Available online at www.sciencedirect.com

ScienceDirect

Comput. Methods Appl. Mech. Engrg. 289 (2015) 409–428

Computer methods in applied mechanics and engineering

www.elsevier.com/locate/cma

X-TMCMC: Adaptive kriging for Bayesian inverse modeling

Panagiotis Angelikopoulos[a], Costas Papadimitriou[b,*], Petros Koumoutsakos[a]

[a] ETH Zürich, Department of Mechanical Engineering, Computational Science and Engineering Laboratory, Clausiusstrasse 33, ZH-8092, Switzerland

[b] University of Thessaly, Department of Mechanical Engineering, Volos 38334, Greece

Received 27 June 2014; received in revised form 22 December 2014; accepted 28 January 2015
Available online 21 February 2015

Abstract

The Bayesian inference of models associated with large-scale simulations is prohibitively expensive even for massively parallel architectures. We demonstrate that we can drastically reduce this cost by combining adaptive kriging with the population-based Transitional Markov Chain Monte Carlo (TMCMC) techniques. For uni-modal posterior probability distribution functions (PDF), the proposed hybrid method can reduce the computational cost by an order of magnitude with the same computational resources. For complex posterior PDF landscapes we show that it is necessary to further extend the TMCMC by Langevin adjusted proposals. The proposed hybrid method exhibits high parallel efficiency. We demonstrate the capabilities of our method on test bed problems and on high fidelity simulations in structural dynamics.

Keywords: Bayesian inference; Transitional MCMC; Langevin diffusions; Surrogates; Kriging; Structural dynamics

1. Introduction

Bayesian inference holds a central position in data driven Uncertainty Quantification (UQ) and robust reliability and safety in engineering sciences [1]. Bayesian tools include Laplace methods of asymptotic approximation [2,3] and Markov Chain Monte Carlo (MCMC) (e.g. [4–9]). Despite their superiority in terms of reduced computational cost, asymptotic approximations may fail to adequately capture multi-modal and complex posterior Probability Density Functions (PDF). In turn, MCMC algorithms are currently amongst the most attractive Bayesian inference tools.

MCMC based algorithms are used for both sampling the target PDF as well as evaluating probability integrals over high-dimensional spaces of the uncertain model parameters [8]. The efficiency of the MCMC and Stochastic Algorithms (SA) tools hinges on the number of required system simulations. Thus limits their applicability for Bayesian UQ in the case of computational expensive model evaluations [10,11]. It is thus important to find ways to reduce the computational budget without deteriorating their efficiency.

[*] Corresponding author. Tel.: +30 2421074006.
E-mail addresses: panagiotis.angelikopoulos@mavt.ethz.ch (P. Angelikopoulos), costasp@uth.gr (C. Papadimitriou), petros@ethz.ch (P. Koumoutsakos).

http://dx.doi.org/10.1016/j.cma.2015.01.015
0045-7825/

One way of relaxing computational costs is by approximating the forward model evaluations using meta-models. A plethora of different surrogate models is available. The most commonly used are based on response surface methods [12], neural networks [13], polynomial chaos expansions [14], and kriging [15]. Surrogate techniques have been used in stochastic optimization algorithms [16–18], in structural reliability estimation using Monte Carlo simulations [19–23] or more efficient subset simulation [24,25], and in reliability-based optimization methods [26,27]. However, the error of the surrogate approximation is not always available online during the sampling process, and also most surrogates introduce bias in their predictions.

In this paper we present an efficient Bayesian inverse modeling framework by integrating kriging into the Transitional Markov Chain Monte Carlo (TMCMC) [28]. Inspired by similar approaches in evolutionary optimization [29–31], we propose an adaptive kriging scheme. This surrogate version of TMCMC (K-TMCMC), takes advantage of the TMCMC annealing property to replace the full simulation runs at a sample point of a TMCMC stage by a kriging estimate using as support points the neighborhood samples from previous stages.

A Langevin extension of TMCMC is also introduced. This extension is shown to improve the original TMCMC algorithm for increasingly complex posterior PDF. This new extension termed L-TMCMC is demonstrated to outperform TMCMC for multi-modal posterior PDFs and cases of unidentifiable parameters. The kriging estimates implemented within L-TMCMC also provide samples of acceptable quality. The L-TMCMC and its surrogate version, termed KL-TMCMC, involve an extra cost of evaluating the derivatives of the objective function with respect to the parameter set. The additional high computational cost for L-TMCMC can be alleviated for problems where adjoint techniques are applicable [32–34]. For KL-TMCMC the kriging algorithm provides the estimation of the gradient at a surrogate point. Finally, due to its population based nature and efficient load balancing, the proposed surrogated algorithms can readily exploit parallel and heterogeneous computer architectures [35]. The adaptive Kriging extensions of TMCMC are termed X-TMCMC and thus include the K-TMCMC (Kriging-TMCMC), the L-TMCMC and KL-TMCMC (Kriging + Langevin-TMCMC).

The paper is organized as follows. Section 2 reviews the Bayesian formulation, the available SA algorithms, the TMCMC algorithm and also introduces the L-TMCMC algorithm. Section 3 presents the adaptive kriging technique and its integration within the X-TMCMC for Bayesian UQ, and discusses parallel implementation issues. Section 4 demonstrates the effectiveness of the proposed adaptive kriging technique in terms of accuracy and computational efficiency. The numerical examples are chosen to cover uni-modal, multi-modal and unidentifiable posterior PDF, as well as a realistic application in structural dynamics in Section 5. Section 6 includes the summary and conclusions of the paper.

2. Bayesian inverse modeling

2.1. Bayesian framework

The Bayesian framework [1,2] is used to compute the posterior PDF $f\left(\underline{\theta}|D, M\right)$ of the parameters $\underline{\theta}$ of a model class M of an engineering system using available data D. This is achieved by using the Bayes theorem

$$f\left(\underline{\theta}|D, M\right) = \frac{f\left(D|\underline{\theta}, M\right)\pi(\underline{\theta}|M)}{f(D|M)} \tag{1}$$

where $f\left(D|\underline{\theta}, M\right)$ is the likelihood of observing the data D from the model class M, $\pi\left(\underline{\theta}|M\right)$ is the prior PDF and $f(D|M)$ is the evidence of the model class M, given by the multi-dimensional integral

$$f(D|M) = \int_{\Theta} f(D|\underline{\theta}, M)\pi(\underline{\theta}|M)d\underline{\theta} \tag{2}$$

over the space Θ of the uncertain model parameters. The evidence $f(D|M)$, is important for model class selection [36,37]. The prior PDF of the model parameters incorporates subjective information on the uncertainty.

The model prediction equation relates observational data $D \equiv \{\underline{\hat{y}}\}$ with model predictions $\underline{g}\left(\underline{\theta}|M\right)$

$$\underline{\hat{y}} = \underline{g}\left(\underline{\theta}|M\right) + \underline{e}. \tag{3}$$

The prediction error term $\underline{e} \sim N(\underline{0}, C)$, accounting for measurement, computational and modeling errors, is normally distributed with zero mean and covariance matrix C. Using (3), the likelihood in (1) is given by:

$$f(D|\underline{\theta}, M) = \frac{|C|^{-1/2}}{(2\pi)^{m/2}} \exp\left[-\frac{1}{2} J(\underline{\theta}; \underline{\hat{y}})\right] \tag{4}$$

where

$$J\left(\underline{\theta}; \underline{\hat{y}}\right) = \left[\underline{\hat{y}} - \underline{g}\left(\underline{\theta}|M\right)\right]^T C^{-1} \left[\underline{\hat{y}} - \underline{g}\left(\underline{\theta}|M\right)\right] \tag{5}$$

is a weighted measure of fit between the system model predictions and the measured data, and $|\cdot|$ denotes determinant.

2.2. MCMC sampling

The reconstruction of the posterior distribution for the posterior PDF (1) is usually performed using Markov Chain Monte Carlo (MCMC) algorithms. A standard variant of MCMC algorithms is the Metropolis–Hastings (MH) [5]. A non exhaustive list of MCMC algorithms include Adaptive MCMC [6,38,39], Differential Evolution Monte Carlo [40] or Differential Evolution Random Subsampling Monte Carlo [41] (DREAM). Other MCMC methods which can be categorized in the framework of Evolutionary Strategy MCMC methods [42] are particle filters [43], the TMCMC [28] and Hamiltonian TMCMC [9].

In this work we extend the TMCMC method, which is a generalization of the algorithm proposed by Beck and Au [8] shown to identify complex supports and adequately sample from multimodal PDFs and parameter manifolds that are unidentifiable. Another advantage of TMCMC is its high parallel efficiency [35]. The annealing structure of TMCMC is exploited at each stage of the algorithm to provide adequate number of local support points corresponding to full model simulations to approximate the value of the posterior PDF at new surrogate points. The TMCMC algorithm is next described in sufficient detail to reveal the algorithmic key points that facilitate the adaptive surrogate design.

2.3. TMCMC algorithm

The TMCMC algorithm [28] relies on the construction of a series of intermediate PDFs

$$f_j(\underline{\theta}) = [f(D|\underline{\theta}, M)]^{p_j} \pi(\underline{\theta}|M), \quad j = 0, \ldots, m, \ 0 = p_0 < p_1 < \cdots < p_m = 1. \tag{6}$$

The steps involved in the TMCMC algorithm are:

0. **INITIALIZE**
   Generate samples $\{\underline{\theta}_{0,k},\ k = 1, \ldots, N_0\}$ from the prior distribution $f_0(\underline{\theta}) \equiv \pi(\underline{\theta}|M)$ and compute the likelihood $f(D|\underline{\theta}_{0,k}, M)$ for each sample.
1. For $j = 1, \ldots, m-1$ **REPEAT**:
   Compute plausibility weights for the samples $\{\underline{\theta}_{j,k},\ k = 1, \ldots, N_j\}$ of the N_j samples from the intermediate PDF $f_j(\underline{\theta})$.

$$w(\underline{\theta}_{j,k}) = \frac{f_{j+1}(\underline{\theta}_{j,k})}{f_j(\underline{\theta}_{j,k})} = [f(D|\underline{\theta}_{j,k}, M)]^{p_{j+1}-p_j}. \tag{7}$$

   Estimate the exponent p_{j+1} so that the coefficient of variation of the plausibility weights $w(\underline{\theta}_{j,k})$ equal to a prescribed tolerance *tolCOV* indicating the population elitism diversification.
2. Estimate the normalized plausibility weights $\tilde{w}(\underline{\theta}_{j,k}) = w(\underline{\theta}_{j,k}) / \sum_{l=1}^{N_j} w(\underline{\theta}_{j,l})$ for each sample $\underline{\theta}_{j,k},\ k = 1, \ldots, N_j$ and compute the interchain covariance matrix

$$\Sigma_j = \sum_{k=1}^{N_j} \tilde{w}(\underline{\theta}_{j,k})[\underline{\theta}_{j,k} - \underline{\mu}_j][\underline{\theta}_{j,k} - \underline{\mu}_j]^T$$

   using the samples $\{\underline{\theta}_{j,k},\ k = 1, \ldots, N_j\}$, where $\underline{\mu}_j = \sum_{k=1}^{N_j} \tilde{w}(\underline{\theta}_{j,k})\underline{\theta}_{j,k}$ is the sample mean. Also, estimate $S_j = (1/N_j) \sum_{k=1}^{N_j} w(\underline{\theta}_{j,k})$, which is an asymptotically unbiased estimator for $E[w(\underline{\theta}_{j,k})]$ and is used to provide an unbiased estimate for the model evidence required for model selection.

3. Resample based on the samples available in stage j in order to generate samples for stage $j+1$ distributed as $f_{j+1}(\underline{\theta})$. Choose $l = 1, \ldots, N_{j+1}$ samples from the samples $\{\underline{\theta}_{j,k},\ k = 1, \ldots, N_j\}$ with probability $\tilde{w}(\underline{\theta}_{j,k})$. The same point can be chosen multiple times using a tournament selection scheme [44]. A selected sample is called leader and is denoted by $\underline{\theta}_{j,0}^{i,leader}$, $i = 1, \ldots, \hat{N}_{j+1}$, where $\hat{N}_{j+1}$ is the number of unique samples selected, while the number of times the sample $\underline{\theta}_{j,0}^{i,leader}$ is selected is denoted by $n_{j+1,i}$.
4. Starting with each unique leader $\underline{\theta}_{j,0}^{i,leader}$, generate a chain by performing $n_{j+1,i}$ Metropolis–Hasting (MH) steps with candidate normal proposal $N(\underline{\theta}, \beta^2 \Sigma_j)$, where β^2 is a user-selected parameter scaling the covariance matrix Σ_j. Thus, for each leader sample $\underline{\theta}_{j,0}^{i,leader}$, $i = 1, \ldots, \hat{N}_{j+1}$, a MCMC chain of length $n_{j+1,i}$ is initiated, generating the samples $\{\theta_{j,l}^{i,leader},\ l = 1, \ldots, n_{j+1,i}\}$.
5. Compute $\prod_{j=0}^{m-1} S_j$, which is an asymptotically unbiased estimate of the evidence $f(D|M)$ of the model class.

Note that the computation of a sample $\underline{\theta}_{j+1,k}$ in any chain in step 4 requires the estimate of the likelihood $f(D|\underline{\theta}_{j+1,k}, M)$ needed in the intermediate PDF $f_{j+1}(\underline{\theta}) = [f(D|\underline{\theta}_{j,k}, M)]^{p_{j+1}} \pi(\underline{\theta}|M)$ that is involved in the MH step performed within a chain. Also, the estimation of the MH samples for a chain corresponding to a leader sample can proceed independently from the other chains and thus all these chains can run independently by exploiting parallel computing architectures to drastically reduce the time-to-solution. In step 4 of the algorithm, we use a MH scheme to rejuvenate the samples, using as a proposal a scaled down version of the sample covariance matrix.

We remark that the global nature of the sample covariance matrix Σ_j, constructed from all the stage points, leads to non-optimal proposal shapes, specifically in the cases of multimodal target PDF, and/or complex supports arising from unidentifiable parameters. In the following we introduce a modification of the original TMCMC algorithm to improve the performance for complex posterior PDF supports using local information of the intermediate posterior PDF.

2.4. L-TMCMC algorithm

The efficiency of the TMCMC can be improved by using information about the shape of the intermediate posterior PDFs. For this, we employ a Metropolis-adjusted Langevin transition kernel [45,46] for our individual chain rejuvenation and refer to this method as L-TMCMC. The Metropolis Adjusted Langevin transition kernel replaces the MH steps in step 4. Specifically, for each unique leader sample $\underline{\theta}_{j,0}^{i,leader}$, perform $n_{j+1,i}$ Metropolis-Adjusted-Langevin steps starting from each leader $\underline{\theta}_{j,0}^{i,leader}$ with candidate proposal density

$$q(\underline{\theta}_{j,l}^{i,leader}, \underline{\theta}_{j,l+1}^{i,leader}) \equiv N\left(\underline{\theta}_{j,l}^{i,leader} + \frac{1}{2}h^2 \nabla \log f_j(\underline{\theta}_{j,l}^{i,leader}), hI\right) \tag{8}$$

where $\underline{\theta}_{j,l}^{i,leader}$ is the l-th sample in chain i, $N(\underline{\mu}, C)$ denotes normal distribution with mean $\underline{\mu}$ and covariance matrix C and h has to be considered a-priori like β^2 in normal TMCMC. As in the MH algorithm, each candidate point $\underline{\theta}_{j,l+1}^{i,leader}$ is accepted with probability

$$\min\left(1, \frac{f_j(\underline{\theta}_{j,l+1}^{i,leader})\, q(\underline{\theta}_{j,l+1}^{i,leader}, \underline{\theta}_{j,l}^{i,leader})}{f_j(\underline{\theta}_{j,l}^{i,leader})\, q(\underline{\theta}_{j,l}^{i,leader}, \underline{\theta}_{j,l+1}^{i,leader})}\right). \tag{9}$$

Note that the gradients taken in the proposal density come from the intermediate posterior bridging functions at each stage.

The Langevin kernel requires the evaluation of the gradients of the objective function with respect to the model parameters. Adjoint techniques [32–34], requiring a single adjoint problem for finding the gradient independently of the dimension of the parameter space, can be used to provide fast and accurate gradients. Adjoint techniques are model intrusive, in certain case inconvenient to develop, while they might not be applicable for complicated models of certain systems. Using adjoint methods the computational cost is approximately doubled since each Langevin step, in addition to a function evaluation, requires also the gradient information from the solution of the adjoint problem. Lack of adjoint formulations for the problem at hand may increase the workload significantly since finite difference approximations of the gradient have to be alternatively used, scaling up the computational effort by a factor proportional to the number of uncertain parameters. Analytical formulations for the gradient performed in parallel simultaneously for all parameters or parallel numerical differentiation tools can be used to minimize the time-to-solution.

3. Surrogate approximations

The TMCMC becomes prohibitively expensive for engineering problems, where each model evaluation necessary for the likelihood function, requires large-scale simulations. Surrogate models can be employed to replace a significant number of time-consuming full system simulations required in the likelihood evaluations by very fast approximate estimates. The objective is to avoid many forward model runs, while maintaining the sampling quality.

Among the different surrogate possibilities, herein we integrate kriging within TMCMC and L-TMCMC. Kriging interpolants do not require regular grid of neighbor support points, provide an error estimate of the kriging approximation and also provide an estimate of the derivative of the approximated function with respect to the model parameters, needed in L-TMCMC.

3.1. Kriging interpolation

Consider m support points $[\underline{\theta}_1, \ldots, \underline{\theta}_m]$ in the parameter space and let $\underline{Y} = [J(\underline{\theta}_1), \ldots, J(\underline{\theta}_m)]^T$ be the available values of the measure of fit function $J(\underline{\theta}_i) \equiv J(\underline{\theta}_i; \underline{\hat{y}})$ at these points. The term support point is used to denote a point in the parameter space where a full model simulation is available. Using kriging, $J(\underline{\theta})$ is approximated at a new point $\underline{\theta}$ in the parameter space in the form

$$J(\underline{\theta}) = \underline{Q}^T(\underline{\theta})\underline{\beta} + \varepsilon(\underline{\theta}) \tag{10}$$

where, $\underline{Q}^T(\underline{\theta}) = [Q_1(\underline{\theta}), \ldots, Q_m(\underline{\theta})]$ are basis functions and $\underline{\beta} = [\beta_1, \ldots, \beta_m]^T$ are regression coefficients. The term $\varepsilon(\underline{\theta})$ is a zero mean stochastic process with covariance $C_\varepsilon = \sigma^2 R(\underline{\varphi}; \underline{\theta}, \underline{\theta}')$ with variance σ^2 and a set of parameters $\underline{\varphi}$, where $\underline{\theta}$ and $\underline{\theta}'$ are points in the parameter space. The choice of the correlation function $R(\underline{\varphi}; \underline{\theta}, \underline{\theta}')$ critically affects how kriging represents the data. A common choice for the correlation function is [15]:

$$R(\underline{\varphi}, \alpha; \underline{\theta}, \underline{\theta}') = \exp\left[-\sum_{k=1}^{m} \varphi_k |\theta_k - \theta_k'|^\alpha\right] \tag{11}$$

with $\varphi_k \geq 0,\ k = 1, \ldots, m$, and $0 \leq \alpha \leq 2$, where θ_k and θ_k' are the k-th components of vectors $\underline{\theta}$ and $\underline{\theta}'$, respectively. The optimal choice of the parameters $\underline{\beta}$ and σ^2 that maximize the likelihood function given the values at the support points is

$$\underline{\hat{\beta}} = (Q^T R^{-1} Q)^{-1} Q^T R^{-1} \underline{Y} \tag{12}$$

$$\hat{\sigma}^2 = \frac{1}{m}(\underline{Y} - Q\underline{\hat{\beta}})^T R^{-1} (\underline{Y} - Q\underline{\hat{\beta}}) \tag{13}$$

where R is the matrix with components $R_{ij} = R(\underline{\varphi}, \alpha; \underline{\theta}_i, \underline{\theta}_j)$. The optimal values $\underline{\hat{\varphi}}$ and $\hat{\alpha}$ are obtained by minimizing the negative log-likelihood function $G(\underline{\varphi}, \alpha) = -\log L(\underline{\varphi}, \alpha, \underline{\hat{\beta}}(\underline{\varphi}, \alpha), \hat{\sigma}^2(\underline{\varphi}, \alpha))$, where

$$L(\underline{\varphi}, \alpha, \underline{\beta}, \sigma^2) = \frac{1}{2}\log|R(\underline{\varphi}, \alpha)| + \frac{m}{2}\log[(\underline{Y} - Q\underline{\beta})^T R^{-1}(\underline{\varphi}, \alpha)(\underline{Y} - Q\underline{\beta})]. \tag{14}$$

The prediction at the point $\underline{\theta}$ in the parameter space and its mean square error are given respectively by [47,48]

$$\hat{J}(\underline{\theta}) = Q^T(\underline{\theta})\underline{\hat{\beta}} + \underline{r}^T \hat{R}^{-1}(\underline{Y} - Q\underline{\hat{\beta}}) \tag{15}$$

and

$$s^2(\underline{\theta}) = \hat{\sigma}^2[1 - \underline{r}^T \hat{R}^{-1}\underline{r} + \underline{u}^T (Q^T \hat{R}^{-1} Q)^{-1}\underline{u}] \tag{16}$$

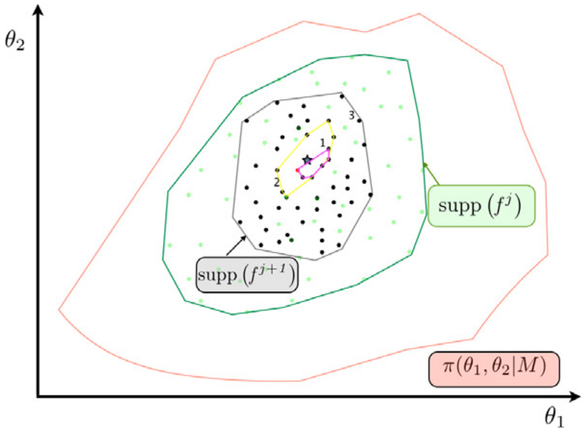

Fig. 1. Local surrogate estimate filtering strategy. The star indicates a candidate point $\underline{\theta}^c$ in the parameter space to be estimated using a surrogate model. The dots indicate support points. Green colored points indicate samples from stage j, whereas black are samples from stage $j+1$. Only the points that belong within the yellow convex hull are used as support points to estimate the likelihood at $\underline{\theta}^c$. The magenta convex hull does not contain the candidate point $\underline{\theta}^c$ and cannot be used to make a surrogate estimate. Note the enrichment of support points (black dots) at stage $j+1$ with the support points from previous stages (green dots corresponding to stage j). Supp defines the boundary support of the intermediate posterior PDF at each stage. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where $\underline{r}$ is a vector with components $r_j = R(\hat{\underline{\varphi}}, \hat{\alpha}; \underline{\theta}_j, \underline{\theta})$, $\hat{R} = R(\hat{\underline{\varphi}}, \hat{\alpha})$, $\underline{u} = F^T R^{-1}\underline{r} - \underline{Q}(\underline{\theta})$, and F is a matrix with components $Q_{ij} = Q_j(\underline{\theta}_i)$. Similar expressions exist for the prediction of the derivatives $\hat{\underline{J}}_\theta(\underline{\theta}) = \underline{\nabla}\hat{J}(\underline{\theta})$ of $\hat{J}(\underline{\theta})$ with respect to $\underline{\theta}$, where $\underline{\nabla}$ is the gradient operator [49].

We integrate the kriging package DACE [43] within the X-TMCMC framework. The estimation of the kriging parameters $\hat{\underline{\varphi}}$ and $\hat{\alpha}$ may involve the solution of a high dimensional minimization problem. The kriging prediction may quickly deteriorate if the minimization is not accurate [49]. A bounded pattern-search method [50] is used in this work to estimate $\hat{\underline{\varphi}}$ and $\hat{\alpha}$.

The root mean squared error or the standard deviation $s(\underline{\theta})$ in (16) represents the predicted deviation of the kriging meta-model from the actual response. In this paper, it is assumed that this predicted deviation has a normal distribution with mean equal to $\hat{J}(\underline{\theta})$ and variance equal to the kriging variance $s^2(\underline{\theta})$. The normalized prediction error $s(\underline{\theta})/\hat{J}(\underline{\theta})$ is a measure of the quality of the kriging estimate.

3.2. *Adaptive kriging for X-TMCMC*

A feature of the TMCMC algorithm is that the support of the intermediate posterior at stage $j+1$ is a subset of that of stage j due to the annealing schedule. Thus, the support points at stage j or previous stages can be conveniently used as support points for making surrogate estimates at new samples generated at stage $j+1$. The full simulation runs available at previous stages allow the enrichment with additional support points of the regions of high probability volume of the current intermediate posterior distribution, hence being able to train a more accurate kriging meta-model. The proposed kriging algorithm is local and adaptive since in order to make an estimate of the objective function at a new sample point it uses the support points that are in the neighborhood of this new sample point. Fig. 1 shows the employed strategy for selecting support points from previous design stages to perform the surrogate estimation at stage $j+1$. The star denotes the location of the candidate point $\underline{\theta}^c$, for which we want to construct the likelihood estimation at stage $j+1$.

In order to control the size of the error introduced in the kriging prediction, a surrogate estimate at stage $j+1$ is performed and accepted if it simultaneously obeys the following rules:

1. The surrogate estimate is based on a user-defined number n_{neigh} of support points $\underline{\theta}_i^{neigh}$, $i = 1, \ldots, n_{neigh}$, which are in the neighbor of the point $\underline{\theta}^c$ with respect to the Euclidean norm or the Mahalanobis distance based on the scaled sample covariance matrix Σ_j in the $\underline{\theta}$ space. The minimum number $n_{\min}$ of support points depends on the dimension m of the uncertain parameter space and the order of the kriging interpolation. For a kriging based on a first and second order polynomial basis, the least number of support points required are $n_{\min} = m + 1 \le n_{neigh}$

and $n_{\min} = (m+1)(m+2)/2 \le n_{neigh}$, respectively. An important consideration when choosing the number of support points is that the support points within a chain corresponding to a leader in TMCMC should be kept fixed. Otherwise, discontinuity in the surrogate objective function evaluation for the samples within the same chain may arise when the surrogate estimates at two neighbor points within the chain are based on two slightly different sets of support points. Such discontinuities affect the quality of the surrogate estimates and should be avoided. Hence, when choosing local kriging, the number of support points is taken to be $n_{neigh} = \alpha n_{\min}$, where the value of α is empirically recommended to be in the range 2–5, and it is associated with the nonlinearity of the target PDF and possible multimodality. Moreover, the support points for a chain are based on the location of the leader in the chain.

2. The surrogate point must belong to the convex hull of the support points, hence allowing only interpolation estimates. We employ the method of Ref. [51] to examine whether the surrogate point $\underline{\theta}^c$ lies in the n-dimensional convex hull of its support points.
3. The kriging support points correspond to full model simulations and not other surrogate estimates from previous stages. Hence they avoid error propagation and subsequent deterioration of the surrogate quality.
4. The surrogate estimate is checked to verify whether its predicted value is within the lower 95% quantile $q_{0.95}$ of all posterior values of the points accounted so far with full model simulations. The purpose of this threshold is to control the sampling procedure and to prevent inaccurate surrogate estimates with likelihood values that are higher than the ones already obtained from real function evaluations. Sample points corresponding to overshooting surrogate estimates are assigned a very high plausibility weight in the TMCMC algorithm (see step 1) which may lead to drawing most samples at the next stage from the region around this surrogate point, something that could be erroneous when the surrogate estimate is inaccurate. This would quickly lead to the break down of the sampling procedure due to the overshooting surrogate estimate. This is similar to the error control introduced in Ref. [52].
5. The surrogate estimate is accepted if the prediction error $s(\underline{\theta}^c)/\hat{J}(\underline{\theta}^c)$, normalized by the surrogate estimate, is smaller than a user specified tolerance value ε. The value of ε affects the convergence speed and bias reduction of the K-TMCMC algorithm. This effect will be demonstrated in the validation examples. To maintain the accuracy of the TMCMC algorithm, the error ε of the function evaluation due to surrogate estimate has to be kept relatively small.

Failure of the surrogate estimate to comply with all the criteria above prompts us to abandon the surrogate trial and perform a full model evaluation for the likelihood instead. The tuneable parameter ε controls the local support point refinement for decreasing the predicted kriging error variance.

Surrogate estimates of $J(\underline{\theta};\underline{\hat{y}})$ are equivalent to using a noisy likelihood. The introduction of the approximate likelihood can be seen within the context of the pseudo-marginal MCMC approach [53]. In order to achieve non-erroneous sampling from the target PDF when using a noisy likelihood, the necessary condition is that there is spatially invariant bias in the mean of the approximate likelihood evaluations. In the case of ordinary kriging, this condition holds. There is a constant bias introduced in our erroneous approximation $\hat{J}(\underline{\theta};\underline{\hat{y}})$. Assuming that $J(\underline{\theta};\underline{\hat{y}}) = \hat{J}(\underline{\theta};\underline{\hat{y}}) + \varepsilon$, ordinary kriging assumes that the errors ε are zero-mean spatially unbiased. Given the exponential transformation (4) leading to the true form of the likelihood, we are introducing a constant spatially independent bias in the variance. In this context, it is important to note that the locality parameter α, must be big enough to accommodate all the MH chain movements using the same support points in the case of a local kriging model. Otherwise the assumption of sampling from the correct distribution will quickly deteriorate, as the erroneous approximations will become spatially dependent due to the local character of the model.

The surrogate estimates can also be conveniently integrated in the L-TMCMC algorithm. In this case the gradient of the objective function required in the proposal PDF (8) is approximated by the kriging estimate for the gradients at the candidate point for which the objective function value is replaced by a kriging estimate. The L-TMCMC algorithm based on surrogate estimate is termed KL-TMCMC.

We provide in a pseudo-code (Algorithm 1) a description of the surrogate implementation flowchart for X-TMCMC (K-TMCMC and KL-TMCMC). A list of the X-TMCMC algorithms with their main features is given in Table 1.

Algorithm 1 (*Likelihood Evaluation $f(D|\underline{\theta}^c, M)$ at X-TMCMC Stage j Given a Candidate Point $\underline{\theta}^c$*).

```
0. Set user defined parameters ($n_{neigh}$, $\alpha$, $\varepsilon$);
   Access database of physical model evaluations up to stage $j$ and compute the lower 95% quantile, $q_{0.95}$, using rule
   4 in Section 3.2.
```

Table 1
List of X-TMCMC algorithms with their main features.

X-TMCMC	Kriging estimates	Gradient evaluations	Comments
TMCMC	No	No	Basic algorithm taken from [28]
K-TMCMC	Yes	No	Basic TMCMC algorithm that includes Kriging evaluations of the likelihood
L-TMCMC	No	Yes	Diffusion kernels from TMCMC are replaced with Langevin diffusions as in Eq. (8)
KL-TMCMC	Yes	Yes	L-TMCMC algorithm that includes in the Langevin transitions the kriging estimates on both likelihood and its gradients

```
1. Begin Surrogate Trial
   Find $n_{neigh}$ closest support points $\underline{\theta}_i^{neigh}$, $i = 1, \ldots, n_{neigh}$, with respect to $\underline{\theta}^c$, following rule 1 in Section 3.2.
   If $\underline{\theta}^c \in convhull(\underline{\theta}_i^{neigh})_{i=1}^{n_{neigh}}$,
         Estimate $\hat{J}(\underline{\theta}^c)$ and $s(\underline{\theta}^c)/\hat{J}(\underline{\theta}^c)$ using Eqs. (15) and (16)
         If $s(\underline{\theta}^c)/\hat{J}(\underline{\theta}^c) < \varepsilon$ and $\hat{J}(\underline{\theta}^c) < q_{0.95}$
              Set $J(\underline{\theta}^c) = \hat{J}(\underline{\theta}^c)$
              For case KL-TMCMC, Estimate $\underline{\nabla}\hat{J}(\underline{\theta}^c)$ from kriging for use in proposal density (8)
         Else
              Evaluate $J(\underline{\theta}^c)$ (and $\underline{\nabla}J(\underline{\theta}^c)$ for case KL-TMCMC) from the physical model;
              Store values $J(\underline{\theta}^c)$ and $\underline{\nabla}J(\underline{\theta}^c)$ to database;
         Endif
   Else Evaluate $J(\underline{\theta}^c)$ (and $\underline{\nabla}J(\underline{\theta}^c)$ for Case KL-TMCMC) from the physical model; Store to database
   Endif
2. Evaluate $f(D|\underline{\theta}^c, M)$ and continue with accept/reject TMCMC and L-TMCMC rules for sample $\underline{\theta}^c$.
```

3.3. Parallelization

The X-TMCMC contains a large number of chains per stage, based on the resampling step of the algorithm. The chains in TMCMC are perfectly parallel. For each stage, the number of computer workers required to run TMCMC does not exceed the number of independent chains. Since the number of samples $n_{j+1,i}$ in each chain $i = 1, \ldots, \hat{N}_{j+1}$ vary, the computing load can be balanced to less than $\hat{N}_{j+1}$ computer workers since one computer worker can handle more than one chains containing fewer samples in relation to the chain with the largest number of samples. Efficient work balancing can be used to minimize the number of computer workers. The load balance in the computer workers can be based on a static job-scheduling scheme under which each worker is assigned a predefined work to complete during the time it is allocated from the job scheduling. This assumes that the time for a likelihood evaluation is approximately the same, independent of the model parameter values assigned in the parameter space. In the case for which the available computer workers equal to the number $\hat{N}_{j+1}$ of chains per stage, where each host would be assigned a single chain, the parallel efficiency is limited by the maximum length chain since the time-to-solution is bottlenecked from the chain with the highest number of samples. The solution of the adjoint formulation required in Langevin chain in the L-TMCMC algorithm for estimating the derivatives of the posterior PDF with respect to all model parameters is expected to double the time-to-solution for the same number of workers used in TMCMC.

For the surrogate K-TMCMC and KL-TMCMC algorithms, the aforementioned static scheduling scheme is non-optimal since it does not take into account the number of fast surrogate function evaluations within each chain. An a-priori knowledge of the surrogate evaluations cannot be anticipated which makes the use of static scheduling schemes inappropriate. Instead, a dynamic scheduling scheme within a shared, queue managed heterogeneous cluster can be beneficial to distribute the likelihood evaluations requiring full system simulations in a pool of workers on a first-come–first-serve basis.

Table 2
Parametric study of the accuracy of the surrogate estimates for different ε values. FE denotes the number of true function evaluations. LE denotes Log-Evidence.

X-TMCMC	ε	$M(\mu)$	$D(\mu)$	$M(\sigma)$	$D(\sigma)$	M(LE)	D(LE)	FE
TMCMC	–	0.0062	0.050	0.9921	0.0420	−13.15	0.022	30,000
K-TMCMC	0.001	−0.0072	0.052	1.0136	0.0462	−13.17	0.032	18,374
K-TMCMC	0.01	0.00921	0.056	0.9892	0.0521	−12.91	0.037	8,229
K-TMCMC	0.1	−0.0141	0.064	0.9857	0.0761	−12.80	0.056	3,621
K-TMCMC	0.5	0.0101	0.071	1.0272	0.0824	−13.21	0.061	1,814
K-TMCMC variable-ε	0.5–0.001	0.0051	0.028	1.0251	0.0541	−13.04	0.037	2,491
L-TMCMC	–	0.0017	0.032	0.9979	0.0320	−13.02	0.015	30,000
KL-TMCMC	0.1	−0.0085	0.036	0.9811	0.0321	−12.79	0.023	1,732

4. Test-bed problems

We demonstrate the proposed algorithms in 3 test-bed problems. We introduce validation measures in order to evaluate the quality of the posterior PDF through the surrogate. We set as q_i a measure of the mean or standard deviation of the marginal distributions of an output quantity of interest, where the subscript i refers to the ith X-TMCMC run. Then the measures

$$M(q) = \frac{1}{N}\sum_{i=1}^{N} q_i \quad \text{and} \quad D(q) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (q_i - M(q))^2} \tag{17}$$

based on N X-TMCMC sample estimates for q_i provide a measure of the statistical mean $M(q)$ and standard deviation $D(q)$ of q. Due to the stochastic nature of the TMCMC algorithm, a sufficiently large number N of X-TMCMC runs need to be performed to get a statistical meaningful estimate for q. We perform $N = 50$ runs in all test cases and the surrogate estimate is performed at the measure of fit $J(\underline{\theta}; \underline{\hat{y}})$ level. The parameters controlling the TMCMC algorithm are chosen to be tolCov $= 1.0$, $\beta^2 = 0.2$. The parameters controlling the L-TMCMC algorithm is tolCOV $= 1.0$ and $h = 1.0$ All results for $M(q)$ and $D(q)$ are presented for the output quantity of interest q which is chosen to be either the sample mean μ or the sample standard deviation σ of the marginal distributions of the model parameters θ_j, $i = 1, \ldots, N_\theta$. In addition choosing q to be the log evidence of the model, an estimate of the accuracy of the surrogate approximation for model selection is provided. We also estimate 5%–95% percent quantiles from spread of the values from the N runs using the numpy.quantile method from scientific python with linear interpolation.

4.1. Uni-modal 10-dimensional posterior PDF

The first example involves a posterior PDF that is represented by a 10-dimensional zero-mean Normal distribution $f(\underline{\theta}|D) = N(\underline{\theta}|\underline{0}, I) \propto \exp[-0.5J(\underline{\theta})]$ with $J(\underline{\theta}) = \underline{\theta}^T\underline{\theta}$. The initial samples are generated from a uniform distribution in the domain $[-10, 10] \times \cdots \times [-10, 10]$. Results for the measures $M(\mu), D(\mu), M(\sigma), D(\sigma)$ quantifying the quality of the posterior are reported in Table 2 for different values of the errors ε allowed for the kriging estimates. For this isotropic distribution, the results for the $M(q)$ and $D(q)$ are also averaged over all 10 independent dimensions. The surrogate estimates in this case are based on a first-order kriging approximation on $J(\underline{\theta}) = \underline{\theta}^T\underline{\theta}$. All versions of X-TMCMC use 5000 samples per stage with 60 neighbors for the surrogate approximation.

In order to judge the accuracy of the K-TMCMC or KL-TMCMC estimates, reference values for the measures $M(\mu), D(\mu), M(\sigma), D(\sigma)$ are also reported in Table 2, computed from $N = 50$ TMCMC or L-TMCMC runs with parameter values of the algorithm chosen to be the same as the ones used for the corresponding surrogate versions. Results for the number of mean true function evaluations (FE), excluding surrogate evaluations, are also reported in the last column of Table 2 for the different algorithms.

Very small values of ε are expected to yield estimates close to the ones obtained without surrogates. As the values of the error ε increases one should expect deterioration in the accuracy of the surrogate approximation. We note that the accuracy in the prediction of the mean and the standard deviation of the model parameters is quite satisfactory even for high values of $\varepsilon = 0.5$. Fig. 2(a) and (b) present the 5% and 95% credible intervals, $M(q)-D(q)$, $M(q)$, $M(q)+D(q)$,

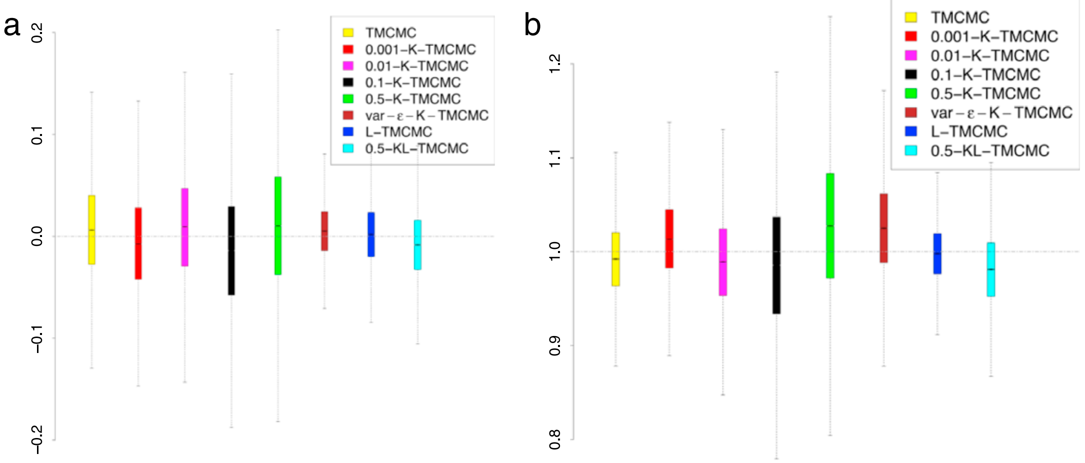

Fig. 2. Statistics (5% quantile, $M(q) - D(q)$, $M(q)$, $M(q) + D(q)$, 95% quantile) for (a) the mean ($q = \mu$) and (b) the standard deviation ($q = \sigma$) of the marginal distribution of the parameters estimated from TMCMC, K-TMCMC for $\varepsilon = 0.001$, 0.01, 0.1 and 0.5, variable-ε K-TMCMC, L-TMCMC, and KL-TMCMC for $\varepsilon = 0.5$.

for the mean ($q = \mu$) and the standard deviation ($q = \sigma$) of the marginal distribution of the parameters. The errors in the mean values when computed by the K-TMCMC algorithm are found to be slightly higher than the one computed for the MCMC algorithm. As the tolerance level ε decreases to as low values as $\varepsilon = 0.001$, the accuracy of the estimates provided from the K-TMCMC algorithm improves slightly. In both cases of K-TMCMC and TMCMC algorithms, the errors in the simplified measures of uncertainty remain very small. Comparing the number of function evaluation in the last column of Table 2, it is observed that for the $\varepsilon = 0.5$ tolerance level case, acceptable accuracy is obtained with a 10% of the number of real model evaluations in K-TMCMC compared to the full model evaluations required for the TMCMC algorithm, without deteriorating the quality of the K-TMCMC estimate. This one order of magnitude reduction in computational effort is considered significant for models requiring several minutes, hours or even days to complete one simulation run.

We also explore an alternative variant of the K-TMCMC, with a variable tolerance level ε across the different stages of the TMCMC algorithm. We start using high error values in the first stages where the parameter space is explored in order to move towards the support of the posterior PDF and adapting ε to smaller values at the final stages where the accuracy of the system simulation is more critical for adequately populating the posterior PDF. For this we adjust the tolerance level ε from as high as 0.5 value in the first stages until the exponent p_j in the annealing schedule of TMCMC becomes 0.2, and then change the ε value to as low as 0.001 until completion of the algorithm. As expected, this variant provides estimates of the mean value and the standard deviation of the uncertainty in the model parameters which are better that the ones provided by the case where ε was kept fixed to $\varepsilon = 0.5$ for all stages. The number of full model simulations is 50% higher than the number obtained for fixed $\varepsilon = 0.5$ error. However, this number is as low as 15% of the number of full model simulations required from the TMCMC algorithm without surrogates.

Results for the L-TMCMC are also shown in Table 2 and Fig. 2. L-TMCMC performs slightly better than the normal TMCMC, at the additional cost of calculating the derivatives of the function. The quality of the estimates provided by the surrogate KL-TMCMC is very good as compared to the L-TMCMC estimates and not worse than the TMCMC estimates. The extra cost from L-TMCMC or KL-TMCMC is justified only when adjoint methods can be used to find the derivatives so that the computational cost of evaluating derivatives is approximately the same as the cost of a function evaluation.

4.2. Multimodal posterior PDF

To showcase the local interpolation character of the methods, together with their ability to capture globally the important regions, we consider a two-dimensional multi-modal posterior distribution function $f(\theta_1, \theta_2) = a \exp(-0.1 \cdot J(\theta_1, \theta_2))$, where $J(\theta_1, \theta_2)$ is the multimodal Himmelblau function given by $J(\theta_1, \theta_2) = (\theta_1^2 + \theta_2 - 11)^2 + (\theta_1 + \theta_2^2 - 7)^2$. The function has four global minima, denoted here as $\hat{\theta}_i$, $i = 1, \ldots, 4$, located in the following

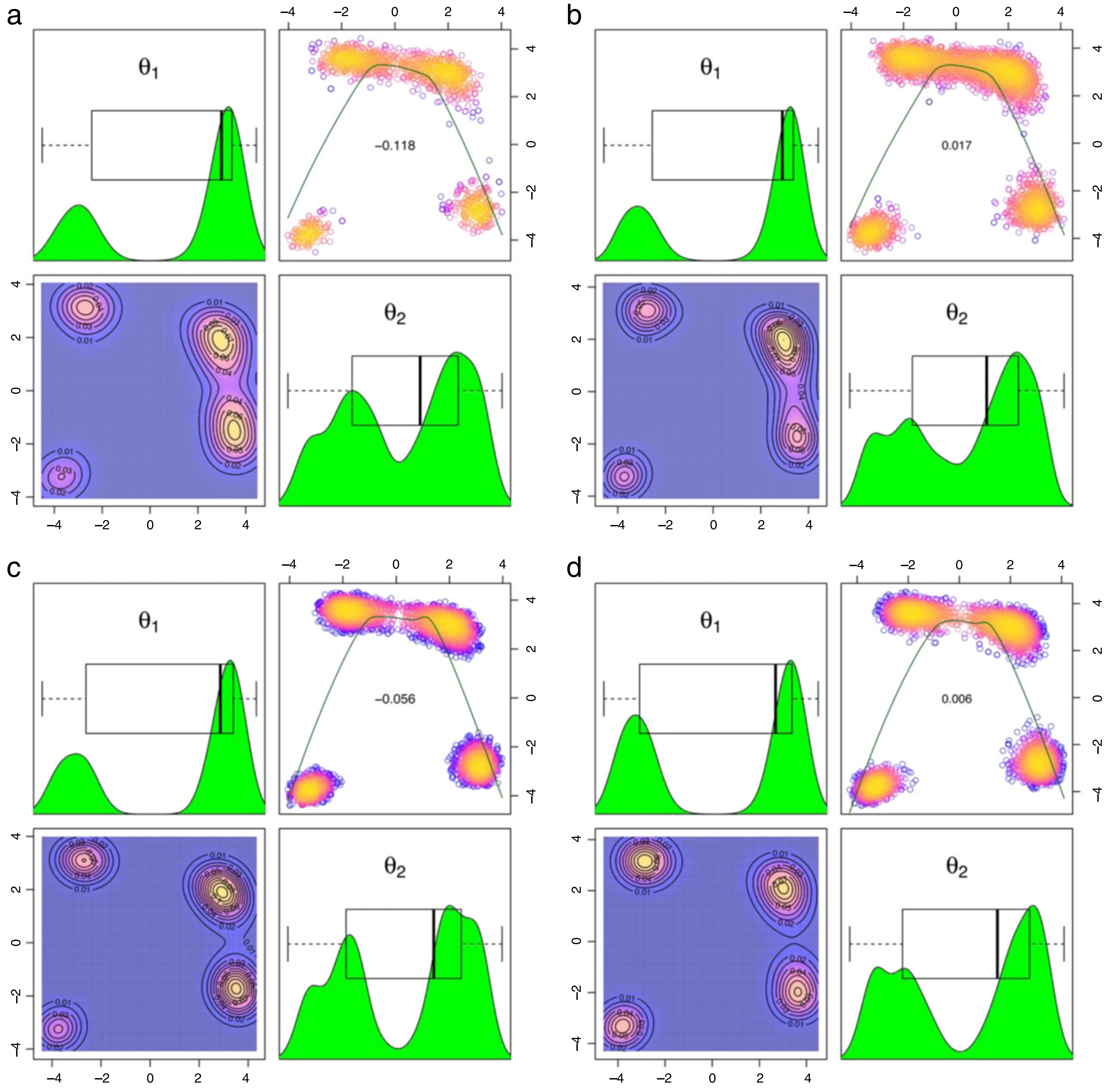

Fig. 3. Reconstructed marginal uncertainties (diagonal panels), posterior samples (upper right panel) and contour plots (lower left panel) of the posterior PDF using (a) K-TMCMC with a tolerance of $\varepsilon = 0.01$, (b) TMCMC with only real function evaluations, (c) KL-TMCMC with a tolerance of $\varepsilon = 0.1$, and (d) L-TMCMC.

points in the two-dimensional parameter space $\underline{\hat{\theta}}_1 = [3, 2]^T, \underline{\hat{\theta}}_2 = [-3.78, -3.28]^T, \underline{\hat{\theta}}_3 = [-2.81, 3.13]^T$ and $\underline{\hat{\theta}}_4 = [3.58, -1.85]^T$. The exact value of the mean of the multimodal posterior distribution is evaluated numerically to be $\underline{\bar{\theta}} = [\bar{\theta}_1, \bar{\theta}_2]^T = [0.9539, 0.3053]^T$.

The X-TMCMC estimates are based on 3000 samples per TMCMC stage with 150 neighbors for the surrogate approximation. The kriging approximation is based on a second-order polynomial basis functions. The initial samples were generated from a uniform distribution in the domain $[-5, 5] \times \cdots \times [-5, 5]$.

The projection of the samples in 2-d parameter space as well as the marginal distribution of the model parameters is shown in Fig. 3 for all four algorithms. Although qualitatively both TMCMC and L-TMCMC identifies the supports related to the four modes, notable differences are observed in the spread and density of the samples around the local

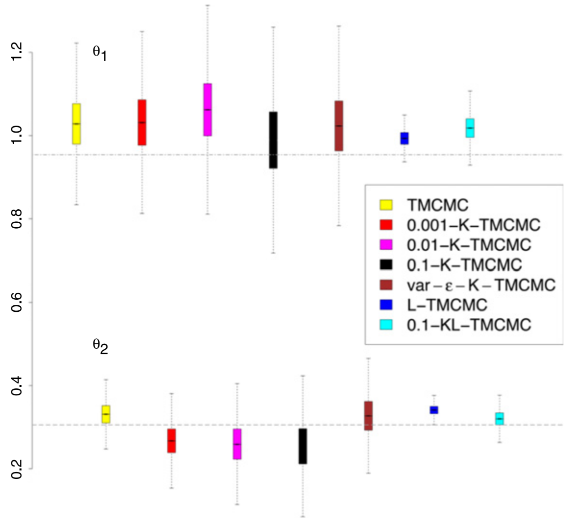

Fig. 4. Statistics (5% quantile, $M(q) - D(q)$, $M(q)$, $M(q) + D(q)$, 95% quantile) for the mean ($q = \mu$) of the marginal posterior distributions of θ_1 and θ_2 estimated from TMCMC, K-TMCMC for $\varepsilon = 0.01$, 0.1 and 0.5, variable-ε K-TMCMC, L-TMCMC, and KL-TMCMC for $\varepsilon = 0.5$.

Table 3
Parametric study of the accuracy of the surrogate estimates for different ε values. FE denotes the number of true function evaluations.

X-TMCMC	ε	$M(\mu)$	$D(\mu)$	$M(\mu)$	$D(\mu)$	FE
		θ_1	θ_1	θ_2	θ_2	
TMCMC	–	1.028	0.072	0.331	0.031	12,000
K-TMCMC	0.001	1.031	0.081	0.267	0.042	6,789
K-TMCMC	0.01	1.062	0.093	0.259	0.054	4,470
K-TMCMC	0.1	0.989	0.101	0.254	0.063	4,121
K-TMCMC variable ε	0.1–0.001	1.023	0.089	0.327	0.051	5,631
L-TMCMC	–	0.993	0.021	0.341	0.013	12,000
KL-TMCMC	0.1	1.018	0.033	0.320	0.021	4,082

modes. However, the K-TMCMC and KL-TMCMC algorithms give results, which are qualitatively very similar to the ones obtained from TMCMC and L-TMCMC algorithms, respectively. Specifically, it is observed that the K-TMCMC identifies correctly and adequately populates the four regions in the parameter space with high probability volume of this challenging multi modal PDF. The quality of the K-TMCMC is considered to be acceptable when compared to the TMCMC estimates. Similar conclusions can be drawn when comparing the quality of the KL-TMCMC with L-TMCMC estimates. This supports the integration of kriging estimates in either TMCMC or L-TMCMC since the quality of the predictions is not significantly deteriorated.

Results for the metrics $M(\mu)$ and $D(\mu)$ of the mean of the marginal posterior distribution of the two model parameters θ_1 and θ_2 are reported in Table 3 for different values of the errors $\varepsilon = 0.001$, 0.01 and 0.1. Reference values are also estimated using TMCMC with 3000 samples per stage. In addition, results for the K-TMCMC with variable ε, selected to be $\varepsilon = 0.1$ for $p_j \leq 0.2$ and $\varepsilon = 0.001$ for $p_j > 0.2$ are also shown. Table 3 also presents the number of real function evaluations needed for the different values of the error ε. Fig. 4 presents the 5% and 95% credible intervals, $M(q) - D(q)$, $M(q)$, $M(q) + D(q)$, for the mean ($q = \mu$) of the marginal distribution of the two parameters. It is observed that the K-TMCMC accuracy in the prediction of the mean of the model parameters is quite satisfactory when compared with the corresponding predictions from the TMCMC even for high values of $\varepsilon = 0.1$. However, predictions of both TMCMC and K-TMCMC algorithms fail to exactly provide good estimates of the ‘exact’ mean values evaluated numerically. The higher the distance of the four modes of the distribution from the mean in this challenging multi-modal PDF, the higher the sensitivity in the sampling mean estimate on the location of the samples in the parameter space. As a result, sampling inaccuracies locally at each mode may affect significantly the estimate of the mean, resulting in small bias in the estimate of TMCMC that corresponds to the smallest size of the confidence interval.

Table 4
Parametric study of the accuracy of the surrogate estimates for different ε value. FE denotes the number of true function evaluations.

X-TMCMC	ε	$M(\mu)$	$D(\mu)$	$M(\mu)$	$D(\mu)$	FE
		θ_1	θ_1	θ_2	θ_2	
TMCMC	–	0.031	0.02	2.56	0.682	12,000
K-TMCMC	0.001	−0.015	0.052	3.16	0.822	2,831
K-TMCMC	0.01	0.052	0.086	4.35	1.281	1,212
K-TMCMC	0.1	0.042	0.094	6.58	2.286	1,075
K-TMCMC variable ε	0.1–0.001	0.038	0.058	4.12	1.158	2,131
L-TMCMC	–	−0.034	0.068	1.42	0.214	16,000
KL-TMCMC	0.1	−0.045	0.035	1.56	0.156	1,883

Results for the L-TMCMC are also shown in Table 3. Note that L-TMCMC performs better than the normal TMCMC, at the additional cost of calculating the derivatives of the function. This is due to the higher acceptance ratios that the L-TMCMC achieves during the chains rejuvenation step (∼81% in this case compared to 31% of the pure TMCMC), which gives a greater diversity to the stage samples. The quality of the estimates provided by the surrogate KL-TMCMC is very good as compared to the L-TMCMC estimates and better that the TMCMC estimates. The number of true function evaluations involved in the surrogate version KL-TMCMC is 30% of the number of true function evaluations required in L-TMCMC.

4.3. *Unidentifiable manifold case using a twisted Gaussian distribution*

The third case study considers an 8-dimensional twisted Gaussian density function [7,54], which is given by the un-normalized density $f_b = f \circ \phi_b$, with $\phi_b(\theta) = (\theta_1, \theta_2 + b\theta_1^2 - 100b, \theta_3, \ldots, \theta_8)$. Here, f signifies the density of a multivariate normal distribution, $N_d(0, \Sigma)$ with $\Sigma = \text{diag}(100, 1, \ldots, 1)$, and ϕ_b is a function that is used to transform f to a twisted distribution resembling a banana-shape support when projecting in the (θ_1, θ_2) parameter space. The value of b is selected to be $b = 0.1$ which corresponds to a highly nonlinear target distribution. The initial samples were generated from a uniform distribution in the domain $[-50, 50] \times \cdots \times [-50, 50]$. All algorithms are run using 3000 samples per TMCMC stage with 150 neighbors for the surrogate approximation.

The projection of the samples in 2-d parameter spaces (θ_1, θ_2), (θ_1, θ_3) and (θ_2, θ_3) as well as the marginal distribution of the model parameters θ_1, θ_2 and θ_3 is shown in Fig. 5 for $\varepsilon = 0.1$ for the four algorithms. TMCMC fails to adequately populate the support of the posterior PDF with high probability volume at the edges, although to a degree it does identify such a support, missing part of the edges. The K-TMCMC algorithm has a similar behavior to TMCMC. Compared to the results provided by L-TMCMC algorithm, it can be observed that the performance of KL-TMCMC is qualitatively acceptable since it identifies and adequately populates with samples the complex support, including the edges, in the (θ_1, θ_2) parameter space with high probability volume. These results indicate that the L-TMCMC and KL-TMCMC improve considerably the sample generation for complex supports such as the ones manifested for unidentifiable parameters. Both algorithms outperform TMCMC and K-TMCMC in populating the edges of the support.

Results for the measures $M(\mu)$ and $D(\mu)$ of the mean of the marginal posterior distribution of the two model parameters θ_1 and θ_2 are reported in Table 4 for the four algorithms. The K-TMCMC results with variable ε correspond to $\varepsilon = 0.1$ for $p_j \leq 0.2$ and $\varepsilon = 0.001$ for $p_j > 0.2$. Fig. 6 presents the 5% and 95% credible intervals, $M(q) - D(q)$, $M(q)$, $M(q) + D(q)$, for the mean ($q = \mu$) of the marginal distribution of the two parameters θ_1 and θ_2.

Similarly to the multi-modal case, L-TMCMC outperforms TMCMC. Predictions from both TMCMC and K-TMCMC algorithms fail to provide acceptable estimates of the exact mean values $\hat{\theta}_1 = \hat{\theta}_2 = 0$. This is because estimates of the mean are highly sensitive on the adequate population of the edges of the support of the posterior PDF. The sampling is inadequate for TMCMC and K-TMCMC as can be seen in Fig. 5(a) and (b), respectively, affecting the accuracy of the mean estimate. However, the quality of the K-TMCMC estimates is considered good compared to the estimates provided by the TMCMC.

The estimates from the L-TMCMC and KL-TMCMC algorithms are significantly closer to the exact ones than the TMCMC estimates. This is due to the local gradient information provided in L-TMCMC, leading to a significantly

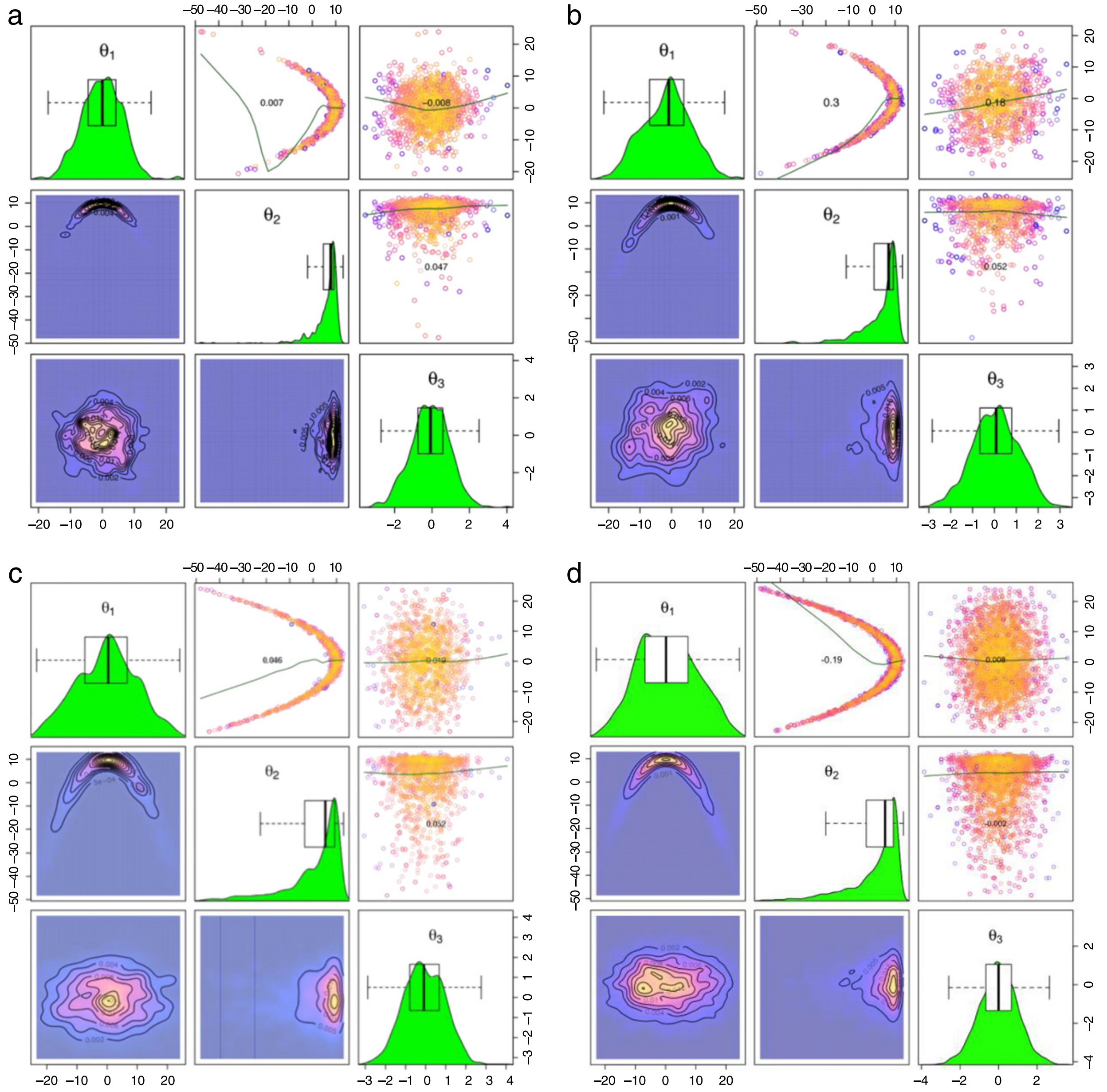

Fig. 5. Reconstructed marginal uncertainties, posterior samples and contour plots of the posterior PDF using (a) K-TMCMC with a tolerance of $\varepsilon = 0.1$, (b) TMCMC with only full model simulations, (c) KL-TMCMC with a tolerance of $\varepsilon = 0.1$, and (d) L-TMCMC.

higher acceptance rate, especially for the samples located in the two reaching tails of the distribution, where the global sample covariance matrix of the TMCMC fails completely to provide a reasonable proposal direction. The KL-TMCMC estimates, in particular, are significantly better than TMCMC and qualitatively give very similar results to the L-TMCMC estimates. The number of true function evaluations in KL-TMCMC is as low as 15% of the corresponding number in L-TMCMC, implicating a substantial reduction in computational time of one order of magnitude for complex systems requiring large computing time to perform a single model simulation.

5. Application: parameter estimation in structural dynamics

The accuracy and computational efficiency of the proposed K-TMCMC algorithm is demonstrated by updating a high fidelity finite element model of the Metsovo bridge. A description of the bridge can be found in [55]. A detailed

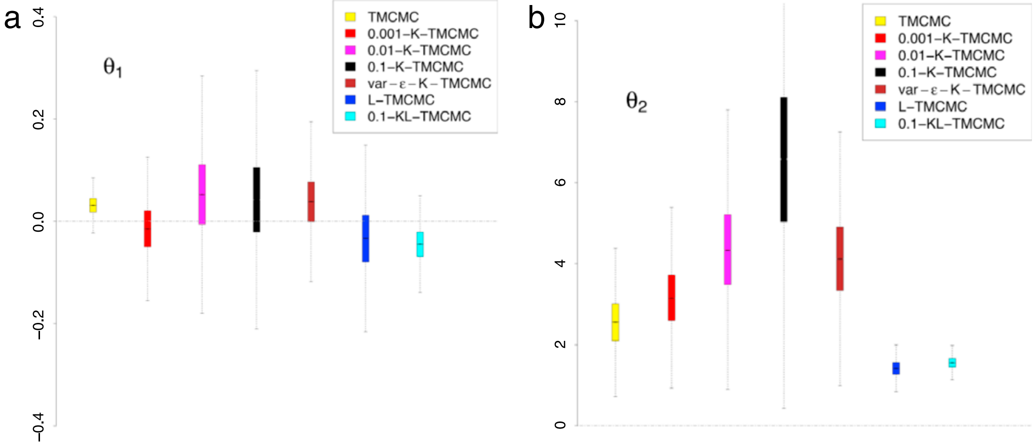

Fig. 6. Statistics (5% quantile, $M(q) - D(q)$, $M(q)$, $M(q) + D(q)$, 95% quantile) for the mean ($q = \mu$) of the marginal posterior distributions of (a) θ_1 and (b) θ_2 estimated from TMCMC, K-TMCMC for $\varepsilon = 0.01$, 0.1 and 0.5, variable-ε K-TMCMC, L-TMCMC, and KL-TMCMC for $\varepsilon = 0.5$.

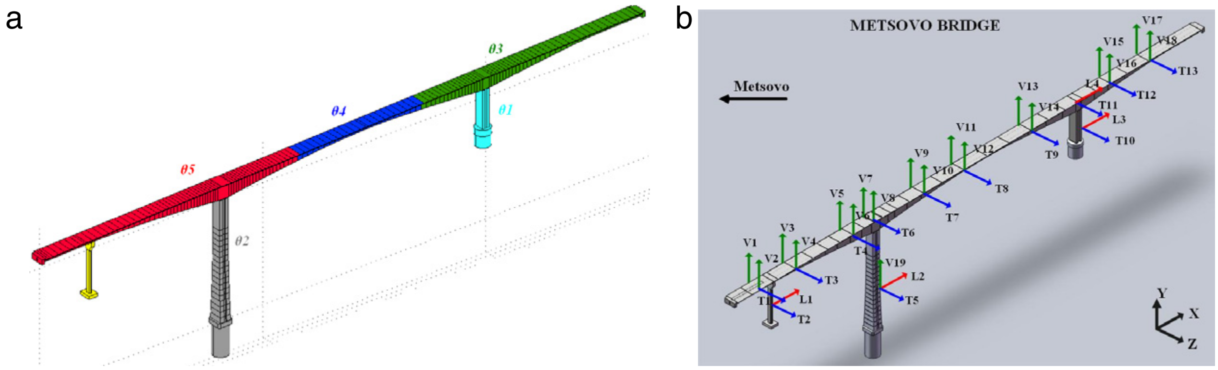

Fig. 7. (a) Parameterization of the finite element model of Metsovo bridge. (b) Sensor locations.

finite element model of the bridge is created using 3-dimensional tetrahedron quadratic Lagrange finite elements. A coarse mesh is chosen to predict the lowest 20 modal frequencies and mode shapes of the bridge. The finite element model has 97,636 elements and 562,101 degrees of freedom (DOF).

The finite element model is parameterized using five parameters associated with the modulus of elasticity of one or more structural components as shown in Fig. 7(a). Specifically, the first two parameters θ_1 and θ_2 account respectively for the modulus of elasticity of two pier components of the bridge. The parameter θ_3, θ_4 and θ_5 account for the modulus of elasticity of structural components of the deck. The model parameters are introduced to scale the nominal values of the properties that they model so that the value of the parameters equal to one corresponds to the nominal value of the finite element model.

Model reduction is used to reduce the computational effort to manageable levels. Specifically, a parameterization-consistent component mode synthesis technique [55] is applied. The dynamics of the six structural components introduced in Fig. 7(a) is represented by the lowest fixed-base modes and the interface constrained modes. In this study the kept modes in each component are the ones with modal frequency less than $\omega_{\max} = \rho\omega_c$, where ω_c is the cut-off frequency selected as the 20th modal frequency ($\omega_c = 4.55$ Hz) of the nominal finite element model, and the ρ value affects the computational efficiency and accuracy of the CMS technique. A further reduction in the number of generalized coordinates is achieved by retaining only a fraction of the constrained interface modes with frequency less than $\omega_{\max} = \nu\omega_c$. In this study, the following choices are made $\rho = 5$ and $\nu = 200$ resulting in a number of 592 generalized coordinates for the reduced model, with errors in the estimates for the lowest 20 modal frequencies to

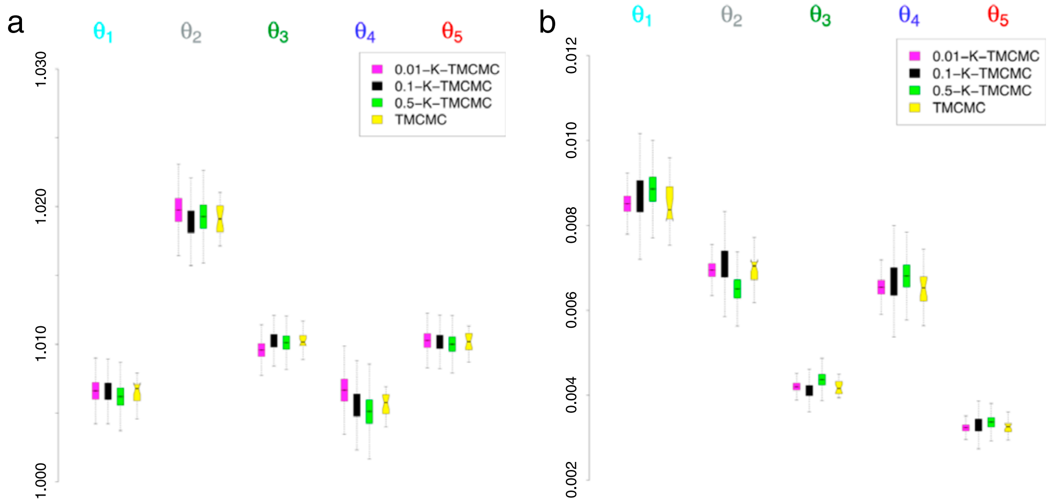

Fig. 8. Statistics (5% quantile, $M(q) - D(q)$, $M(q)$, $M(q) + D(q)$, 95% quantile) for (a) the mean ($q = \mu$) and (b) the standard deviation ($q = \sigma$) of the marginal posterior distributions of five model parameters as estimated from TMCMC, and K-TMCMC algorithm for $\varepsilon = 0.01$, 0.1 and 0.5.

be less than 0.02%. Thus, using CMS a drastic reduction in the number of generalized coordinates is obtained which can exceed three orders of magnitude, without sacrificing in accuracy with which the lowest model frequencies are computed.

Herein, the effectiveness of the proposed kriging technique to further reduce the computational effort when applied for the parameter estimation of the high-fidelity finite element model of the bridge is demonstrated. The estimation of the parameter values and their uncertainties of the finite element model is based on modal frequencies and mode shapes. Simulated, noise contaminated, “measured” modal frequencies and mode shapes are generated by adding a 1% and 3% Gaussian noise to the modal frequencies and modeshape components, respectively, predicted by the nominal non-reduced FE models. 35 sensors are placed on the bridge to monitor vertical and transverse accelerations. Fig. 7(b) shows the sensors' locations used in the identification. The measured data containing the values of the ten lowest modal frequencies and modeshapes are considered.

Applying the Bayesian formulation, the likelihood is given by (18) where the measure of fit function [55]

$$J(\underline{\theta}, M) = \frac{1}{\sigma^2}\sum_{r=1}^{m} \frac{\left[\lambda_r\left(\underline{\theta}\right) - \hat{\lambda}_r\right]^2}{\hat{\lambda}_r} + \frac{w}{\sigma^2}\sum_{r=1}^{m}\frac{\left\|\beta_r\left(\underline{\theta}\right)\underline{\varphi}_r\left(\underline{\theta}\right) - \underline{\hat{\varphi}}_r\right\|^2}{\left\|\underline{\hat{\varphi}}_r\right\|^2} \tag{18}$$

consists of two terms. The first term quantifies the discrepancies between the experimentally identified eigenvalues $\hat{\lambda}_r = \hat{\omega}_r^2$ (the square of the modal frequencies $\hat{\omega}_r$) and the model predicted eigenvalues $\lambda_r\left(\underline{\theta}\right)$, while the second term quantifies the discrepancies between the experimentally identified modeshapes $\underline{\hat{\varphi}}_r$ and the model predicted modeshapes $\underline{\varphi}_r\left(\underline{\theta}\right)$, with $\beta_r(\underline{\theta}) = \underline{\hat{\varphi}}_r^T \underline{\varphi}_r(\underline{\theta}) / \left\|\underline{\varphi}_r(\underline{\theta})\right\|^2$ a normalization constant that guaranties that the measured mode shape $\underline{\hat{\varphi}}_r$ at the measured DOF is closest to the model mode shape $\beta_r(\underline{\theta})\underline{\varphi}_r(\underline{\theta})$ predicted by the particular value of $\underline{\theta}$. In (18) the variable m denotes the number of observed modes. The factor $|C|$ in (4) is $|C| = (\sigma^2)^{m(N_0+1)}$, where N_0 denotes the number of measured DOF. The variable σ is included as unknown in the parameter set $\underline{\theta}$. The factor w is used to weight the relative contribution of the second term in relation to the first term. In the numerical results it is taken to be $w = 1$, although it could generally be included in the set $\underline{\theta}$ to be identified like the parameter σ.

A uniform prior distribution was used with bounds $[0.2, 2] \times \cdots \times [0.2, 2]$ for the five structural model parameters and [0.001,1] for the prediction error parameter σ. Results for simple measures, such as the mean and the standard deviation of the marginal posterior distribution of each one of the five model parameters, are compared in Fig. 8 using TMCMC and K-TMCMC algorithm with kriging for error values $\varepsilon = 0.01$, 0.1 and 0.5. The surrogate estimates in this case are based on a second-order kriging approximation. Both TMCMC and K-TMCMC set-ups include 1000

Table 5
Number of function evaluations for different values of ε for the bridge model. FE denotes the number of true function evaluations. LE denotes Log-Evidence.

X-TMCMC	ε	M(LE)	D(LE)	FE
TMCMC	–	1693.1	0.12	15,000
K-TMCMC	0.01	1694.6	0.74	11,136
K-TMCMC	0.1	1695.0	1.13	4,423
K-TMCMC	0.5	1695.1	0.99	2,232

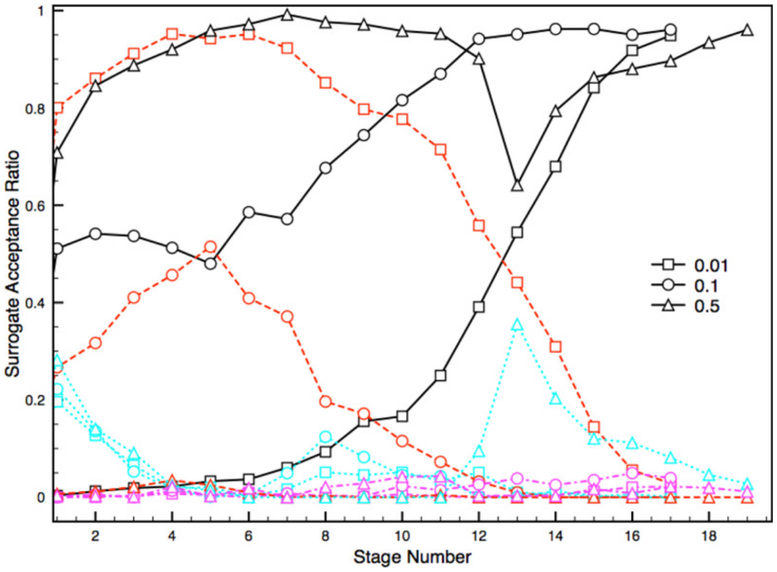

Fig. 9. Insight into various reasons resulting to a surrogate approximation rejection at different error tolerance levels ε. Different symbols indicate different ε. (Black) denotes the total accepted surrogate ratio, whereas rejected surrogates are decomposed on: unmet ε-tolerance level criteria (red, dashed curve), no possible interpolation with neighboring points (magenta, dashed–dotted curve), and overshooting the 95% quantile range (cyan, dotted curve). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

samples per stage and 60 neighbors for the surrogate approximation. The estimates are based on $N = 50$ independent runs with parameter values of the TMCMC algorithm chosen to be the same as the ones used for the K-TMCMC. Results for the number of mean full model simulations, excluding surrogate evaluations, are also reported in Table 5 for the different values of the errors ε used for the kriging estimates.

It is observed that the K-TMCMC prediction of the mean and the standard deviation of the marginal distribution of the model parameters is very close to the results obtained using the TMCMC algorithm, while the accuracy in the predictions is relatively insensitive to the error tolerance values up to $\varepsilon = 0.5$. A slightly higher uncertainty in the values of the mean and the standard deviation observed for the K-TMCMC algorithm is considered acceptable. This adequate accuracy of the K-TMCMC algorithm comes with a significant reduction in the number of full model simulations up to 84% for the $\varepsilon = 0.5$ case as reported in Table 5. For the evidence values, also shown in Table 5, it can be observed that the estimates from the surrogate K-TMCMC algorithm are close to the ones obtained from the TMCMC algorithm.

An insight into the effect of the tolerance ε on the acceptance–rejection of a kriging approximation as a function of K-TMCMC stages is presented in Fig. 9, along with the different criteria that prevail in rejecting a surrogate approximation. In the case of small ε value ($\varepsilon = 0.01$), the support point density in the initial K-TMCMC stages does not suffice to provide a good surrogate approximation over the PDF support of these support points and most of the surrogate points are not accepted.

The criterion that prevails for rejecting the surrogate estimates is failure to meet the ε-tolerance criterion since this tolerance level is too small. Rejection of the surrogate estimates in the initial K-TMCMC stages due to overshooting the 95% quantile range may also occur for a percentage (up to 30%) of the rejected samples. At later K-TMCMC stages, say after stage 10 in Fig. 9, a steep increase of the accepted surrogate points is observed. This occurs due to the accumulation of support points from previous stages, resulting in high density of support points. Hence, the size of the PDF support covered by the neighborhood support points in the parameter space is smaller, the function to be approximated is smoothly varying over this smaller area and thus the surrogate estimate is significantly more accurate, meeting the strict ε-tolerance criterion. Rejection of surrogate points due to overshooting the 95% confidence level seems to be less likely to occur in later stages of the TMCMC algorithm. The total surrogate acceptance ratio is approximately 34%.

In the case of large ε value ($\varepsilon = 0.5$), for the initial K-TMCMC stages a very large percentage of the surrogate points, ranging from 70% in the initial stages to 98% in the stages 6–11, are accepted since the tolerance criterion is very high and it is met by most surrogate points. As a result, only a very low percentage of full model simulations are performed and then used as support points at the next stages. This absence of a sufficient number of full model simulations in the immediate neighborhood of the new MCMC sample creates a problem, which is evident in stage 13 (Fig. 9) by a sharp drop in the acceptance of surrogate points. The few support points are significantly further away from the new MCMC sample so that the surrogate estimate, although it meets the ε-tolerance criterion, results in overshooting the 95% quantile range as it can be clearly seen in the K-TMCMC stages 13–16 in Fig. 9. New support points are thus generated in the intermediate posterior PDF support at the current stage, which are adequate to be used as support points for the next stages without violating the criteria set. Note also that the sharp drop of the acceptance of the surrogate points comes at the extra cost of two additional stages to achieve convergence. In any case, due to the high ε value, rejection of surrogate points due to unmet ε-tolerance criterion is less likely to occur in this case. The total surrogate acceptance ratio is as high as 84%.

The intermediate ε value case ($\varepsilon = 0.1$) seems to be more balanced with a total surrogate ratio of 68%. Almost 50% of the surrogate points are rejected in the initial K-TMCMC stages. The reason that prevails for rejecting the surrogate estimates in the first few K-TMCMC stages is shared between the unmet ε-tolerance criterion and the overshooting of the 95% quantile range. At the later K-TMCMC stages the rejection is mainly due to the unmet ε-tolerance criterion. As the K-TMCMC stages progresses, more and more surrogate points are accepted and also an adequate number of support points are also accumulated in the important PDF support of each intermediate posterior PDF. With increasing K-TMCMC stages, the new support point additions seem to cater for the requested accuracy fairly quickly. In the last stages a very large percentage of the surrogate points are accepted due to the fact that the function is smoothly varying in the PDF support covered by the neighboring support points.

6. Conclusions

We introduce an adaptive kriging technique within TMCMC in an effort to achieve substantial reductions in the number of computationally intensive full model simulations required in the Bayesian inverse modeling. The proposed adaptive kriging exploits the annealing properties of TMCMC and uses the distribution of the spatial structure of TMCMC samples to replace a large number of model simulations by approximate yet fast interpolating estimates. A modified TMCMC algorithm with Langevin adjusted MCMC kernels is also proposed to improve the efficiency for complex posterior distributions with challenging supports such as multi-modal PDFs and manifolds emerging from unidentifiable parameters. A substantial reduction in the number of full model runs was achieved for up to 10-fold order for uni-modal PDFs and up to 3-fold for challenging posterior PDFs. Both algorithms are highly parallel and can exploit the availability of high performance computing facilities to further reduce drastically the time-to-solution. Note that due to the algorithmic similarities between TMCMC and the static particle filter inference based methods e.g [56], the surrogate additions leading to X-TMCMC could also be used within a particle filtering framework.

We remark that the proposed surrogate algorithm can also be extended for uncertainty propagation by replacing a percentage of the full model simulations of an output quantity of interest on the TMCMC or L-TMCMC samples by kriging estimates. A drawback of the proposed method arises at large parameter space dimensions. This is partly due to the difficulty of the multidimensional optimization problem introduced during the kriging training. This can be treated for medium dimensions with the introduction of CMA-ES [57]. Moreover, the possibility of a point lying inside of the convex hull of support points suffers from the course of dimensionality, thus the support point density is adapted to be higher, reducing computational benefits.

Acknowledgments

This research is implemented under the “ARISTEIA” Action of the “Operational Programme Education and Lifelong Learning” and is co-funded by the European Social Fund (ESF) and Greek National Resources. We also acknowledge support by ETH-Zurich.

References

[1] J.L. Beck, Bayesian system identification based on probability logic, Struct. Control Health 17 (2010) 825–847.

[2] J.L. Beck, L. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, ASCE J. Eng. Mech. 124 (1998) 455–461.

[3] L. Tierney, J.B. Kadane, Accurate approximations for posterior moments and marginal densities, J. Amer. Statist. Assoc. 81 (1986) 82–86.

[4] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092.

[5] W.K. Hastings, Monte-Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109.

[6] P.J. Green, A. Mira, Delayed rejection in reversible jump Metropolis–Hastings, Biometrika 88 (2001) 1035–1053.

[7] H. Haario, M. Laine, A. Mira, E. Saksman, DRAM: efficient adaptive MCMC, Stat. Comput. 16 (2006) 339–354.

[8] J.L. Beck, S.K. Au, Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation, ASCE J. Eng. Mech. 128 (2002) 380–391.

[9] S.H. Cheung, J.L. Beck, Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters, ASCE J. Eng. Mech. 135 (2009) 243–255.

[10] P. Angelikopoulos, C. Papadimitriou, P. Koumoutsakos, Data driven, predictive molecular dynamics for nanoscale flow simulations under uncertainty, J. Phys. Chem. B 117 (2013) 14808–14816.

[11] J. Martin, L.C. Wilcox, C. Burstedde, O. Ghattas, A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion, SIAM J. Sci. Comput. 34 (2012) A1460–A1487.

[12] G.E.P. Box, K.B. Wilson, On the experimental attainment of optimum conditions, J. R. Stat. Soc. Ser. B 13 (1951) 1–45.

[13] J.-A. Ting, A. D’Souza, S. Vijayakumar, S. Schaal, Efficient learning and feature selection in high-dimensional regression, Neural Comput. 22 (2009) 831–886.

[14] R.G. Ghanem, P.D. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer, Berlin, 1991.

[15] J. Sacks, S.B. Schiller, W.J. Welch, Designs for computer experiments, Technometrics 31 (1989) 41–47.

[16] N.D. Lagaros, M. Papadrakakis, Learning improvement of neural networks used in structural optimization, Adv. Eng. Softw. 35 (2004) 9–25.

[17] Y. Jin, Surrogate-assisted evolutionary computation: recent advances and future challenges, Swarm Evol. Comput. 1 (2011) 61–70.

[18] S. Kern, N. Hansen, P. Koumoutsakos, Local meta-models for optimization using evolution strategies, in: Lecture Notes in Computer Science, vol. 4193, 2006, pp. 939–948.

[19] M. Papadrakakis, N.D. Lagaros, Reliability-based structural optimization using neural networks and Monte Carlo simulation, Comput. Methods Appl. Math. 191 (2002) 3491–3507.

[20] M. Papadrakakis, V. Papadopoulos, N.D. Lagaros, Structural reliability analysis of elastic–plastic structures using neural networks and Monte Carlo simulation, Comput. Methods Appl. Math. 136 (1996) 145–163.

[21] V. Dubourg, B. Sudret, F. Deheeger, Metamodel-based importance sampling for structural reliability analysis, Probab. Eng. Mech. 33 (2013) 47–57.

[22] C. Bucher, T. Most, A comparison of approximate response functions in structural reliability analysis, Probab. Eng. Mech. 23 (2008) 154–163.

[23] B. Echard, N. Gayton, M. Lemaire, N. Relun, A combined importance sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models, Reliab. Eng. Syst. Saf. 111 (2013) 232–240.

[24] V. Papadopoulos, D.G. Giovanis, N.D. Lagaros, M. Papadrakakis, Accelerated subset simulation with neural networks for reliability analysis, Comput. Methods Appl. Math. 223 (2012) 70–80.

[25] J.M. Bourinet, F. Deheeger, M. Lemaire, Assessing small failure probabilities by combined subset simulation and Support Vector machines, Struct. Saf. 33 (2011) 343–353.

[26] M.A. Valdebenito, G.I. Schueller, A survey on approaches for reliability-based optimization, Struct. Multidiscip. Optim. 42 (2010) 645–663.

[27] V. Dubourg, B. Sudret, J.M. Bourinet, Reliability-based design optimization using kriging surrogates and subset simulation, Struct. Multidiscip. Optim. 44 (2011) 673–690.

[28] J.Y. Ching, Y.C. Chen, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, ASCE J. Eng. Mech. 133 (2007) 816–832.

[29] D. Buche, N.N. Schraudolph, P. Koumoutsakos, Accelerating evolutionary algorithms with Gaussian process fitness function models, IEEE Trans. Syst. Man Cybern. C 35 (2005) 183–194.

[30] S. Kern, S. Müller, N. Hansen, D. Büche, J. Ocenasek, P. Koumoutsakos, Learning probability distributions in continuous evolutionary algorithms—a comparative review, Nat. Comput. 3 (2004) 77–112.

[31] J. Ocenasek, S. Kern, N. Hansen, P. Koumoutsakos, A mixed Bayesian optimization algorithm with variance adaptation, Lect. Notes Comput. Sci. 3242 (2004) 352–361.

[32] D.I. Papadimitriou, C. Papadimitriou, Optimal sensor location for model parameter estimation in CFD, in: Proc. 21st AIAA Computational Fluid Dynamics Conference, American Institute of Aeronautics and Astronautics, 2013, pp. 1–10.

[33] E. Ntotsios, C. Papadimitriou, Multi-objective optimization algorithms for finite element model updating, in: Proc. ISMA2008 International Conference on Noise and Vibration Engineering, Leuven, 2008, pp. 1895–1909.

[34] H.P. Flath, L.C. Wilcox, V. Akcelik, J. Hill, B.V. Waanders, O. Ghattas, Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial hessian approximations, SIAM J. Sci. Comput. 33 (2011) 407–432.

[35] P. Angelikopoulos, C. Papadimitriou, P. Koumoutsakos, Bayesian uncertainty quantification and propagation in molecular dynamics simulations: a high performance computing framework, J. Chem. Phys. 137 (2012) 144103.

[36] J.L. Beck, K.V. Yuen, Model selection using response measurements: Bayesian probabilistic approach, ASCE J. Eng. Mech. 130 (2004) 192–203.

[37] K.-V. Yuen, Bayesian Methods for Structural Dynamics and Civil Engineering, Wiley-Vch Verlag, 2010.

[38] H. Haario, E. Saksman, J. Tamminen, An adaptive Metropolis algorithm, Bernoulli 7 (2001) 223–242.

[39] H. Haario, M. Laine, A. Mira, E. Saksman, DRAM: efficient adaptive MCMC, Stat. Comput. 16 (2006) 339–354.

[40] C. Braak, A Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces, Stat. Comput. 16 (2006) 239–249.

[41] C.J. Ter Braak, J.A. Vrugt, Differential evolution Markov Chain with snooker updater and fewer chains, Stat. Comput. 18 (2008) 435–446.

[42] M. Drugan, D. Thierens, Recombination operators and selection strategies for evolutionary Markov chain Monte Carlo algorithms, Evol. Intell. 3 (2010) 79–101.

[43] A. Doucet, S. Godsill, C. Andrieu, On sequential Monte Carlo sampling methods for Bayesian filtering, Stat. Comput. 10 (2000) 197–208.

[44] C.C. Hsu, C.C. Wong, H.C. Teng, N.J. Li, C.Y. Ho, Localization of mobile robots via an enhanced particle filter incorporating tournament selection and Nelder–Mead simplex search, Int. J. Innov. Comput. Inf 7 (2011) 3725–3737.

[45] G.O. Roberts, J.S. Rosenthal, Optimal scaling of discrete approximations to Langevin diffusions, J. R. Stat. Soc. Ser. B 60 (1998) 255–268.

[46] C.H. Mehl, An empirical study of an adaptive Langevin algorithm for bounded target densities, J. Data Sci. 11 (2013) 501–536.

[47] G. Li, M. Li, S. Azarm, S. Al Hashimi, T. Al Ameri, N. Al Qasas, Improving multi-objective genetic algorithms with adaptive design of experiments and online metamodeling, Struct. Multidiscip. Optim. 37 (2009) 447–461.

[48] M. Li, G. Li, S. Azarm, A kriging metamodel assisted multi-objective genetic algorithm for design optimization, J. Mech. Des. 130 (2008) 1–10.

[49] S.N. Lophaven, H.B. Nielsen, J. Søndergaard, DACE: A MATLAB Kriging Toolbox, in, DTU, DK-2800 Kgs. Lyngby—Denmark, 2002.

[50] R.M. Lewis, V. Torczon, Pattern search algorithms for bound constrained minimization, SIAM J. Optim. 9 (1999) 1082–1099.

[51] C.B. Barber, D.P. Dobkin, H. Huhdanpaa, The Quickhull algorithm for convex hulls, ACM Trans. Math. Software 22 (1996) 469–483.

[52] L. Lin, K.F. Liu, J. Sloan, A noisy Monte Carlo algorithm, Phys. Rev. D 61 (2000).

[53] C. Andrieu, G.O. Roberts, The pseudo-marginal approach for efficient Monte Carlo computations, Ann. Statist. 37 (2009) 697–725.

[54] H. Haario, E. Saksman, J. Tamminen, Adaptive proposal distribution for random walk Metropolis algorithm, Comput. Statist. 14 (1999) 375–395.

[55] C. Papadimitriou, D.C. Papadioti, Component mode synthesis techniques for finite element model updating, Comput. & Structures 126 (2013) 15–28.

[56] N. Chopin, A sequential particle filter method for static models, Biometrika 89 (2002) 539–551.

[57] N. Hansen, S.D. Muller, P. Koumoutsakos, Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), Evol. Comput. 11 (2003) 1–18.