



HAL
open science

A probabilistic hierarchical sub-modelling approach through a posteriori Bayesian state estimation of finite element error fields

J P Rouse, Pierre Kerfriden, M Hamadi

► To cite this version:

J P Rouse, Pierre Kerfriden, M Hamadi. A probabilistic hierarchical sub-modelling approach through a posteriori Bayesian state estimation of finite element error fields. 2021. hal-03462530

HAL Id: hal-03462530

<https://hal.science/hal-03462530>

Preprint submitted on 1 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A probabilistic hierarchical sub-modelling approach through *a posteriori* Bayesian state estimation of finite element error fields.

J.P. Rouse*¹, P. Kerfriden^{2,3}, and M. Hamadi⁴

¹Gas Turbine and Transmission Research Centre (G2TRC), University of Nottingham, Nottingham, Nottinghamshire, NG7 2RD, UK

²Cardiff University, School of Engineering, Cardiff, UK

³Centres des Matériaux, Mines ParisTech/PSL University, France

⁴Airbus, Toulouse, France

Abstract

The present work considers two challenges arising from common multiscale approaches and provides a non-intrusive solution for error estimation. Firstly, the quality of the global mesh/solution and its effect on state estimations at local features is considered. Zhu-Zienkiewicz goal oriented error estimates are used to approximate errors in global deformation fields. These are then propagated across region of interest boundaries to local models and distributions in key parameters are determined. The second challenge follows the observation that local models/features may well appear at several locations in a global model. Furthermore, these locations and the details of the local models may evolve during the design process. The global model remains applicable in all cases, however without some form of interpolation scheme it is not possible to use known error estimates to inform confidence bounds at new feature locations. Gaussian process models that make use of a stochastic differential equation interpretation of the Matérn prior are used to recover the full error field, thereby allowing movement of the local model at marginal expense. The application of goal orientated error estimates and Gaussian processes in multiscale problems of this kind is novel, general, and powerful.

Key Words

Multiscale, Finite Element Methods, Error Estimation, Elasticity

Introduction

Structural analysis, utilised 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 in model development practical analysis problems would quickly become intractable and solutions would not be furnished at a rate required by the iterative design process. Geometric features, material property definitions, and discretisation particulars (i.e. element formulations and refinement) will commonly be varied over the different length scales. Broadly speaking, large, coarse (“macro scale”) models will be solved to address questions relating to “general” stiffness of a structure and to inform unknown driving boundary conditions for smaller scale, detailed (“micro scale”) models that represent localisations where gradient quantities (likely stresses and strains) are of interest. Much has been written about the various approaches adopted for these multiscale problems (a brief summary is given later). All involve the transfer for information (stiffness and boundary conditions in most cases) between macro and micro 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 (however extensions to strongly coupled problems are clear) and asks what are the ramifications of “error” in macro models? Bearing in mind the practical limitations associated with large scale analyses, how can error estimates in large scale models be generated quickly and used to their greatest effect, such that uncertainties can be propagated to local micro models? In the present work, we motivate, propose, and demonstrate a full error field recovery approach that utilises a small number of goal orientated error estimate (GOEE) observations in combination with a Gaussian Process state estimation framework. Importantly, a guided model order reduction technique, enabled by stochastic partial differential equation (PDE) representations of the prior, is proposed that alleviates computational bottlenecks in the Gaussian process steps, thereby expediting full error field estimation (in line with industry requirements).

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 “barrel” section of civil aircraft. It is clear that the real assembly will consist of a multitude of components, from large skin, stringer, and stabiliser 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 barrel section may well be required in order to anticipate the outcome of wing bend tests (among others). Clearly, small scale features (holes, rivets, and through thickness sections, for example) add little to global stiffness, justifying their omission from large scale models. Micro 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. Let us consider a hole feature in a laminate material section by way of example. Discretisation is refined in this micro models to provide better approximation of terms related to gradients in the displacement field. The hole micro model will, of course, be much smaller than the macro barrel model, and it is highly unlikely that any loads or boundary conditions will be known for the hole model a priori, meaning that they will need to be extracted from the barrel model. One may ask, if the barrel model is coarse (and presumably relatively inaccurate in terms of local deformation fields), can the hole model which is driven by it be trusted? Confidence in micro model results could (conceptually) be gained by performing error estimation around the region of interest (RoI) in the barrel model and propagating the uncertainty in driving degrees of freedom (DoFs) to the hold model. This observation alone does not prompt full error field estimation, as it assumes the RoI in the barrel model is well defined. In reality, there will be many hole features in the barrel model at a multitude of locations and analysts may well be unsure as to which holes are critically important. Performing error analysis on all RoI boundaries would be both prohibitively time consuming and wasteful. Each GOEE observation demands its own simulation and, in many cases, observations that are close to one another will, conceivably, be very similar. The additional computation effort associated with this simplistic error field recovery process greatly limits its applicability and the method is not sensitive to the iterative design process. In practice, hole model positions and geometries may change as the design progresses (as refine cable routes are determined, for example). Given that the holes are not represented in the barrel model, it is justifiable to drive the new design hole models from the original barrel model. A brute force approach would be to repeat error estimation at the new RoI boundaries, however this is unsatisfying and impractical. In the present work, we resolve this challenge by considering potential RoI boundaries (to drive the placement of error estimate observations), however we recover a full error field, statistically with confidence intervals, such that any potential RoI boundary or micro model can be considered. Further discussion on the relationship between macro and micro models in aerospace design can be found in the authors’ previous work [1]. A treatment of error estimation in mixed element (i.e. “shell” macro model to “continuum” micro model) multiscale problems is also presented in the author’s previous work, however readers should note that this does not impact the novelty of the present work. Previous work was concerned with error estimation for a particular set of element formulations, whereas the present work takes a small set of these observations and considers how a stress analyst could use them effectively in practical multiscale problems.

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 discretisation methods that are required in order to approximate the solutions to mathematical models. This contrasts with modelling error, which is a measure of how well an abstract model approximates real physical phenomena. Estimates of approximation error may be termed a priori estimates or a posteriori estimates [2], with the former utilising the problem definition and the discretisation 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 [3]. 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 [3]. In addition to accuracy, error estimates should asymptotically tend to zero as the discretisation density increases, produce tight bounds for the error, be computationally inexpensive, be applicable in a wide range of potentially non-linear applications, and inform mesh refinements such that approximate solution processes can be optimised. 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).

Multiscale modelling methods allow for the transfer of information across many length and time scales in structural analysis problems [4, 5]. They are particularly effective in cases where there are multiple physical phenomena of interest (that span a range of length/time scales) and that cannot be practically rationalised in a single model. Multiscale methods can be viewed as a set of approaches which homogenise the heterogeneity observed in all real structures at some length scale. Multiscale modelling techniques have been applied to a wide variety of fields, including aerospace [6, 7, 8], marine [9], and civil [10].

Many different multiscale methods can be found in the literature and several characterisation schemes have been proposed. Weinan bisects the field and categorises multiscale problems as either Type A, in which there are local defects/singularities that require a locally micro-scale model in an otherwise coarser global model, or Type B, which have micro-scale features throughout and require fine-scale modelling everywhere (with some form of computational homogenisation, for example) [5]. Other authors, such as Geers in [11], have used similar categorisations but under different names. In [11], Type A methods are called “Hierarchical” methods while Type B methods are called “Concurrent” methods. There is inevitably some overlap between the techniques applicable to Type A and Type B problems, for example Kim and Swan used adaptive refinement of voxel meshes of representative volume elements within their numerical homogenisation approaches [12]. Type A approaches include “classical” sub-modelling [13], domain decomposition [14, 15], and local mesh refinement including adaptive mesh refinement. The procedures used within the CleanSky2 project MARQUESS, wherein a pre-computed database of solutions enables a single sub-model to be applied to multiple instances of a recurring feature on a global FE model of a composite component, are particularly relevant to the present work [16]. In the approach, pre-computed sub-models are utilised in a bottom-up and top-down approach to identify local failure locations [17, 18, 19, 1]. The methodology is particularly useful as the superposition principle used enables rapid approximation of micro model states. Displacements from the global model can be extracted and used to estimate local stress fields in the micro model by weighting pre-computed solutions that represent particular deformation modes. Type A sub-modelling approaches are the primary concern of the present work. As will be shown later, full error fields will be developed across a macro model that describe uncertainties in driving DoFs. These distributions will in turn be sampled to develop possible loading conditions for local micro models.

An example of a Type B problem is the multiscale modelling of complex fibre architectures, such as 3D woven textile composites. Many techniques can be identified for tackling this type of problem. Computational homogenisation, for example, attempts to determine equivalent properties via representative volume elements within a periodic displacement field. The FE2 approach, on the other hand, utilises a fine mesh discretization linked to the Gauss points of a coarser mesh. The multiscale FE method (MsFEM) uses a fine mesh substructure to replace each element in the global model [5, 20]. Examples from the recent literature include Liang et al.s use of voxel models

(generated using the well-known TexGen software) to analyse woven textile composites with each element being assigned to a particular material component [21]. Shi et al. use a three-scale model involving representative volume elements at the micro and meso-scales to model the fracture of braided composites [22]. Liu et al. use a variant of a voxel FE mesh termed the inhomogeneous FE mesh to model a woven textile composite, with the material varying from integration point to integration point rather than with material boundaries being assumed to follow the mesh [23].

The approximation of errors in elliptical partial differential equations (PDEs), such as those which govern linear elastic stress analysis, has received a great deal of attention [24, 25, 26, 27]. In many cases, error measures are utilised to drive mesh refinement through, for example, polytree decomposition algorithms [25, 28, 29, 30, 31]. 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 [29], immersed surface methods [32], multigrid and composite FEA methods [28, 33, 34], extended FE (XFEM) [2, 35], and stress singularity problems [36]. Oden and Prudhomme’s [37] and Larson and Runesson’s [28] 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 submodelling based on Oden’s work [38, 39] and to [40] 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 [2, 3]. In most cases, error estimators are categorised as energy norm based (including element residual, subdomain residual methods and the constitutive relation error) or recovery based estimators. The latter includes the well know Babuska and Rheinboldt estimator, the Kelly, Gage, Zienkiewicz and Babuska estimator, and the Zienkiewicz-Zhu patch recovery technique [2]. In the present work GOEEs are utilised as, rather than providing a general indication of approximation error, they allow for the quantification of error in a specific quantity of interest (QoI) [41, 3]. GOEEs were proposed in the 1990’s [42] in the work of Prudhomme, Oden, and Ainsworth, [41, 2, 43], Ladevéze [44, 45], Bathe [3], and Cirak [46] to name a few of the most influential authors in this prolific area of research. 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 [47, 48, 49, 50]. In the multiscale setting, a posteriori error estimates have been applied to Type A problems in the work of Tirvaudey et al. [51], wherein a weighted residual based GOEE is applied in non-intrusive sub-modelling problems. A posteriori error estimates for MsFEM problems have been developed in the work of Chung and Chamoin [52, 53]. In the context of the present work, QoIs are driving DoFs extracted from macro 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 sub-models. The statistical recovery process alluded to here is the focus and novelty of the present work. The methods introduced here allows an analyst to compute corrected microscopic solutions by propagating the effect of driving DoFs corrections - the error estimates - from macroscopic to microscopic scales.

The present work concerns itself with an industry relevant problem. We seek to evaluate the influence of macroscopic mesh quality onto the solution of microscopic problem. The solution developed here is non-intrusive in terms of remeshing at the macroscale, as this is typically not feasible in real problems. Furthermore, the solution should allow for redefinition of microscale models and should minimise computational expense. The resolution proposed here uses Zhu-Zienkiewicz (ZZ) error estimates to calculate errors in macroscopic deformation fields at a limited number of locations. Gaussian processes are then trained using a stochastic differential equation interpretation of the Matérn prior [54] to interpolate between different observations of the error, i.e. the ZZ estimates are seen as partial measurements of the full error field in a state estimation framework [55, 56]. Computational bottle necks in the generation of the Matérn prior are circumvented through a GOEE model order reduction method that makes use of hierarchical clustering (enabled through k-medoids type approaches). 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, the use of functional Gaussian processes in multiscale problems, and the model order reduction approaches to make the process tractable in the industrial setting.

1 Problem setting and Formulation

1.1 Linear elasticity

We begin with a general setting for linear problems we will consider in the present work. Let us consider the linear elasticity problem over 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 considers "shell" element formulations [1]). 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 in 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 elasticity is as follows. We look for $u \in \mathcal{U}$ such that

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

In the previous equation, bilinear form a is defined by

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

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

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

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

1.2 Finite element approximation

We now introduce the usual "P1" Lagrange FE method to approximately solve the elasticity problem introduced above. Domain Ω is decomposed into a set \mathcal{T}_h of non-overlapping simplexes. We assume for simplicity that this decomposition is error-free. We now look for FE field $u_h \in \mathcal{U}_h \subset \mathcal{U}$ such that

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

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

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

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 j^{th} canonical 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 enforced.

The FE model may be written in matrix form as follows

$$\mathbf{K}\mathbf{U} = \mathbf{F} \quad (6)$$

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

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

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

and verifies Galerkin orthogonality condition

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

1.3 Hierarchical local submodelling

We now consider a sub-modelling approach which draws on the global model solution to provide more comprehensive results in some RoI. For the present work, sub models (or micro models) are considered to possess sufficiently fine mesh densities such that, compared to the global model, they are error free. In practice, the GOEE approach developed here can easily be applied to micro models as well as macro models, however this is omitted in the present work as to limit tangents that do not add to reader comprehension. It should also be noted that, in practice, an analyst has far greater control over micro models than macro models. In the aerospace example, many engineers across multiple departments may be called upon to produce a global aircraft model. Remeshing in such a case would be unacceptable and the degree of discretisation may be limited by computational resources. Local feature models may well be generated by single analyst however, thereby providing capacity for refinement.

An RoI $\Omega_I \subset \Omega$ is to be re-analyzed using a finer local model (e.g. nonlinear, 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 submodel (see e.g. [41, 57]). In our examples, the model kinematics and FE meshes are compatible at the interface $\partial\Omega_I$. This is not a restriction of the approach but one of our current implementation of the proposed methodology (see Figure 2 for further details on this aspect). A continuous error field will be developed in the following sections, meaning 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 through the chosen sub-modelling technique.

2 Goal oriented error estimation

2.1 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 ZZ recovery-based estimate of $\|e\|_a$ is the following [58]

$$\nu_{zz}^2 = \int_{\Omega} (\epsilon^* - \nabla_s u_h) : C : (\epsilon^* - \nabla_s u_h) d\Omega \quad (9)$$

where ϵ^* 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) [59], or equilibrium-informed recovery procedures [60, 61, 62].

2.2 Error estimates for engineering quantities of interest

The energy is not necessarily the only quantity whose accuracy is of interest to the FE practitioner. In the present context of multilevel modelling, we are interested in a limit-states quantities in hierarchically-defined solution enrichment (i.e. structural zooms), such as maximum stress measures. To address the limitation of a -norm error estimation, the classical GOEE methodology starts by formulating the adjoint (or dual) problem.

We begin by reformulating the original FE problem. We look to find $z \in \mathcal{U}$ such that

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

The linear functional $Q : \mathcal{U} \rightarrow \mathbb{R}$ extracts one of the QoIs from an arbitrary solution field $v \in \mathcal{U}$. Many different forms of $Q(u)$ can of course 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 define errors in driving DoFs in a multiscale problem setting, therefore QoIs will be weighted averaged extractions of components of u . Now, we have that

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

The adjoint field z (the influence function in the work of Prudhomme and Oden [63]) 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

$$\forall v \in \mathcal{U}_h, \quad a(v, z_h) = Q(v) \quad (12)$$

We can therefore write

$$Q(e) = a(e, z_h) + a(e, z - z_h) = r(z_h) + a(e, e_z) \quad (13)$$

The residual term vanishes due to Galerkin orthogonality. The second 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

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

where \hat{Q}_h is an estimate of the QoIs, recovered from the FE field, that approximates the QoI $Q(u)$. If error analyses were limited to a single RoI with a small number of driving DoFs, the above treatment would be sufficient for error propagation. Pointwise QoIs, evaluated using a Dirac delta function for the development of $Q(u)$, could be established along the boundary of the RoI in the macro model and GOEE terms calculated. In the present work, it is assumed that either no such restriction on the placement of adjoints in Ω exists, or that the number of adjoints required to sufficiently constrain all driving DoFs at all possible RoI boundaries is so high that implementing all of them would be prohibitively expensive. As a resolution to this challenge, we will now present a reconstruction of the macroscopic error field by training a Gaussian field using a sparse number of GOEE observations. The resulting probabilistic representation will yield a statistical microscopic error distribution wherever needed.

3 Full error field reconstruction through Bayesian state estimation

3.1 “Truth” reconstruction of finite element error fields using the goal oriented error estimation approaches

We now introduce a corrected model

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

where

$$\hat{\mathbf{U}} = \mathbf{U} + \hat{\mathbf{e}} \quad (16)$$

In the previous expression, $\hat{\mathbf{e}}$ 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. That 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 (this will be discussed later on for the specific case of hierarchical local sub-modelling).

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

- 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 .

It seems appropriate at this stage 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 functional* of the exact error, i.e $Q(e) = \int_{\Omega} \xi_h \cdot e \, d\Omega$, where ξ_h is a vector-valued FE field of \mathcal{U}_h , has the following property

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

where $\pi_{\mathcal{L}_2}^h$ is the $\mathcal{L}_2(\Omega)$ -projector and the above equality stems from standard operations on projectors. Hence, the above observation of the exact error is also an observation 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:

$$\forall i \in \llbracket 1 \ N \rrbracket, \quad \int_{\Omega} \psi_i \cdot e_h \, d\Omega = \mathbf{c}^{iT} \mathbf{M} \hat{\mathbf{e}} \quad (18)$$

where:

$$M_{ij} = \int_{\Omega} \psi_i \cdot \psi_j \, d\Omega \quad (19)$$

and \mathbf{c}^i is the i^{th} canonical 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. We will instead proceed by Gaussian Process functional regression [64, 65], to statistically estimate these N quantities based on the availability a selected few of them.

3.2 Bayesian reconstruction of finite element errors fields from sparse functional observations

3.2.1 Linear state estimation

Bayesian state estimation setting. We assume that we have calculated an array of M ZZ estimates of errors in scalar QoIs. These estimates are seen as noisy observations of the corresponding exact error in QoI, which reads as

$$\forall i \in \llbracket 1 \ M \rrbracket, \quad d_i = \mathbf{Q}^{iT} \hat{\mathbf{e}} + \epsilon_i = \Delta Q_i + \epsilon_i \quad (20)$$

where $\Delta Q_i \approx (Q^i(\pi_{\mathcal{L}_2}^h u) - Q^i(u_h))$ is the true error in the i^{th} QoI. It is vital to note that this is unavailable at this stage and will be given a probability distribution later on in this section, following Bayesian modelling of lack of knowledge. 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 (we have, after all, used the ZZ estimate to this end). 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”¹.

We define the vector of errors in QoIs as

$$\mathbf{d} = \mathbf{\Delta Q} + \boldsymbol{\epsilon} = \begin{pmatrix} \Delta Q_1 \\ \Delta Q_2 \\ \dots \\ \Delta Q_M \end{pmatrix} + \boldsymbol{\epsilon} = \begin{pmatrix} \mathbf{Q}^{1T} \\ \mathbf{Q}^{2T} \\ \dots \\ \mathbf{Q}^{MT} \end{pmatrix} \hat{\mathbf{e}} + \boldsymbol{\epsilon} = \mathbf{H} \hat{\mathbf{e}} + \boldsymbol{\epsilon} \quad (21)$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$ is a M -dimensional random vector containing the errors in ZZ-estimated errors in QoI, $\mathbf{\Delta Q}$ 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, $\mathbf{Q}^{iT} = \mathbf{c}^{I(i)T} \mathbf{M}$, as shown in the previous section.

¹This can be seen as a maximum likelihood statement. The true error in QoI is a Gaussian with zero mean, centered on the ZZ estimate. The distance between the FE estimate of the QoI and the ZZ estimate then provides one data point from which the variance of this Gaussian distribution may be estimated, resulting in the methodology described previously.

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

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

which, together with the distribution of $\boldsymbol{\epsilon}$, encodes all prior knowledge about epistemic sources of uncertainty in our state estimation problem.

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

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

The posterior probability of $\hat{\mathbf{e}}$ is simply obtain by Gaussian conditioning², i.e.

$$\hat{\mathbf{e}}|\mathbf{d} \sim \mathcal{N}(\mathbf{e}^*, \mathbf{\Sigma}^*) \quad (25)$$

where

$$\begin{aligned} \mathbf{e}^* &= \mathbf{\Sigma}\mathbf{H}^T(\mathbf{H}\mathbf{\Sigma}\mathbf{H}^T + \mathbf{\Sigma}_\epsilon)^{-1}\mathbf{d} \\ \mathbf{\Sigma}^* &= \mathbf{\Sigma} - \mathbf{\Sigma}\mathbf{H}^T(\mathbf{H}\mathbf{\Sigma}\mathbf{H}^T + \mathbf{\Sigma}_\epsilon)^{-1}\mathbf{H}\mathbf{\Sigma} \end{aligned} \quad (26)$$

The marginal posterior distribution of any set of linear combinations of the state vector is given by

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

where prediction operator \mathbf{E} is of size n_e by N . In this paper in particular, we are interested in restriction operators, i.e. 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, for instance all the DoFs corresponding to a subdomain $\Omega_I \subset \Omega$, or to the boundary of this subdomain in the context of the Dirichlet hierarchical sub-modelling introduced previously.

3.2.2 Prior covariance structure

Great care is required when choosing the prior covariance. Computational tractability is of course vitally important, however it is key to note that the form of prior covariance will of course dictate how observations are propagated into the posterior distribution. If unrepresentative correlations are utilised across the domain it is entirely likely meaningless posterior distributions will be realised. In the context of structural modelling, a simple distance based prior may well suggest strong correlation between two neighbourhoods that, while physically close, are not mechanically connected and thus should not greatly influence one another. 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 prior structures. Standard (kernel based) and random process based formulations are introduced, with the latter being utilised in the remainder of the work.

Kernel-based covariance. A standard choice is to use Kernels to define the covariance matrix, i.e.

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

²The bayesian conditioning may also be expressed as a Kalman update

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

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

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

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

where K_ν is the modified Bessel function of second kind of order ν , and α and l are hyperparameters of the Gaussian process ³.

Random-process-based covariance In this paper, we define the prior for uncertain state $\hat{\mathbf{e}}$ using a stochastic partial differential equation approach [54, 66]. The distribution of $\hat{\mathbf{e}}$ is implicitly defined through the solution of stochastic linear system

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

which we rewrite as

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

In the previous expressions, taken from the work of Roininen, Huttunen, and Lasanen [54], \mathbf{M} and \mathbf{K} are defined by $\mathbf{M}_{i,j} = \int_\Omega \psi_i(x) \psi_j(x) d\Omega$ and $\mathbf{K}_{i,j} = \int_\Omega \nabla \psi_j(x) \cdot \nabla \psi_i(x) d\Omega$, respectively, and are (proportional to) the standard FE mass and stiffness matrices, respectively. $\mathbf{W} \in \mathbb{R}^N$ is a white noise vector with normally distributed independent components.

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

$$\boldsymbol{\Sigma} = \alpha^2 (\mathbf{M} + \beta^2 \mathbf{K})^{-1} \mathbf{M} (\mathbf{M} + \beta^2 \mathbf{K})^{-1} \quad (32)$$

Notice that mass matrix \mathbf{M} is invertible, under weak assumptions on the quality of the FE discretisation. 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 process ^{4 5}.

3.2.3 Parameter optimisation

Optimising the hyperparameters of the prior in a data-driven manner may be done by maximising the data-likelihood with respect to hyperparameter vector $\boldsymbol{\theta} := (\alpha \beta)^T$. In our case, the data log likelihood reads as [67]

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

where $Z(\boldsymbol{\theta}) = \sqrt{(2\pi)^M |\mathbf{H}^T \boldsymbol{\Sigma}(\boldsymbol{\theta}) \mathbf{H} + \boldsymbol{\Sigma}_\epsilon|}$. Providing that the number of observations remains small, this optimisation problem is tractable. The expression of the sensitivities of the data-likelihood can be found in [67].

For now, this optimisation will be performed “by hand”. We calibrate the overall size and length scale of the Gaussian process in order to match observations, by a procedure of trial-and-error. Fully automatising the algorithm is a perspective of this work but it not implemented here.

³The formulation described above corresponds to the introduction a continuous Gaussian process and its subsequent restriction to the vertices of the FE mesh

⁴Let us indicate that the continuous counterpart of the prior introduced previously reads as

$$\forall v \in \mathcal{U}, \quad (e, v)_{\mathcal{L}_2(\Omega)} + \beta^2 a(e, v) = \alpha \langle g, v \rangle \quad (33)$$

where random linear functional g is such that

$$\forall (v, w) \in \mathcal{U}^2, \quad E(\langle g, v \rangle \langle g, w \rangle) = (v, w)_{\mathcal{L}_2(\Omega)} \quad (34)$$

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

⁵In paper [64, 65], the authors propose, in another context, to make use of the following prior covariance:

$$\boldsymbol{\Sigma} = \alpha^2 \mathbf{K}^{-1} (\mathbf{M} + \beta^2 \mathbf{K}) \mathbf{K}^{-1} \quad (35)$$

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

3.3 Computational bottleneck

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 $\mathbf{E}\Sigma\mathbf{H}^T(\mathbf{H}\Sigma\mathbf{H}^T + \Sigma_\epsilon)^{-1}\mathbf{d}$. Product $\Sigma\mathbf{H}^T$ appearing in innovation matrix $\mathbf{H}\Sigma\mathbf{H}^T + \Sigma_\epsilon$ 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 $(\mathbf{M} + \beta^2\mathbf{K}) \cdot = \cdot$. This operation scales as $o(MN^2)$. The inversion of $(\mathbf{H}\Sigma\mathbf{H}^T + \Sigma_\epsilon)^{-1}$ 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 $\mathbf{E}\Sigma\mathbf{E}^T - \mathbf{E}\Sigma\mathbf{H}^T(\mathbf{H}\Sigma\mathbf{H}^T + \Sigma_\epsilon)^{-1}\mathbf{H}\Sigma\mathbf{E}^T$. 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 submodelling, \mathbf{E} will extract all the DoFs of $\partial\Omega_I$.

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

$$\mathbf{s} = \mathbf{E} \left(\mathbf{x}_0 + \mathbf{G} \left(\tilde{\mathbf{d}} - \mathbf{H}\mathbf{x}_0 \right) \right) \quad (37)$$

where $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \Sigma)$ is drawn from the prior distribution, $\tilde{\mathbf{d}} \sim \mathcal{N}(\mathbf{d}, \Sigma_\epsilon)$ is a randomly perturbed observation, and the Kalman gain is $\mathbf{G} = \Sigma\mathbf{H}^T(\mathbf{H}\Sigma\mathbf{H}^T + \Sigma_\epsilon)^{-1}$. Drawing a sample from the posterior distribution is dominated by the cost of solving two systems involving the matrix $(\mathbf{M} + \beta^2\mathbf{K})$ 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 factorised 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 Gaussian process 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.

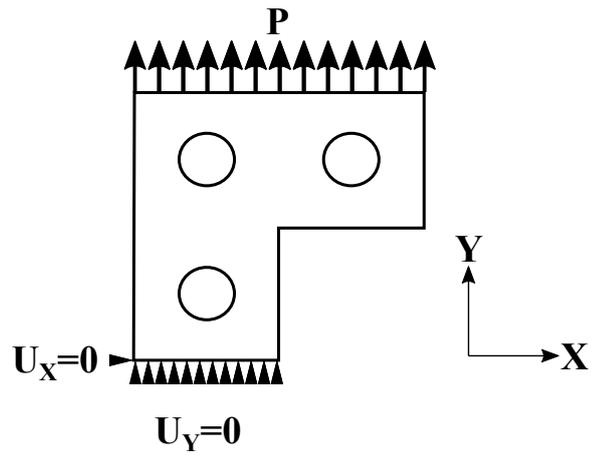
Let us emphasise that the numerical difficulty is due to the dimension of the parameter space (as is often the case in state estimation, when using Kriging or Kalman filters to identify high-dimensional spatial fields from sparse measurements), not the number of observations (as may appear when using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

4 Results - Example Problem definition and application of full rank Bayesian State Estimation

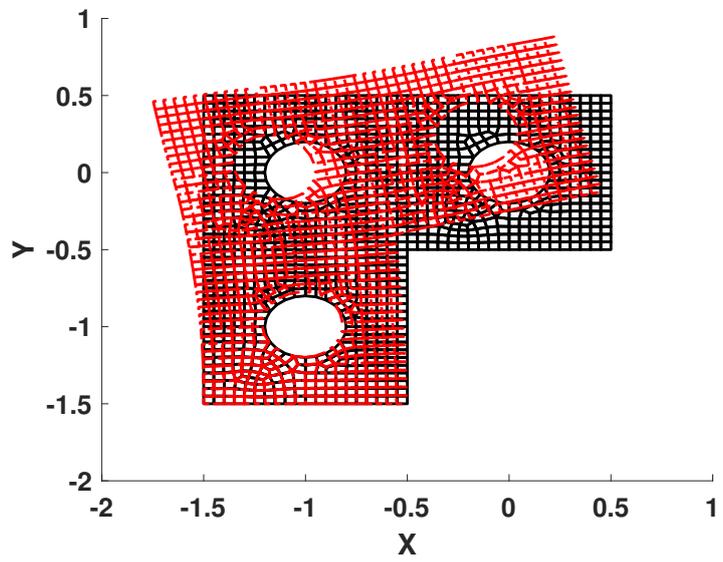
4.1 Context and example problem definition

We begin by illustrating the Gaussian process enhanced GOEE methods developed above through a simple multiscale problem. We here consider a 2d (plane stress) continuum element "perforated L" plate, fixed along the lower edge and loaded in the vertical by a uniformly distributed load (see figure 1 (a)). Note that the "x" dof at the node at the lower left hand corner is constrained to prevent rigid body motion. To clarify the loading arrangement, the deformed (unity scale factor) and undeformed mesh is presented in figure 1 (b) (note that this is the primal problem solution, found by normal FE solution approaches). We consider the situation presented in figure

1 as our macro problem and introduce several potential micro models at one of two locations (model interfaces are shown in blue and red in figure 2) to complete the multiscale setting. The introduction of only a small number macro/micro boundaries clearly simplifies the problem, to the benefit of the error field recovery. In reality, it is highly likely that many potential interface will be possible over the global model, thereby greatly increasing the number of potential QoIs. In the present work, we limit selection of QoIs to the two boundaries highlighted in figure 2. In all cases, QoIs are centred on one of the boundary nodes using a Dirac delta function (furnishing us with point wise error estimates) and consider the resultant of the DoFs in a random direction. In this way, each QoI observation provides information on errors in both x and y . For the meshes presented in figure 2 and given the restrictions listed above, the “all duals” case (i.e. the situation in which all relevant adjoints are solved) amounts to solving an additional 51 problems. Coherency between meshes at the macro/micro model boundaries is maintained, thereby simplifying the propagation of errors through the scales (note however that this is done purely to simplify the implementation and is not a requirement). In the present work, error propagation is only performed using the “two circle” micro model in the blue boundary position. Error fields are recovered over the entire macro model domain however, in the interest of space, repeated micro model implementations that do not add to reader comprehension are omitted. In all cases, linearity in both material stiffness and geometry is assumed.



(a)



(b)

Figure 1: The example macroscopic system, showing (a) boundary conditions and loads, and (b) the deformed (red) and undeformed (black) mesh (with a unity deformation scaling factor).

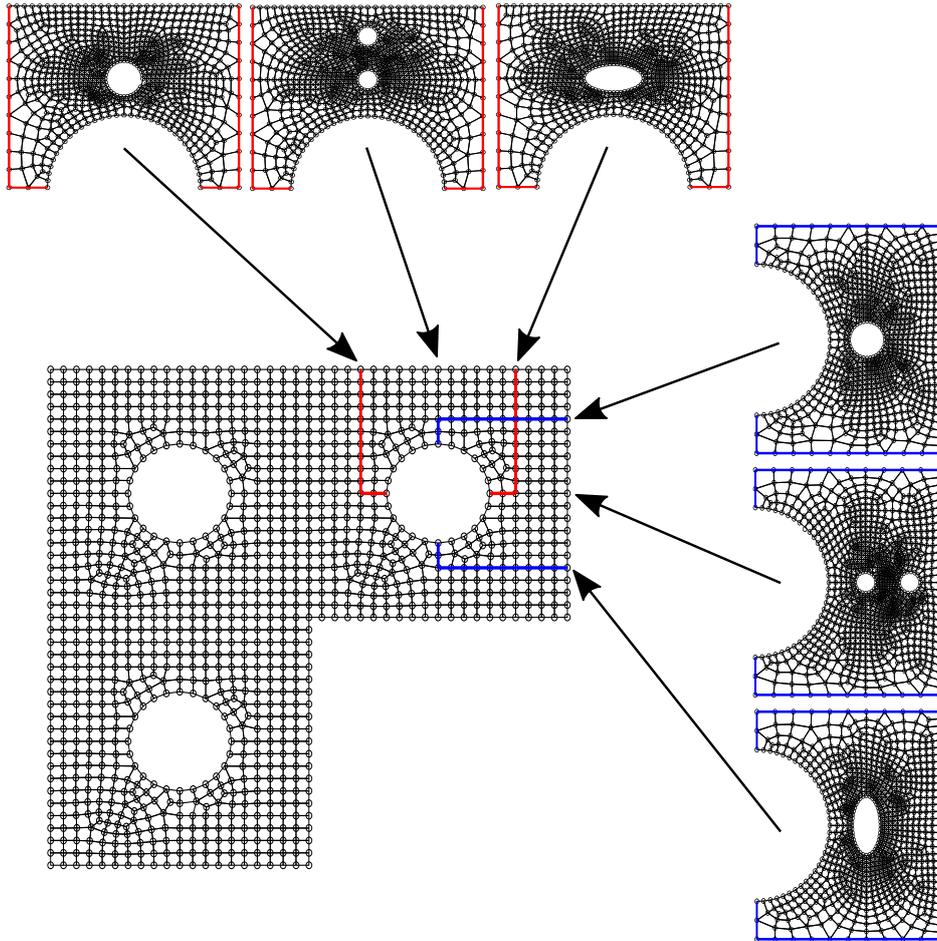


Figure 2: The example multiscale problem considered in the present work. Two potential RoI boundaries are defined (in red and blue), within which a multitude of micro models (that contain local features) can be placed, thereby motivating a description of error in driving DoFs on the RoI boundaries.

4.2 Full Rank Bayesian Recovery of Goal Oriented Error Estimation field

We present here the recovery of error fields over the entire macro domain using the methods outlined above, wherein the stochastic partial differential equation representation of the Matérn prior is used (see equation 32), and consider how the number of observations (duals) influences the posterior distributions. In the following, we randomly select a number of duals from the 51 available (restrictions are outlined and justified in the previous section), optimise the hyperparameter α (see equation 32) by maximising the likelihood given by equation 36 (note β is taken as 1 in all cases, being a length scale on the order of the problem itself), evaluate the prior and posterior, and sample from the resultant to propagate error through the model scales. Given that β is assumed, the hyperparameter optimisation becomes 1d. A starting value of 3.5×10^{-4} is assumed for α in this optimisation (note that this value is preserved in realisations that make use of the prior but no observations) and the process amounts to a simple polynomial fitting approach (applied to evaluations of equation 36 for a wide range of α values around the initial estimate). The optimisation approach implemented here is clearly simplistic and far more sophisticated methods may be utilised for enhanced hyperparameter estimation, however given the scope of the present work the method is judged to be appropriate. It is important to note that some form of hyperparameter optimisation is required here, 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 following. A covariance matrix $\boldsymbol{\Sigma}$ may be decomposed to the matrix \mathbf{A} using Cholesky decomposition, such that

$$\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T \quad (38)$$

A noise vector, \mathbf{z} , may then be constructed whose components are independent and drawn from a standard normal. A sample, $\boldsymbol{\zeta}$ is generated using the mean $\boldsymbol{\mu}$ by

$$\boldsymbol{\zeta} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \quad (39)$$

Note that, in the context of the present work, each sampling $\boldsymbol{\zeta}$ is a realisation of the error field $\hat{\mathbf{e}}$. Figure 3 presents 4 such samplings, all developed using the same \mathbf{z} noise vector (shown in figure 3 (a)). By using the same \mathbf{z} in all samplings, the influence of number of training duals (observations) can be visualised. Figure 3 (b) shows sampling performed on the prior, figure 3 (c) shows sampling performed on a 2 dual trained posterior, figure 3 (d) shows sampling performed on a 16 dual trained posterior, and figure 3 (e) shows sampling performed on a posterior trained by all (51) duals. Note that dual training sets are subsets of one another (that is to say, the 2 duals selected to train that particular posterior are also found in the 16 used to train that particular posterior), meaning features in the posterior resulting from particular observations should be preserved as the number of training cases is increased. Figures 3 (b)-(e) clearly show that, as the number of observations is increased, the posterior tends (as expected) to the case where all duals are considered (figure 3 (e)). 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 neighbourhood is greatly reduced), and note that, as we move away from the RoI boundaries (and hence observations) we tend towards the prior distribution.

The influence of uncertainty at the RoI boundary (and the effect of number of QoI observations on this) can be further visualised by considering stress distributions over the entire the RoI. We consider the cases where 2, 16, and all available duals are used to train the posteriors. Monte-Carlo samplings are taken from each posterior and are used to calculate potential stress states in the micro model (as above). Distributions are then fitted to the results on an integration point by integration basis, allowing for the determination of the mean von Mises stress ($Mean_{VM}$) and its standard deviation (Std_{VM}) over the micro model. Results for the 2, 16, and all available duals training cases are shown in figures 4, 5, and 6, respectively. We may observe that, in these results, the number of training duals 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 driving macro solution the macro solution (which is common to all realisations). 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 duals (the standard deviation of the 2 dual training cases is approximately 3 times greater than that of the all duals training case). This of course can be rationalised as a transition from the uncertainty associated with the prior (which is relatively large and is recovered in the low number of training duals case) to the uncertainty associated with the noise in our error observations (which is propagated in the all training dual 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 micro model and noting how the resulting distributions vary with number of training duals. This is presented in figure 7, 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 duals (from 2 to all available). Note that the standard deviation of the 32 training duals and all training dual distributions are near identical, suggesting that there is limited utility in calculating the additional QoIs.

Our discussions thus far have focused on a fixed set of training duals. That is to say, whenever we make a realisation for a 16 training dual case (for example), we have used the same 16 duals. While these we selected at random from the 51 available over the RoI boundary, one may question how representative the results are of all possible dual selections. Is it possible that the selected duals were particularly “good” (or “bad”) by pure chance? This question is addressed in two ways here. First, we consider different random selections of training duals to develop the peak von Mises

stress distributions, as shown in figure 8. While these distributions are distinct, they all possess the general properties noted above. We see, in all cases, a tendency from large amounts of uncertainty (associated with the prior) to relatively small amounts of uncertainty (associated with the noise in the error observations) as the number of training duals is increased. The Kullback-Leibler divergence (or relative entropy) can be used to formalise these observations. The Kullback-Leibler divergence between the approximate distribution $\mathbf{p}_{\mathbf{L}\bar{\mathbf{e}}}^{n_1}$ obtained using a prior covariance of rank n_1 and the approximate distribution $\mathbf{p}_{\mathbf{L}\bar{\mathbf{e}}}^{n_2}$ obtained using a prior covariance of rank n_2 , is

$$\begin{aligned} \mathcal{D}_{\text{KL}}(\mathbf{p}_{\mathbf{L}\bar{\mathbf{e}}}^{n_1} || \mathbf{p}_{\mathbf{L}\bar{\mathbf{e}}}^{n_2}) &= \frac{1}{2} \left(\text{Tr} \left(\text{Cov}(\mathbf{p}_{\mathbf{L}\bar{\mathbf{e}}}^{n_2})^{-1} \text{Cov}(\mathbf{p}_{\mathbf{L}\bar{\mathbf{e}}}^{n_1}) \right) \right. \\ &\quad \left. - (M + n_e) + \ln \left(\frac{|\text{Cov}(\mathbf{p}_{\mathbf{L}\bar{\mathbf{e}}}^{n_2})|}{|\text{Cov}(\mathbf{p}_{\mathbf{L}\bar{\mathbf{e}}}^{n_1})|} \right) \right) \end{aligned} \quad (40)$$

In the context of the present work, we may use \mathcal{D}_{KL} to compare posteriors against the all dual training case. This is presented in figure 9. Note that, in figure 9, each evaluation of \mathcal{D}_{KL} is made using a posterior trained using a random selection of duals. As expected, as we increase the number of training duals, we see the spread in \mathcal{D}_{KL} reduce to the point where it represents only the noise in the observations. For small numbers of training duals we note an increasingly larger spread in \mathcal{D}_{KL} , suggesting a recovery of the prior distribution. Note that one may be “lucky” (or “unlucky”) when randomly selecting duals and achieve a posterior that has a relatively low (or high) \mathcal{D}_{KL} . It is not possible, however, to ensure that this is the case without deliberate selection of the training duals based on some exterior selection pressure, which is outside the scope of the present work.

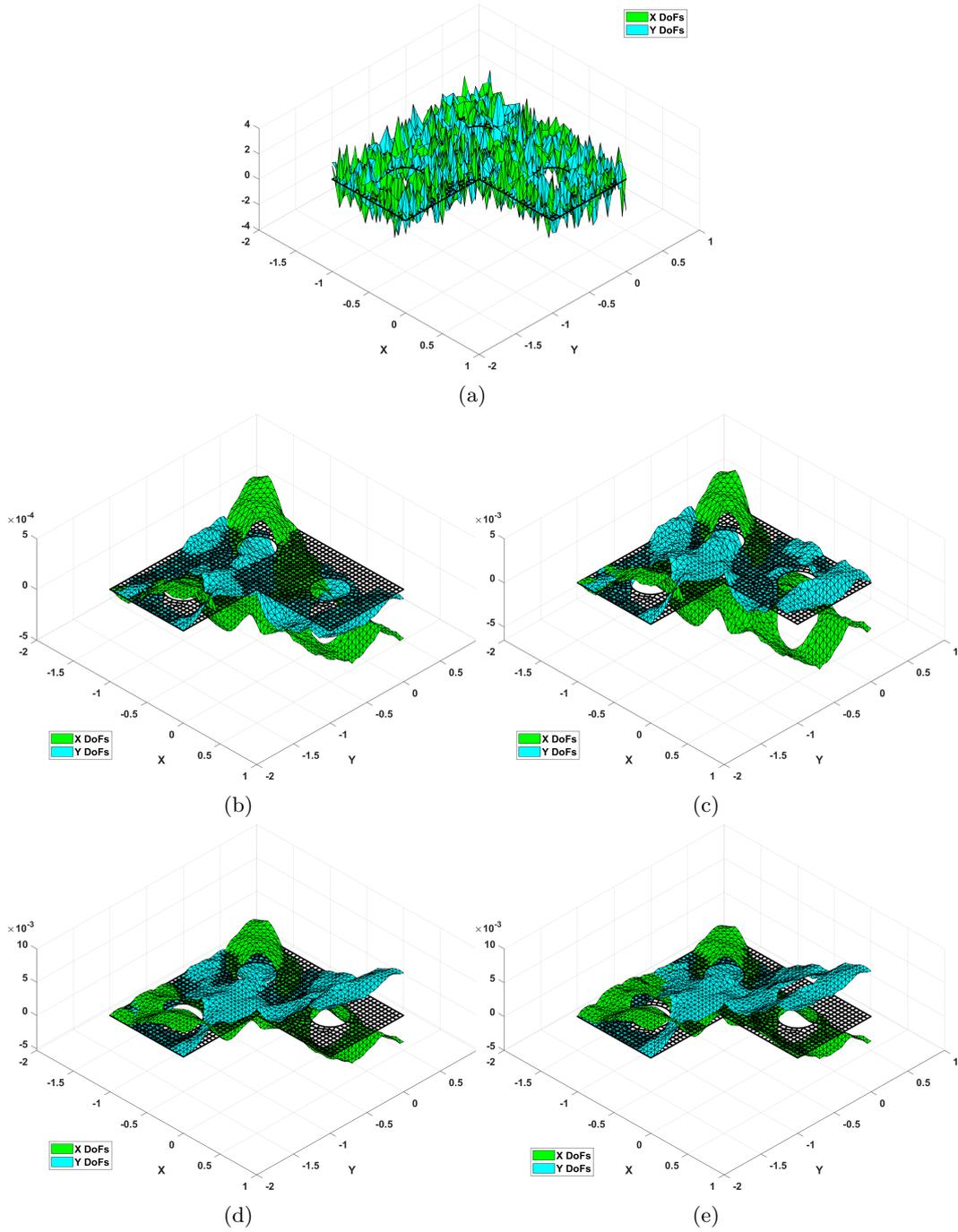


Figure 3: A sampling from GOEE field distributions, showing convergence of the solutions with increasing numbers of training data sets (duals). Note the same noise vector, shown in (a), is used in all cases. Sub figures (b)-(e) show the stochastic differential equation determined Matérn prior, a posterior trained using 2 randomly selected duals, a posterior trained using 16 randomly selected duals, and a posterior trained using all duals on the RoI boundaries, respectively.

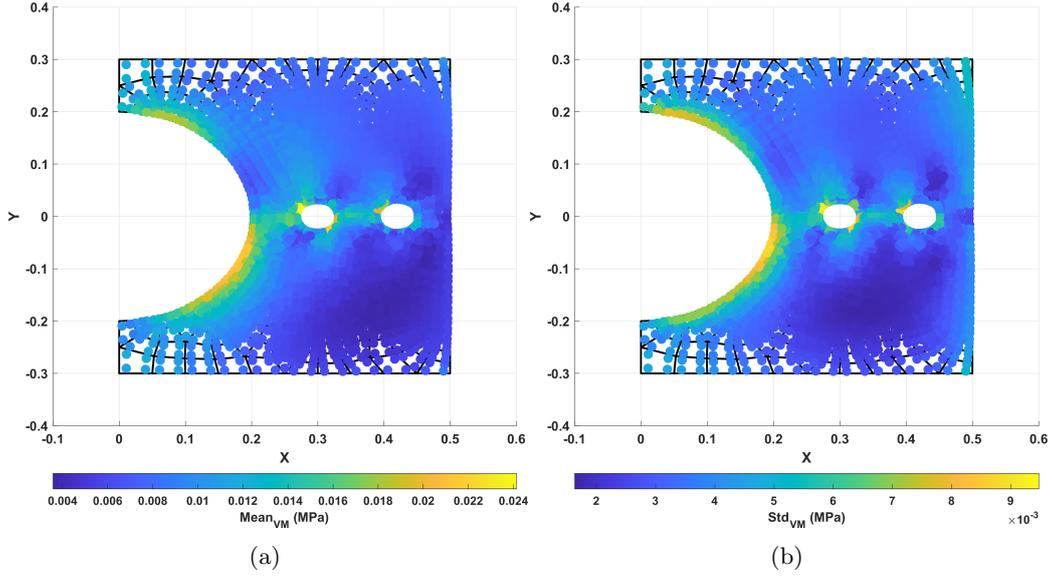


Figure 4: Monte-Carlo error field propagation for the “2 hole” (position 2) microscopic model, showing (a) the mean field and (b) the standard deviation field. Results are based on a posterior trained using 2 randomly selected duals from the ROI boundaries.

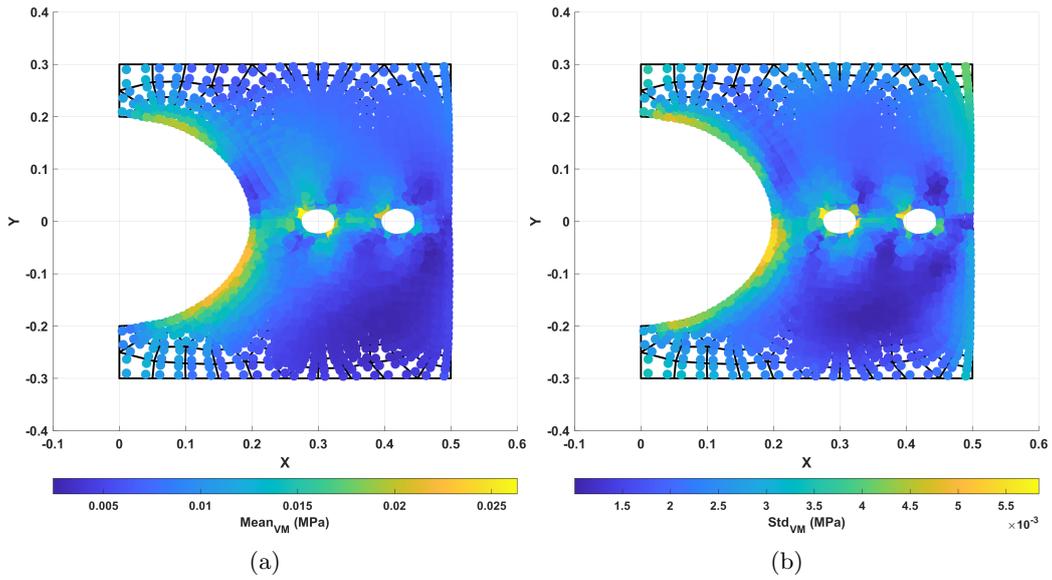


Figure 5: Monte-Carlo error field propagation for the “2 hole” (position 2) microscopic model, showing (a) the mean field and (b) the standard deviation field. Results are based on a posterior trained using 16 randomly selected duals from the ROI boundaries.

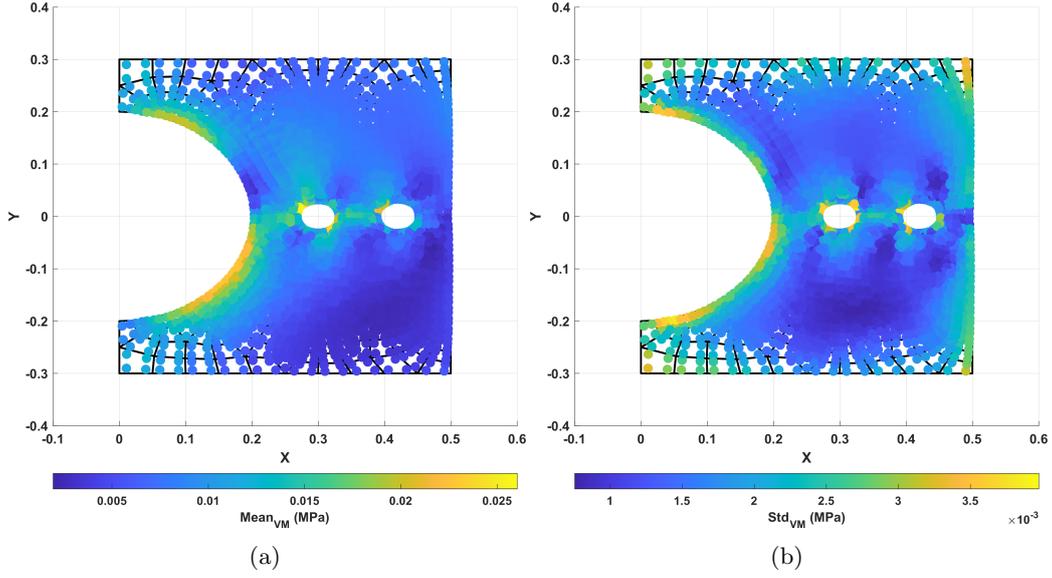


Figure 6: Monte-Carlo error field propagation for the “2 hole” (position 2) microscopic model, showing (a) the mean field and (b) the standard deviation field. Results are based on a posterior trained using all duals from the RoI boundaries.

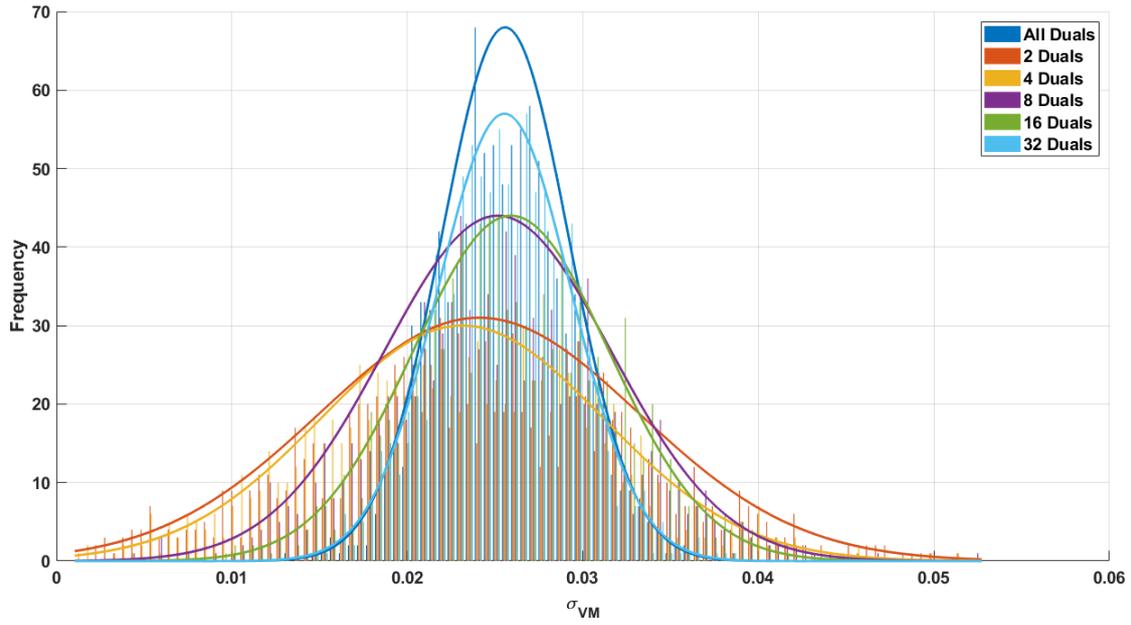
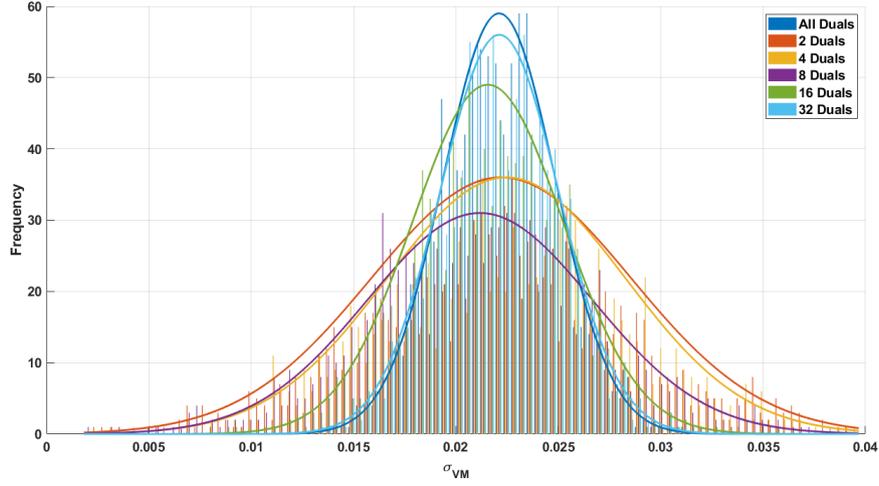
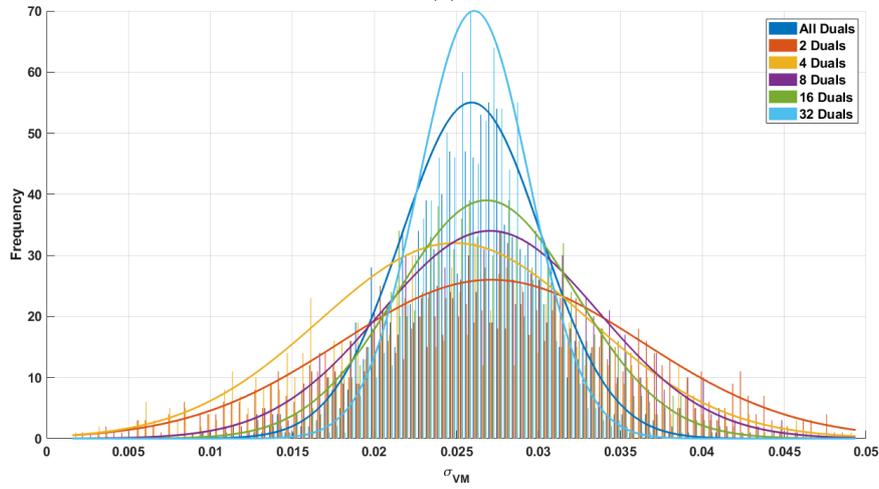


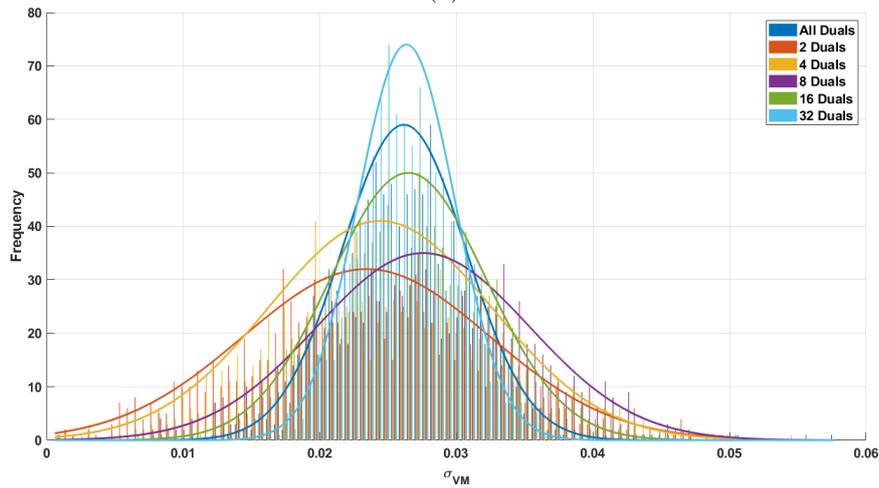
Figure 7: Peak von Mises equivalent stresses (σ_{vM}), taken from a series of Monte-Carlo error field propagations for the “2 hole” (position 2) microscopic model. Microscopic models are driven by DoF fields modified by posterior distributions trained using varying numbers of dual solutions.



(a)



(b)



(c)

Figure 8: Alternative peak von Mises equivalent stresses (σ_{VM}), taken from a series of Monte-Carlo error field propagations for the “2 hole” (position 2) microscopic model. Microscopic models are driven by DoF fields modified by posterior distributions trained using varying numbers of dual solutions. Note that, while selected duals are consistent in any one Monte-Carlo simulation, different randomly selected duals are used in subfigure (a)-(c).

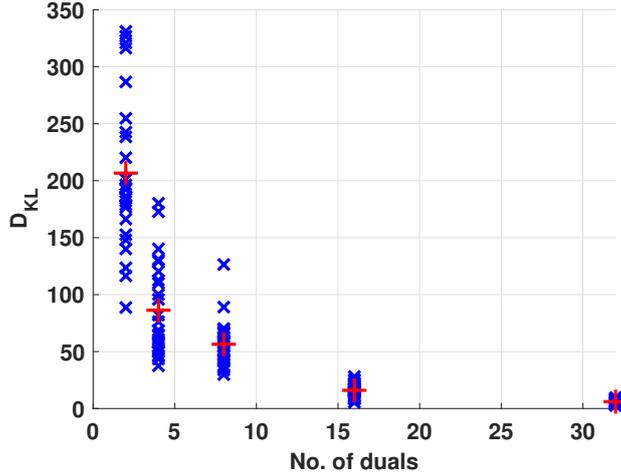


Figure 9: Relative entropy (Kullback-Leibler divergence) for Monte-Carlo developed posterior distributions (i.e. posterior distributions are developed using a random selection of duals from the RoI boundaries).

5 Low-rank Gaussian process by Galerkin Model Order Reduction

5.1 Stochastic reduced basis approximation

We now look to address the computational bottleneck identified in section 3.3 through the projection of the Gaussian process onto a low rank space. We will look for a reduced basis approximation of prior distribution of $\hat{\mathbf{e}}$. The corresponding approximation reads as

$$\hat{\mathbf{e}} \approx \tilde{\mathbf{e}} = \boldsymbol{\phi}\boldsymbol{\gamma} \quad (41)$$

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 (see e.g. [68, 69, 70])

$$\forall \mathbf{b} \in \mathbb{R}^N, \quad \boldsymbol{\phi}^T (\mathbf{b} - \mathbf{A}\tilde{\mathbf{e}}) = \mathbf{0} \quad (42)$$

We find that

$$\tilde{\mathbf{e}} = \boldsymbol{\phi} \left(\boldsymbol{\phi}^T \mathbf{A} \boldsymbol{\phi} \right)^{-1} \boldsymbol{\phi}^T \mathbf{b}. \quad (43)$$

Therefore, the approximate prior reads as

$$\tilde{\mathbf{e}} \sim \mathcal{N} \left(\mathbf{0}, \tilde{\boldsymbol{\Sigma}} := \boldsymbol{\phi} \boldsymbol{\Sigma}_\gamma \boldsymbol{\phi}^T \right) \quad (44)$$

where the covariance matrix for the n_ϕ reduced random variables $\boldsymbol{\gamma}$ is

$$\boldsymbol{\Sigma}_\gamma = \alpha^2 \boldsymbol{\phi} \left(\boldsymbol{\phi}^T (\mathbf{M} + \beta^2 \mathbf{K}) \boldsymbol{\phi} \right)^{-1} \boldsymbol{\phi}^T \mathbf{M} \boldsymbol{\phi} \left(\boldsymbol{\phi}^T (\mathbf{M} + \beta^2 \mathbf{K}) \boldsymbol{\phi} \right)^{-1} \boldsymbol{\phi}^T \quad (45)$$

The approximate prior covariance $\tilde{\boldsymbol{\Sigma}}$ 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, but the corresponding system matrices are fully populated.

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 paper (refer for instance to the usual implementation of the Ensemble Kalman Filter)

5.2 Projection subspaces

5.2.1 Karhunen-Loève transform.

If the basis vectors are made of all the eigenvectors of the full covariance matrix, i.e.

$$\forall i = 1 \dots N, \quad \mathbf{K}\phi_i = \lambda_i \mathbf{M}\phi_i \quad \text{and} \quad \forall j \neq i, \quad \phi_i^T \phi_j = \delta_{ij}, \quad (46)$$

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

$$\tilde{\Sigma} = \alpha^2 \phi \Lambda_m (\Lambda_m + \beta^2 \Lambda_k)^{-2} \phi^T \quad (47)$$

where Λ_m and Λ_k are diagonal matrices with i^{th} 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.

5.2.2 Subspace generated by the set of precomputed dual 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

$$\forall i \in \llbracket 1, M \rrbracket, \quad \mathbf{K}^T \mathbf{z}^i = \mathbf{Q}^i \quad (48)$$

where \mathbf{Q}_i^T is the i^{th} 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 $\mathbf{Z} = (\mathbf{z}^1 \ \mathbf{z}^2 \ \dots \ \mathbf{z}^M) \in \mathbb{R}^{N \times M}$.

5.2.3 RBF enrichment and regularisation

Additional reduced basis functions are required to let the Gaussian process 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 [71] as projection vectors. In this paper, we report results related to the use of a set of analytically defined and numerically optimised 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, contained in matrix $\mathbf{B} \in \mathbb{R}^{N \times n_b}$.

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

$$\phi = (\mathbf{B} \ \mathbf{Z}) \in \mathbb{R}^{N \times n_\Phi} \quad (49)$$

To avoid the potential (near-)linear dependencies of the columns of ϕ , an SVD of ϕ is computed, and ϕ is replaced by the matrix whose columns are the left-singular vectors of this decomposition that are associated non-vanishing singular values up to a numerical tolerance.

5.3 Accuracy measure

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

$$\Delta \mathbf{e} = \hat{\mathbf{e}} - \tilde{\mathbf{e}} = \mathbf{A}^{-1} \left(\mathbf{I}_d - \mathbf{A}\mathbf{B}(\mathbf{B}^T \mathbf{A}\mathbf{B})^{-1} \mathbf{B}^T \right) \mathbf{b} \quad (50)$$

where we recall that \mathbf{b} is a zero-mean Gaussian-distributed vector with covariance matrix $\alpha^2 \mathbf{M}$.

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

$$\mathbf{E}\Delta \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\Delta \mathbf{E}} := \mathbf{E}\Sigma_{\Delta} \mathbf{E}^T) \quad (51)$$

with

$$\Sigma_{\Delta} = \alpha^2 \left(\mathbf{A}^{-1} \left(\mathbf{I}_d - \mathbf{A}\mathbf{B}(\mathbf{B}^T \mathbf{A}\mathbf{B})^{-1} \mathbf{B}^T \right) \right) \mathbf{M} \left(\left(\mathbf{I}_d - \mathbf{B}(\mathbf{B}^T \mathbf{A}\mathbf{B})^{-1} \mathbf{B}^T \mathbf{A}^T \right) \mathbf{A}^{-1} \right) \quad (52)$$

The computation of $E\Sigma_{\Delta}$ is not tractable, as we would need to compute $\mathbf{A}^{-1}\mathbf{E}^T$, which is a matrix whose columns are made of all dual solutions, the computation of which we are precisely trying to avoid by using Gaussian process functional regression. However, choosing a selected few lines of \mathbf{E} can provide valuable GOEE of the accuracy of the low-rank approximation to the full Gaussian process, which we will make use of later on.

6 Adaptive set of radial basis functions using stochastic adjoint-based goal oriented methodology and hierarchical clustering

6.1 Reduced basis vectors defined as finite element interpolants of RBFs

We propose to construct the reduced space as the span of a set of radial bases functions, projected in the FE space. To this end, let us define

$$\forall i \in \llbracket 1, n_b \rrbracket, \forall x \in \Omega, \quad \zeta^i(x) = \omega^{\lceil i/d \rceil} \exp\left(-\frac{1}{2} \frac{(x - c^{\lceil i/d \rceil})^2}{l^{\lceil i/d \rceil 2}}\right) e_{i - (\lceil i/d \rceil - 1)d} \quad (53)$$

$$\zeta^i(x) = \eta^{\lceil i/d \rceil}(x) e_{i - (\lceil i/d \rceil - 1)d} \quad (54)$$

and its FE interpolant

$$\zeta_h^i(x) = \mathcal{I}_h \zeta^i(x) \quad (55)$$

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

$$\forall i \in \llbracket 1, n_b \rrbracket, \forall x \in \Omega, \quad \zeta_h^i(x) = \sum_{j=1}^N \psi_j(x) \gamma_j^i \quad (56)$$

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

- For any $i \in \llbracket 1, n_b/d \rrbracket$, we define l_i as the distance to the N_1^{th} closest point of set $\{c_i\}_{i \in \llbracket 1, n_b/d \rrbracket}$, with $N_1 \in \mathbb{R}^{+*} \setminus \{1\}$, 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 $\{\omega_i\}_{i \in \llbracket 1, n_b/d \rrbracket}$ by

$$\forall i \in \llbracket 1, n_b/d \rrbracket, \quad \omega_i(x) = \frac{1}{\frac{n_b/d}{\sum_{i=1}^{n_b/d} \eta^i(x)}} \quad (57)$$

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

To initialise 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 $\mathcal{C}^0 = \{c_i^0\}_{i \in \llbracket 1, n_b^0/d \rrbracket}$ is computed as the centroid of the i^{th} cluster of mesh vertices.

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

$$\mathbf{B} = (\gamma^1 \ \gamma^2 \ \dots \ \gamma^{n_b}) \quad (58)$$

6.2 Stochastic dual weighted residual approach

We define the following error estimate

$$\nu = \sum_{i=1}^{M^*} \text{Var} \left(\left(\mathbf{Q}^{i^*} \right)^T \Delta \mathbf{e} \right) \quad (59)$$

where $\{\mathbf{Q}^{i^*}\}_{i \in \llbracket 1 \ M^* \rrbracket}$ are an arbitrary set of extractors of scalar QoIs. As mentioned previously, we will use a random selection of the lines of prediction operator \mathbf{H} .

We note that

$$\left(\mathbf{Q}^{i^*} \right)^T \Delta \mathbf{e} = \left(\mathbf{Q}^{i^*} \right)^T \left(\mathbf{A}^{-1} \mathbf{r} \right) = \left(\mathbf{z}^{i^*} \right)^T \mathbf{r} \quad (60)$$

where i^{th} dual \mathbf{z}^{i^*} is such that $\mathbf{A}^T \mathbf{z}^{i^*} = \mathbf{Q}^{i^*}$, and the random residual vector $\mathbf{r} := \mathbf{b} - \mathbf{A} \tilde{\mathbf{e}}$ has expression

$$\mathbf{r} = \mathbf{I} - \mathbf{A} \phi \left(\phi^T \mathbf{A} \phi \right)^{-1} \phi^T \mathbf{b} =: \boldsymbol{\xi} \mathbf{b} \quad (61)$$

Therefore, \mathbf{r} is a zero-mean Gaussian vector with covariance $\text{Cov}(\mathbf{r}) = \boldsymbol{\xi} \mathbf{M} \boldsymbol{\xi}^T$. Hence, (59) reads as

$$\nu = \sum_{i=1}^{M^*} \mathbf{z}^{i^*T} \text{Cov}(\mathbf{r}) \mathbf{z}^{i^*} \quad (62)$$

Equation (60) is usually coined Dual Weighted Residual in the error estimation literature [72, 73], albeit usually employed in a deterministic setting. In order to make this expression small, an owing to Galerkin orthogonality, the suggestion is to use the adjoints $\{\mathbf{z}^{i^*}\}_{i \in \llbracket 1 \ M^* \rrbracket}$ as reduced basis vectors. They are, of course unavailable and we can only try and produce a similar effect by using selected RBFs.

Extending classical error estimation procedures, we use Galerkin orthogonality⁶ to reveal that (62) reads as

$$\nu = \sum_{i=1}^{M^*} \left(\mathbf{z}^{i^*} - \Pi_{\Phi} \mathbf{z}^{i^*} \right)^T \text{Cov}(\mathbf{r}) \left(\mathbf{z}^{i^*} - \Pi_{\Phi} \mathbf{z}^{i^*} \right) \quad (64)$$

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

$$\nu = \sum_{i=1}^{M^*} \left(\mathbf{w}^i \right)^T \mathbf{y}^i \quad (65)$$

with $\mathbf{w}^i := \mathbf{z}^{i^*} - \Pi_{\Phi} \mathbf{z}^{i^*}$ and $\mathbf{y}^i := \text{Cov}(\mathbf{r}) \mathbf{w}^i$. Equation (65) can be seen as an extension of the classical dual weighted residual framework to discrete stochastic Galerkin approximations. As usual in dual weighed residual approaches, we may interpret the previous expression as a discrete error density, as

$$\nu = \sum_{i=1}^{M^*} \sum_{j=1}^N \mathbf{w}_j^i \mathbf{y}_j^i = \sum_{j=1}^N \left(\sum_{i=1}^{M^*} \mathbf{w}_j^i \mathbf{y}_j^i \right) = \sum_{j=1}^N \left(\sum_{i=1}^{M^*} |\mathbf{w}_j^i \mathbf{y}_j^i| \right). \quad (66)$$

We may finally attempt to select the radial basis functions that approximately minimise ν , i.e. 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 $\left\{ \sum_{i=1}^{M^*} |\mathbf{w}_j^i \mathbf{y}_j^i| \right\}_{j \in \llbracket 1 \ N \rrbracket}$, thereby making sure

that for DoFs corresponding to large values of projected dual vectors $\{\mathbf{w}^i\}_{i \in \llbracket 1 \ M^* \rrbracket}$, 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 [72, 73]). This implies refining the projection space in those areas, following Galerkin’s principle.

⁶Galerkin orthogonality:

$$\forall i \in \llbracket 1, n_{\Phi} \rrbracket, \quad \Phi^{iT} \mathbf{r} = 0 \quad (63)$$

6.3 Goal oriented hierarchical clustering

We propose to use the discrete GOEE density described previously to guide the hierarchical refinement of the agglomerative clustering algorithm described in section 6.1.

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

$$q_l^k = \sum_{j=1}^{n_c^l} \sum_{i=1}^{M^*} \left| \mathbf{w}_{\mathcal{I}^l(j)}^{i,k} \mathbf{y}_{\mathcal{I}^l(j)}^{i,k} \right| \quad (67)$$

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^* for refinement, with

$$l^* = \arg \max_l (q_l^k) \quad (68)$$

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

6.4 Rank selection by holdout/cross-validation

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

$$\sqrt{\frac{1}{M^{**}} \sum_{i=1}^{M^{**}} \text{Var} \left((\mathbf{Q}^{i**})^T \Delta \mathbf{e} \right)} \leq \xi \sqrt{\frac{1}{M^{**}} \sum_{i=1}^{M^{**}} \text{Var} \left((\mathbf{Q}^{i**})^T \tilde{\mathbf{e}} \right)} \quad (69)$$

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 $\{\mathbf{Q}^{i**}\}_{i \in \llbracket 1, M^{**} \rrbracket}$. Indeed, there exists a severe risk of over-fitting if using set $\{\mathbf{Q}^{i*}\}_{i \in \llbracket 1, M^* \rrbracket}$ to build the reduced basis *and* monitor convergence. Unfortunately, this requires the computation of additional dual vectors. Alternatively, one could cross-validate the previous methodology, thereby avoiding this additional computational effort at the cost of a heavier implementation.

7 Results of Low Rank Approximation of Full Error Field Gaussian Process

We now implement the goal oriented hierarchical clustering approach described above to the example problem outlined in section 4.1. In the following, we consider a randomly selected 25 training duals 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 normalised to ensure partition of unity). A series of clusters are presented here to demonstrate the refining nature of the hierarchical approach, both at the macro and micro scales.

Clusters for the example macro problem are shown in figure 10, 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 67, here referred to as the clustering density) for a given cluster arrangement (and associated RBF space), noting the cluster within which the weighted summation of q is maximised, and performing k-medoids on the selected cluster. In the results presented here, 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

11 for the clusters in figure 10. The example problem considered in the present work limits the location of QoIs to two RoI boundaries (see figure 2). As a result, the error in low rank QoI estimations (characterised by the density q) is concentrated in the upper right corner of the macro model. This can be clearly seen in figure 11, 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 10, are presented in figure 12.

We must now consider how, and if, the clustering based RBF refinement leads to convergence of the low rank approximation of the Gaussian process prior. We may first consider the entropy of the difference between our “exact” error field (furnished without model order reduction of the prior) and the RBF approximation of it. Recall that equation 52 gives the covariance of the discrepancy field. In lieu of a complete set of QoIs (\mathbf{E}), we use the restricted set \mathbf{H} which is made up of the same QoIs that inform our GOEE observations (see equation 21). Through this substitution, we note that the calculation $\mathbf{A}^{-1}\mathbf{H}^T$ (used to evaluate the distribution in equation 51) becomes trivial, as it is equal to the solutions of the dual problems (\mathbf{z}) we have already solved. The entropy of the difference between exact and low rank error fields is given by

$$\frac{1}{2} \ln \det (2\pi \exp (\boldsymbol{\Sigma}_{\Delta} \mathbf{H})) \quad (70)$$

The evolution of this entropy term with cluster iteration can be seen figure 13. If the hierarchical clustering approach is effective we would expect a continued reduction in the discrepancy entropy, indicating a collapse of the discrepancy distribution as the low rank approximation approach the exact field. This feature is clearly demonstrated in figure 13. The near linear rate at which the entropy reduced with cluster iteration is somewhat surprising and suggests that, at least for the example problem, the improvements offered by cluster refinement are incremental despite the significant changes in q . Note that the entropy of the error field distribution itself is also presented in figure 13 for reference. We may also perform Monte-Carlo analyses on the micro models by sampling error correct DoF distributions that are generated by exact and low rank approximations of the Gaussian process terms (similar to the demonstrations performed in section 4.2). Peak von Mises stress may be extracted from the “2 circle” micro model (as used before) for the low rank approximations associated with each cluster refinement, as shown in figure 14. 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 sub spaces 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 macro model mesh. Given the limited number of DoFs driving the micro model (and therefore the small number of possible QoIs) 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 micro 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.

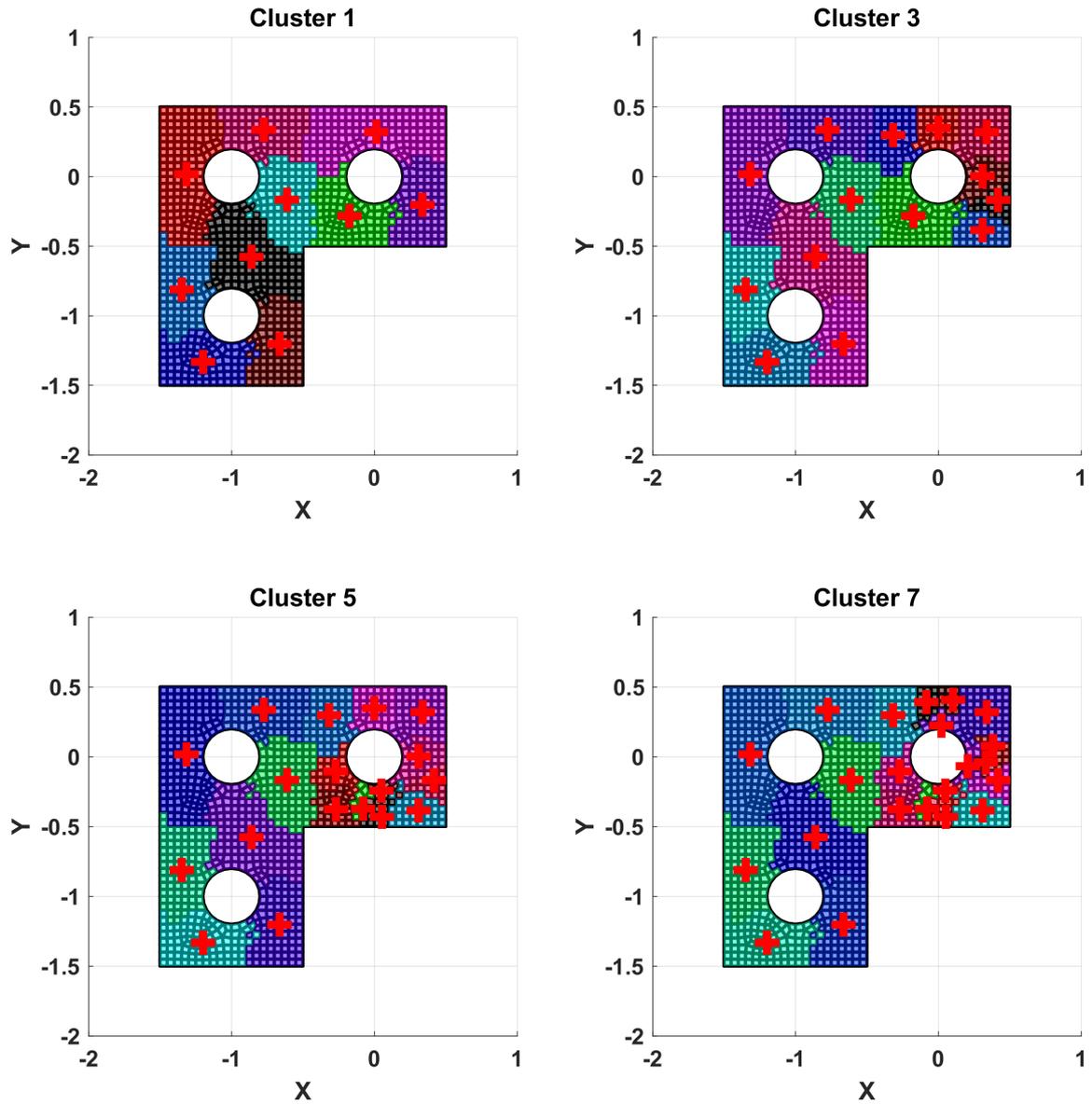


Figure 10: Agglomerative hierarchical clustering development for the example macroscopic problem, using 10 initial clusters. The crosses indicate the centroids of each cluster of vertices. Clusters selected by the defined density quantity are refined into 3 sub-clusters using K-means at each cluster step.

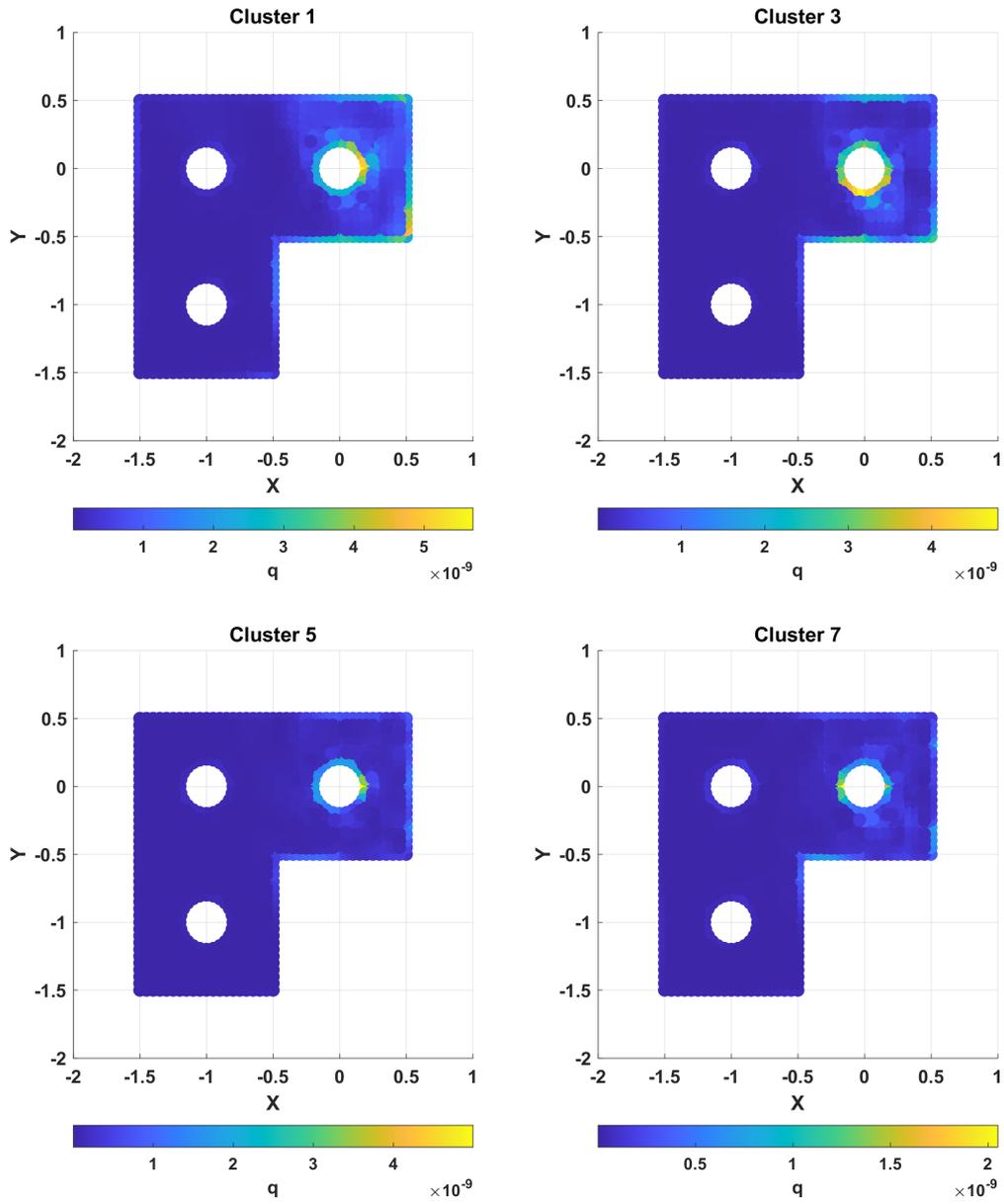


Figure 11: Clustering density (see equation 67) evolution during cluster refinement for the example macroscopic problem.

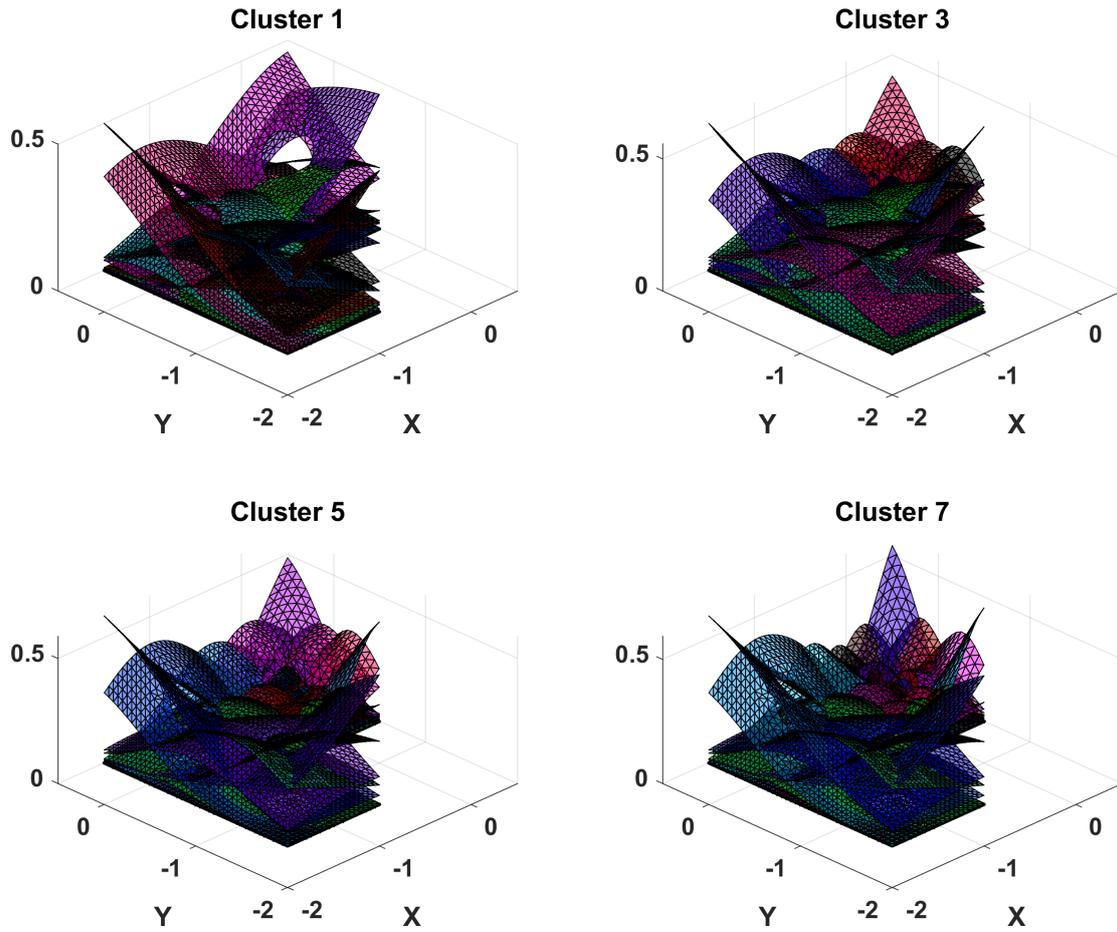


Figure 12: Agglomerative hierarchical clustering development for the example macroscopic problem, using 10 initial clusters, showing RBF definition for each cluster.

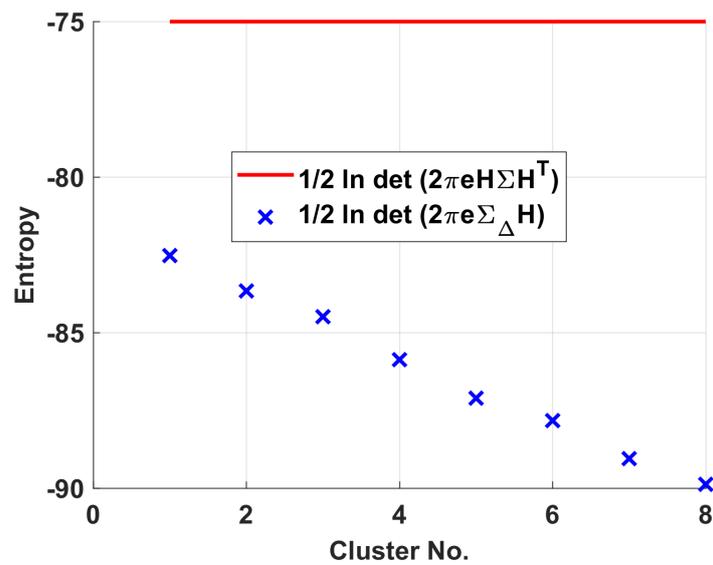


Figure 13: Entropy of the discrepancy between exact error field on the union of all micro boundaries and its approximation using adaptive RBFs.

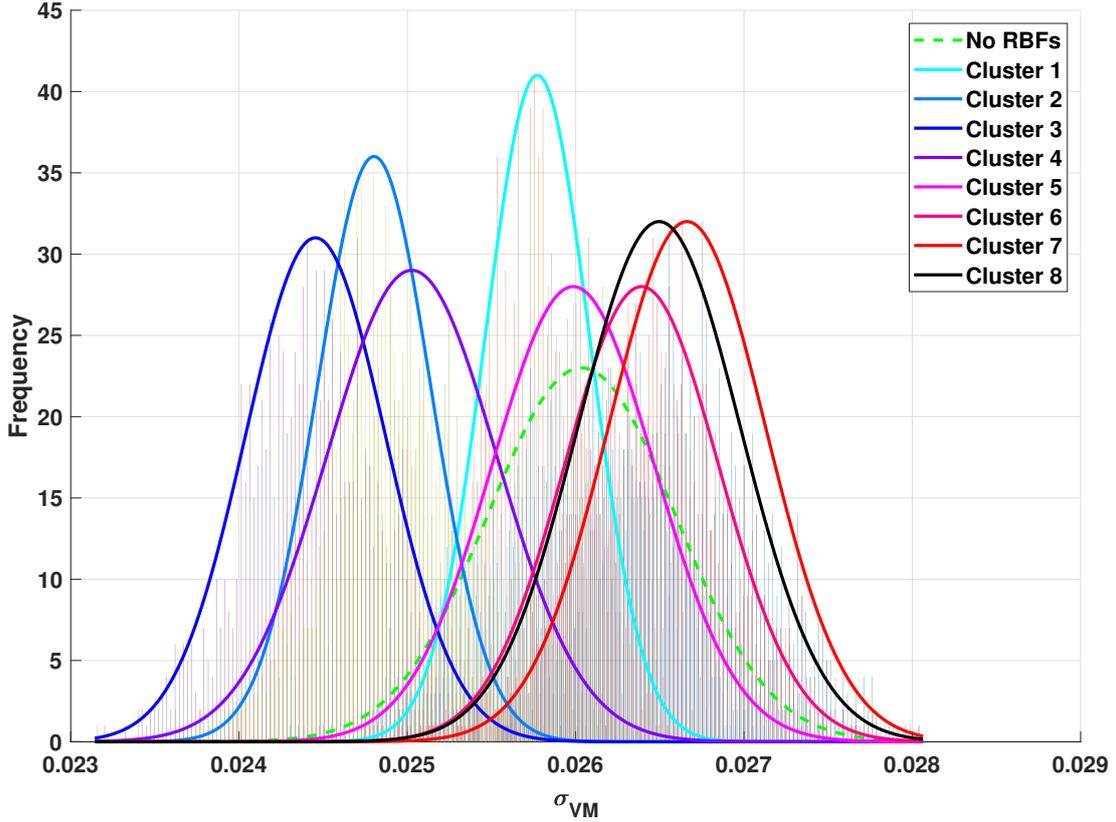


Figure 14: Peak von Mises equivalent stresses (σ_{VM}), taken from a series of Monte-Carlo error field propagations for the “2 hole” (position 2) microscopic model. Error fields that modify driving DoFs result from posterior distributions trained with 25 dual solutions. Projections are made using the cluster based RBF functions (with higher clustering numbers indicating subsequent clusterings that have made use of density driven refinement). A true “No RBFs” distribution is include for reference (this does not make use of any form of model order reduction in the calculation of the prior).

Perspectives and Conclusions

A statistical learning approach is presented here for the estimation of full error fields in large FE models. The method is unobtrusive - it does not require any re-meshing operation and is computationally efficient, owing to the low-rank SPDE-based gaussian process specifically developed in this work. The stochastic PDE representation of the prior error field allows for posterior realisations to be generated using sparse matrix inversions, thereby ensuring the cost of generating such realisations, and that of tuning the GP hyperparameters using maximum marginal likelihood, is comparable to the cost of solving the FEA problem itself. It is important to note that the adoption of a stochastic PDE 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-modelling multiscale problems, the methods are applicable to many alternative cases. For example, consider a large domain where there exist many potential localisation 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 localisation features. The error field recovery approach outlined here would furnish the analyst with confidence intervals for QoIs at the localisation 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 macro scale and features at the micro scale will be of interest. There is a need, therefore, to extract uncertainty information across the macro model such that this can be propagated down to various micro 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 also 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 extended to non-linear problems by linearising constitutive matrices (as in the work of Cirak [46] or that of Ghorashi and Rabczuk [74]). The integration of non-linearity at local levels is straightforward. Duals solutions (at the macro 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 dual solutions from previous increments being used to inform subsequent error field estimations [75]. 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 optimise the QoIs through adaptive learning/Greedy algorithms. Note that this can be accomplished in a simplistic way by characterising QoIs with the problem shape functions, however a more intriguing question lies in the full optimisation of QoIs in the problem domain.

References

- [1] M.S. Bonney, R. Evans, J. Rouse, A. Jones, P. Kerfriden, and M. Hamadi. Goal oriented error estimation in multi-scale shell element finite element problems. *Advanced Modeling and Simulation in Engineering Sciences*, 8, 2021.
- [2] Mark Ainsworth and J.Tinsley Oden. A posteriori error estimation in finite element analysis. *Computer Methods in Applied Mechanics and Engineering*, 142(1):1 – 88, 1997.
- [3] Thomas Grtsch and Klaus-Jrgen Bathe. A posteriori error estimation techniques in practical finite element analysis. *Computers & Structures*, 83(4):235 – 265, 2005.
- [4] Marco A. R. Ferreira and Herbert Lee. *Multiscale modeling: A Bayesian perspective*. Springer, New York, USA, 2007.
- [5] E. Weinan. *Principles of multiscale modeling*. Cambridge University Press, Cambridge, UK, 2011.
- [6] Stéphane Guinard, Robin Bouclier, Mateus Toniolli, and Jean-Charles Passieux. Multiscale analysis of complex aeronautical structures using robust non-intrusive coupling. *Advanced Modeling and Simulation in Engineering Sciences*, 5:1–27, 2018.
- [7] Bassam El Said, Federica Daghia, Dmitry Ivanov, and Stephen R. Hallett. An iterative multiscale modelling approach for nonlinear analysis of 3D composites. *International Journal of Solids and Structures*, 132–133:42–58, 2018.
- [8] R. Sturm, P. Schatrow, and Y. Klett. Multiscale modeling methods for analysis of failure modes in foldcore sandwich panels. *Applied Composite Materials*, 22:857–868, 2015.
- [9] James Lua, William Gregory, and Jagannathan Sankar. Multi-scale dynamic failure prediction tool for marine composite structures. *Journal of Materials Science*, 41:6673–6692, 2006.
- [10] Daniela Addressi and Elio Sacco. A multi-scale enriched model for the analysis of masonry panels. *International Journal of Solids and Structures*, 49:865–880, 2012.
- [11] Marc G. D. Geers, Varvara G. Kouznetsova, Karel Matou, and Julien Yvonnet. Homogenization methods and multiscale modeling: nonlinear problems. *Encyclopedia of Computational Mechanics Second Edition*, pages 1–34, 2017.
- [12] Hyung Joo Kim and Colby C. Swan. Voxel-based meshing and unit-cell analysis of textile composites. *International Journal of Numerical Methods in Engineering*, 56:977–1006, 2003.

- [13] L. Gendre, O. Allix, and P. Gosselet. A two-scale approximation of the schur complement and its use for non-intrusive coupling. *International Journal for Numerical Methods in Engineering*, 87(9):889–905, 2011.
- [14] P. Kerfriden, O. Allix, and P. Gosselet. A three-scale domain decomposition method for the 3D analysis of debonding in laminates. *Computational Mechanics*, 44:343–362, 2009.
- [15] Pierre Gosselet and Christian Rey. Non-overlapping domain decomposition methods in structural mechanics. *Archives of Computational Methods in Engineering*, 13:515–572, 2006.
- [16] Xi Zou, Shibo Yan, James Paul Rouse, I. Arthur Jones, Maxime Hamadi, and Michel Fouinneteau. A computationally efficient approach for analysing the onset of failure in aerospace composite structures. In *ICCM22, Melbourne, Australia*, 2019.
- [17] X. Zou, S. Yan, J. Rouse, M. Matveev, S. Li, I. A. Jones, M. Hamadi, and M. Fouinneteau. The identification of failure initiation hotspots in idealised composite material component models using a bottom-up database method. *Proceedings of the 18th European Conference on Composite Materials*, 2018.
- [18] X. Zou, S. Yan, M. Matveev, J. P. Rouse, I. A. Jones, M. Hamadi, and M. Fouinneteau. Comparison of interface modelling strategies for predicting delamination in composite l-angle sections under four-point bending. composite structures. *Journal of Composite Structures*, 2019. Submitted.
- [19] Matthew S. Bonney, Richard Evans, James Rouse, Arthur Jones, and Maxime Hamadi. Bayesian reconstruction of goal orientated error fields in large aerospace finite element models. *Proceedings of the Aerospace Europe Conference 2020*, 2020.
- [20] D. A. Paladim, J. P. Moitinho-de Almeida, S. P. A. Bordas, and P. Kerfriden. Guaranteed error bounds in homogenisation: an optimum stochastic approach to preserve the numerical separation of scales. *International Journal for Numerical Methods in Engineering*, 110(2):103–132, 2017.
- [21] Biao Liang, Weizhao Zhang, Joel S. Fenner, Jiaying Gao, Yi Shi, Danielle Zeng, Xuming Su, Wing K. Liu, and Jian Cao. Multi-scale modeling of mechanical behavior of cured woven textile composites accounting for the influence of yarn angle variation. *Composites Part A: Applied Science and Manufacturing*, 124:105460, 2019.
- [22] Baohui Shi, Man Zhang, Shengkai Liu, Baozhong Sun, and Bohong Gu. Multi-scale ageing mechanisms of 3D four directional and five directional braided composites impact fracture behaviors under thermo-oxidative environment. *International Journal of Mechanical Sciences*, 155:50–65, 2019.
- [23] Gang Liu, Li Zhang, Licheng Guo, Feng Liao, Tao Zheng, and Suyang Zhong. Multi-scale progressive failure simulation of 3D woven composites under uniaxial tension. *Composite Structures*, 208:233–243, 2019.
- [24] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems i: A linear model problem. *SIAM Journal on Numerical Analysis*, 28:43–77, 1991.
- [25] R. Verfürth. A posteriori error estimation and adaptive mesh-refinement techniques. *Journal of Computational and Applied Mathematics*, 50(1):67 – 83, 1994.
- [26] J. E. Bank and A. Weiser. Some a posteriori error estimators for elliptic partial differential equations. *Mathematics of Computation*, 44:283–301, 1985.
- [27] Pierre Ladevéze and Jean-Pierre Pelle. *Mastering Calculations in Linear and Nonlinear Mechanics*. Springer, 2005.

- [28] Fredrik Larsson and Kenneth Runesson. On two-scale adaptive fe analysis of micro-heterogeneous media with seamless scale-bridging. *Computer Methods in Applied Mechanics and Engineering*, 200(37):2662 – 2674, 2011. Special Issue on Modeling Error Estimation and Adaptive Modeling.
- [29] Junqi Zhang, Sundararajan Natarajan, Ean Tat Ooi, and Chongmin Song. Adaptive analysis using scaled boundary finite element method in 3d. *Computer Methods in Applied Mechanics and Engineering*, 372:113374, 2020.
- [30] Chongmin Song, Ean Tat Ooi, Aladurthi L N Pramod, and Sundararajan Natarajan. A novel error indicator and an adaptive refinement technique using the scaled boundary finite element method. *Engineering Analysis with Boundary Elements*, 94:10 – 24, 2018.
- [31] Alejandro Allendes, Csar Naranjo, and Enrique Otrola. Stabilized finite element approximations for a generalized boussinesq problem: A posteriori error analysis. *Computer Methods in Applied Mechanics and Engineering*, 361:112703, 2020.
- [32] Luca Heltai and Nella Rotundo. Error estimates in weighted sobolev norms for finite element immersed interface methods. *Computers & Mathematics with Applications*, 78(11):3586 – 3604, 2019.
- [33] M. Rech, S. Sauter, and A. Smolianski. Two-scale composite finite element method for dirichlet problems on complicated domains. *Numerische Mathematik*, 102:681–708, 2006.
- [34] Tamal Pramanick and Rajen Kumar Sinha. Error estimates for two-scale composite finite element approximations of parabolic equations with measure data in time for convex and nonconvex polygonal domains. *Applied Numerical Mathematics*, 143:112 – 132, 2019.
- [35] O. A. González-Estrada, J. J. Ródenas, S. P. A. Bordas, E. Nadal, P. Kerfriden, and F. J. Fuenmayor. Locally equilibrated stress recovery for goal oriented error estimation in the extended finite element method. *Computers & Structures*, 152:1–10, 2015.
- [36] Zhiqiang Cai, Seokchan Kim, and Hyung-Chun Lee. Error estimate of a finite element method using stress intensity factor. *Computers & Mathematics with Applications*, 76(10):2402 – 2408, 2018.
- [37] J. T. Oden and S. Prudhomme. Estimation of modeling error in computational mechanics. *Journal of Computational Physics*, 182:496–515, 2002.
- [38] N. Rahimi, P. Kerfriden, F. C. Langbein, and R. R. Martin. Cad model simplification error estimation for electrostatics problems. *Journal on Scientific Computing*, 40:B196–B227, 2018.
- [39] D. A. Paladim, J. P. Moitinho de Almeida, S. P. A. Bordas, and P. Kerfriden. Guaranteed error bounds in homogenisation: an optimum stochastic approach to preserve the numerical separation of scales. *International Journal for Numerical Methods in Engineering*, 110:103–132, 2017.
- [40] L. Chamoin and F. Legoll. Goal-oriented error estimation and adaptivity in msfem computations. *Computational Mechanics*, 67:1201–1228, 2021.
- [41] J. T. Oden and S. Prudhomme. Estimation of modeling error in computational mechanics. *Journal of Computational Mechanics*, 182:496–515, 2002.
- [42] Zhijia Lin, Zhuo Zhuang, Xiaochuan You, Heng Wang, and Dandan Xu. Enriched goal-oriented error estimation applied to fracture mechanics problems solved by xfem. *Acta Mechanica Solida Sinica*, 25(4):393 – 403, 2012.
- [43] M. Ainsworth and J. T. Oden. A unified approach to a posteriori error estimation using element residual methods. *Numerische Mathematik*, 65:23–50, 1993.

- [44] P. Ladevze, Ph. Rougeot, P. Blanchard, and J.P. Moreau. Local error estimators for finite element linear analysis. *Computer Methods in Applied Mechanics and Engineering*, 176(1):231–246, 1999.
- [45] L. Chamoin and P. Ladevéze. A non-intrusive method for the calculation of strict and efficient bounds of calculated outputs of interest in linear viscoelasticity problems. *Computer Methods in Applied Mechanics and Engineering*, 197:994–1014, 2008.
- [46] E. Cirak, F. and Ramm. A posteriori error estimation and adaptivity for linear elasticity using the reciprocal theorem. *Computer Methods in Applied Mechanics and Engineering*, 156:351–362, 1998.
- [47] J.T. Oden and S. Prudhomme. Goal-oriented error estimation and adaptivity for the finite element method. *Computers & Mathematics with Applications*, 41(5):735 – 756, 2001.
- [48] Roland Becker and Rolf Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta numerica*, 10(1):1–102, 2001.
- [49] Octavio Andrés González Estrada, E. Nadal, J. J. Ródenas, Pierre Kerfriden, Stéphane Pierre-Alain Bordas, and F. J. Fuenmayor. Mesh adaptivity driven by goal-oriented locally equilibrated superconvergent patch recovery. *arXiv e-prints*, September 2012.
- [50] Fehmi Cirak and Ekkehard Ramm. A posteriori error estimation and adaptivity for linear elasticity using the reciprocal theorem. *Computer Methods in Applied Mechanics and Engineering*, 156(1):351 – 362, 1998.
- [51] Marie Tirvaudey, Ludovic Chamoin, Robin Bouclier, and Jean-Charles Passieux. A posteriori error estimation and adaptivity in non-intrusive couplings between concurrent models. *Computer Methods in Applied Mechanics and Engineering*, 367:113104, August 2020.
- [52] E. T. Chung, W. T. Leung, and S. Pollock. Goal-oriented adaptivity for gmsfem. *Journal of Computational and Applied Mathematics*, 296:625–637, 2016.
- [53] L. Chamoin and F. Legoll. Goal-oriented error estimation and adaptivity in msfem computations. *arXiv preprint 1908.00367.21*, 2019.
- [54] Lassi Roininen, Janne MJ Huttunen, and Sari Lasanen. Whittle-Matérn priors for Bayesian statistical inversion with applications in electrical impedance tomography. *Inverse Probl. Imaging*, 8(2):561, 2014.
- [55] Mohinder S. Grewal and Angus P. Andrews. *Kalman Filtering: Theory and Practice with MATLAB*. Wiley, 2014.
- [56] Jean-Paul Chilés and Pierre Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. Wiley, 2012.
- [57] D. A. Paladim, J. P. Moitinho de Almeida, S. P. A. Bordas, and P. Kerfriden. Guaranteed error bounds in homogenisation: An optimum stochastic approach to preserve the numerical separation of scales. *International Journal for Numerical Methods in Engineering*, 110:103–132, 2017.
- [58] O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *International Journal for Numerical Methods in Engineering*, 24(2):337–357, 1987.
- [59] O.C. Zienkiewicz and J.Z. Zhu. The superconvergent patch recovery (spr) and adaptive finite element refinement. *Computer Methods in Applied Mechanics and Engineering*, 101:207–224, 1992.
- [60] O.A. González-Estrada, J.J. Ródenas, S.P.A. Bordas, E. Nadal, P. Kerfriden, and F.J. Fuenmayor. Locally equilibrated stress recovery for goal oriented error estimation in the extended finite element method. *Computers & Structures*, 152:1–10, 2015.

- [61] O.A. González-Estrada, J.J. Ródenas, S.P.-A. Bordas, M. Duflot, P. Kerfriden, and E. Giner. On the role of enrichment and statical admissibility of recovered fields in a posteriori error estimation for enriched finite element methods. *Engineering Computations*, 29(8), 2012.
- [62] O.A. González-Estrada, E. Nadal, J.J. Ródenas, P. Kerfriden, S.P.-A. Bordas, and F.J. Fuenmayor. Mesh adaptivity driven by goal-oriented locally equilibrated superconvergent patch recovery. *Computational Mechanics*, 53:957–976, 2013.
- [63] S. Prudhomme and J. T. Oden. On goal-oriented error estimation for elliptic problems: application to the control of pointwise errors. *Computer Methods in Applied Mechanics and Engineering*, 176:131–331, 1999.
- [64] N. C. Nguyen, H. Men, R. M. Freund, and J. Peraire. Gaussian functional regression for state prediction using linear PDE models and observations. *SIAM J. Sci. Comput*, 2015.
- [65] Ngoc Cuong Nguyen, Han Men, Robert M. Freund, and Jaume Peraire. Functional Regression for State Prediction Using Linear PDE Models and Observations. *SIAM Journal on Scientific Computing*, 38(2):B247–B271, 2016.
- [66] Per Sidén, Finn Lindgren, David Bolin, Anders Eklund, and Mattias Villani. Spatial 3D Mat\`ern priors for fast whole-brain fMRI analysis. *arXiv preprint arXiv:1906.10591*, 2019.
- [67] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian Processes for Machine Learning*, volume 2. MIT press Cambridge, MA, 2006.
- [68] C. Prud’homme, D. V. Rovas, K. Veroy, L. Machiels, Y. Maday, A. T. Patera, and G. Turinici. Reliable Real-Time Solution of Parametrized Partial Differential Equations: Reduced-Basis Output Bound Methods. *Journal of Fluids Engineering*, 124(1):70–80, 2002.
- [69] D.B.P. Huynh G. Rozza and A.T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: Application to transport and continuum mechanics. *Archives of Computational Methods in Engineering*, 15(3):229–275, 2008.
- [70] K.C. Hoang, P. Kerfriden, and S.P.A. Bordas. A fast, certified and “tuning free” two-field reduced basis method for the metamodeling of affinely-parametrised elasticity problems. *Computer Methods in Applied Mechanics and Engineering*, 298:121–158, 2016.
- [71] Marian Brezina, R. Falgout, Scott MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge. Adaptive smoothed aggregation (α SA) multigrid. *SIAM review*, 47(2):317–346, 2005.
- [72] Roland Becker and Rolf Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numerica 2001*, 10:1–102, 2001.
- [73] J.Tinsley Oden and Serge Prudhomme. Estimation of Modeling Error in Computational Mechanics. *Journal of Computational Physics*, 182(2):496–515, 2002.
- [74] S. Sh. Ghorashi and T. Rabczuk. Goal-oriented error estimation and mesh adaptivity in 3d elastoplasticity problems. *International Journal of Fracture*, 203:3–19, 2017.
- [75] M. Girolami, E. Febrianto, G. Yin, and F. Cirak. The statistical finite element method (statfem) for coherent synthesis of observation data and model predictions. *Computer Methods in Applied Mechanics and Engineering*, 375:113533, 2021.