling at locations in the input space that maximize the expected improvement

$$\mathbb{E}[I_t(\mathbf{x})] = [f_{\min}^{(t)} - \hat{y}_t(\mathbf{x})] \Phi\left(\frac{f_{\min}^{(t)} - \hat{y}_t(\mathbf{x})}{s_t(\mathbf{x})}\right) + s\phi\left(\frac{f_{\min}^{(t)} - \hat{y}_t(\mathbf{x})}{s_t(\mathbf{x})}\right), \quad (2.11)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal cumulative distribution and density function, respectively, and s_t is the square root of the predictor variance at level t (see equation (2.6)). Consequently, the value of $\mathbb{E}[I_t(\mathbf{x})]$ is large if either the value predicted by $\hat{y}_t(\mathbf{x})$ is smaller than $f_{\min}^{(t)}$ or there is a large amount of uncertainty in the predictor $\hat{y}_t(\mathbf{x})$, hence $s_t^2(\mathbf{x})$ is large. Here we must underline that, although the expected improvement acquisition function is widely used in the literature, it may have some important pitfalls, mainly due to its vulnerability to noise in the observations, especially in the heteroscedastic case. The design of robust and efficient data acquisition techniques is an active area of research and several interesting studies exist in the literature. The work of Picheny *et al.* [27] addresses some of these issues, including parallel batch sampling and robustness to noise using a quantile-based expected improvement criterion. A different class of robust methods has branched out of the entropy search algorithm put forth in [28]. These approaches are good candidates in cases where the simple expected improvement acquisition fails, and can be implemented with minor overhead to the existing computational cost.

Once a recursive cycle has been completed and the final predictor $\hat{y}_s$ and variance $s_s^2(\mathbf{x})$ at level s are known, the expected improvement can be readily computed by equation (2.11). Then, the EGO algorithm suggests to re-train the surrogates by augmenting the design sets D_t with a set of points that correspond to locations where the expected improvement is maximized. This procedure iterates until a stopping criterion is met. Owing to the potentially high computational cost of generating training data from sampling g in equation (1.1), it is common to use the maximum number of function evaluations as the stopping criterion or the convergence rate of the objective function. Another termination criterion stems from setting a target value for the expected improvement, allowing the next cycle to be carried out only if the expected improvement is above the imposed threshold. Although the theoretical analysis of such probabilistic algorithms is challenging, theoretical guarantees of convergence to the global optimum have been studied in the works of [29,30].

2.5. Workflow and computational cost

The EGO algorithm can be summarized in the following implementation steps:

Step 1: Given samples of the true response $y_t = g_t(\mathbf{x}_t)$, $t = 1, \ldots, s$ at all fidelity levels we employ the recursive scheme outlined in §2.2 to construct a family of predictive surrogates $\hat{y}_t(\mathbf{x})$. This step scales cubically with the number of training points using MLE. The number of required training points is generally problem dependent as it is related to the complexity of the underlying response surface we try to approximate. A rule of thumb is to start with $6d$ points at the lowest fidelity level, where d is the dimension of the input space, and halve this number for each subsequent higher fidelity level. This initialization is usually performed using Latin hypercube sampling and has been proven robust in our experiment so far. A reason for this is that in every iteration of the Bayesian optimization loop the acquisition function will choose a new informative sampling location. Therefore, even if the initial sampling plan is inadequate, new informative points will be added in the first couple of iterations and the issue will be resolved.

Step 2: For each surrogate, we compute the set of input configurations that maximize the expected improvement, i.e. we solve the optimization problem

$$\max_{x \in \mathbb{R}^d} \mathbb{E}[I_t(\mathbf{x})]. \quad (2.12)$$

This step involves probing the trained surrogate to perform predictions at new locations, each of which can be obtained in parallel and with linear cost.

Step 3: We evaluate the response $y_t = g_t(\mathbf{x_t})$, $t = 1, \ldots, s$ at the new points suggested by maximizing the expected improvement at each fidelity level. Therefore, this step scales according to the computation cost of computing the given models $g_t(\mathbf{x_t})$.

Step 4: We train again the multi-fidelity surrogates using the augmented design sets and corresponding data to obtain new predictive schemes that resolve in more detail regions of the response surface where the minimum is more likely to occur. Similarly, training the surrogates scales cubically with the data, while predictions of the trained surrogate are obtained in linear cost.

Step 5: Perform step 2 and check if the desired termination criterion is met. This could entail checking whether a maximum number of function evaluations is reached, whether the minimum of the objective function has converged, or whether the maximum expected improvement is less than a threshold value. If the chosen criterion is not satisfied, we repeat steps 3–5.

2.6. Haemodynamics

There exists a wide variety of modelling approaches to study the multi-physics phenomena that characterize blood flow in the human circulation. At the continuum scale, our understanding is mainly based on clinical imaging and measurements, *in vitro* experiments, and numerical simulations. Each of these approaches complements one another as they are able to reveal different aspects of the phenomena under study at different levels of accuracy. Without loss of generality, in what follows we focus our attention on 1D and 3D formulations of blood flow in compliant and rigid arteries, respectively. In particular, we consider three different layers of fidelity as represented by (i) 3D Navier–Stokes models in rigid domains, (ii) nonlinear 1D FSI models, and (iii) linearized 1D-FSI models.

2.6.1. Three-dimensional models in rigid arterial domains

We consider 3D unsteady incompressible flow in a rigid domain Ω, described by the Navier–Stokes equations subject to the incompressibility constraint

$$\left.\begin{array}{c} \dfrac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot (\nabla \mathbf{v}) = -\nabla p + \nu \nabla^2 \mathbf{v} + \mathbf{f} \\ \text{and} \quad \nabla \cdot \mathbf{v} = 0, \end{array}\right\} \tag{2.13}$$

where $\mathbf{v}$ is the velocity vector, p is the pressure, t is time, ν is the kinematic viscosity of the fluid and $\mathbf{f}$ are the external body forces. The system is discretized in space using the spectral/*hp* element method (SEM) [31], according to which the computational domain Ω is decomposed into a set of polymorphic non-overlapping elements $\Omega^{e_i} \subset \Omega$, $i = 1, \ldots$, Nel. Within each element, the solution is approximated as a linear combination of hierarchical, mixed-order, semi-orthogonal Jacobi polynomial expansions. This hierarchical structure consists of separated vertex (linear term) $\Phi_k(\mathbf{x})$, edge $\Psi_k(\mathbf{x})$, face $\Theta_k(\mathbf{x})$ and interior (or bubble) modes $\Lambda_k(\mathbf{x})$. According to this decomposition, the polynomial representation of a field $\mathbf{v}(t, \mathbf{x})$ at any point $\mathbf{x}_j$ is given by the linear combination of the basis functions multiplied by the corresponding modal amplitudes:

$$\mathbf{v}(t, \mathbf{x}_j) = \sum_{k=1}^{Nv} \hat{\mathbf{v}}_k^V(t) \Phi_k(\mathbf{x}_j) + \sum_{k=1}^{Ne} \hat{\mathbf{v}}_k^E(t) \Psi_k(\mathbf{x}_j) + \sum_{k=1}^{Nf} \hat{\mathbf{v}}_k^F(t) \Theta_k(\mathbf{x}_j)$$
$$+ \sum_{k=1}^{Ni} \hat{\mathbf{v}}_k^I(t) \Lambda_k(\mathbf{x}_j).$$

The numerical solution to the above system is based on decoupling the velocity and pressure by applying a high-order time-splitting scheme [32]. We solve for an intermediate velocity field $\mathbf{v}^*$ in physical space (equation (2.14)) before a Galerkin projection is applied to obtain the weak formulation for the pressure

and velocity variables (equations (2.15) and (2.16)), which are computed by solving a Poisson and a Helmholtz problem, respectively. At time step n and sub-iteration step k, we solve equations (2.15) and (2.16) for p_k^n and $\mathbf{v}_k^n$, respectively, using previous time-step solutions $\mathbf{v}^{n-q}$, $q = 1, \ldots, J_e$, where J_e denotes the order of the stiffly stable time integration scheme used,

$$\mathbf{v}^* = \sum_{q=0}^{Je-1} \alpha_k \mathbf{v}^{n-q} - \Delta t(\mathbf{N}_{k-1}^n + \mathbf{f}), \tag{2.14}$$

$$\mathbf{L}\hat{p}_k^n = -\frac{1}{\Delta t}(\nabla \cdot \mathbf{v}^*, \phi) + \left(\frac{\partial p_k^n}{\partial n}, \phi\right) - \mathbf{L}^D \hat{p}^D \tag{2.15}$$

and $\quad \mathbf{H}\hat{\mathbf{v}}_k^n = \dfrac{1}{\gamma_0}(\mathbf{v}^* - \Delta t \nabla p_k^n, \phi) + \dfrac{\Delta t \nu}{\gamma_0}\left(\dfrac{\partial \mathbf{v}_k^n}{\partial n}, \phi\right) - \mathbf{H}^D \hat{\mathbf{v}}^{n+1,D},$

$$\tag{2.16}$$

where we define

$$\mathbf{N}_{k-1}^n = \mathbf{v}_{k-1}^n \cdot \nabla \mathbf{v}_{k-1}^n. \tag{2.17}$$

Also, γ_0, α_k are the coefficients of a backward differentiation formula, β_k are the coefficients of the stiffly stable time integration scheme, $\mathbf{H} = \mathbf{M} - \Delta t \nu / \gamma_0$, and $\mathbf{M}$ and $\mathbf{L}$ are the mass and stiffness matrices, respectively. Moreover, $\hat{\mathbf{v}}$ and $\hat{p}$ are the unknown modal amplitudes of the velocity and pressure variables, respectively, while variables with a D superscript are considered to be known. In particular, the last term in equations (2.15) and (2.16) is due to lifting a known solution for imposing Dirichlet boundary conditions, and the corresponding operators $\mathbf{H}^D$ and $\mathbf{L}^D$ need to be constructed only for these boundaries [31]. The system is closed with a Neumann $\partial \mathbf{v} / \partial n = 0$ boundary condition for the velocity at the outlets, and a consistent Neumann pressure boundary condition at the boundaries with a prescribed Dirichlet condition for the velocity field [32]:

$$\frac{\partial p_k^n}{\partial n} = -\left[\frac{\gamma_0 \mathbf{v}_k^n - \sum_{q=0}^{Je-1} \alpha_k \mathbf{v}^{n-q}}{\Delta t} + \sum_{k=0}^{Je-1}[\beta_k(\mathbf{N}_{k-1}^n + \mathbf{V}_{k-1}^n)]^{n-k}\right] \cdot \mathbf{n},$$
$$\tag{2.18}$$

where $\mathbf{V}_{k-1}^n$ is defined as

$$\mathbf{V}_{k-1}^n = \nu \nabla \times (\nabla \times \mathbf{v}_{k-1}^n). \tag{2.19}$$

At the first sub-iteration step $k = 1$, the solution from previous sub-iterations is not available and $\mathbf{N}_0^n$, $\mathbf{V}_0^n$ are extrapolated in time as follows:

$$\mathbf{N}_0^n = \sum_{q=1}^{Je} \beta_q [\mathbf{v} \cdot \nabla \mathbf{v}]^{n-q} \tag{2.20}$$

and

$$\mathbf{V}_0^n = \nu \sum_{q=1}^{Je} \beta_q [\nabla \times (\nabla \times \mathbf{v})]^{n-q}. \tag{2.21}$$

After computing $\hat{\mathbf{v}}_k^n$ and $\hat{p}_k^n$, we can accelerate the convergence of the iterative procedure by applying the Aitken under-relaxation scheme to the computed velocity field:

$$\tilde{\mathbf{v}}_k^n = \lambda_k \tilde{\mathbf{v}}_{k-1}^n + (1 - \lambda_k)\mathbf{v}_k^n, \tag{2.22}$$

where $\lambda_k \in [0, 1]$ is computed using the classical Aitken rule:

$$\lambda_k = \lambda_{k-1} + (\lambda_{k-1} - 1)\frac{(\mathbf{Q}_{k-1} - \mathbf{Q}_k) \cdot \mathbf{Q}_k}{\|\mathbf{Q}_{k-1} - \mathbf{Q}_k\|^2},$$

$$\mathbf{Q}_k = \tilde{\mathbf{v}}_{k-1}^n - \mathbf{v}_k^n. \tag{2.23}$$

Finally, convergence is checked by requiring $\|\mathbf{v}_k^n - \mathbf{v}_{k-1}^n\|$ and $\|p_k^n - p_{k-1}^n\|$ to be less than a pre-specified tolerance.

For a complete analysis of this scheme, the reader is referred to [32,33]. Results in the context of arterial blood flow simulations can be found in [34], while parallel performance has been documented in [35].
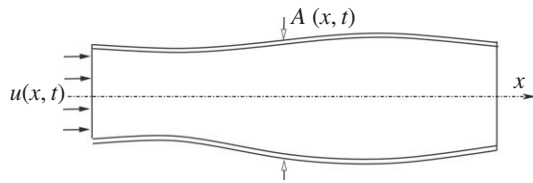
Figure 1. Flow in a 1D compliant artery. (Adapted from [37].)

2.6.2. Nonlinear one-dimensional models in compliant arterial domains

We consider viscous incompressible 1D flow in a compliant tube. The flow dynamics is governed by a nonlinear hyperbolic system of partial differential equations that can be directly derived from the Navier–Stokes equations under the assumptions of axial symmetry, dominance of the axial velocity component, radial displacements of the arterial wall and constant internal pressure on each cross section [5]. Although such averaging introduces approximations that neglect some of the underlying physics, this reduced model has been widely used and validated in the literature, returning reasonable predictions for pulse wave propagation problems [36]. Here, we would like to highlight that such simple models can become extremely useful when employed as low-fidelity models within the proposed multi-fidelity framework, because (despite their inaccuracies) they can capture the right trends at very low computational cost (figure 1).

The conservation of mass and momentum can be formulated in space–time (A, U) variables as (see [37] for a detailed derivation)

$$\left.\begin{array}{c} \dfrac{\partial A}{\partial t} + \dfrac{\partial (AU)}{\partial x} = 0 \\[2mm] \text{and} \qquad \dfrac{\partial U}{\partial t} + U \dfrac{\partial U}{\partial x} = -\dfrac{1}{\rho}\dfrac{\partial p}{\partial x} + K_r \dfrac{U}{\rho A}, \end{array}\right\} \quad (2.24)$$

where x is the axial coordinate across the vessel's length, t is time, $A(x, t)$ is the cross-sectional area of the lumen, $U(x, t)$ is average axial fluid velocity, $Q(x, t) = AU$ is the mass flux, $p(x, t)$ is the internal pressure averaged over the tube's cross section and K_r is a friction parameter that depends on the velocity profile chosen [37]. Here, we use an axisymmetric fluid velocity $u(x, r, t)$ profile that satisfies the no-slip condition,

$$u(x, r, t) = U \frac{\zeta + 2}{\zeta} \left\{ 1 - \left(\frac{r}{R}\right)^\zeta \right\}, \quad (2.25)$$

with $R(x, t)$ being the lumen radius, ζ a constant and r the radial coordinate. Following [5], $\zeta = 9$ gives a good fit to experimental blood flow data and $\zeta = 2$ returns the parabolic flow profile. Moreover, $U = 1/R^2 \int_0^R 2ru\,dr$, and the friction parameter can be expressed as $K_r = (2\mu/U)(A/R)[\partial u/\partial r]_R = -22\mu\pi$ [37], with μ being the blood viscosity that is a function of the lumen radius and the blood haematocrit, based on the findings of Pries et al. [38].

System (2.24) can be solved for (A, U) after introducing a constitutive law for the arterial wall that relates the cross-sectional area A to pressure p. For simplicity, in this work, we have adopted a purely elastic wall model, leading to a pressure–area relation of the form [39]

$$p = p_{ext} + \frac{\beta}{\sqrt{A}} \left(\sqrt{A} - \sqrt{A_0} \right), \quad \beta = \frac{\sqrt{\pi} E h}{(1 - v^2) A_0}, \quad (2.26)$$

where p_{ext} is the external pressure on the arterial wall, E is the Young modulus of the wall, h is the wall thickness and v is the Poisson ratio. More complex constitutive laws accounting for viscoelastic creep and stress relaxation phenomena can be derived based on the theory of quasi-linear viscoelasticity [40],

and the reader is referred to [39] for implementation details in the context of 1D blood flow solvers.

For the numerical solution of system (2.24), we have adopted the discontinuous Galerkin scheme first put forth by Sherwin et al. [37]. Spatial discretization consists of dividing each arterial domain Ω into Nel elemental non-overlapping regions. A continuous solution within each element is retrieved as a linear combination of orthogonal Legendre polynomials, while discontinuities may appear across elemental interfaces. Global continuity is restored by flux upwinding across elemental interfaces, bifurcations and junctions, with conservation of mass, continuity of Riemann invariants and continuity of the total pressure being ensured by the solution of a Riemann problem [37]. Finally, time integration is performed explicitly using a second-order accurate Adams–Bashforth scheme [37].

2.6.3. Linear one-dimensional models in compliant arterial domains

The analysis can be further simplified if one adopts a linear formulation [37,41]. This can be achieved by expressing the system of equations in equation (2.24) in terms of the (A, p, Q) variables, with $Q = AU$, and linearizing them about the reference state $(A_0, 0, 0)$, with β and A_0 held constant along x. This yields

$$C_{1D} \frac{\partial \tilde{p}}{\partial t} + \frac{\partial \tilde{q}}{\partial x} = 0, \quad (2.27)$$

$$L_{1D} \frac{\partial \tilde{q}}{\partial t} + \frac{\partial \tilde{p}}{\partial x} = -R_{1D}\tilde{q} \quad (2.28)$$

and

$$\tilde{p} = \frac{\tilde{a}}{C_{1D}}, \quad (2.29)$$

where $\tilde{a}$, $\tilde{p}$ and $\tilde{q}$ are the perturbation variables for area, pressure and volume flux, respectively, i.e. $(A, p, Q) = (A + \tilde{a}, \tilde{p}, \tilde{q})$, and

$$R_{1D} = \frac{2(\zeta + 2)\pi\mu}{A_0^2}, \quad L_{1D} = \frac{\rho}{A_0} \quad \text{and} \quad C_{1D} = \frac{A_0}{\rho c_0^2} \quad (2.30)$$

are the viscous resistance to flow, blood inertia and wall compliance, respectively, per unit length of vessel [41]. For a purely elastic arterial wall behaviour, it follows that

$$\frac{\partial \tilde{p}}{\partial \tilde{a}} = \frac{\partial p}{\partial A}\bigg|_{A=A_0} = \frac{1}{C_{1D}}. \quad (2.31)$$

Then, the corresponding linear forward and backward Riemann invariant can analytically be expressed as

$$W_{1,2} = \tilde{q} \pm \frac{\tilde{p}}{Z_0}, \quad Z_0 = \frac{\rho c_0}{A_0}, \quad (2.32)$$

where Z_0 is the characteristic impedance of the vessel.

3. Results

Here, we aim to demonstrate the effectiveness of the proposed framework through three benchmark problems. In the first two cases, consider the important aspect of calibrating outflow boundary conditions of blood flow simulations in truncated arterial domains. In the first demonstration, we employ simple nonlinear and linearized 1D-FSI models in a symmetric arterial bifurcation, aiming to provide a pedagogical example that highlights the main aspects of this work. In the second case, we target a more realistic scenario in which we seek to calibrate a 3D Navier–Stokes solver by leveraging cheap evaluations of 1D models in an asymmetric Y-shaped bifurcation topology. Finally, we conclude with a higher dimensional problem that involves inferring the forcing term that gave rise to a given solution of an elliptic partial differential equation.

arterial segment	length (cm)	radius (mm)	β (Pa m^{-1})
1	5	2	6.97×10^7
2	3	1.5	5.42×10^8
3	3	1.5	5.42×10^8
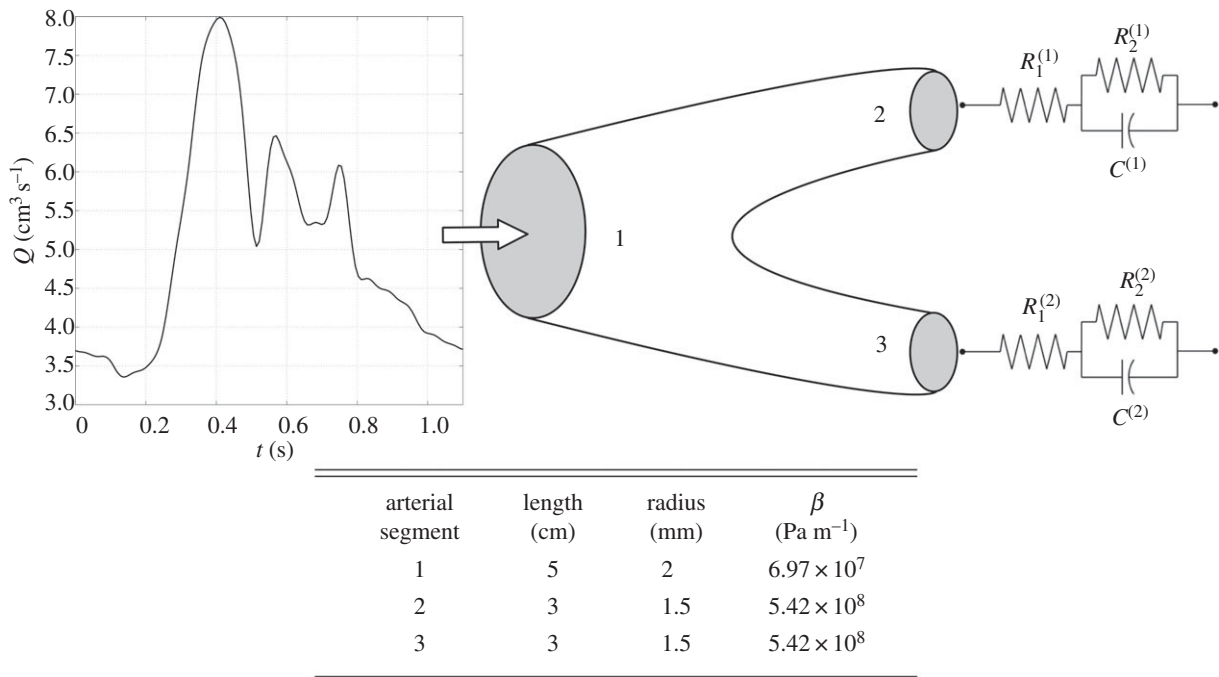
Figure 2. Blood flow in a symmetric Y-shaped bifurcation. Flow is driven by a physiological flow-rate wave at the inlet, while outflow boundary conditions are imposed through three-element Windkessel models. The inset table contains the geometric and elasticity properties used in this simulation.

3.1. Calibration of a nonlinear one-dimensional solver

We consider a benchmark problem for calibrating the outflow parameters in a Y-shaped symmetric bifurcation with one inlet and two outlets. The geometry resembles the characteristic size and properties of an idealized carotid bifurcation, yet is kept symmetric to enhance clarity in our presentation. Figure 2 presents a schematic of the problem set-up. In particular, the effect of the neglected downstream vasculature is taken into account through three-element Windkessel models, while the inflow is driven by a physiologically correct flow-rate wave.

Our goal is to calibrate the outflow resistance parameters to obtain a physiologically correct pressure wave at the inlet. To enable a meaningful visualization of the solution process, we confine ourselves to a two-dimensional input space defined by variability in the total resistances imposed at the two outlets, $R_T^{(i)}$, $i = 1, 2$. The total resistance is defined as $R_T^{(i)} = R_1^{(i)} + R_2^{(i)}$, where $R_1^{(i)}$ corresponds to the characteristic impedance of each terminal vessel (see equation (2.32)). Moreover, the capacitance parameters are empirically set following the findings of Grinberg & Karniadakis [42] as $C^{(i)} = 0.18/R_T^{(i)}$, $i = 1, 2$. To this end, our objective is to identify a pair of $R_T^{(i)}$ that matches a physiologically correct systolic pressure of $p_s^* = 126$ mmHg at the inlet. Assuming no prior knowledge on possible sensible combinations in the input $\mathbf{x} = [R_T^{(1)}, R_T^{(2)}]^T$, we assume a large input space that spans four orders of magnitude: $\mathcal{X} = [10^6, 10^{10}]^2$ measured in (Pa s m^{-3}), and consider the following optimization problem:

$$\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathcal{X}} \frac{(p_s^* - p_s(\mathbf{x}))^2}{(p_s^*)^2}. \tag{3.1}$$

First, we employ a nonlinear 1D-FSI model in order to construct a detailed representation of the response surface of the relative error in systolic pressure, i.e. $g(\mathbf{x}) = (p_s^* - p_s(\mathbf{x}))^2/(p_s^*)^2$, that will be later used as a reference solution to assess the
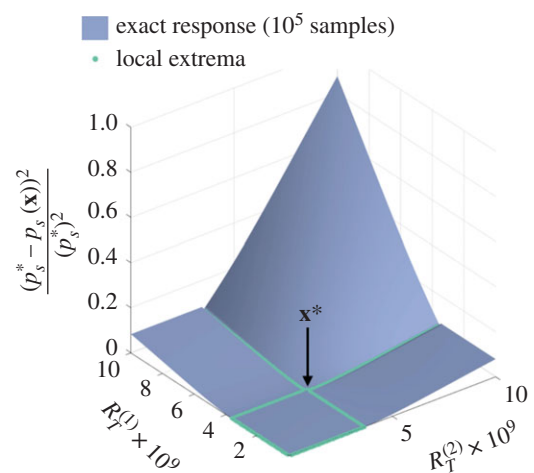
Figure 3. Blood flow in a symmetric Y-shaped bifurcation. Detailed response surface and identification of the minimum error in the inlet systolic pressure as a function of the total resistance parameters imposed at the outlets. The surface is constructed by probing a nonlinear 1D-FSI model on 10 000 uniformly spaced samples in the space of inputs χ. (Online version in colour.)

accuracy and convergence of the proposed multi-fidelity model inversion techniques. Here, the choice of studying haemodynamics using 1D models is motivated by their ability to accurately reflect the interplay between the parametrization of the outflow and the systolic pressure at the inlet, yet at a very low computational cost [37,43]. To this end, figure 3 shows the resulting response surface obtained by probing an accurate nonlinear 1D-FSI solver on 10 000 uniformly spaced samples of the input variables. Since we have considered a symmetric geometry, the resulting response surface exhibits symmetry with respect to the diagonal plane, and can be visibly subdivided using the path of its local extrema into four different subregions. The first part is confined in a flat square region near the origin, suggesting that the solution is relatively
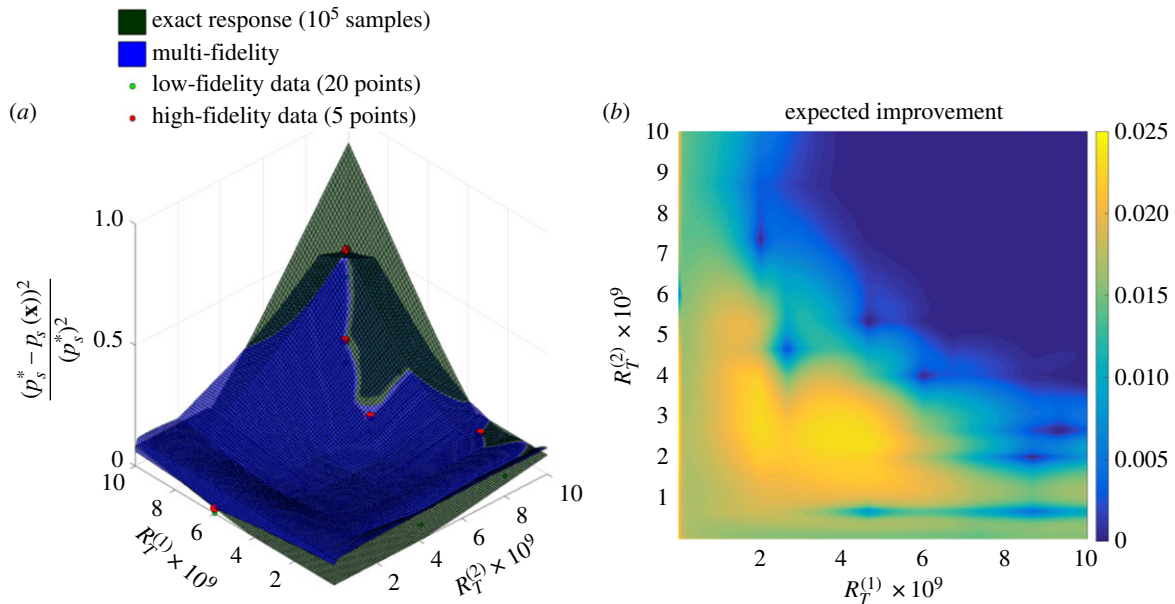
Figure 4. Blood flow in a symmetric Y-shaped bifurcation. Zeroth iteration of the EGO algorithm. (*a*) Exact solution versus the multi-fidelity predictor for the inlet systolic pressure error, trained on 20 low-fidelity and five high-fidelity observations. (*b*) Map of the corresponding expected improvement. (Online version in colour.)

insensitive to the choice of inputs, as any combination in the inputs within this region results in a reasonably small error in the systolic pressure. The second and third parts correspond to the symmetric rectangular regions defined by further increasing one of the two resistance parameters. There, the response starts to exhibit sensitivity to the increasing input parameter, leading to a noticeable error in the systolic pressure. Lastly, in the fourth region, the solution exhibits strong sensitivity on the inputs, leading to an explosive growth of the systolic pressure deviation as the resistance parameters take larger values. Interestingly, a closer look at the path of local extrema reveals a subtle undershoot in the response surface along the interface of the aforementioned subregions. Although any input combination within the first subregion results in a very small systolic pressure error, the global solution x^* of the minimization problem in equation (3.1) resides right at the interface of the four subregions (figure 3). Note that this topology of the global response is likely to pose serious challenges to any gradient-based optimization strategy, leading to a potentially very large number of function evaluations until convergence, especially in the absence of a very accurate initial guess.

Now, we turn our attention to solving the optimization problem introduced by equation (3.1) using the proposed surrogate-based framework. To this end, in order to build multi-fidelity surrogates for the systolic pressure $y = p_s(x)$, we have considered probing two models: high-fidelity solutions are obtained through the aforementioned nonlinear 1D-FSI solver, while the lower fidelity response is measured through a linearized 1D-FSI model. Here, we emphasize that the linearized model has been purposely derived around a biased reference state, returning erroneous predictions that deviate from the correct solution up to 30%. This is to demonstrate that the proposed methodology is robust with respect to misspecification in the lower fidelity observations. It is important to stress here that, despite its inaccurate predictions, the low-fidelity model can still be highly informative as long as it is statistically correlated to the high-fidelity model.

Initially, we perform 20 linearized low-fidelity simulations, supplemented by five high-fidelity realizations, all randomly sampled within the bounds that define the space of inputs. Then, based on the computed response $y_t(x_t)$, we train a two-level surrogate using the recursive scheme presented in §2.2 with a stationary Matérn 5/2 kernel function [13]. The Matérn family of kernels provides a flexible parametrization that has been commonly employed in the literature mainly due to its simplicity and the ability to model a wide spectrum of responses ranging from rough Ornstein–Uhlenbeck paths to infinitely differentiable functions [13]. However, one of its limitations stems from its stationary behaviour (i.e. the Matérn auto-correlation between two spatial locations is translation invariant). Accounting for non-stationary effects has been found to be particularly important in the context of Bayesian optimization, and here we address this by employing the *warping* transformation put forth by Snoek *et al.* [44], leading to kernels of the form $\kappa(h(x), h(x'); \theta)$, where h denotes a nonlinear warping function. From our experience, this step is more crucial than the initial choice of kernel and appears to be quite robust for either the Matérn family or any other popular alternative, such as the radial basis function, spherical and polynomial kernels [13].

The resulting response surface of the error in the systolic pressure at the inlet is presented in comparison with the reference solution in figure 4. Once the surrogate predictor and variance are available, we can compute the spatial distribution of the corresponding expected improvement using equation (2.11). This highlights regions of high expected improvement, as illustrated in figure 4, thus suggesting the optimal sampling locations for the next iteration of the EGO algorithm (see §2.4). Note that the maximum expected improvement in this zeroth iteration is already attained very close to the global minimum of the reference solution (figure 3).

The next step involves performing an additional set of low- and high-fidelity simulations at the suggested locations of high expected improvement (see §2.4). Specifically, in each iteration, we augment the training data with 20 low-fidelity and five high-fidelity observations placed in the
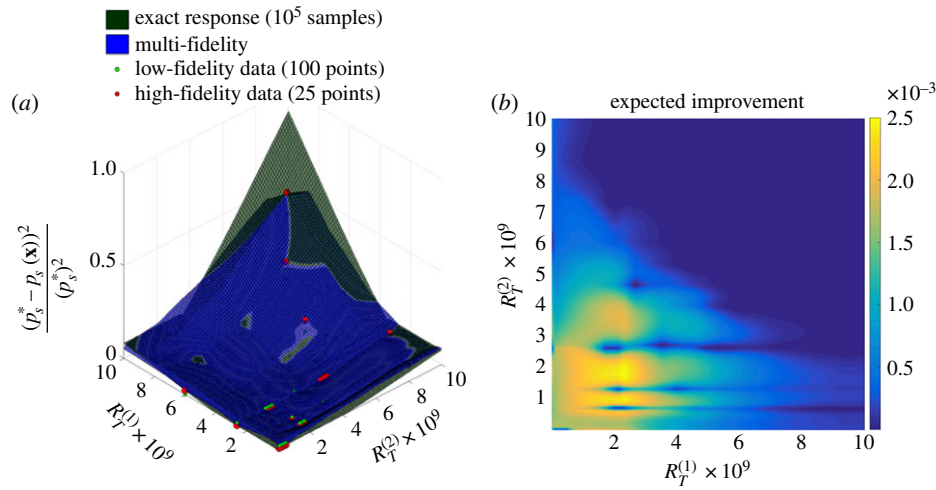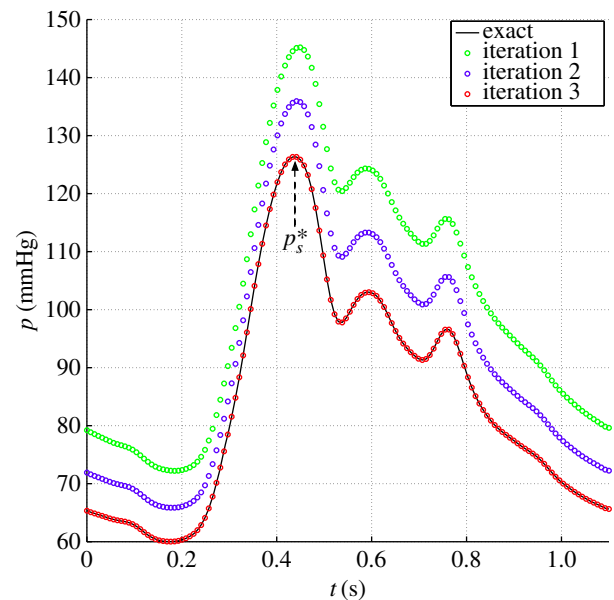
Figure 5. Blood flow in a symmetric Y-shaped bifurcation. Fourth iteration of the EGO algorithm. (*a*) Exact solution versus the multi-fidelity predictor for the inlet systolic pressure error, trained on 100 low-fidelity and 25 high-fidelity observations. (*b*) Map of the corresponding expected improvement. (Online version in colour.)

neighbourhood of local maximizers of the expected improvement. Then, the recursive multi-fidelity predictor is re-trained to absorb this new information, and this procedure is repeated until a termination criterion is met. Here, we have chosen to exit the iteration loop once the minimum of the predicted response surface (i.e. the relative error between the computed inlet systolic pressure and a target value) is less than $(p_s^* - p_s(\mathbf{x}))^2/(p_s^*)^2 < 10^{-3}$.

Figure 5 presents the resulting response surface of the error in the inlet systolic pressure after four iterations of the EGO algorithm. The predictor is trained on 100 low-fidelity and 25 high-fidelity realizations that clearly target resolving the region of the response surface where the global minimum is likely to occur. In figure 5, we also depict the spatial distribution of the expected improvement after the new data have been absorbed. By comparing figures 4 and 5, we observe that within four iterations of the EGO algorithm the expected improvement has been reduced by a factor greater than 5, effectively narrowing down the search for the global minimum in a small region of the response surface. In fact, four iterations of the EGO algorithm were sufficient to meet the termination criterion, and, therefore, identify the set of outflow resistance parameters that returns an inlet systolic pressure matching the target value of $p_s^* = 126$ mmHg. Here, we underline the robustness and efficiency of the proposed scheme, as convergence was achieved using only 25 samples of the high-fidelity nonlinear 1D-FSI model, supplemented by 100 inaccurate low-fidelity samples of the linearized 1D-FSI solver. In the absence of low-fidelity observations, the same level of accuracy was achieved with nine EGO iterations and 50 high-fidelity simulations—a twofold increase in computational cost.
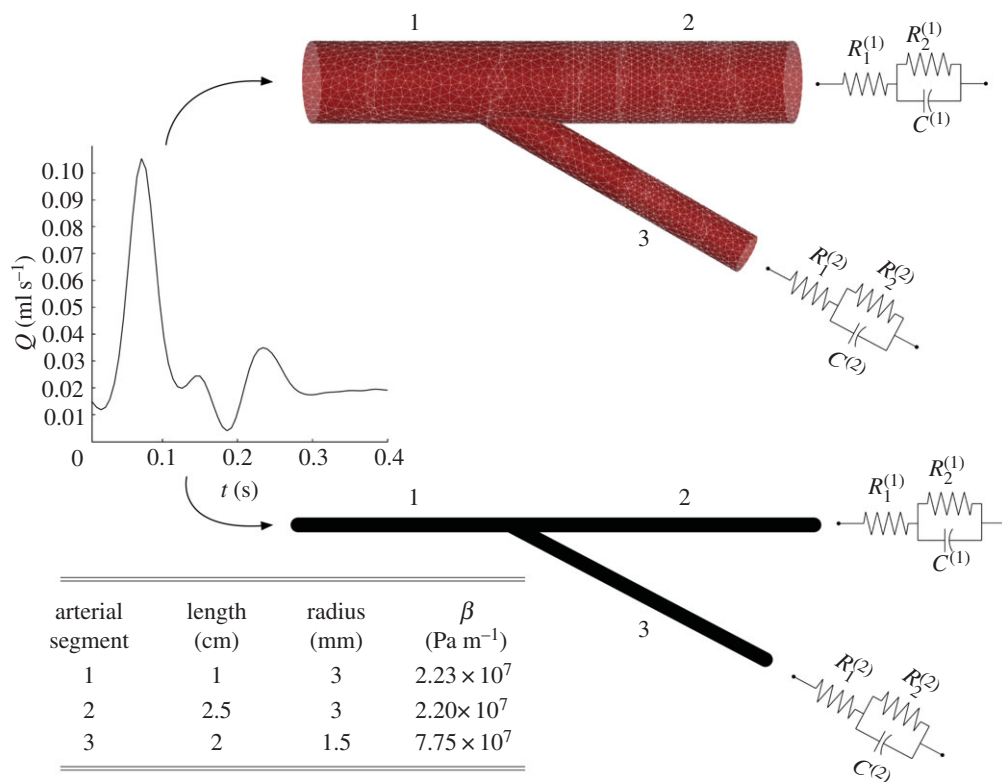
In table 1, we summarize the optimal predicted configuration for the total terminal resistances, along with the relative $\mathbb{L}_2$-error in the corresponding inlet pressure waveform compared with the optimal reference configuration (see $\mathbf{x}^*$ in figure 3), over one cardiac cycle, and for every iteration of the EGO algorithm. The convergence of the inlet pressure waveform to the reference solution is demonstrated in figure 6. As the EGO iterations pursue a match with the target inlet systolic pressure p_s^*, the pressure waveform gradually converges to the reference solution after each iteration of the optimizer.

Figure 6. Blood flow in a symmetric Y-shaped bifurcation. Convergence of the inlet pressure waveform for the first three EGO iterations. The exact solution corresponds to the reference results obtained from 10 000 high-fidelity samples, defining the target inlet systolic pressure p_s^*. (Online version in colour.)

Table 1. Blood flow in a symmetric Y-shaped bifurcation. Optimal terminal resistances and corresponding relative $\mathbb{L}_2$-error in the inlet pressure wave over one cardiac cycle, for each iteration of the EGO algorithm. The comparison is done with respect to the inlet pressure wave p_{opt} corresponding to the optimal configuration of resistances predicted by the reference solution obtained using 10 000 high-fidelity samples.

iteration no.	$R_T^{(1)} \times 10^9$	$R_T^{(2)} \times 10^9$	$\dfrac{\|p - p_{opt}\|_2}{\|p_{opt}\|_2}$
1	0.001	5.9612	0.1949
2	4.6479	2.6275	0.0945
3	2.0214	2.9275	0.0012
4	3.7183	3.6275	2.13×10^{-4}

arterial segment	length (cm)	radius (mm)	β (Pa m^{-1})
1	1	3	2.23×10^7
2	2.5	3	2.20×10^7
3	2	1.5	7.75×10^7

Figure 7. Blood flow in an asymmetric Y-shaped bifurcation. Sketch of the computational domain of the rigid 3D and 1D-FSI models. Flow is driven by a flow-rate wave at the inlet, while outflow boundary conditions are imposed through three-element Windkessel models. The inset table contains the geometric and elasticity properties used in this simulation. (Online version in colour.)

3.2. Calibration of a three-dimensional Navier–Stokes solver

This benchmark is designed to highlight an important aspect of the proposed framework, namely the significant efficiency gains one can obtain by employing low-fidelity models that exhibit correlation to the target high-fidelity model. In particular, we will examine two scenarios; the first involving low- and intermediate-fidelity models that are only weakly correlated to the high-fidelity model, while in the second case we perform an educated adjustment to obtain an intermediate-fidelity model that exhibits strong correlation with the high-fidelity solver. In both cases, we adopt a similar problem set-up to the one above, namely the calibration of outflow resistance parameters of a high-fidelity 3D Navier–Stokes blood flow solver in an asymmetric Y-shaped bifurcation targeting a systolic inlet pressure of $p_s* = 47.5$ mmHg (figure 7). Here, this is done by employing a 3D spectral/hp element solver (high fidelity), a nonlinear 1D-FSI solver (intermediate fidelity) and a linearized 1D-FSI solver (low fidelity). The rigid 3D computational domain is discretized by 45 441 tetrahedral spectral elements of polynomial order 4, and the simulation of one cardiac cycle takes approximately 6 h on 16 cores of an Intel Xeon X56502.67 GHz. The nonlinear and linearized 1D-FSI solvers represent each arterial segment using one discontinuous Galerkin element of polynomial order 4, and the resolution of one cardiac cycle takes 1 min, and 15 s, respectively, on a single core of an Intel Xeon X56502.67 GHz. It is therefore evident that significant computational gains can be achieved if we can calibrate the high-fidelity solver mainly relying on sampling the low- and intermediate-fidelity models.

3.2.1. Weakly correlated models

The aforementioned set-up leads to a discrepancy in the corresponding arterial wall models where a rigid wall is assumed in the 3D case in contrast to the linear elastic wall response in the 1D-FSI cases. Although all models provide a good agreement in predicting the flow-rate wave and pressure drop between the inlet and the outlets, this modelling discrepancy is primarily manifested in the predicted point-wise pressure distribution. As depicted in figure 8, the inlet pressure prediction of the 3D rigid model is up to 62% higher than the 1D-FSI models, a fact attributed to the inherently different nature of wave propagation between the two cases. This inconsistency is directly reflected in the inferred scaling factor ρ_t appearing in the recursive scheme of §2.2 that quantifies how correlated our quantity of interest is (systolic inlet pressure) between the different levels of fidelity. In particular, starting from an initial nested Latin hypercube experimental design in $\mathcal{X} = [10^6, 10^{10}]^2$ containing 20 low-fidelity, 10 intermediate-fidelity and two high-fidelity samples, we obtain a correlation coefficient of 0.91 between the low- and intermediate-fidelity levels, and a correlation of 0.12 between the intermediate- and high-fidelity solvers. These values confirm the strong correlation between levels 1 and 2 in our recursive inference scheme (linearized and nonlinear 1D-FSI models), while also revealing the very weak correlation between levels 2 and 3 (nonlinear 1D-FSI and the rigid 3D model). Essentially, this suggests that in this case one cannot expect the low- and intermediate-fidelity models to be highly informative in the construction of the high-fidelity response surface for $p_s = f(R_T^{(1)}, R_T^{(2)})$.

With our goal being the identification of the optimal configuration of $\mathbf{x} = [R_T^{(1)}, R_T^{(2)}]^T$ such that the relative error
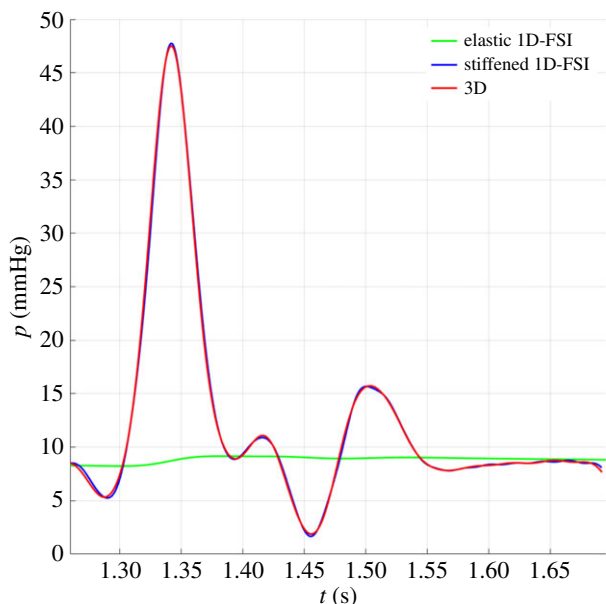
Figure 8. Stiffening the 1D-FSI wall model. Pressure wave predictions at the inlet of an asymmetric Y-shaped bifurcation using a purely elastic 1D-FSI solver, an appropriately stiffened 1D-FSI solver using equation (3.4), and a high-fidelity 3D Navier–Stokes solver in a rigid domain. (Online version in colour.)

between the predicted and target inlet systolic pressure is minimized, we may enter the EGO loop with the aforementioned initial nested design plan and iterate until convergence is achieved. Setting a stopping criterion of 10^{-3} on the expected improvement of the multi-fidelity predictor resulted in six EGO iterations, where in each iteration we added 10 low-fidelity, five intermediate-fidelity and one high-fidelity observation at the locations suggested by the highest local maxima of the expected improvement at each fidelity level. Therefore, a total of 80 low-fidelity, 60 intermediate-fidelity and eight high-fidelity simulations were performed in order to achieve a relative error less than 10^{-3}. The resulting response surface and corresponding expected improvement after the 6th iteration are presented in figure 9. Here, we must note that, despite the low cross-correlation between the high and lower fidelity levels, the multi-fidelity approach was still able to pay some dividends as, in order to achieve the same result by only sampling the high-fidelity solver, the EGO required 10 iterations.

3.2.2. Strongly correlated models

In order to demonstrate the full merits of adopting a multi-fidelity approach, we can enhance the correlation between the rigid 3D model and the nonlinear 1D-FSI model by artificially stiffening the arterial wall response of the 1D model. This is done by appropriately scaling the spring constant β in the 1D pressure–area relation (see equation (2.26)):

$$\beta = \beta_{scal} \frac{\sqrt{\pi}Eh}{(1-\nu^2)A_0}, \quad \beta_{scal} > 1. \tag{3.2}$$

Practically, we can find an appropriate scaling factor β_{scal}^* such that the 1D solver accurately reproduces the results predicted by the full 3D model. However, if the chosen value for β_{scal} is not close to the 'best-fit' value β_{scal}^*, the 1D solver under-predicts the pressure wave for $\beta_{scal} \ll \beta_{scal}^*$ and over-predicts it for $\beta_{scal} \gg \beta_{scal}^*$. The identification of the appropriate scaling factor is case dependent. However,

we can explicitly derive a formula for β_{scal}^* based on the physiological properties of the network. To this end, consider a single vessel in an arterial network and let R_t be the total peripheral resistance downstream, and R_1 the vessel's characteristic impedance:

$$R_1 \doteq \frac{c_0\rho}{A_0}, \quad c_0 = \sqrt{\frac{1}{2}\frac{\beta}{\rho}} A_0^{1/4}. \tag{3.3}$$

Then, we must have that $R_t - R_1 \geq 0$, and, therefore, we can scale β and stiffen the 1D model, respecting the upper bound:

$$R_1 \leq R_t \Rightarrow \beta_{scal} \leq \frac{2R_t^2 A_0^{3/2}}{\beta\rho}. \tag{3.4}$$

The total resistance R_t for each segment can be computed by the properties of the network downstream and the Windkessel resistance parameters imposed at the outlets. This effectively enables the nonlinear 1D-FSI solver to reproduce the pressure wave propagation predicted by the full 3D simulation (figure 8).

In order to estimate the optimal parametrization of the outflow boundary conditions, we enter the EGO loop using an initial nested Latin hypercube sampling plan containing 70 low-fidelity, 15 intermediate-fidelity and two high-fidelity observations. This initialization confirms a relatively weak correlation of 0.3 between the elastic low-fidelity and the stiffened intermediate-fidelity 1D-FSI solvers, but confirms a strong correlation of 0.95 between the stiffened 1D-FSI solver and the high-fidelity 3D Navier–Stokes model. As a consequence, the EGO algorithm converges to the desired accuracy level after only one iteration, as the 20 samples of the intermediate-fidelity solver have been very informative in the construction of the high-fidelity predictive distribution. In figure 10, we demonstrate the resulting response surface for the relative error in systolic inlet pressure after one iteration of the EGO loop, along with the resulting map of the expected improvement criterion. Leveraging the accurate predictions of the stiffened nonlinear 1D-FSI solver we were able to achieve a $4\times$ speed-up in calibrating the outflow parametrization compared with the case where the calibration was carried out by only evaluating the high-fidelity 3D solver.

3.3. Helmholtz equation in 10 dimensions

Consider a Helmholtz boundary value problem in $x \in [0,1]$

$$u_{xx} - \lambda^2 u = \sum_{k=1}^{K} w_k \ \sin(2\,k\pi x) \tag{3.5}$$

and

$$u(0) = 0, \quad u(1) = 0, \tag{3.6}$$

with $K = 10$ forcing term weights $w_k \in [0, 1]$. Now, assume that we are given a reference solution $u^*(x)$ that corresponds to weights generated by a spectrum $s(k) = \alpha \exp(\beta k^\gamma)$ with a random draw of parameters $\alpha = 0.5$, $\beta = 0.5$ and $\gamma = 0.5$ (figure 11). Our goal now is to identify the exact values of w_k that generated the solution $u^*(x)$, assuming no knowledge of the underlying spectrum. This translates to solving the following optimization problem:

$$\min_{w_k \in \mathbb{R}^{10}} \|u(x; w_k) - u^*(x)\|_2. \tag{3.7}$$
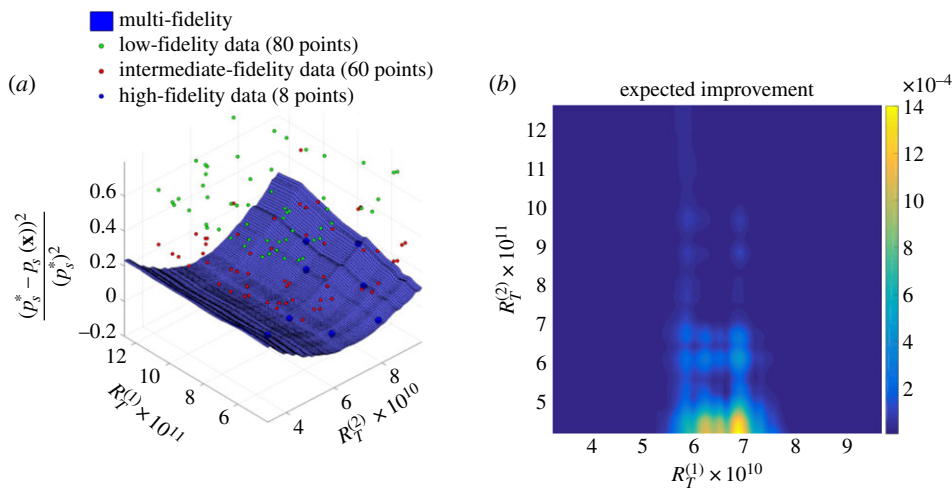
Figure 9. Blood flow in an asymmetric Y-shaped bifurcation. (*a*) Detailed response surface after six iterations of the EGO algorithm, and identification of the minimum relative error in the inlet systolic pressure as a function of the total resistance parameters imposed at the outlets. The response surface is constructed by blending multi-fidelity samples of a high-fidelity 3D solver (eight samples), an intermediate-fidelity nonlinear 1D-FSI model (60 samples) and a low-fidelity linearized 1D-FSI solver (80 samples). (*b*) Map of the corresponding expected improvement. (Online version in colour.)
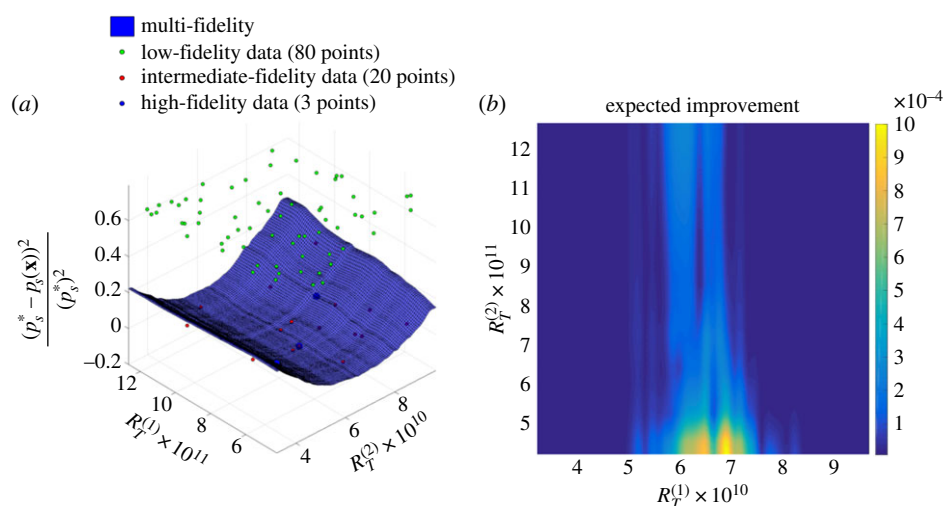

Figure 10. Blood flow in an asymmetric Y-shaped bifurcation. (*a*) Detailed response surface after one iteration of the EGO algorithm, and identification of the minimum relative error in the inlet systolic pressure as a function of the total resistance parameters imposed at the outlets. The response surface is constructed by blending multi-fidelity samples of a high-fidelity 3D solver (three samples), an intermediate-fidelity nonlinear 1D-FSI model (20 samples) and a low-fidelity linearized 1D-FSI solver (80 samples). (*b*) Map of the corresponding expected improvement. (Online version in colour.)
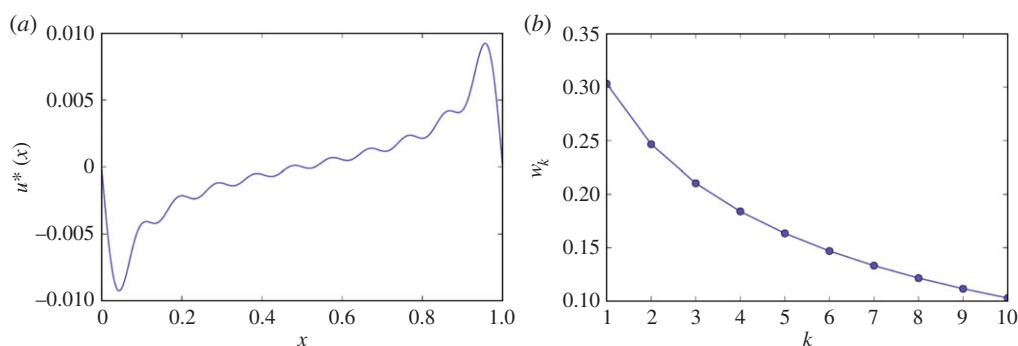
Figure 11. Inverse problem in 10 dimensions. (*a*) The reference solution $u^*(x)$ exhibits non-trivial boundary layers and high-frequency oscillations. (*b*) The weights spectrum corresponding to the reference solution. The slow decay implies that all weights w_k are actively contributing to the forcing term of equation (3.5). (Online version in colour.)

Here we consider three levels of fidelity corresponding to different discretization methods for solving equation (3.5). At the lowest fidelity level, we have chosen a fast centred finite difference scheme on a uniform grid containing 15 points. The medium-fidelity solver is a spectral collocation scheme using trigonometric interpolation on a uniform grid of 30
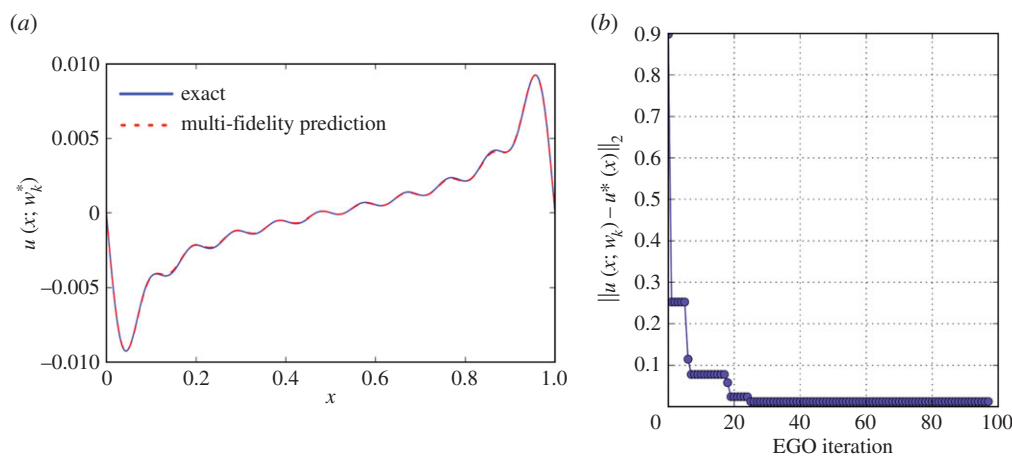
Figure 12. Inverse problem in 10 dimensions. (a) Comparison of the reference solution $u^*(x)$ and the solution corresponding to the optimal weights w_k^* identified by the EGO. (b) Value of the objective function in equation (3.7) at each EGO iteration. (Online version in colour.)

collocation points. This results to a more accurate scheme but is computationally more expensive as it involves the solution of a linear system. Finally, the highest fidelity data are generated by solving equation (3.5) exactly using a Galerkin projection on a basis consisting of 10 sine modes. Our goal now is to construct a 10-dimensional multi-fidelity surrogate for $u(x; w_k)$ mostly using realizations of the lower fidelity solvers supplemented by a few high-fidelity data, and solve equation (3.7) using Bayesian optimization.

Our workflow proceeds as follows. First, we create the training sets for a three-level multi-fidelity surrogate by generating weights w_k using random sampling, and compute the corresponding solution to equation (3.5) using the three different discretization methods outlined above. In particular, we use a nested experimental design consisting of $6K$ points at the lower fidelity level, $3K$ at the intermediate-fidelity level and K points at the highest fidelity level. This collection of input/output pairs provides the initialization of the EGO algorithm that aims to identify the combination of weights w_k that minimizes the objective function of equation (3.7). Then, at each EGO iteration new sampling points are appended to the training set of each fidelity level according to the expected improvement acquisition function.

Figure 12a shows the solution corresponding to the optimal combination of weights w_k identified by the multi-fidelity EGO algorithm after 100 iterations versus the reference target solution $u^*(x)$. As we can see in figure 12b, the algorithm is able to reduce the objective function down to 8.79×10^{-3} only after 25 iterations and identify a sensible configuration of weights using only a total of 35 evaluations of the highest fidelity model.

4. Discussion

We have presented a probabilistic, data-driven framework for inverse problems in haemodynamics. The framework is based on multi-fidelity information fusion and Bayesian optimization algorithms, allowing for the accurate construction of response surfaces through the meaningful integration of variable sources of information. Leveraging the properties of the proposed probabilistic inference schemes, we use the EGO algorithm [26] to perform model inversion using only a few samples of expensive high-fidelity model evaluations, supplemented with a number of cheap and potentially inaccurate

low-fidelity observations. This approach enables a computationally tractable global optimization framework that is able to efficiently 'zoom in' to the response surface in search of the global minimum, while exhibiting robustness to misspecification in the lower fidelity models.

All benchmark cases presented were deterministic inverse problems. Our main motivation was to provide a clear and intuitive presentation of the proposed framework, underlining the conditions under which a multi-fidelity approach can result in substantial computational gains. Such gains are expected to be more profound as the dimensionality of the parameter space is increased, as in such cases a very large number of samples will be required to perform model inversion. An application of this methodology to stochastic inverse problems requires two changes in the proposed methodology. First, the quantities to be estimated are no longer crisp numbers and our goal would be to estimate a proper parametrization of their distribution. For instance, if some properties are treated as independent Gaussian random variables, then the estimation problem would target the computation of their mean and variance, while if they were treated as correlated one should also account for their covariance. In general, a proper parametrization of uncertain parameters can be accommodated at the expense of increasing the dimensionality of the optimization problem. A second change incurred by stochasticity is related to the objective function itself. While in the deterministic case we consider minimizing a quadratic loss function, in the stochastic case the minimization objective would involve minimizing the distance to the mean and/or higher order moments, or, more generally, a moment matching criterion such as the Kullback–Leibler divergence [13] between the distribution of observed outputs and a target distribution. Accounting for variability in the model parameters is indeed a problem of great interest and we plan to address this in future work.

Summarizing, the proposed framework provides a principled workflow that allows one to limit evaluations of expensive high-fidelity solvers to a minimum, and extensively use cheap low-fidelity models to explore such vast parameter spaces. Extending and applying this methodology to realistic high-dimensional inverse problems requires the challenges of constructing response surfaces in high dimensions and potentially in the presence of massive datasets to be addressed. These challenges define an area of active research, and we refer the reader to the recent ideas put forth in [1] for a tractable solution plan.

References

1. Perdikaris P, Venturi D, Karniadakis GE. In press. Multi-fidelity information fusion algorithms for high dimensional systems and massive data-sets. *SIAM J. Sci. Comput.*

2. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. 1989 Design and analysis of computer experiments. *Stat. Sci.* **4**, 409–423. (doi:10.1214/ss/1177012413)

3. Perdikaris P, Venturi D, Royset J, Karniadakis GE. 2015 Multi-fidelity modelling via recursive co-kriging and Gaussian–Markov random fields. *Proc. R. Soc. A* **471**, 20150018. (doi:10.1098/rspa.2015.0018)

4. Kennedy MC, O'Hagan A. 2000 Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **87**, 1–13. (doi:10.1093/biomet/87.1.1)

5. Formaggia L, Quarteroni A, Veneziani A. 2010 *Cardiovascular mathematics: modeling and simulation of the circulatory system*, vol. 1. Milan, Italy: Springer-Verlag Italia.

6. Quick CM, Young WL, Noordergraaf A. 2001 Infinite number of solutions to the hemodynamic inverse problem. *Am. J. Physiol. Heart. Circ. Physiol.* **280**, H1472–H1479.

7. Bertoglio C, Moireau P, Gerbeau J-F. 2012 Sequential parameter estimation for fluid–structure problems: application to hemodynamics. *Int. J. Numer. Method. Biomed. Eng.* **28**, 434–455. (doi:10.1002/cnm.1476)

8. Blanco P, Watanabe S, Feijóo R. 2012 Identification of vascular territory resistances in one-dimensional hemodynamics simulations. *J. Biomech.* **45**, 2066–2073. (doi:10.1016/j.jbiomech.2012.06.002)

9. Lassila T, Manzoni A, Quarteroni A, Rozza G. 2013 A reduced computational and geometrical framework for inverse problems in hemodynamics. *Int. J. Numer. Method. Biomed. Eng.* **29**, 741–776. (doi:10.1002/cnm.2559)

10. Lombardi D. 2014 Inverse problems in 1D hemodynamics on systemic networks: a sequential approach. *Int. J. Numer. Method. Biomed. Eng.* **30**, 160–179. (doi:10.1002/cnm.2596)

11. Melani A. 2013 Adjoint-based parameter estimation in human vascular one dimensional models. PhD thesis, Politecnico di Milano, Italy.

12. Le Gratiet L, Garnier J. 2014 Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *Int. J. Uncertain. Quantif.* **4**, 365–386. (doi:10.1615/Int.J.UncertaintyQuantification.2014006914)

13. Rasmussen CE, Williams CKI. 2006 *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.

14. Snoek J, et al. 2015 Scalable Bayesian optimization using deep neural networks. (http://arxiv.org/abs/1502.05700)

15. Jones DR. 2001 A taxonomy of global optimization methods based on response surfaces. *J. Glob. Optim.* **21**, 345–383. (doi:10.1023/A:1012771025575)

16. Wiener N. 1949 *Extrapolation, interpolation, and smoothing of stationary time series*, vol. 2. Cambridge, MA: MIT Press.

17. Matheron G. 1963 Principles of geostatistics. *Econ. Geol.* **58**, 1246–1266. (doi:10.2113/gsecongeo.58.8.1246)

18. Cressie NA, Cassie NA. 1993 *Statistics for spatial data*, vol. 900. New York: Wiley.

19. Murray I, Adams RP, MacKay DJ. 2009 Elliptical slice sampling. (http://arxiv.org/abs/1001.0175)

20. Garbuno-Inigo A, DiazDelaO F, Zuev K. 2015 Gaussian process hyper-parameter estimation using parallel asymptotically independent Markov sampling. (http://arxiv.org/abs/1506.08010)

21. Snoek J, Larochelle H, Adams RP. 2012 Practical Bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, vol. 25 (eds F Pereira, CJC Burges, L Bottou, KQ Weinberger), pp. 2951–2959. See https://papers.nips.cc/paper/4522-practical-bayesian-optimization-of-machine-learning-algorithms.

22. Snelson E, Ghahramani Z. 2005 Sparse Gaussian processes using pseudo-inputs. In *Advances in neural information processing systems*, vol. 18 (eds Y Weiss, B Schölkopf, JC Platt), pp. 1257–1264. See https://papers.nips.cc/paper/2857-sparse-gaussian-processes-using-pseudo-inputs.

23. Hensman J, Fusi N, Lawrence ND. 2013 Gaussian processes for big data. (http://arxiv.org/abs/1309.6835)

24. Anitescu M, Chen J, Wang L. 2012 A matrix-free approach for solving the parametric Gaussian process maximum likelihood problem. *SIAM J. Sci. Comput.* **34**, A240–A262. (doi:10.1137/110831143)

25. Wilson AG, Nickisch H. 2015 Kernel interpolation for scalable structured Gaussian processes (kiss-gp). (http://arxiv.org/abs/1503.01057)

26. Jones DR, Schonlau M, Welch WJ. 1998 Efficient global optimization of expensive black-box functions. *J. Glob. Optim.* **13**, 455–492. (doi:10.1023/A:1008306431147)

27. Picheny V, Ginsbourger D, Richet Y, Caplin G. 2013 Quantile-based optimization of noisy computer experiments with tunable precision. *Technometrics* **55**, 2–13. (doi:10.1080/00401706.2012.707580)

28. Hernández-Lobato JM, Hoffman MW, Ghahramani Z. 2014 Predictive entropy search for efficient global optimization of black-box functions. In *Advances in neural information processing systems*, vol. 27 (eds Z Ghahramani, M Welling, C Cortes, ND Lawrence, KQ Weinberger), pp. 918–926. See https://papers.nips.cc/paper/5324-predictive-entropy-search-for-efficient-global-optimization-of-black-box-functions.

29. Bull AD. 2011 Convergence rates of efficient global optimization algorithms. *J. Mach. Learn. Res.* **12**, 2879–2904.

30. Vazquez E, Bect J. 2010 Convergence properties of the expected improvement algorithm with fixed mean and covariance functions. *J. Stat. Plan. Inference* **140**, 3088–3095. (doi:10.1016/j.jspi.2010.04.018)

31. Karniadakis GE, Sherwin S. 2013 *Spectral/hp element methods for computational fluid dynamics*. Oxford, UK: Oxford University Press.

32. Karniadakis GE, Israeli M, Orszag S. 1991 High-order splitting methods for the incompressible Navier–Stokes equations. *J. Comput. Phys.* **97**, 414–443. (doi:10.1016/0021-9991(91)90007-8)

33. Baek H, Karniadakis GE. 2011 Sub-iteration leads to accuracy and stability enhancements of semi-implicit schemes for the Navier–Stokes equations. *J. Comput. Phys.* **230**, 4384–4402. (doi:10.1016/j.jcp.2011.01.011)

34. Grinberg L, Cheever E, Anor T, Madsen J, Karniadakis GE. 2011 Modeling blood flow circulation in intracranial arterial networks: a comparative 3D/1D simulation study. *Ann. Biomed. Eng.* **39**, 297–309. (doi:10.1007/s10439-010-0132-1)

35. Grinberg L, Morozov V, Fedosov D, Insley JA, Papka ME, Kumaran K, Karniadakis GE. 2011 A new computational paradigm in multiscale simulations: application to brain blood flow. In *Proc. 2011 Int. Conf. for High Performance Computing, Networking, Storage and Analysis* (SC), Seattle, WA, 12–18 November, pp. 1–12. New York, NY: ACM.

36. Reymond P, Bohraus Y, Perren F, Lazeyras F, Stergiopulos N. 2011 Validation of a patient-specific one-dimensional model of the systemic arterial tree. *Am. J. Physiol. Heart Circul. Physiol.* **301**, H1173–H1182. (doi:10.1152/ajpheart.00821.2010)

37. Sherwin S, Franke V, Peiró J, Parker K. 2003 One-dimensional modeling of a vascular network in space-time variables. *J. Eng. Math*. **47**, 217–250. (doi:10.1023/B:ENGI.0000007979.32871.e2)

38. Pries A, Neuhaus D, Gaehtgens P. 1992 Blood viscosity in tube flow: dependence on diameter and hematocrit. *Am. J. Physiol. Heart Circ. Physiol*. **263**, H1770–H1778.

39. Perdikaris P, Karniadakis GE. 2014 Fractional-order viscoelasticity in one-dimensional blood flow models. *Ann. Biomed. Eng*. **42**, 1012–1023. (doi:10.1007/s10439-014-0970-3)

40. Fung Y-C. 1990 *Biomechanics*. Berlin, Germany: Springer.

41. Alastruey J, Parker K, Peiró J, Sherwin S. 2008 Lumped parameter outflow models for 1-D blood flow simulations: effect on pulse waves and parameter estimation. *Commun. Comput. Phys*. **4**, 317–336.

42. Grinberg L, Karniadakis GE. 2008 Outflow boundary conditions for arterial networks with multiple outlets. *Ann. Biomed. Eng*. **36**, 1496–1514. (doi:10.1007/s10439-008-9527-7)

43. Perdikaris P, Grinberg L, Karniadakis GE. 2014 An effective fractal-tree closure model for simulating blood flow in large arterial networks. *Ann. Biomed. Eng*. **43**, 1432–1442. (doi:10.1007/s10439-014-1221-3)

44. Snoek J, Swersky K, Zemel RS, Adams RP. 2014 Input warping for Bayesian optimization of non-stationary functions. (http://arxiv.org/abs/1402.0929)