

**SAND20XX-XXXXR**

**LDRD PROJECT NUMBER:** 188256

**LDRD PROJECT TITLE:** Adaptive Bayesian Inference for Prediction

**PROJECT TEAM MEMBERS:** Timothy M. Wildey, John D. Jakeman

**ABSTRACT:**

Performing rigorous predictive simulation necessitates that all available data be incorporated into the modeling process in order to accurately quantify uncertainties in predictions of the model. Statistical Bayesian inference is the most commonly used approach for incorporating data into uncertainty estimates. Statistical Bayesian inference uses data and an assumed error model to inform posterior distributions of model input variables and model discrepancies. An explicit characterization of the posterior distribution is not formed, but rather methods such as Markov chain Monte-Carlo (MCMC) are used to draw samples from the posterior. These samples are then used to evaluate statistics, such as integrals, of predicted quantities of interest (QoI). Recently the authors have developed an alternative reformulation of Bayesian inference based upon a measure-theoretic inversion technique. Unlike traditional statistical Bayesian inference, this reformulation allows us to directly interrogate the posterior distribution and, consequently, to generate samples from the posterior without MCMC. In this work we use forward propagation uncertainty quantification techniques to reduce the cost of approximating the posterior distribution. Specifically we investigate goal-oriented adaptive sparse grid approximation methods. These adaptive methods concentrate high-fidelity model evaluations in regions of the parameter space that both significantly contribute to the uncertainty in specified QoI and are informed by the available data. Numerical results are presented to compare our approach with the standard statistical Bayesian approach and to demonstrate the effectiveness of the goal-oriented adaptivity.

**INTRODUCTION:**

Inverse problems are ubiquitous in computational science and engineering. Often, parameters of interest cannot be measured directly and must be inferred from observable data. The mapping between these parameters and the measurable data is often referred to as the forward model and the goal is to use this forward model to gain knowledge about the parameters, i.e., to invert the forward map. Deterministic inverse problems are often posed as: given deterministic data, determine the set of parameters that, when propagated through a computational model, will reproduce the data as closely as possible. Such problems are often solved using optimization techniques and may involve regularization to guarantee a unique solution. Stochastic inverse problems, on the other hand, are often posed as: given a distribution on the data, determine the distribution of the parameters that, when propagated through a computational model, will reproduce the distribution on the data. While optimization-based approaches do exist for the stochastic inverse problem, the pervasive approach in the literature is to perform statistical Bayesian inference to compute a posterior distribution that is conditional on the data distribution and a user-defined prior distribution on the input parameters. This prior distribution is meant to encompass any additional information (prior knowledge) regarding the distribution on the parameters.

Most methods for solving the Bayesian inverse problem require an assumed statistical error model (usually additive noise) to transform a deterministic forward model into a stochastic model. In addition, some form of Markov Chain Monte Carlo (MCMC) sampling is usually required to generate samples from the posterior [Brooks 2011]. While tremendous advances have been in recent years to develop scalable and efficient MCMC sampling algorithms, this often represents the computational bottleneck of the statistical Bayesian approach. In many practical applications, MCMC requires an infeasible number of model evaluations and in certain cases can fail to converge to the true posterior distribution.

An alternative perspective, based on measure theory, has emerged in recent years as a different approach for stochastic inversion [Breidt 2012, Butler 2012]. The measure-theoretic approach aims to construct a distribution on the parameters by directly inverting the forward model. This approach requires only a volume measure on the parameter space rather than a prior distribution on the input parameters. The main drawback to this approach is the requirement that the input and output spaces be decomposed into measurable subsets, such as hyper-cubes, which becomes impractical in high-dimensional spaces.

The goal of this paper is to bridge the gap between the Bayesian and measure-theoretic approaches by utilizing Bayesian concepts within the measure-theoretic framework. The resulting approach provides a means to approximate the posterior distribution in a manner that is consistent with the measure-theoretic approach, but is not restricted to set-based approximations. Moreover, this new approach for stochastic inversion has the following attractive features:

- Has a solid mathematical foundation,
- Does not rely on MCMC sampling to generate samples of the posterior distribution,
- Does not require a statistical assumption on the model error,
- Leverages the adaptive response surface approximations techniques that have been developed at Sandia under the Dakota project for forward uncertainty propagation,
- Easily incorporates multiple sources of correlated data.

In this paper, we compare our new approach with the statistical Bayesian approach by using both approaches to generate samples from the posterior distribution. We also develop a goal-oriented adaptive approach for response surface approximations of the forward map with the specific objective of accurately and efficiently approximating the posterior distribution.

## DETAILED DESCRIPTION OF EXPERIMENT/METHOD:

We let  $m \in \mathcal{A} \subseteq \mathbb{R}^n$  denote the parameters and  $d \in \mathcal{D} \subseteq \mathbb{R}^p$  denote the observed data. We assume that both  $\mathcal{A}$  and  $\mathcal{D}$  are bounded. Let  $F(m)$  denote the mapping between parameter space and the output space defined by a high-fidelity computational forward model. We assume that the forward model is deterministic in the sense that repeated evaluations for a given  $m \in \mathcal{A}$  will give the same results. In general,  $F(m)$  should be replaced by a discretized model and the effect of numerical discretization errors should be incorporated into the analysis [Jakeman 2015]. However, this is beyond the scope of this paper and we will focus on the stochastic inference algorithms.

## The Statistical Bayesian Approach

Following [Marzouk 2009a], we let  $\rho_{\Lambda,prior}(m)$  denote the prior probability density on the input parameters and we use Bayes rule to define the posterior distribution:

$$\rho_{\Lambda,post}(m) = \frac{\rho(d|m) \rho_{\Lambda,prior}(m)}{\int \rho(d|m) \rho_{\Lambda,prior}(m) dm}$$

Where  $\rho(d|m)$  denotes the likelihood function. We define a likelihood function by combining the deterministic model with a statistical model for the measurement error:

$$d = F(m) + \varepsilon$$

If the density of  $\varepsilon$  is given by  $\rho_{\varepsilon}(\varepsilon)$ , then the likelihood function can be written as

$$\rho(d|m) = \rho_{\varepsilon}(d - F(m)).$$

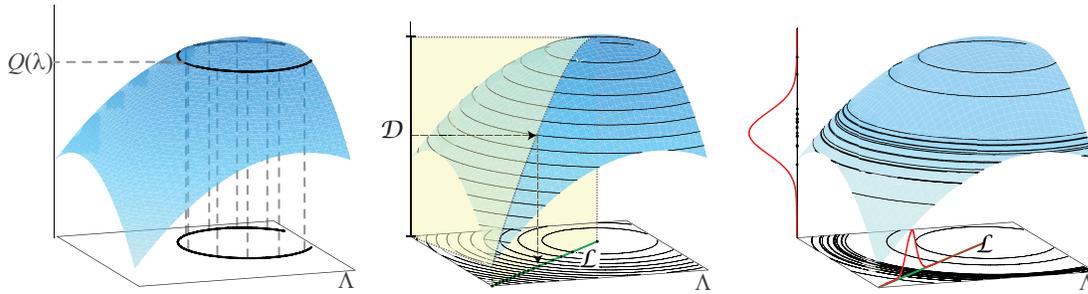
The standard approach for approximating the posterior distribution is to generate samples from the posterior using an MCMC method. A wide variety of MCMC approaches exist [Brooks 2012] and a complete survey is beyond the scope of this paper. The main drawback for all of these approaches is that they tend to require a very large number of evaluations of the forward model. This computational cost can be partially mitigated by using a relatively small number of evaluations of the forward model to construct an emulator model and to then perform MCMC using the emulator [Marzouk 2009a, Marzouk2009b, Dakota].

Approaches that use surrogates for statistical inference can be classified into two main categories: (1) methods that approximate  $F(m)$ ; and (2) methods that approximate the negative of the log likelihood function. For each class adaptively building surrogates can result in significant gains in efficiency over building the surrogate isotropically [Conrad 2015, Li 2014, Pflueger 2010, Zhang 2013].

## The Measure Theoretic Approach

While a full description of the theoretical aspects of the measure-theoretic approach is beyond the scope of this paper, we provide an overview of the computational aspects of this approach and direct the interested reader to [Breidt 2012, Butler 2012] for more details.

The measure-theoretic approach is based on the observation that deterministic inverse problems often have set-valued solutions (see Figure 1 (left)) and a unique solution only exist in an equivalence class of transverse parameterizations that run orthogonally to the contours of the forward map. An analogous argument can be made for stochastic inverse problems: solutions only exist in a set-valued sense. However, a unique solution to the stochastic inverse problem can be determined by utilizing concepts from measure theory and a given volume measure in the parameter space.



**Figure 1: The set-valued inverse of a single output value (left). The invertible map between the set of equivalence classes of transverse parameterizations,  $L$  (middle). The unique inverse distribution in  $L$  (right). Source Butler 2012.**

Following [Breidt 2012, Butler 2012], we assume we are given a probability density on the output space,  $\rho_{obs}(d)$  representing observations and a volume measure,  $\mu_{\Lambda}(m)$ , on the input parameter space. We partition both the input parameter space and output space into measurable subsets,  $\{B_i\}_{i=1}^N$  and  $\{D_j\}_{j=1}^M$ , respectively. We use the Law of Total Probability to write

$$P[B_i] = \sum_{j=1}^M P[B_i|D_j] P_{obs}[D_j]$$

In [Breidt 2012, Butler 2012], the authors use  $A_j$  to denote the inverse image of  $D_j$  in the parameter space obtained by inverting the forward model. We calculate the probability of  $B_i$  given  $D_j$  as:

$$P[B_i | D_j] = \frac{\mu_{\Lambda}(B_i \cap A_j)}{\mu_{\Lambda}(A_j)}$$

which represents the ratio of the volume of  $B_i \cap A_j$  and the volume of  $A_j$ . Of course, this calculation requires that we can approximate the volume measure of  $A_j$ , which is a nonlocal calculation, i.e., the inverse image of  $D_j$  may be a rather large and complicated (non-convex with multiple disconnected sub-regions) portion of the input parameter space. Thus, this calculation requires an approximation of the forward map. The approach in [Breidt 2012, Butler 2012] utilizes a piecewise linear polynomial approximation defined by evaluating the forward model and its gradient via an adjoint-based approach at the centroid of each cell in the partition of the parameter space. This approach works well in low-dimensional spaces but is not practical in high-dimensions.

We can also use the Law of Total Probability and the ratio-of-volumes calculation to construct of piecewise constant approximation (called a simple function approximation in [Breidt 2012, Butler 2012]) of the inverse distribution in the input parameter space:

$$\rho_{\Lambda}(m) \approx \sum_{i=1}^N \sum_{j=1}^M \frac{\mu_{\Lambda}(A_j \cap B_i)}{\mu_{\Lambda}(A_j)} P_{obs}[D_j] \frac{1_{B_i}}{\mu_{\Lambda}(B_i)}$$

where  $1_{B_i}$  denotes the function that is 1 on  $B_i$  and 0 elsewhere.

### A Bayesian Version of the Measure Theoretic Approach

Here, we assume we are given a prior probability density on the input parameter space,  $\rho_{prior}(m)$ . We use the partitions of the input parameter space and the output space defined in the previous section and the Law of Total Probability to write

$$P[B_i] = \sum_{j=1}^M P[B_i|D_j] P_{obs}[D_j]$$

As described in the previous section, the measure-theoretic approach calculates  $P[B_i | D_j]$  using the ratio of the volume of  $B_i \cap A_j$  and the volume of  $A_j$ . Our approach differs in that we utilize Bayes rule for measurable sets:

$$P[B_i | D_j] = \frac{P_{model}[D_j|B_i] P_{prior}[B_i]}{P_{model}[D_j]}$$

where  $P_{model}[D_j]$  denotes the probability of  $D_j$  obtained by propagating the prior probability distribution through the forward model. This set-based Bayesian perspective still requires an approximation of the forward propagation of uncertainty, but it does avoid approximating the inverse image of  $D_j$  (which can be challenging in high-dimensions). We can also use the set-based Bayesian approach to construct a simple function approximation of the posterior distribution in the input parameter space:

$$\rho_{\Lambda,post}(m) \approx \sum_{i=1}^N \sum_{j=1}^M \frac{P_{model}[D_j|B_i] P_{prior}[B_i]}{P_{model}[D_j]} P_{obs}[D_j] \frac{1_{B_i}}{\mu_{\Lambda}(B_i)}$$

Importantly, the set-based Bayesian approach also allows us to consider the limit as the two partitions are refined, leading to the following expression for the posterior distribution defined point-wise in the parameter space:

$$\rho_{\Lambda,post}(m) = \rho_{\Lambda,prior}(m) \frac{\rho_{D,obs}(F(m))}{\rho_{D,model}(F(m))}$$

This result implies that the posterior is equal to the prior times a factor that depends on the ratio of the observed distribution on the data and the model's prediction of the distribution on the data. Both the prior distribution on the input parameters and the observed distribution on the outputs are given (or at least assumed), so we only need to compute distribution on the outputs corresponding to the forward propagation of the prior distribution on the input parameters

through the forward model. We note that if the propagation of the prior distribution through the model matches the data, then the posterior is equal to the prior. However, we emphasize that this is not the same as the statistical Bayesian approach since we utilize the propagation of the prior to compute a likelihood function rather than a statistical error model.

Being able to directly approximate the posterior density at any point in the parameter space, means that we can generate samples of the posterior without relying on MCMC techniques. Instead we can just use rejection sampling to draw samples from the posterior. Unlike MCMC rejection sampling is not conditional on previously sampled points. Consequently rejection sampling will be able to handle multi-modal distributions more easily than MCMC methods and is also embarrassingly parallel.

In contrast to traditional statistical Bayesian inference, our approach does not require a statistical error model to be specified. Rather we require a distribution on the observational data. There appears to be a close connection between statistical error models and densities on the observations, however as yet we have been unable to form this connection.

### **Adaptive sparse grids**

Our measure theoretic Bayesian inference approach requires an approximation of the distribution in the data space corresponding to a forward propagation of the prior distribution on the parameters. Approximating the density can be infeasible when the simulation model is expensive. One way to reduce the computational cost for high-fidelity computational models is to utilize a response surface approximation for the forward propagation of the prior.

Many recent research efforts have focused on developing scalable and adaptive techniques to utilize response surface approximations (also referred to as surrogates or emulators) for forward propagations of uncertainty. The accuracy of the probabilistic predictions (probability density functions, mean, variance, probability of failure, etc.) obtained by sampling a surrogate depends directly on the accuracy of the approximation. Consequently, it is important that the samples used to build the surrogate are chosen judiciously. Numerous adaptive techniques, such as sparse grids, and Gaussian processes, have been used successfully to build surrogates for the forward propagation of uncertainties through a model.

In this report we will leverage response surface techniques for forward propagation of uncertainty to reduce the cost of forming density estimates of  $\rho_{D,obs}$ . Specifically we will utilize locally adapted sparse to develop adaptive response surface approximations that target regions of high posterior density.

Sparse grids represent a function  $f$  as a linear combination of multivariate basis functions.

$$f = \sum_{l \in L} \sum_{i \in I} v_{l,i} \varphi_{l,i}(x)$$

where  $I$  and  $L$  are sets of multidimensional index functions that can be used to control the resolution of the sparse grid. Specifically the approximation is defined on a set of anisotropic grids, which are the tensor product of one-dimensional meshes. The multi-index  $l$  denotes the level of refinement of the grid in each dimension. The multi-index  $i$  denotes the position of a

node within each one-dimensional grid. Here we use multivariate basis functions which are the tensor products of compact one-dimensional piecewise polynomial basis functions defined on equidistant meshes [Jakeman 2013]. Sparse grids are built hierarchically. The coefficients  $v$  are referred to as the hierarchical surplus and are the difference between the function value and the sparse grid at each point before it is added to the grid.

Two main approaches are used for adapting a sparse grid: local refinement [Pflueger 2010, Ma 2009, Jakeman 2013]; and dimension [Hegland 2003, Gerstner 2003, Ganapathysubramanian 2007] refinement. In this paper we will focus on local refinement. Local adaptivity refines a sparse grid one point at a time. At any stage of the algorithm we have two sets of points, those points that have been refined previously, and points that have not been refined. We refer to the latter set of points as the active set. Refinement of the sparse grid proceeds by identifying the point in the active set with the largest error indicator and removing it from the active set. The children of this point are then created and added to the active set. Any single point will have at most  $2d$  children and can be found by traversing a multi-dimensional binary tree [Ma 2009].

The effectiveness of sparse grid adaptivity is dependent on the quality of the error indicator used to prioritize each point in the active set. The hierarchical surplus is a natural error indicator and has been used extensively to efficiently approximate higher-dimensional functions, such as the log-likelihood function [Pflueger 2010, Zhang 2013]. In this paper we will compare adaptive sparse grids built using a hierarchical surplus based error indicator, with an error indicator that estimates the error in the posterior density estimated by our measure-theoretic approach.

## RESULTS:

In this paper, we focus on the following numerical demonstrations:

1. A comparison of the samples of the posterior generated by the proposed Bayesian approach with a statistical Bayesian approach using a Differential Evolution Adaptive Metropolis (DREAM) algorithm as the MCMC sampler.
2. A demonstration of the gains in accuracy and/or efficiency that can be achieved by employing a locally adaptive sparse grid approximation with refinement indicators that incorporate both the prior and the posterior distributions.

We use two very simple model problems to perform the numerical demonstrations. Both model problems have two parameters and one quantity of interest to facilitate a variety of comparisons. Higher-dimensional problems and multiple (correlated) quantities of interest have also been examined, but not as a part of this LDRD project and will not be discussed here.

The first problem we consider is a 2-component nonlinear system of equations with two parameters introduced in [Breidt 2012]:

$$\lambda_1 x_1^2 + x_2^2 = 1$$

$$x_1^2 - \lambda_2 x_2^2 = 1$$

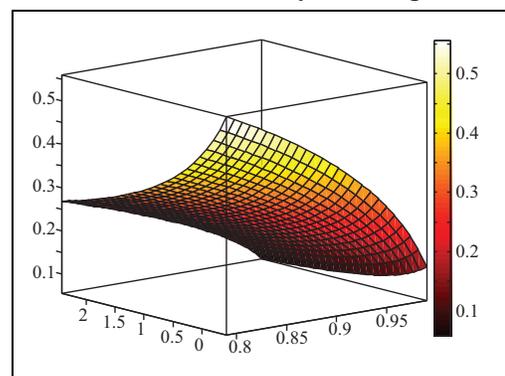


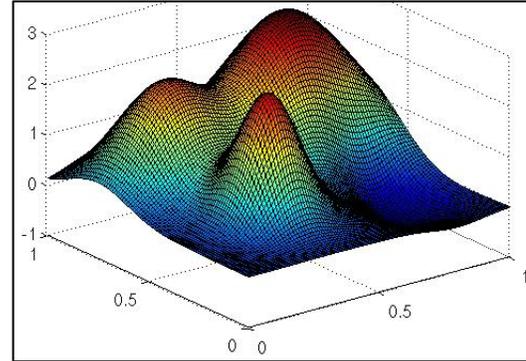
Figure 2: Quantity of interest from the 2-component nonlinear system.

The quantity of interest is the second component. The parameter ranges are given by  $\lambda_1 \in [0.79, 0.99]$  and  $\lambda_2 \in [1 - 4.5\sqrt{0.1}, 1 + 4.5\sqrt{0.1}]$  which are chosen as in [Breidt 2012] to induce an interesting variation in the quantity of interest. The quantity of interest as a function of the parameters over the given ranges is provided in Figure 2.

The second problem we consider is a 1-component function of two parameters,  $\lambda_1 \in [0,1]$  and  $\lambda_2 \in [0,1]$ , given by a sum of four Gaussian peaks:

$$\begin{aligned}
 q = & 2.0e^{-\frac{(\lambda_1-0.25)^2}{2 \cdot 0.15^2}} - \frac{(\lambda_2-0.75)^2}{2 \cdot 0.15^2} \\
 & + 3.0e^{-\frac{(\lambda_1-0.75)^2}{2 \cdot 0.2^2}} - \frac{(\lambda_2-0.75)^2}{2 \cdot 0.2^2} \\
 & + 2.5e^{-\frac{(\lambda_1-0.33)^2}{2 \cdot 0.1^2}} - \frac{(\lambda_2-0.33)^2}{2 \cdot 0.1^2} \\
 & - 1.0e^{-\frac{(\lambda_1-0.8)^2}{2 \cdot 0.1^2}} - \frac{(\lambda_2-0.4)^2}{2 \cdot 0.2^2}
 \end{aligned}$$

The function is shown in Figure 3.



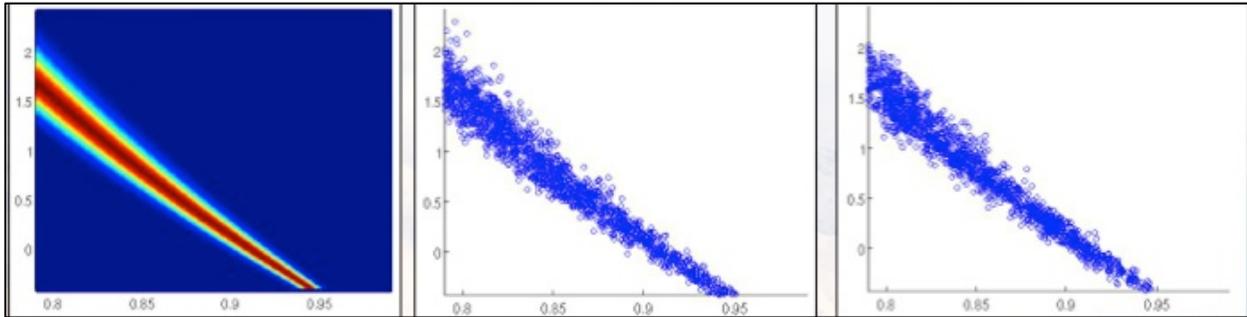
**Figure 3: Quantity of interest from the Gaussian peaks problem.**

### Comparison with Statistical Bayesian with DREAM MCMC

Our first objective is to use the posterior distribution computed using the new Bayesian approach to generate a set of samples from the posterior and to compare these samples with those generated using a statistical Bayesian approach with an advanced MCMC algorithm, namely DREAM. All of the numerical results in this section were generated using DAKOTA [Dakota]. The numerical implementation of the new Bayesian approach in DAKOTA is called WASABI (Weighted Adaptive Surrogate Approximations for Bayesian Inference) and currently requires a surrogate approximation in order to efficiently propagate the prior distribution through the forward model. A non-surrogate approach has also been developed, but will not be discussed in this paper.

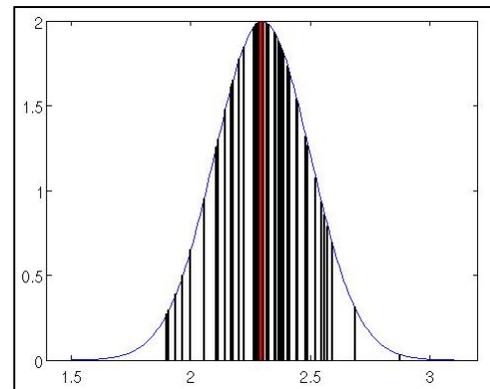
The surrogate model used in this section is a Gaussian process emulator constructed using 100 evaluations of the forward model with a quadratic trend function. Given the surrogate model, the propagation of the prior through the forward model is estimated using a large number (typically 10,000) samples of the surrogate model. Of course, utilizing a surrogate model in this manner introduces an additional error due to the interpolation of the surrogate rather than the evaluation of the true forward model. The effect of this error on probabilistic quantities of interest has been investigated by the authors in previous work [Jakeman 2015] and understanding how these errors affect the posterior distribution will be the subject of future work.

For our first demonstration, we consider the 2-component nonlinear system and assume that the observed distribution on quantity of interest is a normal distribution with mean 0.3 a standard deviation of 0.1. Assuming a Gaussian distribution is not required for WASABI, but the assumption does facilitate an easier comparison with the statistical Bayesian model with an additive Gaussian noise model. We assume a uniform prior distribution on the input parameters.



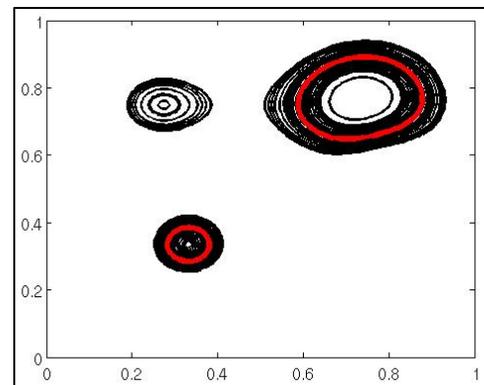
**Figure 4: The approximation of the posterior distribution (left), the samples from the posterior generated by WASABI (middle), and the samples from the posterior generated by MCMC-DREAM (right).**

Using WASABI, we directly construct the posterior distribution on a uniform grid of points in the input parameters space and plot this in Figure 4 (left). Since we can directly interrogate/approximate the posterior distribution, we can easily generate samples from the posterior using rejection sampling. Specifically we sample a point from a proposal distribution then if the value of the posterior at that point is greater than a random uniform number in  $U(0,1)$  we reject the point otherwise we accept it. The samples from the posterior distribution using WASABI are given in Figure 4 (middle). For comparison, the samples from the posterior distribution using MCMC/DREAM are given in Figure 4 (right) For this example, the samples from the posterior are fairly consistent.



**Figure 5: The observed Gaussian distribution on the output. The mean is given in red and samples from the distribution are given in black.**

For our second demonstration, we consider the Gaussian peaks function and assume the observed distribution on the quantity of interest is Gaussian with mean 2.3 and a standard deviation of 0.2. This distribution is shown in Figure 5 along with 50 samples from the distribution. In Figure 6, we plot the generalized contours associated with each of these samples. The posterior distribution computed using the new Bayesian approach with a uniform prior distribution is given in Figure 7 (left). We see that the support of the posterior is consistent with the density of the generalized contours in Figure 6. In Figures 7, we show the samples from the posterior computed using the new Bayesian approach (middle) the statistical Bayesian approach (right). This particular example is more challenging for MCMC due to the multi-modal nature of the posterior distribution. The DREAM algorithm addresses this challenge by running multiple concurrent MCMC chains. Here, we utilize three chains with inter-chain mixing every ten samples. This



**Figure 6: Generalized contours corresponding to the mean of observed data (red) and the samples from the observed distribution (black).**

appeared to be sufficient for MCMC to identify the three disconnected modes of the posterior. We note that WASABI also needs to explore the space in order to locate all of the modes of the posterior, but since we do not rely on MCMC we do not run the risk of getting trapped within one of the modes.

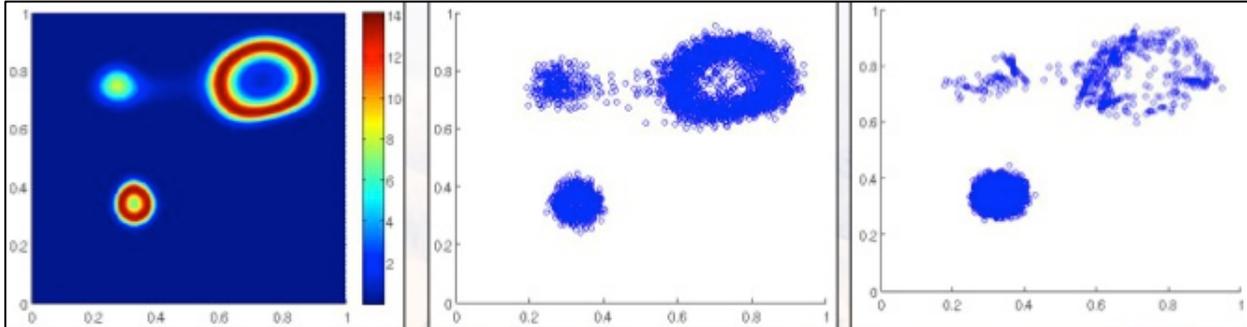


Figure 7: The approximation of the posterior distribution (left), the samples from the posterior using WASABI (middle), and the sample from the posterior using MCMC-DREAM (right).

### Locally Adaptive Sparse Grids

Next, we demonstrate how incorporating an approximation of the posterior distribution into an adaptive refinement criteria can reduce the computational cost required to approximate the posterior to a chosen tolerance.

We demonstrate this adaptive approach on the Gaussian peaks problem introduced in the previous section. We start from an initial grid with a relatively small number of points and explore the effect of various refinement indicators on the accuracy of the posterior distribution. The reference posterior distribution was computed using a high-order tensor product quadrature rule with the aforementioned surrogate-free version of the new Bayesian approach. The computational cost associated with this reference posterior is quite high, but it is free from surrogate approximation errors which is the focus of this demonstration.

In Figure 8, we show the convergence of the L2-norm of the posterior using the following refinement criteria

$$(\alpha\rho_{\lambda,prior} + (1 - \alpha)\rho_{\lambda,post})|v|vol(v)$$

Here the hierarchical surplus  $v$ , and the densities  $\rho_{\lambda,prior}$  and  $\rho_{\lambda,post}$  are estimated when a point is first built and added to the active set. By varying  $0 \leq \alpha \leq 1$  we can balance the need to resolve non-local effects of the model  $f$  on the posterior and the need to concentrate samples in regions of high-posterior probability.

Specifically we consider the following forms of refinement

- Isotropic refinement
- Standard surplus-based refinement with weights based on the prior ( $\alpha = 1$ )
- Surplus-based refinement with weights based only on the posterior ( $\alpha = 0$ )
- Surplus-based refinement with weights based on a convex combination of the prior and posterior. ( $\alpha = 0.1, \alpha = 0.5$ )

We see that the isotropic refinement criteria (dashed black line) converges linearly as one would expect considering we are using linear basis functions. We also see that the standard approach

using the hierarchical surplus weighted by the prior distribution as a refinement indicator (green line) performs better than the isotropic case. We observe that using the hierarchical surplus weighted by the posterior distribution does not converge very well. There are two reasons that this refinement indicator does not perform well. Firstly the initial grid that is used to initiate adaptivity is too coarse to resolve the local minima of the posterior and so refinement stops prematurely in these regions. Secondly, even when the initial grid is sufficiently fine, the  $\alpha = 0$  indicator does not refine outside the regions of non-zero posterior density and thus does not capture the non-local effects of these uncaptured regions on the posterior. This is despite the fact that the initial grid was sufficiently fine to locate all three modes of the posterior. In other words, simply approximating the forward map in regions where the posterior has non-zero support is insufficient to actually approximate the posterior since the values of the posterior depend on the full forward propagation of the prior.

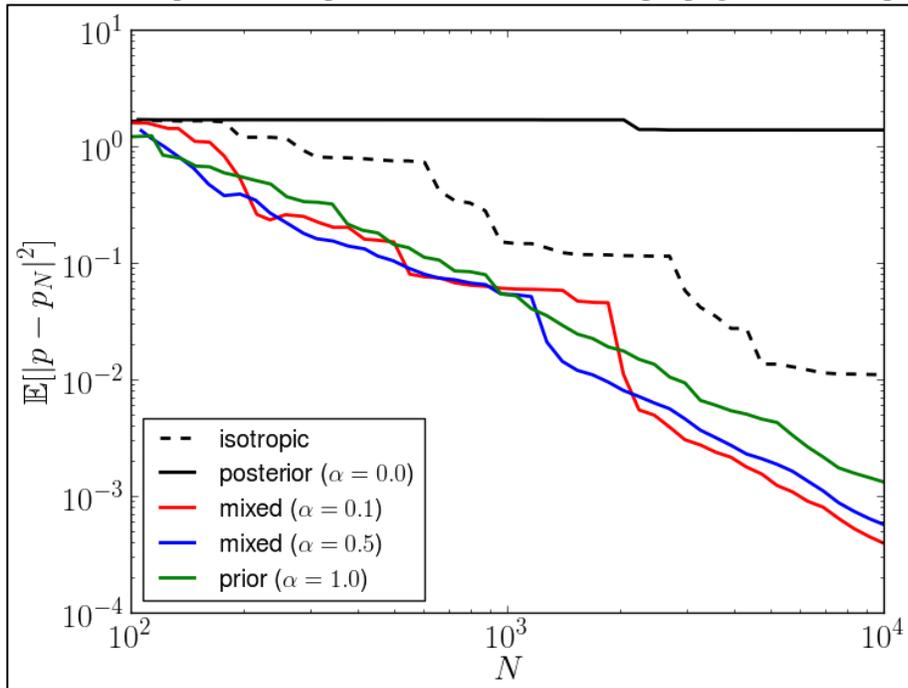


Figure 8: Convergence of the posterior for a variety of adaptive approaches.

To balance nonlocal effect due to the forward propagation of the prior with the local effect of the approximation of the posterior, we considered various convex combinations of prior and posterior density estimates. The results of the balanced approach for two different mixing parameters are also given in Figure 8. We see that a mixing parameter of  $\alpha=0.5$  tends to consistently perform better than the prior weighted case ( $\alpha=1.0$ ), but that the case of  $\alpha=0.1$  tends to perform the best at the higher-number of model evaluations. This is due to the fact that once the nonlocal effects have been sufficiently resolved, it is better to spend the computational effort on the posterior refinement indicator.

## DISCUSSION:

Despite significant recent progress, the effectiveness of state-of-the-art methods for statistical Bayesian inference is still greatly limited by high numbers of model parameters. In high-dimensional parameter spaces, sufficient exploration and maintaining a minimum acceptance

rate has proven challenging for even the most adaptive MCMC schemes [Liebermann 2010]. In addition to the challenges posed by high-dimension, MCMC can have difficulty exploring complex posterior surfaces that contain multiple regions of attraction and numerous local optima. Recently a few algorithms have been developed to address these limitations. Specifically dimension reduction algorithms have been used to increase the acceptance rate of MCMC in high dimensions [Cui 2014] and multiple chains have been used to address the issues of multi-modality in the posterior [Vrugt 2011].

In contrast to traditional statistical Bayesian inference, which typically employs MCMC, our Bayesian inference approach does not require MCMC to draw samples from the posterior. Because our method can directly interrogate the posterior we can simply use rejection sampling to draw samples from the posterior. Provided the support of the proposal distribution encompasses (with high probability) the posterior, the rejection rate of our approach is proportional to volume of the non-zero support of the posterior relative to the support of the proposal distribution. The ability to draw samples from the posterior, however, is not dependent of the number of local optima (modes) in the posterior. More over rejection sampling is embarrassingly parallel and so, unlike MCMC, we can obtain all the samples of the posterior simultaneously.

The number of samples required by our inference approach is equal to the number of samples from the prior propagated through the forward model plus the number of samples used to interrogate the posterior. The accuracy of the estimates of the posterior density however is only dependent on the number of prior samples used. The prior samples are used to evaluate the model and form a density estimate of the response QoI and thus a large number of samples are needed for an accurate estimate of this density. Specifically because we need to form a density estimate of the response QoI, the accuracy of our algorithm will have a Monte Carlo type convergence rate with respect to the number of prior samples. We also remark that the number of samples used to interrogate the posterior may be small if one simply wants to evaluate the density at a few selected points, or large if one wants to generate a large number of samples of the posterior using rejection sampling.

To overcome the slow rate of convergence of Monte-Carlo, we built a surrogate of the forward model that is used in place of the expensive simulation model when propagating prior samples through the forward model. Provided the number of function evaluations required to build the surrogate of sufficient accuracy is less than the number of prior+posterior samples then the using a surrogate is warranted. A similar approach is often used for MCMC based statistical inference, where the MCMC chain is built on a low-cost surrogate.

The accuracy of our estimate of the posterior density is dependent on the accuracy of the surrogate of the forward model. Numerically we have shown that as the error in the surrogate tends to zero the estimate of the posterior density will converge to the posterior density formed without the surrogate, for a fixed set of prior samples. Future work is needed to provide theoretical justification for this result.

Surrogates of the forward model can be built isotropically by treating all dimensions and regions of parameter space equally. However we demonstrated that a reduction in the number of

simulation runs can be achieved with the use of adaptivity. Specifically we used local sparse grid adaptivity with a number of error criterion used to control refinement. It was found that simply refining in regions where the error in the surrogate was high is effective, however by also including estimates of the error in the posterior, further gains can be made. We presented a problem with 3 modes of high-posterior probability and localized support. This problem presents a pathological challenge for adaptive refinement of sparse grids or any surrogate. However it must be noted this is not a limitation of the non-surrogate based inference method. The localized nature of the posterior meant that if adaptivity is started before the local non-zero features have been identified then refinement will stop prematurely or continue in regions of parameter space that are not informed by data and thus degrading the accuracy of the estimated posterior. We believe that adaptivity will be much more effective when the support of the posterior is more global. Future work is needed to test this hypothesis.

## **ANTICIPATED IMPACT:**

### **Future research**

The use of surrogates with our Measure theoretic Bayesian inference approach appears very promising. Future work could explore additional means of reducing the cost of inference, other than sparse grid collocation which is limited to moderate dimensional problems. Future areas of research include: (1) dimension reduction techniques, such as active subspaces, which identify the regions of parameter space that are important for data informed prediction; and (2) multi-fidelity approaches that combine the increased predictive strengths of high-fidelity models with the reduced computational cost of lower fidelity models to enable more extensive exploration of model uncertainties -probability subspaces.

This project developed a formulation of Bayesian inference that is able to rigorously leverage the strengths of forward UQ whilst still constraining prior assumptions of uncertainty using observational data. This concept is a fundamental concept of a DOE early career award pre-proposal submitted by John Jakeman and entitled “Goal-oriented data-informed prediction for large-scale systems subject to uncertainty. The goal of the project is to develop new mathematical and numerical inference methods for quantifying uncertainty in predictions obtained from large-scale physical models. Specifically to develop algorithms that drastically reduce the number of high-fidelity simulations (cost) required to quantify uncertainty whilst attempting to maintain the deterministic prediction accuracy of the high-fidelity model.

The pre-proposal outlined novel mathematical formulations, including but not limited to the work presented here, that integrate inverse and forward propagation of uncertainty, which leverage data and drive computational effort in a manner dictated by the requirements of prediction. The goal-oriented algorithms proposed are based on a reformulation of the sequential inverse-forward UQ problem as a set of purely forward UQ problems. This reformulation allows the methods developed to leverage the strengths of forward UQ methods, such as high-order convergence, and the ability to utilize adjoint information such as gradients and error estimates.

The capabilities developed in this project will be further investigated in an ASC V&V project led by T. Wildey in FY16. The objective of this project will be to pursue goal-oriented adaptive response surface approximations within Dakota for the purposes of stochastic inversion and for inversion for prediction. We will also seek to develop the set-based version of the new Bayesian

approach utilizing the recently developed Voronoi capabilities in DAKOTA.

This set-based version will provide a stronger connection with the measure-theoretic approach in the literature.

### **Software**

The methods developed here will be made available to the DOE community via deployment with the software package Dakota. Dakota is an optimization and uncertainty quantification tool that is widely used throughout the laboratory and by many external research and industrial partners. As such, deployment of our algorithm in Dakota will impact a wide range of customers that have a direct or indirect impact on Sandia's mission.

With the help of this funding, a prototype of our approach has been implemented in Dakota that uses un-adapted surrogates, polynomial chaos expansions and Gaussian process models. With future funding from ASC V&V, we will extend this prototype so use the dimension-adaptive sparse grids in the Dakota package Pecos. Any advances made from future research are also candidates for inclusion in Dakota.

### **Programmatic impact**

The work outlined in this report will assist engineers working on a broad range of applications of interest to the Sandia National Laboratories, to understand and quantify the uncertainties in their predictive models. Example of classes of models that are particularly relevant to this report are:

- Models of fluid structure interactions, important for modeling of nuclear reactor systems,
- Random field flow models, used to predict sea level rise due to ice-sheet melt and model sub-surface flows.
- Shock-hydrodynamic models, which are critical to ensure the safety and reliability of the nuclear weapon stockpile.
- Multi-scale material models, which are required to characterize the structural integrity of components built using additive manufacturing processes.

All of these classes of models have existing either robust or prototype Trilinos-based solvers (Drekar, Albany, MILO) that can be utilized by this project with minimal start-up time and are built upon components that have demonstrated scalability on  $O(100,000)$  cores.

### **CONCLUSION:**

In this paper we have described a recently developed approach for Bayesian inference based on a measure theoretic perspective. We compared the new approach the standard statistical Bayesian approach and demonstrated that the new approach can easily generate comparable sets of samples from the posterior distribution without relying on MCMC. We also discussed and presented numerical results for a new approach for adaptive refinement of a sparse grid surrogate model based on both the prior and the posterior distributions. We believe that this new approach provides a promising alternative to both the measure-theoretic and statistical Bayesian approaches and there are still a number of interesting questions left to be addressed by future work.

## REFERENCES:

[Babuska 2007] I.M. Babuska, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 45(3):1005–1034, 2007.

[Bliznyuk 2012] N. Bliznyuk, D. Ruppert, and C.A. Shoemaker. Local derivative-free approximation of computationally expensive posterior densities. *Journal of Computational and Graphical Statistics*, 21(2):476–495, 2012.

[Breidt 2012] J. Breidt, T. Butler, and D. Estep, A computational measure theoretic approach to inverse sensitivity problems I: Basic method and analysis, *SIAM J. Numer. Analysis*, 49 (2012), pp. 1836–1859.

[Brooks 2011] S. Brooks, A. Gelman, G. Jones, and X. Meng. *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011.

[Butler 2012] T. Butler, D. Estep, and J. Sandelin, A measure-theoretic computational method for inverse sensitivity problems II: A posterior error analysis, *SIAM J. Numer. Analysis*, 50 (2012), pp. 22–45.

[Conrad 2015] P.R Conrad, Y.M. Marzouk, N.S Pillai, and A. Smith. Asymptotically exact mcmc algorithms via local approximations of computationally intensive models. *Journal of the American Statistical Association*, In press, 2015.

[Cui 2014] T. Cui, J. Martin, Y.M. Marzouk, A. Solonen, and A. Spantini. Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems*, 30(11), 2014.

[Dakota] Adams, B.M., Bauman, L.E., Bohnhoff, W.J., Dalbey, K.R., Ebeida, M.S., Eddy, J.P., Eldred, M.S., Hough, P.D., Hu, K.T., Jakeman, J.D., Swiler, L.P., and Vigil, D.M., “DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 5.4 User’s Manual,” Sandia Technical Report SAND2010-2183, December 2009. Updated April 2013.

[Ganapathysubramanian 2007] B. Ganapathysubramanian and N. Zabarar. Modeling diffusion in random heterogeneous media: Data-driven models, stochastic collocation and the variational multiscale method. *Journal of Computational Physics*, 226(1):326–353, SEP 10 2007.

[Gerstner 2003] T. Gerstner and M. Griebel. Dimension-adaptive tensor-product quadrature. *Computing*, 71(1):65–87, SEP 2003.

[Hegland 2003] M. Hegland. Adaptive sparse grids. In K. Burrage and Roger B. Sidje, editors, *Proc. of 10th Computational Techniques and Applications Conference CTAC-2001*, volume 44, pages C335–C353, April 2003.

[Jakeman 2013] J.D. Jakeman and S.G. Roberts. Local and dimension adaptive stochastic

collocation for uncertainty quantification. In Jochen Garcke and Michael Griebel, editors, *Sparse Grids and Applications*, volume 88 of *Lecture Notes in Computational Science and Engineering*, pages 181–203. Springer Berlin Heidelberg, 2013.

[Jakeman 2015] J.D. Jakeman and T. Wildey. Enhancing adaptive sparse grid approximations and improving refinement strategies using adjoint-based a posteriori error estimates. *Journal of Computational Physics*, 280(0):54 – 71, 2015.

[Li 2014] J. Li and Y. M. Marzouk. Adaptive construction of surrogates for the bayesian solution of inverse problems. *SIAM Journal on Scientific Computing*, 36(3):A1163–A1186, 2014.

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

[Ma 2009] X. Ma and N. Zabarar. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *Journal of Computational Physics*, 228:3084–3113, 2009. 1

[Marzouk 2009a] Y. M. Marzouk and D. Xiu. A stochastic collocation approach to bayesian inference in inverse problems. *Communications in Computational Physics*, 6(1):826–847, 2009.

[Marzouk 2009b] Y. M. Marzouk and H. N. Najm. Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6):1862 – 1902, 2009.

[Pfluger 2010a] D. Pfluger, B. Peherstorfer, and H.-J. Bungartz. Spatially adaptive sparse grids for high-dimensional data-driven problems. *Journal of Complexity*, 26(5):508–522, 2010.

[Pfluger 2010b] Dirk Pfluger. *Spatially Adaptive Sparse Grids for High-Dimensional Problems*. Dissertation, Institut für Informatik, Technische Universität München, München, February 2010.

[Vrugt 2011] J. Vrugt, C. ter Braak, C. Diks, A.B. Robinson, J.M. Hyman, and D. Higdon. Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3):273–290, 2011.

[Zhang 2013] G. Zhang, D. Lu, M. Ye, M. Gunzburger, and C. Webster. An adaptive sparse-grid high-order stochastic collocation method for bayesian inference in groundwater reactive transport modeling. *Water Resources Research*, 49:6871–6892, 2013.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.