Received: 27 October 2017 | Revised: 7 March 2019 | Accepted: 8 April 2019

DOI: 10.1002/nme.6078

RESEARCH ARTICLE

WILEY

Enhancing piecewise-defined surrogate response surfaces with adjoints on sets of unstructured samples to solve stochastic inverse problems

Steven A. Mattis[1] | Troy Butler[2]

[1]Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, Texas

[2]Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, Colorado

Correspondence
Steven A. Mattis, Zentrum Mathematik, Technische Universität München, Garching, Germany.
Email: mattis@ma.tum.de

Present Address
Steven A. Mattis, Zentrum Mathematik, Technische Universität München, Garching, Germany

Funding information
United States Department of Energy Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program, Grant/Award Number: DE-SC0009279 and DE-SC0009286; National Science Foundation, Grant/Award Number: DMS-1228206 and DMS-1818941; German Research Foundation, Grant/Award Number: DFG, Project WO 671/11-1

Summary

Many approaches for solving stochastic inverse problems suffer from both stochastic and deterministic sources of error. The finite number of samples used to construct a solution is a common source of stochastic error. When computational models are expensive to evaluate, surrogate response surfaces are often employed to increase the number of samples available for approximating the solution. This leads to a reduction in finite sampling errors while the deterministic error in the evaluation of each sample is potentially increased. The pointwise accuracy of sampling the surrogate is primarily impacted by two sources of deterministic error: the local order of accuracy in the surrogate and the numerical error from the numerical solution of the model. In this work, we use adjoints to simultaneously give a posteriori error and derivative estimates in order to construct low-order, piecewise-defined surrogates on sets of unstructured samples. Several examples demonstrate the computational gains of this approach in obtaining accurate estimates of probabilities for events in the design space of model input parameters. This lays the groundwork for future studies on goal-oriented adaptive refinement of such surrogates.

KEYWORDS

adjoint problems, a posteriori error estimate, inverse problems, Monte Carlo methods, surrogate modeling, uncertainty quantification

1 | INTRODUCTION

One of the most important problems in computational modeling is the identification and estimation of events associated with quantities of interest (QoIs) computed from the solution to a model, eg, conditions for a QoI corresponding to system failure. This represents a type of predictive (ie, forward) modeling problem. When using computational models to analyze the reliability or robustness of a system, we are often tasked with solving the inverse problem of identifying sets of model input parameters that map to such QoI events. Generally speaking, any limitations that exist in solving the forward problem impact the accuracy of the solution to the inverse problem. For example, if the computational model is expensive to evaluate, it may only be possible to evaluate the model at a small number of parameter samples impacting the accuracy of set approximations based on these samples. Moreover, there are often several sources of uncertainties and

errors in the identification of QoI events to be inverted, eg, due to modeling errors of the system and measurement errors of system outputs used to define QoI events.

When probabilities are used to represent uncertainties of QoI events, we solve a type of stochastic inverse problem where the objective is to estimate the probabilities of parameter events defined by applying the inverse of the (vector-valued) QoI map to the QoI events. In the uncertainty quantification (UQ) community, the problem of efficient and accurate propagation of probabilistic uncertainty through QoI maps has received significant attention over the last few decades.[1,2] Monte Carlo techniques[3,4] and their variants (eg, Markov chain Monte Carlo[5,6]) are commonly used to solve UQ problems due to their relative ease of implementation and the fact that they can circumvent, with various degrees of success, the so-called curse of dimensionality. However, convergence is typically slow, and the number of samples required to accurately estimate probabilistic quantities may be prohibitively large for models of even moderate computational cost.

One way to reduce the impact of finite sampling error is to construct a surrogate to the QoI response surface such that sampling has a lower computational cost. The past several decades have witnessed tremendous advancement in the development and use of surrogates to propagate uncertainties, making a full review of such techniques virtually impossible. While it is possible to trace back the roots of much of the work involving stochastic finite element approaches[1,7] for building surrogates to the seminal papers of Wiener[8] and Cameron and Martin,[9] there are now many accessible starting points that use similar ideas for building global polynomial approximations based on stochastic spectral methods.[2,10-16] Tensor grid and sparse grid stochastic collocation methods for building surrogates have also gained in popularity.[17,18] There is also interesting new research on using dimension reduction techniques and reduced-order models for building surrogates.[19] The surrogate modeling approach considered in this work most closely resembles techniques that exploit derivative information or error estimates using adjoints for building piecewise low-order surrogate approximations to improve pointwise accuracy in propagations of uncertainties.[13,20,21]

The surrogate response surface is polluted by two sources of error affecting the accuracy of all samples[15,16,20]: (1) the approximation error of the choice of surrogate and (2) the numerical error in evaluation of the samples used to construct the surrogate. We refer to both these types of error as discretization errors. Thus, using a surrogate can represent a trade-off between reduction in finite sampling error since more samples may be used, but at the expense of an overall increase in discretization error due to the use of the surrogate in obtaining these samples. The end result is that our ability to utilize the solution of a stochastic inverse problem to accurately quantify uncertainties may be compromised by the use of surrogates unless additional steps are taken to reduce the discretization errors.

The derivation of computable and accurate a posteriori estimates of discretization errors based on variational techniques and adjoints dates back several decades.[22-27] Such techniques served as the basis of the error estimates for polynomial chaos and pseudospectral-based surrogates derived in the works of Butler et al.[14,28] Subsequently, in the works of Prudhomme and Bryant[15] and Butler et al,[20] such error estimates were used as part of a Bayesian inference to quantify uncertainties on parameters to evolutionary partial differential equations (PDEs) where QoI response surfaces were approximated with polynomial chaos techniques and enhanced by the error estimates. Recently, such techniques have been applied in a reduced-order setting, giving error estimates for QoIs calculated with full models and reduced-order models[29] and for the development of adaptive schemes for mesh refinement for reduced-order models.[30]

This motivates the major contributions of this work, which are the use of adjoint-based techniques to estimate and correct for numerical error in the surrogate while simultaneously increasing the local order of the surrogate response surface. The use of the resulting enhanced surrogates is 2-fold, where we observe an increase in accuracy and decrease in computational complexity in the computation of probabilities of specified events. To highlight the impact this type of surrogate construction may have on other UQ problems, we consider a nonintrusive sample-based method for constructing pullback measures that require *global* accuracy in the surrogate and a significant number of samples. Moreover, we define the surrogates, implicitly, on unstructured sets of random samples within the parameter space defining a type of Voronoi piecewise surrogate. This lays the groundwork for future studies that focus on adaptive refinement of such surrogates and extensions to other UQ problems and algorithms. For example, Voronoi piecewise surrogates (without adjoint-based enhancement) were recently used[31] for high-dimensional data fitting and applied to several interesting, but distinct, UQ problems than those considered here. Additional studies[32,33] provide templates for both adaptivity techniques and constructing Voronoi piecewise surrogates in arbitrary dimensions.

This paper is organized as follows. We provide some general notation, terminology, and assumptions used in this work in Section 2. A brief review of adjoint-based a posteriori error and derivative estimates along with a list of useful references are provided in Section 3. In Section 3, we also describe how we use such error estimates to enhance surrogates by correcting for persistent local biases due to discretization errors. In Section 4, we provide a brief summary of the theory behind the formulation and solution of the type of stochastic inverse problem considered in this work, summarize the random

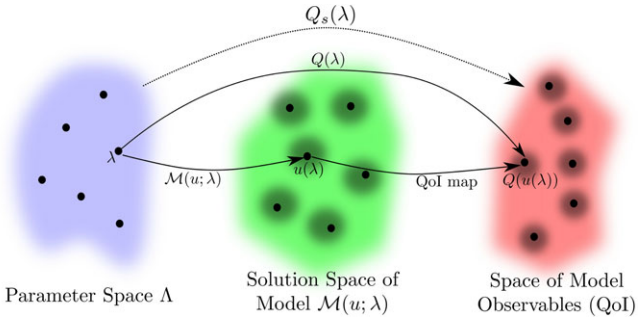

FIGURE 1 Schematic of mappings between relevant spaces. QoI, quantity of interest [Colour figure can be viewed at wileyonlinelibrary.com]

sample-based approximation to pullback measures, and identify sources of errors. We provide some numerical examples in Section 5, and concluding remarks follow in Section 6.

2 | NOTATION AND TERMINOLOGY

2.1 | Spaces and mappings

We assume that a (deterministic) model $\mathcal{M}(u;\lambda) = 0$ of a physical system is given, where u denotes a vector of state variables determined by the solution of the model for a specified vector of (model) parameters λ, which may include initial conditions, boundary data, or source terms. In other words, specification of λ determines the setup of the system being modeled. We assume that the space of possible parameters, denoted by $\mathbf{\Lambda}$, is also known. Then, for each $\lambda \in \mathbf{\Lambda}$, the solution operator of $\mathcal{M}(u;\lambda) = 0$ determines a particular $u(\lambda)$, where we make the dependence of the solution on the model parameters explicit. In Figure 1, we illustrate the mapping by connecting a particular sample, labeled λ, in the parameter space to a particular solution, $u(\lambda)$, by the arrow labeled with $\mathcal{M}(u;\lambda)$.

The vector of (linear) functionals corresponding to model observables defines a (vector-valued) QoI map, denoted by Q, and $Q(u(\lambda))$ represents a particular output datum associated to a particular choice of parameters defined by λ. We then write $Q(\lambda) := Q(u(\lambda))$ to emphasize both the dependence of the output data on the model parameters and the fact that, in an experimental setting, we may be able to control λ to observe $Q(\lambda)$ without fully observing $u(\lambda)$. We let $\boldsymbol{\mathcal{D}} := Q(\mathbf{\Lambda})$ denote the space of model observables. This is illustrated in Figure 1, where a particular solution $u(\lambda)$ in the middle is mapped by the arrow labeled by the QoI map to a model observable on the right denoted by $Q(u(\lambda))$. Oftentimes, we only record information of the map between $\mathbf{\Lambda}$ and $\boldsymbol{\mathcal{D}}$ defined by $Q(\lambda)$, as illustrated by the longer arrow labeled by $Q(\lambda)$ in Figure 1. This is often true in cases where it is not possible to store the entire numerical state space of the model.

We assume that $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}}, \mu_{\mathbf{\Lambda}})$ and $(\boldsymbol{\mathcal{D}}, \mathcal{B}_{\boldsymbol{\mathcal{D}}}, \mu_D)$ are measure spaces where $\mu_{\mathbf{\Lambda}}$ and μ_D are measures used to describe sizes of events in the σ-algebras $\mathcal{B}_{\mathbf{\Lambda}}$ and $\mathcal{B}_{\boldsymbol{\mathcal{D}}}$, respectively. We also assume that Q is at least piecewise differentiable, which implies measurability of the QoI map. Unless otherwise stated, we assume that $\mathbf{\Lambda} \subset \mathbb{R}^n$ and $\boldsymbol{\mathcal{D}} \subset \mathbb{R}^d$ and that $\mu_{\mathbf{\Lambda}}$ and μ_D are the "volume" measures defined by the Lebesgue measures on these spaces.

In the case where the model is solved numerically, we generally compute $Q_h(\lambda) := Q(u_h(\lambda))) \approx Q(\lambda)$. Here, $u_h(\lambda)$ denotes a numerical approximation to $u(\lambda)$, and h denotes a discretization parameter (eg, determined by the mesh size or number of maximum iterations in the solution method). Then,

$$u_h(\lambda) = u(\lambda) + \epsilon_{u,h}(\lambda),$$

where $\epsilon_{u,h}(\lambda)$ is the numerical error in $u_h(\lambda)$. We also write

$$Q_h(\lambda) = Q(\lambda) + \epsilon_{Q,h}(\lambda),$$

where $\epsilon_{Q,h}(\lambda)$ is the numerical error in $Q_h(\lambda)$. We may view these errors as a type of perturbation from the exact solution or data. An a priori error analysis often can be used to determine bounds on $\epsilon_{u,h}(\lambda)$ and $\epsilon_{Q,h}(\lambda)$, which define the maximum perturbations to the exact solutions and data we can expect from numerically solving the model. This is illustrated by the shaded areas around the exact solutions and QoI data in the middle and right spaces of Figure 1, indicating the magnitude of possible perturbations defined by such an a priori error analysis.

2.2 | Surrogate modeling

Let $Q_s(\lambda)$ denote a computationally inexpensive surrogate approximation to $Q(\lambda)$; thus,

$$Q_s(\lambda) = Q(\lambda) + \epsilon_s(\lambda),$$

where $\epsilon_s(\lambda)$ is the error in the surrogate due to limited approximation properties of the surrogate. We use $Q_s(\lambda)$ to more efficiently map large numbers of samples between $\boldsymbol{\Lambda}$ and $\boldsymbol{D}$ as indicated by the dotted-arrow mapping between these spaces in Figure 1. However, constructing $Q_s(\lambda)$ often requires using some particular set of samples of $Q_h(\lambda)$ based on a specific type of sampling in $\boldsymbol{\Lambda}$, eg, using a possibly different set of random samples or using deterministic sampling approaches such as sparse grids.[17,18] We let $Q_{s,h}(\lambda)$ denote this numerically constructed surrogate and $\epsilon_{s,h}(\lambda)$ denote its error. We decompose the error as

$$\epsilon_{s,h}(\lambda) := \epsilon_s(\lambda) + \epsilon_h(\lambda),$$

where $\epsilon_h(\lambda)$ is the error in the surrogate from the numerical solution of the model. In other words, the dotted arrow in Figure 1 is replaced by a mapping that is polluted by multiple sources of error that affect all sets of propagated samples between the spaces.

2.2.1 | Voronoi-based piecewise surrogate

Consider the surrogate $Q_{s,p}(\lambda)$ defined as a local pth-order polynomial function on a set of Voronoi cells $\{\mathcal{V}_k\}_{1\le k\le N_{s,p}}$ *implicitly* defined by a sample set $\{\lambda^{(k)}\}_{1\le k\le N_{s,p}}$, ie,

$$Q_{s,p}(\lambda) = \sum_{1\le k\le N_{s,p}} f_p(\lambda)\chi_{\mathcal{V}_k}(\lambda), \tag{1}$$

where $f_p(\lambda)$ denotes a pth-order polynomial approximating $Q_h(\lambda)$ on $\mathcal{V}_k$. If $p = 0$, then (1) represents a simple function approximation to Q, and it is possible to prove error bounds and adaptive sampling schemes for improving probabilistic quantities computed using $Q_{s,0}$.[34]

We emphasize that we *never construct the Voronoi cells* in order to evaluate $Q_{s,p}$ for any p. Instead, when evaluating $Q_{s,p}$ at a λ not belonging to the sample set, we rely upon nearest-neighbor searches (and in this work, we use the standard Euclidean metric for all examples). We discuss the computational cost of this below.

2.2.2 | Computational considerations

We are interested in problems where evaluation of $Q_h(\lambda)$ is expensive (eg, involving the solution of a finite element model for a PDE). Therefore, the size of the sample set

$$\{\lambda^{(k)}\}_{1\le k\le N_{s,p}},$$

for which we numerically solve the model to determine $Q_h(\lambda)$ and construct $Q_{s,p}$, may be constrained by a computational budget. However, to reduce finite sampling error in computing a probabilistic quantity, we often require $N \gg N_{s,p}$ approximations of the QoI on a different sample set

$$\{\lambda^{(j)}\}_{1\le j\le N}.$$

The following a priori analysis can determine if constructing and utilizing the surrogate is efficient on such a set of N samples. A nearest-neighbor search of $\lambda^{(j)}$ to $\{\lambda^{(k)}\}_{1\le k\le N_{s,p}}$ has an initial cost of $N_{s,p}$ metric computations (eg, approximately $2N_{s,p}n$ FLOPS, where n is the dimension of $\boldsymbol{\Lambda}$) followed by a sort (at a cost of $O(N_{s,p})$ FLOPS using a linear search). Let $\boldsymbol{C}$ denote the number of FLOPS required in evaluating Q_h at a single parameter. Whenever

$$C > 2N_{s,p}n + O(N_{s,p}),$$

it is generally computationally cheaper to construct and evaluate the surrogate than the full computational model to approximate Q_h at a set of N samples. However, while this may make finite sampling error negligible by allowing for significantly larger N, the deterministic error is likely significantly worse since each sample is now polluted by both the errors $\epsilon_s(\lambda)$ and $\epsilon_h(\lambda)$ present in the surrogate.

3 | ADJOINT-BASED A POSTERIORI ERROR ESTIMATION AND SURROGATE ENHANCEMENT

We summarize the general variational analysis and use of adjoint problems to derive a posteriori error estimates for QoI. The interested reader should refer to other works for more information on the theory and implementation of adjoint-based a posteriori error estimates in general[22-27] and on the application to certain classes of surrogate models.[14-16,20,28] For a more thorough introduction to the theory and application of adjoints in general, the reader is also referred to other recommended works.[35-38]

3.1 | Adjoint-based a posteriori error estimates and derivatives

For the sake of simplicity, we initially assume that the solution to the model $\mathcal{M}(u;\lambda)=0$ is defined by the solution to the parameterized linear system

$$A(\lambda)\mathbf{u}=\mathbf{b}(\lambda), \tag{2}$$

where, for each $\lambda\in\mathbf{\Lambda}\subset\mathbb{R}^n$, $\mathbf{b}(\lambda)\in\mathbb{R}^m$ and $A(\lambda)\in\mathbb{R}^{m\times m}$ are invertible. Then, for each $\lambda\in\mathbf{\Lambda}$, there exists a solution $\mathbf{u}(\lambda)\in\mathbb{R}^m$. We also initially assume that the QoI map is given by a single scalar functional defined by $Q(\lambda)=\langle\mathbf{u}(\lambda),\boldsymbol{\psi}\rangle$, where $\boldsymbol{\psi}\in\mathbb{R}^m$ and $\langle\cdot,\cdot\rangle$ denotes the standard Euclidean inner product. We emphasize that even for a linear model and linear functional defining the QoI, the response of the map $Q(\lambda)$ is often *nonlinear* over $\mathbf{\Lambda}$ as we demonstrate in an example below. The adjoint problem to Equation (2) is

$$A(\lambda)^{\top}\boldsymbol{\phi}=\boldsymbol{\psi}, \tag{3}$$

where $\boldsymbol{\phi}(\lambda)$ is the solution to the adjoint problem (often called generalized Green's vector), and $\boldsymbol{\psi}$ is determined by the QoI and independent of λ. Suppose, for a fixed $\lambda\in\mathbf{\Lambda}$, we numerically solve Equation (2) to obtain $\mathbf{u}_h(\lambda)\approx\mathbf{u}(\lambda)$ and subsequently compute $Q_h(\lambda)\approx Q(\lambda)$. Recall that

$$\epsilon_{Q,h}(\lambda):=Q_h(\lambda)-Q(\lambda).$$

Without the exact value of $Q(\lambda)$, $\epsilon_{Q,h}(\lambda)$ is uncomputable. Using a standard variational analysis and properties of inner products and linear operators, we have

$$\begin{aligned}\epsilon_{Q,h}(\lambda)&=\langle\mathbf{u}_h(\lambda)-\mathbf{u}(\lambda),\boldsymbol{\psi}\rangle\\&=\left\langle\mathbf{u}_h(\lambda)-\mathbf{u}(\lambda),A(\lambda)^{\top}\boldsymbol{\phi}(\lambda)\right\rangle\\&=\langle A(\lambda)\mathbf{u}_h(\lambda)-A\mathbf{u}(\lambda),\boldsymbol{\phi}(\lambda)\rangle\\&=\langle A(\lambda)\mathbf{u}_h(\lambda)-\mathbf{b}(\lambda),\boldsymbol{\phi}(\lambda)\rangle\,.\end{aligned} \tag{4}$$

When $\boldsymbol{\phi}(\lambda)$ is given, then Equation (4) is computable and gives the exact error $\epsilon_{Q,h}(\lambda)$. Generally, $\boldsymbol{\phi}(\lambda)$ is approximated by $\boldsymbol{\phi}_h(\lambda)$, and replacement of $\boldsymbol{\phi}(\lambda)$ with $\boldsymbol{\phi}_h(\lambda)$ in Equation (4) gives a computable a posteriori error estimate, which we denote by $e_{Q,h}(\lambda)$. Typically, we compute $\boldsymbol{\phi}_h(\lambda)$ using a higher-order method than used to compute $\mathbf{u}_h(\lambda)$.

Let λ_i denote the ith component of the vector λ for $1\le i\le n$. Then, differentiating Equation (2) with respect to λ_i and following a similar set of steps, we arrive at

$$\partial_{\lambda_i}Q_h(\lambda)=\left\langle\partial_{\lambda_i}\mathbf{b}(\lambda)-\left[\partial_{\lambda_i}A(\lambda)\right]\mathbf{u}_h(\lambda),\boldsymbol{\phi}(\lambda)\right\rangle. \tag{5}$$

The partial derivatives of $\mathbf{b}(\lambda)$ and $A(\lambda)$ can often be determined by algorithmic/automatic differentiation.[39] Subsequently, this implies that the gradient of the QoI with respect to the parameter λ, denoted by $\nabla_\lambda Q(\lambda)$, can be approximated by solving both the model and adjoint exactly once and then computing a finite number of inner products given by Equation (5).

3.1.1 | Generalization to other models and computational costs

The above approach can be applied to most models defined by a linear operator where only a few specific details change. Two excellent and comprehensive works on this subject are those by Bangerth and Rannacher[40] and Becker and Rannacher.[41] When the operator defining the model is nonlinear, we must linearize the model operator prior to defining the adjoint problem. One approach uses the same linear operator that is used in computing a step of Newton's method, in which case the remainder term is typically a higher-order perturbation term that is often neglected.[41] The methodology

presented above is also applicable to problems framed in continuous settings after making some suitable adjustments to the derivations. In a continuous setting, the forward and adjoint solution spaces are infinite dimensional (eg, L^2 spaces). There is an a posteriori error estimate with a similar form to Equation (4) where the Euclidean inner product is replaced by a duality pairing. In the interest of both simplicity and generality, we present the process of obtaining the a posteriori error estimate abstractly for problems of the form

$$\mathcal{L}(\lambda)u(\lambda) = f(\lambda), \tag{6}$$

where $\mathcal{L}(\lambda)$ is a linear differential operator, $f(\lambda)$ is the source data, and both $\mathcal{L}(\lambda)$ and $f(\lambda)$ may depend independently or jointly on various components of the model parameter vector λ. We assume that appropriate boundary/initial conditions are also posted as part of the problem and, for simplicity, that these do not depend on the components of λ. Denote by $\mathcal{L}^*(\lambda)$ the adjoint operator, then the adjoint problem is of the form

$$\mathcal{L}^*(\lambda)\phi(\lambda) = \psi, \tag{7}$$

for some appropriate choice of adjoint data ψ that is determined by the choice of the QoI map and is independent of λ. Suppose the function space for both the forward and adjoint solutions is a Hilbert space X with an inner product $\langle , \rangle_X$. In this setting, the ψ coming from the QoI map is often given by the Riesz representation of the QoI function in the dual space of X. For example, when $X = H_0^1(\Omega)$ (where Ω is some spatial domain), it is common for a QoI to be represented by $\psi \in L^2(\Omega) \subset H^{-1}(\Omega)$, where $H^{-1}(\Omega)$ is the dual space of $H_0^1(\Omega)$.

In a posteriori error analysis for finite elements, it is common to use different finite element spaces $V_f, V_a \subset X$ for the forward and adjoint problems. If $V_f = V_a$, an error estimate is likely zero due to Galerkin orthogonality; hence, it is preferable to use a more refined finite element space for the adjoint problem. A common choice is to use a polynomial approximation one order higher for V_a than for V_f using the same mesh. In this type of formulation of the adjoint problem, the adjoint is often called the *continuous adjoint*. In this setting, the ψ used in computations often comes from an L^2-projection or interpolation of the exact ψ onto the mesh, and the approximate QoI map is defined by

$$Q_h(\lambda) = \langle u_h(\lambda), \psi \rangle_X. \tag{8}$$

Then, the error representation in Equation (4) becomes

$$\epsilon_{Q,h}(\lambda) = \langle \mathcal{L}(\lambda)u_h(\lambda) - f(\lambda), \phi_h(\lambda) \rangle_X, \tag{9}$$

where $u_h \in V_f$ and $\phi_h \in V_a$ and similar changes are made for the derivative calculation in Equation (5). It is also common to formulate an adjoint problem of a PDE system after discretization, which is often called the *discrete adjoint* and is equivalent to the adjoint in the finite-dimensional setting discussed in the previous section, where the operator A is the discretized PDE operator and b is the discretized right-hand side. If an accurate linear solver is used to solve the linear system for the forward problem, it is preferable to pose the adjoint problem in a continuous setting with a refined space in order to obtain a more accurate error estimate. We discuss some of these details in the numerical examples of Section 5, including a detailed derivation of the adjoint operator $\mathcal{L}^*(\lambda)$ for a contaminant transport problem.

A few remarks about the computational costs of using adjoints for differential equations are in order. First, if only derivative information is required, then we do not need to avoid Galerkin orthogonality, and the adjoint has the same general computational complexity and cost as the forward problem. In this work, we solve all adjoints to differential equations by increasing the local polynomial order of the finite element methods by exactly one instead of refining the mesh to avoid Galerkin orthogonality and obtain both error and derivative estimates. Thus, in this work, the solution to the adjoint is approximately 4 to 8 times as expensive as the solution to the forward problem depending on if the solver requires either $\mathcal{O}(n^2)$ or $\mathcal{O}(n^3)$ FLOPS, where n denotes the degrees of freedom (which are generally doubled when going from linear to quadratic order as done in this work).

3.1.2 | Example

We present an illustrative example of a parameterized linear system. Suppose in Equation (2) that

$$A(\lambda) = \begin{bmatrix} e^{\lambda} & \cos(\pi\lambda) \\ \sin(\pi\lambda) & 2e^{\lambda} \end{bmatrix} \tag{10}$$

and

$$\mathbf{b}(\lambda) = \begin{bmatrix} \sin(10\pi\lambda) & 2e^{\lambda} \end{bmatrix}^T, \tag{11}$$

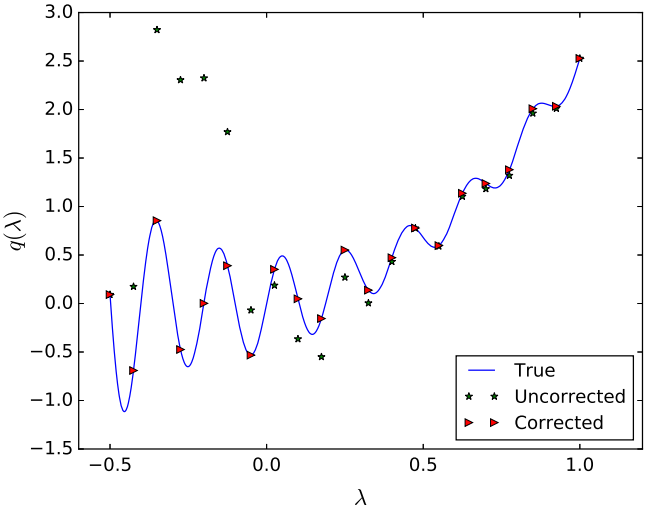

FIGURE 2 The quantity of interest map for Example 1 solved exactly ($Q(\lambda)$) and approximately ($Q_h(\lambda)$) and enhanced with the error estimate ($Q_h(\lambda) + \epsilon_{Q,h}(\lambda)$) [Colour figure can be viewed at wileyonlinelibrary.com]

for $\lambda \in \mathbf{\Lambda} = [-0.5, 1]$. The QoI map Q is defined as the average of the solution $\mathbf{u}$ as follows:

$$Q(\lambda) = \boldsymbol{\psi}^\top \mathbf{u}, \tag{12}$$

where $\boldsymbol{\psi} = [0.5, 0.5]^\top$, defining a nonlinear 1D-to-1D map from $\mathbf{\Lambda}$ to $\boldsymbol{\mathcal{D}} = Q(\mathbf{\Lambda})$. The approximate map $Q_h(\lambda)$ is evaluated by approximately solving the system defined by Equations (10) and (11) for $\mathbf{u}_h$ using two Gauss-Seidel iterations and evaluating the QoI function using Equation (12).

Figure 2 shows that the QoI map is highly nonlinear with respect to λ. The reference $Q(\lambda)$ is calculated by solving Equation (2) using direct numerical inversion. The approximate QoI map $Q_h(\lambda)$ is calculated for several values of λ. We see that, in some regions of $\mathbf{\Lambda}$, $Q_h(\lambda)$ is a very good approximation, and, in some places, it is a very bad approximation. Without performing error estimation, such behavior of the error often cannot be predicted. We solve the discrete adjoint system using direct numerical inversion to compute error estimates given by Equation (4). Using the adjoint-based computed error estimates at the samples in $\mathbf{\Lambda}$, we can improve the pointwise accuracy of the $Q_h(\lambda)$ estimates at each sample, as shown in Figure 2.

3.2 | Enhancing surrogates with solutions to adjoint problems

Traditionally, a posteriori error estimates of QoI from differential equation models derived by a variational analysis and adjoints were used to guide local h- or p-adaptivity, ie, mesh or order refinement, respectively, in the numerical solution to the model (eg, see the work of Becker and Rannacher[41] and the references therein). We consider using the adjoint solutions to enhance the surrogate model $Q_{s,p}$ given by Equation (1) using a two-stage approach that leaves the computational method for solving the model intact (ie, we do not engage in mesh refinement or other alterations to the numerical evaluation of Q_h at any sample point).

For simplicity, we demonstrate p-enhancement of $Q_{s,0}$ to $Q_{s,1}$ and note that there are other ways to perform local p-refinement than what is considered here using adjoints, eg, by using nets of nearby samples to perform regression or interpolation in a particular Voronoi cell.[31] Future research will focus on melding adjoint-based derivative estimates along with other approaches such as those considered in the work of Rushdi et al[31] to enrich p-refinement further and either mitigate or avoid extrapolation errors in the surrogate that can become significant in higher dimensions.

3.2.1 | *p*-Enhancement

In the first stage, let $N_{s,1} \leq N_{s,0}$ denote the number of numerical solutions to the model $\mathcal{M}(u; \lambda) = 0$ such that we can also solve $N_{s,1}$ adjoint problems and compute the necessary inner products to construct $\nabla_\lambda Q_h(\lambda)$ for each of the $N_{s,1}$ samples in

$\mathbf{\Lambda}$. For simplicity, we order the samples in $\mathbf{\Lambda}$ so that for $\lambda^{(k)}$ with $1 \leq k \leq N_{s,1}$ corresponds to a sample where we solved the adjoint problem. Then, we construct the piecewise-defined surrogate

$$Q_{s,1}(\lambda) = \sum_{1\leq k\leq N_{s,1}} \left[Q_h(\lambda^{(k)}) + \nabla_\lambda Q_h(\lambda^{(k)})(\lambda - \lambda^{(k)})\right] \chi_{\mathcal{V}^{(k)}}(\lambda) + \sum_{N_{s,1}<k\leq N_{s,0}} Q_h(\lambda^{(k)})\chi_{\mathcal{V}^{(k)}}(\lambda). \tag{13}$$

This represents a local p-adaptive refinement to the piecewise-constant surrogate $Q_{s,0}$. If $N_{s,1} = N_{s,0}$, then we drop the rightmost term in Equation (13). For simplicity below, we assume that $N_{s,1} = N_{s,0}$.

3.2.2 | Error-enhancement

Assuming we solve the $N_{s,1}$ adjoint problems with a higher-order method to produce a set of reliable a posteriori error estimates,

$$\{e_{Q,h}(\lambda^{(k)})\}_{1\leq k\leq N_{s,1}},$$

and then, we proceed to stage 2. In this stage, we correct for the persistent local bias due to the error $\epsilon_{Q,h}(\lambda^{(k)})$ polluting the evaluation of any $\lambda \in \mathcal{V}_k$ by using the enhanced surrogate

$$\hat{Q}_{s,1}(\lambda) = \sum_{1\leq k\leq N_{s,1}} \left[Q_h(\lambda^{(k)}) + e_{Q,h}(\lambda^{(k)}) + \nabla_\lambda Q_h(\lambda^{(k)})(\lambda - \lambda^{(k)})\right] \chi_{\mathcal{V}^{(k)}}(\lambda). \tag{14}$$

In some cases, calculating the derivatives in stage 1 may be computationally expensive or difficult to implement, whereas the error estimates for stage 2 may be more easily computable. In such cases, an enhanced piecewise-constant surrogate can be constructed by only performing stage 2 by

$$\hat{Q}_{s,0}(\lambda) = \sum_{1\leq k\leq N_{s,1}} \left[Q_h(\lambda^{(k)}) + e_{Q,h}(\lambda^{(k)})\right] \chi_{\mathcal{V}^{(k)}}(\lambda). \tag{15}$$

It is possible to also correct for numerical errors in the approximation of $\nabla_\lambda Q_h(\lambda)$. However, this generally requires solving adjoints to the approximate adjoint in order to estimate such errors. While many computational models are now being developed with adjoint capabilities, they are generally developed without this type of sensitivity error estimation in mind; hence, we neglect errors in the sensitivity computations here.

3.2.3 | Example

Recall the linear system and QoI map discussed in Section 3.1.2. Below, for simplicity, we use $N_s = N_{s,1} = N_{s,0} = 121$ uniform samples. Figure 3 shows the surrogate models where derivative estimates for the piecewise linear surrogates are computing using Equation (5) with

$$\partial_\lambda A(\lambda) = \begin{bmatrix} e^\lambda & -\pi \sin(\pi\lambda) \\ \pi\cos(\pi\lambda) & 2e^\lambda \end{bmatrix} \tag{16}$$

and

$$\partial_\lambda \mathbf{b}(\lambda) = \begin{bmatrix} 10\pi\cos(10\pi\lambda) & 2e^\lambda \end{bmatrix}^\top. \tag{17}$$

The error-enhanced surrogates $\hat{Q}_{s,0}(\lambda)$ and $\hat{Q}_{s,1}(\lambda)$ are clearly better pointwise approximations of $Q(\lambda)$ than the surrogates $Q_{s,0}(\lambda)$ and $Q_{s,1}(\lambda)$. This underscores the importance of estimating and correcting for any underlying numerical error, $\epsilon_h(\lambda)$, since it can create significant biases that persist even if both the number of samples $N_{s,p}$ or order p is increased.

4 | STOCHASTIC INVERSE PROBLEM FORMULATION

We highlight only the main ideas behind the formulation and solution of a general stochastic inverse problem with pullback measures under minimal assumptions. For more information on the theoretical underpinnings related to the existence and uniqueness of pullback measures using the disintegration theorem, we direct the interested reader to the work of Butler et al[42] and the references therein.

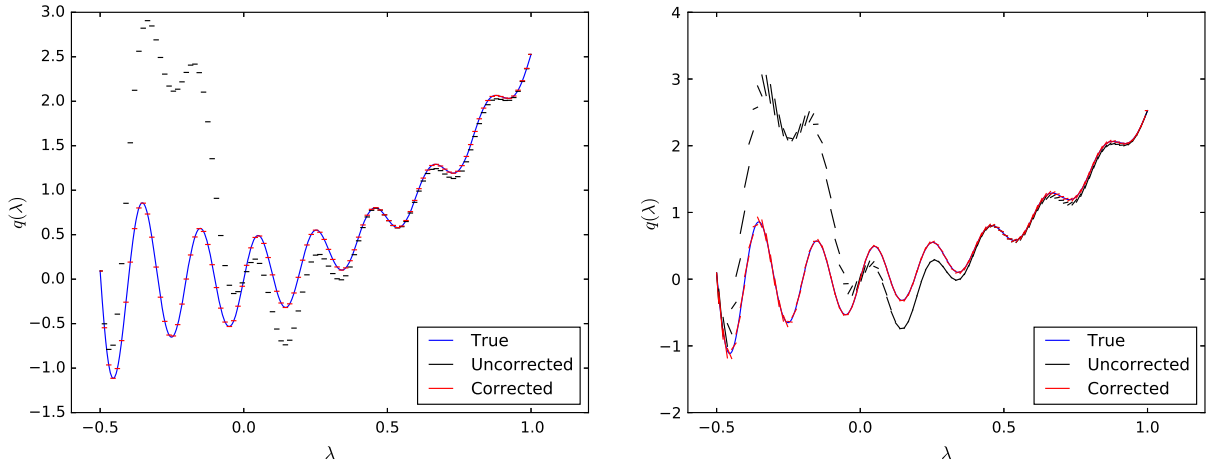

FIGURE 3 Exact quantity of interest (QoI) map ($Q(\lambda)$), surrogate QoI map ($Q_{s,p}(\lambda)$), and error-enhanced surrogate QoI map ($\hat{Q}_{s,p}(\lambda)$), labeled in the legends as True, Uncorrected, and Corrected, respectively, using $N_s = 101$. The left plot is with $p = 0$. The right plot is with $p = 1$ [Colour figure can be viewed at wileyonlinelibrary.com]

4.1 | Pullback measures

Specification of a probability measure $P_{\boldsymbol{D}}$ on $(\boldsymbol{D}, \mathcal{B}_{\boldsymbol{D}})$ (eg, modeling uncertainty in the observed data) defines a particular stochastic inverse problem of determining a *pullback* (probability) measure $P_{\boldsymbol{\Lambda}}$ on $(\boldsymbol{\Lambda}, \mathcal{B}_{\boldsymbol{\Lambda}})$ such that the push-forward measure of $P_{\boldsymbol{\Lambda}}$ matches $P_{\boldsymbol{D}}$, ie,

$$P_{\boldsymbol{\Lambda}}(Q^{-1}(E)) = P_{\boldsymbol{D}}(E), \forall E \in \mathcal{B}_{\boldsymbol{D}}. \tag{18}$$

When $P_{\boldsymbol{D}}$ is absolutely continuous with respect to $\mu_{\boldsymbol{D}}$, then, by an application of the Radon-Nikodym theorem, we can rewrite Equation (18) in terms of probability density functions and integrals. We refer to any measure $P_{\boldsymbol{\Lambda}}$ satisfying Equation (18) a solution to the stochastic inverse problem.

Note that Equation (18) implies that any solution is uniquely determined on the induced σ-algebra

$$\mathcal{C}_{\boldsymbol{\Lambda}} = \{Q^{-1}(E) \,:\, E \in \mathcal{B}_{\boldsymbol{D}}\} \subset \mathcal{B}_{\boldsymbol{\Lambda}}.$$

However, for $A \in \mathcal{B}_{\boldsymbol{\Lambda}} \backslash \mathcal{C}_{\boldsymbol{\Lambda}}$, we require more information than Equation (18) provides for computing $P_{\boldsymbol{\Lambda}}(A)$.

There are several ways that we may determine a specific and unique pullback measure. Here, we consider the use of the disintegration theorem, which requires specification of a family of conditional probabilities on the family of sets

$$\{Q^{-1}(q) \subset \boldsymbol{\Lambda}\}_{q \in \mathcal{D}}.$$

In the absence of prior information, we proceed as follows. First, we consider a disintegration of the volume measure $\mu_{\boldsymbol{\Lambda}}$ to define (lower-dimensional) volume measures on $Q^{-1}(q)$ for almost every $q \in \mathcal{D}$. Then, we use the standard Ansatz introduced in the work of Butler et al[42] to proportion probabilities/densities along events that are subsets of each $Q^{-1}(q)$ according to their relative (lower-dimensional) volume measures. The algorithm shown in Appendix A uses a discretization of this standard Ansatz, but we note that we can easily incorporate any other choice of Ansatz (eg, due to specification of a prior) as discussed in the work of Butler et al.[42]

4.2 | Identifying and reducing sources of error

For a fixed discretization of $P_{\boldsymbol{D}}$, the main sources of error in Algorithm 1 are from (1) the number of samples $N \gg N_{s,p}$ used to discretize $\boldsymbol{\Lambda}$ and (2) the evaluation of the QoI map (or its surrogate) on each of these N samples. Obviously, the use of a surrogate QoI map may allow for significantly larger N in the algorithm. However, errors in evaluation of the surrogate QoI map can result in incorrect binning of output samples, which can lead to persistent errors even in the limit of infinite N. A combination of p- and error-enhancing of the surrogate QoI map can reduce the overall errors such that we may approximate the probability of an arbitrary event in $\mathcal{B}_{\boldsymbol{\Lambda}}$ accurately as illustrated in the example below.

Algorithm 1 Numerical approximation of the inverse density

Choose a discretization partition $\{D_i\}_{1\le i\le M}$ of $\boldsymbol{\mathcal{D}}$.
for $i = 1, \dots, M$ **do**
  Compute $p_{\boldsymbol{\mathcal{D}},i} = P_{\boldsymbol{\mathcal{D}}}(D_i)$.
end for
Choose samples $\{\lambda^{(j)}\}_{j=1}^N \subset \boldsymbol{\Lambda}$ (implicitly) defining a Voronoi partition $\{\mathcal{V}_j\}_{j=1}^N$ of $\boldsymbol{\Lambda}$.
for $j = 1, \dots, N$ **do**
  Compute approximations $Q_j = Q_h(\lambda^{(j)}) \approx Q(\lambda^{(j)})$.
  Bin Q_j in the partition $\{D_i\}_{1\le i\le M}$ and let pointer vector $\mathcal{O}_j = \{i : Q_j \in D_i\}$.
  Compute approximations $V_j \approx \mu_{\boldsymbol{\Lambda}}(\mathcal{V}_j)$.
end for
for $i = 1, \dots, M$ **do**
  Compute pointer vector $C_i = \{j : Q_j \in D_i\}$.
end for
for $j = 1, \dots, N$ **do**
  Compute $p_{\boldsymbol{\Lambda},j} = \left(V_j \big/ \sum_{k\in C_{\mathcal{O}_j}} V_k\right) p_{\boldsymbol{\mathcal{D}},\mathcal{O}_j}$.
end for

4.3 | Example

Recall the linear system, QoI map, surrogate QoI map, and enhanced surrogate QoI map discussed in Sections 3.1.2 and 3.2.3. We now pose a stochastic inverse problem for this model. Suppose the probability measure $P_{\boldsymbol{\mathcal{D}}}$ on the output space $(\boldsymbol{\mathcal{D}}, \mathcal{B}_{\boldsymbol{\mathcal{D}}})$ is uniform on the interval $[-0.25, 0.25]$. Since the distribution is uniform, we choose a rather simple discretization of $P_{\boldsymbol{\mathcal{D}}}$, where the interval $[-0.25, 0.25]$ defines one cell to bin output samples. We use both $Q_{s,0}$ and $Q_{s,1}$ in Algorithm 1 with and without error-enhancement for a range of numbers N_s of uniform independent and identically distributed (i.i.d.) generating samples in $\boldsymbol{\mathcal{D}}$. The resulting probability measures are subsequently used to generate approximations of the probability of the set $A = [-1, 0] \in \mathcal{B}_{\boldsymbol{\Lambda}}$. In applying Algorithm 1, we use $N = 10^6$ uniform i.i.d. samples in $\boldsymbol{\Lambda}$ to evaluate the surrogates.

The reference value of $P(A)$ is calculated by evaluating $Q(\lambda)$ with a high-fidelity method (using direct numerical inversion) at 10^7 i.i.d. uniform samples in $\boldsymbol{\Lambda}$ and Algorithm 1. It should be noted that this value still has a small amount of numerical error due to error in Monte Carlo integration. Table 1 shows the average (over 20 batches) relative errors in calculated probabilities using piecewise-constant and linear surrogates with and without enhancement for the various regularly spaced values of N_s used to construct these surrogates. Notice that for both error-enhanced surrogates, the average error decreases rapidly, and apparently at the same rate, as the number of generating samples increases. It is also apparent that error-enhancement using the surrogate is more important in this example than p-enhancement since p-enhancement alone fails to produce converging estimates of the probability of A. The similarity in the errors using either the $p = 0$ or $p = 1$ surrogates (with or without error-enhancement, respectively) is explained in an analogous way to why the trapezoidal and midpoint rules for integration in one dimension generally give similar results. Specifically, in Figure 3, we observe that $Q_{s,0}$ and $\hat{Q}_{s,0}$ are similar to the simple function approximations obtained by evaluation of $Q_{s,1}$ and $\hat{Q}_{s,1}$, respectively, at the midpoints of each subinterval defined by the implicit Voronoi decomposition on which they

TABLE 1 Average (over 20 batches) of the relative error in probabilities of the set A calculated using $Q_{s,p}$ and error-enhanced $\hat{Q}_{s,p}$ for $p = 0$ and $p = 1$. The errors are calculated with respect to the reference value $P_{\boldsymbol{\Lambda}}(A) = 0.4256$

N_s	$Q_{s,0}$	$\hat{Q}_{s,0}$	$Q_{s,1}$	$\hat{Q}_{s,1}$
10	51.2%	34.8%	44.8%	24.2%
100	36.4%	5.45%	34.7%	6.23%
1000	37.8%	0.45%	38.1%	0.60%
10 000	37.8%	0.06%	37.6%	0.07%

are defined. This is only an approximation since the sample generating a particular Voronoi cell is not necessarily the midpoint of the subinterval it defines (although the distance between these points goes to zero as N_s increases). Subsequently, since exact integrals (and, thus, probability computations) involving $Q_{s,1}$ and $\hat{Q}_{s,1}$ can be obtained by applying either the trapezoidal or midpoint rule on each subinterval, the errors in computed probabilities using either $p = 0$ or $p = 1$ will be similar for sufficiently large N_s.

5 | NUMERICAL RESULTS

The example in the previous section involves a simple discretized linear system, and the adjoint problem, error representation, and derivative calculations are all framed in a discrete setting using the Euclidean inner product. The methods presented can be used for more interesting problems in a continuous setting, as discussed in Section 3.1.1. Such techniques are also often used when dealing with either nonlinear differential equations or situations where the numerical scheme requires stabilization (see, eg, the work of Cyr et al[43] for an excellent review of this topic). We present two examples based on PDEs where an approximation to the continuous adjoint is used. The first is a time-dependent model for contaminant transport in a groundwater aquifer, and the second is an elliptic PDE with a large number of parameters. In these problems, finite element methods are used to discretize the spaces of forward solutions and adjoint solutions. Higher-order spaces are used for adjoint problems, as discussed in Section 3.1.1. In addition, functions ψ, which are determined by the QoI (ie, the Riesz representations of components of QoI maps), are interpolated onto high-order finite spaces for the calculation of Q_h, as in Equation (8). Hence, some discretization error is present due to the representation of ψ on the mesh; however, this error should be relatively small.

In the examples below, we construct pullback measures for uniform probability distributions on the data space as a type of stress test on how the pointwise accuracy of the piecewise surrogates around a "QoI event of interest" impacts the error in computed probabilities in the parameter space. In practice, observed data probability measures are often determined by parametric estimates using available data often resulting in unimodal distributions (eg, Gaussian, chi-squared, or beta). If uncertainty is due to variability in input parameters and the response is strongly nonlinear, then multimodal distributions defined by Gaussian mixture models are commonly used. In either case, these distributions are often "peaked" over relatively small portions of the response surface so that the pointwise accuracy of the surrogate model is only required near the maximum likelihood points of the observed density. This suggests that we may only need to solve adjoints at specific samples used to construct the surrogate or that we adaptively refine the surrogate using adjoints at some subset of the samples meeting threshold criteria. This is outside the scope of this work and is left for future research.

Calculated probabilities of events are compared to reference values that are calculated using p- and error-enhanced versions of the surrogate model presented in this paper using finite element spaces, time discretizations, and/or Voronoi tessellations that are much more refined than the ones that are being analyzed. The computational method for constructing pullback measures (see Appendix A) requires a surrogate model, and as shown in Section 3, this class of surrogates using highly refined spaces is extremely accurate. The relative errors of the computed probability of the QoI event of interest using enhanced and unenhanced surrogates are calculated with respect to the reference value and are analyzed.

5.1 | Contaminant transport problem

This example is a time-dependent PDE modeling groundwater contaminant transport. It has the form of an advection-dispersion problem with an initial approximate point source. Similar systems have been used for performing parameter estimation based on field data with the stochastic inversion framework discussed in Section 4.[44]

5.1.1 | The forward problem, adjoint, and QoI

The forward problem is

$$\begin{cases} \theta \frac{\partial u}{\partial t} + \mathbf{v} \cdot \nabla u - \nabla \cdot (\theta D \nabla u) &= 0, \quad t \in (0, T], (x, y) \in \Omega \\ \theta D \nabla u \cdot \mathbf{n} &= 0, \quad t \in (0, T], (x, y) \in \partial\Omega \\ u &= u_0, \quad t = 0, (x, y) \in \Omega, \end{cases} \tag{19}$$

where the unknown u is the concentration of a contaminant, θ is porosity, $\mathbf{v} = [v_x, v_y]^T$ is the Darcy velocity, and D is the dispersion matrix. D is a symmetric positive-definite matrix defined by

$$D_{xx} = \alpha_L \frac{v_x^2}{|\mathbf{v}|} + \alpha_T \frac{v_y^2}{|\mathbf{v}|} + D^*, \tag{20}$$

$$D_{yy} = \alpha_L \frac{v_y^2}{|\mathbf{v}|} + \alpha_T \frac{v_x^2}{|\mathbf{v}|} + D^*, \tag{21}$$

and $D_{xy} = D_{yx} = 0$, where α_L is the lateral dispersion, α_T is the transverse dispersion, and D^* is the underlying molecular diffusion. The domain is $\Omega = [497000.0, 502000.0] \times [537000.0, 541000.0]$. The initial condition u_0 is defined as an approximate point source at $(x_s, y_s) = (498250.0, 538000.0)$ by

$$u_0 = M \frac{400}{\pi} \exp(-400(x - x_s)^2 - 400(y - y_s)^2), \tag{22}$$

where M is the source magnitude. Suppose that the components of the Darcy velocity v_x and v_y are uncertain and all of the other model parameters are known. Let the space of unknown values of $[v_x, v_y]$ be $\mathbf{\Lambda} = [10, 50]^2$. Suppose we have geophysically reasonable parameters $\theta = 0.25$, $\alpha_L = 70$, $\alpha_T = 7$, $D^* = 0.01$, and $M = 1000$. The QoI is an approximate solution at some point (x_r, y_r) at time $T = 10$ defined by $\langle u, \psi \rangle$, with $\psi(x, y, t) = \psi_\Omega(x, y)\delta_T(t)$, where

$$\psi_\Omega = \frac{400}{\pi} \exp(-400(x - x_r)^2 - 400(y - y_r)^2),$$

δ_T is the Dirac delta in time centered at $t = T$, and $\langle , \rangle$ is a space-time inner product defined below. Let the measurement location be $(x_r, y_r) = (499563.69, 538995.82)$.

We use the open-source software FEniCS[45,46] to solve the system using the space-time continuous Galerkin method with linear elements in space and time on a 50×50 spatial mesh with a time step of 0.1.

The duality pairing between the adjoint and forward problem solutions is given by $\langle , \rangle : X \times Y \to \mathbb{R}$, where $X = L^2(\Omega, [0, T]) \cup L^2(\Omega, 0)$ and $Y = L^2(\Omega, [0, T]) \cup L^2(\Omega, T)$ are spaces for the forward and adjoint solutions, respectively. The derivation of the adjoint operator $\mathcal{L}^*$ follows:

$$\begin{aligned}
< \mathcal{L}u, \phi > &= \int_0^T \int_\Omega \left(\theta \frac{du}{dt}\phi + (\mathbf{v} \cdot \nabla u)\phi - (\nabla \cdot (\theta D \nabla u))\phi \right) d\Omega dt + \int_\Omega \theta u(0)\phi(0) d\Omega \\
&= \int_0^T \int_\Omega \left(-\theta u \frac{d\phi}{dt} \right) d\Omega dt + \left[\int_\Omega \theta u \phi d\Omega \right]_0^T + \int_0^T \int_\Omega (-u(\mathbf{v} \cdot \nabla \phi))\, d\Omega dt + \int_0^T \int_{\partial\Omega} (u(\phi \mathbf{v} \cdot \mathbf{n}))\, ds dt \\
&= \int_0^T \int_\Omega \left(-\theta u \frac{d\phi}{dt} \right) d\Omega dt + \int_0^T \int_\Omega (-u(\mathbf{v} \cdot \nabla \phi))\, d\Omega dt + \int_0^T \int_\Omega (-u \nabla \cdot (\theta D \nabla \phi))\, d\Omega dt \\
&\quad + \theta \int_\Omega u(T)\psi d\Omega + \int_0^T \int_{\partial\Omega} (u(\phi \mathbf{v} \cdot \mathbf{n}))\, ds dt + \int_0^T \int_{\partial\Omega} (-\theta D \nabla u \cdot \mathbf{n})\phi ds dt + \int_0^T \int_{\partial\Omega} u(\theta D \nabla \phi \cdot \mathbf{n}) ds dt \\
&= \int_0^T \int_\Omega \left(-\theta u \frac{d\phi}{dt} - u(\mathbf{v} \cdot \nabla \phi) - u \nabla \cdot (\theta D \nabla \phi) \right) d\Omega dt + \int_\Omega \theta u(T)\phi(T) d\Omega = < u, \mathcal{L}^* \phi > .
\end{aligned}$$

Because of the Dirac delta in the definition of ψ, the adjoint problem is specified as the following initial boundary value problem[47]:

$$\begin{cases}
-\theta \frac{\partial \phi}{\partial t} - \mathbf{v} \cdot \nabla \phi - \nabla \cdot (\theta D \nabla \phi) = 0, & t \in [0, T), (x, y) \in \Omega \\
(\phi \mathbf{v} + \theta D \nabla \phi) \cdot \mathbf{n} = 0, & t \in [0, T), (x, y) \in \partial\Omega \\
\phi = \psi_\Omega, & t = T, (x, y) \in \Omega.
\end{cases} \tag{23}$$

The adjoint problem (23) is solved backward in time on the same 50×50 mesh as the forward problem, but with quadratic elements in space (required for calculating error estimates) while keeping linear elements in time with the same time step of 0.1. The QoI map is evaluated approximately by projecting ψ_Ω onto the finite element space used for the adjoint problem and calculating the $L^2(\Omega)$ inner product with $u_h(T)$. Error estimates are computed as described in Equation (4),

TABLE 2 Average (over 20 batches) of the relative error in probabilities of the set A calculated using $Q_{s,p}$ and error-enhanced $\hat{Q}_{s,p}$ for $p = 0$ and $p = 1$. The errors are calculated with respect to the reference value $P_{\Lambda}(A) = 0.2220$

N_s	$Q_{s,0}$	$\hat{Q}_{s,0}$	$Q_{s,1}$	$\hat{Q}_{s,1}$
50	16.58%	17.43%	20.18%	22.12%
500	3.52%	3.73%	3.21%	3.61%
5000	3.43%	0.43%	3.84%	0.50%

and derivatives are computed as described in Equation (5). For the forward and adjoint problems, the discretized linear systems for both are solved directly with the adjoint computed backward in time. We then have

$$e_{Q,h} = \int_0^T \int_\Omega \left(-\theta \frac{\partial u_h}{\partial t} \phi_h - (\mathbf{v} \cdot \nabla u_h)\phi_h - (\theta D \nabla u_h \cdot \nabla \phi_h) \right) d\Omega dt + \int_\Omega \theta(u_0 - u_h(0))\phi_h(0) d\Omega.$$

5.1.2 | The stochastic inverse problem and surrogate setup

Suppose the probability measure $P_{\mathcal{D}}$ on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ is uniform on the interval $[0.005, 0.021]$, which we discretize into three uniform intervals to bin QoI samples and produce more variation in the pullback measure. We numerically solve the stochastic inverse problem using $Q_{s,p}$ and $\hat{Q}_{s,p}$ for a range of numbers N_s of uniform i.i.d. generating samples in $\mathbf{\Lambda}$. We then approximate the probability measure of a set $A := [25, 35]^2 \subset \mathbf{\Lambda}$, using the computed approximation of P_{Λ} by Algorithm 1 with $N = 10^7$ uniform i.i.d. samples.

5.1.3 | Numerical computations and discussion

Higher-fidelity solutions to the forward problem are calculated by solving the model on a 100×100 mesh with time steps of 0.05 with the same type of time and finite element discretizations. Corresponding adjoint problems (using quadratic finite elements) are also solved and used to calculate both error estimates and derivatives to generate a highly accurate piecewise linear surrogate. A reference value of $P_{\Lambda}(A) = 0.2220$ is subsequently calculated using an enhanced piecewise linear surrogate generated from these high-fidelity results.

Table 2 shows the average (over 20 batches) of the relative error calculated using piecewise-constant and linear surrogates with and without enhancement. The model was run for 10^4 uniformly sampled points in $\mathbf{\Lambda}$, and the values in each batch used to compute values in Table 2 were randomly sampled from these 10^4 samples without replacement. As with the other cases, the errors from using surrogates without error-enhancement level off at a certain point because of the persistent deterministic error polluting all values. Both enhanced surrogates (ie, $\hat{Q}_{s,0}$ and $\hat{Q}_{s,1}$) produce estimates whose error is reduced relatively quickly as seen in the simple 1D example. Moreover, the similarities between the errors when using the piecewise-constant or linear surrogates (with or without enhancement, respectively) suggest that the response surface is approximately linear so that the piecewise-constant surrogate allows for local cancellation of errors in the solution in a way similar to how the midpoint rule gives exact integrals for linear functions. However, we do not expect this to hold in higher dimensions, as both the previous and next examples demonstrate.

5.2 | A 100-dimensional example

This example involves solving a stochastic inverse problem for a higher-dimensional ($n = 100$) parameter space. The model is an elliptic PDE defined in a continuous setting. This example also demonstrates a common feature involving QoI computed from high-dimensional spaces, which we briefly mention here since, at first glance, it may seem that performing stochastic inversion into such a high-dimensional space of parameters is not possible with the types of surrogates considered so far, which often suffer from the curse of dimensionality. However, in many cases, there exist lower-dimension manifolds in the parameter space on which the QoI is predominately sensitive, which are commonly referred to in the literature as active subspaces.[19] While this literature provides a framework for identifying and exploiting this low-dimensional structure explicitly, we instead opt to perform the inversion in this example on the full 100-dimensional parameter space to demonstrate how the p-enhancement effectively improves results in the directions of the active subspace without having to explicitly construct the subspace. This example also demonstrates the benefits of using adjoints for estimating errors

and gradients simultaneously since only one adjoint calculation is required for the QoI for each sample. In other words, the estimation of error and derivatives by adjoints is a computation that is (almost) independent of the dimension of the parameter space. The actual computational cost of calculating error estimates and gradients of QoI only depends on the number of parameters in that it requires an inner-product calculation for each parameter (see Equations (4) and (5)). The computational cost of performing an inner product is usually extremely small compared to performing model solves. Hence, both *p*- and error-enhancement are not considerably more computationally expensive for higher-dimensional parameter spaces compared to lower-dimensional ones.

5.2.1 | The forward problem, adjoint, and the QoI

We consider the elliptic boundary value problem on the unit square

$$\begin{cases} -\nabla \cdot (K(x,y)\nabla u(x,y)) = 0,\ (x,y) \in (0,1)\times(0,1) \\ u(0,y) = 0,\ y \in (0,1) \\ u(1,y) = 1,\ y \in (0,1) \\ K(x,0)\nabla u(x,0)\cdot \mathbf{n} = K(x,1)\nabla u(x,1)\cdot \mathbf{n} = 0,\ x \in (0,1), \end{cases} \tag{24}$$

where $K(x,y)$ is a conductivity field that we treat as a random function. K belongs to an infinite-dimensional space, but truncating a Karhunen-Loève (K-L) expansion is a classical option for deriving finite-dimensional parameterizations for $\log(K)$. We construct the K-L expansion of $Y(x,y)$, where $Y(x,y) := \log[K(x,y)]$. Let $\bar{Y}(x,y)$ be the mean value of $Y(x)$, and suppose that it has an exponential covariance

$$C_Y(x,y) = \sigma_Y^2 \exp(-|x_1 - y_1|/\eta_1 - |x_2 - y_2|/\eta_2), \tag{25}$$

where the variance is $\sigma_Y = 1.0$ and the correlation lengths are $\eta_1 = 0.08$ and $\eta_2 = 0.07$. Hence, $Y(x,y)$ can be written as

$$Y(x,y) = \bar{Y}(x,y) + \sum_{n=0}^{\infty} \xi_n \sqrt{\lambda_n} f_n(x,y), \tag{26}$$

where λ_n and $f_n(x,y)$ are eigenpairs determined by C_Y, and ξ_n are standard normal random variables. In other words, we shift notation slightly in this example to align more closely with the K-L literature so that λ_n is now an eigenvalue and ξ_n is now used as the variable for an uncertain parameter. Truncating the series in Equation (26) at the Nth term gives the finite-dimensional approximation

$$Y(x,y) \approx \bar{Y}(x,y) + \sum_{n=0}^{N} \xi_n \sqrt{\lambda_n} f_n(x,y). \tag{27}$$

For the given unit square domain and covariance, we can calculate the eigenpairs analytically.[48] We use the first 100 K-L terms (ie, $N = 100$) because the eigenvalues above this are observed to be negligible for this correlation length and take $\bar{Y}(x,y) = 1$. Suppose we have a scalar QoI, Q, representing an approximation to the solution of Equation (24) at $(0.5, 0.5)$, using the inner product $\langle u, \psi \rangle_{L^2}$ with a steep Gaussian, ie,

$$\psi = \frac{400}{\pi} \exp(-400(x-0.5)^2 - 400(y-0.5)^2).$$

The forward model is solved on a 20×20 finite element mesh with linear finite elements, and the adjoint problem is solved on the same mesh with quadratic finite elements. Using 1000 samples from a 100-dimensional standard normal distribution (representing prior knowledge of the K-L expansion), solving the forward and adjoint problems, and calculating gradients, the corresponding active subspace is calculated. Figure 4 shows the first 20 eigenvalues of the sensitivity matrix with bootstrap intervals. The small bootstrap intervals indicate that this structure of the eigenvalues is accurate. Notice the large gap between the first and second eigenvalues, which indicates the presence of a one-dimensional active subspace for the QoI.

5.2.2 | The stochastic inverse problem and surrogate setup

Suppose the probability measure P_D is uniform on the interval $[0.3, 0.4]$. As mentioned above, we solve the stochastic inverse problem on the 100-dimensional space of coefficients $\{\xi_i\}_{i=1}^{100}$ instead of the active subspace. Furthermore, to ensure that we built the surrogate over "low-probability" conductivity fields, we chose to uniformly sample from a truncation of

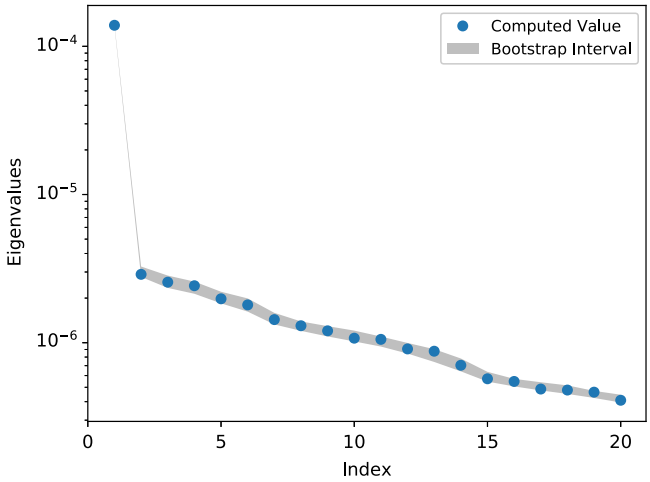

FIGURE 4 First 20 eigenvalues of the sensitivity matrix of the high-dimensional example with bootstrap intervals [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Average (over 10 batches) of the relative error in probabilities of the set A calculated using $Q_{s,p}$ and error-enhanced $\hat{Q}_{s,p}$ for $p = 0$ and $p = 1$. The errors are calculated with respect to the reference value $P_{\Lambda}(A) = 0.608$

N_s	$Q_{s,0}$	$\hat{Q}_{s,0}$	$Q_{s,1}$	$\hat{Q}_{s,1}$
25	100%	100%	25.0%	25.82%
250	100%	100%	6.76%	6.6%
2500	100%	75.99%	2.94%	2.06%
25 000	100%	38.82%	1.63%	1.24%

the 100-dimensional Gaussian cube so that conductivity fields sufficiently different from the mean field would be used to define the QoI map. Specifically, we truncated the Gaussian cube in each dimension by three standard deviations away from the mean to define a 100-dimensional hypercube $\mathbf{\Lambda} = [-3, 3]^{100}$. The stochastic inverse problem is then solved using $Q_{s,p}$ and $\hat{Q}_{s,p}$ for various choices of N_s uniform i.i.d. generating samples in $\mathbf{\Lambda}$. We then approximate the probability measure of a cylinder set $A = [-3, 0.2] \times [-3, 3]^{99} \subset \mathbf{\Lambda}$ using the computed approximation of P_{Λ} by Algorithm 1 with $N = 10^5$ uniform i.i.d. samples.

5.2.3 | Numerical computations and discussion

To compute a reference probability of $P(A) = 0.608$, we implemented Algorithm 1 using a p- and error-enhanced surrogate generated using the forward model and adjoint solutions (to correct for the deterministic error) on the full set of $N = 10^5$ uniform i.i.d. samples. Table 3 shows the average relative error in estimating $P(A)$ over 10 batches for the surrogates constructed with 25, 250, 2500, and 25 000 samples, where the samples in each batch were randomly sampled from the 10^5 samples without replacement. We see that when using $Q_{s,0}$, there is 100% error, and by error-enhancement to $\hat{Q}_{s,0}$, the error is only slightly reduced by using larger N_s. However, with both $Q_{s,1}$ and $\hat{Q}_{s,1}$, the error decreases quickly as more samples are used. We believe the stark difference in these results is attributed to the presence of the active subspace. Thus, this approach to constructing low-order error-enhanced surrogates on unstructured methods appears to apply to high-dimensional problems in which active subspaces are present.

6 | CONCLUSIONS AND FUTURE WORK

We described an approach to constructing and enhancing piecewise-defined low-order surrogates on unstructured sets of samples using adjoints. The two types of enhancement considered here are p-enhancement (ie, improving the local order of the surrogate) and error-enhancement. The utility of these surrogates for solving stochastic inverse problems

is studied in the context of constructing pullback measures and estimating particular events in the parameter space. A simple finite-dimensional linear example is used throughout to illustrate the various concepts presented. Two PDE examples, commonly studied in the literature, demonstrated various aspects of the ideas. Specifically, an elliptic PDE example demonstrated that p-enhancement is required to achieve a steady reduction of errors in higher dimensions. The contaminant transport example demonstrated the benefits of h-enhancement for a more computationally complex problem where the response surface is likely more linear around the event of interest in the parameter domain, which allows for cancellation of errors when using the piecewise-constant surrogates.

There are many interesting future directions to take this work. Perhaps the most obvious direction is to examine piecewise Voronoi surrogate approximations with higher-order polynomial approximations suggested by the work of Rushdi et al.[31] Another direction would be to study piecewise-defined surrogates based on global surrogate approaches as discussed below. While there is much work on the use of polynomial chaos expansions to construct global surrogates,[1,10,11] in order to ensure the error $\epsilon_s(\lambda)$ is bounded by some threshold so that results are useful, a high-order polynomial approximation may be required, which subsequently implies that the value of N_s may be prohibitively large and beyond the limits provided by the computational budget. For example, in polynomial chaos approaches, the total number of coefficients we need to compute a general p-order surrogate over an m-dimensional parameter space is given by $(m + p)!/(m!p!)$.[10,28] Computation of the polynomial coefficients, in turn, requires computation of integrals over the parameter space, which is often done with numerical quadrature. If a QoI is particularly sensitive to λ, then a large order p of the global polynomial chaos surrogate approximation may be required to control $\epsilon_s(\lambda)$. This, in turn, requires higher-order numerical quadrature, which increases the total number N of numerical solutions to the model. In the work of Marzouk et al,[13] for a Bayesian inference problem of parameters to a contaminant transport model with relatively few parameters, it is suggested that a piecewise polynomial chaos surrogate be constructed on a regular partitioning of $\mathbf{\Lambda}$ to improve accuracy. The general idea is that $\epsilon_s(\lambda)$ can be made small if the surrogate is defined by localized low-order surrogate approximations to the QoI response surface and that this is computationally feasible since it is often easier to solve several lower-order problems than a significantly higher-order problem. Accuracy in the global posterior of the Bayesian solution from such a surrogate is subsequently observed. This provides a motivation for future work in the design of low-order polynomial surrogates constructed by other means than those shown in this work for goal-oriented measure-theoretic inversion.

ACKNOWLEDGEMENTS

This material is based upon work supported by the United States Department of Energy Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under award numbers DE-SC0009279 and DE-SC0009286 as part of the DiaMonD Multifaceted Mathematics Integrated Capability Center. The work of T. Butler was supported in part by the National Science Foundation (DMS-1228206 and DMS-1818941). The work of S. Mattis was supported in part by the German Research Foundation (DFG, Project WO 671/11-1).

ORCID


Steven A. Mattis https://orcid.org/0000-0002-3464-335X

REFERENCES

1. Ghanem R, Red-Horse J. Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach. *Phys D*. 1999;133(1-4):137-144. https://doi.org/10.1016/S0167-2789(99)00102-5
2. Le Maître OP, Najm HN, Ghanem RG, Knio OM. Multi-resolution analysis of Wiener-type uncertainty propagation schemes. *J Comput Phys*. 2004;197(2):502-531. https://doi.org/10.1016/j.jcp.2003.12.020
3. Robert CP, Casella G. *Monte Carlo Statistical Methods*. New York, NY: Springer; 2004.
4. Gentle JE. *Random Number Generation and Monte Carlo Methods*. New York, NY: Springer; 2003.
5. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. Boca Raton, FL: Chapman & Hall/CRC; 1996.
6. Geyer CJ. Practical Markov chain Monte Carlo. *Statistical Science*. 1992;7(4):473-483.
7. Ghanem R, Spanos P. *Stochastic Finite Elements: A Spectral Approach*. New York, NY: Springer; 1991.
8. Wiener N. The homogeneous chaos. *Am J Math*. 1938;60(4):897-936. http://www.jstor.org/stable/2371268
9. Cameron RH, Martin WT. The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. *Ann Math*. 1947;48(2):385-392. http://www.jstor.org/stable/1969178

10. Xiu D, Karniadakis GE. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J Sci Comput*. 2002;24(2):619-644. https://doi.org/10.1137/S1064827501387826
11. Wan X, Karniadakis GE. Beyond Wiener–Askey expansions: handling arbitrary pdfs. *J Sci Comput*. 2006;27(1-3):455-464. https://doi.org/10.1007/s10915-005-9038-8
12. Le Maître OP, Knio OM, Najm HN, Ghanem RG. Uncertainty propagation using Wiener-Haar expansions. *J Comput Phys*. 2004;197(1):28-57. https://doi.org/10.1016/j.jcp.2003.11.033
13. Marzouk YM, Najm HN, Rahn LA. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J Comput Phys*. 2007;224(2):560-586. https://doi.org/10.1016/j.jcp.2006.10.010
14. Butler T, Constantine P, Wildey T. A posteriori error analysis of parameterized linear systems using spectral methods. *SIAM J Matrix Anal Appl*. 2012;33:195-209.
15. Prudhomme S, Bryant CM. Adaptive surrogate modeling for response surface approximations with application to Bayesian inference. *Adv Model Simul Eng Sci*. 2015;2(1):1-21. https://doi.org/10.1186/s40323-015-0045-5
16. Bryant CM, Prudhomme S, Wildey T. Error decomposition and adaptivity for response surface approximations from PDEs with parametric uncertainty. *SIAM/ASA J Uncertain Quantif*. 2015;3(1):1020-1045. https://doi.org/10.1137/140962632
17. Almeida RC, Oden JT. Solution verification, goal-oriented adaptive methods for stochastic advection-diffusion problems. *Comput Methods Appl Mech Eng*. 2010;199:2472-2486.
18. Nobile F, Tempone R, Webster CG. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J Numer Anal*. 2008;46(5):2309-2345. https://doi.org/10.1137/060663660
19. Constantine PG, Dow E, Wang Q. Active subspace methods in theory and practice: applications to kriging surfaces. *SIAM J Sci Comput*. 2014;36(4):A1500-A1524. https://doi.org/10.1137/130916138
20. Butler T, Dawson C, Wildey T. Propagation of uncertainties using improved surrogate models. *SIAM/ASA J Uncertain Quantif*. 2013;1:164-191.
21. Estep D, Neckels D. Fast and reliable methods for determining the evolution of uncertain parameters in differential equations. *J Comput Phys*. 2006;213(2):530-556. http://www.sciencedirect.com/science/article/pii/S0021999105003918
22. Babuska I, Rheinboldt WC. *A-posteriori* error estimates for the finite element method. *Int J Numer Methods Eng*. 1978;12(10):1597-1615. https://doi.org/10.1002/nme.1620121010
23. Estep D. A posteriori error bounds and global error control for approximation of ordinary differential equations. *SIAM J Numer Anal*. 1995;32(1):1-48.
24. Ainsworth M, Oden JT. A posteriori error estimation in finite element analysis. *Comput Methods Appl Mech Eng*. 1997;142(1):1-88. http://www.sciencedirect.com/science/article/pii/S0045782596011073
25. Estep D, Tavener S, Wildey T. A posteriori analysis and improved accuracy for an operator decomposition solution of a conjugate heat transfer problem. *SIAM J Numer Anal*. 2008;46(4):2068-2089. https://doi.org/10.1137/060678737
26. Carey V, Estep D, Tavener S. A posteriori analysis and adaptive error control for multiscale operator decomposition solution of elliptic systems I: triangular systems. *SIAM J Numer Anal*. 2009;47(1):740-761. https://doi.org/10.1137/070689917
27. Wildey T, Tavener S, Estep D. *A posteriori* error estimation of approximate boundary fluxes. *Commun Numer Methods Eng*. 2008;24(6):421-434. https://doi.org/10.1002/cnm.1014
28. Butler T, Dawson C, Wildey T. A posteriori error analysis of stochastic differential equations using polynomial chaos expansions. *SIAM J Sci Comput*. 2011;33:1267-1291.
29. Chaudhry JH, Estep D, Gunzburger M. Exploration of efficient reduced-order modeling and a posteriori error estimation. *Int J Numer Methods Eng*. 2017;111(2):103-122. https://doi.org/10.1002/nme.5453
30. Carlberg K. Adaptive *h*-refinement for reduced-order models. *Int J Numer Methods Eng*. 2015;102(5):1192-1210. https://doi.org/10.1002/nme.4800
31. Rushdi AA, Swiler LP, Phipps ET, D'Elia M, Ebeida MS. VPS: Voronoi piecewise surrogate models for high-dimensional data fitting. *Int J Uncertain Quantif*. 2017;7(1):1-21.
32. Ebeida MS, Mitchell SA, Swiler LP, Romero VJ, Rushdi AA. POF-Darts: geometric adaptive sampling for probability of failure. *Reliab Eng Syst Saf*. 2016;155:64-77.
33. Ebeida MS, Rushdi AA. Recursive Spoke Darts: local hyperplane sampling for Delaunay and Voronoi meshing in arbitrary dimensions. *Procedia Engineering*. 2016;163(suppl C):110-122. http://www.sciencedirect.com/science/article/pii/S1877705816333380
34. Butler T, Graham L, Mattis S, Walsh S. A measure-theoretic interpretation of sample based numerical integration with applications to inverse and prediction problems under uncertainty. *SIAM J Sci Comput*. 2017;39(5):A2072-A2098.
35. Marchuk GI. *Adjoint Equations and Analysis of Complex Systems*. Dordrecht, The Netherlands: Kluwer; 1995.
36. Marchuk GI, Agoshkov VI, Shutyaev VP. *Adjoint Equations and Perturbation Algorithms in Nonlinear Problems*. Boca Raton, FL: CRC Press; 1996.
37. Lanczos C. *Linear Differential Operators*. Mineola, NY: Dover Publications; 1997.
38. Cacuci D. *Sensitivity and Uncertainty Analysis: Theory*. Vol. I. Boca Raton, FL: Chapman & Hall/CRC; 1997.
39. Bartlett RA, Gay DM, Phipps ET. Automatic differentiation of C++ codes for large-scale scientific computing. In: *Computational Science – ICCS 2006*. Berlin, Germany: Springer; 2006:525-532. https://doi.org/10.1007/11758549_73
40. Bangerth W, Rannacher R. *Adaptive Finite Element Methods for Differential Equations*. Basel, Switzerland: Birkhäuser Verlag; 2003.
41. Becker R, Rannacher R. An optimal control approach to *a posteriori* error estimation in finite element methods. *Acta Numer*. 2001;10:1-102.

42. Butler T, Estep D, Tavener S, Dawson C, Westerink J. A measure-theoretic computational method for inverse sensitivity problems III: multiple quantities of interest. *SIAM/ASA J Uncertain Quantif*. 2014;2(1):174-202. https://doi.org/10.1137/130930406
43. Cyr E, Shadid J, Wildey T. Approaches for adjoint-based a posteriori analysis of stabilized finite element methods. *SIAM J Sci Comput*. 2014;36(2):A766-A791. https://doi.org/10.1137/120895822
44. Mattis SA, Butler TD, Dawson CN, Estep D, Vessilinov VV. Parameter estimation and prediction for groundwater contamination based on measure theory. *Water Resour Res*. 2015;51:7608-7629.
45. Logg A, Mardal K-A, Wells G, eds. *Automated Solution of Differential Equations by the Finite Element Method*. Berlin, Germany: Springer; 2012.
46. Alnæs M, Blechta J, Hake J, et al. The FEniCS Project Version 1.5. *Arch Numer Softw*. 2015;3(100). http://journals.ub.uni-heidelberg.de/index.php/ans/article/view/20553
47. Butler T, Estep D, Sandelin J. A computational measure theoretic approach to inverse sensitivity problems II: a posteriori error analysis. *SIAM J Numer Anal*. 2012;50(1):22-45.
48. Zhang D, Lu Z. An efficient, high-order perturbation approach for flow in random porous media via Karhunen–Loeve and polynomial expansions. *J Comput Phys*. 2004;194(2):773-794.
49. Hu N, Fish J, McAuliffe C. An adaptive stochastic inverse solver for multiscale characterization of composite materials. *Int J Numer Methods Eng*. 2017;109(12):1679-1700. https://doi.org/10.1002/nme.5341
50. Graham L, Mattis S, Walsh S, Butler T, Pilosov M, McDougall D. BET: Butler, Estep, Tavener Method v2.0.0. 2016. https://doi.org/10.5281/zenodo.59964

How to cite this article: Mattis SA, Butler T. Enhancing piecewise-defined surrogate response surfaces with adjoints on sets of unstructured samples to solve stochastic inverse problems. *Int J Numer Methods Eng*. 2019;1–18. https://doi.org/10.1002/nme.6078

APPENDIX A

COMPUTATIONAL ALGORITHM FOR CONSTRUCTING PULLBACK MEASURES

Algorithm 1 describes a nonintrusive sample-based approach to approximating pullback measures, which proceeds in four stages defined by four separate non-nested for-loops and uses the standard Ansatz. There is a more recently developed algorithm for solving this type of inverse problem applied to multiscale characterization of composite materials,[49] but we focus on the plain version of the algorithm to focus our attention on the accuracy of the surrogate. We note that Algorithm 1 applies to any discretizing set of samples in $\mathbf{\Lambda}$ no matter how the samples are generated. The authors are codevelopers of BET,[50] an open-source Python package for stochastic measure-theoretic problems, which includes implementations of Algorithm 1.

In the first for-loop, we use a partition $\{D_i\}_{1\le i\le M}$ of $\boldsymbol{\mathcal{D}}$ to construct an M-dimensional vector of probabilities with the ith component given by $p_{\boldsymbol{\mathcal{D}},i}$ corresponding to the probability of $P_{\boldsymbol{\mathcal{D}}}(D_i)$. In the second for-loop, we use the QoI map to build a pointer vector between a set of N samples $\{\lambda^{(j)}\}_{j=1}^N$ *implicitly* defining a partition of $\mathbf{\Lambda}$ into Voronoi cells and the partition $\{D_i\}_{1\le i\le M}$ of $\boldsymbol{\mathcal{D}}$. We also compute approximations of the volumes of the implicitly defined Voronoi cells (eg, using Monte Carlo approximations). Then, in the third for-loop, we again use the QoI map to construct a pointer vector between each D_i and the set of all samples (and, thus, Voronoi cells) approximating $Q^{-1}(D_i)$. Finally, in the fourth for-loop, we use the standard Ansatz and the two pointer vectors to distribute the probabilities $p_{\boldsymbol{\mathcal{D}},i}$, for each i, across all Voronoi cells that approximate $Q^{-1}(D_i)$. The output of the algorithm is an N-dimensional vector of probabilities with the jth component given by $p_{\mathbf{\Lambda},j}$ corresponding to the jth sample $\lambda^{(j)} \in \mathbf{\Lambda}$ used to approximate the pullback measure for all $A \in \mathbf{\Lambda}$ using either a counting measure,

$$P_{\mathbf{\Lambda}}(A) \approx \sum_{\lambda^{(j)}\in A} p_{\mathbf{\Lambda},j}, \tag{A1}$$

or the approximate volumes of the Voronoi cells,

$$P_{\mathbf{\Lambda}}(A) \approx \sum_{1\le j\le N} p_{\mathbf{\Lambda},j}\,\mu_{\mathbf{\Lambda}}(\mathcal{V}_j \cap A). \tag{A2}$$