Real-time Bayesian data assimilation with data selection, correction of model bias, and on-the-fly uncertainty propagation

Paul-Baptiste Rubio, Ludovic Chamoin, François Louf

Real-time Bayesian data assimilation with data selection, correction of model bias, and on-the-fly uncertainty propagation

Paul-Baptiste Rubio[a], Ludovic Chamoin[a,*], François Louf[a]

[a] *LMT (ENS Paris-Saclay, CNRS, Université Paris-Saclay), 61 Avenue du Président Wilson, 94235 Cachan, France*

Abstract

The work introduces new advanced numerical tools for data assimilation in structural mechanics. Considering the general Bayesian inference context, the proposed approach performs real-time and robust sequential updating of selected parameters of a numerical model from noisy measurements, so that accurate predictions on outputs of interest can be made from the numerical simulator. The approach leans on the joint use of Transport Map sampling and PGD model reduction into the Bayesian framework. In addition, a procedure for the dynamical and data-based correction of model bias during the sequential Bayesian inference is set up, and a procedure based on sensitivity analysis is proposed for the selection of the most relevant data among a large set of data, as encountered for instance with full-field measurements coming from digital image/volume correlation (DIC/DVC) technologies. The performance of the overall numerical strategy is illustrated on a specific example addressing structural integrity on damageable concrete structures, and dealing with the prediction of crack propagation from a damage model and DIC experimental data.

Keywords: Data assimilation, Bayesian inference, Model reduction, Modeling error, Real-time simulations, Full-field measurements, Uncertainty quantification

2010 MSC: 65C20, 74S30, 74S60

1. Introduction

Data assimilation for model updating has become a common practice in structural mechanics activities. It enables to produce relevant numerical simulations that are as close as possible to the reality, and consequently that enable to predict at best the behavior of physical systems of interest. The procedure is usually performed from a set of experimental observations which are assimilated in a single shot, as for the identification of material constitutive laws for instance, but recent applications also lean on the sequential (in time) assimilation of data obtained on-the-fly. In particular, this latter approach is a key concept of Dynamic Data Driven Application Systems (DDDAS) in which a continuous interaction between *in situ* experimental data and simulation tools is implemented for control purposes [6]. In such a context, the numerical model constantly needs to precisely describe the evolving physical system, so that a real-time dynamical updating

*Corresponding author
Email address: ludovic.chamoin@ens-paris-saclay.fr (Ludovic Chamoin)

of model parameters is required inside the retroactive control loop.

It is well-known that the determination of model parameters from indirect and noisy observations usually leads to ill-posed inverse problems. In order to address this issue, and directly take the various uncertainty sources (modeling error, measurement noise,..) into account, a convenient and powerful stochastic framework based on Bayesian inference can be employed [10, 23, 22]. It constitutes a natural regularization procedure in which a probability density, obtained from the Bayes formula, is assigned to the set of parameters to be identified. These parameters are thus considered as random variables and the result of the inference is the associated probability density function (pdf), so that assimilation uncertainties may be further propagated through the model. However, a well-known drawback of Bayesian inference is its computational complexity which makes real-time simulations and sequential assimilation some difficult tasks. Indeed, the inverse approach requires solving the numerical model usually described by PDEs for many combinations of the parameters, which is a costly multi-query process. Furthermore, the posterior pdfs have to be explored to derive useful information such as mean, standard deviation, maximum, or marginals, which requires the computation of multi-dimensional integrals over the parameter space. This is classically performed using Markov Chain Monte-Carlo (MCMC) [10] or Sequential Monte-Carlo [1] methods which are also computationally intensive [17]. Consequently, the use of Bayesian inference remains nowadays intractable in real-time applications such as those envisioned for DDDAS.

In this work, we first present a new formulation of Bayesian inference which is compatible with real-time sequential model updating. It is based on the joint use of two advanced numerical techniques inside the Bayesian framework. First, and as an alternative to Monte-Carlo techniques, Transport Map sampling [7] is introduced to simplify the post-processing of posterior pdfs. It defines a deterministic application between a posterior probability measure to be sampled and a simple reference probability measure (e.g. standard normal distribution) [14, 21], so that all computations (sampling, numerical integration, ...) can be performed in a straightforward manner from this latter measure. The construction of the application is based on a mapping with polynomial structure and results in the solution of a minimization problem. Transport Map sampling goes with convenient sampling error estimates and clear convergence criteria. Furthermore, the natural composition of transport maps is particularly suited to sequential data assimilation.

Second, we resort to model reduction by means of the Proper Generalized Decomposition (PGD) technique [4] to highly lower the computation time. This technique describes the multi-parametric model solution by means of a modal representation with separated variables and explicit dependency on model parameters. Constructed in an offline phase, the PGD solution is then used at two levels in the online phase of the Bayesian data assimilation procedure. On the one hand, due to a straightforward model evaluation for any parameter set, it is employed to compute likelihood functions associated with posterior pdfs [2, 19] and propagate uncertainties on outputs of interest at low cost. On the other hand, the PGD representation yields explicit gradient and Hessian information, which yields a large speed-up in the computation of transport

maps [20].

In order to further increase robustness, we also propose here a procedure to correct model bias during the sequential Bayesian inference. This procedure is based on the introduction of a data-based enrichment term, constructed from a comparison between the predicted model outputs and the actual data along the assimilation process. The model enrichment is defined dynamically and in a stochastic setting; it is propagated in time by means of specific extrapolation procedures. We mention that the proposed approach can be viewed as a stochastic version of some preliminary works conducted in a deterministic context [13, 5].

Eventually, and in order to circumvent technical issues when a large amount of experimental data is available (as in the case of full-field measurements coming from digital image/volume correlation (DIC/DVC) technologies), we address the topic of data selection. A selection of the most relevant data for the model updating purpose is here performed by means of sensitivity analysis, using information directly available in the offline phase.

The overall numerical strategy which is proposed leads to an attractive and very efficient approach to address real-time, robust, and sequential data assimilation. Its performance is here illustrated on a specific application in the context of structural integrity for large-scale structures. The objective in this application is the real-time prediction of crack propagation (e.g. final crack length, or failure probability) in concrete beams by means of a PGD-reduced damage model and experimental data sequentially provided by DIC measurements.

The paper is structured as follows: Bayesian data assimilation and posterior sampling using the Transport Map framework are presented in Section 2; PGD model order reduction and its benefits in the present context are explained in Section 3; the procedure for model bias correction is detailed in Section 4; data selection is addressed in Section 5; numerical results are reported in Section 6; eventually, conclusions and prospects are drawn in Section 7.

2. Transport Map sampling in Bayesian data assimilation

2.1. Basics on Bayesian inference

The purpose of Bayesian inference is to characterize the posterior pdf $\pi(\mathbf{p}|\mathbf{d}^{\text{obs}})$ of a d-dimensional vector of model parameters $\mathbf{p} \in P$ by means of indirect noisy measurements $\mathbf{d}^{\text{obs}}$. In this context, the Bayesian formulation of the inverse problem reads [10]:

$$\pi(\mathbf{p}|\mathbf{d}^{\text{obs}}) = \frac{1}{C}\pi(\mathbf{d}^{\text{obs}}|\mathbf{p}).\pi_0(\mathbf{p}) \tag{1}$$

where $C = \int \pi(\mathbf{d}^{\text{obs}}|\mathbf{p}).\pi_0(\mathbf{p})\mathrm{d}\mathbf{p}$ is a normalization constant, $\pi_0(\mathbf{p})$ is the prior pdf that is related to the *a priori* knowledge on the parameters (i.e. before the assimilation of data $\mathbf{d}^{\text{obs}}$), and $\pi(\mathbf{d}^{\text{obs}}|\mathbf{p})$ is the so-called likelihood function. This latter function corresponds to the probability for the model $\mathcal{M}$ to predict

observations $\mathbf{d}^{\text{obs}}$ given a value of the parameter set $\mathbf{p}$. With the classical assumption of additive measurement noise, with pdf π_{meas}, the likelihood function reads:

$$\pi(\mathbf{d}^{\text{obs}}|\mathbf{p}) = \pi_{\text{meas}}(\mathbf{d}^{\text{obs}} - \mathcal{M}(\mathbf{p})) \tag{2}$$

and requires multiple runs of the model in order to get $\mathcal{M}(\mathbf{p})$.

In the case of sequential assimilation of measurements $\mathbf{d}_i^{\text{obs}}$ at time points t_i, $i \in \{1, ..., N_t\}$, the Bayesian formulation is given by considering the prior at time t_i as the posterior at time t_{i-1}:

$$\pi(\mathbf{p}|\mathbf{d}_1^{\text{obs}}, ..., \mathbf{d}_i^{\text{obs}}) \propto \left(\prod_{j=1}^{i} \pi_{t_j}(\mathbf{d}_j^{\text{obs}}|\mathbf{p}) \right) .\pi_0(\mathbf{p}) \tag{3}$$

with $\pi_{t_j}(\mathbf{d}_j^{\text{obs}}|\mathbf{p}) = \pi_{\text{meas}}\left(\mathbf{d}_j^{\text{obs}} - \mathcal{M}\left(\mathbf{p}, t_j\right)\right)$ for a given set of measurements $\mathbf{d}_j^{\text{obs}}$, and with the same assumption on measurement noise as before. We highlight that in the previous formulations, no assumption is made on the pdfs or on the model.

From the implicit expression of $\pi(\mathbf{p}|\mathbf{d}^{\text{obs}})$ (or $\pi(\mathbf{p}|\mathbf{d}_1^{\text{obs}}, ..., \mathbf{d}_i^{\text{obs}})$), additional exploration with quantities of interest such as means, variances, or first-order marginals, is in practice operated and exploited. These quantities, which are based on posterior pdf sampling and large-dimension integrals, are classically computed using Monte-Carlo integration-based techniques such as Markov Chain Monte-Carlo (MCMC). However, such multi-query procedures are much time consuming and incompatible with real-time computations; we thus propose an alternative approach in the following section.

2.2. Transport Map sampling

2.2.1. Principle

Transport Map sampling originates from pioneering works dealing with optimal transport [24] which were later adapted to Bayesian inference with effective computation tools [21]. The general idea is to build a deterministic mapping M between a reference probability measure ν_ρ (associated with pdf ρ) and a target measure ν_π (associated with pdf π), the purpose being then to perform a change of variables such that:

$$\int g \mathrm{d}\nu_\pi = \int g \circ M \mathrm{d}\nu_\rho \tag{4}$$

In this framework, samples drawn according to the reference pdf ρ can be thus transported by means of the mapping M in order to become samples drawn according to the target pdf π (Figure 1). In the same way, a quadrature rule for ρ can be transported to a quadrature rule for π. For the considered Bayesian inference context, the target pdf π corresponds to the posterior pdf $\pi(\mathbf{p}|\mathbf{d}^{\text{obs}})$, while a standard normal Gaussian pdf is chosen as the reference pdf ρ.

From the reference pdf ρ, the purpose is therefore to build the map $M : \mathbb{R}^d \to \mathbb{R}^d$ such that:

$$\nu_\pi \approx M_\sharp \nu_\rho = \rho \circ M^{-1} |\det \nabla M^{-1}| \tag{5}$$

where $\sharp$ denotes the push forward operator. To quantify the difference between ν_π and $M_\sharp \nu_\rho$, the Kullback-Leibler (K-L) divergence $\mathcal{D}_{KL}$ is introduced:

$$\mathcal{D}_{KL}(M_\sharp \nu_\rho || \nu_\pi) = \mathbb{E}_\rho \left[\ln \frac{\nu_\rho}{M_\sharp^{-1} \nu_\pi} \right] = \int_P \left[\log(\rho(\mathbf{p})) - \log([\pi \circ M](\mathbf{p})) - \log(|\det \nabla M(\mathbf{p})|) \right] \rho(\mathbf{p}) \mathrm{d}\mathbf{p} \tag{6}$$

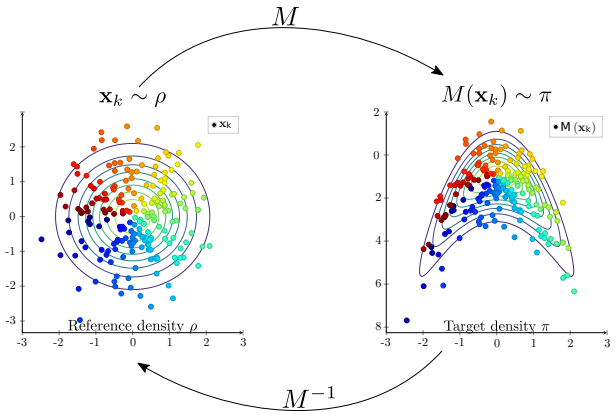


Figure 1: Illustration of the Transport Map principle for sampling a target pdf.

2.2.2. Computation of the map

The map M is searched among Knothe-Rosenblatt rearrangements (i.e lower triangular and monotonic maps). This particular choice of structure is motivated by the properties of unique minimizer of (6), optimality with weighted quadratic cost, and computational feasibility [14, 21]. The map M is thus parameterized as:

$$M(\mathbf{p}) = \begin{bmatrix} M^1(\mathbf{a}_c^1, \mathbf{a}_e^1, p_1) \\ M^2(\mathbf{a}_c^2, \mathbf{a}_e^2, p_1, p_2) \\ \vdots \\ M^d(\mathbf{a}_c^d, \mathbf{a}_e^d, p_1, p_2, ..., p_d) \end{bmatrix} \tag{7}$$

with $M^k(\mathbf{a}_c^k, \mathbf{a}_e^k, \mathbf{p}) = \Phi_c(\mathbf{p})\mathbf{a}_c^k + \int_0^{p_k} (\Phi_e(p_1, ..., p_{k-1}, \theta)\mathbf{a}_e^k)^2 \mathrm{d}\theta$. Functions Φ_c and Φ_e are Hermite polynomials associated with coefficients $\mathbf{a}_c$ and $\mathbf{a}_e$ to be set. With such a parameterization, the optimal map M is obtained by minimizing the K-L divergence (6). Using a specific quadrature rule $(\omega_i, \mathbf{p}_i)_{i=1}^N$ for the pdf ρ (Gaussian quadrature usually), the associated minimization problem reads:

$$\min_{\mathbf{a}_c^{1,...,d}, \mathbf{a}_e^{1,...,d}} \sum_{i=1}^N \omega_i \left[-\log(\bar{\pi} \circ M(\mathbf{a}_c^{1,...,d}, \mathbf{a}_e^{1,...,d}, \mathbf{p}_i) - \log(|\det \nabla M(\mathbf{a}_c^{1,...,d}, \mathbf{a}_e^{1,...,d}, \mathbf{p}_i))|) \right] \tag{8}$$

where $\bar{\pi}$ is the non-normalized version of the target pdf. This problem is fully deterministic and can be solved using classical algorithms (such as BFGS) with the computation of derivatives (gradient, Hessian) of $\bar{\pi}(\mathbf{p})$. Once the map M is obtained, the quality of the approximation $M_\sharp \nu_\rho$ of the measure ν_π can be assessed

through the convergence criterion ϵ_σ (variance diagnostic, see [21]) defined as:

$$\epsilon_\sigma = \frac{1}{2}\mathbb{V}\mathrm{ar}_\rho\left[\ln\frac{\nu_\rho}{M_\sharp^{-1}\nu_\pi}\right] \tag{9}$$

Therefore, an adaptive strategy regarding the order of the map can be derived from the criterion ϵ_σ in order to reach a prescribed error tolerance. It is fruitful to notice that the numerical cost associated with the computation of ϵ_σ is very low as integration is performed on the reference pdf ρ and with the same quadrature rule as the one used in the computation of the K-L divergence.

2.2.3. Sequential data assimilation with transport maps

In the case of sequential inference (assimilation of new data $\mathbf{d}_i^{\text{obs}}$ at each time point t_i), Transport Map sampling beneficially exploits the Markov structure of the posterior pdf (3) to yield a computation time which is almost constant for each assimilation procedure. Indeed, instead of being fully computed, the map between the reference pdf ρ and the posterior pdf at time t_i is obtained by composition of low-order maps:

$$(M_1 \circ .... \circ M_i)_\sharp\,\rho(\mathbf{p}) = (M_i^c)_\sharp\,\rho(\mathbf{p}) \approx \pi(\mathbf{p}|\mathbf{d}_1^{\text{obs}}, ..., \mathbf{d}_i^{\text{obs}}) \tag{10}$$

The map M_1 represents the coupling between the pdf $\rho(\mathbf{p})$ and the first posterior pdf $\pi(\mathbf{p}|\mathbf{d}_1^{\text{obs}}) \propto \pi_{t_1}(\mathbf{d}_1^{\text{obs}}|\mathbf{p}).\pi(\mathbf{p})$. Then, each map M_i, $i \in \{2, ..., N_t\}$, is computed between ρ and the pdf π_i^* defined as:

$$\pi_i^*(\mathbf{p}) = \pi_{t_i}(\mathbf{d}_i^{\text{obs}}|M_{i-1}^c(\mathbf{p})).\rho(\mathbf{p}) \tag{11}$$

and corresponding to a posterior pdf affected by an inverse transformation using maps already computed for all previous assimilation steps.

The schematic principle of the sequential computation of transport maps is given in Figure 2. We mention that for the assimilation of the first data set $\mathbf{d}_1^{\text{obs}}$, an additional linear map L is computed (Laplace approximation). This linear transformation step, which acts as a normalization of the parameter space, enables to build an intermediate pdf that is closer to the standard normal reference pdf (in the sense that it has approximatively zero mean and identity covariance matrix) which improves convergence in the computation of transport maps M_i.

3. Use of PGD model order reduction

3.1. Basics on PGD

Due to the increasing number of high-dimensional approximation problems in engineering activities, model reduction techniques have been the object of a growing interest in research and industry to keep such problems tractable. Here we deal with model reduction techniques that use low-rank tensor formats, and which are prominent tools for the effective computation and representation of high-dimensional solutions. We focus on an appealing technique based on low-rank canonical format and referred to as Proper Generalized Decomposition (PGD), which was introduced and successfully used in many applications of computational

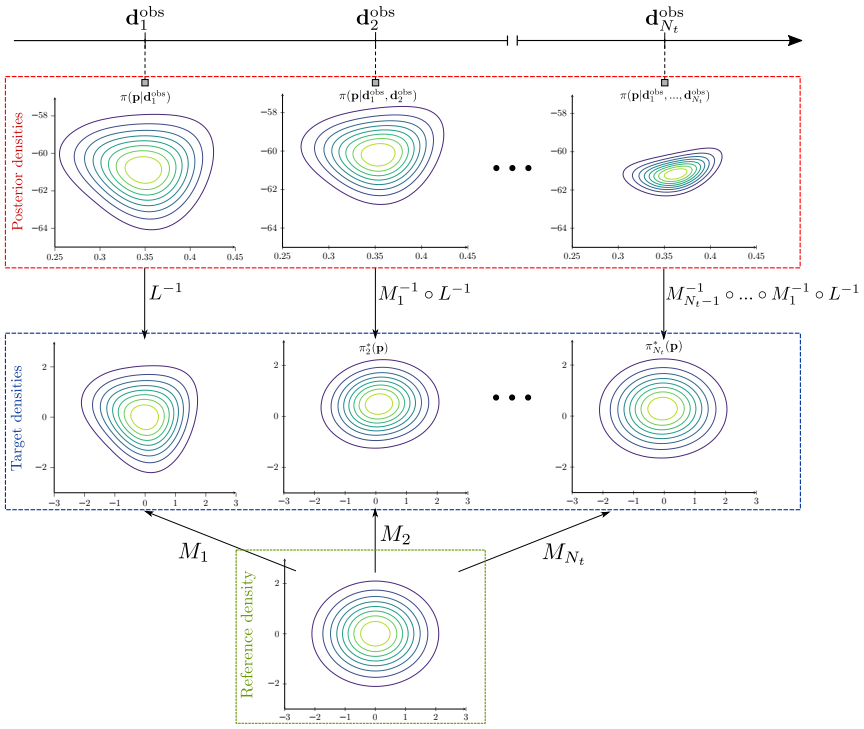


Figure 2: Schematic principle of sequential inference with transport maps.

mechanics [4]. Contrary to POD, the PGD approximation does not require any knowledge on the solution. It operates in an iterative strategy in which basis functions (or modes) are computed on the fly, by solving eigenvalue problems.

In the classical PGD framework, the reduced model is directly built in an offline phase from the global weak formulation of the problem of interest. It leads to an approximate solution $\mathbf{u}_m$ (PGD representation at order m) of the exact model solution $\mathbf{u}$, with a separated form [4]:

$$\mathbf{u}_m(\mathbf{x}, t, \mathbf{p}) = \sum_{k=1}^{m} \mathbf{\Lambda}_k(\mathbf{x}) \lambda_k(t) \prod_{i=1}^{d} \alpha_k^i(p_i) \tag{12}$$

where $\mathbf{x}$, t, and $\mathbf{p}$ are space, time, and parameter coordinates, respectively. This solution can then be easily evaluated online, computing products and sums of mono-parameter functions.

3.2. PGD in Bayesian inference

In the formulation (3), the posterior pdf can be explicitly expressed as a function of parameters $\mathbf{p}$ only when the model is also explicit with respect to $\mathbf{p}$. However, in most engineering applications, the model is derived from some Partial Differential Equations (PDEs) with implicit dependency on parameters $\mathbf{p}$. This usually requires the full computation of the model solution for a many values of the parameter set $\mathbf{p}$, which is a costly procedure not compatible with real-time constraints.

In order to circumvent this issue, the PGD technique can be advantageously employed. Indeed, once the PGD approximation $\mathbf{u}_m(\mathbf{x}, t, \mathbf{p})$ is built, an explicit formulation of the non-normalized posterior pdf can be

derived. Owing to the observation operator $\mathcal{O}$, the set of observables $\mathbf{d}(\mathbf{p},t) = \mathcal{O}\left(\mathbf{u}_m(\mathbf{x},t,\mathbf{p})\right)$ is extracted from the PGD field $\mathbf{u}_m(\mathbf{x},t,\mathbf{p})$, with explicit dependency on $\mathbf{p}$. Therefore, the non-normalized posterior pdf $\bar{\pi}$ reads (considering sequential data assimilation here):

$$\bar{\pi}\left(\mathbf{p}|\mathbf{d}_1^{\text{obs}},...,\mathbf{d}_i^{\text{obs}}\right) = \prod_{j=1}^{i} \pi_{\text{meas}}\left(\mathbf{d}_j^{\text{obs}} - \mathbf{d}\left(\mathbf{p},t_j\right)\right).\pi_0(\mathbf{p}) \tag{13}$$

which leads to cost-effective evaluations of this latter pdf from sampling methods [2, 19].

On the other hand, having available a PGD solution is an interesting asset to perform uncertainty quantification and propagation. Considering a quantity of interest (QoI) q defined from the operator $\mathcal{Q}$ applied on the model solution $\mathbf{u}$, an approximation of q with explicit dependency on $\mathbf{p}$ can be obtained by applying the operator $\mathcal{Q}$ to the PGD solution:

$$q(\mathbf{p}) \approx \mathcal{Q}\left(\mathbf{u}_m(\mathbf{x},t,\mathbf{p})\right) \tag{14}$$

Once measurements have been assimilated with Bayesian inference, parameter samples $\mathbf{p}_k$ ($k \in \{1,...,K\}$) can be drawn from the posterior pdf, and associated QoI samples q_k ($k \in \{1,...,K\}$) can be computed in a straightforward manner as $q_k = q(\mathbf{p}_k)$. Therefore, the pdf of q can be easily obtained (using for instance a kernel density estimation (KDE) from samples q_k) and post-processed for design or control purposes.

3.3. Transport Map sampling with PGD models

An additional benefit of PGD model reduction can be found in the sampling procedure using transport maps and described in Section 2. We remind (see Section 2.2) that the computation of transport maps leans on the minimization of the functional defined in (8). With the PGD formulation, partial derivatives of the model solution with respect to parameters $\mathbf{p}$ can be directly and explicitly recovered as:

$$\frac{\partial^n \mathbf{u}_m}{\partial p_j^n}(\mathbf{x},t,\mathbf{p}) = \sum_{k=1}^{m} \mathbf{\Lambda}_k(\mathbf{x})\lambda_k(t)\frac{\partial^n \alpha_k^j}{\partial p_j^n}(p_j)\prod_{\substack{i=1\\ i\neq j}}^{d}\alpha_k^i(p_i) \tag{15}$$

then stored in the offline phase. Parameter modes α_k^i being usually described by means of finite element functions, the derivations are thus performed on one-dimensional shape functions. Furthermore, owing to the PGD representation with separated variables, cross-derivatives can be computed from combinations of univariate mode derivatives.

As a result, first- and second-order derivatives of the pdf $\bar{\pi}$ are easily accessible which enables the effective (fast convergence), and at low cost, solution of the minimization problem (8) by means of dedicated algorithms using gradient or Hessian information (BFGS, trust-region, ...). This represents a much better option for the computation of transport maps compared to concurrent approaches that use function information alone to solve (8) (as in the simplex method).

4. Correction of model bias

4.1. General idea

In the classical Bayesian inference formulation, the forward model $\mathcal{M}$ that is evaluated in the likelihood function is supposed to be exact in the sense that, once model parameters are updated, numerical outputs should be consistent (no bias) with experimental observations. However, engineering applications are very often associated with modeling bias due to the difficulty in representing the complexity of physical phenomena of interest. In the present context of stochastic data assimilation, the inaccuracy of the model can have a major impact on the quality of the posterior pdf (e.g. the posterior mean may diverge in the case of sequential data assimilation). In order to keep a robust approach, it is thus of prime importance to address this issue. Some recent works proposed first model correction strategies, as in [3] where modeling error computed between low- and high-fidelity models (and thus known *a priori*) is introduced to improve the Bayesian formulation. Nevertheless, quantifying modeling error or assessing it by means of error estimates requires the definition of a high-fidelity model, which is often unreachable or unknown.

Here, we propose an alternative strategy in which the corrective modeling error term that is integrated in the Bayesian formulation is directly defined from the data. This data-based strategy shares similarities with those developed in the PBDW method [13] or in the definition of hybrid twins [5], even though it is here implemented in a stochastic framework.

We introduce the stochastic estimation residual $\mathbf{B}$, defined at the assimilation time point t_i as:

$$\mathbf{B}(\mathbf{x}^{\text{obs}}, t_i) = \mathbf{d}_i^{\text{obs}} - \mathbf{e}_{\text{meas}} - \mathcal{M}(\mathbf{x}^{\text{obs}}, t_i, \mathbf{p}) \tag{16}$$

$\mathbf{x}^{\text{obs}}$ represents the vector of spatial coordinates of the measurement points, and $\mathbf{e}_{\text{meas}}$ is the additive measurement noise. This residual, which represents the discrepancy between data and model, is fully computable after data assimilation at time t_i. Its pdf is obtained from a straightforward propagation of uncertainties on $\mathbf{e}_{\text{meas}}$ and $\mathbf{p}$.

In the context of sequential Bayesian inference, we decide to take into account at time t_{i+1} information obtained at time t_i on the discrepancy between model and data, by considering the corrected model output $\mathcal{M}^{\text{corr}}$ defined as:

$$\mathcal{M}^{\text{corr}}(\mathbf{x}^{\text{obs}}, \mathbf{p}, t_{i+1}) = \mathcal{M}(\mathbf{x}^{\text{obs}}, \mathbf{p}, t_{i+1}) + \hat{\mathbf{B}}_{i \to i+1}(\mathbf{x}^{\text{obs}}) \tag{17}$$

where the model bias term $\hat{\mathbf{B}}_{i \to i+1}(\mathbf{x}^{\text{obs}})$ is taken as the extrapolation at time t_{i+1} of the estimation residual $\mathbf{B}(\mathbf{x}^{\text{obs}}, t_i)$ computed at time t_i. Therefore, the likelihood function at time t_{i+1} reads:

$$\pi(\mathbf{d}_{i+1}^{\text{obs}} | \mathbf{p}) = \pi_{\hat{B}}(\mathbf{d}_{i+1}^{\text{obs}} - \mathcal{M}(\mathbf{x}^{\text{obs}}, \mathbf{p}, t_{i+1})) \tag{18}$$

where $\pi_{\hat{B}}$ is the pdf associated with the extrapolated model bias $\hat{\mathbf{B}}_{i \to i+1}$.

4.2. Extrapolation procedure

Several techniques may be used to extrapolate the model bias over an assimilation time step. A simple one would be to linearly extrapolate the mean and standard deviation of the residual vector $\mathbf{B}(\mathbf{x}^{\text{obs}}, t_i)$, independently at each observation point of $\mathbf{x}^{\text{obs}}$, and approximate the pdf $\pi_{\hat{B}}$ by a set of individual Gaussian pdfs. However, a drawback of this procedure is that it relies on data alone without any physical consideration, so that it may lead to inconsistent results. We prefer here using a more global extrapolation technique that also involves the model being supposed to provide an informative representation of the studied physics, with limited model bias. This way, physics is taken into account and measurement noise is filtered by the model. We first introduce the matrices of mean and standard deviation of the model bias, defined at assimilation time t_i as:

$$\begin{aligned} \mathbb{B}_{\text{mean}} &= \left[\text{mean}\left(\mathbf{B}(\mathbf{x}^{\text{obs}}, t_1)\right), ..., \text{mean}\left(\mathbf{B}(\mathbf{x}^{\text{obs}}, t_i)\right)\right] \\ \mathbb{B}_{\text{std}} &= \left[\text{std}\left(\mathbf{B}(\mathbf{x}^{\text{obs}}, t_1)\right), ..., \text{std}\left(\mathbf{B}(\mathbf{x}^{\text{obs}}, t_i)\right)\right] \end{aligned} \tag{19}$$

with mean(.) and std(.) the operators that compute the mean and standard deviation of a random variable using a quadrature rule. These matrices are thus constructed from the collection of mean and standard deviation vectors of $\mathbf{B}(\mathbf{x}^{\text{obs}}, t_j)$, defined in (16) and which have the size of $\mathbf{x}^{\text{obs}}$ (potentially large, as for full-field measurements obtained with DIC/DVC). The extrapolation is then performed from space and time SVD modes of $\mathbb{B}_{\text{mean}}$ and $\mathbb{B}_{\text{std}}$. In other terms, the SVD decompositions such that:

$$\mathbb{B}_{\text{mean}} = \mathbb{U}_{\text{mean}}\mathbb{D}_{\text{mean}}\mathbb{V}_{\text{mean}}^T \quad ; \quad \mathbb{B}_{\text{std}} = \mathbb{U}_{\text{std}}\mathbb{D}_{\text{std}}\mathbb{V}_{\text{std}}^T \tag{20}$$

are computed. Vectors $\mathbf{U}_{\text{mean}}$ and $\mathbf{U}_{\text{std}}$ (resp. $\mathbf{V}_{\text{mean}}$ and $\mathbf{V}_{\text{std}}$) that compose $\mathbb{U}_{\text{mean}}$ and $\mathbb{U}_{\text{std}}$ (resp. $\mathbb{V}_{\text{mean}}$ and $\mathbb{V}_{\text{std}}$) represent the space (resp. time) dependency of the model bias. Then, the mean and standard deviation of the extrapolated vector $\hat{\mathbf{B}}_{i\to i+1}(\mathbf{x}^{\text{obs}})$ at time t_{i+1} are computed using linear extrapolations $\hat{\mathbf{V}}_{\text{mean}}$ and $\hat{\mathbf{V}}_{\text{std}}$ of SVD vectors $\mathbf{V}_{\text{mean}}$ and $\mathbf{V}_{\text{std}}$, respectively. Eventually, the pdf of $\hat{\mathbf{B}}_{i\to i+1}(\mathbf{x}^{\text{obs}})$ is recovered as a multidimensional Gaussian distribution.

We mention that truncation is usually applied in this extrapolation procedure, so that only the first SVD modes are in practice used. Furthermore, in order to avoid the increase of computation and storage costs along the sequential data assimilation, and thus keep a numerical strategy which remains compatible with the real-time simulation constraint, the so-called Sequential Karhunen-Loeve (SKL) algorithm [18] is used. It enables the fast SVD decomposition of a matrix $[\mathbb{M}_{[1,i-1]}\ \mathbb{M}_i]$ obtained from the horizontal concatenation of matrices $\mathbb{M}_{[1,i-1]}$ and $\mathbb{M}_i$, knowing the SVD decomposition of $\mathbb{M}_{[1,i-1]}$. The algorithm is based on a QR decomposition in order to retrieve the span of the SVD modes already computed. In the present case, $\mathbb{M}_{[1,i-1]}$ corresponds to the collection of means or standard deviations of the model bias computed until time t_{i-1}, while $\mathbb{M}_i$ corresponds to the mean or standard deviation of the model bias computed at time t_i.

4.3. Detection of a model bias with model evidence

The normalization constant $C = \int \pi(\mathbf{d}^{\text{obs}}|\mathbf{p}).\pi_0(\mathbf{p})\text{d}\mathbf{p} = \pi(\mathbf{d}^{\text{obs}})$ involved in (1) is also an indicator, denoted model evidence, on the quality of the model. It can be used in practice for model selection in order

to determine which model among a class of models is the most likely to have generated the data. Indeed, considering N models $\{\mathcal{M}_1, ..., \mathcal{M}_N\}$ with each model $\mathcal{M}_k$ depending on the parameter set $\mathbf{p}_k$, the posterior pdf that $\mathcal{M}_k$ generated the data $\mathbf{d}^{\text{obs}}$ is:

$$\pi(\mathcal{M}_k|\mathbf{d}^{\text{obs}}) = \frac{\pi(\mathbf{d}^{\text{obs}}|\mathcal{M}_k).\pi(\mathcal{M}_k)}{\sum_{k=1}^{N} \pi(\mathbf{d}^{\text{obs}}|\mathcal{M}_k).\pi(\mathcal{M}_k)} \tag{21}$$

where $\pi(\mathcal{M}_k)$ is the prior pdf on the model and $\pi(\mathbf{d}^{\text{obs}}|\mathcal{M}_k)$ is the integrated likelihood function of $\mathcal{M}_k$. This latter pdf is computed by marginalization:

$$\pi(\mathbf{d}^{\text{obs}}|\mathcal{M}_k) = \int \pi(\mathbf{d}^{\text{obs}}, \mathbf{p}_k|\mathcal{M}_k) d\mathbf{p}_k = \int \pi(\mathbf{d}^{\text{obs}}|\mathbf{p}_k, \mathcal{M}_k).\pi(\mathbf{p}_k|\mathcal{M}_k) d\mathbf{p}_k = C_k \tag{22}$$

where C_k is the model evidence associated to the Bayesian inference performed with the model $\mathcal{M}_k$.
Therefore, in the case where the prior pdfs $\pi(\mathcal{M}_k)$, $k \in \{1, ..., N\}$, are equal, the Bayes factor F_{ij} between two models $\mathcal{M}_i$ and $\mathcal{M}_j$ reads:

$$F_{ij} = \frac{\pi(\mathcal{M}_i|\mathbf{d}^{\text{obs}})}{\pi(\mathcal{M}_j|\mathbf{d}^{\text{obs}})} = \frac{C_i}{C_j} \tag{23}$$

and indicates, when $F_{ij} > 1$ (resp. $F_{ij} < 1$), that the model $\mathcal{M}_i$ is more (resp. less) likely than the model $\mathcal{M}_j$ to produce the considered data $\mathbf{d}^{\text{obs}}$. An empirical interpretation of the Bayes factor is given by the Jeffreys scale of evidence [9].

In the present context, the evolution of model evidence C can be monitored along the sequential data assimilation process in order to identify when the model $\mathcal{M}$ becomes less reliable. In the case where the model evidence drops drastically (within a prescribed range), the correction of model bias proposed in Section 4.1 is implemented. Consequently, the model correction procedure is used only when necessary.

The computation of the model evidence is not straightforward in practice. In classical sampling methods that aim at characterizing the posterior pdf, this quantity is indeed not directly available. However, when using Transport Map sampling, the model evidence can be interpreted as the distance between the reference pdf ρ and the posterior pdf pulled-back by the transport map M (see [7]):

$$C = \exp\left(\mathbb{E}_\rho\left[\log\left(T_\sharp^{-1}\pi\right) - \log(\rho)\right]\right) \tag{24}$$

This quantity is easy to compute as a quadrature rule with respect to ρ is supposed to be known.
Another advantage of Transport Map sampling in sequential data assimilation, as detailed in Section 2.2.3, is that if the model is found to be inaccurate in a given time range, the corresponding maps can be removed in order to retrieve the error committed in the estimation of model parameters. Missed measurements can then be re-assimilated once the model bias is corrected.

5. Data selection for effective data assimilation

We consider here the framework of full-field measurements obtained from the Digital Image Correlation (DIC) technique [8]. In such a framework, rich experimental information is available but it may be difficult to

handle and post-process in the context of real-time Bayesian data assimilation. Indeed, as measurements are compared with model outputs, large data sets require more complex computations in the inference procedure. We propose here a numerical strategy that circumvents this difficulty by filtering the large amount of data and keeping the most relevant ones alone.

5.1. DIC principles

The purpose of DIC is to identify a displacement field $\mathbf{u}(\mathbf{x})$ that links two gray level pictures, a reference picture $f(\mathbf{x})$ and a deformed picture $g(\mathbf{x})$, $\mathbf{x}$ representing here the pixel coordinates. Invoking the local gray-level conservation between f and g, and considering a global DIC approach applied over the whole measurement zone (Region Of Interest or ROI) made of a large number N_p of pixels, the field $\mathbf{u}(\mathbf{x})$ is determined by minimizing the following nonlinear correlation residual:

$$\frac{1}{N_p}\sum_{\mathbf{x}\in ROI}\left(f(\mathbf{x})-g(\mathbf{x}+\mathbf{u}(\mathbf{x}))\right)^2 \tag{25}$$

For mechanical regularization purposes, the displacement field is expressed in a given FE basis $\{\psi_n\}$, i.e. $\mathbf{u}(\mathbf{x})=\sum_n \mathbf{u}_n\psi_n(\mathbf{x})=\mathbf{\Psi}^T(\mathbf{x})\mathbf{U}$.
To solve (25), the first-order Taylor expansion $g(\mathbf{x}+\mathbf{u}(\mathbf{x}))\approx g(\mathbf{x})+\mathbf{u}(\mathbf{x}).\nabla f(\mathbf{x})$ is employed, and an iterative construction of $\mathbf{u}(\mathbf{x})$ is performed by minimizing:

$$\sum_{\mathbf{x}\in ROI}\left(f(\mathbf{x})-\tilde{g}(\mathbf{x})-\nabla f\cdot\delta\mathbf{u}^{(k)}(\mathbf{x})\right)^2 \tag{26}$$

where $\tilde{g}(\mathbf{x})=g(\mathbf{x}+\mathbf{u}^{(k)}(\mathbf{x}))$ is the updated deformed image, and $\delta\mathbf{u}^{(k)}=\mathbf{u}^{(k+1)}-\mathbf{u}^{(k)}$ is the correction at iteration k. The minimization of (25) thus comes down to the successive solution of linear systems of the form $\mathbb{M}_{DIC}\delta\mathbf{U}^{(k)}=\mathbf{b}_{DIC}^{(k)}$ with:

$$(\mathbb{M}_{DIC})_{mn}=\sum_{\mathbf{x}\in ROI}\nabla f(\mathbf{x})\psi_m(\mathbf{x})\cdot\nabla f(\mathbf{x})\psi_n(\mathbf{x}) \quad ; \quad (\mathbf{b}_{DIC}^{(k)})_m=\sum_{\mathbf{x}\in ROI}(f(\mathbf{x})-\tilde{g}(\mathbf{x}))\nabla f(\mathbf{x})\psi_m(\mathbf{x}) \tag{27}$$

The vector $\mathbf{b}_{DIC}^{(k)}$ is the residual vector updated at each iteration k, while $\mathbb{M}_{DIC}$ is a symmetric positive matrix which is computed once for all as it only depends on the reference picture f and the chosen basis $\{\psi_n\}$.

Measurement uncertainty is a crucial aspect when dealing with data assimilation, and Bayesian inference supposes that the pdf on measurement error is known. In connection with this point, the previously described global DIC method provides a clear estimation of the measurement noise. Indeed, assuming a white noise e_f with zero mean and variance γ_f^2 affecting both pictures f and g, and noticing that the resulting noise $\mathbf{e}$ on the measured dofs $\mathbf{U}$ reads $\mathbf{e}=\mathbb{M}_{DIC}^{-1}\sum_{\mathbf{x}\in ROI}2e_f\mathbf{\Psi}^T\{\nabla f\}$, it is straightforward that $\text{mean}(\mathbf{e})=\mathbf{0}$ and that the covariance matrix on measurement noise is $\mathbb{C}_{\mathbf{e}}=\text{mean}(\mathbf{e}\otimes\mathbf{e})=2\gamma_f^2\mathbb{M}_{DIC}^{-1}$. As $\mathbb{M}_{DIC}^{-1}$ is not diagonal, the DIC measurement noise is spatially correlated; the DIC matrix $\mathbb{M}_{DIC}$ thus highlights spatial correlations in the measurement noise due to the picture texture and the chosen FE basis.

5.2. Selection of DIC data from sensitivity fields

In the context of model updating, the relevance of data is indicated by their sensitivity with respect to parameters of interest. Considering a model $\mathcal{M}(\mathbf{p})$, the sensitivity of the model output to a given parameter p_j in $\mathbf{p}$ reads $\mathbf{S}_j(\mathbf{x},\mathbf{p}) = \partial\mathbf{u}(\mathbf{x},\mathbf{p})/\partial p_j$ (for time-dependent problems, the mean value in time $t^\diamond$ may be chosen to compute sensitivity quantities). We notice that $\mathbf{S}_j$ can be easily approximated from the computed PGD representation (12), as:

$$\mathbf{S}_j(\mathbf{x},\mathbf{p}) \approx \frac{\partial \mathbf{u}_m(\mathbf{x},t^\diamond,\mathbf{p})}{\partial p_j} = \sum_{k=1}^{m} \mathbf{\Lambda}_k(\mathbf{x})\lambda_k(t^\diamond)\frac{\partial \alpha_k^j(p_j)}{\partial p_j}\prod_{\substack{i=1\\ i\neq j}}^{d}\alpha_k^i(p_i) \tag{28}$$

Cross-sensitivities $\mathbf{S}_{ij}(\mathbf{x},\mathbf{p}) = \partial^2\mathbf{u}(\mathbf{x},\mathbf{p})/\partial p_i \partial p_j$ are calculated the same way.

The sensitivity quantities can thus be computed *a priori*, in the offline phase of data assimilation and once for all, before being used to select the location of the most sensitive measurements for parameter identification. In order to achieve this latter target in the context of DIC, and to take DIC measurement noise into account, it is chosen to weight the sensitivity fields $\mathbf{S}_j$ by the measurement sensitivity (given by the matrix $\mathbb{M}_{DIC}$) which represents the signal-to-noise ratio. We thus define new sensitivity fields $\mathbf{S}_j^{DIC}$ as $\mathbf{S}_j^{\mathrm{DIC}} = \mathbb{M}_{DIC}\mathbf{S}_j$; this way, the measurement uncertainties are propagated to the sensitivity fields. Fields $\mathbf{S}_j^{DIC}$ are in practice computed by projecting fields $\mathbf{S}_j$ on the DIC mesh.

Once fields $\mathbf{S}_j^{DIC}$ are obtained, they are evaluated for a reference value of $\mathbf{p}$, and the selection of most sensitive data is merely performed by considering the nodes on which the field magnitude is the higher. In the case of multi-parameter identification, several options can be chosen such as considering the products of individual sensitivity fields, or the minimal/maximal sensitivity among those obtained with dimensionless parameters.

6. Illustrative application: real-time prediction of structural integrity for concrete structures

In this section, we illustrate the overall data assimilation methodology which was set up in the previous sections. The target application is a real test-case that involves a large-scale concrete beam with initial crack submitted to a bending loading. It is instrumented with DIC measurement devices in order to predict crack propagation, and possibly the collapse of the structure, using real-time Bayesian data assimilation associated with a damage model.

6.1. Experimental campaign

The considered experimental test is a three-point bending test on a concrete specimen (Figure 3). The specimen is a prismatic 840 mm $\times$ 100 mm $\times$ 100 mm beam notched at mid-span on 20% of its height. It is supported on the bottom side by two steel cylinders, steel plates being inserted between the cylinders and the beam to prevent indentation due to stress concentration. A controlled displacement u_{imp} is prescribed on the top side of the specimen by a tensile test machine equipped with an hydraulic actuator. The upper

support of the test machine is fixed, and a load cell measures the reaction force on this support.

After three adjustment cycles which are carried out in order to fill all gaps related to positioning, a monotonic increasing displacement is imposed at the rate of 3×10^{-3} mm/s until the total fracture of the specimen is observed. The associated load/displacement curve, measured by the load cell and a LVTD sensor, is shown in Figure 3.

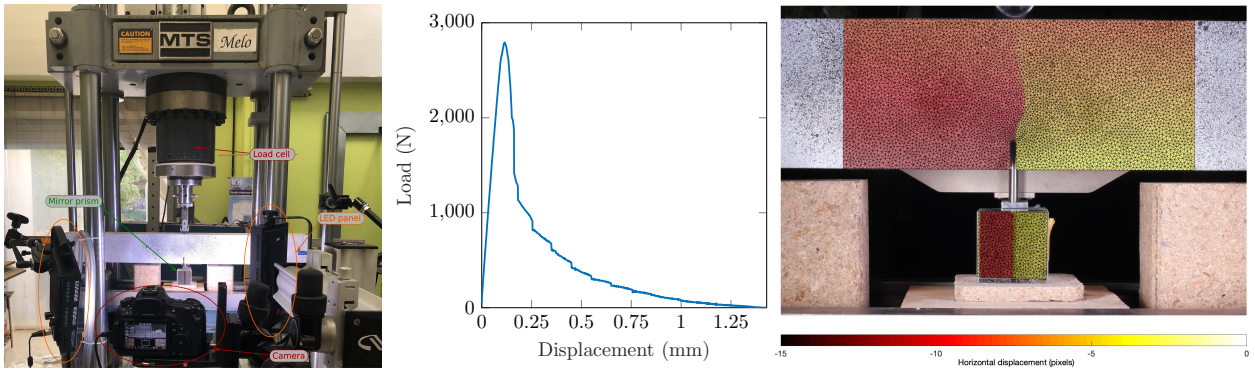


Figure 3: Front view of the experimental setting (left), registered load/displacement curve (center), and measured DIC displacement field in the central part of the beam (right).

The test is instrumented with DIC technology, full-fields measurements being performed on the central part of the specimen (Figure 3). For this purpose, a painted speckle pattern is sprayed on the considered surface of the concrete beam, and LED panels are added in order to adjust the exposition and contrast of the picture. Furthermore, a prism including a 45 degree-oriented mirror is placed under the notch in order to be able to use DIC on the bottom side with the same camera. That way, the reflection of the bottom of the beam is in the camera axis and the crack opening is visible. DIC pictures are taken every 5 s. The *Corelli* software [12] is then used to post-process the pictures by building a FE mesh with linear triangle elements in the zone of interest. We notice that the *in situ* DIC measurements are used here for sequential data assimilation, even though the experimental test is not performed with real-time post-processing due to practical constraints.

Figure 3 shows an illustration of the displacement field (x coordinate) measured from DIC at an advanced damage state of the specimen during the test. The crack inside the specimen is clearly visible owing to the observed discontinuity in the experimental displacement field.

6.2. Mathematical model and PGD numerical approximation

In agreement with the experimental setting, a mathematical model is defined on a 3D domain and with boundary conditions detailed in Figure 4. The material behavior is based on an isotropic damage elastic model adapted from [16] (the crack closing behavior is not considered here). It is described by means of the following constitutive relations:

$$\sigma = (1-d)\mathbf{C}\epsilon \quad ; \quad d(Y, A_d, Y_0) = 1 - \frac{1}{1 + A_d(Y - Y_0)} \tag{29}$$

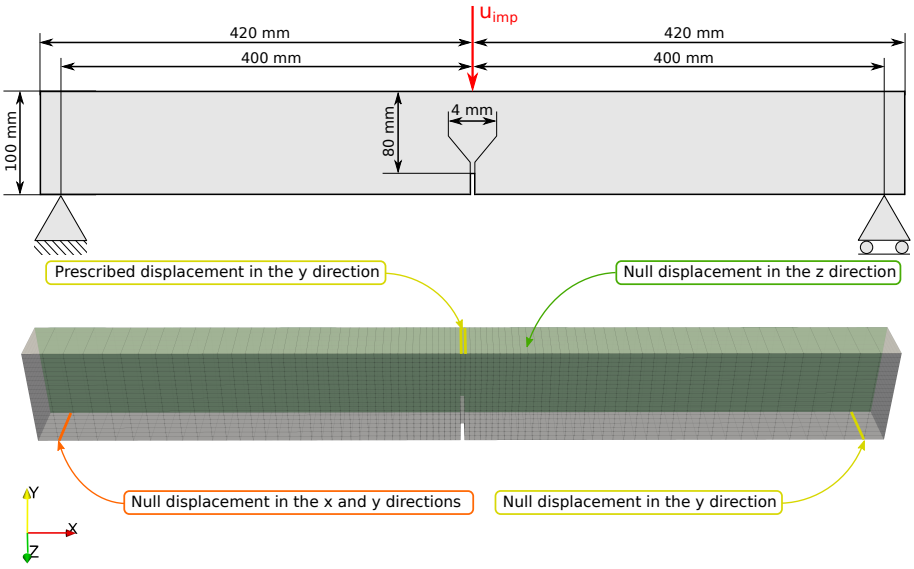


Figure 4: Geometry and boundary conditions for the model problem (top), and mesh and boundary conditions used for the computation of the PGD solution (bottom).

in which:

- σ (resp. ϵ) is the Cauchy stress tensor (resp. linearized strain tensor)
- $\mathbf{C}$ is the Hooke tensor
- d is the scalar damage variable
- $Y = \frac{1}{2}\langle\epsilon\rangle_+ : \mathbf{C} : \langle\epsilon\rangle_+$ is the released energy rate, $\langle\epsilon\rangle_+$ denoting the positive part of ϵ
- Y_0 is the initial threshold on energy rate for damage initiation
- A is a scalar brittleness parameter that drives the post-peak behavior

Consequently, the damage law depends on parameters Y_0 and A_d which will be later inferred in real-time from DIC measurements during the sequential data assimilation process. The influence of these two parameters is shown in Figure 5 by considering the response $\sigma_t = f(\epsilon_t)$ to a unidirectional traction test, with σ_t the traction stress component and ϵ_t the longitudinal strain component.

Following the works on the formulation of multi-parametric nonlinear PGD models (see [25]), an approximate PGD solution of the previous mathematical model is computed in the offline phase. Its expression at order m reads:

$$\mathbf{u}_m(\mathbf{x}, t, Y_0, A_d) = \sum_{k=1}^{m} \mathbf{\Lambda}_k(\mathbf{x})\lambda_k(t)\alpha_k^1(Y_0)\alpha_k^2(A_d) \tag{30}$$

The computation is performed using the LATIN-PGD algorithm [11]. In this context, the geometry is discretized with linear cubic (Q8) elements. The displacement u_{imp} is prescribed as a linearly increasing loading from 0 to 4×10^{-4} m; it is applied through 100 loading steps. The reference values for material

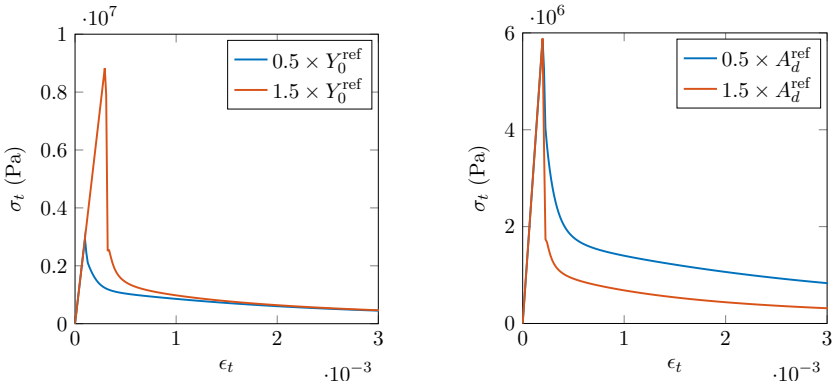


Figure 5: Influence of the damage law parameters Y_0 (left, for fixed A_d) and A_d (right, for fixed Y_0).

parameters are set to $E = 30$ GPa (Young modulus), $\nu = 0.23$ (Poisson ratio), $Y_0^{\text{ref}} = 216$ J.m^{-3}, and $A_d^{\text{ref}} = 2.25 \times 10^{-3}$ J^{-1}.m^3; they stem from an educated guess using experimental campaigns on similar materials. The variability of damage parameters is chosen to be centered on the reference values $(Y_0^{\text{ref}}, A_d^{\text{ref}})$ with a 50% variation magnitude. The dimensionless quantities $\bar{Y}_0 = Y_0/Y_0^{\text{ref}}$ and $\bar{A}_d = A_d/A_d^{\text{ref}}$ are introduced for comparison purposes.

An accurate PGD approximation is obtained with $m = 6$ modes. The corresponding space modes $\mathbf{\Lambda}_k(\mathbf{x})$, time modes $\lambda_k(t)$, and parameter modes $\alpha_k^1(\bar{Y}_0)$ and $\alpha_k^2(\bar{A}_d)$ are shown in Figures 6 and 7. All modes except time modes are normalized, which explains the decrease in the magnitude of time modes when m increases (the contribution of each PGD then becomes smaller and smaller). We also notice that the first PGD mode corresponds to a full elasticity mode as the corresponding parameter modes (α_1^1, α_1^2) are unit functions.

6.3. Data assimilation with synthetic measurements

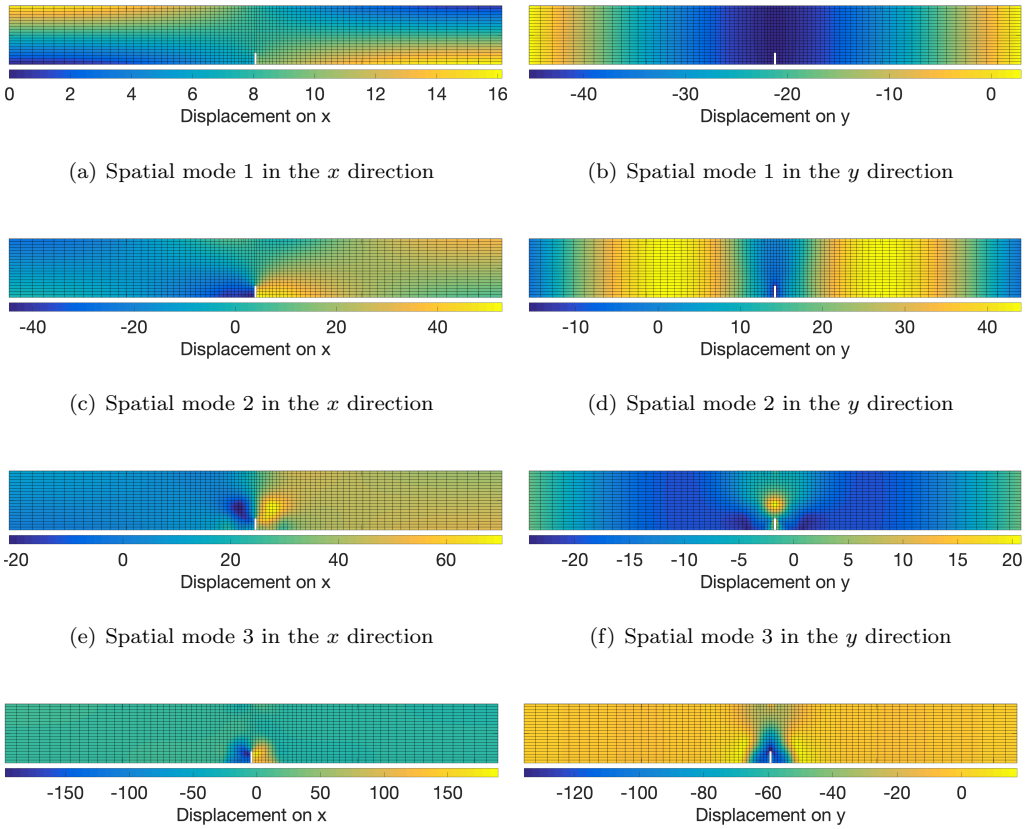
6.3.1. Sequential updating of damage law parameters

From the PGD model detailed above, we now implement the data assimilation strategy based on Transport Map sampling and PGD model reduction described in Sections 2 and 3. Dimensionless parameters $\bar{Y}_0$ and $\bar{A}_d$ are thus sequentially updated within the Bayesian framework. Consequently, at each assimilation time t_i and from data $\mathbf{d}_j^{\text{obs}}$, $1 \leq j \leq i$, the following posterior pdf is estimated:

$$\pi(\bar{Y}_0, \bar{A}_d | \mathbf{d}_1^{\text{obs}}, \ldots, \mathbf{d}_i^{\text{obs}}) \propto \prod_{j=1}^{i} \pi(\mathbf{d}_j^{\text{obs}} | \bar{Y}_0, \bar{A}_d).\pi_0(\bar{Y}_0, \bar{A}_d) \tag{31}$$

Data should in practice correspond to some mesh dofs computed from DIC. However, for validation purposes, they are here simulated from the PGD model. 30 nodes are randomly chosen as measurement points in the central part of the beam (see Figure 8), and displacement data at these nodes are obtained with parameter values $Y_0 = Y_0^{\text{ref}}$ and $A_d = A_d^{\text{ref}}$ (i.e. $\bar{Y}_0 = \bar{A}_d = 1$). A Gaussian white noise with a standard deviation of 10^{-6} m is then added to those displacements in order to simulate the DIC measurement noise.

The results in terms of sequentially obtained marginals along the assimilation process are shown in Figure 9. The prior pdf $\pi(\bar{Y}_0, \bar{A}_d)$ is chosen as a Gaussian distribution with mean $(0.9, 1.1)$ and covariance



(a) Spatial mode 1 in the x direction

(b) Spatial mode 1 in the y direction

(c) Spatial mode 2 in the x direction

(d) Spatial mode 2 in the y direction

(e) Spatial mode 3 in the x direction

(f) Spatial mode 3 in the y direction

(g) Spatial mode 4 in the x direction

(h) Spatial mode 4 in the y direction

Figure 6: Representation of the first spatial modes.

matrix $0.2\,\mathbb{I}$. We observe that the obtained marginals have a large width during the first assimilation time steps, due to the fact that the structure is then in its undamaged elastic regime so that displacement data give very poor information on damage parameters. However, maxima of the marginals at the final time give the estimation $(\bar{Y}_0, \bar{A}_d) = (1.04, 1.00)$ of the parameters, which is very close to the reference values. We also observe that the convergence in the estimation of $\bar{Y}_0$ is slower than that for $\bar{A}_d$, due to different parameter sensitivities. At the final time, the standard deviation of the marginal on $\bar{Y}_0$ is 0.041 while it is 0.022 for the marginal on $\bar{A}_d$.

The associated computation costs are also given in Figure 9. On the one hand, the dashed line corresponds to the CPU time required to compute each transport map using the explicit formulation of the functional alone (zero-order derivative) in the solution of the minimization problem (8). On the other hand, the solid line refers to computations using the explicit formulation of both functional and its gradient (first-order derivative) computed thanks to the PGD representation (30). In addition, the bar chart indicates the final map order which is required at each assimilation step in order to achieve the given tolerance on the variance diagnostic ($\epsilon_\sigma = 10^{-3}$). Results clearly show that using the functional gradient highly speeds up the computation of the transport maps, especially when the map order becomes large.

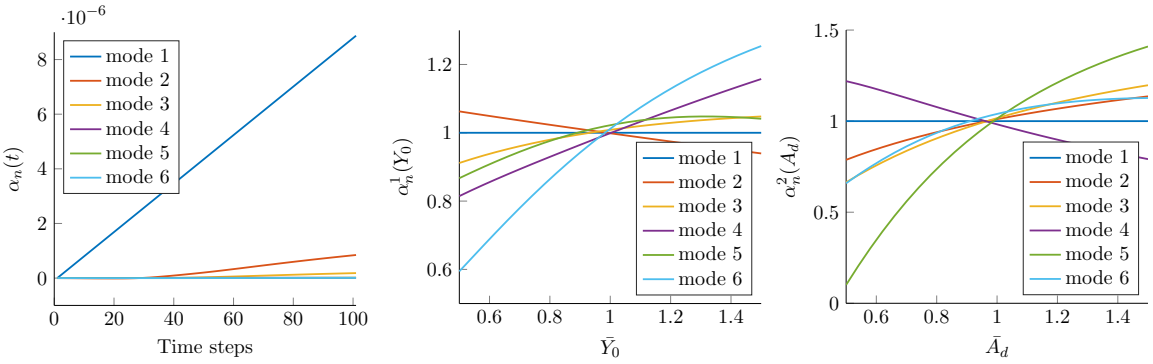


Figure 7: Representation of the first time modes (left), parameter $\bar{Y}_0$ modes (center), and parameter $\bar{A}_d$ modes (right).

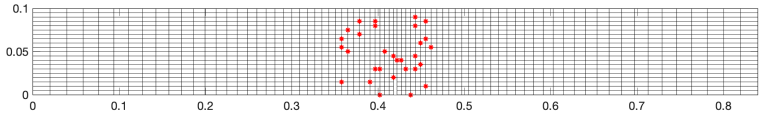


Figure 8: Location of the nodes considered as measurement points.

We can also explain the changes of map order occurring at time points 32, 36, 38 and 39. During the first time steps, the structure is in its undamaged elastic regime (data thus do not provide any knowledge on the parameters) so that the posterior pdfs remain very close to the prior Gaussian pdf and first-order transport maps are sufficient to represent the mapping with the reference standard normal pdf. Between time points 30 and 40, the change in the posterior pdf is large as the structure is subjected to a nonlinear damage behavior. Therefore, higher-order maps are necessary to represent the mapping between consecutive posterior pdfs. After this critical stage, and due to short time steps, the evolution of the posterior pdf is smooth and first-order transport maps can be reused.

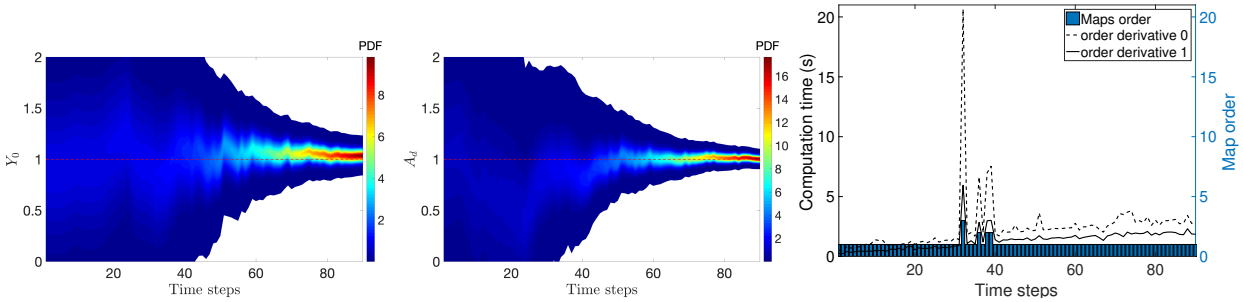


Figure 9: Successive posterior marginals on $\bar{Y}_0$ (left) and $\bar{A}_d$ (center), and cost for transport map computation (right), along the data assimilation process.

6.3.2. On-the-fly prediction of the crack length

Once the posterior pdfs on parameters are updated at each assimilation time point, uncertainties can be propagated trough the PGD model in order to predict the behavior of the system, in terms of some model outputs, for future times. The output which is considered here is the crack length l_T at final time T.

In order to do the link between the PGD damage model and fracture mechanics, a kinematic bridge is built. It involves the elastic solution of a beam, with similar geometry and boundary conditions as previously, but in which a vertical crack with variable length l is inserted. We assume that the crack is initiated in the right corner of the notch. Different FE meshes with n_{dof} dofs are created (their topology depends on the value of l) and the corresponding elastic solutions with prescribed unitary displacement are computed in the offline phase. In the present study, the crack length range varies from $l = 0$ mm to $l = 80$ mm, and it is discretized in 80 points so that 80 elastic computations are performed. Figure 10 shows some of the obtained solutions for the particular crack lengths $l = 0$ mm, $l = 25$ mm, and $l = 75$ mm.

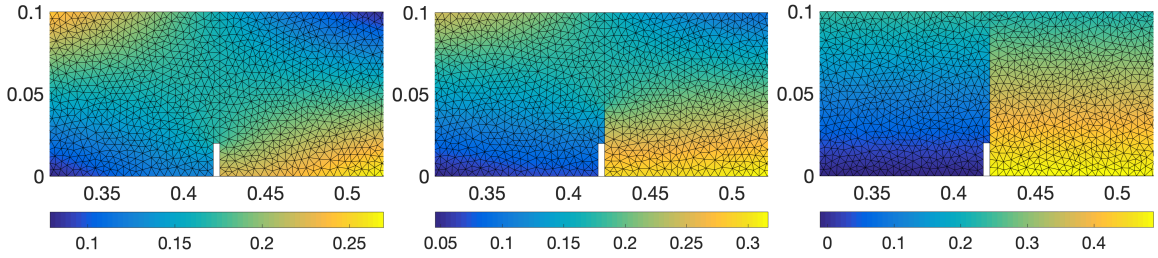


Figure 10: Elastic solution in terms of x-component of the displacement field for crack lengths $l = 0$ mm (left), $l = 25$ mm (center), and $l = 75$ mm (right).

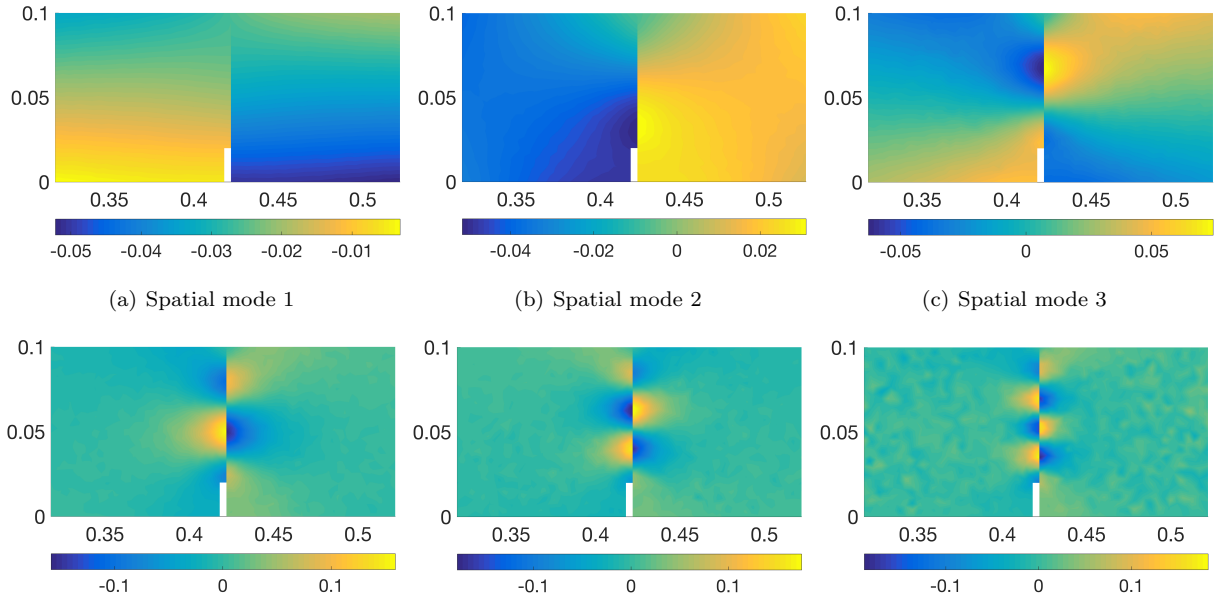
A meta-model is then constructed, still in the offline phase, from the 80 available snapshots. For this purpose, snapshots are stored in a matrix $\mathbb{Y} = \{\mathbf{Y}_1, ..., \mathbf{Y}_{80}\}$ with $\mathbf{Y}_j \in \mathbb{R}^{n_{\text{dof}}}$ the displacement vector for the elastic solution j, and the SVD of $\mathbb{Y}$ is computed as $\mathbb{Y} = \mathbb{U}\mathbb{D}\mathbb{V}^T$. $\mathbb{U} = \{\mathbf{U}_1, ..., \mathbf{U}_{n_{\text{dof}}}\}$ is a $n_{\text{dof}} \times n_{\text{dof}}$ matrix, $\mathbb{V} = \{\mathbf{V}_1, ..., \mathbf{V}_{80}\}$ is a 80×80 matrix, and $\mathbb{D}$ is a $n_{\text{dof}} \times 80$ diagonal matrix containing singular values σ_j, $j \in \{1, ..., 80\}$. By truncating the SVD decomposition to N_{SVD} modes, the displacement field $\mathbf{u}_{\text{SVD}}(\mathbf{x}, l)$ in the structure with respect to the crack length l can be approximated as:

$$\mathbf{u}_{\text{SVD}}(\mathbf{x}, l) = \sum_{k=1}^{N_{\text{SVD}}} \sigma_k \mathbf{u}_k(\mathbf{x}) v_k(l) \tag{32}$$

where functions $\mathbf{u}_k(\mathbf{x})$ and $v_k(l)$ are interpolations obtained from vectors $\mathbf{U}_k$ and $\mathbf{V}_k$, respectively. The first six spatial SVD modes $\mathbf{u}_k(\mathbf{x})$ are shown in Figure 11, and the corresponding parameter SVD modes $v_k(l)$ (weighted by the mode magnitude σ_k) are shown in Figure 12. We observe that a SVD truncation is relevant after a few modes (as singular values σ_k rapidly decrease to zero), that the first two SVD modes correspond to the kinematics with a full-length crack, and that the other SVD modes can be seen as localized stitching patches acting along the crack.

From the previous SVD meta-model, the PGD approximation (30), and the continuously updated pdfs on Y_0 and A_d, the final crack length l_T can be estimated in a stochastic sense. The associated pdf reads:

$$\pi(l_T) = \pi_{\text{u}}(\mathbf{u}^{\text{SVD}}(l_T)).\pi_0(l_T) \tag{33}$$



(a) Spatial mode 1 (b) Spatial mode 2 (c) Spatial mode 3

(d) Spatial mode 4 (e) Spatial mode 5 (f) Spatial mode 6

Figure 11: Spatial SVD modes in terms of x-component of the displacement field.

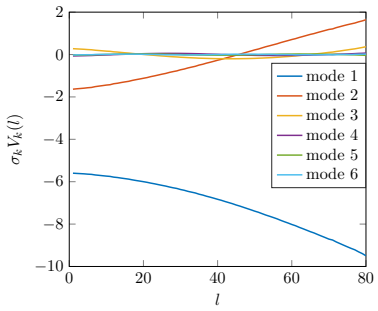


Figure 12: Weighted parameter SVD modes associated to the parameter l.

where $\pi_0(l_T)$ is the prior pdf (based on *a priori* knowledge on l_T), and $\pi_{\mathrm{u}}(\mathbf{u}^{\mathrm{SVD}}(l))$ is the likelihood function obtained by the propagation of uncertainties on parameters Y_0 and A_d through the PGD model evaluated at time T and at the considered measurement points (i.e. $\mathbf{u}_m(\mathbf{x}^{\mathrm{obs}}, T, Y_0, A_d)$). Such a propagation is in practice performed efficiently using the additional use of inverse transport maps [15].

Choosing the prior pdf $\pi_0(l_T)$ as a Gaussian pdf with mean 0.04 and standard deviation 0.013, results on the sequential estimation of l_T by means of successive posterior pdfs $\pi(l_T)$ are reported in Figure 13. We observe that during the first time steps, the estimation of the final crack length l_T is very coarse, with a large variance, then it improves and converges to a mean estimation $l_T = 0.0372$ and a maximum *a posteriori* estimation $l_T = 0.0376$. In addition, the cumulative computation time required to perform both parameter inference and output prediction is given in Figure 13. The blue part corresponds to the CPU time associated with the updating of parameters (Y_0, A_d), while the red part corresponds to the CPU time associated with the estimation of l_T. We notice that all iterations can be performed in less than 5 s (on a standard laptop)

which is the duration between two successive picture acquisitions. Therefore, the approach can be considered as real-time in the present context.

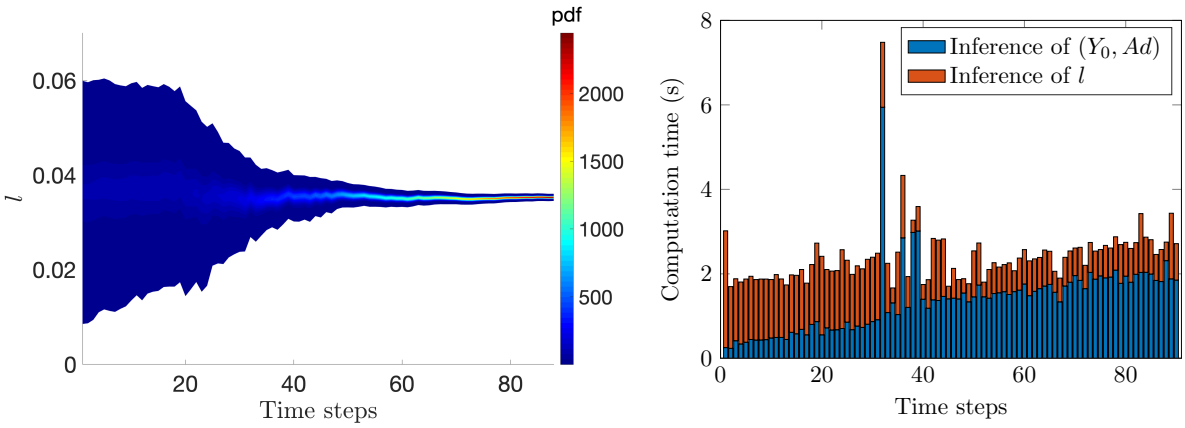


Figure 13: Successive posterior pdfs on the output l_T (left), and total computation cost at each time step (right).

6.4. Correction of model bias

Keeping the same data assimilation process as in the previous section, a model bias is now introduced by means of a non-negligible truncation error in the PGD representation $\mathbf{u}_m$ used to perform Bayesian inference. We choose $m = 3$, keeping synthetic data computed from an accurate PGD model with $m = 6$. The obtained results in terms of successive marginals for parameters $\bar{Y}_0$ and $\bar{A}_d$ are given in Figure 14. We observe that the influence of the model bias remains limited during the first time steps, due to the fact that only the first PGD mode has a contribution in the undamaged elastic regime, but it later becomes important, leading to shifted marginals and wrong estimation of the model parameters (means are not even converging).

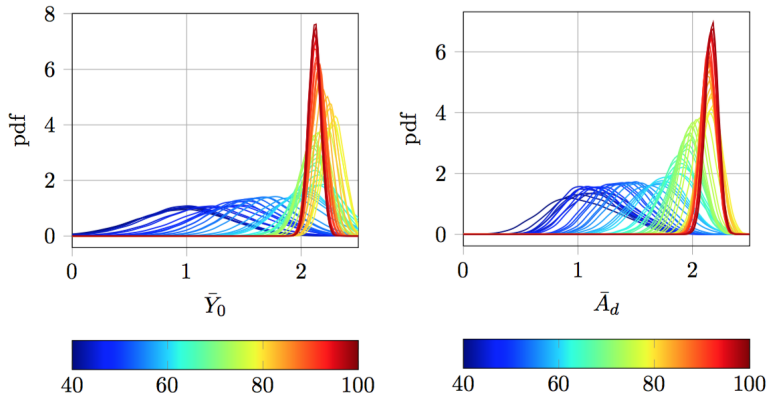


Figure 14: Posterior marginals on dimensionless parameters $\bar{Y}_0$ (left) and $\bar{A}_d$ (right). The color map indicates the correspondence with assimilation time points.

To correct the model bias, we implement the procedure proposed in Section 4. The bias is estimated at the 30 measurement points (i.e. 60 dofs) and the stochastic correction is computed by the extrapolation of the SVD modes for residual mean and variance. Here, only two SVD modes are kept to represent the model correction. The results obtained with this correction of the model bias are given in Figure 15. We observe that the proposed model correction procedure highly improves the quality of the sequential model updating, eliminating the effect of model bias in the posterior pdfs. Information on model evidence is also displayed in Figure 15, comparing evolutions for the reference data assimilation (performed with $m = 6$) and for the biased data assimilation (performed with $m = 3$) with or without model bias correction. All evolutions are smoothed with a moving average filter. We observe that the model evidence slowly drops off from the reference value when using a 3-mode PGD model without any correction. Adding the correction improves the model evidence except in the first time steps when the model bias is the lowest (as mostly measurement noise is then extrapolated).

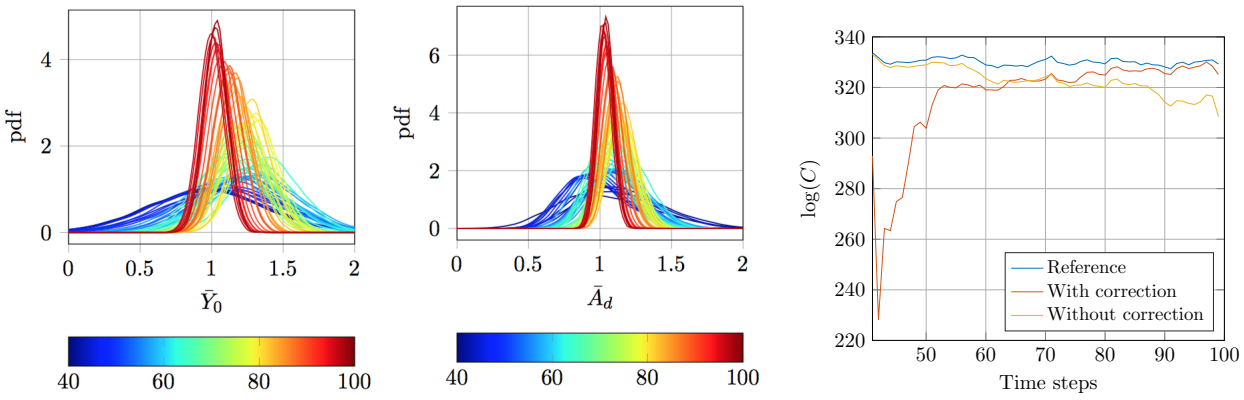


Figure 15: New posterior marginals on dimensionless parameters $\bar{Y}_0$ (left) and $\bar{A}_d$ (center) obtained after model bias correction, and evolution of the smoothed model evidences (right).

6.5. Selection of most relevant DIC measurements

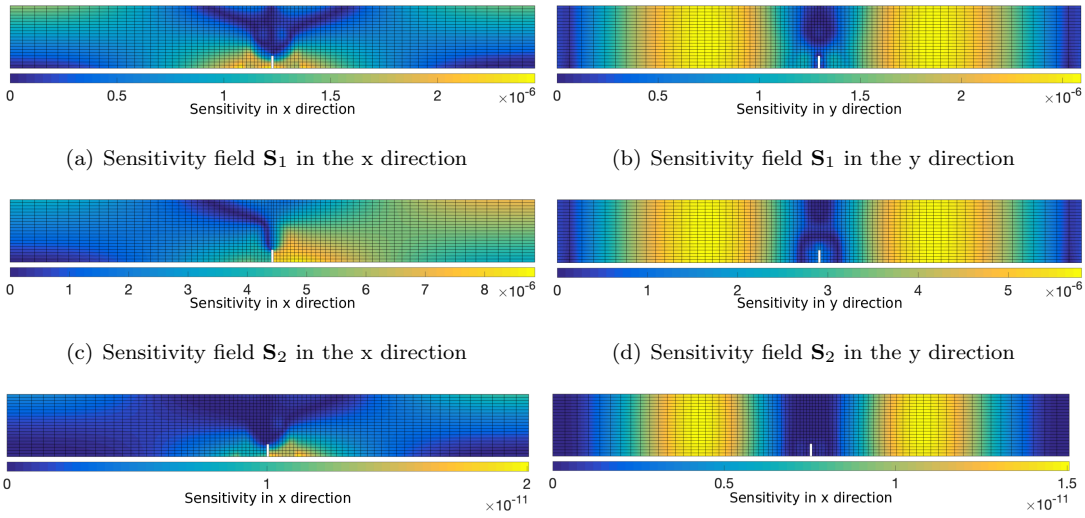
In this section, we investigate the procedure proposed in Section 5 in order to select the most relevant data among all full-field DIC measurements available. For convenience reasons and to keep notations clear, we denote $p_1 = \bar{Y}_0$ and $p_2 = \bar{A}_d$. The first step is to compute sensitivity fields $\mathbf{S}_i = \partial \mathbf{u} / \partial p_i$ $(i \in \{1, 2\})$; this is performed in the offline phase and from the PGD model as:

$$\mathbf{S}_1(\mathbf{x}) = \left| \sum_{k=1}^{m} \mathbf{\Lambda}_k(\mathbf{x}) \lambda_k(t^\diamond) \frac{\partial \alpha_k^1}{\partial \bar{Y}_0}(\bar{Y}_0^\diamond) \alpha_k^2(\bar{A}_d^\diamond) \right| \quad ; \quad \mathbf{S}_2(\mathbf{x}) = \left| \sum_{k=1}^{m} \mathbf{\Lambda}_k(\mathbf{x}) \lambda_k(t^\diamond) \alpha_k^1(\bar{Y}_0^\diamond) \frac{\partial \alpha_k^2}{\partial \bar{A}_d}(\bar{A}_d^\diamond) \right| \tag{34}$$

where $t^\diamond$, $\bar{Y}_0^\diamond$, and $\bar{A}_d^\diamond$ are mean values of time t and parameters $\bar{Y}_0$ and $\bar{A}_d$, respectively. In addition, the cross-sensitivity field $\mathbf{S}_{12} = \mathbf{S}_1 \times \mathbf{S}_2$ is defined in order to get an average sensitivity of the model with respect to both parameters. Indeed, when identifying strictly independent parameters, the data selected with respect

to the sensitivity of one parameter may be insensitive to the other parameters.

Figure 16 shows the sensitivity fields in both directions x (horizontal) and y (vertical).



(a) Sensitivity field $\mathbf{S}_1$ in the x direction

(b) Sensitivity field $\mathbf{S}_1$ in the y direction

(c) Sensitivity field $\mathbf{S}_2$ in the x direction

(d) Sensitivity field $\mathbf{S}_2$ in the y direction

(e) Sensitivity field $\mathbf{S}_{12}$ in the x direction

(f) Sensitivity field $\mathbf{S}_{12}$ in the y direction

Figure 16: Visualization of the sensitivity fields.

From the computation of sensitivity fields, the most relevant DIC data are selected from weighted fields $\mathbf{S}_i^{\text{DIC}} = \mathbb{M}_{DIC}\mathbf{S}_i$ ($i \in \{1, 2, 12\}$). Figure 17 shows the map of the field $\mathbf{S}_{12}^{\text{DIC}}$. The brighter nodes correspond to measurements points with the higher sensitivity for the objective of parameter updating under measurement uncertainties. The 30 most sensitive dofs among all the DIC dofs are then selected and represented in Figure 17. These are all located on the right-hand side of the structure which is the area with the larger horizontal displacement. This set of locations is highly non-symmetric due to the boundary conditions applied in the mathematical model (the x-component of the displacement field being imposed to zero on the left boundary of the structure).

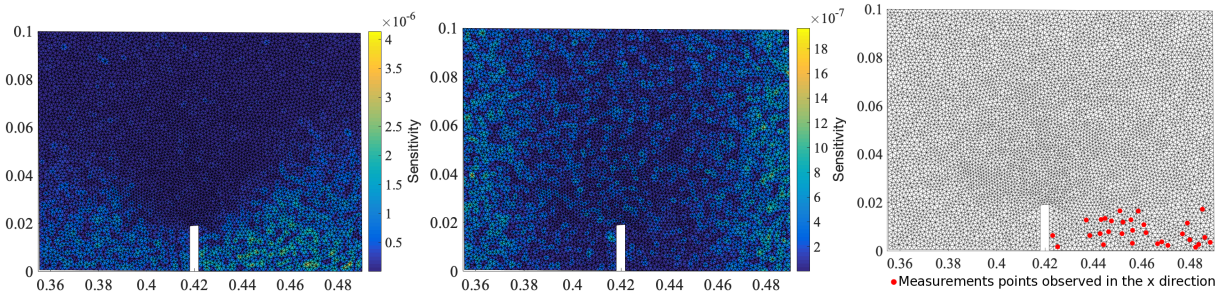


Figure 17: Sensitivity $\mathbf{S}_{12}^{\text{DIC}}$ along the x-direction (left) and the y-direction (center), and selected 30 most sensitive dofs (right).

7. Conclusions

We presented a new numerical strategy for real-time Bayesian data assimilation, mixing Transport Map sampling and reduced-order modeling (PGD) tools. It was complemented with procedures for data selection, correction of model bias, and uncertainty propagation. The overall strategy, resorting to computations performed in an *offline* phase, leads to an effective approach in engineering activities. It was here implemented and illustrated on a real-life test dealing with crack propagation in a concrete beam equipped with *in-situ* DIC measurements. Based on its innovative and attractive aspects, the proposed strategy should pave the way for future research works on the topic.

References

[1] Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking, IEEE Transactions on Signal Processing;50(2):174–188.

[2] Berger J, Orlande HRB, Mendes N (2017) Proper Generalized Decomposition model reduction in the Bayesian framework for solving inverse heat transfer problems, Inverse Problems in Science and Engineering;25(2):260–278.

[3] Calvetti D, Dunlop M, Somersalo E, Stuart A (2018) Iterative updating of model error for Bayesian inversion, Inverse Problems;34(2).

[4] Chinesta F, Keunings R, Leygue A (2014) The Proper Generalized Decomposition for Advanced Numerical Simulations: A Primer, SpringerBriefs in Applied Sciences and Technology.

[5] Chinesta F, Cueto E, Abisset-Chavanne E, Duval J-L, Khaldi F.E (2018) Virtual, digital and hybrid twins: A new paradigm in data-based engineering and engineered Data, Archives of Computational Methods in Engineering;online.

[6] Darema F (2004) Dynamic Data Driven Applications Systems: A new paradigm for application simulations and measurements, Computational Science - ICCS:662–669.

[7] El Moselhy TA, Marzouk Y (2012) Bayesian inference with optimal maps, Journal of Computational Physics;231(23):7815–7850.

[8] Hild F, Roux S (2006) Digital image correlation: From displacement measurement to identification of elastic properties - A review, Strain;42(2):69–80.

[9] Jeffreys H (1961) Theory of Probability, Oxford, Clarendon Press, 3rd edition.

[10] Kaipio J, Somersalo E (2004) Statistical and Computational Inverse Problems, Springer-Verlag New York.

[11] Ladevèze P (2016) On reduced models in nonlinear solid mechanics, European Journal of Mechanics - A/Solids;60:227–237.

[12] Leclerc H, Neggers J, Matthieu F, Hild F, Roux S (2015) Correli 3.0, IDDN.FR.001.520008.000.S.P.2015.000.31500, Agence pour la Protection des Programmes, Paris.

[13] Maday Y, Patera A.T, Penn J.D, Yano M (2014) A parametrized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics, International Journal for Numerical Methods in Engineering;102(5):933–965.

[14] Marzouk Y, Moselhy T, Parno M, Spantini A (2016) Sampling via measure transport: an introduction, Handbook of Uncertainty Quantification,1–41.

[15] Parno M.D, Marzouk Y.M (2018) Transport map accelerated Markov Chain Monte-Carlo, SIAM-ASA Journal on Uncertainty Quantification;6(2):645–682.

[16] Richard B, Ragueneau F (2013) Continuum damage mechanics based model for quasi brittle materials subjected to cyclic loadings: Formulation, numerical implementation and applications, Engineering Fracture Mechanics;98:383–406.

[17] Robert CP, Casella G (2004) Monte Carlo Statistical Methods, Springer Texts in Statistics, New York.

[18] Ross D.A, Lim J, Lin R.S, Yang M.H (2008) Incremental learning for robust visual tracking, International Journal of Computer Vision;77(1-3):125–141.

[19] Rubio PB, Louf F, Chamoin L (2018) Fast model updating coupling Bayesian inference and PGD model reduction, Computational Mechanics;62(6):1485–1509.

[20] Rubio PB, Louf F, Chamoin L (2019) Transport Map sampling with PGD model reduction for fast dynamical Bayesian data assimilation, International Journal in Numerical Methods in Engineering, online.

[21] Spantini A, Bigoni D, Marzouk Y (2018) Inference via low-dimensional couplings, Journal of Machine Learning Research;19:1–71.

[22] Stuart AM (2010) Inverse problems: a Bayesian perspective, Acta Numerica;19:451–559.

[23] Tarantola A (2005) Inverse Problem Theory and Methods for Model Parameter Estimation, Society for Industrial and Applied Mathematics.

[24] Villani C (2008) Optimal Transport: Old and New, Springer.

[25] Vitse M, Néron D, Boucard P-A (2019) Dealing with a nonlinear material behavior and its variability through PGD models: Application to reinforced concrete structures, Finite Elements in Analysis and Design;153:22–37.