Adaptive Sampling for Non-intrusive Reduced Order Models Using Multi-Task Variance

Abhijnan Dikshit[1][0000−0002−8745−6873], Leifur Leifsson[1][0000−0001−5134−870X], Slawomir Koziel[2,3][0000−0002−9063−2647], and Anna Pietrenko-Dabrowska[3][0000−0003−2319−6782]

[1] School of Aeronautics and Astronautics, Purdue University, West Lafayette, Indiana 47907, USA
{adikshit@purdue.edu, leifur@purdue.edu}
[2] Engineering Optimization & Modeling Center, Department of Engineering, Reykjavík University, Menntavegur 1, 102 Reykjavík, Iceland
koziel@ru.is
[3] Faculty of Electronics Telecommunications and Informatics, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland
anna.dabrowska@pg.edu.pl

Abstract. Non-intrusive reduced order modeling methods (ROMs) have become increasingly popular for science and engineering applications such as predicting the field-based solutions for aerodynamic flows. A large sample size is, however, required to train the models for global accuracy. In this paper, a novel adaptive sampling strategy is introduced for these models that uses field-based uncertainty as a sampling metric. The strategy uses Monte Carlo simulations to propagate the uncertainty in the prediction of the latent space of the ROM obtained using a multi-task Gaussian process to the high-dimensional solution of the ROM. The high-dimensional uncertainty is used to discover new sampling locations to improve the global accuracy of the ROM with fewer samples. The performance of the proposed method is demonstrated on the environment model function and compared to one-shot sampling strategies. The results indicate that the proposed adaptive sampling strategies can reduce the mean relative error of the ROM to the order of 8×10^{-4} which is a 20% and 27% improvement over the Latin hypercube and Halton sequence sampling strategies, respectively at the same number of samples.

Keywords: Adaptive sampling · Field-based uncertainty · Reduced order modeling · Multi-task Gaussian process · Monte Carlo simulation.

1 Introduction

The need to reduce the computational cost of high-fidelity simulations has been of great interest in the fields of science and engineering. A computationally cheaper alternative to high-fidelity simulations has been identified in the form of reduced order modeling (ROM) strategies. ROMs can be intrusive [2], i.e., they alter the governing equations of a problem and offer a simpler alternative that

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

is faster to solve, or they can be non-intrusive [24], i.e., they can be data-driven models that use high-fidelity data and machine learning techniques to create an approximation of simulation output. Intrusive ROMs typically require access to the simulation source code which may be difficult to acquire. Implementing intrusive ROMs also requires extensive knowledge and effort to manipulate the governing equations and simulation source code. These challenges can be overcome by using non-intrusive ROMs that only require access to the data generated by a high-fidelity simulation. Non-intrusive ROMs have gained popularity and have been extensively applied in the domain of fluid mechanics to predict high-dimensional field-based variables [20, 12, 6].

A common issue with the creation of ROMs is that they often require a large number of samples to achieve the desired approximation quality of high-dimensional field-based variables. These samples are usually generated in one-shot using, for example, Latin hypercube sampling (LHS) [22] or low discrepancy sequences, such as Halton sequences [17]. These methods are usually space-filling design of experiments, however, these methods do not generate sampling plans that are tailored to the approximation quality of non-intrusive ROMs. As a result, a higher number of samples are required to lower the prediction error of the ROM leading to a large number of high-fidelity simulations.

To overcome this disadvantage, adaptive sampling strategies can be used to choose samples within a parameter space that directly target the prediction quality of the ROM. An extensive overview of adaptive sampling and different types of sampling strategies is given in [15] and [9]. The initial work on adaptive sampling of ROMs was focused on the sampling of POD-based ROMs. Braconnier et al. [5] and Guenot' et al. [11] developed sampling strategies based on a measure of quality of POD basis generation and the prediction quality of the low-dimensional space interpolation model of the ROM. Wang et al. [7] extended the work of Guenot et al. [11] by developing a fully-automated approach that utilizes the proposed adaptive sampling metrics in a combined manner.

Yang and Xiao [23] proposed an adaptive sampling algorithm for a POD-based ROM that used the variance estimate and gradient information of the low-dimensional space modeled by a Gaussian process regression model. They employ a weighting strategy for the variance of the prediction of each modal coefficient and the magnitudes of the gradients. The weights are based on the singular values of the modal coefficients. This weighting strategy restricts the use of this method to only POD-based ROMs. The work of Karcher and Franz [14] and Franz et al. [8] developed adaptive sampling strategies that can be implemented for both POD-based and manifold learning ROMs. Some of the sampling metrics used in these studies include manifold-filling metrics, residual of governing equations, and error estimations, such as leave-one-out cross-validation error and mean squared error of the interpolation model of the low-dimensional space.

Most of the previous work on adaptive sampling strategies for ROMs has typically utilized the approximation quality of the low-dimensional space directly. This approach does not account for the approximation quality in the

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

high-dimensional space which is the actual quantity of interest in a ROM prediction. The current work will quantify the variance in the high-dimensional solution of the ROM by the use of forward propagation through the backmapping procedure of the ROM. The high-dimensional variance estimate will then be utilized as a sampling objective in an adaptive sampling algorithm. It is hypothesized that the introduction of high-dimensional variance information will improve the performance of the adaptive sampling.

In this work, a novel adaptive sampling strategy is proposed for ROMs based on the multi-task variance of the latent space obtained by the free-form parameterization multi-task Gaussian process (MTGP) formulation [4]. In this case, the latent space is modeled using a MTGP model to predict all the latent space coefficients simultaneously. Monte Carlo sampling is used to propagate the multi-task variance of the latent space to the high-dimensional solution space to estimate the high-dimensional variance and employ it to discover new sampling locations to improve the global accuracy of the non-intrusive ROM. The proposed adaptive sampling strategy is a novel approach as it uses high-dimensional variance information as an infill criteria to search the design space for new samples rather than using only the latent space variance. This strategy will introduce more information in the form of field-based variance information into the search for new samples that could lead to a better search and possibly higher accuracy with fewer samples. The proposed method is also agnostic to the dimensionality reduction method of the ROM. To illustrate the possible benefits, the proposed algorithm is benchmarked against traditional one-shot sampling strategies, in particular, LHS and Halton sequences.

The next section will describe the novel adaptive sampling algorithm proposed in this paper. The section following that will describe the application of the adaptive sampling algorithm to a 2D analytical test function and a performance comparison with benchmarking methods. The final section of the paper will present the main conclusions and possible future work.

2 Field-based adaptive sampling

This section introduces the high-dimensional field-based adaptive sampling strategy for ROMs proposed in this work.

2.1 Proposed adaptive sampling algorithm

The general process of adaptive sampling is driven by an optimization problem that chooses a new sampling point by maximizing a sampling objective. This can be mathematically represented as

$$\mathbf{x}_{new} = \text{argmax } f(\mathbf{x}), \tag{1}$$

where $\mathbf{x} = (x_1, x_2, ...x_n)$ is a vector of model parameters, $f(\mathbf{x})$ denotes the sampling objective and $\mathbf{x}_{new}$ is the new sampling point. The sampling objective

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

itself is based on a measure of the prediction uncertainty of the ROM. In the case of the field-based adaptive sampling proposed here, each point in the field will be assigned a prediction uncertainty rather than an aggregated prediction uncertainty being used for the entier field prediction. However, to facilitate the calculation of a sampling objective the prediction uncertainty from all points in the field must be aggregated using mathematical operations such as averaging. This means that in the case of field-based adaptive sampling proposed here, the sampling problem should be represented as

$$\max_{\mathbf{x}} f = g(\hat{\sigma}(\mathbf{x})), \tag{2}$$

where $\hat{\sigma}(\mathbf{x})$ is the prediction uncertainty field for a given set of parameters and $g(\hat{\sigma}(\mathbf{x}))$ is an aggregation function to aggregate the prediction uncertainty in the field in a meaningful way.

The main contribution of this work is to introduce high-dimensional information in the adaptive sampling algorithm. To do this, the variance estimate of the latent space prediction provided by an MTGP model will be propagated to the high-dimensional solution to attain a high-dimensional variance estimate. The proposed propagated uncertainty sampling algorithm is shown in Algorithm 1. The main features of the algorithm are discussed in detail in the following subsections. The algorithm described in these subsections will be repeated until a fixed computational budget is reached. Here, the computational budget refers to the number of samples that can be obtained within the limits of computational cost.

2.2 Dimensionality reduction

To start the algorithm, an initial sampling plan $\mathbf{X} = \{\mathbf{x}^{(1)}, ..., \mathbf{x}^{(n)}\}$ is generated using a one-shot sampling strategy. In this work, Halton sequences are used to generate the initial sampling plan that contains only a few number of samples. This is used as the base sampling plan to create the initial ROM and calculate variance estimates of the ROM. The high-dimensional field-based solutions $\mathbf{Y} = \{\mathbf{y}^{(1)}, ..., \mathbf{y}^{(n)}\} \in \mathbb{R}^{m \times n}$ are calculated for each sample in $\mathbf{X}$ to

Algorithm 1 Adaptive sampling using field-based variance estimate.

Require: initial data sets $(\mathbf{X}, \mathbf{Y})$, computational budget N_{max}
  repeat
    obtain POD modes $\mathbf{\Phi}$ through SVD
    project data to low dimensional space, $\mathbf{z} \leftarrow \mathbf{\Phi}^T\mathbf{Y}$
    fit $MTGP_z$ to the data set $(\mathbf{X}, \mathbf{z})$
    $\mathbf{x}_{new} \leftarrow$ argmax $g(\hat{\sigma}(\mathbf{x}))$
    $\mathbf{X} \leftarrow \mathbf{X} \cup \mathbf{x}_{new}$
    get field-based data $(\mathbf{y}_{new})$ at $\mathbf{x}_{new}$
    $\mathbf{Y} \leftarrow \mathbf{Y} \cup \mathbf{y}_{new}$
  until computational budget expires
  return final sampling plan, $\mathbf{X}$

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

create the data for the ROM. $\mathbf{Y}$ is a matrix containing n fields of dimensionality m. To focus on the adaptive sampling algorithm itself, the simplest possible dimensionality reduction method is used to create the ROM in this work. The proposed algorithm, however, is agnostic of the dimensionality reduction method and any dimensionality reduction method could be applied along with it. POD is applied to the high-dimensional solutions $\mathbf{Y}$ to generate the low-dimensional latent space for the data. The modes of the high-dimensional space, $\mathbf{\Phi} = [\boldsymbol{\phi}^{(1)}, ..., \boldsymbol{\phi}^{(d)}] \in \mathbb{R}^{m \times d}$, are generated using singular value decomposition (SVD) and the truncation of the modes is performed according to the relative information criterion [19]. $\mathbf{\Phi}$ represents the matrix of POD modes with $d << n$ and $d << m$ as the dimensionality of the latent space coefficients will be much lower than the original dimensionality. The optimal number of modes for the POD algorithm is chosen according to a relative information content of 0.9999.

2.3 Multi-task Gaussian process models

After POD is applied to the high-dimensional solutions, a low-dimensional space is generated for the data. A parametric mapping is created between the input parameters and the low-dimensional space using a MTGP model that can simultaneously predict multiple latent space coefficients. This model is denoted as $MTGP_z$ from here onwards. MTGPs are chosen as they will provide a variance estimate of the low-dimensional space that can be propagated to the high-dimensional solution using the truncated POD modes and Monte Carlo sampling.

MTGPs are multi-task learning models that are applied to Gaussian processes [4]. They aim at learning multiple correlated tasks simultaneously and boost the prediction capability of the model by modelling the correlations among the different outputs that are trying to be predicted. The MTGP models used in this work are created using GPyTorch [10], a Python implementation of Gaussian process models built on the framework of PyTorch [18]. The MTGP models were trained using a radial basis function kernel wrapped with a multi-task kernel and the hyperparameters were optimized using the maximum log-likelihood criterion. These models provide an individual variance estimate of the prediction of each output.

2.4 Uncertainty propagation

The main feature of the proposed adaptive sampling algorithm is the propagation of the variance estimate provided by $MTGP_z$ from the low-dimensional space to the high-dimensional solution space. The uncertainty propagation algorithm used in this work is given in Algorithm 2. One of the most well-known ways of propagating uncertainty is to use Monte Carlo sampling [16]. Monte Carlo sampling uses a large number of samples drawn from the distribution of uncertain variables and uses these samples to estimate the statistics of an output that is dependent on the uncertain variables. In this work, a collection of m samples, $z^{(m)}(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, is drawn from the posterior distribution of the $MTGP_z$ model. Each sample is then projected to the high-dimensional space to obtain

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

Algorithm 2 Propagating variance estimate to high-dimensional space.

Require: point in parameter space $\mathbf{x}$, fitted $MTGP_z$ model, number of Monte-Carlo samples M
  for $m = 1, ..., M$ **do**
    draw sample from $MTGP_z$ posterior of low-dimensional space at point of interest $z^{(m)}(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
    predict high-dimensional solution, $\hat{\mathbf{y}}(\mathbf{x}) = \boldsymbol{\Phi} z^{(m)}(\mathbf{x})^T$
  end for
  estimate $\hat{\sigma}(\mathbf{x})$ from data of Monte Carlo samples

the predicted high-dimensional field $\hat{\mathbf{y}}(\mathbf{x})$. Once the high-dimensional solutions of all the samples have been found, the standard deviation, $\hat{\sigma}(\mathbf{x})$, of the prediction of the ROM at each point of the solution domain will be estimated. Essentially, each point of the solution domain is treated as having a distribution with a mean value and standard deviation. The resulting standard deviation field will be used as an estimate of the uncertainty in the high-dimensional field-solution predicted by the ROM.

2.5 Scalarized sampling objective

The propagated uncertainty will provide a measure of variance or standard deviation for every grid point in the high-dimensional solution. A scalarization or aggregation method is required to enable the use of an optimization algorithm to find the next infill point for the algorithm. In this work, two different scalarization methods are employed. One of them involves using the average of the variance while the other involves using the maximum value of the variance obtained. The average variance is calculated as

$$\hat{\sigma}_{avg}(\mathbf{x}) = \frac{1}{m} \sum_{i=1}^{m} \hat{\sigma}_i(\mathbf{x}), \tag{3}$$

where $\hat{\sigma}_i(\mathbf{x})$ is the standard deviation of the i^{th} grid point in the high-dimensional field and m is the total number of grid points in the field. Similarly, the maximum value of the variance can be determined as

$$\hat{\sigma}_{max}(\mathbf{x}) = \max_{i=1,2,...,m} \hat{\sigma}_i(\mathbf{x}), \tag{4}$$

Once the standard deviation estimate has been calculated, the sampling objective can be defined. To promote the exploration of the parameter space by the adaptive sampling algorithm, the minimum Euclidean distance between sampling points will be included in the sampling objective [1, 13]. A further consideration in the design of the sampling objective is the fact that non-intrusive ROMs are regression models and therefore, the standard deviation may not be zero at an already sampled point as is the case for interpolation models. To prevent placement of new infill points near existing sampling points, a threshold distance

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

is defined below which the standard deviation is artificially set to zero. This will prevent an optimization algorithm from looking in regions close to existing samples while maximizing the sampling objective. After including all of these considerations, the sampling objective for the algorithm is defined as

$$f(\mathbf{x}) = \begin{cases} d(\mathbf{x}, \mathbf{X})\hat{\sigma}_{scalar}(\mathbf{x}), & d(\mathbf{x}, \mathbf{X}) \geq \alpha D_{min}^{max} \\ 0, & \text{otherwise} \end{cases}, \tag{5}$$

where $\hat{\sigma}_{scalar}(\mathbf{x})$ is the scalarized field standard deviation from either 3 or 4, $d(\mathbf{x}, \mathbf{X})$ is the minimum Euclidean distance between the point of interest and existing sampling points and αD_{min}^{max} is the threshold distance. D_{min}^{max} represents the maximum of the minimum Euclidean distance between any two existing sampling points and α is a parameter used to tune the value of the threshold distance. This sampling objective will be maximized to locate the new infill location, $\mathbf{x}_{new}$, for the adaptive sampling algorithm.

3 Numerical experiments

This section discusses the test case used to assess the proposed method and the results obtained.

3.1 2D Environment model function

The analytical environment model function (EMF) [3] will be used to test the proposed adaptive sampling algorithm. EMF aims to model the spill of a pollutant caused by a chemical accident over a spatial and temporal coordinate grid. Since this function is a multi-output function, it can provide a good test bed for understanding the behavior of adaptive sampling algorithms for ROMs without the need to run high-fidelity simulations. ROMs will be trained to predict the outputs of this function on a 10 x 10 grid leading to the requirement of modelling 100 outputs in total. The EMF is defined as

$$h(s, t|\mathbf{x}) = \frac{M}{\sqrt{4\pi Dt}} \exp\left(\frac{-s^2}{4Dt}\right) + \frac{1_{t>\tau} M}{\sqrt{4\pi D(t-\tau)}} \exp\left(\frac{-(s-L)^2}{4D(t-\tau)}\right), \tag{6}$$

where $\mathbf{x} = (M, D, L, \tau)$ are the model parameters, and s and t represent the locations of a grid point in the 10 x 10 grid. The original function has four input parameters, however, in this work, a 2D version of the problem is considered which fixes two parameters of the function. A 2D version is considered as it is easy to study and visualize different aspects of the proposed algorithm in 2D. In this 2D version, the values of L and τ are fixed at their calibrated values of 1.505 and 30.125, respectively [3]. The parameters $M \in [7, 13]$ and $D \in [0.02, 0.12]$ are considered to be the variables in the parameter space.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

3.2 Comparison methods and metric

The results obtained from the adaptive sampling algorithm are compared to Halton sequences [17] and the mean performance of 10 LHS sampling plans [22] at various numbers of samples. The mean relative error in the parameter space is used as a measure of the overall prediction quality of a ROM built by using different types of sampling plans. The mean relative error is calculated as

$$e_{rel} = \frac{1}{N}\sum_{i=1}^{N}\frac{||\mathbf{W}^{pred} - \mathbf{W}^{*}||_2}{||\mathbf{W}^{*}||_2}, \tag{7}$$

where $\mathbf{W}^{pred}$ is the prediction of the field solution from the ROM, $\mathbf{W}^{*}$ is the true field solution and N is the number of testing samples for the ROM.

3.3 Setup of the sampling algorithm

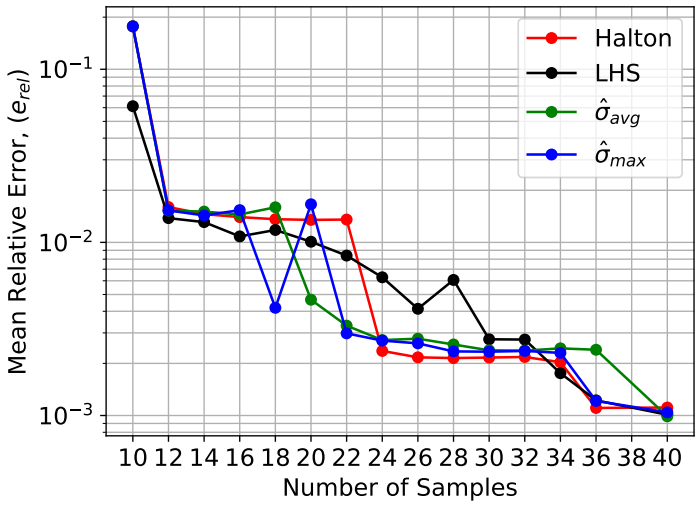
The initial sampling plans, created using Halton sequences, consist of 10 and 20 samples. The testing data to assess the prediction quality of the ROM built from different sampling plans is an LHS sampling plan of 100 samples and common to each of the methods of generating samples. The initial sample sizes were chosen based on heuristics that have been suggested based on the dimensionality of the parameter space and computational budget of the algorithm [15]. These samples are used to train the initial ROM and propagate the uncertainty from the low-dimensional space to the high-dimensional solution space. The Monte Carlo sampling for uncertainty propagation uses 1000 samples drawn from the posterior of the MTGP model to estimate the field standard deviation of the prediction of the pollutant spill multiplied by the minimum Euclidean distance between samples as described in 5. The value of α is set to 0.10 for all numerical experiments. A fixed budget of 40 samples is chosen as the stopping criteria. Both the mean standard deviation and the maximum standard deviation in the field prediction of the pollutant spill were used as the sampling objectives of the problem. The sampling objective is maximized using differential evolution [21] with a population size of 200, a recombination constant of 0.7 and a mutation rate of 0.9.

3.4 Results

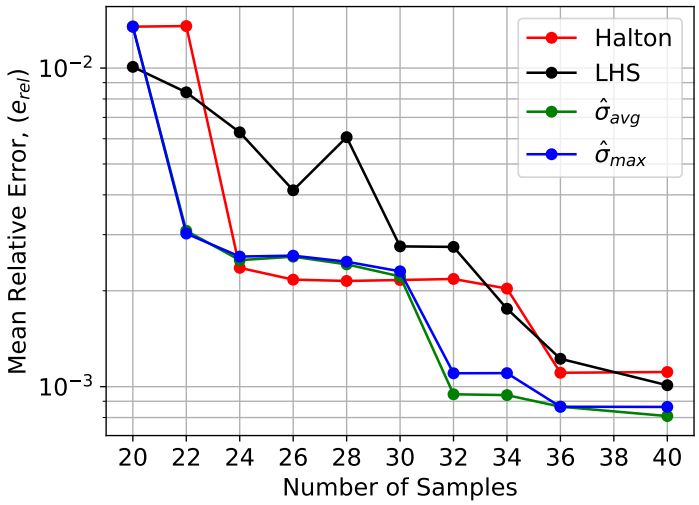
Figure 1 shows the variation of the mean relative error of the prediction of the ROM starting from 10 and 20 initial samples for a fixed budget of 40 samples. It can be seen that the adaptive sampling strategies quickly reduce the prediction error of the ROM as compared to the LHS and Halton sequence sampling plans, especially when using 20 initial samples.

With 10 initial samples, the sampling plans driven by $\hat{\sigma}_{avg}(\mathbf{x})$ and $\hat{\sigma}_{max}(\mathbf{x})$ achieve a final mean relative error of $9.892 \cdot 10^{-4}$ and $1.042 \cdot 10^{-3}$, respectively. This shows that both the sampling strategies achieved a similar performance with 10 initial samples. The LHS sampling and Halton sequence sampling strategy are

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8



(a)



(b)

Fig. 1. Mean relative error of the prediction of the ROM constructed using the proposed sampling methods and benchmark one-shot sampling methods for (a) 10 initial samples and (b) 20 initial samples.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

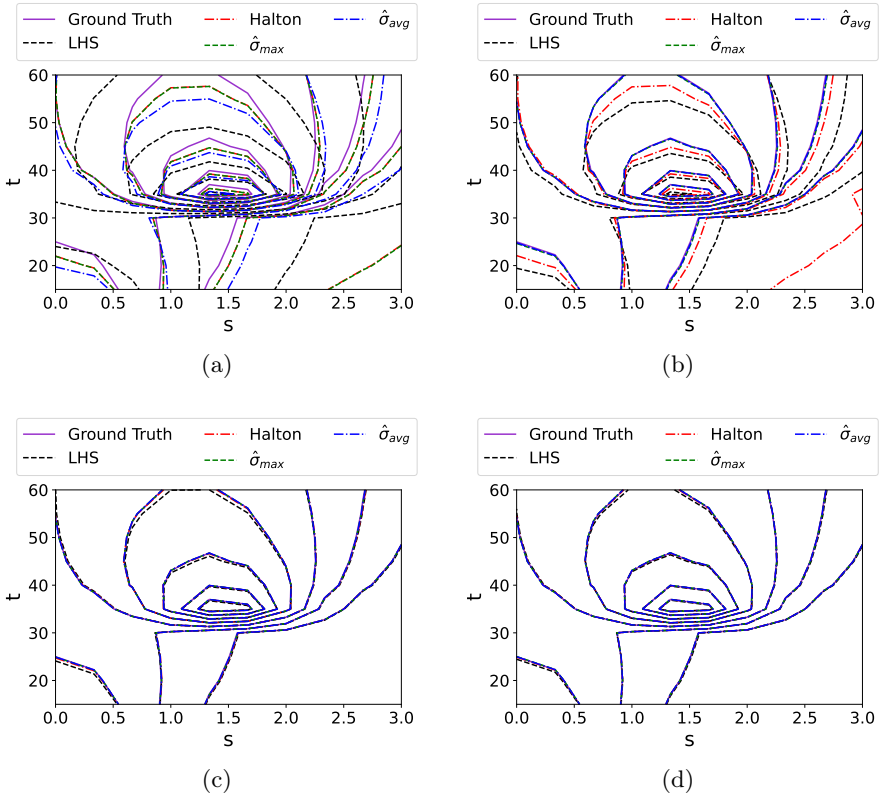


Fig. 2. True and predicted contour lines of EMF for $M = 11.3830$ and $D = 0.0204$ with (a) 10 initial samples, (b) 20 samples, (c) 30 samples and (d) 40 samples.

able to achieve a mean relative error of 1.010×10^{-3} and 1.111×10^{-3}, respectively. This means that in this case the adaptive sampling strategies perform on par with the one-shot sampling strategies.

If the adaptive sampling algorithm is started with an initial sample size of 20, there is an improvement in the prediction accuracy of the ROM generated using the adaptive sampling plan. The $\hat{\sigma}_{avg}(\mathbf{x})$ and $\hat{\sigma}_{max}(\mathbf{x})$ strategies are able to reduce the mean relative error of the ROM to 8.086×10^{-4} and 8.638×10^{-4}, respectively. The sampling plan produced using $\hat{\sigma}_{avg}(\mathbf{x})$ reduces the prediction error of the ROM slightly more than the sampling plan generated using $\hat{\sigma}_{max}(\mathbf{x})$. The adaptive sampling strategies achieve a 20% improvement over LHS and a 27% improvement over Halton sequences in terms of the mean relative error at the computational budget of 40 samples. Both adaptive sampling strategies can also reduce the prediction error of the ROM quicker than the one-shot sampling strategies. The proposed sampling strategies can overtake the one-shot sampling strategies in terms of prediction error by a sample size of 32 samples.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

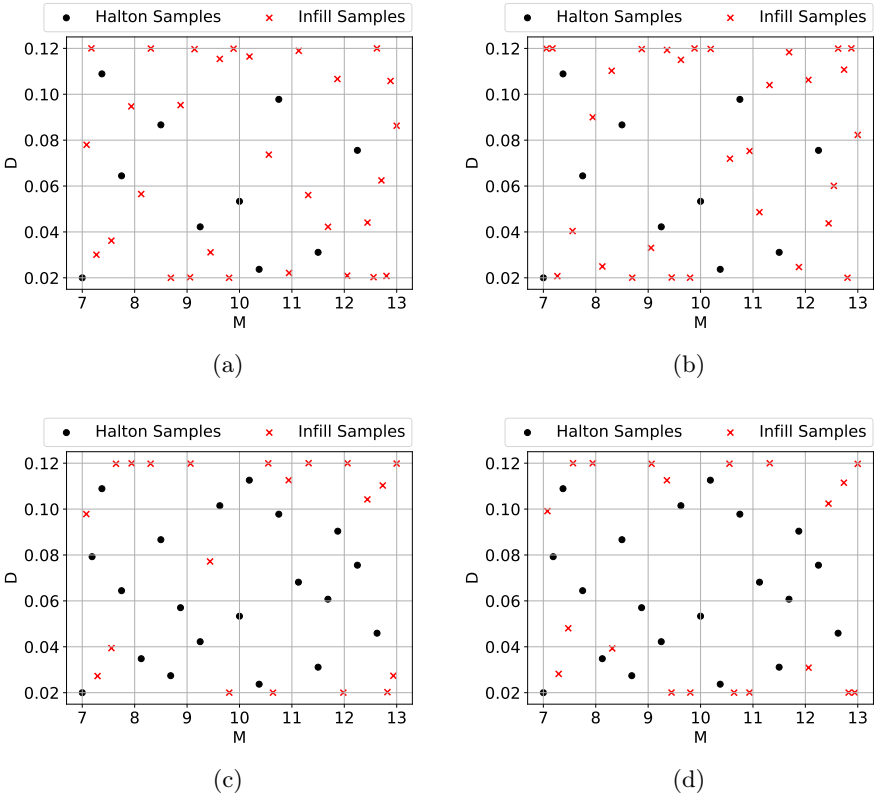


Fig. 3. Sampling plans of the 2D parameter space for the environment model function generated using adaptive sampling strategies using (a) mean propagated variance and 10 initial samples, (b) maximum propagated variance and 10 initial samples, (c) mean propagated variance and 20 initial samples, and (d) maximum propagated variance and 20 initial samples.

The nature of the prediction error of the ROM generated using different sampling strategies is also shown in Fig. 2. The figure shows a comparison between the predicted EMF contour lines from a ROM generated using all of the sampling strategies used in this work and the true EMF contour lines. The figure illustrates that the adaptive sampling strategies can produce a ROM that predicts the EMF more accurately with a fewer number of samples than a ROM generated using one-shot sampling strategies. This is because the contour lines predicted using the adaptive sampling strategies are closer to the true contours as compared to the contour lines predicted using one-shot sampling strategies.

Figure 3 shows the sampling plans that are generated using the adaptive sampling strategies. For comparison, Fig. 4 presents sampling plans of 40 samples that were generated using the benchmark one-shot sampling strategies. It can be

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

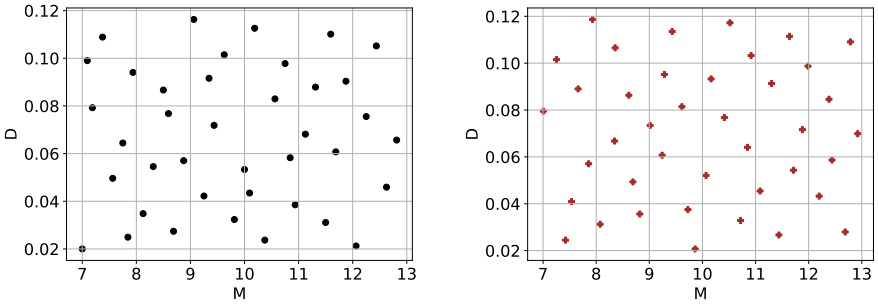


Fig. 4. Benchmark one-shot sampling plans generated using (a) Halton sequences and (b) LHS.

seen that the adaptive sampling strategies place many samples on the edge of the parameter space in addition to the samples placed in the central region. There may be an influence of the distance criteria in placing the samples near the edge of the sample space as this will tend to maximize the minimum distance between samples. This is, however, also a beneficial aspect of the sampling strategy as it tends to explore the edge of the parameter space more than typical one-shot sampling strategies.

The evolution of the contours of $\hat{\sigma}_{avg}(\mathbf{x})$ with increasing number of samples is shown in Fig. 5. The contours show that higher values of $\hat{\sigma}_{avg}(\mathbf{x})$ occur near the edges of the parameter space rather than the central region. This is also an influencing factor in the placement of infill samples near the edge of the parameter space. It is also evident that as the number of samples increases the value of $\hat{\sigma}_{avg}(\mathbf{x})$ decreases across the parameter space. This illustrates that the average field uncertainty of the ROM decreases as the number of samples increases in the sampling algorithm.

4 Conclusion

This paper proposes an adaptive sampling strategy for non-intrusive ROMs based on forward propagation of the multi-task variance information of the latent space. The posterior distribution of the multi-task Gaussian process model of the latent space is sampled using Monte Carlo sampling and the standard deviation of the high-dimensional field solution is calculated. The overall field standard deviation is scalarized using the average value or the maximum value. A sampling objective is designed that maximizes the product of the variance estimate and the minimum distance between the samples. The adaptive sampling process is carried out for a fixed computational budget.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

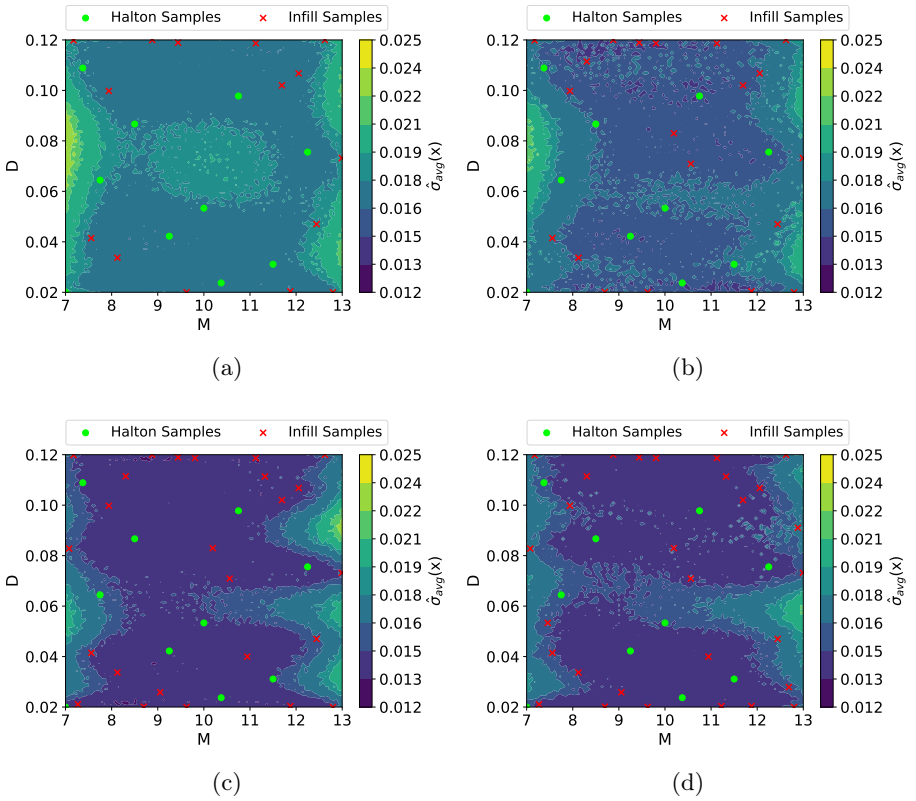


Fig. 5. Contours of $\hat{\sigma}_{avg}(\mathbf{x})$ plotted for (a) 25 samples, (b) 30 samples, (c) 35 samples, and (d) 40 samples starting with 10 initial samples.

The proposed adaptive sampling algorithm is demonstrated on the environment model function. The proposed sampling strategies outperform both of the one-shot strategies especially with an initial sample size of 20 samples. In the case of the EMF, using the average value of the field uncertainty yields slightly lower prediction errors at the termination of the adaptive sampling algorithm as compared to using the maximum value of the field uncertainty. Both adaptive sampling strategies tend to place more samples near the edge of the parameter space as compared to the one-shot sampling strategies. This is due to the Euclidean distance criterion included in the sampling objective and larger field uncertainty in locations near the edges of the parameter space.

Future work will aim at applying the adaptive sampling strategies to higher dimensional and complex engineering problems, such as modeling flow around an airfoil or wing. The sampling objective of the strategy is a very important aspect of the method that will be given more attention in future work. Multiple

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

different scalarization methods and objective formulations will be investigated to improve the performance of the algorithm.

Acknowledgments

This work was supported in part by the U.S. National Science Foundation (NSF) award number 2223732 and by the Icelandic Centre for Research (RANNIS) award number 239858.

References

1. Aute, V., Saleh, K., Abdelaziz, O., Azarm, S., Radermacher, R.: Cross-validation based single response adaptive design of experiments for kriging metamodeling of deterministic computer simulations. Structural and Multidisciplinary Optimization **48**, 581–605 (9 2013). https://doi.org/10.1007/s00158-013-0918-5
2. Benner, P., Gugercin, S., Willcox, K.: A survey of projection-based model reduction methods for parametric dynamical systems. SIAM Review **57**, 483–531 (2015). https://doi.org/10.1137/130932715
3. Bliznyuk, N., Ruppert, D., Shoemaker, C., Regis, R., Wild, S., Mugunthan, P.: Bayesian calibration and uncertainty analysis for computationally expensive models using optimization and radial basis function approximation. Journal of Computational and Graphical Statistics **17**, 270–294 (6 2008). https://doi.org/10.1198/106186008X320681
4. Bonilla, E.V., Chai, K., Williams, C.: Multi-task gaussian process prediction. In: Advances in Neural Information Processing Systems. vol. 20. Vancouver, Canada (3-6 December, 2007)
5. Braconnier, T., Ferrier, M., Jouhaud, J.C., Montagnac, M., Sagaut, P.: Towards an adaptive pod/svd surrogate model for aeronautic design. Computers and Fluids **40**, 195–209 (1 2011). https://doi.org/10.1016/j.compfluid.2010.09.002
6. Decker, K., Iyengar, N., Rajaram, D., Perron, C., Mavris, D.: Manifold alignment-based nonintrusive and nonlinear multifidelity reduced-order modeling. AIAA Journal **61**, 454–474 (1 2023). https://doi.org/10.2514/1.J061720
7. Du, X., Wang, J., Martins, J.R.: A fully automated adaptive sampling strategy for reduced-order modeling of flow fields. In: AIAA SciTech 2023 Forum. National Harbor, MD (23-27 January 2023). https://doi.org/10.2514/6.2023-0534
8. Franz, T., Zimmermann, R., Görtz, S.: Adaptive sampling for nonlinear dimensionality reduction based on manifold learning. Modeling, Simulation and Applications **17**, 255–269 (2017). https://doi.org/10.1007/978-3-319-58786-8_16
9. Fuhg, J.N., Fau, A., Nackenhorst, U.: State-of-the-art and comparative review of adaptive sampling methods for kriging. Archives of Computational Methods in Engineering **28**, 2689–2747 (6 2021). https://doi.org/10.1007/s11831-020-09474-6
10. Gardner, J.R., Pleiss, G., Bindel, D., Weinberger, K.Q., Wilson, A.G.: Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. pp. 7587–7597. Montreal, Canada (3-8 December, 2018)
11. Guénot, M., Lepot, I., Sainvitu, C., Goblet, J., Coelho, R.F.: Adaptive sampling strategies for non-intrusive pod-based surrogates. Engineering Computations (Swansea, Wales) **30**, 521–547 (2013). https://doi.org/10.1108/02644401311329352

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8

12. Halder, R., Fidkowski, K.J., Maki, K.J.: Non-intrusive reduced-order modeling using convolutional autoencoders. International Journal for Numerical Methods in Engineering **123**, 5369–5390 (11 2022). https://doi.org/10.1002/nme.7072
13. Jiang, P., Shu, L., Zhou, Q., Zhou, H., Shao, X., Xu, J.: A novel sequential exploration-exploitation sampling strategy for global metamodeling. IFAC-PapersOnLine **48**, 532–537 (2015). https://doi.org/10.1016/j.ifacol.2015.12.183
14. Karcher, N., Franz, T.: Adaptive sampling strategies for reduced-order modeling. CEAS Aeronautical Journal **13**, 487–502 (4 2022). https://doi.org/10.1007/s13272-022-00574-6
15. Liu, H., Ong, Y.S., Cai, J.: A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. Structural and Multidisciplinary Optimization **57**, 393–416 (1 2018). https://doi.org/10.1007/s00158-017-1739-8
16. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. The Journal of Chemical Physics **21**, 1087–1092 (1953). https://doi.org/10.1063/1.1699114
17. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. SIAM (1992). https://doi.org/10.1137/1.9781611970081.fm
18. Paszke, A., Gross, S., Massa, F., Lerer, A., Google, J.B., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Xamla, A.K., Yang, E., Devito, Z., Nabla, M.R., Tejani, A., Chilamkurthy, S., Ai, Q., Steiner, B., Bai, L.F.J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Vancouver, Canada (8-14 December, 2019)
19. Rajaram, D., Perron, C., Puranik, T.G., Mavris, D.N.: Randomized algorithms for non-intrusive parametric reduced order modeling. AIAA Journal **58**, 5389–5407 (2020). https://doi.org/10.2514/1.J059616
20. Sabater, C., Stürmer, P., Bekemeyer, P.: Fast predictions of aircraft aerodynamics using deep-learning techniques. AIAA Journal **60**, 5249–5261 (9 2022). https://doi.org/10.2514/1.J061234
21. Storn, R., Price, K.: Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization **11**, 341–359 (1997). https://doi.org/https://doi.org/10.1023/A:1008202821328
22. Tang, B.: Orthogonal array-based latin hypercubes. Journal of the American Statistical Association **88**(424), 1392–1397 (1993). https://doi.org/10.1080/01621459.1993.10476423
23. Yang, M., Xiao, Z.: Pod-based surrogate modeling of transitional flows using an adaptive sampling in gaussian process. International Journal of Heat and Fluid Flow **84** (8 2020). https://doi.org/10.1016/j.ijheatfluidflow.2020.108596
24. Yu, J., Yan, C., Guo, M.: Non-intrusive reduced-order modeling for fluid problems: A brief review. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering **233**, 5896–5912 (12 2019). https://doi.org/10.1177/0954410019890721

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_8