RESEARCH ARTICLE

WILEY

A probabilistic data assimilation framework to reconstruct finite element error fields from sparse error estimates: Application to sub-modeling

James Paul Rouse[1] | Pierre Kerfriden[2,3] | Maxime Hamadi[4]

[1]Gas Turbine and Transmission Research Centre (G2TRC), University of Nottingham, Nottingham, Nottinghamshire, UK

[2]Centres des Matériaux, Mines Paris - PSL, Paris, France

[3]School of Engineering, Cardiff University, Cardiff, UK

[4]Airbus, Toulouse, France

Correspondence
James Paul Rouse, Gas Turbine and Transmission Research Centre (G2TRC), University of Nottingham, Nottingham, Nottinghamshire NG7 2RD, UK.
Email: james.rouse@nottingham.ac.uk

Abstract

The present work proposes a computational approach that recovers full finite element error fields from a small number of estimates of errors in scalar quantities of interest. The approach is weakly intrusive and is motivated by large scale industrial applications wherein modifying the finite element models is undesirable and multiple regions of interest may exist in a single model. Error estimates are developed using a Zhu-Zienkiewicz estimator coupled with the adjoint methodology to deliver goal-oriented results. A Bayesian probabilistic estimation framework is deployed for full field estimation. An adaptive, radial basis function based reduced order modeling strategy is implemented to reduce the cost of calculating the posterior. The Bayesian reconstruction approach, accelerated by the proposed model reduction technology, is shown to yield good probabilistic estimates of full error fields, with a computational complexity that is acceptable compared to the evaluation of the goal-oriented error estimates. The novelty of the work is that a set of computed error estimates are considered as partial observations of an underlying error field, which is to be recovered. Future improvements of the method include the optimal selection of goal-oriented error measures to be acquired prior to the error field reconstruction.

KEYWORDS
error estimation, finite element method, Gaussian process

1 | INTRODUCTION

Structural analysis, utilized to answer many key questions in the design process of complex engineering systems, often requires the abstraction of features at many different length scales. Without some level of length scale limit (applied during model development), practical analysis problems would quickly become intractable and solutions would not be generated at a rate required by the iterative design process. Still, a range of length scales will almost always be required to appropriately represent a real structure. Geometric features, material property definitions, and discretization particulars (element formulations and refinement levels) will commonly vary over the different length scales. Broadly speaking, large, coarse ("global scale") models will be solved to address questions relating to the "general" stiffness of a structure.

Global scale results may well also inform unknown driving boundary conditions for smaller scale detailed ("local scale") models that represent localizations where gradient quantities (likely stresses and strains) are of interest. Much has been written about the various approaches adopted for these multiscale problems and all (definitionally) involve the transfer of information between global and local models, either in a strongly coupled sense (where information flow is bidirectional) or in an weakly coupled sense (where information flow is unidirectional). The present work concerns itself mainly with weakly coupled multiscale problems and asks what are the ramifications of "error" in global models? Paying particular attention to the practical limitations associated with large scale analyses, how can error estimates in large models be generated quickly and used to their greatest effect, such that uncertainties can be propagated to local models? In the present work, we motivate, propose, and demonstrate a full error field recovery approach that utilizes a small number of goal-orientated error estimate (GOEE) observations in combination with a Gaussian process (GP) state estimation framework. Importantly, a guided model order reduction technique, enabled by a stochastic partial differential equation (SPDE) prior, is proposed that alleviates computational bottlenecks in the GP steps, thereby expediting full error field estimation.

Full field error estimation, that is to say the estimation of a continuous error field over an entire domain rather than at several discrete locations, is now motivated through a simple aerospace example. Consider a large airframe assembly, such as the fuselage panel section of civil aircraft (see Figure 1). It is clear that the real assembly will consist of a multitude of components, from large skin, stringer, and stabilizer components that are vital in the approximation of stiffness, to small "features" such as holes, rivets, and fillet sections, which are prime failure initiation sites that must be given careful consideration if a safe design is to be produced. An approximation of global stiffness of the panel section may well be required in order to anticipate the outcome of, say, wing bend tests. Clearly, small scale features (holes, rivets, and through thickness sections, for example) add little to global stiffness, justifying their omission from large scale models. Local models, where small scale features (holes and composite material laminates, for example) are explicitly represented, may be used to answer concerns over the potential for premature failure. It is highly unlikely that any loads or boundary conditions will be known for a local model a priori. This motivates the multiscale approach and suggest that boundary conditions should be extracted from a global model. Implicit in this approach is the assumption that either boundary conditions are known with a good level of certainty for the global model, or that local models are sufficiently remote from them that any uncertainty has a limited effect.

An important question naturally arises. If the global model is coarse (and presumably relatively inaccurate in terms of local deformation fields), can the local model results be trusted? Confidence in local model could (conceptually) be gained by performing error estimation around the region of interest (RoI) in the global model and propagating the uncertainty in driving degrees of freedom (DoFs). This general observation motivates the present work. Before expanding on the proposed resolution, the nuance of the problem will be discussed and the work will be contextualized by considering error estimation in the literature.

The present work considers error estimation in finite element (FE) problems, as FE is by far the most commonly employed approach for engineering stress analysis. Approximation errors are the inaccuracies which are inherent to the discretization methods that are required in order to approximate the solutions to mathematical models. This contrasts with modeling error, which is a measure of how well an abstract model approximates real physical phenomena. Estimates
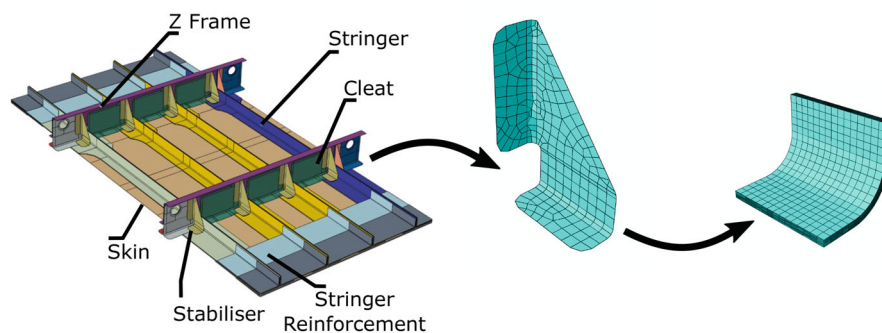

FIGURE 1 An example multiscale aerospace problem. The large scale structure (on the left) made up of repeated instances of a small number of components (stabilizers, for example, shown in the middle). An engineer may be interested in the stress state at some local feature (a fillet section, for example, as shown on the right). In the context of the present work, the structure is a global problem that can be realized with coarse elements, whereas detailed component/feature models (driven by the structural assembly model) are local problems.

of approximation error may be termed a priori estimates or a posteriori estimates,[1] with the former utilizing the problem definition and the discretization to estimate error and the latter using the solution approximation itself to estimate error. A further distinction may also be made between the approximation of error bounds and the estimation of error itself.[2] The former can be guaranteed but may be inaccurate. Error bounds may be large, for example, although it is important to note that this is not always the case and bounds may be sufficiently narrow that they function as estimates. The latter is usually not guaranteed but the range of values associated with it will typically be narrower than for error bounds. Grätch and Blathe note that error estimates should possess several properties.[2] In addition to accuracy, error estimates should asymptotically tend to zero as the discretization density increases, should produce tight bounds for the error, should be computationally inexpensive, should be applicable in a wide range of potentially non-linear applications, and should inform mesh refinements such that approximate solution processes can be optimized. At present, no one estimator can satisfy all of these requirements however, for the applications considered in the present work, several relevant methods can be found in the literature (as will be discussed later).

The approximation of errors in elliptical partial differential equations, such as those which govern linear elastic stress analysis, has received a great deal of attention.[3-6] In many cases, error measures are utilized to drive mesh refinement through, for example, polytree decomposition algorithms.[4,7-10] Error estimates have been developed for a wide range of solution approximation methods. In addition to "conventional" FE, error estimates have been derived for boundary element methods,[8] immersed surface methods,[11] multigrid and composite FE methods,[7,12,13] extended FE (XFEM),[1,14] and stress singularity problems.[15] Oden and Prudhomme's[16] and Larson and Runesson's[7] contributions are particularly relevant to the present work due to its treatment of error estimation in multiscale problems. The interested reader may also refer to our previous work on sub-modeling based on Oden's work[17,18] and to Reference 19 for a contribution related to the so-called multiscale FE method (MsFEM).

Numerous reviews of a priori and a posteriori error estimators are available in the literature.[1,2] In most cases, error estimators are categorized as energy norm based (including element residual, sub-domain residual methods and the constitutive relation error) or recovery based estimators. The latter includes the well known Babuska and Rheinboldt estimator, the Kelly, Gage, Zienkiewicz and Babuska estimator, and the Zienkiewicz-Zhu patch recovery technique.[1] In the present work goal-oriented error estimates (GOEEs) are utilized as, rather than providing a general indication of approximation error, they allow for the quantification of error in a specific quantity of interest (QoI).[2,16] GOEEs were proposed in the 1990's[20] in the work of Prudhomme, Oden, and Ainsworth,[1,16,21] Ladevéze,[22,23] Bathe,[2] and Cirak[24] to name a few of the most influential authors in this prolific area of research. GOEEs extend the energy norm error estimator to represent linear engineering QoIs. QoIs must be predefined (in the present work this is done at the local length scale) and observations will typically be sparse (given that GOEEs are furnished using solutions to additional adjoint problems). Such features of the method usually limit its use. For example, GOEEs are not typically applied in the study of errors in maximum equivalent stress or in the analysis of complete error fields at the global scale. The present work will offer a response to these limitations through the application of a GP data assimilation framework.

In many cases, the GOEE approach has been used to represent the uncertainty in a physically relevant quantity that may be turned into an error on a failure criterion.[24-27] In the multiscale setting, a posteriori error estimates have been applied to Type A problems in the work of Tirvaudey et al.,[28] wherein a weighted residual based GOEE is applied in non-intrusive sub-modeling problems. A posteriori error estimates for MsFEM problems have been developed in the work of Chung and Chamoin.[29,30] In the context of the present work, QoIs are driving DoFs extracted from global scale models. If meaningful error estimates for those quantities can be derived, it is the straightforward to sample from the resulting distributions and propagate uncertainties to continuum element local sub-models. The statistical recovery process alluded to here is the focus and novelty of the present work. The methods introduced here allow an analyst to compute corrected local solutions by propagating the effect of driving DoFs corrections—the error estimates—from global to local scales.

The present work concerns itself with an industry relevant problem. We seek to evaluate the influence of global mesh quality onto the solution of a local problem. The solution developed here is non-intrusive in terms of re-meshing at the global, as this is typically not feasible in real problems. Furthermore, the solution allows for redefinition of local models and attempts to alleviate computational bottlenecks. It should be noted that, in the case of a single QoI that is defined for a single (pre-defined) local scale model, the GOEE approach can be used with little modification. This is not the focus of the present work however. We here consider a case where errors in driving DoFs at the global scale need to be assessed, with a view to propagating epistemic uncertainty to local length scale models. The approach proposed here uses Zhu-Zienkiewicz (ZZ) error estimates to calculate errors in global model deformation fields. Due to the additional computational expense associated with solving adjoints, only a small number of observations are considered. GPs are then trained using a SPDE prior[31] to interpolate between different observations of the error, that is, the ZZ estimates are seen as partial measurements

of the full error field in a state estimation framework.[32,33] Computational bottle necks in the generation of the prior are circumvented though a GOEE model order reduction method that makes use of hierarchical clustering. The SPDE prior naturally incorporates measures of proximity and its use avoids the $o(N^3)$ complexity associated with Kernel approaches. Errors in local model boundary conditions (driving DoFs) can be evaluated by Monte-Carlo sampling of the recovered error field on the RoIs. Novelty in the work is derived from the application of ZZ error estimates in the multiscale setting, the use of ZZ evaluated adjoint problems in the GOEE setting for assessing the quality of local deformations, and the use of functional GPs in multiscale problems.

2 | PROBLEM SETTING AND FORMULATION

2.1 | Linear elasticity

We begin with a general setting for the linear problems of elasticity that we will consider in the present work. Let us consider domain $\Omega \in \mathbb{R}^d$, with $d \in \{2, 3\}$ (only two dimensional problems will be considered in the present work, however extensions to three dimensions are trivial and the authors' previous work further considers "shell" element formulations[34]). We look for a displacement field $u \in \mathcal{U} = \mathcal{H}_1(\Omega)$, the space of fields of $\mathbb{R}^d$ with values in $\mathbb{R}^d$ whose derivative are square-integrable over Ω. The solution is assumed to satisfy homogeneous Dirichlet boundary conditions $u = 0$ on $\partial\Omega_u$, which is of non-zero measure. The complementary part of boundary $\partial\Omega$, over which Neumann boundary conditions are applied, is denoted $\partial\Omega_t$. The weak form of the problem of linear elasticity is as follows. We look for $u \in \mathcal{U}$ such that

$$\forall v \in \mathcal{U}, \quad a(u, v) = l(v). \tag{1}$$

In the previous equation, bi-linear form a is defined by

$$a(u, v) := \int_\Omega \nabla_s u : C : \nabla_s v \, d\Omega, \tag{2}$$

where ∇_s is the symmetrized gradient operator, C is the usual fourth-order Hooke tensor corresponding to isotropic linear elasticity, and linear form l is defined by

$$l(v) := \int_\Omega f \cdot v \, d\Omega + \int_{\partial\Omega_t} T \cdot v \, d\Gamma, \tag{3}$$

where f is a known source field, and T is a known field of prescribed boundary tractions.

2.2 | Finite element approximation

We now introduce the standard "P1" Lagrange FE method to approximately solve the elasticity problem introduced above. Domain Ω is tessellated into a set $\mathcal{T}_h$ of simplexes. We now look for FE field $u_h \in \mathcal{U}_h \subset \mathcal{U}$ such that

$$\forall v_h \in \mathcal{U}_h, \quad a(u_h, v_h) = l(v_h). \tag{4}$$

$\mathcal{U}_h$ is the space of continuous, piecewise linear FE displacement fields that satisfy the homogeneous Dirichlet conditions. Mathematically,

$$\mathcal{U}_h = \left\{ u_h(x) \in \mathcal{U} \mid \exists \mathbf{U} \in \mathbb{R}^N, \ u_h(x) = \sum_{i=1}^{\hat{N}} \sum_{j=1}^{d} \left(\phi_i(x) e_j \right) U_{(i-1)d+j} = \sum_{i=1}^{N} \psi_i(x) U_i \right\}, \tag{5}$$

where the $\hat{N} = N/d$ shape functions ϕ_i defined over Ω and with values in $\mathbb{R}$ are the standard piecewise linear "hat" functions associated with the vertices of tessellation $\mathcal{T}_h$, and e_j is the jth canonical basis vector of $\mathbb{R}^d$. In the previous

expression, only the hat functions that vanish over Dirichlet boundary $\partial\Omega_d$ are considered so that the homogeneous boundary conditions are automatically satisfied.

The FE model may be written in matrix form as follows:

$$\mathbf{KU} = \mathbf{F}, \tag{6}$$

where $K_{ij} = a(\psi_i, \psi_j)$ is an element of stiffness matrix $\mathbf{K}$ and $F_i = l(\psi_i)$ is an element of force vector $\mathbf{F}$.

2.2.1 | Finite element error

The FE error $e_h = u - u_h \in \mathcal{U}$ satisfies the weak form

$$\forall v \in \mathcal{U}, \quad a(e_h, v) = r(v) := l(v) - a(u_h, v), \tag{7}$$

and satisfies the Galerkin orthogonality condition

$$\forall v_h \in \mathcal{U}_h, \quad a(e_h, v_h) = 0. \tag{8}$$

2.3 | Sub-modeling

We now consider a sub-modeling approach which draws on the solution of a global FE model (or global/coarse model) to provide more comprehensive results in a RoI.

For the present work, sub-models (or local/fine models) are considered to possess sufficiently fine mesh densities such that, compared to the global model, they are error free. In another words, we are interested in quantifying the effect of discretization errors from the global FE model onto the solution of the sub-models. It should be noted that, in practice, an analyst has far greater control over local models than global models. In the aerospace industry for example, many engineers across multiple departments may be called upon to produce a global aircraft model. Re-meshing in such a case would be unacceptable and the degree of discretization may be limited by computational resources. Nonetheless, local feature models may well be generated by a single analyst however, thereby providing capacity for refinement.

The RoI $\Omega_I \subset \Omega$ is to be re-analyzed using a finer local model (e.g., non-linear, time-dependent model, 3D extrusion of a global plate/shell model). To achieve this, the displacements computed using the coarse FE model are extracted from $\partial\Omega_I$, and are subsequently applied to the boundary of the sub-model.[16,18] In all of our examples, the mesh of the global model and that of the sub-models are compatible: they have the same nodes on the boundary of the RoI $\partial\Omega_I$. This is not a restriction of the approach but one of our current implementation of the proposed methodology (see Figure 3 for further details on this aspect). A continuous error field approximation will be developed in the following sections, which means that Ω_I can be located without restriction within Ω.

Our goal is to estimate the error in the boundary field to be transferred from the coarse to the fine model, and propagate this error though the chosen sub-modeling technique.

3 | GOAL-ORIENTED ERROR ESTIMATION

3.1 | Gradient recovery based error estimation

We wish to estimate $\|e_h\|_a = \sqrt{a(e_h, e_h)}$ (error in energy norm) without computing too expensive an estimate for exact error field e_h. The Zienkiewicz-Zhu (ZZ) recovery based estimate of $\|e\|_a$ is the following[35]

$$v_{zz}^2 = \int_\Omega (\epsilon^\star - \nabla_s u_h) : C : (\epsilon^\star - \nabla_s u_h) \, d\Omega, \tag{9}$$

where $\epsilon^\star$ is a continuous, smoothed recovered strain field that is obtained by post processing the discontinuous, piecewise constant tensor field $\nabla_s u_h$. One may simply average the values of the strain tensor corresponding to all the elements connected to a node, for each node of the mesh, and use the FE shape functions to obtain a continuous strain field. More advanced techniques include the super convergent patch recovery approach (SPR),[36] or equilibrium informed recovery procedures,[14,37,38] which yield improved performances of the error estimation procedure, but will not be considered in the present contribution.

3.2 | Error estimates for engineering quantities of interest (QoI)

The error in energy norm is usually not of prime interest sub-modeling, we are interested in a limit states quantities in our structural zooms, such as maximum stress measures. To address the practical limitation of a-norm error estimation, the GOEE methodology is classically deployed to extend error estimation in a-norm to error estimation in engineering quantities of interest.

We begin by introducing the so-called adjoint (or dual) problem.

We look for adjoint solution $z \in \mathcal{U}$ (the influence function in the work of Prudhomme and Oden[39]) such that

$$\forall v \in \mathcal{U}, \qquad a(v, z) = Q(v). \tag{10}$$

The linear functional $Q : \mathcal{U} \to \mathbb{R}$ extracts one of the QoIs from an arbitrary displacement field $v \in \mathcal{U}$. Many different forms of $Q(u)$ can be chosen and the most appropriate choice will depend on which QoIs relate to the analyst's objectives. For instance, $Q(u) = \frac{1}{|\Omega_I|} \int_{\Omega_I} e_1 \cdot u \, d\Omega$ is the average of u in the first spatial dimension, in region $\Omega_I \subset \Omega$. $Q(u) = \frac{1}{|\Omega_I|} \int_{\Omega_I} \Sigma : C : \nabla_s u \, d\Omega$ extracts a weighted average of the component of the stress field in Ω_I. For the present work, we look to quantify errors in the nodal displacement values for all nodes located on the boundary of the sub-model.

In this context, QoIs will be weighted averages of the components of u corresponding to this boundary.

Now we have, following standard derivations from the GOEE literature,

$$\forall v \in \mathcal{V}, \qquad a(e, z) = Q(e) = Q(u) - Q(u_h) = r(z). \tag{11}$$

The adjoint field z is unknown and must be approximated by using the FE method. We will look for FE field $z_h \in \mathcal{U}_h$ such that

$$\forall v_h \in \mathcal{U}_h, \qquad a(v_h, z_h) = Q(v_h). \tag{12}$$

We can therefore write that

$$Q(e) = a(e, z_h) + a(e, z - z_h) = r(z_h) + a(e, e_z) = a(e, e_z). \tag{13}$$

The residual term in the last equation vanished due to Galerkin orthogonality. The remaining term contains both exact errors (in the direct and adjoint problems), and may be evaluated by a ZZ type recovery based error estimate, as follows

$$Q(u) - Q(u_h) \approx \hat{Q}_h - Q(u_h) := \int_\Omega (\epsilon^\star - \nabla_s u_h) : C : (\epsilon_z^\star - \nabla_s z_h) \, d\Omega, \tag{14}$$

where $\hat{Q}_h$ is an estimate of the QoIs, recovered from the FE field, that approximates the QoI $Q(u)$.

If the error analysis was limited to a single RoI with a small number of FE DoFs for the boundary of the sub-model, the above treatment would be appropriate. One adjoint problem would be created for each of these DoFs, yielding an error estimate for every one of them, separately. This "truth" solution is presented in the following subsection. However, when several RoIs are considered, or when the boundary of the sub-model possesses a large number of DoFs, computing all the corresponding errors in quantities of interest becomes intractable. Recall that solving an adjoint problem with a direct solver typically involves performing forward and backward substitution with the Cholesky factor (the factorization itself has already been computed for the original, forward FE problem). These, operations are followed by a ZZ gradient recovery process (Equation 14), whose associated cost cannot be neglected.

As a solution to tackle this challenge, we will develop a full field reconstruction technique, using a GP as prior statistical model for the error field and a small number of selected GOEE as "data points" for model fitting.

4 │ FULL ERROR FIELD RECONSTRUCTION THROUGH BAYESIAN FUNCTIONAL REGRESSION

4.1 │ "Truth" reconstruction of finite element error fields using the goal-oriented error estimation approaches

We now introduce a corrected FE model

$$\hat{u}_h(x) = \sum_{i=1}^{N} \psi_i(x)\hat{U}_i, \tag{15}$$

where

$$\hat{\mathbf{U}} = \mathbf{U} + \hat{\mathbf{e}}. \tag{16}$$

In the previous expressions, $\hat{\mathbf{e}}$ is a vector of FE DoF values corresponding to a FE representation of the exact error field $e_h = u - u_h$. This is to be understood in the sense that the corrected FE displacement should yield QoIs that are closer to that delivered by an infinitely refined FE model. This means that we aim to estimate the pollution error, whilst the interpolation error itself is assumed to be small enough to be ignored, or is to be estimated by other means.

Two relevant FE representations of the exact error field naturally come to mind for $\hat{e}_h = \hat{u}_h(x) - u_h(x) = \sum_{i=1}^{N} \psi_i(x)\hat{e}_i$

- the FE interpolant $\mathcal{I}_h e_h$ of the exact error field e_h in $\mathcal{U}_h$,
- the $\mathcal{L}_2(\Omega)$-projection of the exact error field e_h onto $\mathcal{U}_h$.

We choose to require $\hat{\mathbf{e}}$ to be the vector of nodal values of the $\mathcal{L}_2(\Omega)$ projection of the exact error field onto the FE space. Indeed, any FE linear functional of the exact error, that is, $Q(e) = \int_\Omega \xi_h \cdot e_h \, d\Omega$, where $\xi_h \in \mathcal{U}_h$ is a vector valued FE field, has the following property:

$$\int_\Omega \xi_h \cdot e_h \, d\Omega = \int_\Omega \xi_h \cdot \pi_{\mathcal{L}_2}^h e_h \, d\Omega, \tag{17}$$

where $\pi_{\mathcal{L}_2}^h$ is the $\mathcal{L}_2(\Omega)$ projector and the above equality stems from standard operations on projectors. Hence, measure $Q(e)$ of the exact error is also a measure of its $\mathcal{L}_2$ projection. As a consequence, it is possible to construct a ZZ based approximation of the $\mathcal{L}_2$ projection of the exact error onto the FE space by noting that

$$\forall i \in [\![1 \; N]\!], \qquad \int_\Omega \psi_i \cdot e_h \, d\Omega = \int_\Omega \psi_i \cdot \hat{e}_h \, d\Omega = \mathbf{c}^{i^T} \mathbf{M} \hat{\mathbf{e}}, \tag{18}$$

where

$$M_{ij} = \int_\Omega \psi_i \cdot \psi_j \, d\Omega, \tag{19}$$

and $\mathbf{c}^i$ is the ith canonical basis vector of $\mathbb{R}^d$. Hence, knowing these N GOEE ZZ estimates, an inversion of the mass matrix $\mathbf{M}$ gives access to $\hat{\mathbf{e}}$. A similar conclusion can be made if only part of the error field is to be reconstructed from GOEE. Of course, computing N GOEE ZZ estimates is out of the question in all but simplest of demonstrative models. AS mentioned above, we will instead proceed by GP functional regression,[40,41] to statistically estimate these N quantities based on the availability of a selected few of them.

4.2 | Bayesian reconstruction of finite element errors fields from sparse functional observations

4.2.1 | Field reconstruction via Gaussian process functional regression

Bayesian estimation setting

We now assume that we have calculated an array of M ZZ estimates of errors in scalar QoIs. For the purpose of Bayesian full field reconstruction, these estimates are seen as noisy observations of the corresponding exact error in QoI, which reads as

$$\forall i \in [\![1, M]\!], \qquad d_i = \mathbf{Q}^{i^T}\hat{\mathbf{e}} + \epsilon_i = \Delta Q_i + \epsilon_i, \tag{20}$$

where $\Delta Q_i \approx (Q^i(\pi_{\mathcal{L}_2}^h u) - Q^i(u_h))$ is the true error in the ith QoI.

Above, $\epsilon_i \sim \mathcal{N}(0, \sigma^i)$ is a white noise that allows us to represent the fact that the error estimate is not exact, that is, the ZZ estimate is only an estimate of the true error in QoI. We will set $\sigma^i = |\Delta Q_i|^2$, which qualitatively encodes the fact that "the typical error in the error estimate is of the order of the error estimate itself"[*].

We define the vector of errors in QoIs as

$$\mathbf{d} = \Delta\mathbf{Q} + \epsilon = \begin{pmatrix} \Delta Q_1 \\ \Delta Q_2 \\ \dots \\ \Delta Q_M \end{pmatrix} + \epsilon = \begin{pmatrix} \mathbf{Q}^{1^T} \\ \mathbf{Q}^{2^T} \\ \dots \\ \mathbf{Q}^{M^T} \end{pmatrix}\hat{\mathbf{e}} + \epsilon = \mathbf{H}\hat{\mathbf{e}} + \epsilon, \tag{21}$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma_\epsilon)$ is a M dimensional random vector containing the errors in ZZ estimated errors in QoI, $\Delta\mathbf{Q}$ is a M dimensional vector containing the true errors in QoIs, and $\mathbf{H}$ is a M by N linear observation operator. In our examples, we will use FE shape functions as extractor of quantities of interest. In this case, $\mathbf{Q}^{i^T} = \mathbf{c}^{\mathcal{I}(i)^T}\mathbf{M}$, as shown in the previous section ($\mathcal{I}$ is an arbitrary index mapping). Notice that consistently with the previous paragraph, the data covariance matrix Σ_ϵ is diagonal, with known entries.

We further assume that the unknown error vector $\hat{\mathbf{e}}$ is zero mean and Gaussian distributed, as follows

$$\hat{\mathbf{e}} \sim \mathcal{N}(\mathbf{0}, \Sigma), \tag{22}$$

which, together with the distribution of ϵ, encodes all prior knowledge about epistemic sources of uncertainty in our Bayesian estimation problem.

Posterior distributions

We can now look for the posterior probability distribution of $\hat{\mathbf{e}}$, given the M noisy measurements. Following standard techniques in GP inference, the distribution of the state and the noisy observations is jointly and consistently Gaussian, which is summarized by

$$\begin{pmatrix} \mathbf{d} \\ \hat{\mathbf{e}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{H}\Sigma\mathbf{H}^T + \Sigma_\epsilon & \mathbf{H}\Sigma \\ \Sigma\mathbf{H}^T & \Sigma \end{pmatrix} \right). \tag{23}$$

The posterior probability of $\hat{\mathbf{e}}$ is simply obtain by Gaussian conditioning[†], that is,

$$\hat{\mathbf{e}}|\mathbf{d} \sim \mathcal{N}\left(\mathbf{e}^\star, \Sigma^\star \right), \tag{25}$$

where

$$\mathbf{e}^\star = \Sigma\mathbf{H}^T(\mathbf{H}\Sigma\mathbf{H}^T + \Sigma_\epsilon)^{-1}\mathbf{d}$$
$$\Sigma^\star = \Sigma - \Sigma\mathbf{H}^T(\mathbf{H}\Sigma\mathbf{H}^T + \Sigma_\epsilon)^{-1}\mathbf{H}\Sigma. \tag{26}$$

The marginal posterior distribution of any set of linear combinations of the nodal values of the error field is given by

$$\mathbf{E}(\hat{\mathbf{e}}|\mathbf{d}) \sim \mathcal{N}(\mathbf{E}\mathbf{e}^\star, \mathbf{E}\Sigma^\star \mathbf{E}^T), \tag{27}$$

where prediction operator $\mathbf{E}$ is of size n_e by N. In this article, we are interested in restriction operators, that is, Boolean rectangular matrices with a single non-zero element per line. In practice, this amounts to predicting the components of the error vector corresponding to a particular subset of the FE DoFs, typically all the DoFs corresponding to a sub-domain $\Omega_I \subset \Omega$, or to the boundary of this subdomain in the context of FE sub-modeling.

4.2.2 | Prior covariance structure

Great care is required when choosing the prior distribution. Computational tractability of the posterior is of course vitally important. However, the form of the prior covariance will also dictate how scalar error estimates will be assimilated to produce complete error fields. If unrepresentative correlations are utilized across the domain, poor posterior distributions are to be expected. In the context of structural modeling, a simple prior based on Euclidean distances may well suggest strong correlation between two neighborhoods that, while physically close, are not mechanically connected and thus should not be thought to greatly influence one another *a priori*. A simple example of this challenge is a "U" channel section, wherein the top edges of the section may be relatively close to one another but this proximity does not (necessarily) suggest strong correlation. What follows is a brief discussion of Matérn type priors. We first introduce standard kernel based priors, which defines covariances pairwise, based on the Euclidean distance between pairs of points. This type of classical priors will not be used in this article. Instead, we will use implicitly defined priors via the solution of SPDE, which automatically takes geodesic proximity into account when establishing the correlation structure. In addition, using mechanical operators to define the SPDE naturally couples the dimensions of the prior random field, while such a coupling would have to be developed from scratch when using a Kernel based approach.

Kernel based covariance
A standard choice is to use Kernels to define the covariance matrix, that is,

$$\mathrm{Cov}(\mathbf{c}^{i^T}\hat{\mathbf{e}}, \mathbf{c}^{j^T}\hat{\mathbf{e}}) = k\left(P_{\lceil \frac{i}{d} \rceil}, P_{\lceil \frac{j}{d} \rceil}\right) \delta\left(\frac{i}{d} - \left\lfloor \frac{i}{d} \right\rfloor, \frac{j}{d} - \left\lfloor \frac{j}{d} \right\rfloor\right), \tag{28}$$

where $\mathbf{c}^i$ is the ith canonical basis vector of $\mathbb{R}^N$ and $P_k \in \mathbb{R}^d$ is the position of the FE vertex associated with the kth linear shape function. For instance, the Whittle Matérn Kernel reads as

$$k\left(P_i, P_j\right) = \alpha^2 \left(\frac{\|P_i - P_j\|_2}{l}\right)^\nu K_\nu \left(\frac{\|P_i - P_j\|_2}{l}\right), \tag{29}$$

where K_ν is the modified Bessel function of second kind of order ν, and α and l are hyperparameters of the GP. The formulation described here corresponds to the introduction a continuous GP and its subsequent restriction to the nodes of the FE mesh.

Random process based covariance
In this article, we define the prior for error field $\hat{\mathbf{e}}$ using a SPDE approach, as suggested in References 31 and 42. The distribution of $\hat{\mathbf{e}}$ is implicitly defined through the solution of stochastic linear system

$$\left(\mathbf{M} + \beta^2 \mathbf{K}\right) \hat{\mathbf{e}} = \alpha \sqrt{\mathbf{M}} \mathbf{W}, \tag{30}$$

which we rewrite as

$$\mathbf{A}\hat{\mathbf{e}} = \mathbf{b} \quad \text{with} \quad \mathbf{A} = \left(\mathbf{M} + \beta^2 \mathbf{K}\right) \quad \text{and} \quad \mathbf{b} = \alpha \sqrt{\mathbf{M}} \mathbf{W}. \tag{31}$$

In the previous expressions, taken from the work of Roininen et al.,[31] $\mathbf{M}$ and $\mathbf{K}$ are defined by $\mathbf{M}_{ij} = \int_{\Omega} \psi_i(x)\psi_j(x)\,d\Omega$ and $\mathbf{K}_{ij} = \int_{\Omega} \nabla_s\psi_i(x) : C : \nabla_s\psi_j(x)\,d\Omega$, respectively, and are the standard FE mass and stiffness matrices, respectively. $\mathbf{W} \in \mathbb{R}^N$ is a white noise vector with normally distributed independent components. α and β are the two parameters of the Gaussian prior (also called hyperparameters in Bayesian statistics).[‡]

By applying the standard rules of linear operators applied to Gaussian distributed multivariate variables, we find the prior covariance for error vector $\hat{\mathbf{e}}$ is

$$\boldsymbol{\Sigma} = \alpha^2 \left(\mathbf{M} + \beta^2\mathbf{K}\right)^{-1}\mathbf{M}\left(\mathbf{M} + \beta^2\mathbf{K}\right)^{-1}. \tag{34}$$

Notice that mass matrix $\mathbf{M}$ is invertible, under weak assumptions on the quality of the FE discretization. Moreover, both $\mathbf{M}$ and $\mathbf{K}$ are symmetric and positive. Hyperparameter β is a covariance length, which controls the smoothness of the random field, while hyperparameter α controls the overall amplitude of the GP.[§]

4.2.3 | Parameter optimization

Optimizing the hyperparameters of the prior in a data driven manner may be done by maximizing the marginal likelihood with respect to hyperparameter vector $\boldsymbol{\theta} := (\alpha\ \beta)^T$. In our case, the data log likelihood reads as[43]

$$\log(\mathcal{L}_{\theta}(\mathbf{d})) = -\frac{1}{2}\mathbf{d}^T(\mathbf{H}\boldsymbol{\Sigma}(\theta)\mathbf{H}^T + \boldsymbol{\Sigma}_{\epsilon})^{-1}\mathbf{d} - \log(Z(\theta)), \tag{36}$$

where $Z(\theta) = \sqrt{(2\pi)^M|\mathbf{H}^T\boldsymbol{\Sigma}(\theta)\mathbf{H} + \boldsymbol{\Sigma}_{\epsilon}|}$. Providing that the number of observations remains small, this optimization problem is tractable (the cost of assembling innovation matrix ($\mathbf{H}\boldsymbol{\Sigma}(\theta)\mathbf{H}^T + \boldsymbol{\Sigma}_{\epsilon}$), which dominates the computational cost associated with the present optimization process, is discussed in the next section).

For now, this optimization will be performed "by hand." We calibrate the overall size and length scale of the GP in order to match observations, by a procedure of trial and error. Fully automatizing the algorithm is a perspective of this work but it not implemented here.

4.2.4 | Sampling the posterior distribution

To estimate the impact of the discretization errors on quantities of interest, we sample the posterior distribution of error fields using a Monte-Carlo process.

Sampling the posterior distribution may be done without assembling the full posterior covariance matrix. This is done as follows. One first draws a sample $\mathbf{s}_0 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ from the prior distribution. The sample is transformed into a sample $\mathbf{s} \sim \mathcal{N}(\mathbf{m}^\star, \boldsymbol{\Sigma}^\star)$ of the posterior distribution using the following operation

$$\mathbf{s} = \mathbf{s}_0 + \mathbf{G}\left(\tilde{\mathbf{d}} - \mathbf{H}\mathbf{s}_0\right), \tag{37}$$

where $\tilde{\mathbf{d}} \sim \mathcal{N}(\mathbf{d}, \boldsymbol{\Sigma}_{\epsilon})$ is a randomly perturbed observation, and the Kalman gain $\mathbf{G}$ is defined as

$$\mathbf{G} = \boldsymbol{\Sigma}\mathbf{H}^T(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T + \boldsymbol{\Sigma}_{\epsilon})^{-1}. \tag{38}$$

Statistics of any number of engineering QoIs can be obtained after repeating this operation many time (typically more than 50 times).

In the context of sub-modeling, we will, for every sample, extract the trace of the reconstructed error field on the boundary of the RoI, and add it to the displacement solution obtained by solving the initial global FE problem. This corrected boundary condition will subsequently be applied to the sub-model to deliver corrected local solutions. By repeating this operation for all available global samples, empirical distributions for any local QoI may be constructed. The shift between the median of this empirical distribution and the uncorrected local result will represent a quantified bias introduced by FE errors at the global scale, while the variance of this distribution will encode both our uncertainty in the accuracy of

the ZZ estimate, and the additional epistemic uncertainty related to the Bayesian reconstruction method introduced in this section.

4.3 | Computational cost

The procedure of full field reconstruction described previously is unfortunately prohibitively expensive to be used without additional numerical developments. The difficulty is that multiplications by the prior covariance matrix $\boldsymbol{\Sigma}$, which are required to assemble the innovation matrix $\mathbf{S} = \left(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T + \boldsymbol{\Sigma}_\epsilon\right)$ and further required to compute the posterior mean and draw samples from the posterior error field distribution according to Equation (37), involves solving a sparse linear system of size N, that is, a system as large as that yielded by the global FE problem[¶]. To simplify the discussion, we will count the necessary solutions of sparse linear systems of size N as an indication of the cost of the error reconstruction approach proposed previously.

First, consider operation sequence

$$\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T = \alpha^2 \mathbf{H}\left(\mathbf{M} + \beta^2\mathbf{K}\right)^{-T}\mathbf{M}\left(\mathbf{M} + \beta^2\mathbf{K}\right)^{-1}\mathbf{H}^T, \tag{39}$$

which needs performing in order to assemble and factorized the innovation matrix, which is a typically a very small, fully populated matrix, as the number of ZZ estimates is small. The above expression involves solving M systems of linear equations with system matrix $\left(\mathbf{M} + \beta^2\mathbf{K}\right)$ and with the columns of $\mathbf{H}^T$ as right hand side.

Once the innovation matrix has been inverted, the posterior mean may be computed as

$$\mathbf{e}^\star = \alpha^2\left(\mathbf{M} + \beta^2\mathbf{K}\right)^{-T}\mathbf{M}^{-1}\left(\mathbf{M} + \beta^2\mathbf{K}\right)^{-1}\mathbf{H}^T\mathbf{S}^{-1}\mathbf{d}, \tag{40}$$

which involves solving an additional linear system of equations with $\left(\mathbf{M} + \beta^2\mathbf{K}\right)$ as system matrix (the rightmost inversion may be inherited from the computation of the innovation matrix).

A sample $\mathbf{s}$ drawn from the posterior distribution is obtained by evaluating the expression

$$\mathbf{s} = \mathbf{s}_0 + \alpha^2\left(\mathbf{M} + \beta^2\mathbf{K}\right)^{-T}\mathbf{M}^{-1}\left(\mathbf{M} + \beta^2\mathbf{K}\right)^{-1}\mathbf{H}^T\mathbf{S}^{-1}\left(\tilde{\mathbf{d}} - \mathbf{H}\mathbf{s}_0\right), \tag{41}$$

which, once $\mathbf{s}_0$ is drawn from the prior distribution, requires solving another system of equations with system matrix $\left(\mathbf{M} + \beta^2\mathbf{K}\right)$ (for large number of MC draws compared to the number of ZZ estimates, further pooling of operations may be beneficial). Drawing a sample from the prior distribution also requires solving such a system. Indeed, a sample $\mathbf{s}_0$ is obtained by evaluating the following expression:

$$\mathbf{s}_0 = \alpha\left(\mathbf{M} + \beta^2\mathbf{K}\right)^{-T}\sqrt{\mathbf{M}}\mathbf{W}, \tag{42}$$

where $\mathbf{W} \sim \mathcal{N}(0, \mathbf{I}_d)$ and $\sqrt{\mathbf{M}}$ denotes the Cholesky factor of the mass matrix.

To summarize our findings, a lower bound for the number of times $\left(\mathbf{M} + \beta^2\mathbf{K}\right)$ needs to be inverted is M (for the assembly of the innovation matrix) plus n_{MC}, the number of Monte-Carlo draws from the posterior distribution. Of course, multiple right hand side strategies may be deployed to accelerate these operations. For a direct solver, the Cholesky decomposition, whose associated complexity scales as $\mathcal{O}(N^2)$, may be performed once and for all, subsequent solves requiring only forward and backward substitutions, which scale as $\mathcal{O}(N)$ for sparse systems. For iterative solvers, techniques may also be deployed to fasten repeated solutions to system of equation with a fixed system matrix, for instance projections in subspaces spanned by previous solutions, reuse of Krylov subspaces or appropriate initialization strategies from previously computed solutions. Nonetheless, the overall cost of the proposed GP based error field reconstruction approach exceeds that of computing the M adjoint solutions (and exceeds by far the cost of the original FE problem), which is undesirable. In the second part of the article we will propose a reduced modeling strategy to accelerate the GP based reconstruction methodology using a reduced basis method that circumvents the need to solve linear systems of size N.

It is now possible to compute the marginal distribution of the set of unseen linear combinations of components of $\hat{\mathbf{e}}$ that we are interested in. The posterior mean of $\mathbf{E}\hat{\mathbf{e}}$, is $\mathbf{E}\boldsymbol{\Sigma}\mathbf{H}^T(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T + \boldsymbol{\Sigma}_\epsilon)^{-1}\mathbf{d}$. Product $\boldsymbol{\Sigma}\mathbf{H}^T$ appearing in innovation matrix $\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T + \boldsymbol{\Sigma}_\epsilon$ dominates the computational cost. It may be evaluated by considering the lines of observation operator $\mathbf{H}$, one by one, and left applying the covariance operator to each individual observation vector. This involves solving M linear

problems of the type $\left(\mathbf{M} + \beta^2\mathbf{K}\right) . = ..$ This operation scales as $o(MN^2)$. The inversion of $(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T + \boldsymbol{\Sigma}_\epsilon)^{-1}$ is of negligible cost as we wish to recover full error fields based on a limited number of observations, with typically $M \leq 20$.

The computation of the posterior covariance is only tractable for low rank prediction operators $\mathbf{E}$. Indeed, the posterior covariance for the unseen quantities to be predicted is $\mathbf{E}\boldsymbol{\Sigma}\mathbf{E}^T - \mathbf{E}\boldsymbol{\Sigma}\mathbf{H}^T(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T + \boldsymbol{\Sigma}_\epsilon)^{-1}\mathbf{H}\boldsymbol{\Sigma}\mathbf{E}^T$. The second term can be computed by taking advantage of the fact that the number of observations is small. However, if prediction operator $\mathbf{E}$ has a large number of lines n_e, the first term becomes intractable. In our case of hierarchical Dirichlet sub-modeling, $\mathbf{E}$ will extract all the DoFs of $\partial\Omega_I$.

We may generate realizations of $\mathbf{Ee}|\mathbf{d}$ directly, without actually constructing the posterior covariance matrix. Such a sample $\mathbf{s} \sim \mathcal{N}(\mathbf{m}^\star, \boldsymbol{\Sigma}^\star)$ is obtained as follows:

$$\mathbf{s} = \mathbf{E}\left(\mathbf{x}_0 + \mathbf{G}\left(\tilde{\mathbf{d}} - \mathbf{H}\mathbf{x}_0\right)\right), \tag{43}$$

where $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is drawn from the prior distribution, $\tilde{\mathbf{d}} \sim \mathcal{N}(\mathbf{d}, \boldsymbol{\Sigma}_\epsilon)$ is a randomly perturbed observation, and the Kalman gain is $\mathbf{G} = \boldsymbol{\Sigma}\mathbf{H}^T(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T + \boldsymbol{\Sigma}_\epsilon)^{-1}$. Drawing a sample from the posterior distribution is dominated by the cost of solving two systems involving the matrix $\left(\mathbf{M} + \beta^2\mathbf{K}\right)$ on the left hand side (once to draw a sample from the prior distribution, and once to apply the Kalman gain). In this analysis, we consider that the innovation is already assembled and factorized to compute the posterior mean.

This analysis of the numerical tractability of the Gaussian regression process is unfortunate news. Indeed, the goal of the previous developments was to replace the expensive computation of many observations of the error field via the adjoint method by a cheap interpolation between a few selected observations. However, we see here that the GP interpolation itself requires solving a series of FE systems, each as expensive to solve as an adjoint problem (M systems to build the innovation matrix, one more to compute the posterior mean and 2 per sample to evaluate the variability via Monte-Carlo). The method can only be of benefit if the cost of evaluating the posterior distribution is made orders of magnitude cheaper than that of actually computing the all QoIs directly, which we will do in the following section by proposing a dedicated low rank approximation strategy.

Let us emphasize that the numerical difficulty is due to the dimension of the parameter space (as is often the case in state estimation, when using Kriging or Kalman filters to identify high dimensional spatial fields from sparse measurements), not the number of observations (as may appear when using GP in machine learning, for instance to perform regression in relatively small feature spaces but with high volumes of data).

5 | RESULTS PART I: EXAMPLE PROBLEM DEFINITION AND APPLICATION OF FULL RANK GP BASED ERROR FIELD RECOVERY

5.1 | Example problem definition

We begin by illustrating the GP enhanced GOEE methodology developed above through a simple sub-modeling problem. We here consider a 2D (plane stress) "perforated L" plate, clamped along the lower edge and loaded in the vertical direction by a uniformly distributed load (see Figure 2A). Note that the node at the lower left hand corner is fixed along the horizontal direction in order to prevent rigid body motion. The deformed (unity scale factor) and undeformed mesh for the primal problem is presented in Figure 2B.

We consider the situation presented in Figure 2 as our global problem and introduce local sub-models at two locations (model interfaces are shown in blue and red in Figure 3).

The quantities of interest are defined as $\Delta Q_i = \int_\Omega \psi_i \cdot \hat{e}_h \, d\Omega = \mathbf{c}^{iT}\mathbf{M}\hat{\mathbf{e}}$ where indexes i's correspond to either one of, or both of, the two DoFs (displacement in "x" and/or displacement in "y") associated with sparsely selected nodes located on the boundary of the sub-models. QoIs therefore represent a linear combination of displacements. Note that the information gathered by these QoIs is the information needed to assess errors in driving DoFs in the multiscale setting. For the meshes presented in Figure 3, the "all duals" case (i.e., the situation in which all adjoint problems corresponding to the boundary of the RoI are solved). Noting that, in the present work, we consider a single adjoint centered at each RoI boundary node that is a linear combination of DoFs (see above), the "all duals" case amounts to solving an additional 51 linear problems (note the shared node in the red and blue RoI boundaries). As mentioned previously, only forward and backward substitution, followed by the ZZ gradient recovery are required, as the Cholesky factor has already been computed when solving the direct FE problem.
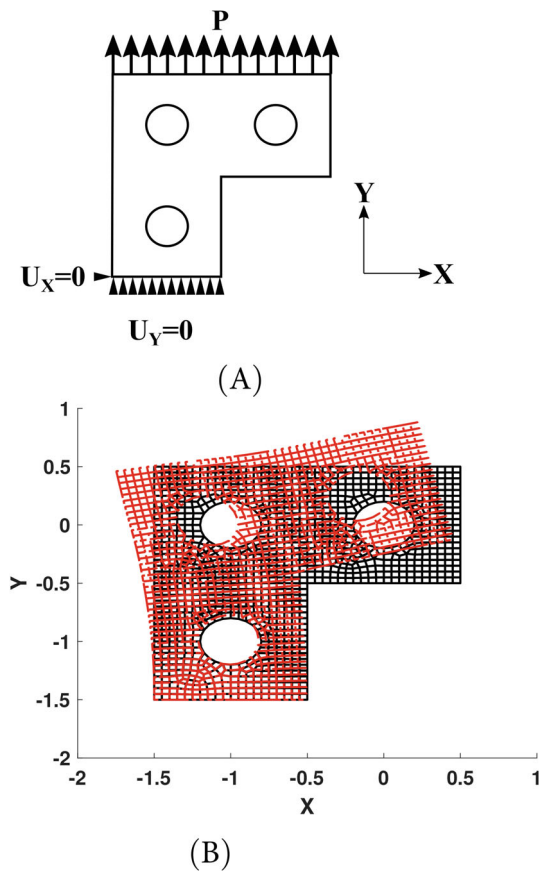
FIGURE 2 The example global FE problem, showing (A) boundary conditions and loads, and (B) the deformed (red) and undeformed (black) mesh (with a unity deformation scaling factor).
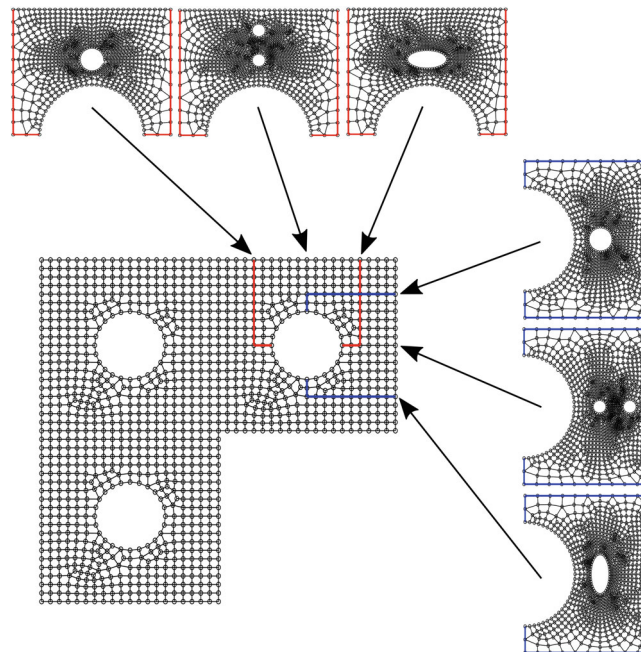

FIGURE 3 The example multiscale problem considered in the present work. Two potential RoI boundaries are defined (in red and blue), within which a multitude of local models (that contain local features) may be placed. We wish to estimate the global scale error over the sub-models boundaries, and propagate the error down to the local models.

In the present work, error propagation is only performed using the "two circular holes" local model represented with the blue boundary in Figure 3. Error fields are recovered over the entire global model domain however.

5.2 | Full rank Bayesian recovery of error fields

We present here the recovery of error fields over the entire global model domain using the methods outlined above.

In the following, we randomly select a number of adjoints from the 51 available, as described in the previous section, select a value for the hyperparameter α (see Equation 34) such that the likelihood given by Equation (36) is maximized, evaluate the prior and posterior, and sample form the resultant to propagate error though the model scales.

In the following, the hyperparameter β is taken as 1 in all cases, being a length scale on the order of the computational domain itself. Given this fixing of β, the term "optimization" is avoided when discussing the selection of α. Clearly, multivariate optimization schemes could be implemented to maximize likelihood. The present work features several compounding methods that address the core error field recovery problem described in the introduction. To limit the complexity of the present work, discussions on true hyperparameter optimization have been omitted. Future work will look to include these processes as part of developing a complete multiscale tool that can be deployed in large simulation problems.

With the above in mind, in the present work, selection of an α value is achieved by fitting a polynomial to evaluations of likelihood (36) for a range of trial α values and then searching for a stationary point in the tested range. The range of tested α values used here is 3.5×10^{-2} to 3.5×10^{-6}. The polynomial fitting approach described here can be implemented in any case where training data (GOEEs) are provided. Note that, in cases where we wish to draw samples from the prior alone (i.e., where there is no data to influence hyperparameter value determination), values of $\beta = 1$ and $\alpha = 3.5 \times 10^{-4}$.

It is important to note that some form of hyperparameter selection is required as more training information is added, as the use of a single set of hyperparameters (regardless of number of observations) infers some form of prior knowledge that fixes the amplitude of the prior.

We may sample from multivariate normal distributions ($\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$) using standard methods, such as the Monte-Carlo method described in Section 4.2.4. Note that, in the context of the present work, each sampling $\boldsymbol{s}$ (see Equation 37) is a realization of the error field $\hat{\mathbf{e}}$. Figure 4 presents 4 such samplings, all developed using the same $\tilde{\mathbf{d}}$ noise vector (shown in Figure 4A). By using the same $\tilde{\mathbf{d}}$ in all samplings, the influence of number of training adjoints (observations) can be visualized. Figure 4B shows sampling performed on the prior, Figure 4C shows sampling performed on a 2 adjoint trained posterior, Figure 4D shows sampling performed on a 16 adjoint trained posterior, and Figure 4E shows sampling performed on a posterior trained by all (51) adjoints. Note also that adjoint training sets are subsets of one another (that is to say, the 2 adjoints selected to train that particular posterior are also found in the 16 used to train that particular posterior). The effect of this restriction is that features which manifest in the posterior from the inclusion of a particular observation are preserved as more training data is included. Figure 4B–E clearly show that, as the number of observations is increased, the posterior tends to the case where all adjoints are considered (as expected, see Figure 4E). We can also see the influence of the Dirichlet boundary conditions in all samples (clearly, we know these when generating the prior, hence variance in this neighborhood is greatly reduced). Note also that, as we move away from the RoI boundaries (and hence observations), we tend towards the prior distribution (again as expected).

The influence of uncertainty at the RoI boundary (and the effect of number of QoI observations on it) can be further visualized by considering stress distributions over the entire the RoI. We consider the cases where 2, 16, and all available adjoints are used to train the posteriors. Monte-Carlo samplings are taken from each posterior and are used to calculate stress states in the local model (as above). Distributions are then fitted to the results on an integration point by integration point basis, allowing for the determination of the mean von Mises stress ($Mean_{VM}$) and its standard deviation (Std_{VM}) over the local model. Results for the 2, 16, and all available adjoints training cases are shown in Figures 5–7, respectively. We may observe that, in these results, the number of training adjoints has little influence on the mean of the stress field. This is of course to be expected as this field will be mainly governed by the global primal solution (which is common to all realizations). Similar features are observed in all standard deviation fields (the largest deviations are located around stress concentrations), however the magnitude of the standard deviations decreases with an increasing number of training adjoints (the standard deviation of the 2 adjoint training cases is approximately 3 times greater than that of the all adjoints training case). This can be rationalized as a transition from the uncertainty associated with the prior (which is relatively large and is recovered in the low number of training adjoints cases) to the uncertainty associated with the noise in our
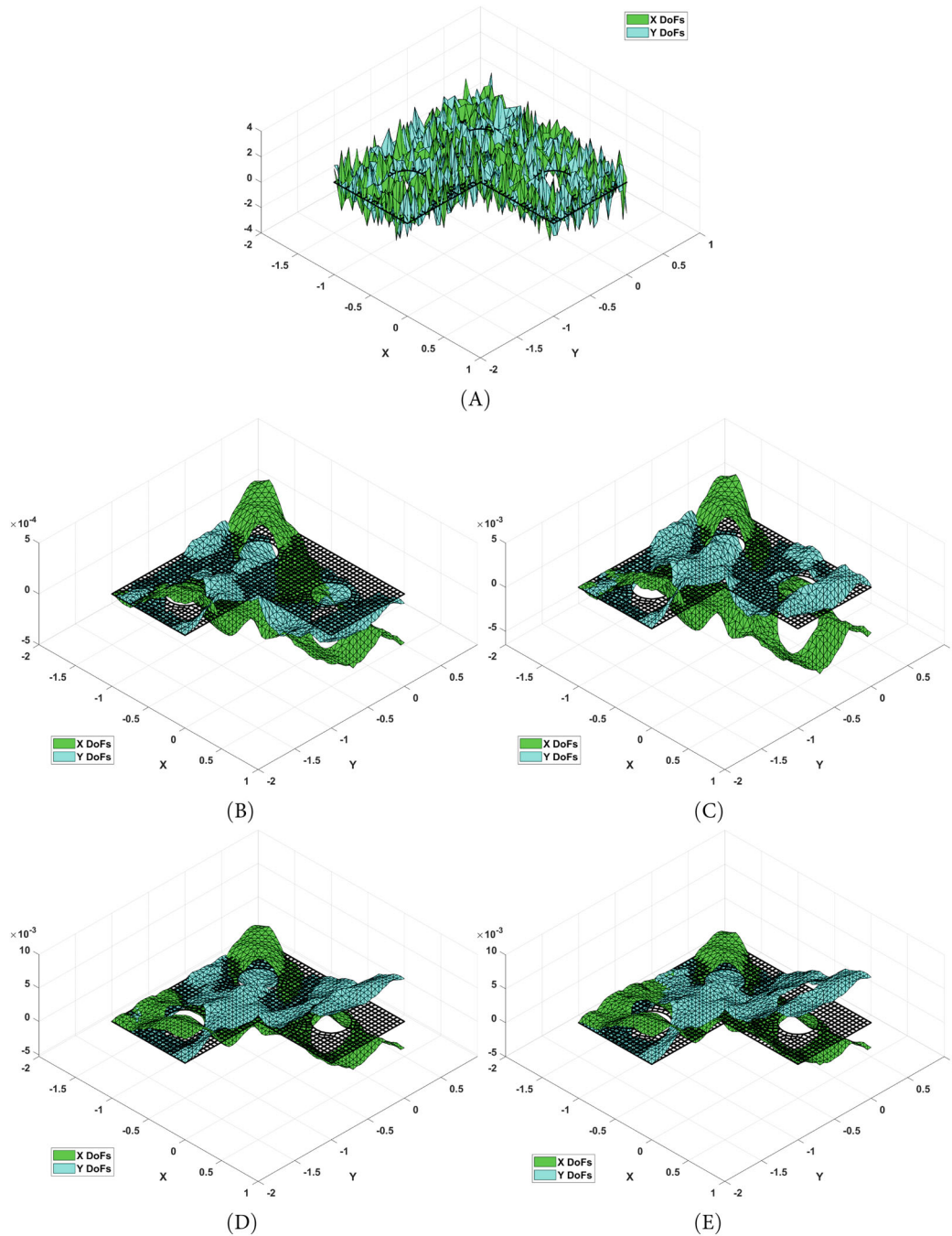
FIGURE 4 A sample from GOEE random field distributions, showing convergence of the solutions with increasing numbers of training data sets (ZZ estimates in QoI). Note the same noise vector, shown in (A), is used in all cases. (B–E) show, respectively, a sample from the SPDE prior for the error field, a sample from the posterior trained using 2 randomly selected QoIs, a posterior trained using 16 randomly selected QoIs, and a posterior trained using all the adjoint problems corresponding to DOFS located on the RoI boundaries.

error observations (which is propagated in the all training adjoint case). Observations regarding the development of the mean and standard deviation of the von Mises stress fields can be further demonstrated by considering only the peak von Mises stresses in the local model and noting how the resulting distributions vary with number of training adjoints. This is presented in Figure 8, wherein we again see a preservation of the mean and a reduction in the standard deviation of the peak von Mises stress as we increase the number of training adjoints (from 2 to all available). Note that the standard deviation of the 32 training adjoints and all training adjoints distributions are near identical, suggesting that there is limited utility in calculating the additional QoIs.
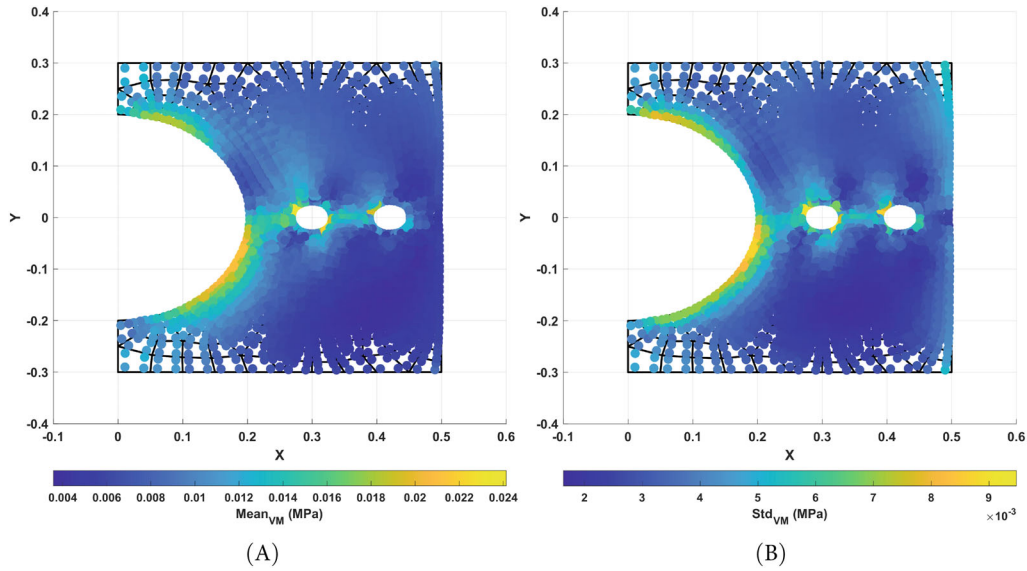
FIGURE 5 Monte-Carlo propagation of the coarse scale FE error to the "2 hole" (position 2) sub-model, showing (A) the mean von Mises stress field and (B) the standard deviation of the von Mises stress field. Results are based on a posterior trained using two randomly selected QoIs corresponding to DoFs located on the RoI boundaries.
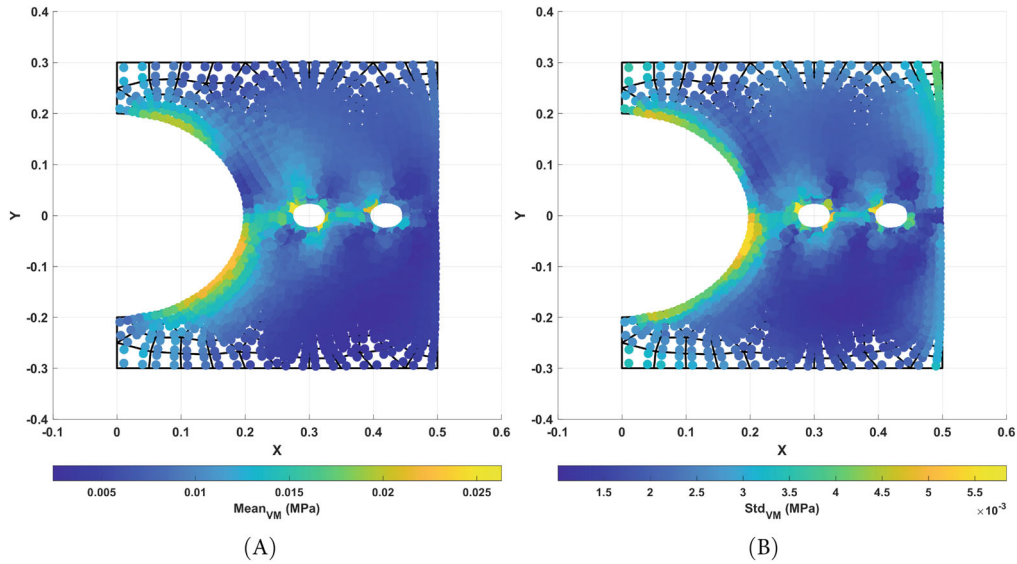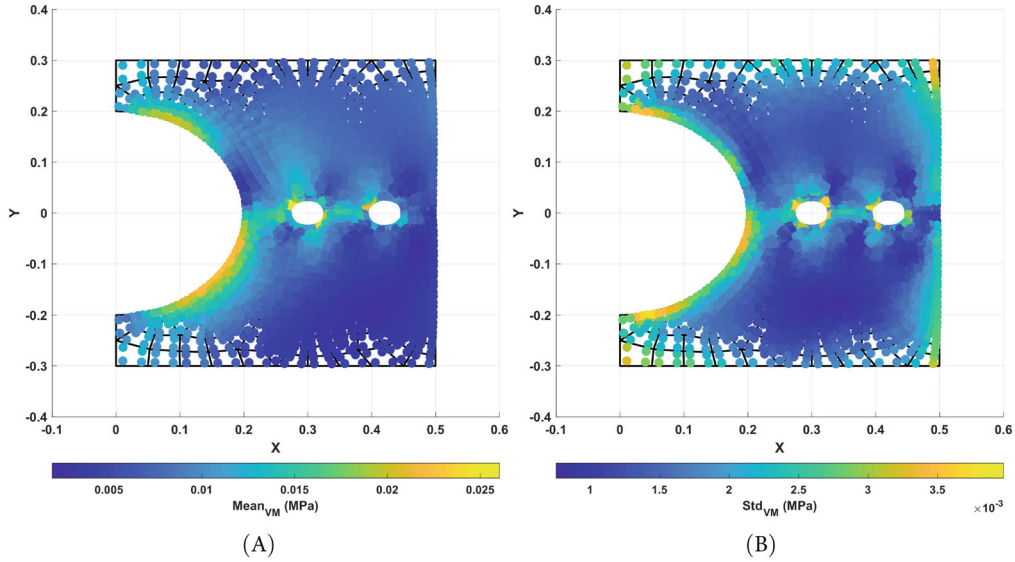

FIGURE 6 Monte-Carlo error field propagation for the "2 hole" (position 2) sub-model, showing (A) the mean field and (B) the standard deviation field. Results are based on a posterior trained using 16 randomly selected QoIs corresponding to DoFs located on the RoI boundaries.

6 | LOW RANK GAUSSIAN PROCESS BY GALERKIN MODEL ORDER REDUCTION

6.1 | Stochastic reduced basis approximation

We now seek to address the computational bottleneck identified in Section 4.3 via the projection of the GP into a low dimensional subspace. We will look for a reduced basis approximation of the SPDE prior distribution for $\hat{\mathbf{e}}$. The corresponding approximation reads as

$$\hat{\mathbf{e}} \approx \tilde{\mathbf{e}} = \boldsymbol{\phi}\boldsymbol{\gamma}, \tag{44}$$

FIGURE 7 Monte-Carlo error field propagation for the "2 hole" (position 2) sub-model, showing (A) the mean field and (B) the standard deviation field. Results are based on a posterior trained using all the adjoint problems corresponding to DOFS located on the RoI boundaries.
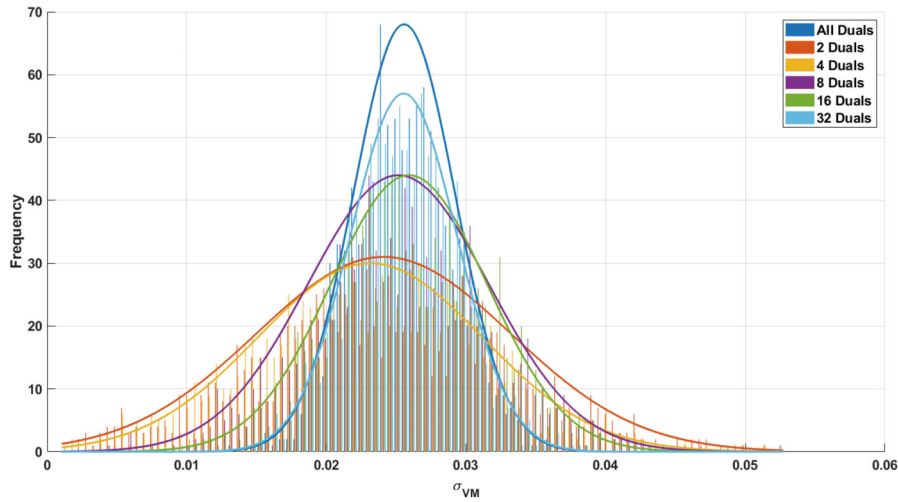

FIGURE 8 Peak von Mises equivalent stresses (σ_{VM}), computed using Monte-Carlo posterior sampling of the global error field and propagated to the "2 hole" (position 2) sub-model. The sub-model is driven by DoF fields modified by posterior distributions trained using varying numbers of QoIs.

where the n_ϕ columns of $\boldsymbol{\phi}$ are the deterministic reduced bases. Random vector $\boldsymbol{\gamma}$ may be optimally computed using Galerkin's principle[44-46]

$$\forall\, \mathbf{b} \in \mathbb{R}^N, \quad \boldsymbol{\phi}^T \left(\mathbf{b} - \mathbf{A}\tilde{\mathbf{e}} \right) = \mathbf{0}. \tag{45}$$

We find that

$$\tilde{\mathbf{e}} = \boldsymbol{\phi} \left(\boldsymbol{\phi}^T \mathbf{A} \boldsymbol{\phi} \right)^{-1} \boldsymbol{\phi}^T \mathbf{b} . \tag{46}$$

Therefore, the approximate prior reads as

$$\tilde{\mathbf{e}} \sim \mathcal{N} \left(\mathbf{0}, \tilde{\boldsymbol{\Sigma}} := \boldsymbol{\phi} \boldsymbol{\Sigma}_\gamma \boldsymbol{\phi}^T \right), \tag{47}$$

where the covariance matrix for the n_ϕ reduced random variables γ is

$$\Sigma_\gamma = \alpha^2 \phi \left(\phi^T (\mathbf{M} + \beta^2 \mathbf{K}) \phi \right)^{-1} \phi^T \mathbf{M} \phi \left(\phi^T (\mathbf{M} + \beta^2 \mathbf{K}) \phi \right)^{-1} \phi^T. \tag{48}$$

The approximate prior covariance $\tilde{\Sigma}$ given in the expression above is rank deficient. The systems of equations involved in computing the posterior mean and co-variance are now reduced to n_ϕ equations, with fully populated system matrices if the reduced basis vectors correspond FE functions that are not locally supported (i.e., as opposed to locally supported "hat" functions).

6.1.1 | Reduction of computational operations

The low-rank approximation can be taken advantage of when computing posterior distributions, typically by making use of the Sherman–Morrison formula. These computational tricks are classical and therefore not detailed in the article (the interested reader may refer, for instance, to the usual implementation of the ensemble Kalman filter[47]).

6.2 | Projection subspaces

6.2.1 | Karhunen–Loève transform

If the basis vectors are made of all the eigenvectors of the full covariance matrix, that is,

$$\forall\, i = 1 \,\ldots\, N, \qquad \mathbf{K}\phi_i = \lambda_i \mathbf{M}\phi_i \quad \text{and} \quad \forall\, j \neq i, \quad \phi_i^T \phi_j = \delta_{ij}, \tag{49}$$

then the previous derivation is the Karhunen–Loève transform: the random coefficients are uncorrelated (and independent in the Gaussian case) and the covariance matrix Σ_γ is diagonal. More precisely, we have

$$\tilde{\Sigma} = \alpha^2 \phi \Lambda_m (\Lambda_m + \beta^2 \Lambda_k)^{-2} \phi^T, \tag{50}$$

where Λ_m and Λ_k are diagonal matrices with ith diagonal entry equal to $\phi_i^T \mathbf{M} \phi_i$ and $\phi_i^T \mathbf{K} \phi_i$, respectively. The Karhunen–Loeve transform provides a natural method to truncate reduced basis expansions, but requires computing the spectrum of the discrete Laplacian operator. The associated computational cost is not acceptable in the present context.

6.2.2 | Subspace generated by the set of precomputed adjoint solutions

What can we choose as good projection subspace? An obvious first choice is to reuse what is already at our disposal: the set of all adjoint states $\mathbf{z}_i$ defined by

$$\forall i \in [\![1, M]\!], \qquad \mathbf{K}^T \mathbf{z}^i = \mathbf{Q}^i, \tag{51}$$

where $\mathbf{Q}_i^T$ is the ith line of observation operator $\mathbf{H}$. These linear systems of equations have been solved to compute the GOEE corresponding to each component of data vector $\mathbf{d}$. We will store concatenation $\mathbf{Z} = (\mathbf{z}^1 \; \mathbf{z}^2 \; \ldots \; \mathbf{z}^M) \in \mathbb{R}^{N \times M}$.

6.2.3 | RBF enrichment and regularization

Additional reduced basis functions are required to let the GP express uncertainty. Several options have been explored during the early phase of this work. In particular, we have successfully used coarse functions generated by Algebraic Multigrid Solvers[48] as reduced basis vectors. In this article, we report results related to the use of a set of analytically defined and numerically optimized radial basis functions (RBFs). This will be fully detailed in the following chapter. For now, we assume that these n_b functions are represented in the FE space, and can be encoded by their nodal values, which are stored in matrix $\mathbf{B} \in \mathbb{R}^{N \times n_b}$.

Finally, the set of adjoints and RBFs are concatenated as follows:

$$\boldsymbol{\phi} = (\mathbf{B} \ \mathbf{Z}) \in \mathbb{R}^{N \times n_\Phi}. \tag{52}$$

To avoid the potential (near-)linear dependencies of the columns of $\boldsymbol{\phi}$, an SVD of $\boldsymbol{\phi}$ is computed, and $\boldsymbol{\phi}$ is replaced by the matrix whose columns are the left singular vectors that are associated with non-vanishing singular values up to a numerical tolerance.

Remark

Note that only using $\mathbf{B}$, that is, discarding the adjoint vectors when building the reduced basis, would result in a sub-optimal strategy with undesired properties. Indeed, including the adjoint vectors in the reduced basis ensures that the posterior density of the QoIs measured using the ZZ estimates is exactly reproduced by the low rank approximation of the GP strategy. The proof is as follows. Starting from the Galerkin orthogonality, with $\mathbf{A} = \mathbf{M} + \beta^2 \mathbf{K}$ and $\mathbf{b}$ a zero-mean Gaussian distributed vector with covariance matrix $\alpha^2 \mathbf{M}$, we have

$$\mathbf{Z}^T (\mathbf{b} - \mathbf{A}\tilde{\mathbf{e}}) = \mathbf{0}, \tag{53}$$

$$\mathbf{H}\mathbf{A}^{-T} (\mathbf{b} - \mathbf{A}\tilde{\mathbf{e}}) = \mathbf{0}, \tag{54}$$

and therefore, the announced property reads as

$$\mathbf{H} (\hat{\mathbf{e}} - \tilde{\mathbf{e}}) = \mathbf{0}. \tag{55}$$

6.3 | Accuracy measure

The discrepancy between the exact error vector and its low rank approximation is a random variable defined by

$$\boldsymbol{\Delta e} = \hat{\mathbf{e}} - \tilde{\mathbf{e}} = \mathbf{A}^{-1} \left(\mathbf{I}_d - \mathbf{A}\boldsymbol{\phi}(\boldsymbol{\phi}^T\mathbf{A}\boldsymbol{\phi})^{-1}\boldsymbol{\phi}^T \right) \mathbf{b}. \tag{56}$$

Therefore, the prior distribution of discrepancies in errors in predicted QoIs $\mathbf{E}\hat{\mathbf{e}}$ is given by

$$\mathbf{E}\boldsymbol{\Delta e} \sim \mathcal{N} \left(\mathbf{0}, \boldsymbol{\Sigma}_{\Delta E} := \mathbf{E}\boldsymbol{\Sigma}_\Delta \mathbf{E}^T \right) \tag{57}$$

with

$$\boldsymbol{\Sigma}_\Delta = \alpha^2 \left(\mathbf{A}^{-1} \left(\mathbf{I}_d - \mathbf{A}\boldsymbol{\phi}(\boldsymbol{\phi}^T\mathbf{A}\boldsymbol{\phi})^{-1}\boldsymbol{\phi}^T \right) \right) \mathbf{M} \left(\left(\mathbf{I}_d - \boldsymbol{\phi}(\boldsymbol{\phi}^T\mathbf{A}\boldsymbol{\phi})^{-1}\boldsymbol{\phi}^T\mathbf{A}^T \right) \mathbf{A}^{-1} \right). \tag{58}$$

The computation of $\mathbf{E}\boldsymbol{\Sigma}_\Delta$ is not tractable, as we would need to compute $\mathbf{A}^{-1}\mathbf{E}^T$, which is a matrix whose columns are made of desired adjoint solutions, the computation of which is precisely what we aim to avoid by using GP functional regression. However, choosing a selected few lines of $\mathbf{E}$ can provide valuable goal-oriented estimates of the accuracy of the low rank approximation to the full GP, which we will make use of later on.

7 | ADAPTIVE RADIAL BASIS FUNCTIONS USING STOCHASTIC ADJOINT BASED OPTIMIZATION AND HIERARCHICAL CLUSTERING

7.1 | Reduced basis vectors as finite element interpolants of RBFs

We propose to construct the reduced basis contained in $\mathbf{B}$ as the FE nodal values corresponding to analytically defined radial bases functions. To this end, let us define

$$\forall i \in [\![1, n_b]\!], \ \forall x \in \Omega, \quad \zeta^i(x) = \omega^{\lceil i/d \rceil} \exp\left(-\frac{1}{2} \frac{\left(x - c^{\lceil i/d \rceil}\right)^2}{l^{\lceil i/d \rceil 2}} \right) e_{i - (\lceil i/d \rceil - 1)d}, \tag{59}$$

$$\zeta^i(x) = \eta^{\lceil i/d \rceil}(x)\, e_{i-(\lceil i/d \rceil - 1)d}, \tag{60}$$

and its FE interpolant

$$\zeta_h^i(x) = \mathcal{I}_h \zeta^i(x), \tag{61}$$

which may be represented by basis vector $\gamma^i \in \mathbb{R}^N$, the component of which are such that

$$\forall\, i \in [\![1, n_b]\!], \ \forall\, x \in \Omega, \quad \zeta_h^i(x) = \sum_{j=1}^N \psi_j(x)\gamma_j^i. \tag{62}$$

We assume in these developments that n_b is a multiple of d. Set $\{c_i\}_{i \in [\![1, n_b/d]\!]} \in \left(\mathbb{R}^d\right)^{n_b/d}$ is composed of the n_b/d centers of the set of radial bases functions $\{\eta_i\}_{i \in [\![1, n_b/d]\!]} \in \mathbb{R}^{n_b/d}$. Set $\{\omega_i\}_{i \in [\![1, n_b/d]\!]} \in \mathbb{R}^{n_b/d}$ is a set of weights that allows the approximation to reproduce constant vectors, and $\{l_i\}_{i \in [\![1, n_b/d]\!]} \in \mathbb{R}^{n_b/d}$ is the set of length scales associated with the RBFs. Sets $\{c_i\}_{i \in [\![1, n_b/d]\!]}$, $\{\omega_i\}_{i \in [\![1, n_b/d]\!]}$ and $\{l_i\}_{i \in [\![1, n_b/d]\!]}$ are chosen as follows.

- For any $i \in [\![1, n_b/d]\!]$, we define l_i as the distance to the N_lth closest point of set $\{c_i\}_{i \in [\![1, n_b/d]\!]}$, with $N_l \in \mathbb{R}^{+*} \setminus \{1\}$ and l a parameter of the approach, in the sense of the standard Euclidean distance. This technique ensures that the set of RBFs are overlapping to a satisfying degree.
- We define weight set $\{\omega_i\}_{i \in [\![1, n_b/d]\!]}$ by

$$\forall\, i \in [\![1, n_b/d]\!], \quad \omega_i(x) = \frac{1}{\sum_{i=1}^{n_b/d} \eta^i(x)}, \tag{63}$$

  which ensures that the RBFs realize a partition of unity.

- The set of centers $C = \{c_i\}_{i \in [\![1, n_b/d]\!]}$ is constructed adaptively, using a goal-oriented (divisive) hierarchical clustering approach, as described in details in the next section.

  To initialize the adaptation process, a standard coarse clustering of the vertices of FE tessellation $\mathcal{T}_h$ is first performed. We use the agglomerative hierarchical clustering tree implemented in Matlab to partition the set of vertices of the mesh into n_b^0/d. Then, each element of set $C^0 = \{c_i^0\}_{i \in [\![1, n_b^0/d]\!]}$ is computed as the centroid of the ith cluster of the corresponding mesh vertices.

Once the n_b/d radial basis functions are constructed, we can build projection matrix $\mathbf{B}$ by concatenation:

$$\mathbf{B} = (\gamma^1\ \gamma^2\ \dots\ \gamma^{n_b}). \tag{64}$$

7.2 | Stochastic adjoint weighted residual approach

In this section, we provide details about the algorithm that we developed to refine the initial RBF center set $C^0 = \{c_i^0\}_{i \in [\![1, n_b^0/d]\!]}$. This id done in a greedy, goal-oriented manner. This means that RBF centers are only added during the refinement process (i.e., not moved nor removed, although this would be an interesting extension), and that this will be done such that the discrepancy measure is as approximately minimized for the set of all QoIs for which an estimate of error is desired. As described previously, in the sub-modeling case, we wish to quantify the error in the $\mathcal{L}_2$ projection of the entire displacement field over the boundary of the RoI.

To achieve this goal, we first define the following error estimate

$$\nu = \sum_{i=1}^{M^\star} \mathrm{Var}\left(\left(\mathbf{Q}^{i^\star}\right)^T \Delta \mathbf{e}\right), \tag{65}$$

where $\{\mathbf{Q}^{i^\star}\}_{i \in [\![i\ M^\star]\!]}$ are an arbitrary set of extractors of scalar QoIs. As mentioned above, we will use a random selection of the lines of prediction operator $\mathbf{H}$ (i.e., QoIs for which we wish to estimate the accuracy using the GP based

functional regression approach, which may or may not have been estimated as part of the M ZZ goal-oriented error estimations).

We note that

$$\left(\mathbf{Q}^{i\star}\right)^{T}\Delta\mathbf{e} = \left(\mathbf{Q}^{i\star}\right)^{T}\left(\mathbf{A}^{-1}\mathbf{r}\right) = \left(\mathbf{z}^{i\star}\right)^{T}\mathbf{r}, \tag{66}$$

where i^{th} adjoint solution $\mathbf{z}^{i\star}$ is such that

$$\mathbf{A}^{T}\mathbf{z}^{i\star} = \mathbf{Q}^{i\star}, \tag{67}$$

and the random residual vector $\mathbf{r} := \mathbf{b} - \mathbf{A}\tilde{\mathbf{e}}$ has expression

$$\mathbf{r} = \left(\mathbf{I} - \mathbf{A}\boldsymbol{\phi}\left(\boldsymbol{\phi}^{T}\mathbf{A}\boldsymbol{\phi}\right)^{-1}\boldsymbol{\phi}^{T}\right)\mathbf{b} =: \boldsymbol{\xi}\mathbf{b}. \tag{68}$$

Therefore, $\mathbf{r}$ is a zero mean Gaussian vector with covariance $\text{Cov}(\mathbf{r}) = \boldsymbol{\xi}\mathbf{M}\boldsymbol{\xi}^{T}$. Hence, (65) reads as

$$\nu = \sum_{i=1}^{M^{\star}}\mathbf{z}^{i\star\,T}\text{Cov}(\mathbf{r})\mathbf{z}^{i\star}. \tag{69}$$

Equation (66) is usually coined dual weighted residual in the error estimation literature,[16,26] albeit usually employed in a deterministic setting. In order to make this expression small, and owing to Galerkin orthogonality, the suggestion is to use the adjoints $\{\mathbf{z}^{i\star}\}_{i\in\llbracket i\ M^{\star}\rrbracket}$ as reduced basis vectors. They are, of course unavailable and we can only try and produce a similar effect by using a tailored selection algorithm for the RBF centers.

Extending classical error estimation procedures, we use Galerkin orthogonality to reveal that (69) reads as

$$\nu = \sum_{i=1}^{M^{\star}}\left(\mathbf{z}^{i\star} - \Pi_{\boldsymbol{\Phi}}\mathbf{z}^{i\star}\right)^{T}\text{Cov}(\mathbf{r})\left(\mathbf{z}^{i\star} - \Pi_{\boldsymbol{\Phi}}\mathbf{z}^{i\star}\right), \tag{70}$$

where $\Pi_{\boldsymbol{\Phi}} = \boldsymbol{\Phi}\left(\boldsymbol{\Phi}^{T}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{T}$ is the norm-two projector onto span($\boldsymbol{\Phi}$), which we may further write as

$$\nu = \sum_{i=1}^{M^{\star}}\left(\mathbf{w}^{i}\right)^{T}\mathbf{y}^{i} \tag{71}$$

with $\mathbf{w}^{i} := \mathbf{z}^{i\star} - \Pi_{\boldsymbol{\Phi}}\mathbf{z}^{i\star}$ and $\mathbf{y}^{i} := \text{Cov}(\mathbf{r})\mathbf{w}^{i}$. Equation (71) can be seen as an extension of the classical dual weighted residual framework to discrete stochastic Galerkin approximations.

As per usual in dual weighed residual approaches, we may interpret the previous expression as a discrete error density, as

$$\nu = \sum_{i=1}^{M^{\star}}\sum_{j=1}^{N}\mathbf{w}_{j}^{i}\mathbf{y}_{j}^{i} = \sum_{j=1}^{N}\left(\sum_{i=1}^{M^{\star}}\mathbf{w}_{j}^{i}\mathbf{y}_{j}^{i}\right) = \sum_{j=1}^{N}\left(\sum_{i=1}^{M^{\star}}|\mathbf{w}_{j}^{i}\mathbf{y}_{j}^{i}|\right). \tag{72}$$

We will finally select the radial basis function centers such that approximately minimize ν, that is, our chosen measure of the distance between exact and low rank Gaussian priors for all QoIs. This can be done by cancelling large elements of set $\left\{\sum_{i=1}^{M^{\star}}\left|\mathbf{w}_{j}^{i}\mathbf{y}_{j}^{i}\right|\right\}_{j\in\llbracket 1\ N\rrbracket}$, thereby making sure that for DoFs corresponding to large values of projected dual vectors $\{\mathbf{w}^{i}\}_{i\in\llbracket 1\ M^{\star}\rrbracket}$, the associated components of "residual" vector $\{\mathbf{y}^{i}\}_{i\in\llbracket 1\ M\rrbracket}$ are small (this is the central idea of the dual weighted residual approach[16,26]). This implies refining the projection space in those areas, following Galerkin's principle.

7.3 │ Goal-oriented hierarchical clustering

We propose to use the discrete GOEE density described previously to guide the hierarchical refinement of the hierarchical clustering algorithm described in Section 7.1.

This is done as follows. Given the current clustering of the mesh vertices at iteration k of the RBF construction process, which is associated with centroid set $C^k = \{c_i^k\}_{i \in [\![1, n_b^k/d]\!]}$, we compute, for every cluster index $l \in [\![1 \; n_c^l]\!]$

$$q_l^k = \sum_{j=1}^{n_c^l} \sum_{i=1}^{M^\star} \left| \mathbf{w}_{\mathcal{I}^l(j)}^{i,k} \mathbf{y}_{\mathcal{I}^l(j)}^{i,k} \right|, \tag{73}$$

where n_c^l is d times the number of vertices of cluster l, and $\mathcal{I}^l$ is an index map between the cluster wise and global numbering of FE DoFs. In words, q_l^k is the contribution to the GOEE metric of cluster l at iteration k of the construction algorithm.

Then, we set cluster $l^\star$ for refinement, with

$$l^\star = \arg\max_l \; (q_l^k). \tag{74}$$

This cluster is subdivided into n_r sub-clusters using K-means, the centroids of which will define new RBF centers, yielding the updated center set $C^{k+1} = \{c_i^{k+1}\}_{i \in [\![1, n_b^{k+1}/d]\!]}$ with $n_b^{k+1} = n_b^k + n_r$.

7.4 | Rank selection by holdout/cross validation

To make sure that the distribution of reduced rank GP $\mathbf{E}\tilde{e}$ is close enough to "truth" distribution of $\mathbf{E}\hat{e}$, we will monitor the size of $\mathbf{\Sigma}_{\Delta E}$ through the greedy construction process. The RBF selection algorithm stops if assertion

$$\sqrt{\frac{1}{M^{\star\star}} \sum_{i=1}^{M^{\star\star}} \mathrm{Var}\left(\left(\mathbf{Q}^{i^{\star\star}}\right)^T \Delta \mathbf{e} \right)} \leq \xi \sqrt{\frac{1}{M^{\star\star}} \sum_{i=1}^{M^{\star\star}} \mathrm{Var}\left(\left(\mathbf{Q}^{i^{\star\star}}\right)^T \tilde{\mathbf{e}} \right)} \tag{75}$$

with tolerance $0 < \xi \ll 1$, is satisfied. In the previous equation, the ** notation indicates that we use a new set of randomly selected lines of $\mathbf{E}$ to populate QoI extractors $\{\mathbf{Q}^{i^{\star\star}}\}_{i \in [\![i \; M^{\star\star}]\!]}$. Indeed, there exists a severe risk of over fitting if using set $\{\mathbf{Q}^{i^\star}\}_{i \in [\![i \; M^\star]\!]}$ to build the reduced basis *and* monitor convergence. Unfortunately, this requires the computation of additional adjoint vectors. Alternatively, one could cross validate the previous methodology, thereby avoiding this additional computational effort at the cost of a somewhat more complex implementation.

7.5 | Reduction in computational complexity

Sampling from the approximate posterior density using the proposed reduced order modeling scheme only involves solving linear systems of size n_ϕ instead of N. However, $M^\star$ linear systems of equations with system matrix $\left(\mathbf{M} + \beta^2 \mathbf{K}\right)$ have to be solved to build the goal-oriented error density that drives the refinement process of the RBF basis, as per Equation (67).

8 | RESULTS PART II: LOW RANK APPROXIMATION FOR THE GP BASED ERROR FIELD RECOVERY

We now implement the goal-oriented hierarchical clustering approach described above to the example problem outlined in Section 5.1. In the following, we consider a randomly selected 25 training adjoints in all cases and look to train a posterior based on a low rank approximation of the prior. RBFs define a projection space for the prior. We here use a hierarchical clustering approach to locate RBFs in the solution domain (RBFs are normalized to ensure partition of unity). A series of clusters are presented to demonstrate the refining nature of the hierarchical approach, both at the global and local scales.

Clusters for the example global problem are shown in Figure 9, with "Cluster 1" illustrating the initial cluster arrangement (due to k-medoids) and Clusters 3, 5, and 7 relating to the 3rd, 5th, and 7th refinement, respectively. Cluster refinement is achieved by evaluating the quantity q (as per Equation (73), here referred to as the clustering density) for a
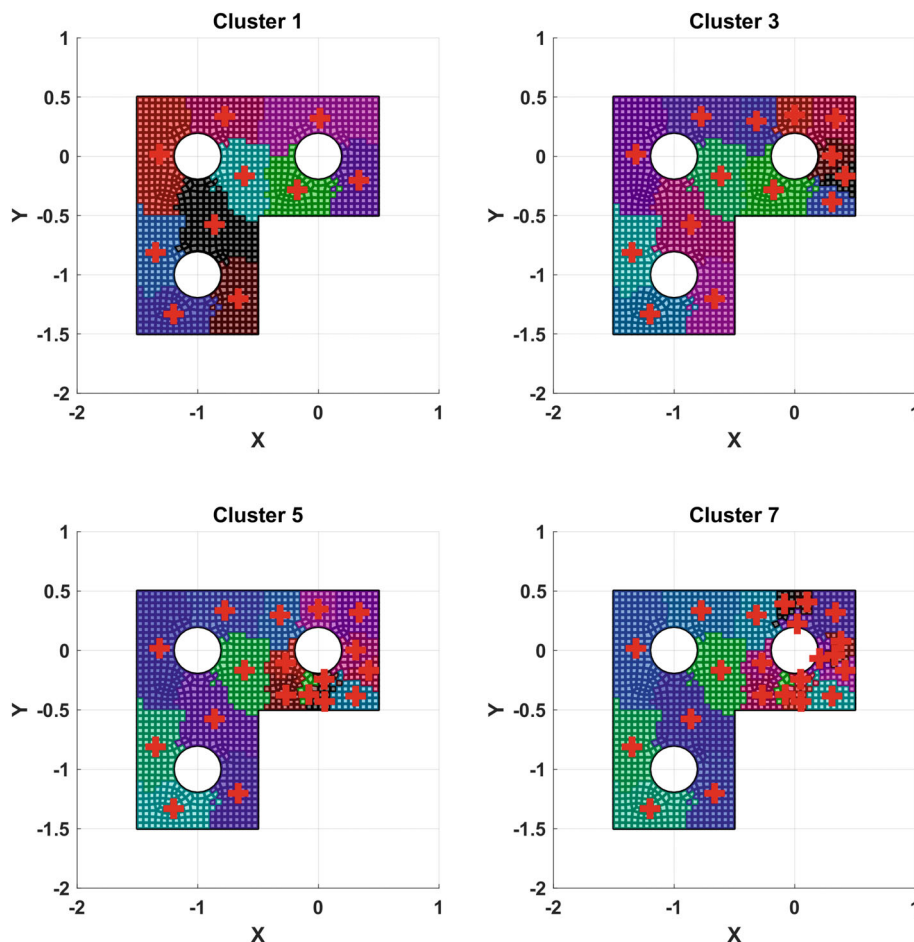
FIGURE 9   Goal-oriented hierarchical clustering using 10 initial clusters. The crosses indicate the centroids of each cluster of vertices. Clusters having the largest contribution to the error discrepancy density are refined into three sub-clusters using standard K-means at each cluster step.

given cluster arrangement (and associated RBF space), noting the cluster within which the weighted summation of q is maximized, and performing k-medoids on the selected cluster. An initial 10 clusters (Cluster 1) are created to begin the hierarchical process, with subsequent refinements partitioning the selected cluster into 3. Example plots of the density q are presented in Figure 10 for the clusters in Figure 9. The example problem considered in the present work limits the location of QoIs to two RoI boundaries (see Figure 3). As a result, the error in low rank QoI estimations (characterized by the density q) is concentrated in the upper right corner of the global model. This can be clearly seen in Figure 10, wherein there is not only the desired reduction in the magnitude of q with cluster refinement but also a concentration of q on the boundaries of clusters. For completeness example RBFs, resulting from the clusters in Figure 9, are presented in Figure 11.

We may perform Monte-Carlo analyses on the local models by sampling error correct DoF distributions that are generated by exact and low rank approximations of the GP terms (similar to the demonstrations performed in Section 5.2). Peak von Mises stress may be extracted from the "2 circle" local model (as before) for the low rank approximations associated with each cluster refinement. Resultant distributions in peak von Mises stress are shown in Figure 12. Note that as cluster refinement progresses we see a general broadening in peak stress distributions, resulting from the greater fidelity in error field approximation offered by higher dimensional projection spaces. Furthermore, we note a refinement of the distribution mean and standard deviation, again resulting from the great degree of variation captured in the refined cluster projections. In the example problem, it is clear that convergence of the subspaces is oscillatory (as opposed to monotonic). It is unclear why this is the case, however we suggest that it is a symptom of the relatively coarse global model mesh. Given the limited number of DoFs driving the local model, it is conceivable that, given some particular projection space, the propagation of certain QoIs over regions that in reality are not controlled by these quantities happens to result in preferable global states. This feature of the low rank approximation is not explored here, however the convergent nature of the process should be noted. This satisfies the original motivation for the goal-oriented projection.
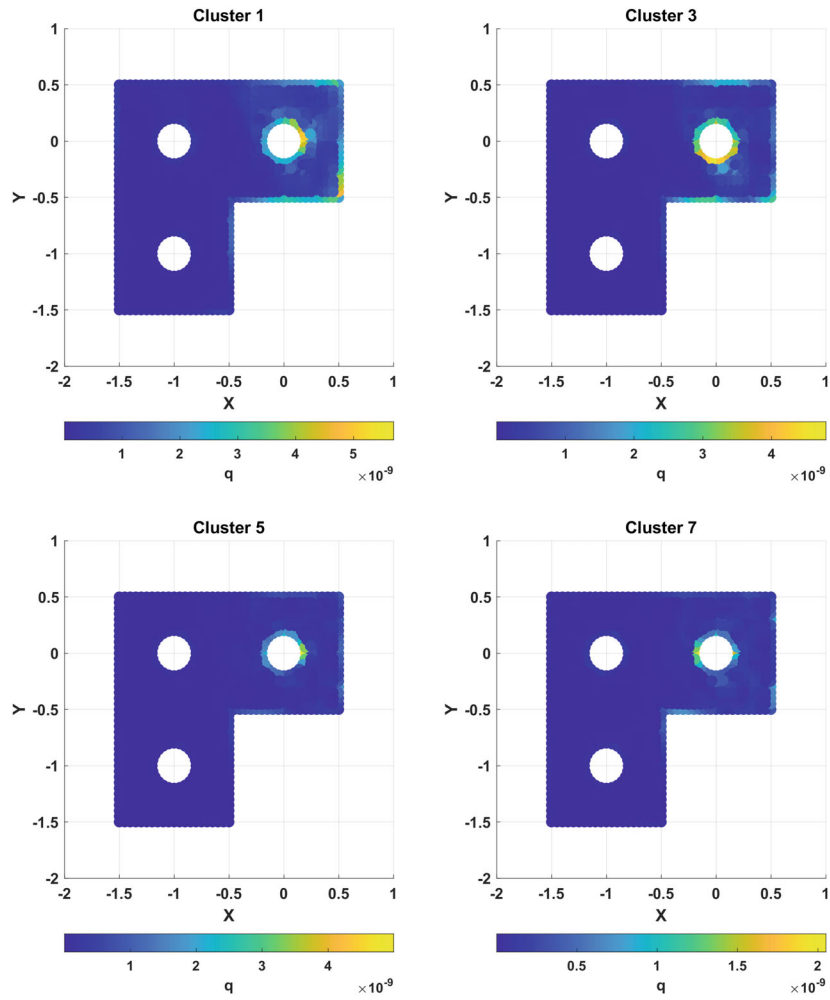
FIGURE 10 Evolution of the clustering density (see Equation 73) during cluster refinement.
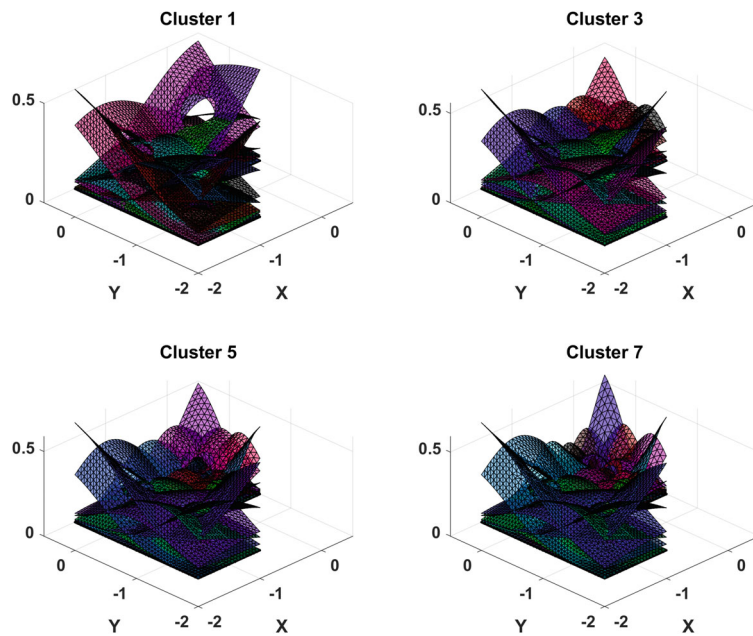

FIGURE 11 Goal-oriented hierarchical clustering using 10 initial clusters, showing the resulting RBFs at various iteration count of the greedy process.
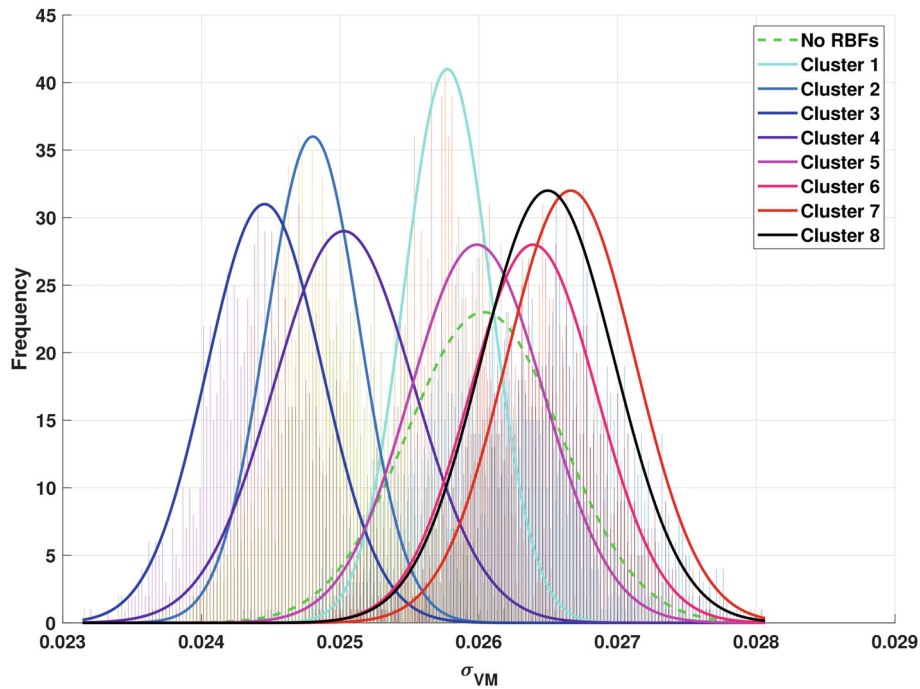
FIGURE 12 Peak von Mises equivalent stresses (σ_{VM}) in the "2 hole" (position 2) sub-model, obtained by Monte-Carlo sampling of the global error field and subsequent propagation to the local model. The GP is trained using ZZ estimates. Projections are made using the cluster based RBF approach (with higher clustering numbers indicating goal-oriented hierarchical refinement). A true "No RBFs" distribution is included for reference (this does not make use of any form of model order reduction to approximate the prior).

9 | PERSPECTIVES AND CONCLUSIONS

A statistical learning approach is presented here for the estimation of full error fields in FE models. A particular emphasis is placed on multiscale applications, where full error field estimation allows for the propagation of errors through the length scales with relative ease. The method is unobtrusive. It does not require any re-meshing operation and is computationally efficient, owing to the low rank SPDE based GP specifically developed in this work. The SPDE representation of the prior error field allows for posterior realizations to be generated using sparse matrix inversions, thereby ensuring the cost of generating such realizations, and that of tuning the GP hyperparameters using maximum marginal likelihood, is comparable to the cost of solving the FE problem itself. It is important to note that the adoption of a SPDE representation of the prior is of little help if done in isolation. It is the sparse approximation of this prior through a guided reduced basis projection that alleviates computational bottlenecks and enables the overall methodology to be computationally affordable.

Although attention is given to classical sub-modeling multiscale problems, the methods are applicable to many alternative cases. For example, consider a large domain where there exist many potential localization features but the analyst does not know which is of the greatest concern. Consider also that, due to some limit on computational overhead, the mesh cannot be refined at all localization features. The error field recovery approach outlined here would furnish the analyst with confidence intervals for QoIs at the localization features, thereby indicating which should be investigated in "dive deeper" exercises. While it is highly unlikely for an analyst to be completely ignorant of potential RoIs, it is very likely in complex systems that many potential regions at the global scale and features at the local scale will be of interest. There is a need, therefore, to extract uncertainty information across the global model such that this can be propagated down to various local models. This observation drives the creation of approaches like the one developed in the present work. A strong constraint on computational cost is industry inspired and motivates the development of an efficient analysis method; a consideration often omitted in similar studies. Attention has been limited to small strain linear (elastic) problems here, however the methods may be readily extended to more complicated non-linear conditions. GOEE concepts may, for example, be applied to non-linear problems by linearization constitutive matrices (as in the work of Cirak[24] or that of Ghorashi and Rabczuk[49]). The integration of non-linearity at local levels is straightforward. Adjoint solutions (at the global scale) can be updated with the iterative primal solution, with each leading to a modification of the recovered

error field. Note that QoIs do not need to be the same between solution increments in the non-linear case. State estimation in space and time can provide inspiration, with adjoint solutions from previous increments being used to inform subsequent error field estimations.[50] Active learning can be implemented here to strategically define QoIs between increments. QoIs are defined arbitrarily in the present work, however future work should look to optimize the QoIs through adaptive learning/Greedy algorithms. Note that this can be accomplished in a simplistic way by characterizing QoIs with the problem shape functions, however a more intriguing question lies in the full optimization of QoIs in the problem domain.

DATA AVAILABILITY STATEMENT

Utility codes (developed in MATLAB) and supplementary data that support the fidnings of this study are available from the corresponding author upon reasonable request.

ENDNOTES

*This may be interpreted as a maximum likelihood statement. The true error in QoI is a Gaussian with zero mean, centered on its ZZ corrected estimate. The distance between the FE estimate of the QoI and the ZZ corrected estimate then provides one data point from which the variance of this Gaussian distribution may be estimated, resulting in the methodology described previously. Theoretically, it would be possible to calibrate the epistemic uncertainty parameters using maximum marginal likelihood, which will be done for other parameters of the GP model later on. However, this would require generating error estimates that are closer to the true QoI than the ZZ corrections, for example, by refining the computational mesh, which is undesirable.

†The Bayesian conditioning may also be expressed as a Kalman update

$$
\begin{aligned}
\mathbf{e}^{\star} &= \mathbf{0} - \mathbf{G}(\mathbf{\Delta Q} - \mathbf{H}\,\mathbf{0}) \\
\mathbf{\Sigma}^{\star} &= (\mathbf{I} - \mathbf{GH})\mathbf{\Sigma}\,,
\end{aligned}
\tag{24}
$$

where the Kalman gain is defined as $\mathbf{G} = \mathbf{\Sigma H}^T(\mathbf{H\Sigma H}^T + \mathbf{\Sigma}_\epsilon)^{-1}$.

‡Let us indicate that the continuous counterpart of the prior introduced previously reads as

$$
\forall v \in \mathcal{U}, \qquad (e,v)_{\mathcal{L}_2(\Omega)} + \beta^2 a\,(e,v) = \alpha\langle g,v\rangle
\tag{32}
$$

where random linear functional g is such that

$$
\forall(v,w) \in \mathcal{U}^2, \qquad E\,(\langle g,v\rangle\langle g,w\rangle) = (v,w)_{\mathcal{L}_2(\Omega)}\,,
\tag{33}
$$

and $e \in \mathcal{U} \times \Xi$ is the random continuous field of interest, and is defined over the set of all outcomes Ξ.

§In article,[40,41] the authors propose, in another context, to make use of the following prior covariance:

$$
\mathbf{\Sigma} = \alpha^2\mathbf{K}^{-1}\left(\mathbf{M} + \beta^2\mathbf{K}\right)\mathbf{K}^{-1}.
\tag{35}
$$

In our preliminary experiments with this prior, we failed to eliminate the spurious mesh dependencies that seem to appear when regularization parameter β is non-vanishing. As a consequence, we did not manage to generate sufficient freedom to optimize this prior in a data driven manner.

¶Let us emphasize that the numerical difficulty is due to the dimension of the FE (as is often the case in state estimation problems, when using Kriging or Kalman filters to identify high dimensional spatial fields from sparse measurements), not the number of observations (as may appear when using GP in machine learning, for instance to perform regression in relatively small feature spaces but with high volumes of data).

ORCID

James Paul Rouse https://orcid.org/0000-0003-1568-9122
Pierre Kerfriden https://orcid.org/0000-0002-7749-3996

REFERENCES

1. Ainsworth M, Oden JT. A posteriori error estimation in finite element analysis. *Comput Methods Appl Mech Eng*. 1997;142(1):1-88.
2. Grätsch T, Bathe K-J. A posteriori error estimation techniques in practical finite element analysis. *Comput Struct*. 2005;83(4):235-265.
3. Eriksson K, Johnson C. Adaptive finite element methods for parabolic problems I: a linear model problem. *SIAM J Numer Anal*. 1991;28:43-77.
4. Verfürth R. A posteriori error estimation and adaptive mesh-refinement techniques. *J Comput Appl Math*. 1994;50(1):67-83.
5. Bank JE, Weiser A. Some a posteriori error estimators for elliptic partial differential equations. *Math Comput*. 1985;44:283-301.

6. Ladevéze P, Pelle J-P. *Mastering Calculations in Linear and Nonlinear Mechanics*. Springer; 2005.

7. Larsson F, Runesson K. On two-scale adaptive FE analysis of micro-heterogeneous media with seamless scale-bridging. *Comput Methods Appl Mech Eng*. 2011;200(37):2662-2674. Special Issue on Modeling Error Estimation and Adaptive Modeling.

8. Zhang J, Natarajan S, Ooi ET, Song C. Adaptive analysis using scaled boundary finite element method in 3D. *Comput Methods Appl Mech Eng*. 2020;372:113374.

9. Song C, Ooi ET, Pramod ALN, Natarajan S. A novel error indicator and an adaptive refinement technique using the scaled boundary finite element method. *Eng Anal Bound Elem*. 2018;94:10-24.

10. Allendes A, Naranjo C, Otárola E. Stabilized finite element approximations for a generalized Boussinesq problem: a posteriori error analysis. *Comput Methods Appl Mech Eng*. 2020;361:112703.

11. Heltai L, Rotundo N. Error estimates in weighted Sobolev norms for finite element immersed interface methods. *Comput Math Appl*. 2019;78(11):3586-3604.

12. Rech M, Sauter S, Smolianski A. Two-scale composite finite element method for Dirichlet problems on complicated domains. *Numer Math*. 2006;102:681-708.

13. Pramanick T, Sinha RK. Error estimates for two-scale composite finite element approximations of parabolic equations with measure data in time for convex and nonconvex polygonal domains. *Appl Numer Math*. 2019;143:112-132.

14. González-Estrada OA, Ródenas JJ, Bordas SPA, Nadal E, Kerfriden P, Fuenmayor FJ. Locally equilibrated stress recovery for goal oriented error estimation in the extended finite element method. *Comput Struct*. 2015;152:1-10.

15. Cai Z, Kim S, Lee H-C. Error estimate of a finite element method using stress intensity factor. *Comput Math Appl*. 2018;76(10):2402-2408.

16. Oden JT, Prudhomme S. Estimation of modeling error in computational mechanics. *J Comput Phys*. 2002;182(2):496-515.

17. Rahimi N, Kerfriden P, Langbein FC, Martin RR. Cad model simplification error estimation for electrostatics problems. *J Sci Comput*. 2018;40:B196-B227.

18. Paladim DA, Moitinho de Almeida JP, Bordas SPA, Kerfriden P. Guaranteed error bounds in homogenisation: an optimum stochastic approach to preserve the numerical separation of scales. *Int J Numer Methods Eng*. 2017;110:103-132.

19. Chamoin L, Legoll F. Goal-oriented error estimation and adaptivity in MSFEM computations. *Comput Mech*. 2021;67:1201-1228.

20. Lin Z, Zhuang Z, You X, Wang H, Xu D. Enriched goal-oriented error estimation applied to fracture mechanics problems solved by XFEM. *Acta Mech Solida Sin*. 2012;25(4):393-403.

21. Ainsworth M, Oden JT. A unified approach to a posteriori error estimation using element residual methods. *Numer Math*. 1993;65:23-50.

22. Ladevèze P, Rougeot P, Blanchard P, Moreau JP. Local error estimators for finite element linear analysis. *Comput Methods Appl Mech Eng*. 1999;176(1):231-246.

23. Chamoin L, Ladevéze P. A non-intrusive method for the calculation of strict and efficient bounds of calculated outputs of interest in linear viscoelasticity problems. *Comput Methods Appl Mech Eng*. 2008;197:994-1014.

24. Cirak F, Ramm E. A posteriori error estimation and adaptivity for linear elasticity using the reciprocal theorem. *Comput Methods Appl Mech Eng*. 1998;156(1):351-362.

25. Oden JT, Prudhomme S. Goal-oriented error estimation and adaptivity for the finite element method. *Comput Math Appl*. 2001;41(5):735-756.

26. Becker R, Rannacher R. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numer*. 2001;10(1):1-102.

27. González Estrada O, Nadal E, Ródenas JJ, Kerfriden P, Pierre-Alain Bordas S, Fuenmayor FJ. *Mesh Adaptivity Driven by Goal-Oriented Locally Equilibrated Superconvergent Patch Recovery*. arXiv e-prints, September 2012.

28. Tirvaudey M, Chamoin L, Bouclier R, Passieux J-C. A posteriori error estimation and adaptivity in non-intrusive couplings between concurrent models. *Comput Methods Appl Mech Eng*. 2020;367:113104.

29. Chung ET, Leung WT, Pollock S. Goal-oriented adaptivity for GMSFEM. *J Comput Appl Math*. 2016;296:625-637.

30. Chamoin L, Legoll F. Goal-oriented error estimation and adaptivity in msfem computations. arXiv preprint 1908.00367.21, 2019.

31. Roininen L, Huttunen JMJ, Lasanen S. Whittle-Matérn priors for Bayesian statistical inversion with applications in electrical impedance tomography. *Inverse Probl Imaging*. 2014;8(2):561.

32. Grewal MS, Andrews AP. *Kalman Filtering: Theory and Practice with MATLAB*. Wiley; 2014.

33. Chilés J-P, Delfiner P. *Geostatistics: Modeling Spatial Uncertainty*. Wiley; 2012.

34. Bonney MS, Evans R, Rouse J, Jones A, Kerfriden P, Hamadi M. Goal oriented error estimation in multi-scale shell element finite element problems. *Adv Model Simul Eng Sci*. 2021;8.

35. Zienkiewicz OC, Zhu JZ. A simple error estimator and adaptive procedure for practical engineering analysis. *Int J Numer Methods Eng*. 1987;24(2):337-357.

36. Zienkiewicz OC, Zhu JZ. The superconvergent patch recovery (SPR) and adaptive finite element refinement. *Comput Methods Appl Mech Eng*. 1992;101:207-224.

37. González-Estrada OA, Ródenas JJ, Bordas SP-A, Duflot M, Kerfriden P, Giner E. On the role of enrichment and statistical admissibility of recovered fields in a posteriori error estimation for enriched finite element methods. *Eng Comput*. 2012;29(8):814-841.

38. González-Estrada OA, Nadal E, Ródenas JJ, Kerfriden P, Bordas SP-A, Fuenmayor FJ. Mesh adaptivity driven by goal-oriented locally equilibrated superconvergent patch recovery. *Comput Mech*. 2013;53:957-976.

39. Prudhomme S, Oden JT. On goal-oriented error estimation for elliptic problems: application to the control of pointwise errors. *Comput Methods Appl Mech Eng*. 1999;176:131-331.

40. Nguyen NC, Men H, Freund RM, Peraire J. Gaussian functional regression for state prediction using linear PDE models and observations. *SIAM J Sci Comput*. 2016;38(2):247-271.

41. Nguyen NC, Men H, Freund RM, Peraire J. Functional regression for state prediction using linear PDE models and observations. *SIAM J Sci Comput*. 2016;38(2):B247-B271.

42. Sidén P, Lindgren F, Bolin D, Eklund A, Villani M. Spatial 3D Mat\'ern priors for fast whole-brain fMRI analysis. arXiv preprint arXiv:1906.10591, 2019.

43. Williams CKI, Rasmussen CE. *Gaussian Processes for Machine Learning*. Vol 2. MIT Press; 2006.

44. Prud'homme C, Rovas DV, Veroy K, et al. Reliable real-time solution of parametrized partial differential equations: reduced-basis output bound methods. *J Fluids Eng*. 2002;124(1):70-80.

45. Huynh DBP, Rozza G, Patera AT. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: application to transport and continuum mechanics. *Arch Comput Methods Eng*. 2008;15(3):229-275.

46. Hoang KC, Kerfriden P, Bordas SPA. A fast, certified and "tuning free" two-field reduced basis method for the metamodelling of affinely-parametrised elasticity problems. *Comput Methods Appl Mech Eng*. 2016;298:121-158.

47. Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geograph Res Oceans*. 1994;99:10143-10162.

48. Brezina M, Falgout R, MacLachlan S, Manteuffel T, McCormick S, Ruge J. Adaptive smoothed aggregation (α SA) multigrid. *SIAM Rev*. 2005;47(2):317-346.

49. Ghorashi SS, Rabczuk T. Goal-oriented error estimation and mesh adaptivity in 3D elastoplasticity problems. *Int J Fract*. 2017;203:3-19.

50. Girolami M, Febrianto E, Yin G, Cirak F. The statistical finite element method (STATFEM) for coherent synthesis of observation data and model predictions. *Comput Methods Appl Mech Eng*. 2021;375:113533.