Solving Bayesian inverse problems with expensive likelihoods using constrained Gaussian processes and active learning

To cite this article: Maximilian Dinkel *et al* 2024 *Inverse Problems* **40** 095008

View the article online for updates and enhancements.

Solving Bayesian inverse problems with expensive likelihoods using constrained Gaussian processes and active learning

Maximilian Dinkel[1,*] ⓘ **, Carolin M Geitner**[1] ⓘ **,**
Gil Robalo Rei[1] ⓘ **, Jonas Nitzler**[1] ⓘ **and Wolfgang A Wall**[1,2] ⓘ

[1] Institute for Computational Mechanics, Technical University of Munich, Munich, Germany
[2] Munich Data Science Institute, Technical University of Munich, Munich, Germany

E-mail: maximilian.dinkel@tum.de, carolin.geitner@tum.de, gil.rei@tum.de, jonas.nitzler@tum.de and wolfgang.a.wall@tum.de

Abstract

Solving inverse problems using Bayesian methods can become prohibitively expensive when likelihood evaluations involve complex and large scale numerical models. A common approach to circumvent this issue is to approximate the forward model or the likelihood function with a surrogate model. But also there, due to limited computational resources, only a few training points are available in many practically relevant cases. Thus, it can be advantageous to model the additional uncertainties of the surrogate in order to incorporate the epistemic uncertainty due to limited data. In this paper, we develop a novel approach to approximate the log likelihood by a constrained Gaussian process based on prior knowledge about its boundedness. This improves the accuracy of the surrogate approximation without increasing the number of training samples. Additionally, we introduce a formulation to integrate the epistemic uncertainty due to limited training points into the posterior density approximation. This is combined with a state of the art active learning strategy for selecting training points, which allows to approximate posterior densities in higher dimensions very efficiently. We demonstrate the fast convergence of our approach for a benchmark problem and infer a random field that is discretized by 30 parameters using only about 1000 model evaluations. In a practically

---

* Author to whom any correspondence should be addressed.

relevant example, the parameters of a reduced lung model are calibrated based on flow observations over time and voltage measurements from a coupled electrical impedance tomography simulation.

Keywords: Bayesian inverse problem, Gaussian process, active learning

1. Introduction

Computational modeling of physical phenomena has gained huge importance in all fields of engineering and applied science. However, most models rely on parameters that can not be measured directly and have to be calibrated. Some selected recent applications from our own group range from photoacoustic image reconstruction [1], calibration of multiphase porous media model parameters to predict tumor growth [2], the determination of diffusion parameters for laser powder bed fusion models [3] to the analysis of material parameters in coupled multi-physics biofilm models [4]. A common approach is to solve this problem by formulating it as a least-square minimization problem with a regularization term [5, 6]. However, this leads to ill-posed problems and a.o., the resulting point estimates do not capture measurement errors and can be misleading. By formulating the calibration task as a Bayesian inverse problem, the uncertainty of the measurements can be incorporated, and additionally, prior knowledge can be integrated. Rather than a point estimate, the Bayesian solution provides a probability distribution for the uncertain parameters given the noisy measurements. For practically relevant problems, the inference task can not be solved analytically. Nevertheless, the distribution can be approximated using sampling [7–9], particle [10–13], or variational methods [14–17]. Generally, sampling and particle methods require a large number of evaluations of the computational model, which can be prohibitive for expensive models. Most variational methods, on the other hand, need the gradient of the model outputs with respect to the uncertain parameters, which can be impractical for complex, potentially coupled solvers.

In scenarios where the likelihood function is intractable, Approximate Bayesian Computation (ABC) methods can be applied to estimate posterior distributions [18–20]. ABC methods generate samples from the posterior distribution by comparing observed data to data simulated from the model, typically using a distance metric and accepting samples that fall within a predefined threshold. While ABC is a powerful tool for complex models, it often requires a large number of model evaluations to generate sufficiently many accepted samples, especially in high-dimensional parameter spaces. Additionally, the choice of the distance metric and acceptance threshold can significantly affect the efficiency and accuracy of the results. In this work, we assume that the likelihood function is tractable but requires the evaluation of an expensive simulation model.

In order to reduce the number of computationally expensive forward model evaluations, various surrogate modeling approaches have emerged, ranging from polynomial chaos expansion [21, 22], adaptive sparse grid collocation methods [23], Bayesian neural networks [24], to Gaussian process regression [4, 25–28]. Nevertheless, in higher stochastic dimensions, surrogate methods often fail to capture the true posterior as the necessary number of forward evaluations increases exponentially with the number of input dimensions in order to cover the whole input space sufficiently.

In addition to conventional black box surrogate modeling approaches, there has been an increasing trend within the research community towards physics-constrained surrogate models. These models integrate prior knowledge of the underlying physics, governing equations,

or constraints into the modeling process. Due to the additional information also higher dimensional inverse problems or uncertainty quantification tasks can be tackled as demonstrated in [29–31]. However, for the application of these approaches, the surrogate model has to be adjusted to each investigated forward model, which is often inconvenient when, like in our case, quite different classes of problems should be tackled. Hence, our attempt in this paper is to develop and present efficient black box surrogate approaches.

In [27], it is shown how the epistemic uncertainty due to a limited number of forward solver evaluations can be incorporated into the posterior of the uncertain parameters using a Bayesian surrogate such as a Gaussian process. Similarly, [32] models the surrogate uncertainties and extends on this idea by choosing the training samples systematically in high posterior probability regions to reduce the necessary number of forward model calls. Nevertheless, for high dimensional and independent model outputs, even the Gaussian process model can become prohibitively expensive as an independent Gaussian process has to be trained and evaluated for each output. In such cases, it can be beneficial to approximate the (unnormalized) log likelihood function instead of the model outputs, as only a scalar value has to be approximated. Kandasamy *et al* [33] follow this approach and propose to select new training samples by maximizing the exponentiated variance of the surrogate model response. Similarly, [34] chooses new training samples by maximizing the entropy of the approximated posterior distribution. However, both approaches do not consider the boundedness of the log likelihood function, which can serve as substantial additional information for the surrogate approximation without the necessity for additional forward solver evaluations.

In this work, we approximate the log likelihood function by a Gaussian process to reduce the number of computationally expensive forward model evaluations. Further, we exploit the knowledge about the boundedness of the log likelihood function by constraining the surrogate response for improved predictive accuracy. To avoid overconfident and potentially inaccurate results, we incorporate the epistemic uncertainty of the surrogate due to limited training points into the posterior distribution of the uncertain parameters. The scalability of our approach to higher dimensions is additionally enabled by an adaptive sampling scheme adopted from [33].

The rest of the paper is structured as follows: In section 2, we introduce our novel approach to solve Bayesian inverse problems. In section 3, we investigate three numerical examples to test our approach. Finally, we provide a conclusion in section 4.[3]

2. Methodology

Consider a computationally expensive forward model $\mathcal{M} : \mathbb{R}^{n_x} \to \mathbb{R}^{n_{\text{obs}}}$ which takes the parameters $\boldsymbol{x} \in \mathbb{R}^{n_x}$ as inputs. For given noisy observations of the model output $\boldsymbol{y}_{\text{obs}} \in \mathbb{R}^{n_{\text{obs}}}$, the goal is to infer the corresponding input parameters $\boldsymbol{x}$. The one-dimensional vector $\boldsymbol{y}_{\text{obs}}$ can consist of measurements at different times, at different locations, of vector-valued quantities, or taken from different fields. For simplicity, we assume that the measurement error $\boldsymbol{\epsilon}$ is additive:

$$\boldsymbol{y}_{\text{obs}} = \mathcal{M}(\boldsymbol{x}) + \boldsymbol{\epsilon}. \tag{1}$$

In a Bayesian framework, we aim to find a posterior distribution of the input parameters:

$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}}) \propto p(\boldsymbol{y}_{\text{obs}}|\boldsymbol{x}) p(\boldsymbol{x}), \tag{2}$$

---

[3] Generally, we use plain letters for scalar, boldface letters for vector-valued, and capital letters for matrix-valued quantities. The code for this paper is implemented in the in-house code QUEENS [35].

where $p(\boldsymbol{y}_{\text{obs}}|\boldsymbol{x})$ denotes the likelihood and $p(\boldsymbol{x})$ the prior distribution. Here, we model the likelihood by a normal distribution with covariance Σ_n:

$$p\left(\boldsymbol{y}_{\text{obs}}|\boldsymbol{x}\right) \propto \exp\left(-\frac{1}{2}\left\|\boldsymbol{y}_{\text{obs}} - \mathcal{M}\left(\boldsymbol{x}\right)\right\|_{\Sigma_n}^2\right). \tag{3}$$

This is just one popular choice and the extension of our approach to other distributions that are based on a distance measure between $\boldsymbol{y}_{\text{obs}}$ and $\mathcal{M}(\boldsymbol{x})$ is straightforward.

2.1. Surrogate Modeling of the Log Likelihood

For complex forward models, the posterior distribution in (2) is typically not available analytically. Furthermore, classical sampling methods such as Markov Chain Monte Carlo (MCMC) would require an enormous amount of likelihood evaluations, which is infeasible for computationally expensive forward models. Therefore, we propose to approximate the log likelihood (up to a constant) by a surrogate model in order to reduce the number of likelihood evaluations drastically compared to classical sampling methods. To be precise, the exponent of the likelihood function $f(\boldsymbol{x}) = -\frac{1}{2}\|\boldsymbol{y}_{\text{obs}} - \mathcal{M}(\boldsymbol{x})\|_{\Sigma_n}^2$ is approximated by a Gaussian process $f_s \sim \mathcal{GP}(\mu(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$, which is defined by a mean function $\mu(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ [36]:

$$\begin{aligned}\mu(\boldsymbol{x}) &= \mathbb{E}\left[f_s(\boldsymbol{x})\right] \\ k(\boldsymbol{x}, \boldsymbol{x}') &= \mathbb{E}\left[(f_s(\boldsymbol{x}) - \mu(\boldsymbol{x}))(f_s(\boldsymbol{x}') - \mu(\boldsymbol{x}'))\right].\end{aligned} \tag{4}$$

The distribution of the function values $\boldsymbol{f}_s$ at N given points $X \in \mathbb{R}^{N \times n_x}$ can be derived as:

$$\boldsymbol{f}_s|X \sim \mathcal{N}\left(\boldsymbol{\mu}\left(X\right), K\left(X, X\right)\right), \tag{5}$$

where the covariance matrix is evaluated as $K(X, X)_{i,j} = k(X_i, X_j)$. For a given set of training observations $\mathcal{D} = \{(\boldsymbol{x}^{(i)}, f^{(i)})|i = 1, \ldots, N)\}$ the function value f_s at a point $\boldsymbol{x}$ follows a normal distribution [36, 37]:

$$\begin{aligned}f_s|\boldsymbol{x}, \mathcal{D} &\sim \mathcal{N}\left(m, s^2\right) \\ m &= \mu\left(\boldsymbol{x}\right) + k\left(\boldsymbol{x}, X\right)\left[K\left(X, X\right) + \sigma_f^2 I_N\right]^{-1}\left(\boldsymbol{f} - \boldsymbol{\mu}\left(X\right)\right), \\ s^2 &= k\left(\boldsymbol{x}, \boldsymbol{x}\right) - k\left(\boldsymbol{x}, X\right)\left[K\left(X, X\right) + \sigma_f^2 I_N\right]^{-1}k\left(X, \boldsymbol{x}\right)\end{aligned} \tag{6}$$

where σ_f^2 denotes an independent and identically distributed Gaussian measurement noise on the observed function values $\boldsymbol{f}$. Here, we only consider deterministic forward solvers, meaning we can observe the log likelihood function without uncertainty, as we do not consider model inaccuracies. Nevertheless, we optimize the variance term as this can often lead to improved predictive accuracy in sparse data scenarios [38]. As a covariance function, we use the squared exponential kernel with signal variance σ_s^2 and characteristic lengthscales $\boldsymbol{l} \in \mathbb{R}^{n_x}$:

$$k\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \sigma_s^2 \exp\left(-\sum_{i=1}^{n_x}\frac{(x_i - x_i')^2}{2l_i^2}\right), \tag{7}$$

but there is no restriction to choose another kernel. Besides the training set $\mathcal{D}$, the surrogate response f_s is also dependent on the hyperparameters $\boldsymbol{\theta} = [\sigma_f^2, \sigma_s^2, \boldsymbol{l}]$.

Evidently, the approximated unnormalized log likelihood function $f(\boldsymbol{x}) = -\frac{1}{2}\|\boldsymbol{y}_{\text{obs}} - \mathcal{M}(\boldsymbol{x})\|_{\Sigma_n}^2$ can not be larger than zero as $\|\boldsymbol{y}_{\text{obs}} - \mathcal{M}(\boldsymbol{x})\|_{\Sigma_n}^2 \geqslant 0$. We can even find a lower estimate $\hat{b} < 0$ for the upper bound of $f(\boldsymbol{x})$, which further improves the predictive accuracy of the surrogate. The estimate is based on the distribution of

$$b(\boldsymbol{y}_{\text{obs}}) = -\frac{1}{2}\|\boldsymbol{y}_{\text{obs}} - \mathcal{M}(\boldsymbol{x})\|_{\Sigma_n}^2 = -\frac{1}{2}\gamma, \qquad \gamma \sim \chi^2(n_{\text{obs}}), \tag{8}$$

where χ^2 denotes the chi-squared distribution (see appendix). We choose a conservative estimate for the upper bound $\hat{b}$ that satisfies $Pr(b \leqslant \hat{b}) = 0.95$, meaning that the probability that the actual upper bound b is larger than $\hat{b}$ is only 5%. In the unlikely event that we encounter a training point with a value larger than $\hat{b}$, we update the upper bound to this value. The constraint is enforced on the Gaussian process based on virtual observation points as described in [37, 39]. To enable efficient predictions, the approach is simplified by using only one virtual observation point $\boldsymbol{x}^v$ that coincides with the point $\boldsymbol{x}$, meaning that we do not enforce the constraint in the whole domain, but only at $\boldsymbol{x}^v = \boldsymbol{x}$. Conditioned on the event $\zeta := f_s(\boldsymbol{x}^v) \leqslant \hat{b}$ this leads to a truncated normal predictive distribution (see appendix):

$$p(f_s|\boldsymbol{x}, \zeta, \boldsymbol{\theta}, \mathcal{D}) = \mathcal{TN}(m, s^2, -\infty, \hat{b}))$$
$$= \frac{1}{0.5(1 + \text{erf}(\frac{\hat{b}-m}{\sqrt{2}s}))} \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(f_s - m)^2}{2s^2}\right) \mathbb{I}_{f_s \leqslant \hat{b}}, \tag{9}$$

where the mean m and the variance s^2 are defined as in equation (6) and depend on the hyperparameters $\boldsymbol{\theta}$. The posterior of the hyperparameters can be obtained using Bayes' rule:

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{10}$$

with the likelihood $p(\mathcal{D}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{0}, K(X,X) + \sigma_f^2 I_N)$ [36]. For the prior $p(\boldsymbol{\theta})$, we choose independent exponential distributions with small rate parameters, such that the prior is rather uninformative but still has a regularizing effect. The posterior $p(\boldsymbol{\theta}|\mathcal{D})$ is analytically intractable, but we can draw samples from it using sampling methods such as Hamiltonian Monte Carlo [8, 9]. The computational cost of evaluating the likelihood $p(\mathcal{D}|\boldsymbol{\theta})$ increases cubically with the number of training samples N due to the inversion of the covariance matrix $K(X,X)$. With more elaborate techniques, the computational cost can be reduced to $\mathcal{O}(N^2)$ [40]. However, this is beyond the scope of this work, as we only consider problems where the evaluation of the forward model for more than several thousand times would be infeasible. Similar to [37], we neglect the dependency of the hyperparameters on the constraint ζ to reduce the numerical complexity of the surrogate model, i.e. $p(\boldsymbol{\theta}|\mathcal{D}, \zeta) \approx p(\boldsymbol{\theta}|\mathcal{D})$. With the approximated posterior $p(\boldsymbol{\theta}|\mathcal{D}, \zeta)$ the hyperparameters can be marginalized to propagate their uncertainty to the posterior of f_s:

$$p(f_s|\boldsymbol{x}, \zeta, \mathcal{D}) = \int p(f_s|\boldsymbol{x}, \zeta, \boldsymbol{\theta}, \mathcal{D})p(\boldsymbol{\theta}|\mathcal{D}, \zeta)\,\mathrm{d}\boldsymbol{\theta}$$
$$\approx \frac{1}{n_\theta}\sum_{j=1}^{n_\theta} p(f_s|\boldsymbol{x}, \zeta, \boldsymbol{\theta}_j, \mathcal{D}), \quad \boldsymbol{\theta}_j \sim p(\boldsymbol{\theta}|\mathcal{D}). \tag{11}$$

This is referred to as *fully Bayesian Gaussian process regression*, whereas the approach of taking a maximum likelihood estimate for the hyperparameters is called *type II maximum*

likelihood or ML-II [41]. Analogously we will refer to the approach using the maximum-a-posteriori (MAP) estimate for the hyperparameters as MAP-II.

In the following, we derive an approach to incorporate the surrogate uncertainties into the approximation of the posterior $p(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}})$ in (2). As we approximate the likelihood in the log space, we have to transform the uncertainties into the original space where we solve the Bayesian inverse problem. Therefore we introduce the random variable $g := \exp(f_s)$, which serves as our approximation of the unnormalized likelihood. Using the rule for the transformation of random variables, the probability density function for g can be derived as (see appendix):

$$p\left(g|\boldsymbol{x},\zeta,\mathcal{D}\right) \approx \frac{1}{n_\theta} \sum_{j=1}^{n_\theta} \frac{2}{g\sqrt{2\pi}s_j\left(1+\mathrm{erf}\left(\frac{\hat{b}-m_j}{\sqrt{2}s_j}\right)\right)} \exp\left(-\frac{(\ln g - m_j)^2}{2s_j^2}\right) \mathbb{I}_{0\leqslant g\leqslant\exp(\hat{b})}, \quad (12)$$

where m_j and s_j^2 are defined as the mean and the variance in equation (6) for a given hyperparameter sample $\boldsymbol{\theta}_j$. The cumulative density function (CDF) of this distribution is given as:

$$F_g\left(g|\boldsymbol{x},\zeta,\mathcal{D}\right) \approx \frac{1}{n_\theta} \sum_{j=1}^{n_\theta} \frac{1+\mathrm{erf}\left(\frac{\ln g - m_j}{\sqrt{2}s_j}\right)}{1+\mathrm{erf}\left(\frac{\hat{b}-m_j}{\sqrt{2}s_j}\right)}. \quad (13)$$

The CDF provides the probability that the unnormalized likelihood is smaller or equal to a certain value g. So we can prescribe $F_g(g|\boldsymbol{x},\zeta,\mathcal{D}) = q$, solve the equation for the corresponding value g, and use this value as our approximator for the unnormalized likelihood $Zp(\boldsymbol{y}_{\mathrm{obs}}|\boldsymbol{x})$, where Z is a constant. In other words, the inverse CDF $F_g^{-1}(q) = g$ gives us an approximation g of the unnormalized likelihood, which we are certain about with probability q, that the true unnormalized likelihood is smaller or equal to this value. With increasing quantile q the values of the approximation of the unnormalized likelihood grow in regions with high epistemic uncertainties due to limited training points. The idea is here that we prefer to overestimate the likelihood instead of underestimating it in order to reduce bias and explore regions with high epistemic uncertainty due to limited data points in an adaptive sampling setting (see section 2.3). Analogously, the inverse CDF or upper confidence bound is also commonly used in Bayesian optimization as an acquisition function to choose new training points for the Gaussian process [42–44]. Although the inverse CDF of (13) can not be written in closed form, it can be solved efficiently using iterative methods such as Newton's or Chandrupatla's method [45]. As the cost of evaluating our estimator of the likelihood increases linearly with the number of hyperparameter samples n_θ, we also consider a more efficient approximator based on the MAP estimate of $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{\theta}} p\left(\boldsymbol{\theta}|\mathcal{D}\right). \quad (14)$$

If we approximate the posterior of the hyperparameters with $p(\boldsymbol{\theta}|\mathcal{D}) \approx \delta(\boldsymbol{\theta}-\boldsymbol{\theta}_{\mathrm{MAP}})$ the inverse CDF can be written as:

$$F_g^{-1}(q) \approx \exp\left[\mathrm{erf}^{-1}\left\{q\left(1+\mathrm{erf}\left(\frac{\hat{b}-m_{\mathrm{MAP}}}{\sqrt{2}s_{\mathrm{MAP}}}\right)\right)-1\right\}\sqrt{2}s_{\mathrm{MAP}}+m_{\mathrm{MAP}}\right]. \quad (15)$$

Finally, the posterior distribution of the input parameters in (2) is approximated as:

$$p\left(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}},\mathcal{D}\right) \tilde{\propto} F_g^{-1}(q)\,p\left(\boldsymbol{x}\right). \quad (16)$$
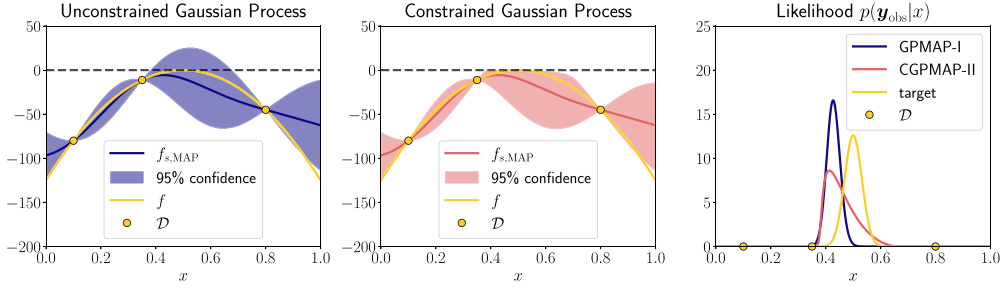
Figure 1. Approximation of the likelihood $p(\mathbf{y}_{\text{obs}}|\mathbf{x}) = \mathcal{N}(0.5, 0.001)$ with three training points. The left plot shows the unconstrained Gaussian process approximation of the log likelihood (up to a constant). The middle plot shows the constrained Gaussian process approximation. The right plot shows the target likelihood and the approximations using the GPMAP-I and CGPMAP-II approach with $q = 0.90$.
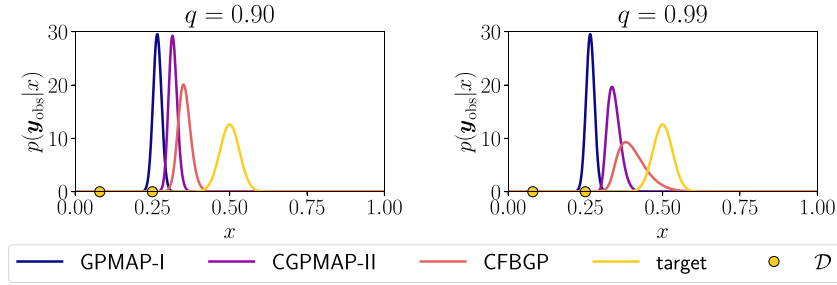

Figure 2. Approximation of the likelihood $p(\mathbf{y}_{\text{obs}}|\mathbf{x}) = \mathcal{N}(0.5, 0.001)$ with two training points. It shows the target likelihood and the approximations using the GPMAP-I, CGPMAP-II, and CFBGP ($n_\theta = 100$) approach with $q = 0.90$ (left) and $q = 0.99$ (right).

In the following, we denote the approach using the MAP estimate of the unconstrained Gaussian process predictions f_s, and therefore not considering any surrogate uncertainties, as GPMAP-I. Whereas, if the Gaussian process is constrained, the MAP estimate for the hyperparameters is used, and the uncertainties are incorporated using the inverse CDF in (15) we refer to this approach as CGPMAP-II. Accordingly, if the hyperparameters are marginalized we refer to the constrained fully Bayesian Gaussian process approach as CFBGP.

In figure 1, the approximation of a generic one-dimensional Gaussian likelihood $p(\mathbf{y}_{\text{obs}}|\mathbf{x}) = \mathcal{N}(0.5, 0.001)$ using three training points is shown. The left plot shows the unconstrained Gaussian process approximation of the unnormalized log likelihood f. The middle plot shows the constrained Gaussian process approximation, which narrows the confidence intervals and reduces the epistemic uncertainty. The right plot shows the corresponding true target likelihood and the approximations using the GPMAP-I and CGPMAP-II approach with $q = 0.90$. It can be seen that due to the consideration of the surrogate uncertainty, the support of the likelihood approximation is wider and covers the whole support of the target likelihood when using the CGPMAP-II approach compared to GPMAP-I. The approximation using GPMAP-I is overconfident and can lead to more biased posterior approximations.

Figure 2 visualizes the different approximations of the same likelihood but with only two training points. The approximation is most biased when using the GPMAP-I approach and the

least biased when using the CFBGP approach. The bias is also reduced when using a higher quantile $q = 0.99$ compared to $q = 0.90$ for CGPMAP-II or CFBGP.

2.2. Sequential Monte Carlo for Bayesian Inference

Generally, the normalized posterior in (16) can not be solved analytically. Nevertheless, we can obtain a particle representation of the form [27]:

$$p\left(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}, \mathcal{D}\right) \approx \sum_{i=1}^{n_p} w_i \delta\left(\boldsymbol{x} - \boldsymbol{x}_i\right) \tag{17}$$

using Sequential Monte Carlo (SMC). Several variations of this method have been studied, but the key idea is to start from a particle presentation of the known prior distribution and sequentially move these particles toward the posterior distribution via bridging distributions. Most algorithms consist of an importance sampling step between the bridging distributions, a resampling step based on the weights w_i, and rejuvenation steps to mitigate degenerate weights. The rejuvenation steps are usually performed using MCMC. For technical details the interested reader is referred to [10–13]. Here we use the waste-free algorithm of the Python particles library, which uses the samples of the intermediate MCMC steps as particles [11, 46]. Nevertheless, also other methods such as variational inference or gradient-based Monte Carlo methods can be used to compute an approximation of the posterior distribution in (16).

2.3. Adaptive sampling strategy

The accuracy of the approximated posterior distribution of the input parameters in (2) is strongly influenced by the choice of the training points for the surrogate. A common approach is to draw (quasi-)random samples from the prior distribution and use them as training points. In low stochastic dimensions, this can work well. However, the necessary number of training points increases exponentially with the number of input dimensions with this approach. For expensive forward models, this can become infeasible rapidly. Numerous criteria have been proposed for the selection of optimal training points, such as maximizing the mutual information between the chosen points and the points that are not selected [47] or minimizing the posterior integrated variance [48].

The main difference between the use case here and most other regression problems is that a high accuracy of the surrogate model is only required in regions of $\boldsymbol{x}$ with high posterior probability. In [34], the authors use this insight and propose to choose a new training point that maximizes the entropy of the posterior distribution approximation of $\boldsymbol{x}$, which is a random variable itself. Similarly, [33] proposes to choose a new training point where the exponentiated variance of the Gaussian process that approximates the log of the joint probability $p(\boldsymbol{x}, \boldsymbol{y}_{\mathrm{obs}})$ is maximal.

Here, we adapt the approach proposed in [32] as shown in algorithm 1: we first draw a few initial samples from the prior distribution to train the surrogate. Here, this means drawing samples from the hyperparameter posterior $p(\boldsymbol{\theta}|\mathcal{D})$ or calculating $\boldsymbol{\theta}_{\mathrm{MAP}}$. Based on this surrogate, we can draw samples from the approximated posterior distribution in (16) and use a few of these samples as additional training samples. Again the surrogate is constructed using all evaluated samples of the forward model to draw samples from the approximated posterior distribution. This process is repeated until the difference between the approximated posterior distributions between the iterations is negligible.

Algorithm 1. Adaptive sampling strategy.

1. Set $j = 0$
2. Draw samples from prior distribution and evaluate forward model to obtain training set
   $\mathcal{D}_j = \{(\boldsymbol{x}_i, f(\boldsymbol{x}_i))\}_{i=1}^{N_0}, \boldsymbol{x}_i \sim p(\boldsymbol{x})$
3. Train surrogate model based on $\mathcal{D}_j$
4. Use SMC to obtain particle representation of intermediate posterior distribution
   $p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}}, \mathcal{D}_j) \approx \sum_{i=1}^{n_p} w_i \delta(\boldsymbol{x} - \boldsymbol{x}_i)$
5. **do**
   i. Set $j = j + 1$
   ii. Draw samples from intermediate posterior distribution and evaluate forward model to
       obtain training set $\hat{\mathcal{D}}_j = \{(\boldsymbol{x}_i, f(\boldsymbol{x}_i))\}_{i=1}^{N_{\text{ada}}}, \boldsymbol{x}_i \sim p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}}, \mathcal{D}_{j-1})$
   iii. Train surrogate model based on $\mathcal{D}_j = \{\hat{\mathcal{D}}_j, \mathcal{D}_{j-1}\}$
   iv. Use SMC to obtain particle representation of intermediate posterior distribution
       $p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}}, \mathcal{D}_j) \approx \sum_{i=1}^{n_p} w_i \delta(\boldsymbol{x} - \boldsymbol{x}_i)$
   while $\hat{D}_{\text{CS}}(p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}}, \mathcal{D}_j), p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}}, \mathcal{D}_{j-1})) > \alpha_{\text{tol}}$

To assess the difference between the particle representation of two distributions, we introduce a discrepancy measure that is reliable, computationally cheap, and symmetric. First, we compute a kernel density estimate (KDE) of the particle representation of the distribution with a Gaussian kernel. Subsequently, we can calculate the Cauchy–Schwarz divergence between the two KDEs in closed form [49]. Still, for a large number of mixture components, meaning particles, this becomes computationally expensive, and in high dimensions, a large number of particles is necessary for an accurate representation of the distribution. Hence, we only compute the divergence measure between the marginal distributions and just use a subset of 5000 particles, which is sufficient for an accurate representation of the distribution in one dimension. In essence, we evaluate the maximal Cauchy–Schwarz divergence between the marginal distributions of the two compared distributions:

$$
\hat{D}_{\text{CS}}(p(\boldsymbol{x}), q(\boldsymbol{x})) := \max_{x_j \in \{1, \dots, n_x\}} D_{\text{CS}}(p(x_j) q(x_j))
$$

$$
= \max_{x_j \in \{1, \dots, n_x\}} -\ln\left(\frac{\int p(x_j) q(x_j) \, dx_j}{\sqrt{\int p(x_j)^2 \, dx_j \int q(x_j)^2 \, dx_j}}\right), \quad (18)
$$

using KDE approximations of the particle representations. In the following, we use the discrepancy measure defined in (18) as a convergence measure by evaluating the measure between intermediate posterior distributions (see algorithm 1). Further, we use the measure to assess the accuracy of posterior approximations by comparing them to a reference posterior approximation.

The adaptive sampling strategy using CFBGP with $n_\theta = 100, N_0 = 5, N_{\text{ada}} = 5$ is visualized in figure 3. The current training points $\mathcal{D}_j$ and newly drawn training points $\hat{\mathcal{D}}_j$ are indicated by white dots for each iteration j. For the underlying problem the prior distribution $p(\boldsymbol{x})$ is uniform with bounds $[-5, 5]$ and the likelihood is taken from [15]: $p(\boldsymbol{y}_{\text{obs}}|\boldsymbol{x}) \propto \exp(-U(\boldsymbol{x}))$, where $U(\boldsymbol{x}) = \frac{1}{2}\left(\frac{\|\boldsymbol{x}\| - 2}{0.4}\right)^2 - \ln(\exp\{-\frac{1}{2}[\frac{x_1 - 2}{0.6}]^2\} + \exp\{-\frac{1}{2}[\frac{x_1 + 2}{0.6}]^2\})$.

Using adaptive sampling, the number of forward model evaluations can be reduced drastically compared to naively drawing samples from the prior distributions. Nevertheless, there is a chance that not all regions of the true posterior distribution with significant probability are
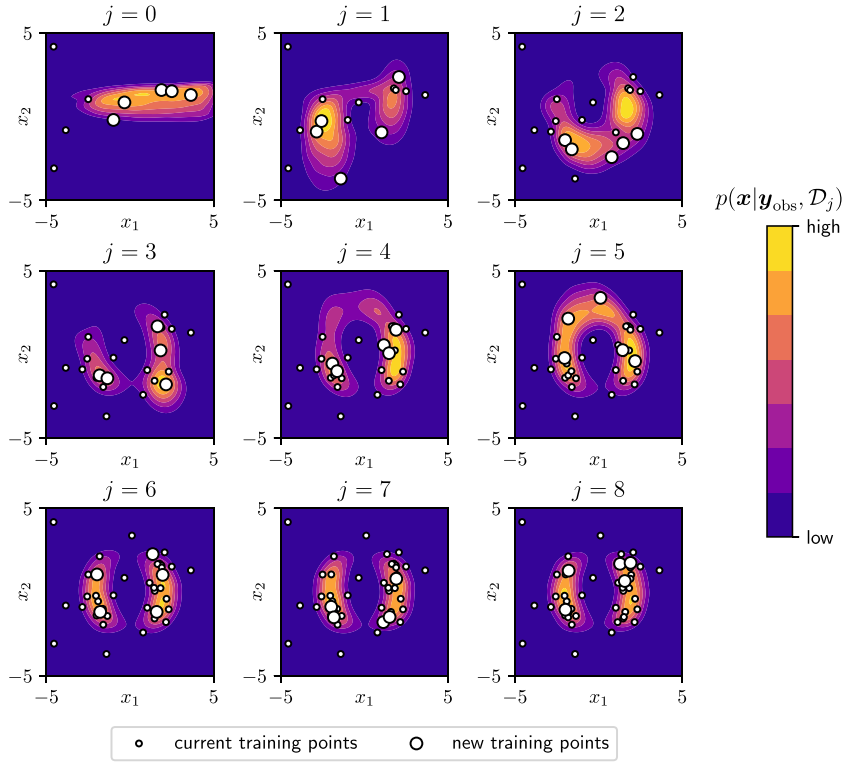
Figure 3. Visualization of adaptive sampling strategy using CFBGP with $n_\theta = 100$, $N_0 = 5$, $N_{ada} = 5$. Current training points $\mathcal{D}_j$ and new training points $\hat{\mathcal{D}}_j$ are shown for each iteration j. The prior distribution $p(\boldsymbol{x})$ is uniform with bounds $[-5,5]$ and the likelihood is taken from [15]: $p(\boldsymbol{y}_{\text{obs}}|\boldsymbol{x}) \propto \exp(-U(\boldsymbol{x}))$, where $U(\boldsymbol{x}) = \frac{1}{2}\left(\frac{\|\boldsymbol{x}\|-2}{0.4}\right)^2 - \ln\left(\exp\left\{-\frac{1}{2}\left[\frac{x_1-2}{0.6}\right]^2\right\} + \exp\left\{-\frac{1}{2}\left[\frac{x_1+2}{0.6}\right]^2\right\}\right)$.

discovered. This risk can be reduced by choosing a high quantile q in (13) to weigh the surrogate uncertainties stronger. This underlines the importance of modeling the uncertainties of the surrogate model in order to capture the true posterior distribution with a feasible amount of forward solver evaluations. In the adaptive sampling setting, the quantile q can be viewed as a trade-off factor between exploitation and exploration, as high values of q lead to more exploration, and low values of q lead to more exploitative sampling. Additionally, we would like to point out that the used adaptive sampling strategy allows for batch-wise evaluations of the forward model and consequently the possibility of using the available computational resources efficiently.

3. Numerical examples

In the following, we test our approach on three numerical examples. First, we investigate the convergence of the algorithm to the true solution for a generic Gaussian likelihood. Subsequently, we infer the diffusivity field for a diffusion problem and finally, we look at a practically relevant example with a coupled forward solver.
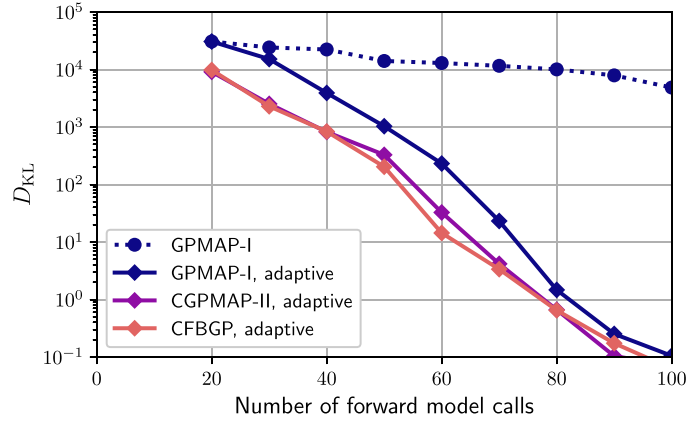
Figure 4. KL-divergence over the number of solver calls for $n_x = 10$. The number of initially and adaptively drawn training samples is set to $N_0 = 20$ and $N_{\mathrm{ada}} = 10$.

3.1. Artificial inverse problem with generic Gaussian likelihood

To test the convergence to the true posterior, we analyze an artificial inverse problem where the posterior distribution of the uncertain parameters can be derived analytically. The prior is given as $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{0}, I_{n_x})$ and the likelihood follows a normal distribution $p(\boldsymbol{y}_{\mathrm{obs}}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{a}, \sigma^2 I_{n_x})$, where the mean vector is sampled from a standard normal distribution $\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{0}, I_{n_x})$ and the likelihood variance is set to $\sigma^2 = 10^{-4}$. The posterior distribution can be derived as $p(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}) = \mathcal{N}(\frac{1}{1+\sigma^2}\boldsymbol{a}, \frac{\sigma^2}{\sigma^2+1}I_{n_x})$. We quantify the accuracy of the approximated posterior distribution $p(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}, \mathcal{D})$ by evaluating the Kullback–Leibler (KL) divergence [50]:

$$D_{\mathrm{KL}}\left(p\left(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}\right), p\left(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}, \mathcal{D}\right)\right) = \int p\left(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}\right) \ln \frac{p\left(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}\right)}{p\left(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}, \mathcal{D}\right)} \mathrm{d}\boldsymbol{x}. \tag{19}$$

To evaluate the KL divergence, we approximate the particle representation of $p(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}, \mathcal{D})$ by a normal distribution, parameterized with the empirical mean and covariance of the particles.

Subsequently, we can compute the KL divergence between the approximated and the true posterior analytically. In figure 4, the KL-divergence is plotted over the number of solver calls for $n_x = 10$. For better comparability, the KL-divergence is averaged over ten runs of the algorithm with different training point initializations drawn from the prior. The number of initially and adaptively drawn training samples is set to $N_0 = 20$ and $N_{\mathrm{ada}} = 10$. We compare the convergence for the three considered approaches GPMAP-I, CGPMAP-II, and CFBGP with quantile $q = 0.90$ and $n_\theta = 100$ hyperparameter samples and use 2000 particles and 25 rejuvenation steps for the SMC algorithm. It can be seen that when the samples are drawn naively from the prior distribution, the convergence of the approximation to the true posterior distribution is comparatively slow. On the other hand, when the samples are drawn adaptively using Algorithm 1, all considered approximations converge to a good approximation of the true posterior within less than 100 solver calls. The marginal posterior distributions of the true posterior and an approximation using the CFBGP approach with 100 solver calls are shown in figure 5. The results might imply that the uncertainty of the Gaussian process does not have to be modeled. However, the log likelihood function of the investigated problem is very simple, which is not the case in more realistic applications, such as in the following examples.
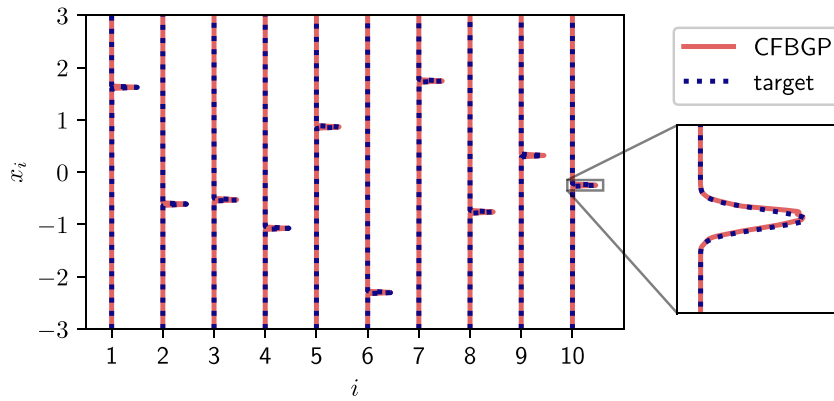
Figure 5. KDE of marginal posterior distributions of the true posterior $p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}})$ and an approximation using the CFBGP approach with 100 solver calls and $n_\theta = 100$.

3.2. Inference of diffusivity field

In the second example, we consider the following diffusion problem:

$$\nabla \cdot (D(\boldsymbol{c}) \nabla u(\boldsymbol{c})) = 10, \quad \boldsymbol{c} \in \Omega \tag{20}$$

$$u(\boldsymbol{c}) = 0, \quad \boldsymbol{c} \in \partial\Omega \tag{21}$$

where $\boldsymbol{c}$ denotes the coordinates in the domain Ω, $D(\boldsymbol{c})$ denotes the diffusivity coefficient at location $\boldsymbol{c}$ and $u(\boldsymbol{c})$ is a spatially varying density. The PDE is solved using the finite element FEniCS library [51, 52] with a 20×20 mesh of bilinear quadrilateral elements. The prior of the diffusivity coefficient is modeled as a lognormal random field based on a normal random field G with zero mean and covariance function

$$\text{Cov}[G(\boldsymbol{c}), G(\boldsymbol{c}')] = \exp\left[-\frac{\|\boldsymbol{c} - \boldsymbol{c}'\|_2^2}{2l_G^2}\right] \tag{22}$$

where $l_G = 0.2$. The prior of the field is discretized using the Karhunen–Loève expansion [53]:

$$D(\boldsymbol{c}|\boldsymbol{x}) = \exp(G(\boldsymbol{c}|\boldsymbol{x})) = \exp\left(\sum_{i=1}^{k} x_i \phi_i(\boldsymbol{c})\right) \tag{23}$$

where ϕ_i are the eigenfunctions of the covariance function, weighted by the square root of the corresponding eigenvalues, and the coefficients x_i are normally distributed random variables with unit variance. The number of expansion terms is set to $k = 30$, which allows to preserve 99% of the variance of the field G. The goal is to infer a posterior distribution of the coefficients x_i for given observations of the density u. In order to create artificial observations, we draw a random sample from the prior distribution $p(\boldsymbol{x})$, reconstruct the diffusivity field using equation (23), solve the discretized forward problem in (21), evaluate the density u at all 361 nodes inside the domain Ω, add independent and normally distributed noise with variance 10^{-4} and collect the observations in the vector $\boldsymbol{y}_{\text{obs}}$. The generated diffusivity field and corresponding solution field are shown in figure 6.
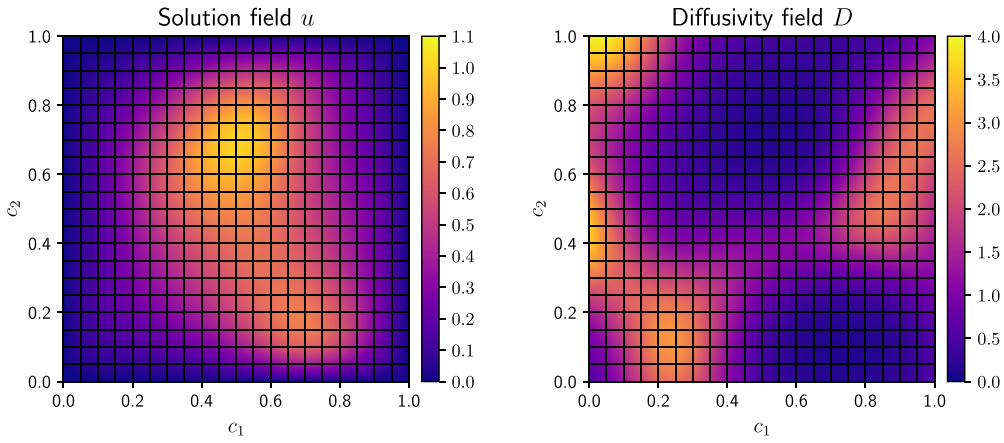
Figure 6. Ground truth of the diffusivity field *D* and the corresponding solution *u*.
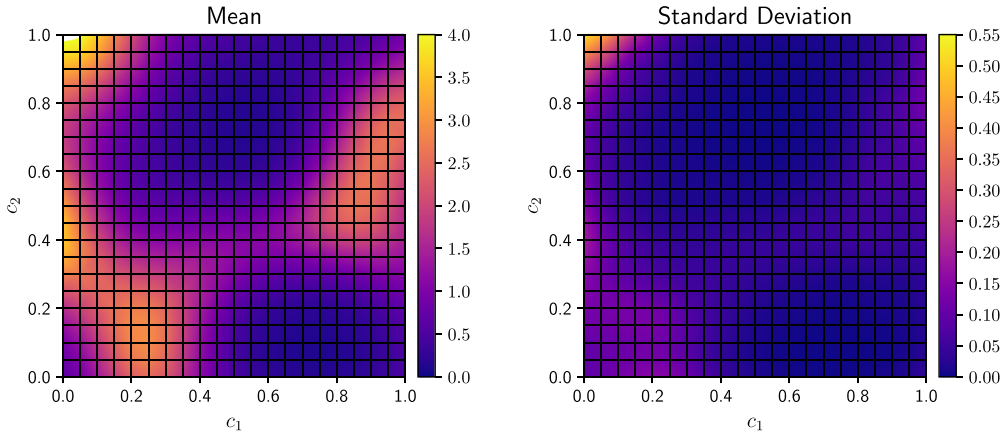

Figure 7. Reference solution of the inverse problem using SMC. The mean and the standard deviation of the posterior of the diffusivity field *D* are shown.

As a reference solution, we solve the inverse problem without a surrogate using SMC with 3000 particles and 30 rejuvenation steps, which leads to almost three million model evaluations. Figure 7 shows the mean and the standard deviation of the obtained posterior distribution. It can be seen that the mean aligns well with the ground truth solution and that the uncertainty of the posterior is maximal in the bottom corners.

For our proposed approach, we set $N_0 = 40$ and $N_{\text{ada}} = 20$ and use 3000 particles and 30 rejuvenation steps for the SMC algorithm. As a convergence criterion, we choose $\alpha_{\text{tol}} = 10^{-2}$. The convergence of the algorithm using the CGPMAP-II approach and using the CFBGP approach with $q = 0.90$ and $n_\theta = 100$ hyperparameter samples is visualized in figure 8. It shows the discrepancy measure defined in (18) between the intermediate posterior distributions $\hat{D}_{\text{CS}}\left(p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}}, \mathcal{D}_j), p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}}, \mathcal{D}_{j-1})\right)$ that is used as a convergence measure. Further, it shows the discrepancy measure between the posterior approximations and the reference solution using MCMC $\hat{D}_{\text{CS}}\left(p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}}, \mathcal{D}_j), p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}})\right)$, which serves as an accuracy measure.
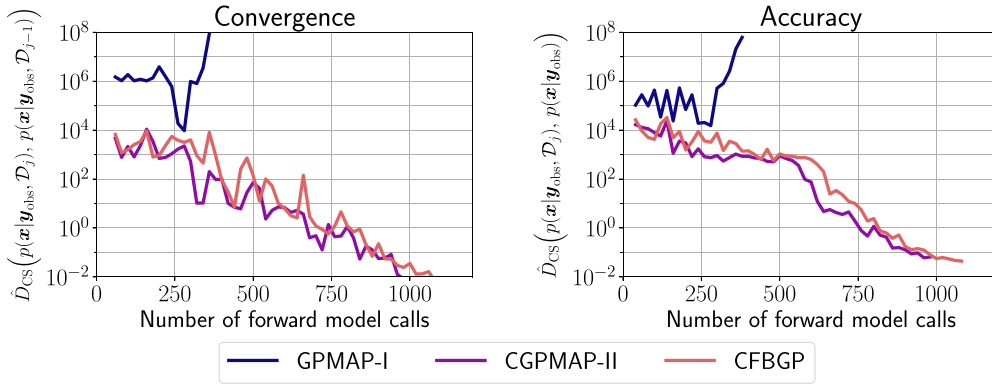
Figure 8. Discrepancy measure defined in (18) between the intermediate posterior distributions $\hat{D}_{CS}\left(p(\boldsymbol{x}|\boldsymbol{y}_{obs},\mathcal{D}_j),p(\boldsymbol{x}|\boldsymbol{y}_{obs},\mathcal{D}_{j-1})\right)$ that is used as a convergence measure (left). Discrepancy measure between the posterior approximations and the reference solution using MCMC $\hat{D}_{CS}\left(p(\boldsymbol{x}|\boldsymbol{y}_{obs},\mathcal{D}_j),p(\boldsymbol{x}|\boldsymbol{y}_{obs})\right)$, which serves as an accuracy measure (right).

The convergence criterion is met after 980 solver calls using the CGPMAP-II approach and after 1080 solver calls if the CFBGP approach is used. However, the discrepancy $\hat{D}_{CS}$ to the reference solution indicates that the accuracy of the approximated posterior is slightly higher when using the CFBGP approach. Nevertheless, the approximated posterior mean and standard deviation of the field are very accurate in both cases, as they align well with the reference solution (see figures 7 and 9). The high accuracy is also evident when looking at the posterior marginals of the inferred coefficients in figure 10. It can be seen that the marginal posterior approximations align well with the reference solution. As the coefficients x_i correspond to the weighted eigenfunctions ϕ_i with decreasing eigenvalue with increasing index i, the marginal posterior variance of x_i increases with increasing index i. It must be noted that due to the added measurement noise and the used prior information, the ground truth does not usually align with the mode of the posterior marginals but lies within regions of significant probability mass.

On the other hand, when using the GPMAP-I approach, the SMC algorithm becomes unstable and terminates due to numerical issues after 320 solver calls. Due to the exploitative nature of the GPMAP-I approach, the algorithm puts new training points in local maxima of the posterior far off the global maximum, and due to the unboundedness and exponentiation of the Gaussian process, the posterior approximations become very sharp with small variance. Ultimately, this leads to a failed Cholesky decomposition in the MCMC step of the SMC algorithm due to extremely close training points. This highlights the importance of considering the surrogate uncertainties for more complex likelihoods if the adaptive sampling strategy in algorithm 1 is applied, as this encourages exploration and prevents getting stuck in local optima.

3.3. Inference of stiffness field in a human lung model

To this day, respiratory diseases are one of the leading causes of death and affected patients often require mechanical ventilation as a life-saving measure. Especially for diseased, i.e. heterogeneous lungs, adequate adjustment of ventilation parameters is a balancing act between sufficient oxygenation of the blood and removal of carbon dioxide from the lungs, and minimizing damage to lung tissue, a potential side effect triggered and exacerbated by ventilation.
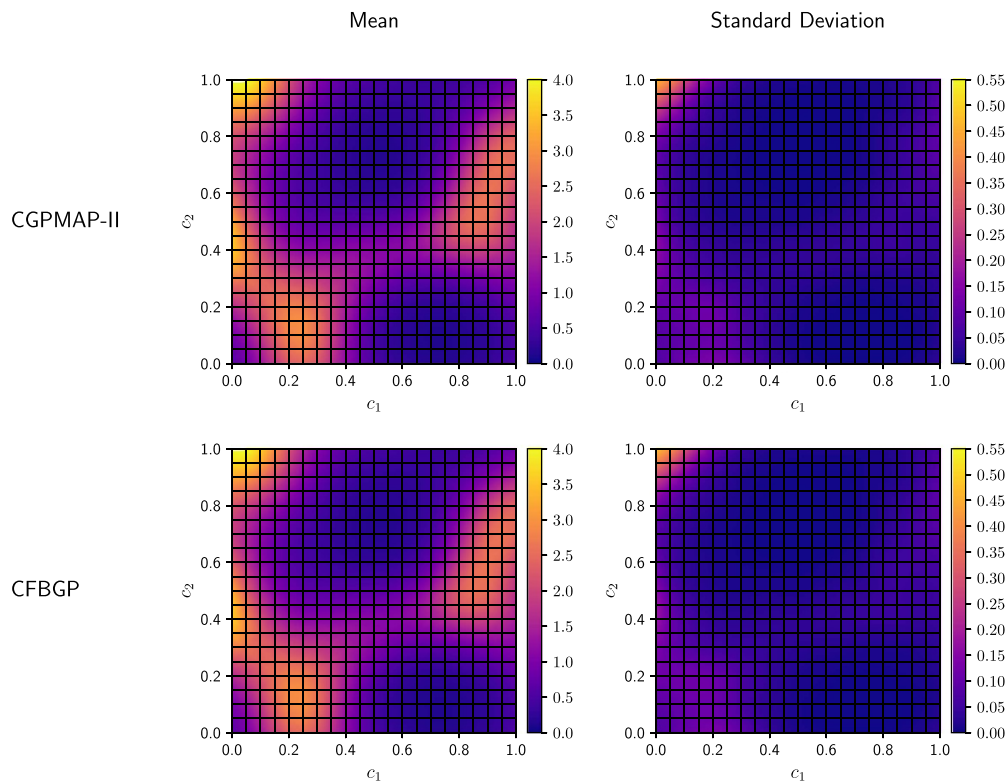
Figure 9. Posterior mean and standard deviation of the diffusivity field *D* using the CGPMAP-II approach (top row) and using the CFBGP approach with $n_\theta = 100$ hyper-parameter samples (bottom row).

Due to the high inter-patient variability, a patient-specific ventilation strategy is crucial, but also challenging: at the bedside, physicians usually only have the global ventilation measurements at hand to customize the ventilation parameters, which do not deliver insight into the local lung behavior. However, it is the individual heterogeneity of diseased lungs that requires deeper local insight and understanding. Computational and patient-specific models can help provide more information about local phenomena and tissue conditions and ultimately advance the development of protective and patient-specific ventilation strategies.

With the presented method to solve inverse problems, we investigate in a simplified application-oriented example a well-established reduced dimensional model of the human lungs. Assuming measurements observable at the bedside (ventilation measurements and electrical impedance tomography, see below), we determine with the presented approach inversely the material parameters of the lung model, which eventually may allow to draw conclusions about the local conditions in a patient's lungs. The lung model was developed in our group (see e.g. [54, 55]) and is implemented in the in-house code 4C [56]. The human lungs are supplied with air by a conducting airway tree that begins at the trachea and gradually branches to distal regions of the organ. In each so-called generation, an airway branches into two smaller airways. In total, the human lungs comprise around 23 generations, with the first 17 generations acting as a purely conducting airway network. Under various assumptions, the Navier–Stokes
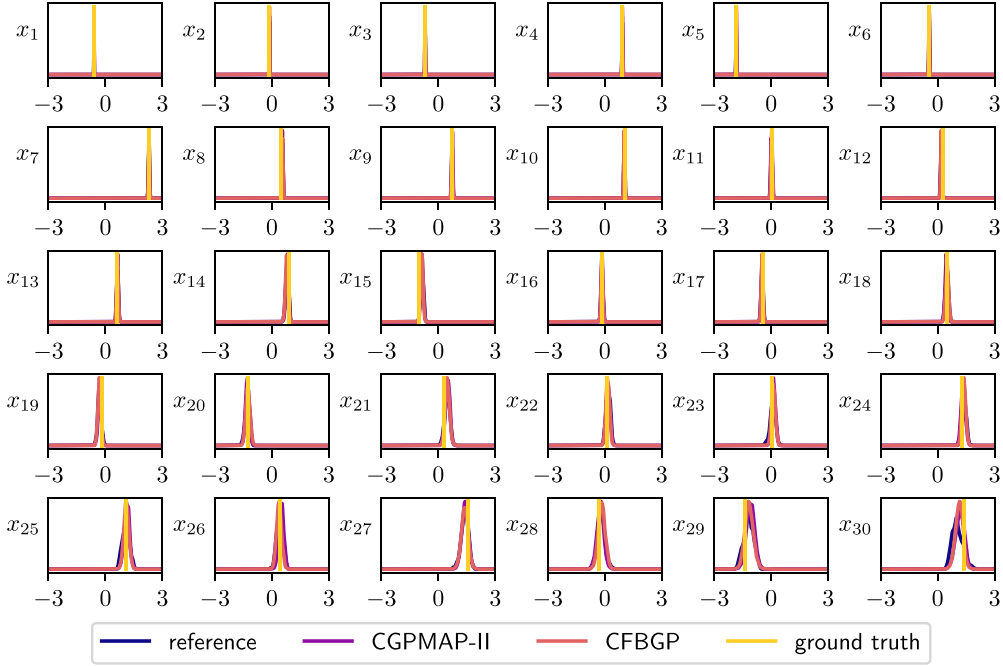
Figure 10. KDE of marginal posterior distributions of the field coefficients x of the reference solution, using the CGPMAP-II approach and using the CFBGP approach with $n_\theta = 100$ hyperparameter samples and the ground truth.

equations that describe the 3D fully resolved air flow in conducting airways and their visco-elastic wall mechanics can be integrated along the coordinates of the idealized cylindrical airways. This leads to the following simplified equations for each airway element of the model [54, 55, 57]:

$$C\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\left(P_{\mathrm{in}} + P_{\mathrm{out}}\right) - \widetilde{P}_{\mathrm{ext}}\right) + Q_{\mathrm{out}} - Q_{\mathrm{in}} + C \cdot R_{\mathrm{visc}}\frac{\mathrm{d}}{\mathrm{d}t}\left(Q_{\mathrm{out}} - Q_{\mathrm{in}}\right) = 0,$$

$$\frac{I}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(Q_{\mathrm{in}} + Q_{\mathrm{out}}\right) + \frac{1}{2}\left(R_\mu + R_{\mathrm{conv}}\right)\cdot\left(Q_{\mathrm{in}} + Q_{\mathrm{out}}\right) + P_{\mathrm{out}} - P_{\mathrm{in}} = 0, \quad (24)$$

where C describes capacitive, I inductive and R resistive effects of the system. The airways are coupled by the interfacing pressures $P_{\mathrm{in}}, P_{\mathrm{out}}$ and flow rates $Q_{\mathrm{in}}, Q_{\mathrm{out}}$ at the in- and outlets. In this example, the airways are resolved down to the fifth generation, and the remaining generations are modeled by terminal units which are described by a four-element Maxwell model reproducing the nonlinear viscoelastic behavior of the distal lung regions. For such a reduced-dimensional terminal unit, the following relation can be derived from an Ogden-type material law [58] to mimic its elastic component:

$$P_{\mathrm{a}} - P_{\mathrm{pl/intr}} = \frac{\kappa}{\eta}\left(\frac{V_{\mathrm{a}}}{V_{\mathrm{a},0}}\right)^{-1}\left(1 - \left(\frac{V_{\mathrm{a}}}{V_{\mathrm{a},0}}\right)^{-\eta}\right), \quad (25)$$

where P_{a} denotes the internal alveolar pressure, $P_{\mathrm{pl/intr}}$ the external pleural pressure, V_{a} the current volume of the terminal unit and $V_{\mathrm{a},0}$ the initial, stress-free volume. κ and η determine

the stiffness and curvature of this pressure-volume relation, respectively. For further details on the lung model, the reader is referred to [54, 55, 57]. We simulate the mechanical ventilation of an endotracheally intubated patient and therefore prescribe the inlet pressure at the proximal end of the tube, connected to the ventilator. Besides the measurement of the resulting flow at the ventilator, the physicians use also electrical impedance tomography (EIT) [59] at the bedside to get insight into the local ventilation and to adapt the medical therapy accordingly. Therefore, a belt with 32 electrodes is placed on the chest and small voltages are applied. The measured voltages at the remaining electrodes depend on the local conductivity inside the lungs. We can use the strains in the terminal units in a specific time step to evaluate the local resistivity ρ_{eff} based on the conductivity of the alveolar tissue $\sigma_{\text{alv}} = 0.7284\,\Omega^{-1}\text{m}^{-1}$ and the tortuosity $\bar{\tau} = 1.71$ [55, 60]:

$$\rho_{\text{eff}} = \frac{\bar{\tau}}{\sigma_{\text{alv}}} \left(5\frac{V_{\text{a}}}{V_{\text{a},0}} - 4 \right). \tag{26}$$

As the current volume of the terminal unit V_{a} is time dependent, we can solve the Laplace equation for a corresponding time step:

$$\nabla \cdot \left(\frac{1}{\rho_{\text{eff}}(\boldsymbol{c})} \nabla u(\boldsymbol{c}) \right) = 0, \quad \boldsymbol{c} \in \Omega, \tag{27}$$

using EIDORS [61] where u denotes the voltage and Ω the thorax domain. Diseased lungs of patients that undergo mechanical ventilation often exhibit a very heterogeneous state of lung tissue resulting in inhomogeneous straining and thus ventilation of the organ. This alternating material behavior is included in the model by a local variation of the stiffness parameter κ similar to [55]. This functions in the following as quantity of interest in the inverse problem as these are crucial parameters for clinical application and patient-specific models. However, in this study, we will not elaborate further on the clinical case but just use this highly relevant problem as motivation to construct a challenging example. Similarly to the previous numerical example, the prior of the stiffness parameter is modeled as a lognormal random field based on a normal random field G with zero mean and the covariance function in (22) with $l_G = 100$. Again, the prior of the field is discretized using the Karhunen–Loéve expansion:

$$\kappa(\boldsymbol{c}|\boldsymbol{x}) = \kappa_{\text{min}} + \kappa_{\text{scale}} \exp\left(G(\boldsymbol{c}|\boldsymbol{x})\right) = 500\,[\text{Pa}] + 2000\,[\text{Pa}] \exp\left(\sum_{i=1}^{k} x_i \phi_i(\boldsymbol{c})\right). \tag{28}$$

We set $k = 11$ such that 99% of the variance of the field G can be reproduced. The choice for the parameters $\kappa_{\text{min}} = 500\,\text{Pa}$ and $\kappa_{\text{max}} = 2000\,\text{Pa}$ is based on experiments about the mechanical behavior of lung tissue [62, 63]. For the inverse analysis, we use flow measurements at the ventilator at 539 time points as well as 928 EIT voltage measurements at eight time points as observations. In this paper, we focus on demonstrating the suitability of our approach for such kinds of problems. Hence, we refrain from using and explaining in detail, real clinical measurements. Thus, analogously to the previous example, the artificial observations are generated by drawing a random sample from the prior $p(\boldsymbol{x})$, solving the reduced dimensional lung model, solving the Laplace equation at the defined eight time steps (see figure 12), and adding independent and normally distributed noise to the observations. In total, this yields 7424 voltage measurements and 539 flow measurements. We set the noise variance for the voltage measurements to $10^{-8}\,[\text{V}^2]$ and for the flow measurements to $10^{-4}\,[(\text{l/s})^2]$. The realization of the stiffness field κ that is used to generate the observations is shown in figure 11 and the obtained measurements are visualized in figure 12.
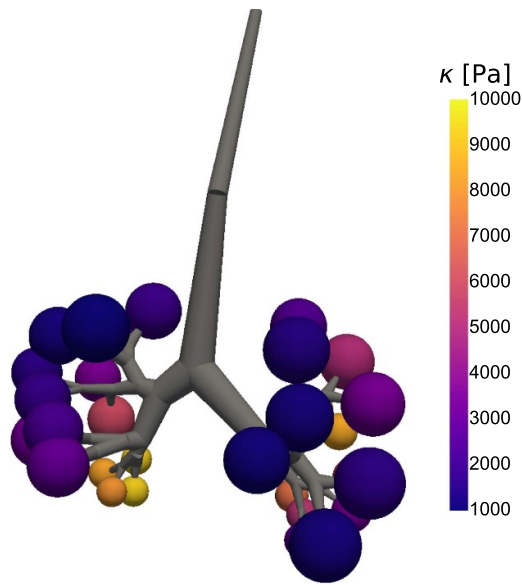
Figure 11. Ground truth of the stiffness field κ. The airway branches are shown in grey and the colored spheres represent the terminal units.
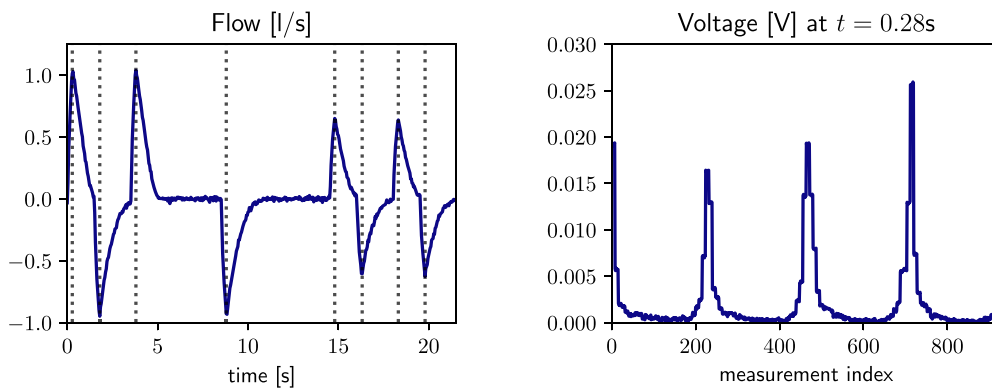

Figure 12. Flow measurements at the ventilator (left). The dotted lines indicate the eight points in time when the EIT measurements are conducted. The right plot shows the measurements of all 32 electrodes at the first time of measurement. In total, there are 32 configurations with 29 measurements leading to 928 measurements at each time step, which are indicated here by the measurement index.

We create a reference solution of the posterior $p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}})$ without a surrogate model using SMC with 2000 particles and 25 rejuvenation steps, which leads to 652 000 model evaluations. For our proposed method, we set $N_0 = 16$, $N_{\text{ada}} = 8$, $\alpha_{\text{tol}} = 10^{-2}$, $q = 0.90$ and use 2000 particles and 25 rejuvenation steps for the SMC algorithm. Figure 13 shows that the convergence criterion is met after 312 solver calls using the CGPMAP-II approach and after 304 solver calls using the CFBGP approach with $n_\theta = 100$. Whereas, when using the GPMAP-I approach, the solution oscillates and does not reach the convergence criterion within 400 model evaluations. In this example, the algorithm converges slightly faster when we marginalize
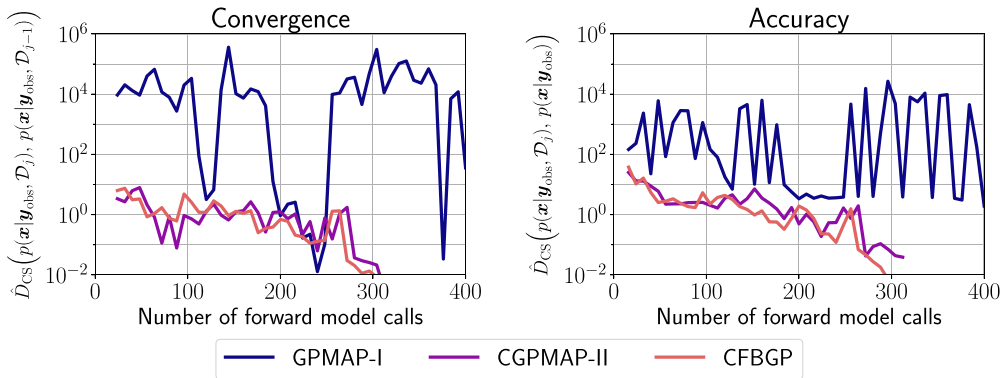
Figure 13. Discrepancy measure defined in (18) between the intermediate posterior distributions $\hat{D}_{\mathrm{CS}}\left(p(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}, \mathcal{D}_j), p(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}, \mathcal{D}_{j-1})\right)$ that is used as a convergence measure (left). Discrepancy measure between the posterior approximations and the reference solution using SMC $\hat{D}_{\mathrm{CS}}\left(p(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}}, \mathcal{D}_j), p(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}})\right)$, which serves as an accuracy measure (right).
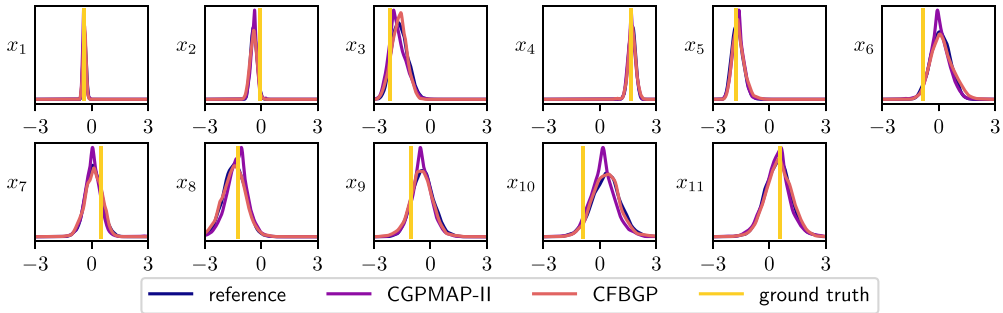

Figure 14. KDE of marginal posterior distributions of the field coefficients $\boldsymbol{x}$ of the reference solution, using the CGPMAP-II approach, using the CFBGP approach with $n_\theta = 100$ hyperparameter samples and the ground truth.

the hyperparameters (CFBGP) compared to using the MAP estimate of the hyperparameters (CGPMAP-II). Moreover, the accuracy of the approximated posterior distribution is considerably higher when using the CFBGP approach. Still, both approaches (CFBGP and CGPMAP-II) deliver a good approximation, which is also visible from the posterior marginals of the field coefficients $\boldsymbol{x}$ in figure 14, which shows that the marginal posterior approximations align well with the reference solution using SMC.

Further, the posterior marginals of the field κ at the terminal units in figure 15 support the impression that the approximations align well with the reference solution. Nevertheless, the marginal posterior distributions in figure 14 also reveal that the approximation using CFBGP is slightly more accurate compared to CGPMAP-II and that the posterior approximation using CGPMAP-II is a little overconfident due to the neglection of the hyperparameter uncertainty. As in the previous example, the ground truth does not usually align with the mode of the posterior marginals but lies within regions of significant probability mass due to the added measurement noise and the used prior information. There is a trend of increasing marginal posterior
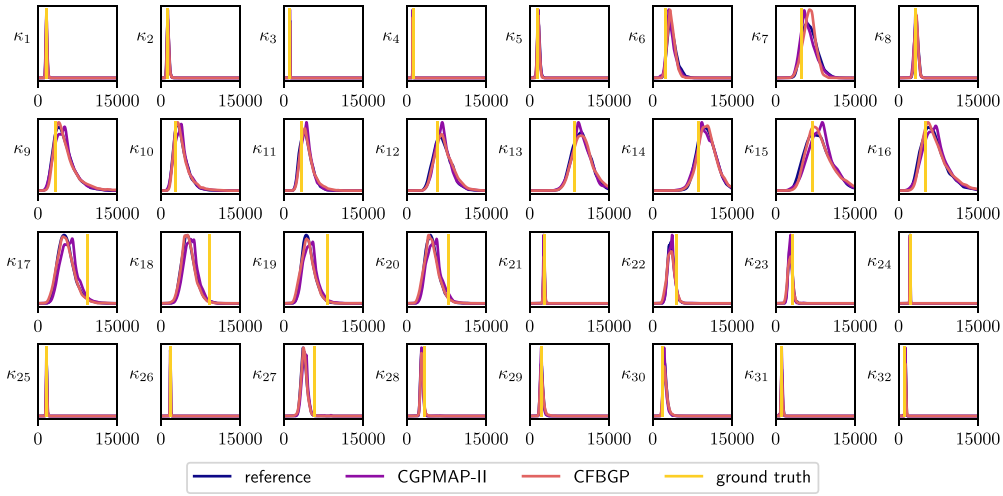
Figure 15. KDE of marginal posterior distributions of the field κ at the terminal units (e.g. $\kappa_1 := \kappa(\boldsymbol{c}_1|\boldsymbol{x})$, where $\boldsymbol{c_1}$ is the location of terminal unit 1) of the reference solution, using the CGPMAP-II approach, using the CFBGP approach with $n_\theta = 100$ hyperparameter samples and the ground truth.

variance of the coefficients x_i with increasing index i, as they correspond to the weighted eigenfunctions ϕ_i with decreasing eigenvalue with increasing index i. However, the variance of the marginal posteriors of x_i is not strictly increasing with index i due to locally varying influence on the observations, especially to the EIT observations. Generally, the variance of the posterior marginals of the field κ at the terminal units increases with increasing distance to the plane of the EIT electrode belt due to the decreasing influence on the measured voltages.

In this example, we also conduct a comparison with waste-free SMC [46] in terms of the number of forward model evaluations as well as the overall computation time (see figure 16). For the comparison, we parameterize the SMC algorithm with 256 particles and use $n_{\mathrm{rs}} = 5$ and $n_{\mathrm{rs}} = 10$ rejuvenation steps, respectively. We prescribe that at most eight forward models can be executed in parallel on a remote CPU cluster, and the training and evaluation of the surrogate models is conducted on a local GPU (Quadro RTX 4000). It can be seen that SMC with $n_{\mathrm{rs}} = 5$ yields the least accurate result and that the accuracy of SMC with $n_{\mathrm{rs}} = 10$ is comparable to the accuracy of the CGPMAP-II approach. However, the number of forward model evaluations and the total computation time of SMC with $n_{\mathrm{rs}} = 10$ is about two orders of magnitude higher than for CGPMAP-II. In this example, the computational overhead due to training and evaluating the surrogate model is relatively small for CGPMAP-II, and consequently, the reduced number of forward model evaluations also translates to a large reduction in overall computation time. The accuracy of CFBGP is significantly higher than for the other approaches, and the number of forward model evaluations is more than two orders of magnitude lower than for SMC with $n_{\mathrm{rs}} = 10$. Nevertheless, the computational overhead due to the surrogate model is significantly larger than for CGPMAP-II. Still, the computational time is around 38 times lower than for SMC with $n_{\mathrm{rs}} = 10$. It can be expected that for an SMC configuration that yields the same degree of accuracy, the factor by which the computational time is reduced is significantly higher. Additionally, the computational time of one forward model evaluation is still rather small at approximately 50 seconds in this example. This is due to the deliberate choice of the rather small fidelity of the reduced lung model, which still allows
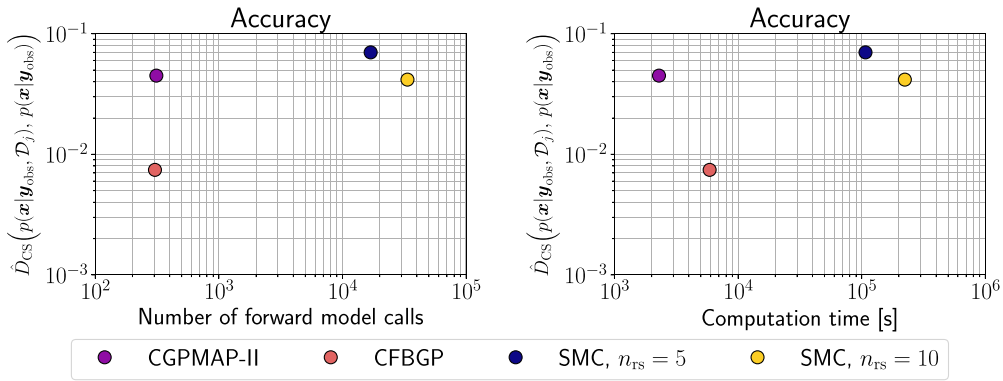
Figure 16. Discrepancy measure between the posterior approximations and the reference solution using SMC $\hat{D}_{\mathrm{CS}}\left(p(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}},\mathcal{D}_j),p(\boldsymbol{x}|\boldsymbol{y}_{\mathrm{obs}})\right)$, which serves as an accuracy measure. Accuracy is plotted against the number of forward model calls (left) and against the total computation time (right). As a comparison, SMC with 256 particles and $n_{\mathrm{rs}}=5$ and $n_{\mathrm{rs}}=10$ (rejuvenation steps) is shown.

us to generate an accurate reference solution in order to quantify the quality of our proposed approach. For models with higher fidelity, the computational time of one model evaluation can take several hours, drastically reducing the relative computational time overhead due to the surrogate model.

4. Conclusion

In this work, we propose an approach to solve Bayesian inverse problems with expensive likelihood models and high dimensional model outputs using Gaussian processes. In this novel approach, the number of likelihood evaluations is drastically reduced by choosing the training samples of the Gaussian process adaptively by sampling from intermediate posterior distributions. In addition, we incorporate the prior knowledge about the boundedness of the log likelihood function in our regression model for increased accuracy of the predictions. As a further improvement, we incorporate the epistemic uncertainty of the surrogate model due to limited training points, where we use the inverse CDF of the transformed predictive distribution. We show accuracy and efficiency in terms of the number of forward model evaluations of the method for a generic problem with analytical solution, as well as for a diffusion problem with 30 uncertain parameters. In a practically relevant example, the stiffness field of a reduced dimensional human lung model that is coupled with an EIT simulation is inferred using only a small number of model evaluations. The given examples, investigating the proposed method in the context of a diffusion problem and a human lung model, underlined the necessity to model the surrogate model uncertainties in high dimensional and complex inverse problems, as the adaptive algorithm does not converge within the given computational bounds when the uncertainty and boundedness of the Gaussian process were neglected (GPMAP-I). Integrating out the hyperparameters increases the computational cost of the Gaussian process predictions significantly, such that only a small number of hyperparameter samples can be used to solve the integral. Therefore, significant improvements are only visible in the final and most complex example when marginalizing the hyperparameters (CFBGP) compared to using the MAP estimate of the hyperparameters (CGPMAP-II).

The efficiency of our proposed black box surrogate algorithm relies on the combination of three techniques, namely adaptive sampling, the consideration of epistemic uncertainties, and the enforcement of the constraint on the surrogate model. The first numerical example shows how the adaptive selection of training samples reduces the necessary number of forward model evaluations, especially in higher stochastic dimensions. The second and third numerical examples show how important the consideration of the surrogate uncertainties is for the stability of the algorithm in combination with the active learning approach. Finally, the constraint on the Gaussian process helps to improve the accuracy of the approximation without any additional forward model evaluations.

In this paper, we use SMC as inference scheme and Gaussian processes as surrogate models. It must be noted that for the proposed approach, also other inference schemes besides SMC can be used, such as variational inference. In addition, other surrogate models besides Gaussian processes can be employed as long as they possess the ability to quantify the predictive uncertainty and accommodate the upper bound constraint. Here, we only consider Gaussian likelihoods, but the approach can be extended to other likelihoods by modifying the upper bound constraint.

Data availability statement

The data cannot be made publicly available upon publication due to legal restrictions preventing unrestricted public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

Appendix

A.1. Upper bound estimate

The random variable $b(\boldsymbol{y}_{\mathrm{obs}}) = -\frac{1}{2} \|\boldsymbol{y}_{\mathrm{obs}} - \mathcal{M}(\boldsymbol{x})\|^2_{\Sigma_n}$ follows a chi-squared distribution:

$$
\begin{aligned}
b(\boldsymbol{y}_{\mathrm{obs}}) &= -\frac{1}{2} \|\boldsymbol{y}_{\mathrm{obs}} - \mathcal{M}(\boldsymbol{x})\|^2_{\Sigma_n}, && \boldsymbol{y}_{\mathrm{obs}} \sim \mathcal{N}(\mathcal{M}(\boldsymbol{x}), \Sigma_n) \\
&= -\frac{1}{2} \boldsymbol{\epsilon}^\top \Sigma_n^{-1} \boldsymbol{\epsilon}, && \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \Sigma_n) \\
&= -\frac{1}{2} \sum_{i=1}^{n_{\mathrm{obs}}} \hat{\epsilon}_i^2, && \hat{\epsilon}_i \sim \mathcal{N}(0, 1) \\
&= -\frac{1}{2} \gamma, && \gamma \sim \chi^2(n_{\mathrm{obs}}) \qquad (29)
\end{aligned}
$$

A.2. Simplified constrained Gaussian process

Following [37] the predictive distribution of the constrained Gaussian process with one virtual observation point $\boldsymbol{x}^v = \boldsymbol{x}$ and the event $\zeta := f_s(\boldsymbol{x}^v) \leqslant \hat{b}$ is given as:

$$f_s | \boldsymbol{x}, \zeta, \mathcal{D} \sim \mathcal{N}\left(\mu(\boldsymbol{x}) + A\left(\zeta - \mu(\boldsymbol{x}^v)\right) + B\left(\boldsymbol{f} - \mu(X)\right), \Sigma\right), \tag{30}$$

where

$$\zeta \sim \mathcal{TN}\left(\mu(\boldsymbol{x}^v) + A_1\left(\boldsymbol{f} - \mu(X)\right), B_1, -\infty, \hat{b}\right), \tag{31}$$

and

$$
\begin{aligned}
A_1 &= k\left(\boldsymbol{x}^v, X\right)\left(K(X, X) + \sigma_f^2 I_N\right)^{-1} \\
A_2 &= k\left(\boldsymbol{x}, X\right)\left(K(X, X) + \sigma_f^2 I_N\right)^{-1} = A_1 \\
B_1 &= k\left(\boldsymbol{x}^v, \boldsymbol{x}^v\right) - A_1 k(X, \boldsymbol{x}^v) \\
B_2 &= k\left(\boldsymbol{x}, \boldsymbol{x}\right) - A_2 k(X, \boldsymbol{x}) = B_1 \\
B_3 &= k\left(\boldsymbol{x}, \boldsymbol{x}^v\right) - A_2 k(X, \boldsymbol{x}^v) = B_1 \\
A &= B_3 B_1^{-1} = I_N \\
B &= A_2 - A A_1 = 0 \\
\Sigma &= B_2 - A B_3^{\top} = 0.
\end{aligned} \tag{32}
$$

Consequently, the predictive distribution simplifies to:

$$f_s | \boldsymbol{x}, \zeta, \mathcal{D} \sim \mathcal{TN}\left(\mu(\boldsymbol{x}) + A_2\left(\boldsymbol{f} - \mu(X)\right), B_2, -\infty, \hat{b}\right). \tag{33}$$

A.3. Transformation of continuous random variable

$$
\begin{aligned}
g &= \exp\left(f_s\right) =: T(f_s) \\
p_g(g) &= p_{f_s}\left(T^{-1}(g)\right)\left|\frac{\mathrm{d}}{\mathrm{d}g} T^{-1}(g)\right| \\
p_g(g | \boldsymbol{x}, \mathcal{D}) &= p_{f_s}(\ln g | \boldsymbol{x}, \mathcal{D}) \frac{1}{g} \\
&= \frac{1}{n_\theta} \sum_{j=1}^{n_\theta} \frac{2}{g \sqrt{2\pi} s_j \left(1 + \mathrm{erf}\left(\frac{\hat{b} - m_j}{\sqrt{2} s_j}\right)\right)} \exp\left(-\frac{(\ln g - m_j)^2}{2 s_j^2}\right) \mathbb{I}_{0 \leqslant g \leqslant \exp(\hat{b})} \quad (34)
\end{aligned}
$$

A.4. Consistency of likelihood estimation

For an infinite number of training points N, the predictive distribution of the Gaussian process in (6) becomes a Dirac distribution with exact mean, independent of the hyperparameter sample. Consequently, the addends in (13) become identical to each other and for $q \in (0, 1)$ the inverse CDF can be written as:

$$\lim_{\substack{s \to 0 \\ m \to \ln g}} \exp\left[\text{erf}^{-1}\left\{q\left(1 + \text{erf}\left(\frac{\hat{b} - m}{\sqrt{2}s}\right)\right) - 1\right\}\sqrt{2}s + m\right] = g,$$

which shows that our chosen estimator for the unnormalized log likelihood is consistent.

ORCID iDs

Maximilian Dinkel ⬤ https://orcid.org/0009-0003-2908-2951
Carolin M Geitner ⬤ https://orcid.org/0000-0002-0370-2265
Gil Robalo Rei ⬤ https://orcid.org/0000-0002-6336-6718
Jonas Nitzler ⬤ https://orcid.org/0000-0002-7258-5664
Wolfgang A Wall ⬤ https://orcid.org/0000-0001-7419-3384

References

[1] Schoeder S, Kronbichler M and Wall W A 2017 Photoacoustic image reconstruction: material detection and acoustical heterogeneities *Inverse Problems* **33** 055010

[2] Hervas-Raluy S *et al* 2023 Tumour growth: an approach to calibrate parameters of a multiphase porous media model based on in vitro observations of Neuroblastoma spheroid growth in a hydrogel microenvironment *Comput. Biol. Med.* **159** 106895

[3] Nitzler J, Meier C, Müller K W, Wall W A and Hodge N E 2021 A novel physics-based and data-supported microstructure model for part-scale simulation of laser powder bed fusion of Ti-6Al-4V *Adv. Model. Simul. Eng. Sci.* **8** 16

[4] Willmann H and Wall W A 2022 Inverse analysis of material parameters in coupled multi-physics biofilm models *Adv. Model. Simul. Eng. Sci.* **9** 7

[5] Tikhonov A N, Goncharsky A V, Stepanov V V and Yagola A G 1995 *Numerical Methods for the Solution of Ill-Posed Problems* (Springer)

[6] Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems* (*Applied Mathematical Sciences* vol 160) (Springer)

[7] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 2004 Equation of state calculations by fast computing machines *J. Chem. Phys.* **21** 1087–92

[8] Neal R M 1994 An improved acceptance procedure for the hybrid Monte Carlo algorithm *J. Comput. Phys.* **111** 194–203

[9] Hoffman M D and Gelman A 2014 The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo *J. Mach. Learn. Res.* **15** 1593–623

[10] Chopin N 2002 A sequential particle filter method for static models *Biometrika* **89** 539–52

[11] Chopin N and Papaspiliopoulos O 2020 *An Introduction to Sequential Monte Carlo (Springer Series in Statistics)* (Springer)

[12] Del Moral P, Doucet A and Jasra A 2006 Sequential Monte Carlo samplers *J. R. Stat. Soc.* B **68** 411–36

[13] Doucet A and De Freitas N 2001 Neil James Gordon and others *Sequential Monte Carlo Methods in Practice* vol 1 (Springer)

[14] Blei D M, Kucukelbir A and McAuliffe J D 2017 Variational inference: a review for statisticians *J. Am. Stat. Assoc.* **112** 859–77

[15] Rezende D and Mohamed S 2015 Variational Inference with Normalizing Flows *Proc. 32nd Int. Conf. on Machine Learning* (*Proc. of Machine Learning Research* vol 37) (*Lille, France*) ed F Bach and D Blei (PMLR) pp 1530–8

[16] Kingma D P, Salimans T and Welling M 2015 Variational dropout and the local reparameterization trick *Advances in Neural Information Processing Systems* vol 28, ed C Cortes, N Lawrence, D Lee, M Sugiyama and R Garnett (Curran Associates, Inc.)

[17] Ranganath R, Gerrish S and Blei D 2014 Black box variational inference *Proc. 17th Int. Conf. on Artificial Intelligence and Statistics* (*Proc. Machine Learning Research* vol 33) (Reykjavik, Iceland) ed S Kaski and J Corander (PMLR) pp 814–22

[18] Beaumont M A, Zhang W and Balding D J 2002 Approximate Bayesian computation in population genetics *Genetics* **162** 2025–35

[19] Sisson S A, Fan Y and Beaumont M A (eds) 2018 *Handbook of Approximate Bayesian Computation* 1st edn (Chapman and Hall/CRC Press)

[20] Rubio F J and Johansen A M 2013 A simple approach to maximum intractable likelihood estimation *Electron. J. Stat.* **7** 1632–54

[21] Marzouk Y M, Najm H N and Rahn L A 2007 Stochastic spectral methods for efficient Bayesian solution of inverse problems *J. Comput. Phys.* **224** 560–86

[22] Marzouk Y and Xiu D 2009 A Stochastic collocation approach to Bayesian inference in inverse problems *Commun. Comput. Phys.* **6** 826–47

[23] Ma X and Zabaras N 2009 An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method *Inverse Problems* **25** 035013

[24] David J C M 1995 Bayesian neural networks and density networks *Nucl. Instrum. Methods Phys. Res.* A **354** 73–80

[25] Kennedy M C and O'Hagan A 2001 Bayesian calibration of computer models *J. R. Stat. Soc.* B **63** 425–64

[26] Rasmussen C E 2003 Gaussian processes to speed up hybrid Monte Carlo for expensive Bayesian integrals *Bayesian Statistics 7* (Oxford University Press) pp 651–60

[27] Bilionis I and Zabaras N 2014 Solution of inverse problems with limited forward solver evaluations: a Bayesian perspective *Inverse Problems* **30** 015004

[28] Semler P and Weiser M 2023 Adaptive Gaussian process regression for efficient building of surrogate models in inverse problems (arXiv:2303.05824)

[29] Zhu Y, Zabaras N, Koutsourelakis P-S and Perdikaris P 2019 Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data *J. Comput. Phys.* **394** 56–81

[30] Rixner M and Koutsourelakis P-S 2021 A probabilistic generative model for semi-supervised training of coarse-grained surrogates and enforcing physical constraints through virtual observables *J. Comput. Phys.* **434** 110218

[31] Raissi M, Perdikaris P and Karniadakis G E 2019 Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations *J. Comput. Phys.* **378** 686–707

[32] Zhang J, Zheng Q, Chen D, Wu L and Zeng L 2020 Surrogate-based Bayesian inverse modeling of the hydrological system: an adaptive approach considering surrogate approximation error *Water Resour. Res.* **56** e2019WR025721

[33] Kandasamy K, Schneider J and Poczos B 2015 Bayesian active learning for posterior estimation *Proc. 24th Int. Joint Conf. on Artificial Intelligence (IJCAI '15)* pp 3605–11

[34] Wang H and Li J 2018 Adaptive Gaussian process approximation for bayesian inference with expensive likelihood functions *Neural Comput.* **30** 3072–94

[35] QUEENS—a general purpose framework for uncertainty quantification, physics-informed machine learning, Bayesian optimization, inverse problems and simulation analytics on distributed computer systems 2024 (available at: https://queens_community.pages.gitlab.lrz.de/website/)

[36] Edward Rasmussen C and Williams C K I 2006 *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (MIT Press)

[37] Agrell C 2019 Gaussian processes with linear operator inequality constraints *J. Mach. Learn. Res.* **20** 1–36

[38] Gramacy R B and Lee H K H 2012 Cases for the nugget in modeling computer experiments *Stat. Comput.* **22** 713–22

[39] Swiler L P, Gulian M, Frankel A L, Safta C and Jakeman J D 2020 A survey of constrained Gaussian process: approaches and implementation challenges *J. Mach. Learn. Model. Comput.* **1** 119–56

[40] Wang K, Pleiss G, Gardner J, Tyree S, Weinberger K Q and Gordon Wilson A 2019 Exact Gaussian processes on a million data points *Advances in Neural Information Processing Systems* vol 32, ed H Wallach, H Larochelle, A Beygelzimer, F d'Alché-Buc, E Fox and R Garnett (Curran Associates, Inc.)

[41] Lalchand V and Edward Rasmussen C 2020 Approximate inference for fully bayesian gaussian process regression *Proc. 2nd Symp. on Advances in Approximate Bayesian Inference* (*Proc. of Machine Learning Research* vol 118) ed C Zhang, F Ruiz, T Bui, A B Dieng and D Liang (PMLR) pp 1–12

[42] Snoek J, Larochelle H and Adams R P 2012 Practical Bayesian optimization of machine learning algorithms *Advances in Neural Information Processing Systems* vol 25, ed F Pereira, C J Burges, L Bottou and K Q Weinberger (Curran Associates, Inc.)

[43] Auer P 2003 Using confidence bounds for exploitation-exploration trade-offs *J. Mach. Learn. Res.* **3** 397–422

[44] Srinivas N, Krause A, Kakade S and Seeger M 2010 Gaussian process optimization in the bandit setting: no regret and experimental design *Proc. 27th Int. Conf. on Int. Conf. on Machine Learning (ICML'10) (Madison, WI, USA)* (Omnipress) pp 1015–22

[45] Chandrupatla T R 1997 A new hybrid quadratic/bisection algorithm for finding the zero of a nonlinear function without using derivatives *Adv. Eng. Softw.* **28** 145–9

[46] Dau H-D and Chopin N 2021 Waste-free sequential Monte Carlo *J. R. Stat. Soc.* B **84** 114–48

[47] Krause A, Singh A and Guestrin C 2008 Near-optimal sensor placements in gaussian processes: theory, efficient algorithms and empirical studies *J. Mach. Learn. Res.* **9** 235–84

[48] Gorodetsky A and Marzouk Y 2016 Mercer Kernels and integrated variance experimental design: connections between gaussian process regression and polynomial approximation *SIAM/ASA J. Uncertain. Quantification* **4** 796–828

[49] Kampa K, Hasanbelliu E and Principe J C 2011 Closed-form cauchy-schwarz PDF divergence for mixture of Gaussians *The 2011 Int. Joint Conf. on Neural Networks* pp 2578–85

[50] Kullback S and Leibler R A 1951 On information and sufficiency *Ann. Math. Stat.* **22** 79–86

[51] Scroggs M W, Baratta I A, Richardson C N and Wells G N 2022 Basix: a runtime finite element basis evaluation library *J. Open Source Softw.* **7** 3982

[52] Alnaes M S, Logg A, Ølgaard K B, Rognes M E and Wells G N 2014 Unified Form Language: A domain-specific language for weak formulations of partial differential equations *ACM Trans. Math. Softw.* **40** 1–37

[53] Ghanem R G and Spanos P D 2003 *Stochastic Finite Elements: A Spectral Approach* (Courier Corporation)

[54] Ismail M, Comerford A and Wall W A 2013 Coupled and reduced dimensional modeling of respiratory mechanics during spontaneous breathing *Int. J. Numer. Methods Biomed. Eng.* **29** 1285–305

[55] Roth C J, Becher T, Frerichs I, Weiler N and Wall W A 2017 Coupling of EIT with computational lung modeling for predicting patient-specific ventilatory responses *J. Appl. Physiol.* **122** 855–67

[56] 4C: a comprehensive multi-physics simulation framework 2024 (available at: https://www.4c-multiphysics.org)

[57] Geitner C M, Becher T, Frerichs I, Weiler N, Bates J H T and Wall W A 2023 An approach to study recruitment/derecruitment dynamics in a patient-specific computational model of an injured human lung *Int. J. Numer. Methods Biomed. Eng.* **39** e3745

[58] William Ogden R and Hill R 1972 Large deformation isotropic elasticity—on the correlation of theory and experiment for incompressible rubberlike solids *Proc. R. Soc.* A **326** 565–84

[59] Brown B H 2003 Electrical impedance tomography (EIT): a review *J. Med. Eng. Technol.* **27** 97–108

[60] Roth C J, Ehrl A, Becher T, Frerichs I, Schittny J C, Weiler N and Wall W A 2015 Correlation between alveolar ventilation and electrical properties of lung parenchyma *Physiol. Meas.* **36** 1211–26

[61] Adler A and William R B L 2006 Uses and abuses of EIDORS: an extensible software base for EIT *Physiol. Meas.* **27** S25–S42

[62] Birzle A M, Martin C, Uhlig S and Wall W A 2019 A coupled approach for identification of nonlinear and compressible material models for soft tissue based on different experimental setups—exemplified and detailed for lung parenchyma *J. Mech. Behav. Biomed. Mater.* **94** 126–43

[63] Sicard D, Haak A J, Moo Choi K, Craig A R, Fredenburgh L E and Tschumperlin D J 2018 Aging and anatomical variations in lung tissue stiffness *Am. J. Physiol.* **314** L946–55