Accurate solution of Bayesian inverse uncertainty quantification problems combining reduced basis methods and reduction error models *

A. Manzoni† S. Pagani‡ T. Lassila§

December 9, 2015

Abstract

Computational inverse problems related to partial differential equations (PDEs) often contain nuisance parameters that cannot be effectively identified but still need to be considered as part of the problem. The objective of this work is to show how to take advantage of a reduced order framework to speed up Bayesian inversion on the identifiable parameters of the system, while marginalizing away the (potentially large number of) nuisance parameters. The key ingredients are twofold. On the one hand, we rely on a reduced basis (RB) method, equipped with computable a posteriori error bounds, to speed up the solution of the forward problem. On the other hand, we develop suitable reduction error models (REMs) to quantify in an inexpensive way the error between the full-order and the reduced-order approximation of the forward problem, in order to gauge the effect of this error on the posterior distribution of the identifiable parameters. Numerical results dealing with inverse problems governed by elliptic PDEs in the case of both scalar parameters and parametric fields highlight the combined role played by RB accuracy and REM effectivity.

1 Introduction

The efficient solution of inverse problems governed by partial differential equations (PDEs) is a relevant challenge from both a theoretical and a computational standpoint. In these problems, unknown or uncertain parameters related to a PDE model have to be estimated from indirect observations of suitable quantities of interest. Being able to design efficient solvers for inverse problems is paramount in several applications, ranging from life sciences (e.g., electrical impedance tomography [26, 32], characterization of myocardial ischemias and identification of blood flow parameters [42, 21]), to material sciences (e.g. scattering problems [10] or subsurface damage detection [3]) and environmental sciences (e.g., identification of permeability in groundwater flows [29], or basal sliding in ice dynamics [34]).

In a parametrized context, a forward problem consists in evaluating some outputs of interest (depending on the PDE solution) for specified parameter inputs. Whenever some parameters are uncertain, they can be identified by considering either a deterministic or a statistical framework. In the former case, we solve an optimization problem by minimizing (e.g. in the least-square sense) the discrepancy between the output quantities predicted by the PDE model and the observations. In the latter case, we assess the relative likelihood of the parameters which are consistent with the

*We are grateful to Prof. Anna Maria Paganoni (Politecnico di Milano) for her valuable insights and careful remarks, and to Prof. Alfio Quarteroni (EPFL and Politecnico di Milano) for his useful suggestions. This work was supported by the Italian "National Group of Computing Science" (GNCS-INDAM).

†CMCS-MATHICSE-SB, Ecole Polytechnique Fédérale de Lausanne, Station 8, CH-1015 Lausanne, Switzerland, andrea.manzoni@epfl.ch (corresponding author)

‡MOX, Dipartimento di Matematica, Politecnico di Milano, P.za Leonardo da Vinci 32, I-20133 Milano, Italy, stefano.pagani@polimi.it

§CISTIB, Department of Electronic and Electrical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom, t.lassila@sheffield.ac.uk

observed output, that is, we need to quantify uncertainties associated with the identifiable parameters due to measurement errors and to nuisance parameters. Such a problem can be referred to as *inverse uncertainty quantification (UQ)* problem. By relying on a *Bayesian approach*, we model the unknown parameters as random variables and characterize their *posterior* probability density function, which includes information both on *prior* knowledge on the parameters distribution and on the model used to compute the PDE-based outputs [44]. In this way, inference of unknown parameters from noisy data accounts for the information coming from (possibly complex and nonlinear) physical models [45, 19]. Nevertheless, we need to face some key numerical challenges, related to *(i)* parametric dimensionality, *(ii)* slow Markov chain Monte Carlo (MCMC) convergence and *(iii)* many forward queries; see e.g. [16, 40] for a general introduction to MCMC techniques. While the first issue can be addressed by considering a parametric reduction (e.g. through a modal decomposition exploiting a Karhunen-Loève expansion [26] or suitable greedy algorithms [25]), several techniques have emerged in the last decade to speed up both MCMC sampling algorithms [28, 23] and the solution of the forward problem.

In this work we focus on this latter aspect, showing how a low-dimensional, projection-based *reduced-order model (ROM)* can be exploited to speed up the solution of Bayesian inverse problems dealing with parametrized PDEs. Among projection-based ROMs, reduced basis (RB) methods built through greedy algorithms [6, 21] or proper orthogonal decomposition (POD) [14, 26, 29, 39] have been already successfully exploited in this field during the last decade. Such methods approximate the solution of the forward problem by a handful of snapshots of the full-order problem – i.e., solutions computed for properly selected parameter values; as a result, each forward query becomes relatively inexpensive and the overall computational cost of Bayesian inversion can be greatly reduced. Very recently, a possible way to compute snapshots adaptively from the posterior distribution, yielding a data-driven ROM, has been shown in [12]. Proper generalized decomposition has also been combined with stochastic spectral methods to deal with dynamical systems in the presence of stochastic parametric uncertainties [11]. Besides projection-based ROMs, a low-fidelity model can be also built according to simplified physics, coarser discretizations, or multiscale formulations. Such a model can also be equipped with correction functions using global polynomials in term of the stochastic parameters. For instance, non-intrusive polynomial chaos using orthogonal polynomials [15] and stochastic collocation using interpolation polynomials [2, 47] have been developed in conjunction with physics-based low fidelity models [31]. See, e.g., [6] for a detailed discussion on the use of low-fidelity or surrogate models to speed up inverse problems.

A relevant question, arising when a ROM is exploited to solve inverse UQ problems, is related to the propagation of *reduction errors* along the inversion process. In other words, we need to quantify those uncertainties due to the use of a ROM and associated with the identifiable parameters, to which we can refer to as *ROM uncertainty*. This latter can be seen as a form of *epistemic uncertainty* (i.e., pertaining to uncertainty about a deterministic quantity that nevertheless cannot be fully ascertained) and its quantification is essential in order to obtain precise and robust solutions to the inverse UQ problem. ROM uncertainty is indeed quite similar to the so-called *model form uncertainty*, arising whenever a limited understanding of the modeled process is available [18, 17]. Concerning the modeling of the reduction error itself, different approaches to approximation have been considered very recently: the so-called approximation error model [1, 26], Gaussian process (GP)-based calibration (or GP machine learning) [39], interpolant-based error indicators [33], and regression-based error indicators [13]. In all of these cases, a statistical representation of the ROM error through calibration experiments is used to model the difference between the full-order and the lower-order model. In the approximation error model, an additive Gaussian noise term is introduced in the model to represent the reduction error; in GP machine learning a Gaussian process is used, estimating the covariance function through the calibration experiments. However, error characterization through well-defined probability distributions may impose a large amount of (sometimes unjustified) statistical structure, thus leading to inaccurate results. For instance, in the approximation error model, the uncertainty yielded by the ROM is described as an independent Gaussian error term, which might be a restrictive assumption (see Sect. 7). On the other hand, GP machine learning techniques entail more severe computational costs; a possible reduction technique in this context has been recently presented in [13] and takes advantage of a posteriori error bounds, weighted with a Gaussian process to generate a more accu-

rate correction. One of our proposed reduction error models (see Sect. 4.3) shares some similarities with this approach, but enables variance estimations directly from a regression analysis and an easier treatment of sign effects/corrections, this latter aspect being not treated in [13].

In contrast to the approaches taken in [13, 33], where the objective was to construct reduction error models that could be used to train or adapt the ROM accordingly in the case that no other error estimator was readily available, our proposed approach is instead aimed at the accurate solution of inverse problems using ROMs, and may use existing ROM error bounds as additional information, if available. In previous work [21] we have shown that using low-fidelity ROMs as surrogates for solving inverse problems can lead to biased and overly optimistic posterior distributions. Our goal is to exploit the existing features of the ROM, such as rigorous a posteriori error estimators, and then to correct for the reduction error within the Bayesian inversion process through suitable *reduction error models* (REMs). By extending some preliminary ideas presented in [21], we show how to take advantage of RB error bounds to gain a strong computational speedup without necessarily increasing the size of the reduced basis or the resulting online computational cost. Hence, we *(i)* propose three reduction error models (REMs) to manage ROM uncertainties, *(ii)* include them within the Bayesian computational framework, and *(iii)* show how they allow to obtain posterior distributions that are free of bias and more reliable than those provided by the ROM alone.

First, we test the approximation error model of [1] (called REM-1 from here on), which requires no a posteriori error bounds to be constructed but however may perform quite poorly when dealing with problems involving nuisance parameters and/or error empirical distributions far from the Gaussian case. Then, we present two alternatives. In the second approach (REM-2), a radial basis interpolant allows to evaluate a surrogate reduction error over the parameter space, relying on few calibration experiments. A special feature of this interpolant-based approach compared to other presented in literature is that instead of directly interpolating the ROM error we interpolate the (inverse) effectivity of the ROM error estimator. In the third approach (REM-3), we fit a log-linear regression model to explain the variability of the reduction errors by means of the corresponding error bounds.

The structure of the paper is as follows. In Sect. 2 we provide a general formulation of the class of problems we are interested to, whereas in Sect. 3 we recall the main properties of RB methods. In Sect. 4, we introduce the three reduced error models considered, and we show how to incorporate them into the Bayesian estimator and in Sect. 5 we sum up the proposed computational procedure for the solution of a reduced inverse problem. In Sect. 6 we prove some results related to the effectivity of the corrections made by the REMs, on the reduced-order likelihood function and the corresponding posterior distribution. We assess the performance of the proposed framework on two numerical examples of inverse problems governed by elliptic PDEs in Sect. 7, and finally draw some conclusions in Sect. 8.

2 Bayesian inverse problems governed by PDEs

Let us introduce the abstract formulation of the forward problem modeling our system of interest, and recall the basic features of the Bayesian framework for the solution of inverse problems governed by PDEs.

2.1 Abstract formulation of forward problem

In this paper we consider systems modelled by linear parametrized PDEs and (a set of) linear outputs of interest. Let us denote by X a Hilbert space of functions defined over a domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$. In the case of second-order, elliptic PDE operators typically $(H_0^1(\Omega))^\nu \subset X \subset (H^1(\Omega))^\nu$, where $\nu = 1$ (resp. $\nu = d$) in the case of scalar (resp. vectorial) problems. Here we restrict ourselves to the scalar case; the extension to the vectorial case is straightforward, concerning the formulation of an inverse problem. The forward problem depends on a finite-dimensional set of parameters $\boldsymbol{\mu} \in \mathcal{P} \subset \mathbb{R}^P$.

Its abstract formulation reads as follows: given $\boldsymbol{\mu} \in \mathcal{P}$, find $u(\boldsymbol{\mu}) \in X$ such that

$$\begin{cases} a(u(\boldsymbol{\mu}), v; \boldsymbol{\mu}) = f(v; \boldsymbol{\mu}) \quad \forall v \in X & \text{(state equation)} \\ \mathbf{s}(\boldsymbol{\mu}) = \boldsymbol{\ell}(u(\boldsymbol{\mu})) & \text{(observation equation).} \end{cases} \tag{1}$$

Here $a : X \times X \times \mathcal{P} \to \mathbb{R}$ and $f : X \times \mathcal{P} \to \mathbb{R}$ are a parametrized bilinear (resp. linear) form, $\mathbf{s} : \mathcal{P} \to \mathbb{R}^s$ and $\boldsymbol{\ell} = (\ell^1, \ldots, \ell^s)$, being $\ell^j : X \to \mathbb{R}$ a linear form for any $j = 1, \ldots, s$. This assumption is not as restrictive as one would think, since for example field-valued inverse problems can be treated by discretizing the underlying field with sufficient accuracy ($P \gg 1$) or by employing a truncated Karhunen-Loève expansion. In particular, we assume that the parameter vector is divided into two parts: $\boldsymbol{\mu} = (\boldsymbol{\gamma}, \boldsymbol{\zeta}) \in \mathbb{R}^{P_\gamma + P_\zeta}$; we denote by $\boldsymbol{\gamma}$ the *identifiable parameters*, and by $\boldsymbol{\zeta}$ the *nuisance parameters*. For the cases at hand, both identifiable and nuisance parameters will be related to physical properties of the system, thus entering the differential operator and possibly boundary conditions and source terms. See e.g. [21] for further details about the case of geometrical parameters.

We require that the forward problem is well-posed for any choice of the parameter vector $\boldsymbol{\mu} \in \mathcal{P}$. To this aim, we assume that $a(\cdot, \cdot; \boldsymbol{\mu})$ is continuous and uniformly coercive over X, for any $\boldsymbol{\mu} \in \mathcal{P}$, and that $f(\cdot; \boldsymbol{\mu})$ is continuous, that is, $f(\cdot; \boldsymbol{\mu}) \in X'$ for any $\boldsymbol{\mu} \in \mathcal{P}$, being X' the dual space of X. We also require that $\ell^j \in X'$ for any $j = 1, \ldots, s$. The Lax-Milgram lemma ensures uniqueness of the solution (and its continuous dependence from data) for any $\boldsymbol{\mu} \in \mathcal{P}$. This framework can also be adapted to stable problems in the sense of an *inf-sup* condition, by using the Babuška-Nečas theorem; see e.g. [36] for further details.

A numerical approximation of the forward problem (1) can be obtained by introducing, e.g., a Galerkin-finite element (FE) method relying on a finite-dimensional space $X_h \subset X$ of (possibly very large) dimension $dim(X_h) = N_h$; here h denotes a FE mesh parameter. Hence, the full-order model (FOM) related to the forward problem reads as follows: given $\boldsymbol{\mu} \in \mathcal{P}$, find $u_h(\boldsymbol{\mu}) \in X_h$ such that

$$\begin{cases} a(u_h(\boldsymbol{\mu}), v_h; \boldsymbol{\mu}) = f(v_h; \boldsymbol{\mu}) \quad \forall v_h \in X_h & \text{(FOM state),} \\ \mathbf{s}_h(\boldsymbol{\mu}) = \boldsymbol{\ell}(u_h(\boldsymbol{\mu})) & \text{(FOM observation).} \end{cases} \tag{2}$$

Under the above assumptions, problem (2) is well-posed and admits a unique solution for any $\boldsymbol{\mu} \in \mathcal{P}$. In particular, the following continuous dependence on data estimate holds:

$$\|u_h(\boldsymbol{\mu})\|_X \leq \frac{\|f(\cdot; \boldsymbol{\mu})\|_{X'}}{\alpha_h(\boldsymbol{\mu})} \qquad \forall \boldsymbol{\mu} \in \mathcal{P}$$

being $\alpha_h(\boldsymbol{\mu})$ the (discrete) stability factor related to the PDE operator, that is

$$\alpha_h(\boldsymbol{\mu}) = \inf_{v_h \in X_h} \frac{a(v_h, v_h; \boldsymbol{\mu})}{\|v_h\|_X^2} \geq \alpha_0 > 0 \qquad \forall \boldsymbol{\mu} \in \mathcal{P},$$

for a suitable $\alpha_0 \in \mathbb{R}$. Being able to efficiently evaluate a tight lower bound $0 < \alpha_h^{LB}(\boldsymbol{\mu}) \leq \alpha_h(\boldsymbol{\mu})$ of the stability factor for any $\boldsymbol{\mu} \in \mathcal{P}$ plays a key role in the a posteriori error bounds related to the RB approximation and, finally, in the solution of the inverse problem.

If the forward problem consists of solving (2) to predict the outcome of an experiment – by computing $u(\boldsymbol{\mu})$ and evaluating the output $\mathbf{s}(\boldsymbol{\mu})$ – in an inverse problem observed data or measurements $\mathbf{s}^*$ are used to estimate unknown parameters $\boldsymbol{\mu}$ characterizing the physical system. Such a problem can be cast in the optimal control framework, see e.g. [21] for further details. If $\mathbf{s}^*$ is an experimental measure, possibly polluted by measurement error, we need to rely instead on a statistical framework in order to quantify uncertainties (due to both measurement errors and nuisance parameters) on the estimated parameter values.

2.2 Bayesian framework

We consider a Bayesian framework [19, 44, 45] for the solution of the inverse UQ problems. We model both the observations $\mathbf{s}^*$ and the parameters $\boldsymbol{\mu}$ as random variables, by introducing suitable probability density functions (PDFs). The solution of the inverse problem is given by a point or

interval estimation computed on the basis of the *posterior probability* density $\pi_{post} : \mathcal{P} \times Y \to \mathbb{R}_0^+$, i.e. the probability density of the parameter $\boldsymbol{\mu}$ given the measured value of $\mathbf{s}^*$, which can be obtained as

$$\pi_{post}(\boldsymbol{\mu} \mid \mathbf{s}^*) = \frac{\pi(\mathbf{s}^* \mid \boldsymbol{\mu})\, \pi_{prior}(\boldsymbol{\mu})}{\eta(\mathbf{s}^*)} \tag{3}$$

thanks to *Bayes' theorem*. Here $\pi_{prior} : \mathcal{P} \to \mathbb{R}_0^+$ is the *prior probability* density, expressing all available information on $\boldsymbol{\mu}$ independently of the measurements on $\mathbf{s}^*$ that will be considered as data; $\pi : Y \times \mathcal{P} \to \mathbb{R}_0^+$ is the *likelihood function* of $\mathbf{s}^*$ conditionally to $\boldsymbol{\mu}$; finally

$$\eta(\mathbf{s}^*) = \int_{\mathcal{P}} \pi(\mathbf{s}^* \mid \boldsymbol{\mu})\, \pi_{prior}(\boldsymbol{\mu}).$$

In order to describe measurement errors, we consider an *additive noise* model, that is, if we suppose that $\boldsymbol{\mu}$ is the true parameter, the outcome of the experiment is

$$\mathbf{s}^* = \mathbf{s}_h(\boldsymbol{\mu}) + \boldsymbol{\varepsilon}_{\text{noise}} = \boldsymbol{\ell}(u_h(\boldsymbol{\mu})) + \boldsymbol{\varepsilon}_{\text{noise}}, \tag{4}$$

where the measurement noise $\varepsilon_{\text{noise}}$ follows a probability distribution π_{noise}. In this way, our data are d noisy s-variate measures $\{\mathbf{s}_1^*, \dots, \mathbf{s}_d^*\}$, $\mathbf{s}_i^* \in \mathbb{R}^s$ for any $i = 1, \dots, d$, modelled by assuming that the outcome of the numerical model is given by the output evaluated for the *true* parameter value. The most typical description of experimental uncertainties is the Gaussian model, that is, we deal with normally distributed, uncorrelated errors $\varepsilon_{\text{noise}} \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 \delta_{ij})$, $i, j = 1, \dots, s$, with known variances σ_i^2, independent of $\boldsymbol{\mu}$. We also denote the likelihood function appearing in (3) by highlighting the dependence on the FE space, as

$$\pi(\mathbf{s}^* | \boldsymbol{\mu}) = \pi^h(\mathbf{s}^* | \boldsymbol{\mu}) = \pi_\varepsilon(\mathbf{s}^* - \mathbf{s}_h(\boldsymbol{\mu})) \tag{5}$$

so that the expression of the posterior PDF given by (3) is as follows:

$$\pi_{post}^h(\boldsymbol{\mu} \mid \mathbf{s}^*) = \frac{\pi^h(\mathbf{s}^* | \boldsymbol{\mu})\, \pi_{prior}(\boldsymbol{\mu})}{\eta_h(\mathbf{s}^*)}, \quad \text{being} \quad \eta_h(\mathbf{s}^*) = \int_{\mathcal{P}} \pi^h(\mathbf{s}^* | \boldsymbol{\mu})\, \pi_{prior}(\boldsymbol{\mu}). \tag{6}$$

We highlight the dependence of the posterior PDF on the FE space, too. If the output depends linearly on the parameters and we choose a Gaussian prior, the posterior is also Gaussian. Instead, as soon as $\boldsymbol{\mu} \mapsto \mathbf{s}(\boldsymbol{\mu})$ is a nonlinear map, the expression of the likelihood function yields a posterior distribution which cannot be written in closed form, requiring instead an exhaustive exploration of the parameter space. This becomes very hard to perform, above all if the parameter space has a large dimension. We then need to rely on MCMC to sample the posterior PDF, such as the well-known Metropolis-Hastings or Gibbs sampling techniques [16, 24]. These methods are exploited to draw a sequence of random samples from a (multi-dimensional) PDF which cannot be expressed in closed form. This is meant in order not only to approximate the posterior PDF, but also to compute integrals related to this distribution. Then, since we are not interested in the nuisance parameters $\boldsymbol{\zeta}$, we proceed to marginalize them. This leads to computing the conditional marginal distribution

$$\pi_{post}^h(\boldsymbol{\gamma} \mid \mathbf{s}^*) = \frac{1}{\eta_h(\mathbf{s}^*)} \int_{\mathcal{P}_\zeta} \pi^h(\mathbf{s}^* | \boldsymbol{\mu})\ \pi_{prior}(\boldsymbol{\gamma}, \boldsymbol{\zeta})\, d\boldsymbol{\zeta}. \tag{7}$$

MCMC methods are needed to evaluate (possibly) high-dimensional integrals like the one in (7). These methods involve repeated evaluations of the likelihood function $\pi^h(\mathbf{s}^* \mid \boldsymbol{\gamma}, \boldsymbol{\zeta})$ – and thus repeated evaluations of the forward problem (2) – so that relying on the FOM would be too expensive also in the case of linear elliptic problems.

Therefore, we seek to replace (2) with a computationally less expensive, reduced-order model providing an inexpensive approximation $u_n(\boldsymbol{\mu})$ to $u_h(\boldsymbol{\mu})$. This allows to compute a reduced-order (and inexpensive) approximation $\mathbf{s}_n(\boldsymbol{\mu})$ to the full-order output $\mathbf{s}_h(\boldsymbol{\mu})$. Replacing the full-order likelihood function π^h with its ROM approximation

$$\pi^n(\mathbf{s}^* | \boldsymbol{\mu}) = \pi_\varepsilon(\mathbf{s}^* - \mathbf{s}_n(\boldsymbol{\mu})) \tag{8}$$

clearly affects the posterior distribution, which changes as follows:

$$\pi^n_{post}(\boldsymbol{\mu} \mid \mathbf{s}^*) = \frac{\pi^n(\mathbf{s}^*|\boldsymbol{\mu})\, \pi_{prior}(\boldsymbol{\mu})}{\eta_n(\mathbf{s}^*)}, \quad \text{being} \quad \eta_n(\mathbf{s}^*) = \int_{\mathcal{P}} \pi^n(\mathbf{s}^*|\boldsymbol{\mu})\, \pi_{prior}(\boldsymbol{\mu}). \tag{9}$$

Consequently, the marginal PDF of the identifiable parameters becomes

$$\pi^n_{post}(\boldsymbol{\gamma} \mid \mathbf{s}^*) = \frac{1}{\eta_n(\mathbf{s}^*)} \int_{\mathcal{P}_\zeta} \pi^n(\mathbf{s}^*|\boldsymbol{\mu})\ \pi_{prior}(\boldsymbol{\gamma}, \boldsymbol{\zeta})\, d\boldsymbol{\zeta}. \tag{10}$$

Nevertheless, being able to quantify the ROM effect on the inversion procedure by relying on a suitable measure of the error $\mathbf{s}_h(\boldsymbol{\mu}) - \mathbf{s}_n(\boldsymbol{\mu})$ is paramount: the main goal of this paper is to set up a suitable procedure to answer this question.

3 Reduced order models and a posteriori error bounds

Solving large PDE systems for several parameter values may require huge computational resources, unless efficient and reliable ROMs for parametrized PDEs are used. In this paper we rely on a certified *reduced basis method*, whose basic ingredients are recalled in this section. A general introduction to this method can be found e.g., in [37]; see also [41].

3.1 Reduced subspaces and projection-based ROMs

The RB method is a projection-based ROM which allows to compute an approximation $u_n(\boldsymbol{\mu})$ of the solution $u_h(\boldsymbol{\mu})$ (as well as an approximation $\mathbf{s}_n(\boldsymbol{\mu})$ of the output $\mathbf{s}_h(\boldsymbol{\mu})$) through a Galerkin projection onto a reduced subspace X_n; given $\boldsymbol{\mu} \in \mathcal{P}$, find $u_n(\boldsymbol{\mu}) \in X_n$ s.t.

$$\begin{cases} a(u_n(\boldsymbol{\mu}), v_n; \boldsymbol{\mu}) = f(v_n; \boldsymbol{\mu}) \quad \forall v_n \in X_n & \text{(ROM state)} \\ \mathbf{s_n}(\boldsymbol{\mu}) = \boldsymbol{\ell}(u_n(\boldsymbol{\mu})) & \text{(ROM observation)}, \end{cases} \tag{11}$$

where $\dim(X_n) = n \ll N_h$. The reduced subspace X_n is constructed from a set of (well-chosen) full-order solutions, usually by exploiting one of the following techniques [22, 38]:

- *Greedy algorithm* [35, 46]. Basis functions are obtained by orthonormalizing a set of full-order solutions, corresponding to a specific choice $S_n = \{\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^n\}$ of parameter values, built by means of the following *greedy* procedure. Let us denote $\Xi_{train} \subset \mathcal{P}$ a (sufficiently rich) finite training sample, selected from $\mathcal{P}$ according to a uniform distribution. Given a prescribed $\boldsymbol{\mu}^1 \in \Xi_{train}$ and a sharp, inexpensive error bound $\Delta_n(\boldsymbol{\mu})$ (see Sect. 3.3) such that

  $$\|u_h(\boldsymbol{\mu}) - u_n(\boldsymbol{\mu})\|_X \leq \Delta_n(\boldsymbol{\mu}) \quad \text{for all } \boldsymbol{\mu} \in \mathcal{P},$$

  we choose the remaining parameter values as

  $$\boldsymbol{\mu}^n := \arg \max_{\boldsymbol{\mu} \in \Xi_{train}} \Delta_{n-1}(\boldsymbol{\mu}), \quad \text{for } n = 2, \dots, n_{max}$$

  until $\Delta_{n_{max}}(\boldsymbol{\mu}) \leq \varepsilon^{RB}_{tol}$ for all $\boldsymbol{\mu} \in \Xi_{train}$, being ε^{RB}_{tol} a suitably small tolerance.

- *Proper orthogonal decomposition* [4, 43]. In this case, the reduced subspace X_n is given by the first n (left) singular vectors of the snapshot matrix $S = [u_h(\boldsymbol{\mu}^1) \mid \dots \mid u_h(\boldsymbol{\mu}^{N_s})] \in \mathbb{R}^{N_h \times N_s}$, corresponding to the largest n singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$. Here $u_h(\boldsymbol{\mu}^1), \dots, u_h(\boldsymbol{\mu}^{N_s})$ are N_s full-order solutions of the forward problem, computed for a random sample $\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^{N_s}$. By construction, the POD basis is orthonormal; moreover, the error in the basis is equal to the squares of the singular values corresponding to the neglected modes, and the maximum subspace dimension is such that $\sum_{i=n+1}^{r} \sigma_i^2 \leq \varepsilon^{POD}_{tol}$, $r = \min\{N_s, N_h\}$, being ε^{POD}_{tol} a suitably small tolerance.

3.2 Affine parametric dependence and Offline/Online decomposition

Constructing the reduced subspace requires several evaluation of the FOM, which are performed only once, during the so-called *Offline* stage. Each *Online* evaluation of the reduced solution (and

related output) requires to solve a problem of very small dimension $n \ll N_h$. Such an Offline/Online decomposition is made possible under the assumption that a suitable *affine parametric dependence* property is fulfilled by the $\boldsymbol{\mu}$-dependent operators. Hence, we require that $a(\cdot,\cdot;\boldsymbol{\mu})$, $f(\cdot;\boldsymbol{\mu})$ can be written as a separable expansion of $\boldsymbol{\mu}$-independent bilinear/linear forms:

$$a(u,v;\boldsymbol{\mu}) = \sum_{q=1}^{Q_a} \Theta_q^a(\boldsymbol{\mu}) a_q(u,v), \qquad f(v;\boldsymbol{\mu}) = \sum_{q=1}^{Q_f} \Theta_q^f(\boldsymbol{\mu}) f_q(v)$$

for some integers Q_a, Q_f. A similar decomposition would be required also on the linear forms ℓ^j, $j = 1,\dots,s$, if they were $\boldsymbol{\mu}$-dependent, too.

3.3 A posteriori error bounds

We can easily derive an *a posteriori* (residual-based) error bound with respect to the full-order solution, for both the PDE solution and linear outputs [37, 38]. Let us denote by $r(w;\boldsymbol{\mu}) = f(w;\boldsymbol{\mu}) - a(u_n(\boldsymbol{\mu}),v;\boldsymbol{\mu})$, for any $w \in X_h$, the residual of the state equation (evaluated on the RB solution $u_n(\boldsymbol{\mu})$) and its dual norm by $\|r(\cdot;\boldsymbol{\mu})\|_{X'} = \sup_{v\in X_h} r(v;\boldsymbol{\mu})/\|v\|_X$. Then, the error bound on the solution reads as follows:

$$\|u_h(\boldsymbol{\mu}) - u_n(\boldsymbol{\mu})\|_X \le \Delta_n(\boldsymbol{\mu}) := \frac{\|r(\cdot;\boldsymbol{\mu})\|_{X'}}{\alpha_h^{LB}(\boldsymbol{\mu})} \qquad \forall \boldsymbol{\mu} \in \mathcal{P}. \tag{12}$$

We remark that also the computation of the dual norm of residuals, as well as of the lower bound α_h^{LB} to the stability factors, takes advantage of a similar *Offline-Online* stratagem, allowing to get an inexpensive evaluation of the error bound for each $\boldsymbol{\mu} \in \mathcal{P}$; see e.g. [38].

Regarding the error bound on the output, which is relevant to the Bayesian inversion, we recall here its expression in the case of linear, noncompliant outputs, as the numerical test cases presented in the remainder deal with this situation. For any $\ell^j \in X'$, let us introduce the following (full-order approximation of the) *dual* problem: find $\psi_h^j(\boldsymbol{\mu}) \in X_h$ such that

$$a(v_h, \psi_h^j(\boldsymbol{\mu});\boldsymbol{\mu}) = -\ell^j(v_h) \quad \forall\, v_h \in X_h.$$

In addition to the reduced space for the forward problem (2) (which can be referred to as the *primal* problem), let us define a reduced subspace for each dual problem, by using the same algorithm (either greedy-RB or POD) chosen for the primal problem. Here X_n^j denotes the dual subspace related to the j-th output, although the dimension of each of these subspaces can be different, and differ from n. The resulting RB approximation $\psi_n^j(\boldsymbol{\mu}) \in X_n^j$ solves

$$a(v_n, \psi_n^j(\boldsymbol{\mu});\boldsymbol{\mu}) = -\ell^j(v_n) \quad \forall v_n \in X_n^j$$

and is required to get the following error bound on the output:

$$|s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu})| \le \Delta_n^j(\boldsymbol{\mu}) \equiv \frac{\|r(\,\cdot\,;\boldsymbol{\mu})\|_{(X_h)'}}{(\alpha_h^{\mathrm{LB}}(\boldsymbol{\mu}))^{1/2}} \, \frac{\|r^j(\,\cdot\,;\boldsymbol{\mu})\|_{(X_h)'}}{(\alpha_h^{\mathrm{LB}}(\boldsymbol{\mu}))^{1/2}} \qquad \forall \boldsymbol{\mu} \in \mathcal{P} \tag{13}$$

where $r^j(w;\boldsymbol{\mu}) = -\ell^j(w) - a(w, \psi_n^j(\boldsymbol{\mu});\boldsymbol{\mu})$, $\forall w \in X_h$, is the dual residual related to the j-th output. Here $s_h^j(\boldsymbol{\mu}) = \ell^j(u_h(\boldsymbol{\mu}))$ denotes the full-order j-th output, whereas $s_n^j(\boldsymbol{\mu}) = \ell^j(u_n(\boldsymbol{\mu}))$ is the corresponding ROM output. In the same way, for the RB dual solution we have:

$$\|\psi_h^j(\boldsymbol{\mu}) - \psi_n^j(\boldsymbol{\mu})\|_X \le \Delta_n^{\psi^j}(\boldsymbol{\mu}) := \frac{\|r^j(\cdot;\boldsymbol{\mu})\|_{X'}}{\alpha_h^{LB}(\boldsymbol{\mu})} \qquad \forall \boldsymbol{\mu} \in \mathcal{P}.$$

As already remarked for the primal residuals, also dual residuals can be efficiently evaluated by taking advantage of the affine $\boldsymbol{\mu}$-dependence. We also denote by

$$\Phi_n^j(\boldsymbol{\mu}) = \frac{\Delta_n^j(\boldsymbol{\mu})}{\mathbf{s}_h^j(\boldsymbol{\mu}) - \mathbf{s}_n^j(\boldsymbol{\mu})} \qquad \forall \boldsymbol{\mu} \in \mathcal{P},\ 1 \le j \le s$$

the effectivity of the estimator $\Delta_n^j(\boldsymbol{\mu})$. It is possible to show that $1 \le |\Phi_n^j(\boldsymbol{\mu})| \le M_h(\boldsymbol{\mu})/\alpha_h^{LB}(\boldsymbol{\mu})$, where $M_h(\boldsymbol{\mu})$ is the FE continuity constant of $a(\cdot,\cdot;\boldsymbol{\mu})$, see e.g. [37] for further details.

4 Reduction Error Models

Being able to evaluate the output quantity of a PDE system at a greatly reduced cost is essential to speed up the solution of inverse UQ problems within a Bayesian framework. Our goal, once a RB approximation has been built in the offline stage, is to exploit its fast and inexpensive online queries to speed up the evaluation of the posterior PDF, of related (point or interval) estimates, and of MCMC integrals like (7) or (10). Not only, by taking into account reduction errors or available error bounds, we can obtain reliable solutions at the end of the inversion process, too. Although ROMs have been exploited to speed up the solution of inverse problems in several works, very few papers have focused on the analysis of reduction error propagation, see, e.g., [23, 8].

In particular, we wish to incorporate a model for the error engendered by the ROM into the Bayesian estimator. To this end, we provide suitable (both deterministic and statistical) reduction error models (REMs), possibly by exploiting available error bounds on the outcome of the forward problem. A basic observation is made possible by expressing, when dealing with *linear* (with respect to the PDE solution) outputs, the measurement equation (4) as

$$\mathbf{s}^* = \boldsymbol{\ell}(u_n(\boldsymbol{\mu})) + [\boldsymbol{\ell}(u_h(\boldsymbol{\mu})) - \boldsymbol{\ell}(u_n(\boldsymbol{\mu}))] + \boldsymbol{\varepsilon}_{\text{noise}} = \boldsymbol{\ell}(u_n(\boldsymbol{\mu})) + \boldsymbol{\delta}(\boldsymbol{\mu}) \tag{14}$$

where $\boldsymbol{\delta}(\boldsymbol{\mu}) = \boldsymbol{\varepsilon}_{\text{ROM}}(\boldsymbol{\mu}) + \boldsymbol{\varepsilon}_{\text{noise}}$ and

$$\boldsymbol{\varepsilon}_{\text{ROM}}(\boldsymbol{\mu}) = [s_h^1(\boldsymbol{\mu}) - s_n^1(\boldsymbol{\mu}) \ , \ \dots \ , \ s_h^s(\boldsymbol{\mu}) - s_n^s(\boldsymbol{\mu})]^T \tag{15}$$

is the reduction error, that is, the error due to ROM approximation of the forward problem and related output. Although in principle $\boldsymbol{\varepsilon}_{\text{ROM}}$ is deterministic, in practice its evaluation is out of reach since it would require, for any $\boldsymbol{\mu}$, the solution of the full-order model. Here we propose three approaches for approximating the reduction error $\boldsymbol{\varepsilon}_{\text{ROM}}(\boldsymbol{\mu})$ by a suitable indicator $\tilde{\boldsymbol{\varepsilon}}_{\#}(\boldsymbol{\mu})$, $\# = 1, 2, 3$, according to suitable reduction error models which can be easily incorporated in the Bayesian framework. In particular:

- in **[REM-1]** we treat ROM errors as epistemic uncertainties, represented through random variables, following the so-called *approximation error model* [1];
- in **[REM-2]** we provide a deterministic approximation of ROM errors by considering *radial basis* interpolants of the error over the parameter space;
- in **[REM-3]** we express ROM errors through a linear model depending on output error bounds, fitted through regression analysis.

Hence, we end up with error indicators which can be either deterministic – that is, $\tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\mu}) = \mathbf{m}_{\text{ROM}}(\boldsymbol{\mu})$, being $\mathbf{m}_{\text{ROM}}(\boldsymbol{\mu})$ a suitable function of $\boldsymbol{\mu}$ – or expressed through a random variable $\tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\mu})$, whose distribution $\pi_{\tilde{\boldsymbol{\varepsilon}}}$ is characterized by $\mathbf{m}_{\text{ROM}}(\boldsymbol{\mu}) = \mathbb{E}[\tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\mu})]$ and $\Sigma_{\text{ROM}}(\boldsymbol{\mu}) = \text{Cov}[\tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\mu})]$. Correspondingly, we end up with a *corrected* reduced-order likelihood

$$\tilde{\pi}^n(\mathbf{s}^*|\boldsymbol{\mu}) = \begin{cases} \pi_{\boldsymbol{\varepsilon}}(\mathbf{s}^* - \mathbf{s}_n(\boldsymbol{\mu}) - \mathbf{m}_{\text{ROM}}(\boldsymbol{\mu})) & \text{deterministic REM} \\ \pi_{\tilde{\boldsymbol{\delta}}}(\mathbf{s}^* - \mathbf{s}_n(\boldsymbol{\mu}) - \mathbf{m}_{\text{ROM}}(\boldsymbol{\mu})) & \text{statistical REM} \end{cases} \tag{16}$$

being $\tilde{\boldsymbol{\delta}}(\boldsymbol{\mu}) = \tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\mu}) + \boldsymbol{\varepsilon}_{\text{noise}}$ and $\text{Cov}[\tilde{\boldsymbol{\delta}}(\boldsymbol{\mu})] = \Sigma_{\text{noise}}(\boldsymbol{\mu}) + \Sigma_{\text{ROM}}(\boldsymbol{\mu})$, by assuming that ROM errors and measurement noise are independent. Correspondingly, we obtain the following *corrected* reduced-order posterior PDF

$$\tilde{\pi}^n_{post}(\boldsymbol{\mu} \mid \mathbf{s}^*) = \frac{\tilde{\pi}^n(\mathbf{s}^*|\boldsymbol{\mu})\, \pi_{prior}(\boldsymbol{\mu})}{\tilde{\eta}_n(\mathbf{s}^*)}, \quad \text{being} \quad \tilde{\eta}_n(\mathbf{s}^*) = \int_{\mathcal{P}} \tilde{\pi}^n(\mathbf{s}^*|\boldsymbol{\mu})\, \pi_{prior}(\boldsymbol{\mu}), \tag{17}$$

yielding to a similar correction in the marginal PDF of the identifiable parameters (10). In the following subsections we discuss the construction of these three REMs.

4.1 REM-1: approximation error model

Following the so-called *approximation error model* of Kaipio et al. [1], we assume that ROM errors are the outcome of a Gaussian random variable, so that $\boldsymbol{\varepsilon}_{\text{ROM}}(\boldsymbol{\mu})$ is replaced by

$$\tilde{\boldsymbol{\varepsilon}}_1(\boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{m}_{\mathrm{ROM}}, \Sigma_{\mathrm{ROM}}) \tag{18}$$

being $\mathbf{m}_{\mathrm{ROM}} \in \mathbb{R}^s$ and $\Sigma_{\mathrm{ROM}} \in \mathbb{R}^{s\times s}$ the sample constant mean and covariance, respectively, obtained by sampling N_{cal} errors $\{\mathbf{s}_h(\boldsymbol{\mu}^k) - \mathbf{s}_n(\boldsymbol{\mu}^k)\}_{k=1}^{N_{cal}}$:

$$\mathbf{m}_{\mathrm{ROM}} = \frac{1}{N_{cal}} \sum_{k=1}^{N_{cal}} \boldsymbol{\varepsilon}_{\mathrm{ROM}}(\boldsymbol{\mu}^k), \quad \Sigma_{\mathrm{ROM}} = \frac{1}{N_{cal}-1} \sum_{k=1}^{N_{cal}} (\boldsymbol{\varepsilon}_{\mathrm{ROM}}(\boldsymbol{\mu}^k) - \mathbf{m}_{\mathrm{ROM}})(\boldsymbol{\varepsilon}_{\mathrm{ROM}}(\boldsymbol{\mu}^k) - \mathbf{m}_{\mathrm{ROM}})^T.$$

We refer to the random sample $S_{N_{cal}} = \{\boldsymbol{\mu}^1, \ldots, \boldsymbol{\mu}^{N_{cal}}\}$ as the *calibration set*, since the computational experiments leading to $\boldsymbol{\varepsilon}_{\mathrm{ROM}}(\tilde{\boldsymbol{\mu}})$, $\tilde{\boldsymbol{\mu}} \in S_{N_{cal}}$, are additional queries to both the FOM and the ROM, required to characterize our REM-1. In this case, the correction does not depend on $\boldsymbol{\mu}$ by construction. If we assume in addition that measurement errors are Gaussian – that is, $\boldsymbol{\varepsilon}_{\mathrm{noise}} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathrm{noise}})$ – and independent from ROM errors, we have that

$$\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\mu}) := \boldsymbol{\varepsilon}_{\mathrm{noise}} + \tilde{\boldsymbol{\varepsilon}}_1(\boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{m}_{\mathrm{ROM}}, \Sigma_{\mathrm{noise}} + \Sigma_{\mathrm{ROM}}). \tag{19}$$

Indeed, the effect of the ROM error results in a shift of the likelihood function and an additional contribution to its variance, provided the normality assumption on the errors evaluated over $S_{N_{cal}}$ is fulfilled. In this case, only a slight modification of the numerical solver is required within the MCMC process, thus making this REM particularly easy to implement. Nevertheless, very often ROM errors do not show a Gaussian distribution, so that further operations are required in order to use such a REM.

To overcome this fact, we can generalize the previous framework by considering any (parametric) distribution for the ROM errors conveniently fitted on the set of calibration experiments, that is, $\tilde{\boldsymbol{\varepsilon}}_1(\boldsymbol{\mu}) \sim \pi_{\tilde{\boldsymbol{\varepsilon}}}(\boldsymbol{\mu})$ possibly depending on a set of shape parameters. Although we assume that ROM errors are independent from (Gaussian) measurement errors, the distribution of $\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\mu}) = \boldsymbol{\varepsilon}_{\mathrm{noise}} + \tilde{\boldsymbol{\varepsilon}}_1(\boldsymbol{\mu})$ cannot be found, in general, in a closed form. In these cases, at each MCMC step, we quantify $\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\mu})$ by calculating the realization of the sum of two random variables by the following convolution:

$$\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\mu}) = \int_{\mathcal{P}} \pi_{\tilde{\boldsymbol{\varepsilon}}}(\boldsymbol{\mu} + \boldsymbol{\nu}) \pi_{\boldsymbol{\varepsilon}}(\boldsymbol{\nu}) d\boldsymbol{\nu}. \tag{20}$$

The evaluation of this integral can be performed, e.g., by an internal MCMC algorithm, which does not feature expensive extra calculations at each step of the outer MCMC process.

Thus, in the case of REM-1, a global approximation of the ROM error over the parameter space is provided – hence, not depending on $\boldsymbol{\mu}$ – by prescribing the distribution of a random variable fitted over a sample of *calibration experiments*.

4.2 REM-2: radial basis interpolation

Despite being straightforward, REM-1 can perform badly, for instance when ROM errors can not be explained by means of a sample statistical distributions in closed form, because of their complex variability over the parameter space. For this reason, we turn instead to a local error model, still exploiting the errors $\{\mathbf{s}_h(\boldsymbol{\mu}^k) - \mathbf{s}_n(\boldsymbol{\mu}^k)\}_{k=1}^{N_{cal}}$ computed over a calibration set $S_{N_{cal}}$, by considering a radial basis interpolant for each output component. In particular, a deterministic $\boldsymbol{\mu}$-dependent correction can be obtained as

$$s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu}) \simeq \Pi^j(\boldsymbol{\mu}) \Delta_n^j(\boldsymbol{\mu}), \qquad j = 1, \ldots, s,$$

where $\Pi^j(\boldsymbol{\mu})$ is a weighted combination of radial basis functions (RBF), i.e.

$$\Pi^j(\mu) = \sum_{k=1}^{N_{cal}} w_k \phi(\|\boldsymbol{\mu} - \boldsymbol{\mu}^k\|), \qquad j = 1, \ldots s,$$

and the coefficients $\{w_k\}_{k=1}^{N_{cal}}$ are determined so that Π^j fulfills the following interpolation constraints over the calibration set:

$$\frac{s_h^j(\boldsymbol{\mu}^k) - s_n^j(\boldsymbol{\mu}^k)}{\Delta_n^j(\boldsymbol{\mu}^k)} = \Pi^j(\boldsymbol{\mu}^k), \qquad k = 1, \dots, N_{cal},\ j = 1, \dots, s.$$

In other words, we compute an interpolant of the inverse effectivities $(\Phi_n^j(\boldsymbol{\mu}))^{-1} \in [-1, 1]$, for each $j = 1, \dots, s$. This choice yields more accurate results than those obtained by interpolating ROM errors over the calibration set, that is, by considering $s_h^j(\boldsymbol{\mu}^k) - s_n^j(\boldsymbol{\mu}^k) = \tilde{\Pi}^j(\boldsymbol{\mu}^k)$, $k = 1, \dots, N_{cal},\ j = 1, \dots, s$ (see Sect. 7.1). While $\phi : \mathbb{R}_0^+ \to \mathbb{R}$ is a fixed shape function, radial with respect to the Euclidean distance $\|\cdot\|$ over $\mathbb{R}^P$. Here we use multiquadric basis functions $(\phi(r) = \sqrt{1 + r^2})$, see e.g. [9] for other available options. Thus, in our deterministic REM-2 we replace ROM errors $\boldsymbol{\varepsilon}_{\mathrm{ROM}}(\boldsymbol{\mu})$ by $\tilde{\boldsymbol{\varepsilon}}_2(\boldsymbol{\mu})$, with

$$\mathbf{m}_{\mathrm{ROM}}(\boldsymbol{\mu}) = [\Pi^1(\boldsymbol{\mu})\Delta_n^1(\boldsymbol{\mu})\ ,\ \dots\ ,\ \Pi^s(\boldsymbol{\mu})\Delta_n^s(\boldsymbol{\mu})]^T, \tag{21}$$

which features a $\boldsymbol{\mu}$-dependent shifting of the likelihood function. If in addition we want to take into account the variability of the ROM errors, we can incorporate the sample covariance evaluated over the calibration set as in REM-1, and thus consider

$$\tilde{\boldsymbol{\delta}}_2(\boldsymbol{\mu}) := \boldsymbol{\varepsilon}_{\mathrm{noise}} + \tilde{\boldsymbol{\varepsilon}}_2(\boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{m}_{\mathrm{ROM}}(\boldsymbol{\mu}), \Sigma_{\mathrm{noise}} + \Sigma_{\mathrm{ROM}}(\boldsymbol{\mu})).$$

Assuming that ROM errors are Gaussian might be very limiting, as highlighted in the previous section. Here we consider this argument just as an heuristic correction.

Although very simple to be characterized, REM-2 suffers from the usual *curse of dimensionality* of multivariate interpolation, which arises from the fact that the simple choices of interpolation node sets (such as tensor product girds) grow exponentially in size as the parametric dimension $p = \dim(\mathcal{P})$ is increased, so that both sampling the parameter space and evaluating calibration experiments rapidly becomes less and less affordable even for relatively small P (say $P > 5$). Sparse grid techniques provide partial but not totally satisfactory resolution to this problem [5]. Nevertheless, RBF interpolation is suitable also for scattered data and, for the cases at hand, shows a very good compromise between accuracy and simplicity.

4.3 REM-3: linear regression model

A possible way to overcome the curse of dimensionality is to rely on a model where the *surrogate* ROM error is computed as a function of a scalar quantity depending on $\boldsymbol{\mu}$, rather than on $\boldsymbol{\mu}$ itself, no matter which is the parametric dimension P of $\boldsymbol{\mu} \in \mathcal{P} \subset \mathbb{R}^P$. In fact, the quantity which provides a good, inexpensive and readily available representation of the ROM error is the a posteriori error bound (13).

In order to derive a REM depending on a posteriori error bounds, we remark that a linear dependence between (the absolute values of the) output errors and related error bounds is shown when considering a logarithmic transformation, as already pointed out in a recent contribution [13]. Thus, we can in principle consider the following model:

$$\ln|s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu})| = \beta_0^j + \beta_1^j \ln(\Delta_n^j(\boldsymbol{\mu})) + \delta_{reg}^j, \qquad j = 1, \dots, s \tag{22}$$

being $\delta_{reg}^j \sim \mathcal{N}(0, \sigma_{reg,j}^2)$, and fit it to the datasets $\{\Delta_n^j(\boldsymbol{\mu}^k), s_h^j(\boldsymbol{\mu}^k) - s_n^j(\boldsymbol{\mu}^k)\}_{k=1}^{N_{cal}}$ obtained by sampling errors and corresponding error bounds for each output, over a calibration set $S_{N_{cal}}$. By doing this, we get the estimates $\hat{\beta}_0^j$, $\hat{\beta}_1^j$ of the coefficients by exploiting standard linear regression theory, as well as the estimate of the variances $\hat{\sigma}_{reg,j}^2$ through the corresponding mean square errors. Thus, by fitting model (22) we obtain the following relation for the absolute value of the ROM error:

$$|s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu})| = \exp(\hat{\beta}_0^j + \hat{\beta}_1^j \ln(\Delta_n^j(\boldsymbol{\mu})) + \hat{\delta}_{reg}^j), \qquad j = 1, \dots, s$$

being $\hat{\delta}_{reg}^j \sim \mathcal{N}(0, \hat{\sigma}_{\delta,j}^2)$. If we consider the deterministic REM-3, the absolute value of ROM errors $|\boldsymbol{\varepsilon}_{\mathrm{ROM}}(\mu)|$ can be replaced by

$$\mathbf{m}_{\mathrm{ROM}}(\boldsymbol{\mu}) = \left[\exp\left(\hat{\beta}_0^1 + \hat{\beta}_1^1 \ln(\Delta_n^1(\boldsymbol{\mu}))\right)\ ,\ \dots\ ,\ \exp\left(\hat{\beta}_0^s + \hat{\beta}_1^s \ln(\Delta_n^s(\boldsymbol{\mu}))\right)\right]^T. \tag{23}$$

Instead, by taking into account the variability shown by ROM errors, $|\boldsymbol{\varepsilon}_{\mathrm{ROM}}(\boldsymbol{\mu})|$ would be replaced by the following log-normal random variables:

$$\tilde{\boldsymbol{\varepsilon}}_3(\boldsymbol{\mu}) \sim \left[log\mathcal{N}(\hat{\beta}_0^1 + \hat{\beta}_1^1 \ln(\Delta_n^1(\boldsymbol{\mu})), \hat{\sigma}^2_{\mathrm{reg},1}(\boldsymbol{\mu})) \ , \ \dots \ , \ log\mathcal{N}(\hat{\beta}_0^s + \hat{\beta}_1^s \ln(\Delta_n^s(\boldsymbol{\mu})), \hat{\sigma}^2_{\mathrm{reg,s}}(\boldsymbol{\mu})) \right]^T . \quad (24)$$

However, in both cases no indications about the *sign* of the error are provided by the error bound (13), so that a correction based on (22) would be necessarily too poor. Once we have fitted the linear models, we shall infer the error sign (on each output component $j = 1, \dots, s$) from the calibration set, in order to replace $\tilde{\varepsilon}^j_{ROM}(\boldsymbol{\mu})$ by $\rho_j \tilde{\varepsilon}^j_3(\boldsymbol{\mu})$, where $\tilde{\varepsilon}^j_3(\boldsymbol{\mu})$ is given by (23) or (24) and $\rho_j \in \{-1, +1\}$ is determined by adopting one of the following strategies:

1. *Nearest neighbor*, assigning to $\tilde{\varepsilon}^j_3(\boldsymbol{\mu})$ the sign of the sampled error in the calibration set closest to $\boldsymbol{\mu}$, that is

$$\rho_j = \mathrm{sgn}\left(s^j_h(\hat{\boldsymbol{\mu}}^k) - s^j_n(\hat{\boldsymbol{\mu}}^k) \right), \qquad \hat{\boldsymbol{\mu}}^k = \underset{\boldsymbol{\mu}^k \in S_{N_{cal}}}{\arg\min} \|\boldsymbol{\mu} - \boldsymbol{\mu}^k\|;$$

2. *Bernoulli trial*, assigning to $\tilde{\varepsilon}^j_3(\boldsymbol{\mu})$ the sign:

$$\rho_j = -1 + 2X_j, \qquad X_j \sim \mathcal{B}e(\hat{p}^j), \qquad \hat{p}^j = \frac{|S^j_+|}{N_{cal}} \in [0, 1],$$

being $S^j_+ = \{s^j_h(\boldsymbol{\mu}) - s^j_n(\boldsymbol{\mu}) > 0 : \boldsymbol{\mu} \in S_{N_{cal}}\}$ and $S^j_- = \{s^j_h(\boldsymbol{\mu}) - s^j_N(\mu) < 0 : \boldsymbol{\mu} \in S_{N_{cal}}\}$ i.e. weighting the Bernoulli trial according to the observed distribution of signs in the calibration sample. While heuristic in nature, this rule does correctly treat the cases where the reduction error is heavily biased towards one sign or another instead of simply assigning, say, a positive sign to all errors.

The splitting of the calibration set in two subsets S^j_+ and S^j_- suggests to consider a regression model over each subset, depending on the computed errors sign. In conclusion, we obtain for each $\boldsymbol{\mu} \in \mathcal{P}$ the random variable:

$$\tilde{\boldsymbol{\delta}}_3(\boldsymbol{\mu}) := \boldsymbol{\varepsilon}_{\mathrm{noise}} + \boldsymbol{\rho}\tilde{\boldsymbol{\varepsilon}}_3(\boldsymbol{\mu}) \quad (25)$$

which is the sum of two random variables with different distributions (and can be computed similarly to (20)).

5 Inversion procedure

Let us now summarize the whole numerical procedure we use to solve a parametric inverse UQ problem. A first *offline* stage (Algorithm 1) consists in the computation of the reduced space and in the additional calibration set evaluation. More advanced versions of the greedy algorithm where the two greedy loops are interlaced could also be used, but since this step is independent of the REM-step that follows, here we have presented for simplicity a basic version. During the *online* stage (Algorithm 2), a MCMC procedure is performed by considering at each iteration the reduced output $\mathbf{s}_n(\boldsymbol{\mu})$ instead of the full-order output $\mathbf{s}_h(\boldsymbol{\mu})$.

The selection of the basis functions is performed through a *greedy (RB)* algorithm. This procedure requires the efficient evaluation of the a posteriori error bound (12), for which a suitable offline/online procedure is exploited (see [27] for further details). The calibration procedure consists in evaluating the difference between the ROM output $\mathbf{s}_n(\boldsymbol{\mu})$ and the full-order output $\mathbf{s}_h(\boldsymbol{\mu})$ for each parameter in the calibration set $S_{N_{cal}}$. Finally, the REM construction is performed accordingly to the procedure described in Sect. 4.

During the *online* stage, the *posterior* distribution is sampled through a Metropolis–Hastings algorithm, which generates a sequence of sample values, whose distribution converges to the desired corrected distribution $\tilde{\pi}^n_{post}$. Each MCMC iteration entails an *online* query, which is performed in an efficient way by the ROM and the REM. The quality of the sampling sequence is finally improved by performing a subsequent *burn-in* and *thinning*, in order to reduce the autocorrelation between the sampled points; see e.g. [16, 24] for further details.

Algorithm 1 Offline procedure

procedure BASIS COMPUTATION
FE matrices:
  $A_q^h,\ F_q^h \leftarrow$ *state problem*
  $L_{q,s}^h \leftarrow$ *ouput evaluation & dual problems*
Lower bound:
  $\alpha_{LB}(\boldsymbol{\mu}) \leftarrow$ successive constraint method / heuristic strategies
Greedy procedure state problem:
  while $\max_{i\in\Xi_{train}} \Delta_n(\boldsymbol{\mu}_i) > \varepsilon_{RB}^{tol}$ **do**
    $\boldsymbol{\mu}^n = \arg\max_{i\in\Xi_{train}} \Delta_n(\boldsymbol{\mu}_i);\quad S_n = S_{n-1} \cup \mathrm{span}\{u_h(\boldsymbol{\mu}^n)\}$
    $A_q^n,\ F_q^n \leftarrow$ compute *reduced state* matrices
Greedy procedure dual problems: **for** each output $j = 1, \ldots, s$
  while $\max_{i\in\Xi_{train}} \Delta_n^j(\boldsymbol{\mu}_i) > \varepsilon_{RB}^{tol}$ **do**
    $\boldsymbol{\mu}_j^n = \arg\max_{i\in\Xi_{train}} \Delta_n^j(\boldsymbol{\mu}_i);\ S_n^j = S_{n-1}^j \cup \mathrm{span}\{\psi_h^j(\boldsymbol{\mu}_j^n)\}$
    $A_{q,j}^n,\ L_{q,j}^n \leftarrow$ compute *reduced dual* matrices
procedure REM CALIBRATION
  for j $= 1 : N_{cal}$ **do**
    $\mathbf{s}_h(\boldsymbol{\mu}^j),\ u_h(\boldsymbol{\mu}^j) \leftarrow$ FE state problem
    $\mathbf{s}_n(\boldsymbol{\mu}^j),\ u_n(\boldsymbol{\mu}^j),\ \Delta_n^{(1:s)}(\boldsymbol{\mu}^j), \leftarrow$ RB state and dual problems
    $err_{j,1:s} = \mathbf{s}_h^{(1:s)}(\boldsymbol{\mu}^j) - \mathbf{s}_n^{(1:s)}(\boldsymbol{\mu}^j)$
  compute REM

Algorithm 2 Online procedure

procedure METROPOLIS SAMPLING
  $\boldsymbol{\mu}^{(1)} \leftarrow$ initial value
sampling loop:
  for cont $= 2 : K$ **do**
    $\bar{\boldsymbol{\mu}} \leftarrow$ random walk
    $[\mathbf{s}_n(\bar{\boldsymbol{\mu}}), \Delta_n(\bar{\boldsymbol{\mu}})] \leftarrow$ compute RB *state + dual problems*
    $\mathbf{m}_{\mathrm{ROM}}(\bar{\boldsymbol{\mu}}) \leftarrow$ evaluate REM mean
    if REM is *deterministic* **then**
      $\tilde{\pi}^n \leftarrow \pi_\varepsilon(\mathbf{s}^* - \mathbf{s}_n(\bar{\boldsymbol{\mu}}) - \mathbf{m}_{\mathrm{ROM}}(\bar{\boldsymbol{\mu}}))$
    if REM is *statistical* **then**
      $\Sigma_{\mathrm{ROM}}(\bar{\boldsymbol{\mu}}) \leftarrow$ evaluate REM covariance matrix
      $\tilde{\pi}^n \leftarrow \pi_{\tilde{\boldsymbol{\delta}}}(\mathbf{s}^* - \mathbf{s}_n(\bar{\boldsymbol{\mu}}) - \mathbf{m}_{\mathrm{ROM}}(\bar{\boldsymbol{\mu}}))$
    $\tilde{\pi}_{post}^n(\bar{\boldsymbol{\mu}}|\mathbf{s}^*) \leftarrow$ Bayes' formula
    $\gamma \leftarrow \tilde{\pi}_{post}^n(\bar{\boldsymbol{\mu}}|\mathbf{s}^*)/\tilde{\pi}_{post}^n(\boldsymbol{\mu}^{(k)}|\mathbf{s}^*)$
    $y \leftarrow$ random sampling from $U(0,1)$
    if $y < \gamma$ **then**
      $\boldsymbol{\mu}^{(k+1)} \leftarrow \bar{\boldsymbol{\mu}};\ k \leftarrow k+1$
burn-in:
    eliminate first M simulations ~~$\boldsymbol{\mu}^{(1:M)}$~~
thinning:
    keep every d-th draw of the chain $\boldsymbol{\mu}^{(1:d:end)}$

6 Effectivity of proposed Reduction Error Models

Let us now analyze the effectivity of the *corrections* made on the reduced-order likelihood function thanks to the proposed reduction error models. In particular, we aim at stating some conditions to be fulfilled by REM corrections in order to guarantee that the corresponding posterior PDF $\tilde{\pi}_n$ is more robust and closer to the full-order PDF π_h than the reduced-order PDF π_n without corrections.

To this end, let us recall the notion of Kullback-Leibler (KL) divergence, which is a non-symmetric measure of the difference between two probability distributions π_A and π_B:

$$D_{KL}(\pi_A||\pi_B) = \int \pi_A(z) \log\left(\frac{\pi_A(z)}{\pi_B(z)}\right) dz. \tag{26}$$

Clearly, $D_{KL}(\pi_A||\pi_B) \geq 0$ whereas $D_{KL}(\pi_A||\pi_B) = 0$ if $\pi_A = \pi_B$ almost surely. This notion has already been used to compare approximations of posterior distributions obtained through generalized polynomial chaos representations, see e.g. [30, 8] for further details.

6.1 Consistency result

Before comparing our REMs and showing their effect on the reduced-order posterior PDFs, we prove that the reduced-order likelihood function π^n approximates the full-order one π^h in a *consistent* way, as long as the ROM dimension increases:

Proposition 1. *Let us consider the additive Gaussian noise model* (4) *and the RB approximation* $\mathbf{s}_n(\boldsymbol{\mu})$ *of the output* $\mathbf{s}_h(\boldsymbol{\mu})$ *defined by* (11) *and* (2)*, by assuming an analytic* $\boldsymbol{\mu}$*-dependence in the bilinear/linear forms. Then, for any* $\boldsymbol{\mu} \in \mathcal{P}$*,*

$$D_{KL}(\pi^h||\pi^n) = \sum_{j=1}^{s} \frac{1}{2\sigma_j^2}(s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu}))^2, \tag{27}$$

so that $\lim_{n\to N_h} D_{KL}(\pi^h||\pi^n) = 0$ *exponentially.*

Proof. The solution $u_n(\boldsymbol{\mu}) \in X_n$ of (2) is obtained as a Galerkin projection over X_n, then

$$\|u_h(\boldsymbol{\mu}) - u_n(\boldsymbol{\mu})\|_X \leq \left(\frac{\bar{M}}{\alpha_0}\right)^{1/2} \inf_{w_n \in X_n} \|u_h(\boldsymbol{\mu}) - w_n\|_X$$

being $M(\boldsymbol{\mu}) \leq \bar{M}$ the continuity constant of $a(\cdot,\cdot;\boldsymbol{\mu})$ and $\alpha_0 > 0$ such that $\alpha_h(\boldsymbol{\mu}) \geq \alpha_0$ for any $\boldsymbol{\mu} \in \mathcal{P}$. Thus we have that $\|u_h(\boldsymbol{\mu}) - u_n(\boldsymbol{\mu})\|_X \to 0$ when $n \to N_h$. By the same argument, $\|\psi_h(\boldsymbol{\mu}) - \psi_n(\boldsymbol{\mu})\|_X \to 0$ when $n \to N_h$. Moreover, the Kolmogorov n-width[1] of the solution set $\mathcal{M}_h$ converges exponentially:

$$d_n(\mathcal{M}_h; X) \leq Ce^{-\alpha n} \quad \text{for some } C, \alpha > 0, \tag{28}$$

provided the $\boldsymbol{\mu}$–dependence in the bilinear/linear forms is analytic; see [20] for further details. Regarding the output, by exploiting twice Galerkin orthogonality we have that

$$\begin{aligned} s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu}) = &\quad \ell^j(u_h(\boldsymbol{\mu})) - \ell^j(u_n(\boldsymbol{\mu})) = a(u_n(\boldsymbol{\mu}), \psi_n(\boldsymbol{\mu}); \boldsymbol{\mu}) - a(u_h(\boldsymbol{\mu}), \psi_h(\boldsymbol{\mu}); \boldsymbol{\mu}) \\ = &\quad a(u_n(\boldsymbol{\mu}) - u_h(\boldsymbol{\mu}), \psi_n(\boldsymbol{\mu}) - \psi_h(\boldsymbol{\mu}); \boldsymbol{\mu}). \end{aligned}$$

Then, $|s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu})| \leq \bar{M}\|u_h(\boldsymbol{\mu}) - u_n(\boldsymbol{\mu})\|_X \|\psi_h(\boldsymbol{\mu}) - \psi_n(\boldsymbol{\mu})\|_X$ so that, when $n \to N_h$, $|s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu})| \to 0$ for any $j = 1, \ldots, s$; furthermore, the order of convergence of the ROM output to the FEM output is twice larger than the one of the state solution. In particular, by considering an additive Gaussian noise model, for any $\boldsymbol{\mu} \in \mathcal{P}$

$$D_{KL}(\pi^h||\pi^n) = \int_{\mathbb{R}^s} \pi^h(\mathbf{s}|\boldsymbol{\mu}) \log\left(\frac{\pi^h(\mathbf{s}|\boldsymbol{\mu})}{\pi^n(\mathbf{s}|\boldsymbol{\mu})}\right) d\mathbf{s} = \sum_{j=1}^{s} \frac{1}{2\sigma_j^2}(s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu}))^2,$$

thanks to the definition (26) of Kullback-Leibler divergence, so that $\lim_{n\to N_h} D_{KL}(\pi^h||\pi^n) \to 0$; finally, convergence takes place at exponential rate thanks to (28). □

The consistency property can be extended to the posterior PDFs according to the

Proposition 2. *Under the assumptions of Proposition 1,* $\lim_{n\to N_h} D_{KL}(\pi^h_{post}||\pi^n_{post}) \to 0$ *exponentially.*

Proof. Thanks to the definition of KL divergence,

$$\begin{aligned} D_{KL}(\pi^h_{post}||\pi^n_{post}) = &\quad \int_{\mathcal{P}} \frac{\pi^h(\mathbf{s}^*|\boldsymbol{\mu})\pi_{prior}(\boldsymbol{\mu})}{\eta_h(\mathbf{s}^*)} \log\left(\frac{\pi^h(\mathbf{s}^*|\boldsymbol{\mu})}{\pi^n(\mathbf{s}^*|\boldsymbol{\mu})}\frac{\eta_n(\mathbf{s}^*)}{\eta_h(\mathbf{s}^*)}\right) d\boldsymbol{\mu} \\ = &\quad \log\left(\frac{\eta_n(\mathbf{s}^*)}{\eta_h(\mathbf{s}^*)}\right) + \int_{\mathcal{P}} \frac{\pi^h(\mathbf{s}^*|\boldsymbol{\mu})\pi_{prior}(\boldsymbol{\mu})}{\eta_h(\mathbf{s}^*)} \log\left(\frac{\pi^h(\mathbf{s}^*|\boldsymbol{\mu})}{\pi^n(\mathbf{s}^*|\boldsymbol{\mu})}\right) d\boldsymbol{\mu}. \end{aligned} \tag{29}$$

By using the definition of π^h and π^n, and the Lipschitz-continuity of $\exp(-\mathbf{s})$ for $\mathbf{s} \geq \mathbf{0}$ (that is, $|e^{-\mathbf{s}} - e^{-\mathbf{t}}| \leq \Lambda|\mathbf{s} - \mathbf{t}|$ for any $\mathbf{s}, \mathbf{t} > \mathbf{0}$, with $\Lambda = 1$), we obtain

[1] The Kolmogorov n-width $d_n(\mathcal{M}_h; X)$ measures how a finite dimensional subspace uniformly approximates the manifold $\mathcal{M}_h = \{u_h(\boldsymbol{\mu}), \boldsymbol{\mu} \in \mathcal{P}\}$ of the PDE solutions: $d_n(\mathcal{M}_h; X) = \inf_{X_n \subset X} \sup_{u_h \in \mathcal{M}_h} \inf_{w_n \in X_n} \|u_h - w_n\|_X$ where the first infimum is taken over all linear subspaces $X_n \subset X$ of dimension n.

$$\begin{aligned}\left|\pi^h(\mathbf{s}^*|\boldsymbol{\mu}) - \pi^n(\mathbf{s}^*|\boldsymbol{\mu})\right| &= \prod_{j=1}^{s} \frac{1}{\sqrt{2\pi\sigma_j^2}} \left| \exp\left(-\frac{(s_j^* - s_n^j(\boldsymbol{\mu}))^2}{2\sigma_j^2}\right) - \exp\left(-\frac{(s_j^* - s_h^j(\boldsymbol{\mu}))^2}{2\sigma_j^2}\right)\right| \\ &\leq \prod_{j=1}^{s} \frac{1}{\sqrt{2\pi\sigma_j^2}} \left| -\frac{(s_j^* - s_n^j(\boldsymbol{\mu}))^2}{2\sigma_j^2} + \frac{(s_j^* - s_h^j(\boldsymbol{\mu}))^2}{2\sigma_j^2}\right| \\ &\leq \prod_{j=1}^{s} \frac{1}{\sqrt{2\pi\sigma_j^2}2\sigma_j^2} \left|s_n^j(\boldsymbol{\mu}) - s_h^j(\boldsymbol{\mu})\right| \left|2s_j^* - s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu})\right|\end{aligned}$$

so that, for any $\boldsymbol{\mu} \in \mathcal{P}$, $\left|\pi^h(\mathbf{s}^*|\boldsymbol{\mu}) - \pi^n(\mathbf{s}^*|\boldsymbol{\mu})\right| \to 0$ when $n \to N_h$ because $|s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu})| \to 0$ for any $j = 1, \ldots, s$. In the same way, $|\eta_n(\mathbf{s}^*) - \eta_h(\mathbf{s}^*)| = \left|\int_{\mathcal{P}} (\pi^h(\mathbf{s}^*|\boldsymbol{\mu}) - \pi^n(\mathbf{s}^*|\boldsymbol{\mu}))\pi_{prior}(\boldsymbol{\mu})d\boldsymbol{\mu}\right| \to 0$ for any given $\mathbf{s}^* \in \mathbb{R}^s$. Thus, both terms in the second line of (29) vanish for $n \to N_h$. □

6.2 A result of effectivity for the proposed REMs

Since we are mainly interested in the case where the ROM dimension n is fixed (and possibly small) we want to show that performing a correction according to a REM improves the quality of the reduced posterior (in terms of the the KL divergence).

By following the structure of the previous section, we first state a result dealing with the approximation of the likelihood function by a corrected ROM:

Proposition 3. *Under the assumptions of Proposition 1, if there exists $C_j < 1$ such that, for any $j = 1, \ldots, s$,*

$$|s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu}) - m_{ROM}^j(\boldsymbol{\mu})| \leq C_j |s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu})| \qquad \forall \boldsymbol{\mu} \in \mathcal{P} \tag{30}$$

then

$$D_{KL}(\pi^h||\tilde{\pi}^n) \leq (\max_{j=1,\ldots,s} C_j^2) D_{KL}(\pi^h||\pi^n) \tag{31}$$

provided that the correction is made according to a deterministic REM.

Proof. In analogy with relation (27), a correction operated by means of a deterministic REM affects just $\mathbb{E}[\mathbf{s}^*|\boldsymbol{\mu}]$, so that

$$D_{KL}(\pi^h||\tilde{\pi}^n) = \sum_{j=1}^{s} \frac{1}{2\sigma_j^2} (s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu}) - m_{ROM}^j(\boldsymbol{\mu}))^2. \tag{32}$$

Thus, under condition (30), (31) directly follows. □

By means of (30), we require that the correction provided by a REM is *effective*, that is, it yields a reduction in the KL divergence between the reduced-order and the full-order posterior PDFs, when in the former case a correction through a deterministic REM is considered. Instead, when relying on statistical REMs, we need to distinguish between two cases:

- the correction shall result in a normal random variable (REM-1 or REM-2), with mean $\mathbf{m}_{ROM}$ and covariance matrix $(\Sigma_{ROM})_{ij} = (\sigma_j^{ROM})^2 \delta_{ij}$. In this case, we would obtain

$$D_{KL}(\pi^h||\tilde{\pi}^n) = \frac{1}{2} \sum_{j=1}^{s} \left(\frac{(s_h^j(\boldsymbol{\mu}) - s_n^j(\boldsymbol{\mu}) - \tilde{\varepsilon}^j(\boldsymbol{\mu}))^2}{\sigma_j^2 + (\sigma_j^{ROM})^2} + \frac{\sigma^2}{\sigma_j^2 + (\sigma_j^{ROM})^2} - 1 - \log\left(\frac{\sigma_j^2}{\sigma_j^2 + (\sigma_j^{ROM})^2}\right)\right) \tag{33}$$

  instead of (32). Thus, in order to ensure that a relation like (31) still holds, we need to further require that $(\sigma_j^{ROM})^2$ is sufficiently small compared to σ_j^2, $j = 1, \ldots, s$ – in practice problems only arise if $\sigma_j = 0$ (i.e. the jth likelihood function is a true delta-function), whereas any realistic inverse problem will inherently experience finite variance in its outputs arising from measurement noise, numerical approximations etc., and so provided that the ROM approximation is convergent a sufficiently large reduced basis can always be found such that remainder terms in (33) are negligible;

- the correction shall result in a non normal distributed random variable (such as, e.g., when using REM-3, where the correction is a multivariate log-normal random variable). In this case we cannot provide a closed-form expression for the KL divergence. However, also in this case the key factors affecting the comparison between $D_{KL}(\pi^h||\tilde{\pi}^n)$ and $D_{KL}(\pi^h||\pi^n)$ are the same as in the previous case (as shown by the numerical results in the following section).

Let us now turn to evaluate how the corrections introduced through our REMs impact on the posterior PDFs. First of all, let us remark that, by taking the expectation of the KL divergence between $\pi^h(\mathbf{s}^*|\boldsymbol{\mu})$ and $\pi^n(\mathbf{s}^*|\boldsymbol{\mu})$, and changing the order of integration, we obtain

$$\mathbb{E}[D_{KL}(\pi^h||\pi^n)] = \int_{\mathcal{P}} D_{KL}(\pi^h||\pi^n)\pi_{prior}(\boldsymbol{\mu})d\boldsymbol{\mu}. \tag{34}$$

Moreover, thanks to the positivity of the KL divergence and relation (31), we get

$$\mathbb{E}[D_{KL}(\pi^h||\tilde{\pi}^n)] \leq \left(\max_{j=1,\dots,s} C_j^2\right) \mathbb{E}[D_{KL}(\pi^h||\pi^n)]. \tag{35}$$

6.3 Posterior comparison for fixed n

We now want to compare the KL divergences between the full-order and the corrected/uncorrected posterior PDFs for small n. We can show the following

Proposition 4. *Under the assumptions of Proposition 3 and provided that $\eta_n(\mathbf{s}) \sim \eta_h(\mathbf{s})$ for $n \to N_h$, for any $\mathbf{s} \in \mathbb{R}^s$, we have that*

$$\mathbb{E}[D_{KL}(\pi^h_{post}||\tilde{\pi}^n_{post})] \leq \mathbb{E}[D_{KL}(\pi^h_{post}||\pi^n_{post})] \tag{36}$$

if the correction is made according to a deterministic REM.

Proof. Let us express the right-hand side of (36) as

$$\mathbb{E}[D_{KL}(\pi^h_{post}||\pi^n_{post})] = \int_{\mathbb{R}^s} \left(\log\left(\frac{\eta_n(\mathbf{s})}{\eta_h(\mathbf{s})}\right) + \int_{\mathcal{P}} \frac{\pi^h(\mathbf{s}|\boldsymbol{\mu})\pi_{prior}(\boldsymbol{\mu})}{\eta_h(\mathbf{s})} \log\left(\frac{\pi^h(\mathbf{s}|\boldsymbol{\mu})}{\pi^n(\mathbf{s}|\boldsymbol{\mu})}\right) d\boldsymbol{\mu}\right) \eta_h(\mathbf{s})d\mathbf{s}. \tag{37}$$

In the same way, the left-hand side of (36) becomes

$$\mathbb{E}[D_{KL}(\pi^h_{post}||\tilde{\pi}^n_{post})] = \int_{\mathbb{R}^s} \left(\log\left(\frac{\tilde{\eta}_n(\mathbf{s})}{\eta_h(\mathbf{s})}\right) + \int_{\mathcal{P}} \frac{\pi^h(\mathbf{s}|\boldsymbol{\mu})\pi_{prior}(\boldsymbol{\mu})}{\eta_h(\mathbf{s})} \log\left(\frac{\pi^h(\mathbf{s}|\boldsymbol{\mu})}{\tilde{\pi}^n(\mathbf{s}|\boldsymbol{\mu})}\right) d\boldsymbol{\mu}\right) \eta_h(\mathbf{s})d\mathbf{s}. \tag{38}$$

We proceed by analyzing separately the two terms of the right-hand side of (37). The second term coincides with (34), i.e.

$$\begin{aligned}\int_{\mathbb{R}^s} \left(\int_{\mathcal{P}} \frac{\pi^h(\mathbf{s}|\boldsymbol{\mu})\pi_{prior}(\boldsymbol{\mu})}{\eta_h(\mathbf{s})} \log\left(\frac{\pi^h(\mathbf{s}|\boldsymbol{\mu})}{\pi^n(\mathbf{s}|\boldsymbol{\mu})}\right) d\boldsymbol{\mu}\right) \eta_h(\mathbf{s})d\mathbf{s} = \\ = \int_{\mathcal{P}} \left(\int_{\mathbb{R}^s} \pi^h(\mathbf{s}|\boldsymbol{\mu}) \log\left(\frac{\pi^h(\mathbf{s}|\boldsymbol{\mu})}{\pi^n(\mathbf{s}|\boldsymbol{\mu})}\right) d\mathbf{s}\right) \pi_{prior}(\boldsymbol{\mu})d\boldsymbol{\mu} = \mathbb{E}[D_{KL}(\pi^h||\pi^n)].\end{aligned}$$

In the same way, the second term of the right-hand side of (38) is such that

$$\int_{\mathbb{R}^s} \left(\int_{\mathcal{P}} \frac{\pi^h(\mathbf{s}|\boldsymbol{\mu})\pi_{prior}(\boldsymbol{\mu})}{\eta_h(\mathbf{s})} \log\left(\frac{\pi^h(\mathbf{s}|\boldsymbol{\mu})}{\tilde{\pi}^n(\mathbf{s}|\boldsymbol{\mu})}\right) d\boldsymbol{\mu}\right) \eta_h(\mathbf{s})ds = \mathbb{E}[D_{KL}(\pi^h||\tilde{\pi}^n)].$$

On the other hand, by developing the first term of (37) with a Taylor expansion, we obtain

$$\begin{aligned}\int_{\mathbb{R}^s} \log\left(\frac{\eta_n(\mathbf{s})}{\eta_h(\mathbf{s})}\right) \eta_h(\mathbf{s})ds &= \int_{\mathbb{R}^s} \left(\left(\frac{\eta_n(\mathbf{s})}{\eta_h(\mathbf{s})} - 1\right) - \frac{1}{2}\left(\frac{\eta_n(\mathbf{s})}{\eta_h(\mathbf{s})} - 1\right)^2 + \mathcal{O}\left(\frac{\eta_n(\mathbf{s})}{\eta_h(\mathbf{s})} - 1\right)^3\right) \eta_h(\mathbf{s})d\mathbf{s} \\ &= \int_{\mathbb{R}^s} (\eta_n(\mathbf{s}) - \eta_h(\mathbf{s}))d\mathbf{s} - \int_{\mathbb{R}^s} \frac{1}{2}\left(\frac{\eta_n(\mathbf{s})^2}{\eta_h(\mathbf{s})} - 2\eta_n(\mathbf{s}) + \eta_h(\mathbf{s})\right) d\mathbf{s} + \int_{\mathbb{R}^s} \mathcal{O}\left(\frac{\eta_n(\mathbf{s})}{\eta_h(\mathbf{s})} - 1\right)^3 \eta_h(\mathbf{s})d\mathbf{s}.\end{aligned}$$

The first term of the last sum can be rewritten as

$$\int_{\mathbb{R}^s} (\eta_n(\mathbf{s}) - \eta_h(\mathbf{s}))ds = \int_{\mathcal{P}} \left(\int_{\mathbb{R}^s} \pi^n(\mathbf{s}|\boldsymbol{\mu})d\mathbf{s}\right) \pi_{prior}(\boldsymbol{\mu})d\boldsymbol{\mu} - \int_{\mathcal{P}} \left(\int_{\mathbb{R}^s} \pi^h(\mathbf{s}|\boldsymbol{\mu})d\mathbf{s}\right) \pi_{prior}(\boldsymbol{\mu})d\boldsymbol{\mu}, \tag{39}$$

and it is vanishing, since

$$\int_{\mathcal{P}} \left(\int_{\mathbb{R}^s} \pi^h(\mathbf{s}|\boldsymbol{\mu}) d\mathbf{s} \right) \pi_{prior}(\boldsymbol{\mu}) d\boldsymbol{\mu} = \int_{\mathcal{P}} \left(\int_{\mathbb{R}^s} \pi^n(\mathbf{s}|\boldsymbol{\mu}) d\mathbf{s} \right) \pi_{prior}(\boldsymbol{\mu}) d\boldsymbol{\mu} = \int_{\mathcal{P}} \pi_{prior}(\boldsymbol{\mu}) d\boldsymbol{\mu} = 1.$$

In this way

$$\int_{\mathbb{R}^s} \log\left(\frac{\eta_n(\mathbf{s})}{\eta_h(\mathbf{s})}\right) \eta_h(\mathbf{s}) d\mathbf{s} = -\frac{1}{2} \int_{\mathbb{R}^s} \left(\frac{\eta_n(\mathbf{s})^2}{\eta_h(\mathbf{s})} - 1 \right) d\mathbf{s} + \int_{\mathbb{R}^s} \mathcal{O}\left(\frac{\eta_n(\mathbf{s})}{\eta_h(\mathbf{s})} - 1 \right)^3 \eta_h(\mathbf{s}) d\mathbf{s},$$

considering the integral of the remainder term of the Taylor expansion to be sufficient small when $\eta_n \sim \eta_h$. Similarly, the first term of the right-hand side of (38) is negligible, so that

$$\mathbb{E}[D_{KL}(\pi^h_{post}||\tilde{\pi}^n_{post})] = -\frac{1}{2} \int_{\mathbb{R}^s} \left(\frac{\tilde{\eta}_n(\mathbf{s})^2}{\eta_h(\mathbf{s})} - 1 \right) d\mathbf{s} + \mathbb{E}[D_{KL}(\pi^h||\tilde{\pi}^n)] + \int_{\mathbb{R}^s} \mathcal{O}\left(\frac{\tilde{\eta}_n(\mathbf{s})}{\eta_h(\mathbf{s})} - 1 \right)^3 \eta_h(\mathbf{s}) d\mathbf{s}$$

$$\mathbb{E}[D_{KL}(\pi^h_{post}||\pi^n_{post})] = -\frac{1}{2} \int_{\mathbb{R}^s} \left(\frac{\eta_n(\mathbf{s})^2}{\eta_h(\mathbf{s})} - 1 \right) d\mathbf{s} + \mathbb{E}[D_{KL}(\pi^h||\pi^n)] + \int_{\mathbb{R}^s} \mathcal{O}\left(\frac{\eta_n(\mathbf{s})}{\eta_h(\mathbf{s})} - 1 \right)^3 \eta_h(\mathbf{s}) d\mathbf{s}.$$

In conclusion, by using (35), inequality (36) follows under the following condition:

$$\begin{aligned} \frac{1}{2} \int_{\mathbb{R}^s} \left(\frac{\eta_n(\mathbf{s})^2}{\eta_h(\mathbf{s})} - \frac{\tilde{\eta}_n(\mathbf{s})^2}{\eta_h(\mathbf{s})} \right) d\mathbf{s} + & \int_{\mathbb{R}^s} \mathcal{O}\left(\frac{\tilde{\eta}_n(\mathbf{s})}{\eta_h(\mathbf{s})} - 1 \right)^3 \eta_h(\mathbf{s}) d\mathbf{s} \\ & \leq \left(1 - \max_{j=1,\dots,s} C_j^2 \right) \mathbb{E}[D_{KL}(\pi^h||\pi^n)] + \int_{\mathbb{R}^s} \mathcal{O}\left(\frac{\eta_n(\mathbf{s})}{\eta_h(\mathbf{s})} - 1 \right)^3 \eta_h(\mathbf{s}) d\mathbf{s}, \end{aligned}$$

which can be seen as a robustness condition on the correction entailed by the REM. □

Remark 5. *According to* (6), (9), $\eta_n \sim \eta_h$ *as soon as the likelihood functions* π^h *and* π^n *are very close to each other, that is,* $D_{KL}(\pi^h||\pi^n) < \varepsilon$ *for any given, small* $\varepsilon > 0$.

7 Numerical results and discussion

We present two numerical examples illustrating the properties and the performance of the correction strategies proposed in Sect. 4. We recall that our final goal is to exploit fast and inexpensive *online* ROM queries by improving their accuracy through suitable REMs. For the cases at hand, we consider a model problem dealing with the scalar heat equation, where uncertain parameters describe either the scalar conductivities over different subdomains or the continuous conductivity field over the whole domain. By measuring the average of the state solution over three boundary portions, identifiable parameters are reconstructed by marginalizing away the nuisance parameters.

7.1 Test case 1

We consider in $\Omega = (0, 1.5)^2$ the following diffusion problem:

$$\begin{cases} \nabla \cdot (k(\mathbf{x}, \boldsymbol{\mu}) \nabla u) = 0 & in\ \Omega \\ k(\mathbf{x}, \boldsymbol{\mu}) \nabla u \cdot \mathbf{n} = 0 & on\ \Gamma_w \\ k(\mathbf{x}, \boldsymbol{\mu}) \nabla u \cdot \mathbf{n} = 1 & on\ \Gamma_b \\ u = 0 & on\ \Gamma_t, \end{cases} \tag{40}$$

where $\partial\Omega = \Gamma_w \cup \Gamma_b \cup \Gamma_t$ (see Fig. 1, left). Here $k(\mathbf{x}, \boldsymbol{\mu})$ is a parametrized diffusion coefficient:

$$k(\mathbf{x}, \boldsymbol{\mu}) = 0.1 \mathbb{I}_{\Omega_0}(\mathbf{x}) + \sum_{i=1}^{3} \mu_i \mathbb{I}_{\Omega_i}(\mathbf{x}),$$

$\mathbb{I}_{\Omega_i}$ is the characteristic function of Ω_i, being $\Omega_1 = (0, 0.5)^2 \cup (1, 1.5)^2$, $\Omega_2 = (0, 0.5) \times (1, 1.5) \cup (1, 1.5) \times (0, 0.5)$, $\Omega_3 = (0.5, 1)^2$ and $\Omega_0 = \Omega \setminus \bigcup_{i=1}^3 \Omega_i$; the outputs are

$$s^j(\boldsymbol{\mu}) = \int_{\Gamma_j} u(\boldsymbol{\mu}) d\Gamma, \quad j = 1, 2, 3.$$

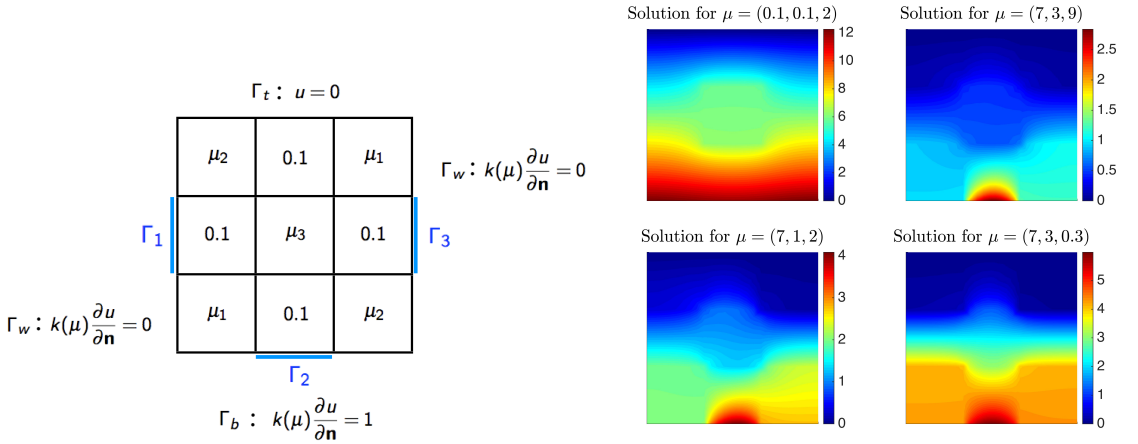

Figure 1: Test case 1: domain, boundary conditions (left) and different sample solutions of 40 (right).

Our objective is to identify $\gamma = \mu_3$ by observing the outputs $s^j(\boldsymbol{\mu})$, $j = 1, 2, 3$, in presence of two nuisance parameters $\boldsymbol{\zeta} = (\mu_1, \mu_2)$. We suppose that the target value $\mathbf{s}^*$ corresponds to the full order output vector evaluated for $\gamma^* = 2$ and a random value of $\boldsymbol{\zeta} \in (0.01, 10)^2$.

The forward problem (40) is first discretized using the FE approach with linear $\mathbb{P}_1$ finite elements, which generates a total of $\mathcal{N}_h = 1,056$ degrees of freedom. In view of the application of the MCMC algorithm, we adopt the RB method: we stop the greedy algorithm after selecting $n = 10$ basis functions for the forward problem and for each dual problem related to one of the three outputs, in view of the analysis of the corrections effects. We intentionally select few basis functions (satisfying a tolerance $\epsilon_{RB}^{tol} = 5 \cdot 10^{-2}$) in order to assess the capability of our REMs in correcting the outcome of a possible inaccurate ROM. Note that for more complex (e.g. non-affine and/or non-linear) problems a larger number of RB functions might be required to guarantee a sufficient accuracy, with a consequent loss of efficiency. As a matter of fact, we reach in our case a considerable speedup ($n/\mathcal{N}_h \simeq 1/100$). Nevertheless, output evaluations are affected by ROM errors, which have to be taken into account by the correction methods. To this aim, we construct a calibration set of $N_{cal} = 100$ points in the parameter space from a Gauss-Patterson sparse grid, from which we calculate the full-order FE as well as the RB solutions and the relative outputs. In this way, we obtain a sample of outputs ROM errors, upon which we calibrate the proposed REMs. This operation can be performed in parallel and does not impact significantly on the offline complexity. We highlight that in our test cases very similar results in terms of calibration performances can be obtained by relying on a random sampling of the parameter space, due to the small parametric dimension.

Before analyzing the quality of the corrections, we provide few details related to REMs construction. By a direct inspection of the errors sample generated over the calibration set, we remark that their distribution is far from being Gaussian (Fig. 2).

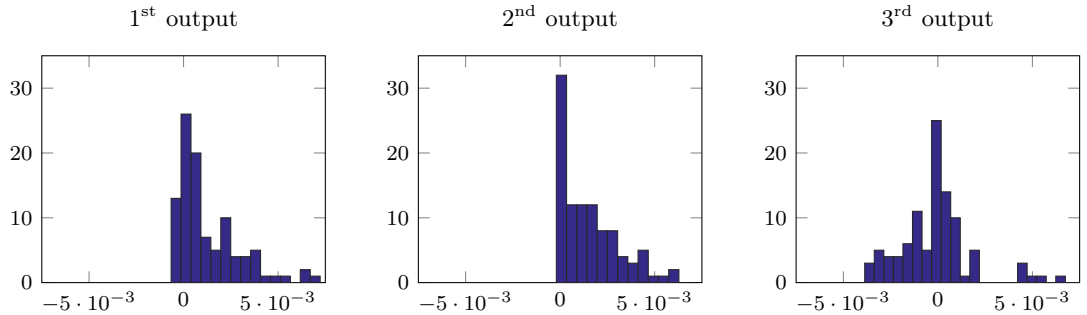

Figure 2: Histogram of the ROM errors for each output $s^j(\boldsymbol{\mu})$, $j = 1, 2, 3$.

The Shapiro-Wilk test rejects the null hypothesis that the errors come from a normal distribution (*p-value* $< 10^{-5}$ for each $j = 1, 2, 3$). Moreover, errors are not equally distributed over

the parameter space: for this reason $\boldsymbol{\mu}$-dependent approaches (like REM-2 and REM-3) should better capture the error variability. Concerning REM-2, we point out that the approach relying on error bound effectivities is preferable to the one based on ROM errors if we aim at minimizing the maximum norm of the interpolation error (see Fig. 3).

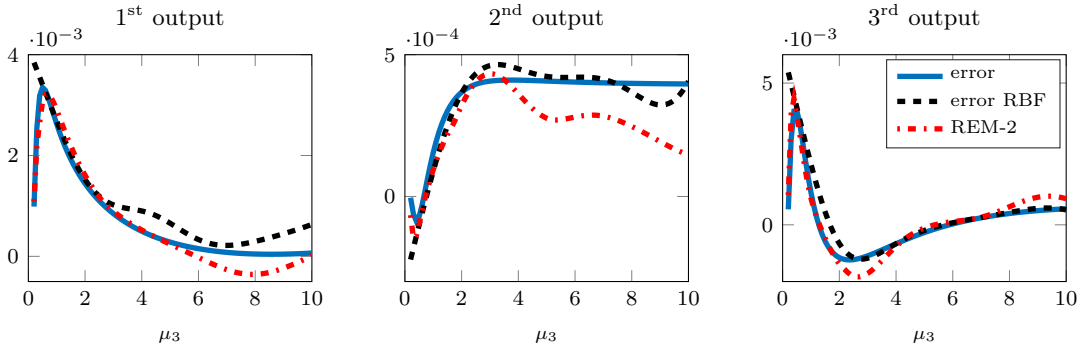

Figure 3: REM-2 construction: comparison between the absolute values of the error (in blue), of the RBF interpolation of the errors (in black, dashed) and of the REM-2 reconstructed errors (in red, dashed), for each output $s^j(\boldsymbol{\mu})$, $j = 1, 2, 3$, on varying μ_3 ($\mu_1 = 7$, $\mu_2 = 3$).

Furthermore, for REM-3, a linear regression model on the log-variable is fitted for each output and by distinguishing the errors sign. (see Fig. 4). It is not surprising that models providing a better fitting are those built on larger datasets, even if the Bernoulli sign trick proposed in Sect. 4.3 contributes to filter out those models built on smaller samples.

A primary test for assessing the quality of our REMs has been made by considering a random sample of 5000 points in $\mathcal{P}$ and computing the frequencies of ROM errors and the corresponding corrections obtained with the three REMs, when (for the sake of comparison) $N_{cal} = 50$ or $N_{cal} = 200$ points in the calibration sample are considered. (see Fig. 5).

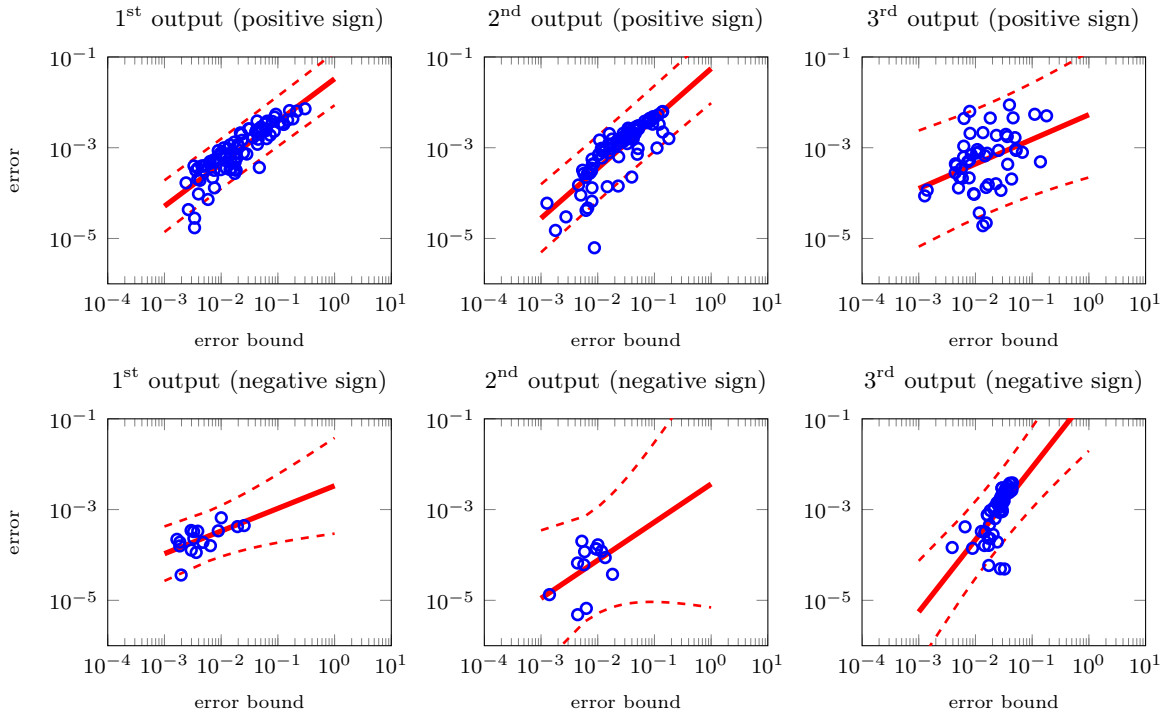

Figure 4: REM-3 construction: linear regression (in the log / log space) of the ROM errors against error bounds and 95% confidence intervals. A linear model (22) in the log / log space is fitted over each subset of computed error bounds and corresponding (positive or negative) errors S^j_+, S^j_-, $j = 1, 2, 3$.

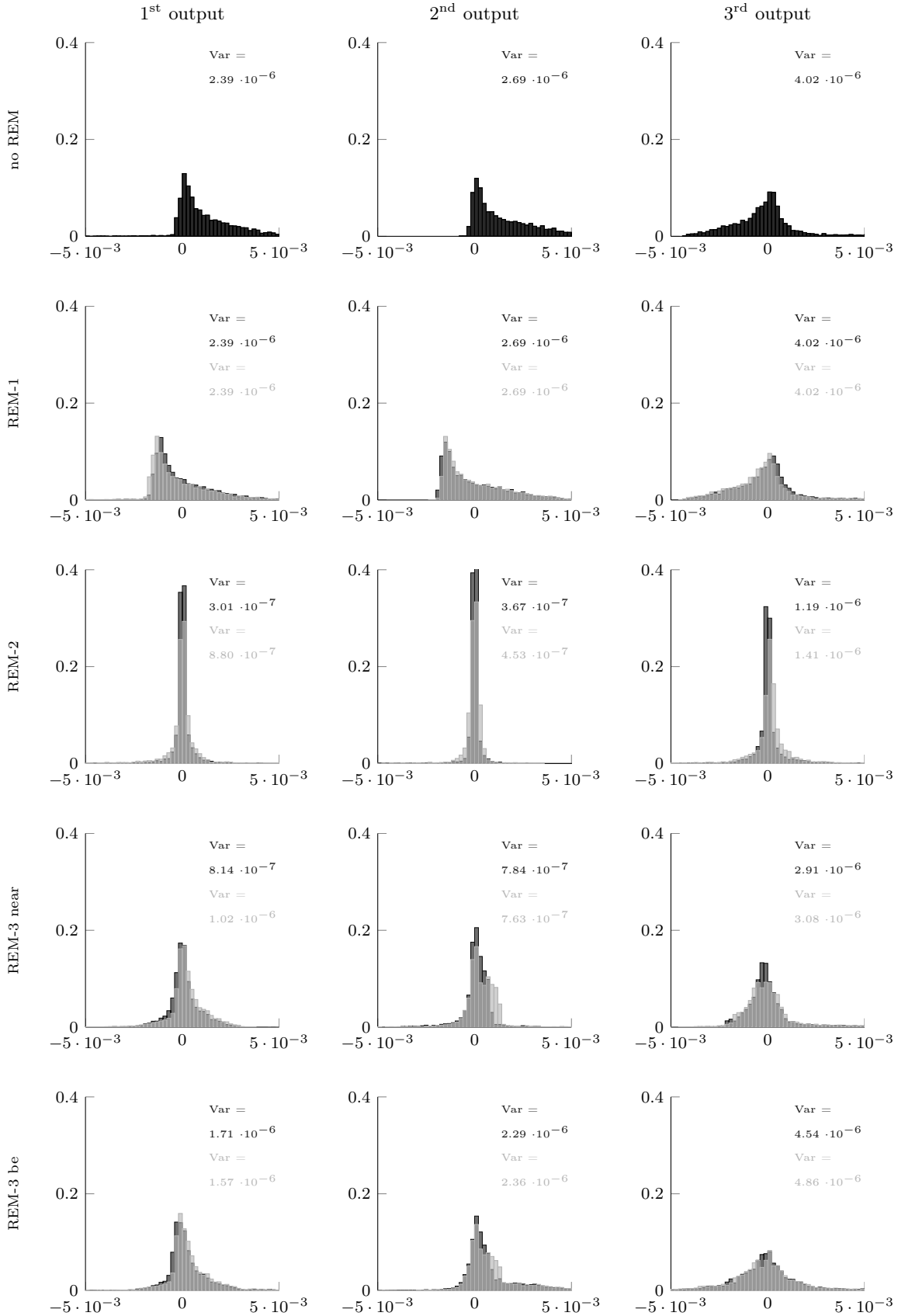

Figure 5: Comparison between the REM reconstruction of the errors for the outputs s^1, s^2 and s^3 (from top to bottom: without REM, REM-1, REM-2, REM-3 with both options for the sign model). Two different calibration samples have been considered: $N_{cal} = 50$ for the gray distribution $N_{cal} = 200$ for the black one.

As expected, REM-1 results in a simple shift of the errors distribution. On the other hand, the distributions of the errors in presence of $\boldsymbol{\mu}$-dependent approaches (REM-2 and REM-3) show a more symmetric shape, with mean closer to zero and smaller variance than in the ROM errors distributions. Hence, the $\boldsymbol{\mu}$-dependent approaches allow to reduce the ROM error by at least one order of magnitude. REM-2 performs very well in this case, although it can suffer from the curse of dimensionality.

Then, we have applied the MCMC algorithm to reconstruct the unknown parameters values, by starting from a uniform prior on the parameters and sampling the RB posterior corrected by the three different REM. A closer look to the results reported in Figs. 5–7 allows to conclude that:

- the correction provided by REM-1 does not improve the accuracy of the RB posterior since the distribution of the calibration ROM errors is far from being Gaussian. Even by assuming log-normal calibration ROM errors, we do not get a significant improvement on the accuracy of the corrected posterior (see Fig. 6);
- REM-2 is quite good in terms of performance when compared purely in terms of model error estimation to REM-1 and REM-3, see Fig. 5. As a matter of fact, REM-2 enhances also the evaluation of reduced posteriors; the same conclusion can also be drawn for REM-3, provided a suitable sign model is taken into account (see Fig. 7, left);
- by considering also statistical correction (Fig. 7, right), we get maximum a-posteriori estimates close to the full-order ones, even if this procedure generates heavy-tailed *posterior* distributions and consequently less tight a posteriori prediction intervals.

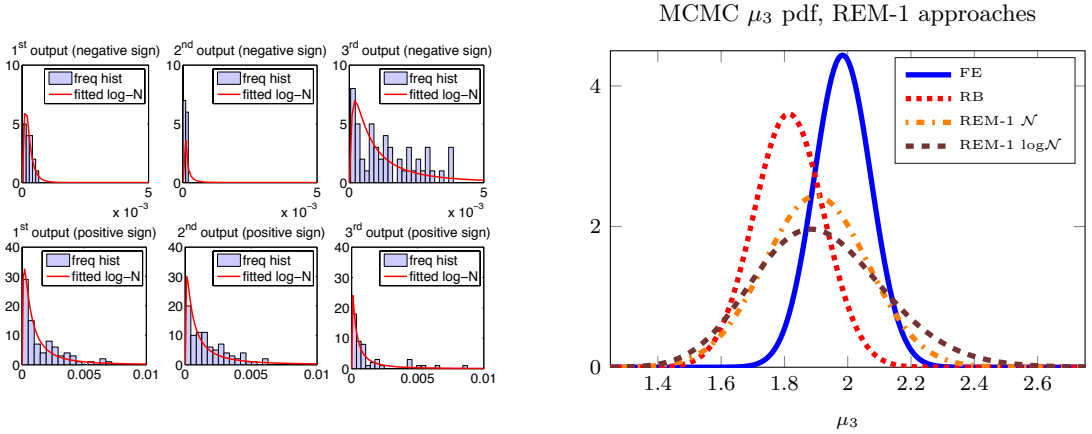

Figure 6: Distribution of the output errors divided by the sign (left) and the corrected reduced posterior obtained by REM-1 correction (right), starting from a uniform *prior* distribution. REM-1 does not yield a relevant improvement of the posterior distribution obtained with the RB method.

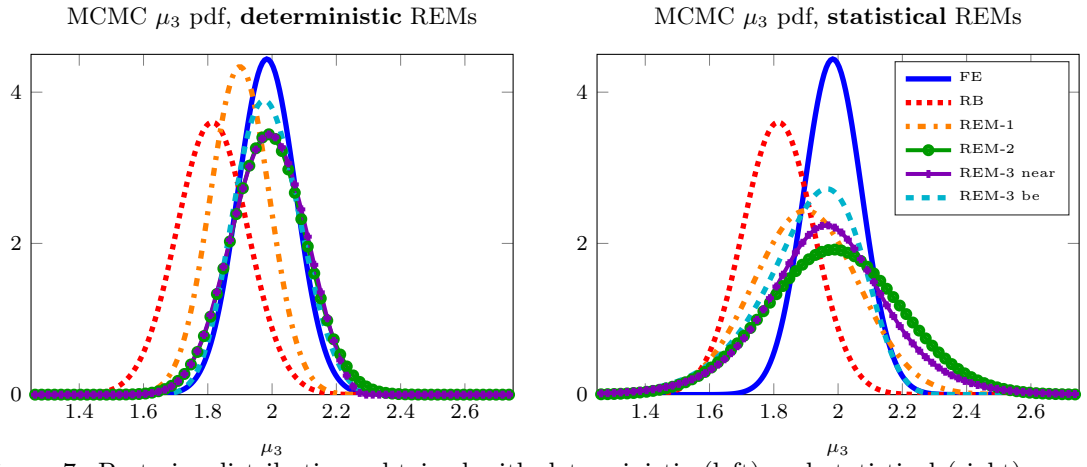

Figure 7: Posterior distributions obtained with deterministic (left) and statistical (right) corrections, using a uniform prior distribution. REM-2 and REM-3 yield a much relevant improvement of the posterior distribution obtained with the RB method than REM-1.

So far we have considered a uniform prior distribution. Turning to a Gaussian *prior* on the parameters, numerical results still confirm a good behavior of REM-2 and REM-3, and a worse performance of REM-1. In this case, we have tested a situation of wrong a priori input with low confidence on it, i.e. $\mu_{p,3} = 3$ and variance $\sigma_i^2 \delta_{ii} = 1,\ 1 \leq i \leq s$. As before, we are able to get a good reconstruction of the full-order posterior distribution for REM-2 and REM-3 (see Fig. 8).

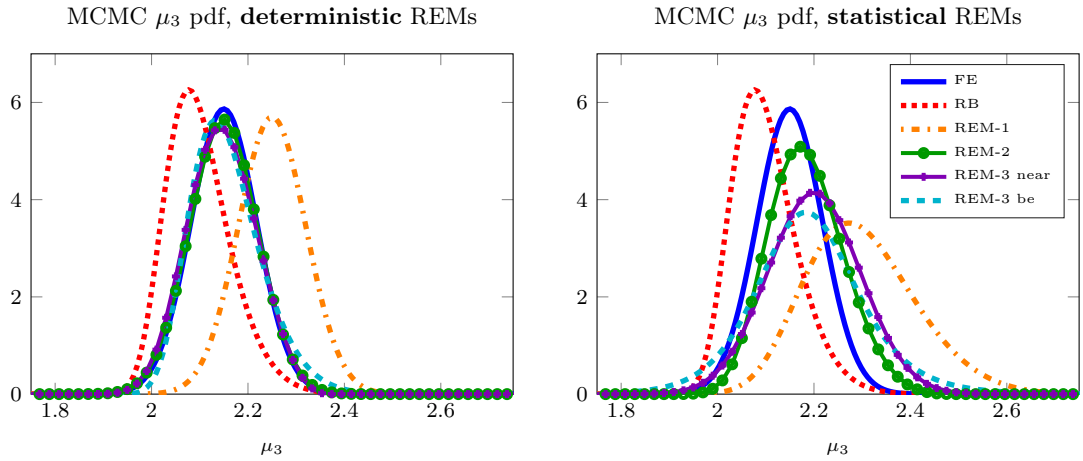

Figure 8: Posterior distributions obtained with deterministic (left) and statistical (right) REMs in the case of Gaussian *prior* distribution. Also in this case REM-2 and REM-3 yield a much relevant improvement of the posterior distribution obtained with the RB method than REM-1.

Finally, as a validation of the theoretical results presented in Sect. 6, we report in Table 1 the maximum a-posteriori (MAP) estimate, the 95% prediction interval (PI) and the KL-divergences for the three REMs and both prior distributions. We are able to verify the relation between REMs corrections and KL divergences of Proposition 4. However, this relation does not hold in presence of a significant variance associated with the correction (in this case the KL divergence increases).

	uniform prior			Gaussian prior		
	μ_3^{MAP}	PI 95%	KL-dist	μ_3^{MAP}	PI 95%	KL-dist
FE	1.9837	(1.8075, 2.1598)		2.1498	(2.0166, 2.2830)	
RB	1.8139	(1.5968, 2.0311)	1.2152	2.0776	(1.9879, 2.2499)	0.2936
REM-1	1.8997	(1.5783, 2.2213)	0.3829	2.2726	(2.0986, 2.5446)	1.0856
REM-2	1.9881	(1.7608, 2.2150)	0.0555	2.1493	(2.0110, 2.2876)	0.0014
REM-3 (near)	1.9917	(1.7623, 2.1898)	0.0516	2.1420	(2.0102, 2.2818)	0.0213
REM-3 (near+var)	1.9719	(1.4932, 2.4501)	0.3423	2.2121	(2.0224, 2.4017)	0.3064
REM-3 (BE)	1.9762	(1.8002, 2.1850)	0.0406	2.1331	(2.0329, 2.3228)	0.0124
REM-3 (BE+var)	1.9645	(1.5546, 2.3723)	0.2474	2.1830	(1.9553, 2.4464)	0.2538

Table 1: MAP estimates, 95% prediction intervals and KL divergences from the posterior FE distribution.

The use of the RB method allows to obtain a reduction of order 10^2 in terms of both RB dimension and required CPU time for each *Online* forward query. Moreover, the proposed REMs yield a KL divergence on the posterior distributions which is $50 \div 100$ times smaller in the case of REM-corrected posteriors, with respect to the case without correction.

Regarding computational costs, a REM-1 correction can be *inexpensively* obtained, whereas REM-2 and REM-3 entail a CPU time for the correction comparable to the one required by the solution of the RB system (see Table 6). Concerning the *Offline* CPU time, the calibration stage required by any REM entails about the 10% of the CPU time required by the *Offline* RB stage if $n = 10$ RB functions and $N_{cal} = 100$ calibration points are used.

The numerical performances in terms of both efficiency and accuracy obviously depend on the RB dimension n and the size N_{cal} of the calibration sample. To assess this fact, we have considered different combinations (n, N_{cal}) for the test case discussed along this section.

By comparing the KL divergences in Table 2, it clearly turns out that a better RB approximation yields a more accurate solution of the inverse UQ problem. Building a RB approximation of dimension $n = 10$ (resp. $n = 40$) requires an *Offline* CPU time of $134\,s$ (resp. $226\,s$), whereas its

		$n = 10$			$n = 40$		
		μ_3^{MAP}	PI 95%	KL-dist	μ_3^{MAP}	PI 95%	KL-dist
FE		1.9837	(1.8075, 2.1598)		1.9837	(1.8075, 2.1598)	
RB		1.8139	(1.5968, 2.0311)	1.2152	2.0645	(1.8892, 2.3114)	0.2806
REM-1	$N_{cal} = 50$	1.9192	(1.7872, 2.1104)	0.3524	1.9418	(1.8206, 2.2137)	0.0512
	$N_{cal} = 100$	1.8997	(1.5783, 2.2213)	0.3829	1.9618	(1.8357, 2.2300)	0.0675
	$N_{cal} = 200$	1.93393	(1.8535, 2.1022)	0.2464	1.9671	(1.8463, 2.2054)	0.0459
REM-2	$N_{cal} = 50$	1.9657	(1.8238, 2.2168)	0.0850	1.9752	(1.8175, 2.1902)	0.0430
	$N_{cal} = 100$	1.9881	(1.7608, 2.2150)	0.0555	1.9928	(1.7981, 2.2025)	0.0380
	$N_{cal} = 200$	1.9881	(1.8335, 2.1429)	0.0541	1.9782	(1.8076, 2.1735)	0.0344
REM-3 (near)	$N_{cal} = 50$	1.9713	(1.8260, 2.1166)	0.1075	1.9630	(1.8140, 2.1604)	0.0634
	$N_{cal} = 100$	1.9917	(1.7623, 2.1898)	0.0516	1.9923	(1.8180, 2.2136)	0.0351
	$N_{cal} = 200$	1.9760	(1.8002, 2.1851)	0.0405	1.9779	(1.8196, 2.1572)	0.0337
REM-3 (BE)	$N_{cal} = 50$	1.95873	(1.8637, 2.1482)	0.0947	1.9488	(1.8101, 2.1726)	0.0967
	$N_{cal} = 100$	1.9762	(1.8002, 2.1850)	0.0406	1.9665	(1.8138, 2.1370)	0.0344
	$N_{cal} = 200$	1.9814	(1.7929, 2.1699)	0.0377	1.9818	(1.8271, 2.1966)	0.0328

Table 2: MAP estimates, 95% prediction intervals and KL divergences from the posterior FE distribution obtained with a uniform prior, different RB dimensions $n = 10, 40$ and sizes $N_{cal} = 50, 100, 200$ of the calibration sample.

Online evaluation requires $2 \cdot 10^{-4}$ (resp. $7 \cdot 10^{-4}\, s$). This latter fact has a major drawback in the overall CPU time required by the MCMC procedure, which grows of the same factor 3.5. On the other hand, considering a calibration sample of increasing dimension $N_{cal} = 50, 100, 200$ yields better results in terms of (corrected) posterior distributions, however entailing a weaker increase of both (*Offline* and *Online*) costs: evaluating the errors over the calibration sample requires $6, 13, 25\, s$, respectively, during the *Offline* stage, whereas the *Online* correction requires almost the same time in the case of REM-2 ($7.5 \cdot 10^{-4}\, s$) and REM-3 with Bernoulli sign model ($4 \cdot 10^{-4}\, s$); in the case of REM-3 with nearest neighbor sign model, the cost varies from $4.75 \cdot 10^{-4}$ ($N_{cal} = 50$) to $1.05 \cdot 10^{-3}$ ($N_{cal} = 200$).

A REM correction thus proves to be necessary also in the case where a more accurate RB approximation is considered. Nevertheless, this latter may feature substantially higher costs, especially when dealing with more complex problems than the one considered in this paper. In this respect, a larger size N_{cal} of the calibration sample may yield more accurate results, without showing a too large impact on the computational efficiency. In any case, both n and N_{cal} have to be chosen according to the problem at hand; from our experience, to reach the same level of accuracy, it may be preferable to deal with *less* accurate (but cheaper) RB approximations and larger calibration samples; see, e.g., also the discussion reported in [7].

7.2 Test case 2

In order to consider a higher dimensional problems in terms of parametric dimension P, we modify problem (40) by considering a parametric field description of the diffusion coefficient $k(\mathbf{x})$: all the possible configurations of the field are generated from a standard multivariate Gaussian of dimension $\mathcal{N}_h$ (we also deal in this case with a discretization made by linear $\mathbb{P}_1$ finite elements). To reduce this complexity, we assume the field as generated by a multivariate Gaussian distribution with covariance matrix

$$C_{ij} = a \exp^{-\frac{\|x_i - x_j\|}{2b^2}} + c\delta_{ij} \qquad \forall i, j = 1, \ldots, \mathcal{N}_h,$$

$a, b, c > 0$ and $\{x_i\}_{i=1}^{\mathcal{N}_h}$ are the nodes of the computational mesh. A further simplification is obtained by considering a Karhunen-Loève expansion of the random field, which identifies the d most relevant (independent) eigenmodes $\xi_1, \xi_2, \ldots, \xi_d$ of the covariance operator C corresponding to the largest eigenvalues $\lambda_1(C) \geq \lambda_2(C) \geq \ldots \geq \lambda_d(C)$. The same result can be achieved by computing the POD of a set of random fields generated accordingly to the proposed distribution. The field description then reduces to

$$k(\mathbf{x}, \boldsymbol{\mu}) = 3.5 + \sum_{i=1}^{d} \mu_i \sqrt{\lambda_i} \xi_i \tag{41}$$

where μ_i, $1 \leq i \leq d$, will play the role of identifiable parameters (each one a priori distributed as a standard Gaussian). In this way the sampling is done in a d-dimensional space, with $d \ll \mathcal{N}_h = 1,572$. By taking for the covariance matrix $a = 1$, $b = 0.6$ and $c = 10^{-8}$, we explain about the 90% of the variance by taking $d = 4$ modes (see Fig. 9, top). We also consider the presence of four nuisance parameters (see Fig. 9, bottom), which describe the presence of localized distortions z_i of the parametric field given by

$$\sum_{i=1}^{4} \boldsymbol{\mu}_{4+i} z_i(\mathbf{x}) = \sum_{i=1}^{4} \boldsymbol{\mu}_{4+i} \exp\left(-\frac{(x-x_i)^2 + (y-y_i)^2}{0.025}\right), \tag{42}$$

where $\boldsymbol{x} = [0.25, 0.25, 1.25, 1.25]$ and $\boldsymbol{y} = [0.25, 1.25, 1.25, 0.25]$.

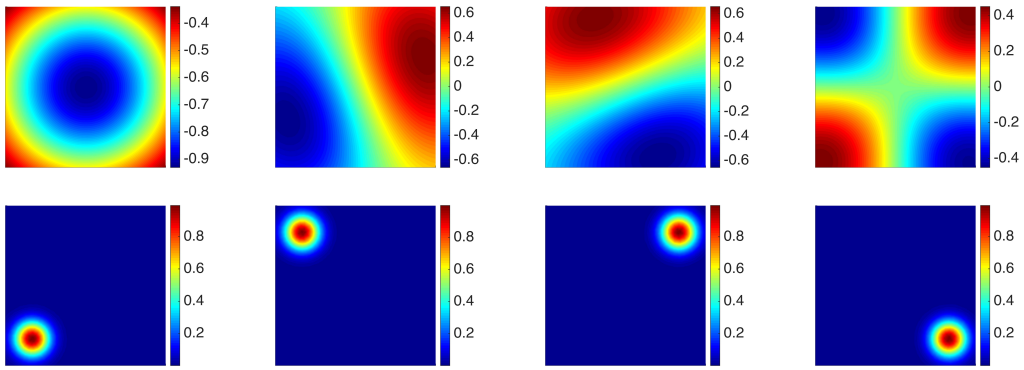

Figure 9: Top: first most relevant modes u_i, $i = 1, \ldots, 4$, of the Karhunen-Loève expansion. Bottom: Four nuisance modes z_i, $i = 1, \ldots, 4$.

In this case, the inverse problem consists in identifying the *posterior* distributions of μ_i, $1 \leq i \leq 4$, by observing the outputs $s^j(\boldsymbol{\mu})$, $j = 1, 2, 3$, in presence of the four nuisance parameters μ_i, $5 \leq i \leq 8$. We consider as a target values $\mathbf{s}^* = \mathbf{s}(\boldsymbol{\mu}^*)$, where $\boldsymbol{\mu}^* = [\boldsymbol{\gamma}^*, \boldsymbol{\zeta}^*]$ and $\boldsymbol{\gamma}^* = (-1, 0.5, -0.6, -0.8)$ whereas $\boldsymbol{\zeta}^*$ is chosen randomly. We build a RB approximation of the state and the dual problems associated to each of the three outputs; $n = 20$ basis functions are selected for each of these four problems. This yields a reduction error on the outputs that has been treated with the proposed REMs, built on a calibration set of $N_{cal} = 100$ values.

We then apply the MCMC procedure and we evaluate the effect of the proposed REMs on the posterior distributions. Similar conclusions to those reported in Sect. 7 can be drawn also for this second test case. As shown in Fig. 10, REM-2 is more effective than REM-1; moreover, REM-3 shows to be the most effective model in terms of correction of the RB posterior, possibly due to the increased parametric dimension (see Fig. 11 for the case of the *Bernoulli* sign model; similar results are obtained with the *nearest neighbor* sign model).

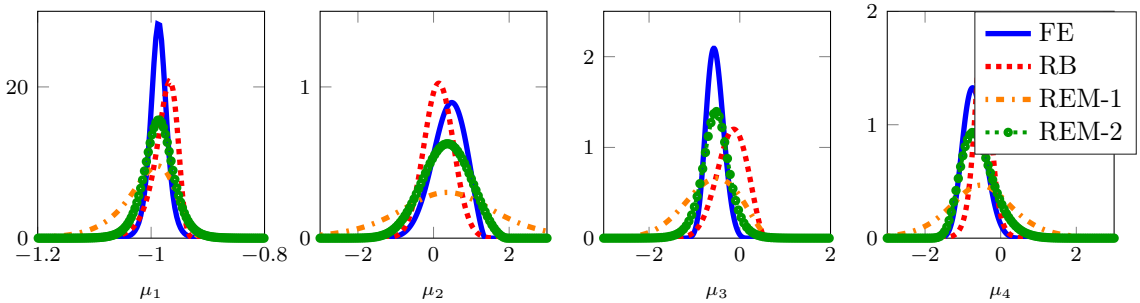

Figure 10: Marginal posterior distribution for the four parameters of interest obtained with the RB method and REM-1, REM-2 corrections. REM-2 yields a better correction (and less skewed distributions) than REM-1.

We then report the MAP estimates for each identifiable parameters and the KL-divergences for each REM in Table 3; the 95% prediction intervals for the MAP estimates are reported in Table 4; these results confirm the efficacy of the deterministic correction provided by REM-2 and REM-3.

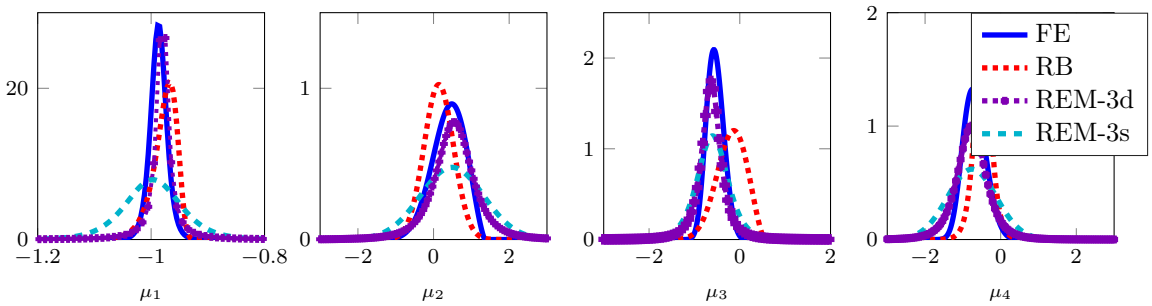

Figure 11: Marginal posterior distribution for the four parameters of interest obtained with the RB method and REM-3 using a Bernoulli sign model. Both the deterministic (d) and the statistical (s) version of the REM yield a very good correction of the posterior distribution.

Moreover, regarding the identification of μ_3 and μ_4 the presence of astatistical correction does not lead to an increasing of KL divergence values.

	μ_1^{MAP}	D_{KL}	μ_2^{MAP}	D_{KL}	μ_3^{MAP}	D_{KL}	μ_4^{MAP}	D_{KL}
FE	-0.987		0.482		-0.571		-0.743	
RB	-0.967	0.327	0.139	0.241	-0.132	0.849	-0.476	0.439
REM-1	-0.988	0.575	0.356	0.713	-0.504	0.722	-0.528	0.657
REM-2	-0.986	0.205	0.3047	0.132	-0.526	0.177	-0.732	0.121
REM-3 (near)	-0.997	0.288	0.584	0.166	-0.657	0.179	-0.831	0.142
REM-3 (near+var)	-0.953	0.530	0.614	0.409	-0.687	0.432	-0.852	0.309
REM-3 (BE)	-0.980	0.221	0.546	0.154	-0.632	0.195	-0.770	0.146
REM-3 (BE+var)	-0.997	0.753	0.488	0.330	-0.601	0.313	-0.760	0.390

Table 3: MAP estimates for the identifiable parameters and KL-divergences from the FE posterior distributions.

	μ_1^{MAP} PI	μ_2^{MAP} PI	μ_3^{MAP} PI	μ_4^{MAP} PI
FE	$(-1.02, -0.95)$	$(-0.52, 1.12)$	$(-0.89, -0.16)$	$(-1.23, -0.07)$
RB	$(-1.032, -0.944)$	$(-0.624, 0.902)$	$(-0.864, 0.361)$	$(-0.967, 0.015)$
REM-1	$(-1.130, -0.938)$	$(-2.675, 3.000)$	$(-1.902, 0.335)$	$(-2.488, 1.433)$
REM-2	$(-1.045, -0.928)$	$(-1.392, 1.707)$	$(-1.182, 0.129)$	$(-1.321, 0.458)$
REM-3 (near)	$(-1.070, -0.924)$	$(-0.906, 2.135)$	$(-1.354, -0.021)$	$-1.8094, 0.103$
REM-3 (near+var)	$(-1.017, -0.890)$	$(-1.203, 2.371)$	$(-1.433, 0.119)$	$(-2.059, 0.396)$
REM-3 (BE)	$(-1.067, -0.892)$	$(-0.930, 2.021)$	$(-1.395, 0.133)$	$(-1.806, 0.266)$
REM-3 (BE+var)	$(-1.113, -0.881)$	$(-1.151, 2.127)$	$(-1.403, 0.200)$	$(-2.020, 0.502)$

Table 4: 95% prediction intervals for each MAP estimator for different REMs.

Finally, we evaluate the diffusivity field (41) for the different MAP estimates, recovered a posteriori by considering different REMs (see Fig. 12). In order to quantify the difference between the *target* field and the fields obtained through the inversion procedure, we compute the $L^2(\Omega)$-norm of their difference (see Table 5). As expected, the most distant field is the one obtained with the RB method without corrections, while the presence of the REMs improves substantially the reconstruction of the field; the best performance in term of parameters identification is provided by REM-3 based on a sign correction strategy.

	RB	REM-1	REM-2	REM-3(near)	REM-3(Be)
$\|\cdot\|_{L^2(\Omega)}$	0.328	0.146	0.099	0.060	0.045

Table 5: $L^2(\Omega)$-norm of the distance between the reconstructed field for different REMs and the *target* one.

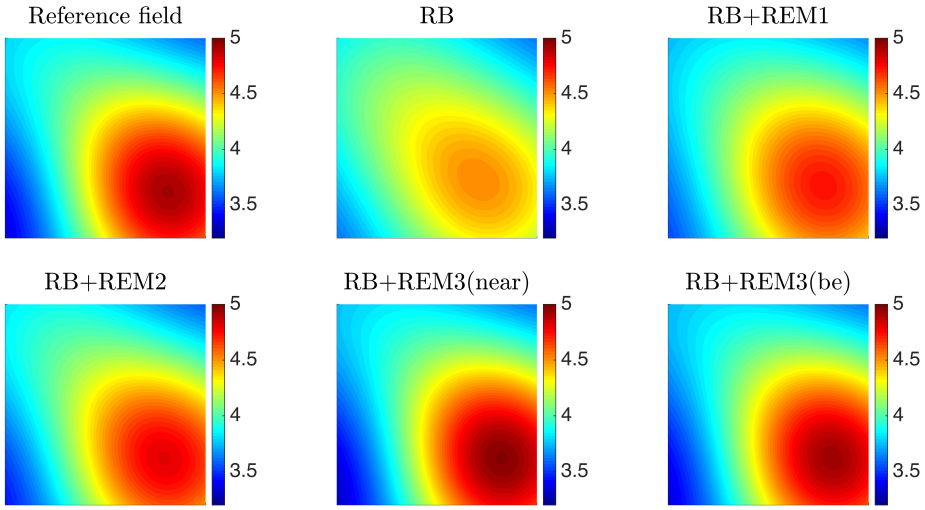

Figure 12: Diffusivity field reconstructed by means of the different REMs (parameters values identified with maximum a posteriori estimators).

Approximation data	test 1	test 2	Performances	test 1	test 2
Number of FE dofs $\mathcal{N}_h$	$1,056$	$1,572$	burn-in M	500	1000
Number of RB dofs n	10	20	thinning d	50	100
Dofs reduction	101:1	79:1	ROM solution	$2 \cdot 10^{-4}$ s	$3 \cdot 10^{-4}$ s
Number of parameters	3	8	ROM error bound	$4 \cdot 10^{-4}$ s	$8.2 \cdot 10^{-4}$ s
FE solution	$4 \cdot 10^{-2}$ s	$1.3 \cdot 10^{-1}$ s	REM-1	≈ 0 s	≈ 0 s
Offline: basis computation	134 s	776 s	REM-2	$7.5 \cdot 10^{-4}$ s	$1.32 \cdot 10^{-3}$ s
Offline: REM calibration	13 s	24 s	REM-3 (near)	$7 \cdot 10^{-4}$ s	$1.12 \cdot 10^{-3}$ s
MCMC iterations	$5 \cdot 10^5$	10^6	REM-3 (Be)	$4 \cdot 10^{-4}$ s	$8.2 \cdot 10^{-4}$ s

Table 6: Computational performances of the proposed framework for both test cases

8 Conclusions

The combined RB/REM methodology proposed in this paper shows how to speed up the solution of an inverse UQ problem dealing with PDEs without affecting the accuracy of the posterior estimates. Provided the RB dimension n is sufficiently small, inexpensive online evaluations can be performed during the MCMC algorithm; this latter takes e.g. less than 1h in the first test case we considered, compared to more than 6h when relying on the full-order model. On the other hand, the REM calibration – which has a negligible impact on the *offline* phase – is instrumental to avoid bias in the computed posterior distributions. Indeed, provided that the REM is trained on a sufficiently large calibration set, the small RB dimension can be compensated with a suitable REM strategy, like the proposed REM-2 or REM-3 approaches. This latter turns out to be a promising option also in view of higher-dimensional problems.

The general trend arising from all our numerical experiments is therefore that REM-1 may be treated as a catch-all method that can be used in the most general case without access to any a posteriori error estimators, but which relies heavily on the assumption of Gaussianity and may not give good results when this assumption is strongly violated. On the other hand, the REM-2 makes no a priori structural assumptions (except smoothness of the inverse effectivity function) and provides excellent accuracy on low-dimensional problems, but relies on multivariate interpolation that does not scale well to high-dimensional problems. The REM-3 produces less accurate reduction error approximations, but only uses regression and so it is more robust in terms of parametric dimension; this method too can exploit existing a posteriori error estimators in correcting the posterior distributions.

References

[1] S. R. Arridge, J. P. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, and V. Vauhkonen, Approximation errors and model reduction with an application in optical diffusion tomography, *Inverse Problems*, 22(1):175–195, 2006.

[2] I. Babuška, F. Nobile, and R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, *SIAM Review*, 52(2):317–355, 2010.

[3] H. T. Banks, M. L. Joyner, B. Wincheski, and W. P. Winfree, Nondestructive evaluation using a reduced-order computational methodology, *Inverse Problems*, 16(4):929, 2000.

[4] G. Berkooz, P. Holmes, J. L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, *Annu. Rev. Fluid Mech.*, 25(1):539–575, 1993.

[5] H. J. Bungartz, M. Griebel, Sparse grids, *Acta Numerica*, **13**: 147–269, 2004.

[6] L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, B. van Bloemen Waanders, and K. Willcox (Eds.), *Large-Scale Inverse Problems and Quantification of Uncertainty*, pages 123–149. John Wiley & Sons, Ltd, 2010.

[7] J. Biehler, M. W. Gee, and W. A. Wall, Towards efficient uncertainty quantification in complex and large-scale biomechanical problems based on a Bayesian multi-fidelity scheme, *Biomech. Model. Mechanobiol.*, 14(3): 489–513, 2015.

[8] A. Birolleau, G. Poëtte, and D. Lucor, Adaptive Bayesian inference for discontinuous inverse problems, application to hyperbolic conservation laws, *Commun. Comput. Phys.*, 16(1):1–34, 2014.

[9] M. D. Buhmann, Radial basis functions, *Acta Numerica*, 9: 1–38, 2000.

[10] T. Bui-Thanh, and O. Ghattas, An analysis of infinite dimensional Bayesian inverse shape acoustic scattering and its numerical approximation, *SIAM J. Uncert. Quant.*, submitted, 2012.

[11] M. Chevreuil, and A. Nouy, Model order reduction based on proper generalized decomposition for the propagation of uncertainties in structural dynamics, *Int. J. Numer. Meth. Engng.*, 89(2):241–268, 2012.

[12] T. Cui, Y. M. Marzouk, and K. Willcox, Data-Driven Model Reduction for the Bayesian Solution of Inverse Problems, *arXiv:1403.4290 e-prints*, 2014.

[13] M. Drohmann, and K. C. Carlberg, The ROMES method for statistical modeling of reduced-order-model error, *SIAM/ASA J. Uncer. Quantif.*, 3(1):116–145, 2015.

[14] D. Galbally, K. Fidkowski, K. Willcox, and O. Ghattas, Nonlinear model reduction for uncertainty quantification in large-scale inverse problems, *Int. J. Numer. Methods Engrg*, 81(12):1581–1608, 2010.

[15] R. G. Ghanem, and P. D. Spanos, Stochastic finite elements: a spectral approach, *Springer*, 1991.

[16] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Markov Chain Monte Carlo in practice, *CRC press*, volume 2., 1995.

[17] J. C. Helton, J. D. Johnson, C. J. Sallaberry, and C. B. Storlie, Survey of sampling-based methods for uncertainty and sensitivity analysis, *Reliab. Eng. Syst. Saf.*, 91(10):1175–1209, 2006.

[18] J. Jakeman, M. Eldred, and D. Xiu, Numerical approach for quantification of epistemic uncertainty, *J. Comput. Phys.,* 229:4648–4663, 2010.

[19] J. Kaipio, and E. Somersalo, Statistical and computational inverse problems, *Springer Science+Business Media, Inc.*, 2005.

[20] T. Lassila, A. Manzoni, A. Quarteroni, and G. Rozza, Generalized reduced basis methods and n-width estimates for the approximation of the solution manifold of parametric PDEs. *Bollettino UMI*, Serie IX, Vol. VI (1):113–135, 2013.

[21] T. Lassila, A. Manzoni, A. Quarteroni, and G. Rozza, A reduced computational and geometrical framework for inverse problems in haemodynamics, *Int. J. Numer. Methods Biomed. Engng.*, 29(7):741–776, 2013.

[22] T. Lassila, A. Manzoni, A. Quarteroni, and G. Rozza, Model order reduction in fluid dynamics: challenges and perspectives. In A. Quarteroni and G. Rozza, editors, *Reduced Order Methods for Modeling and Computational Reduction*, volume 9, pages 235–274. *Springer, MS&A Series*, 2014.

[23] Y. Li, and Y. M. Marzouk Adaptive construction of surrogates for the Bayesian solution of inverse problems, *arXiv:1309.5524 e-prints*, 2013.

[24] J. S. Liu Monte Carlo strategies in Scientific Computing, *Springer*, New York, 2001.

[25] C. Lieberman, K. Willcox, and O. Ghattas, Parameter and state model reduction for large-scale statistical inverse problems, *SIAM J. Sci. Comput.*, 32(5):2523–2542, 2010.

[26] A. Lipponen, A. Seppänen ,and J. P. Kaipio, Reduced-order model for electrical impedance tomography based on proper orthogonal decomposition, *arXiv:1207.0914 e-prints*, 2012.

[27] A. Manzoni, and F. Negri, Rigorous and heuristic strategies for the approximation of stability factors in nonlinear parametrized PDEs, *Adv. Comput. Math.*, 41(5):1255–1288, 2015.

[28] J. Martin, L. C. Wilcox, C. Burstedde, and O. Ghattas, A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion, *SIAM J. Sci. Comput.*, 34(3):A1460–A1487, 2012.

[29] Y. M. Marzouk, and H. N. Najm, Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems, *J. Comput. Phys.*, 228:1862–1902, 2009.

[30] Y. M. Marzouk, and D. Xiu, A Stochastic Collocation Approach to Bayesian Inference in Inverse Problems, *Commun. Comput. Phys.*, 6(4):826–847, 2009.

[31] L. W. T. Ng, and M. S. Eldred, Multifidelity Uncertainty Quantification Using Nonintrusive Polynomial Chaos and Stochastic Collocation, *paper AIAA-2012-1852 in Proceedings of the 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, 2012.

[32] A. Nissinen, V. Kolehmainen, and J. P. Kaipio, Reconstruction of domain boundary and conductivity in electrical impedance tomography using the approximation error approach, *Int. J. Uncert. Quant.*, 1(3):203–222, 2011.

[33] A.Paul-Dubois-Taine, and D. Amsallem, An adaptive and efficient greedy procedure for the optimal training of parametric reducedâĂŘorder models, *Int. J. Numer. Meth. Engng.*, 102(5):1262–1292, 2015.

[34] N. Petra, J. Martin, G. Stadler, and O. Ghattas, A computational framework for infinite-dimensional Bayesian inverse problems: Part II. Stochastic Newton MCMC with application to ice sheet flow inverse problems, *arXiv:1308.6221 e-prints*, 2013.

[35] C. Prud'homme, D. V. Rovas, K. Veroy, L. Machiels, Y. Maday, A. T. Patera, G. Turinici, Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods, *J. Fluid Eng.* 124(1): 70–80, 2002.

[36] A. Quarteroni, Numerical Models for Differential Problems, volume 8, *Springer, MS&A Series*, 2014.

[37] A. Quarteroni, A. Manzoni, and F. Negri, Reduced Basis Methods for Partial Differential Equations. An Introduction, volume 92, *Springer, Unitext Series*, 2016.

[38] A. Quarteroni, G. Rozza, and A. Manzoni, Certified Reduced Basis Approximation for Parametrized Partial Differential Equations in Industrial Applications, *J. Math. Ind.*, 3(1), 2011.

[39] O. Roderick, M. Anitescu, and Y. Peet, Proper orthogonal decompositions in multifidelity uncertainty quantification of complex simulation models, *Int. J. Comput. Math.*, 91(4):748–769, 2014.

[40] C. Robert and G. Casella, Monte Carlo Statisical Methods, *Springer*, 2004.

[41] G. Rozza, D. B. P. Huynh, and A. T. Patera, Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations, *Arch. Comput. Methods Engrg.*,15:229–275, 2008.

[42] Y. Serinagaoglu, D. H. Brooks, and R. S. MacLeod, Improved performance of Bayesian solutions for inverse electrocardiography using multiple information sources, *IEEE Trans. Biomed. Engng.*, 53(10):2024–2034, 2006.

[43] L. Sirovich, Turbulence and the dynamics of coherent structures, part i: Coherent structures, *Quart. Appl. Math.*, 45(3):561–571, 1987.

[44] A. M. Stuart, Inverse problems: a Bayesian Perspective, *Acta Numer.*, 19(1):451–559, 2010.

[45] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, *Society for Industrial and Applied Mathematics*, 2004.

[46] K. Veroy, C. Prud'homme, D. V. Rovas, A. T. Patera, A posteriori error bounds for reduced basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations. In: *Proceedings of the 16th AIAA Computational Fluid Dynamics Conference (2003)*. Paper 2003-3847.

[47] D. Xiu, and J. S. Hesthaven, High-order collocation methods for differential equations with random inputs, *SIAM J. Sci. Comput.*, 27(3):1118–1139, 2005.