Available online at www.eccomasproceedia.org
Eccomas Proceedia UNCECOMP (2021) 246-257

ECCOMAS
Proceedia

UNCECOMP 2021
4th ECCOMAS Thematic Conference on
Uncertainty Quantification in Computational Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Streamed from Athens, Greece, 28 -30 June 2021

A SEQUENTIAL MULTI-POINT SAMPLING PROCEDURE FOR SURROGATE MODELS

Matthias Fischer[1], **Carsten Proppe**[1]

[1]Chair of Engineering Mechanics, Karlsruhe Institute of Technology
Kaiserstr. 10, Bdg. 10.23, 76131 Karlsruhe, Germany
e-mail: {matthias.fischer, proppe}@kit.edu

Abstract. *A sequential sampling procedure is introduced for Gaussian process regression and polynomial chaos expansion. The procedure consists of several sequential sample sets, each with a certain number of sampling points. A grid-based method for sample selection in the context of Gaussian process regression is proposed which aims to improve the model accuracy. The demonstrated methods are investigated for a test case. The obtained surrogate models are validated after each added sample set where the benefit of the proposed sampling methods becomes evident.*

Keywords: Surrogate Models, Sequential Sampling, Parallel Sampling, Gaussian Process Regression, Polynomial Chaos Expansion.

ISSN:2623-3339
Peer-review under responsibility of the organizing committee of UNCECOMP 2021.
doi: 10.7712/120221.8036.18925

1 INTRODUCTION

In many research areas, simulations have become considerably more computationally intensive over time. In the context of surrogate modeling, space-filling designs such as Latin hypercube sampling or Sobol sequences have been applied to ensure good coverage of the input space. Furthermore, sequential sampling methods have been given more focus in the past. In the context of *active learning*, numerous methods have been developed to achieve an optimal experimental design [1]. Various methods aim to add new samples in the input space based on the location of existing samples. For example, distance measures for samples or nested Latin hypercube sampling may be applied [2, 3]. However, these methods do not take the model evaluations for existing samples into account. The goal of sequential sampling methods may thus be extended to generate new samples based on an existing experimental design, model evaluations and a surrogate model such that the quality of the surrogate model can be optimally improved by new samples and model evaluations. Depending on the applied surrogate method, different sampling methods are available or preferable. For instance, active learning for Gaussian process regression is investigated in [4] and [5]. For polynomial chaos expansion, active learning has been frequently applied in structural reliability analysis, e. g. in [6]. In this paper, sampling methods for Gaussian process regression and polynomial chaos expansion are investigated and compared.

Because computation cost has become an increasing factor, the opportunity of parallelization has become more attractive. Parallelized sequential sampling procedures have therefore become a promising possibility [5, 7]. A certain amount of new samples is generated in each cycle of the sampling procedure so that model evaluations for the obtained set of samples can be run in parallel.

For Gaussian process regression, new samples are usually selected at points in the input space where the maximum prediction variance is present [8]. However, this is not necessarily the best choice in order to achieve an optimal model improvement. Furthermore, those points are likely to occur on the boundaries of the input space, especially in high dimensional problems with a small sample size. A new sample selection technique for Gaussian process regression is introduced that is based on expected Gaussian process regression models with respect to new samples.

For polynomial chaos expansion, new samples may be selected based on the information matrix [9]. This matrix is composed of polynomial basis evaluations for each sample, respectively. For example, D-optimal sampling or S-optimal sampling can be applied to compute new samples.

The objective of this paper is to put different sequential sampling methods for surrogate models in a general framework and to include the proposed sample selection technique for Gaussian process regression. In section 2, essentials about applied surrogate models are outlined. In section 3, the sampling procedure and sample selection techniques are described. In section 4, a test case is investigated and discussed. Finally, conclusions with regard to the sampling methods are given in section 5.

2 SURROGATE MODELS

The quantity to be estimated is of the form $f : \mathbb{R}^p \mapsto \mathbb{R}$ which maps inputs $\mathbf{x}$ of dimension p to scalar-valued outputs y. The inputs are assumed to be independent and uniformly distributed on the interval $[0,\, 1]$. If this is not the case, isoprobabilistic transformation, such as the Rosenblatt transformation, can be applied. The goal is to find a set of sequentially added

samples $\mathbf{x}_i$ which yield an optimal surrogate model, i. e. a model with minimum validation error. The surrogate model is built based on a set of samples and corresponding evaluations $\{\mathbf{X}, \mathbf{y}\} : \{(\mathbf{x}_i, y_i), i = 1, \ldots, n\}$. In the following, essentials about the applied surrogate models are given.

2.1 Gaussian process regression (GPR)

Gaussian Process regression, also known as kriging, is a powerful statistical regression technique originating from geostatistics where the input data is treated as a spatial Gaussian process. [10] One crucial advantage of this surrogate method is that, aside from the prediction, it provides the prediction variance as error indicator. The prediction variance can be used in sample selection methods.
The prediction is defined as a weighted sum of observations

$$\mathcal{M}^{\text{GPR}}(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i(\mathbf{x}) y_i \tag{1}$$

where the weights λ_i depend on the position $\mathbf{x}$ in the input space. The weights are calculated by applying two requirements to the prediction. The prediction $\mathcal{M}^{\text{GPR}}(\mathbf{x})$ is assumed to be unbiased with respect to the function f and the variance of the difference between prediction $\mathcal{M}^{\text{GPR}}(\mathbf{x})$ and function f is assumed to be minimum. This leads to the prediction value and prediction variance

$$\mathcal{M}^{\text{GPR}}(\mathbf{x}) = \mu + \mathbf{k}^\top \mathbf{K}^{-1}(\mathbf{y} - \mu\,\mathbf{I}) \tag{2}$$

$$\sigma^2(\mathbf{x}) = \sigma_0^2 - \mathbf{k}^\top \mathbf{K}^{-1}\mathbf{k}. \tag{3}$$

Here $\mu = \frac{\mathbf{I}^\top \mathbf{K}^{-1}\mathbf{y}}{\mathbf{I}^\top \mathbf{K}^{-1}\mathbf{I}}$ is the kriging mean value and $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $k_i = k(\mathbf{x}, \mathbf{x}_i)$, $\sigma_0 = k(\mathbf{x}_i, \mathbf{x}_i)$ where $k(\mathbf{x}_i, \mathbf{x}_j)$ is a valid kernel function. In this paper, the radial basis function

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-(\mathbf{x} - \mathbf{x}')^\top \mathbf{M}(\mathbf{x} - \mathbf{x}')\right) \tag{4}$$

is chosen as anisotropic kernel function. In this expression, the diagonal matrix $\mathbf{M} = \text{diag}(\frac{1}{2l_1^2} \ldots \frac{1}{2l_p^2})$ contains the length scale parameters l_i that will be treated as hyperparameters $\boldsymbol{\theta} = (l_1 \ldots l_p)$. The hyperparameters $\boldsymbol{\theta}$ are either defined based on prior knowledge or determined by maximizing the log marginal likelihood [10]

$$\log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2}\,\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}| - \frac{n}{2}\log 2\pi. \tag{5}$$

2.2 Polynomial chaos expansion (PCE)

A truncated polynomial expansion

$$\mathcal{M}^{\text{PCE}}(\tilde{\mathbf{X}}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \beta_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\tilde{\mathbf{X}}) \tag{6}$$

is considered to approximate the function f where $\tilde{\mathbf{X}} = \{\tilde{X}_1 \ldots \tilde{X}_p\}$ denotes the input variables as independent random variables with marginal probability density functions $\{f_{\tilde{X}_i}(x_i), i = 1 \ldots p\}$. $\beta_{\boldsymbol{\alpha}} \in \mathbb{R}$ are the expansion coefficients and $\psi_{\boldsymbol{\alpha}}(\tilde{\mathbf{X}})$ are the multivariate polynomials with index $\boldsymbol{\alpha}$ that identifies the polynomials in the finite set $\mathcal{A}$. The polynomials $\psi_{\boldsymbol{\alpha}}$ are chosen

such that they are orthogonal with respect to the probability density function of $\tilde{\mathbf{X}}$. Since all random variables $\tilde{X}_i$ are assumed to be uniformly distributed, the Legendre polynomials as the corresponding class of multivariate orthogonal polynomials are used.
The information matrix $\mathbf{A} \in \mathbb{R}^{n \times n_{\boldsymbol{\alpha}}}$ with $A_{ij} = \psi_j(x_i)$ is obtained by evaluating all $n_{\boldsymbol{\alpha}}$ polynomial basis functions for each sample, respectively. This matrix can be used for sample selection techniques as will be demonstrated later.
The focus of this contribution is on the case where computer model f is computationally expensive and therefore a limited number of model evaluations is available. A crucial point of polynomial chaos expansion is that the number of samples must be considerably (about 2 or 3 times [11]) greater than the number of polynomial basis functions. Therefore, a sparse polynomial chaos expansion is favorable. In this paper, least-angle regression [12] is used to find a sparse polynomial basis $\mathcal{A}$ that yields an optimal fit of the sample data while maintaining a limited number of polynomials. The regression coefficients $\boldsymbol{\beta} = \{\beta_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathcal{A}\}$ are calculated by a least squares fit according to

$$\boldsymbol{\beta} = \left(\mathbf{A}^{\top}\mathbf{A}\right)^{-1}\mathbf{A}^{\top}\mathbf{y}. \tag{7}$$

The approach of a sparse expansion is suitable especially for high dimensional problems where full and even truncated designs are problematic regarding the high number of polynomials.

2.3 Model validation

Validation of the surrogate models is conducted by a validation set. The validation set consists of samples from the input parameters distribution and corresponding model evaluations $\{(\mathbf{x}_{\text{val},i}, y_{\text{val},i}), i = 1, \ldots, n_{\text{val}}\}$. The usage of a validation set requires a large amount of additional model evaluations, which is not appropriate for expensive models. However, in this work a validation set is used in order to achieve a more accurate surrogate model assessment. For this purpose, the relative mean square error

$$\varepsilon_{\text{RMSE}} = \frac{1}{\sigma_y^2\, n_{\text{val}}} \sum_{i=1}^{n_{\text{val}}} \left(y_{\text{val},i} - \mathcal{M}(x_{\text{val},i})\right)^2 \tag{8}$$

is calculated, where σ_y^2 is the variance of the model evaluations

$$\sigma_y^2 = \frac{1}{n_{\text{val}}} \sum_{i=1}^{n_{\text{val}}} \left(y_{\text{val},i} - \mu_y\right)^2 \quad \text{with} \quad \mu_y = \frac{1}{n_{\text{val}}} \sum_{i=1}^{n_{\text{val}}} y_{\text{val},i}. \tag{9}$$

When using demonstrated methods for computationally expensive models, cross-validation techniques, such as leave-one-out cross-validation or k-fold cross-validation, may be used instead. So, no additional model runs are necessary. For polynomial chaos expansion the leave-one-out error can be calculated analytically from a single surrogate model [12].

3 SEQUENTIAL SAMPLING

3.1 Sampling procedure

In the beginning, a space-filling sampling method is applied in order to generate a sample set that is used to build the first surrogate model. In this work, a maximin Latin hypercube design

with a small number of samples n_0 is used. The term *maximin* refers to a design that aims to maximize the minimum distance between any two samples in order to improve space-filling properties [13].
Let n_i be the number of desired samples in the i-th sample set. Each sample within a sample set is added one at a time. After each added sample a new surrogate model is generated based on the previous sample set and the new sample $\mathbf{x}_i$. As evaluation $f(\mathbf{x}_i)$, the predicted value from the previous model $\mathcal{M}(\mathbf{x}_i)$ is taken. The goal is to find a new sample $\mathbf{x}_i$ that improves the surrogate model based on certain criteria. The idea behind this procedure is known as *expected improvement* [7]. In this paper, the treatment of model parameters is emphasized, i.e. hyperparameters of Gaussian process regression and the polynomial basis of polynomial chaos expansion.
After each added sample within a sample set, no new information about the true model f is taken into account. It would be unreasonable to update the model parameters then. Therefore, they are assumed to remain unchanged in these steps. In case of Gaussian process regression, the hyperparameters remain unchanged. In case of polynomial chaos expansion, the polynomial basis remains unchanged.
After applying the space-filling design in the beginning and after each sample set, the function f is evaluated at all new n_i samples. A new surrogate model is built based on all available samples $\mathbf{x}_i$ and evaluations f. In these steps, for Gaussian process regression the hyperparameters are updated by maximizing the log marginal likelihood (eq. 5) and for polynomial chaos expansion, the polynomial basis is redefined by least-angle regression. The process is stopped when a defined cross-validation condition is fulfilled or when the cost limit for the number of sample sets m is reached. The sampling procedure is illustrated in Algorithm 1.

3.2 Sample selection technique

Based on the current model $\mathcal{M}_{\text{guess}}$ (see Algorithm 1), a new sample $\mathbf{x}_{\text{new}}$ is selected based on a selection technique that is available for the surrogate method.

3.2.1 Gaussian process regression

A new sample is usually desired at the point in the input space where the maximum prediction variance σ^2 (eq. 3) occurs. In order to find that point, a highly nonlinear optimization problem with usually many local maxima has to be solved. In this work, particle swarm optimization is used as it has shown good performance in such cases. [14]
However, these points do not yield the best model improvement in general. Here, the model improvement is assessed by the prediction variance over the whole input space. Especially in high dimensions, the greatest prediction variance often occurs on the boundary of the input space. The model improvement is therefore limited to one side in relation to the added samples. It is thus likely that other adjusted points yield a lower global prediction variance.
A new grid based selection technique is introduced as follows. A Cartesian grid $\mathbf{X}^G$ with n_G grid points per input dimension is defined on the input space. The total number of grid points is ${n_G}^p$. These grid points are used to determine the prediction variance of the obtained surrogate models. The goal is to find a new sample $\mathbf{x}_0$ in the input space which yields the greatest mean prediction variance reduction of all grid points.
The surrogate model $\mathcal{M}_{\text{guess}}$ is evaluated at point $\mathbf{x}_0$ in the input space that will be determined by optimization. The prediction value at this point $\mathcal{M}_{\text{guess}}(\mathbf{x}_0)$ is used to construct a new surrogate

Algorithm 1 Sequential multi-point sampling procedure for Gaussian process regression and polynomial chaos expansion.

generate space-filling design $\mathbf{X} = \{\mathbf{x}_i, i = 1 \dots n_0\}$
$\mathbf{y} = f(\mathbf{X})$
build surrogate model $\mathcal{M}$ based on $\{\mathbf{X}, \mathbf{y}\}$
  ↳ store hyperparameters $\boldsymbol{\theta}$ (GPR) / polynomial basis $\mathcal{A}$ (PCE)
for *sample set* $i = 1, \dots, m$:
  initialize $\mathbf{X}_{\text{guess}} = \{\,\}$, $\mathbf{y}_{\text{guess}} = \{\,\}$, $\mathcal{M}_{\text{guess}} = \mathcal{M}$
  for *sample point* $j = 1, \dots, n_i$:
    find $\mathbf{x}_{\text{new}}$ based on selection technique for model $\mathcal{M}_{\text{guess}}$
    append $\mathbf{x}_{\text{new}}$ to $\mathbf{X}_{\text{guess}}$
    if $j < n_i$:
      append $\mathcal{M}(\mathbf{x}_{\text{new}})$ to $\mathbf{y}_{\text{guess}}$
      build new surrogate model $\mathcal{M}_{\text{guess}}$ based on $\{(\mathbf{X}, \mathbf{X}_{\text{guess}}), (\mathbf{y}, \mathbf{y}_{\text{guess}})\}$
        ↰ use previous hyperparameters $\boldsymbol{\theta}$ (GPR) / polynomial basis $\mathcal{A}$ (PCE)
    if $j = n_i$:
      append $\mathbf{X}_{\text{guess}}$ to $\mathbf{X}$
      append $f(\mathbf{X}_{\text{guess}})$ to $\mathbf{y}$
      build new surrogate model $\mathcal{M}$ based on $\{\mathbf{X}, \mathbf{y}\}$
        ↳ update hyperparameters $\boldsymbol{\theta}$ (GPR) / polynomial basis $\mathcal{A}$ (PCE)
  validate model $\mathcal{M}$
  break if validation criterion is reached

model $\mathcal{M}_0$. The model quality of $\mathcal{M}_0$ is assessed based on grid $\mathbf{X}^G$. The mean prediction variance σ^2 of $\mathcal{M}_0$ over all grid points $\mathbf{X}^G$ is used as objective function to be minimized with respect to $\mathbf{x}_0$. Again, this is a highly nonlinear optimization problem. Therefore, particle swarm optimization is used to find $\mathbf{x}_0$. Since a new surrogate model $\mathcal{M}_0$ has to be built in each iteration of the optimization, the computation cost is considerably higher compared to the conventional method. However, since hyperparameters $\boldsymbol{\theta}$ remain unchanged in the optimization process, the cost can be clearly reduced. In case of expensive functions f, this effort may still be worthwhile. The method is summarized in Algorithm 2.
If modified importance should be assigned to certain regions of the input space or in case of nonuniform input distributions, a weight function with respect to $\mathbf{x}$ may be multiplied to the prediction variance values for all grid points. Since this is not the case here, i. e. uniform distributions are assumed, this will not be elaborated further.

3.2.2 Polynomial chaos expansion

Sequential sampling methods for polynomial chaos expansion are described in [11]. These methods incorporate information matrix $\mathbf{A}$ (section 2.2) to find new samples according to an optimality criterion. D-optimal sampling aims at maximizing the determinant $D(\mathbf{A}) = \det(\frac{1}{n}\mathbf{A}^\top\mathbf{A})^{\frac{1}{n\alpha}}$. This is related to minimizing the variance of the PCE coefficients $\beta_{\boldsymbol{\alpha}}$ (eq. 6).

Algorithm 2 Grid-based sample selection technique for Gaussian process regression.

Input: model $\mathcal{M}_{\text{guess}}^{\text{GPR}}$, $\{(\mathbf{X}, \mathbf{X}_{\text{guess}}), (\mathbf{y}, \mathbf{y}_{\text{guess}})\}$, hyperparameters $\boldsymbol{\theta}$
generate Cartesian grid $\mathbf{X}^G = \{\mathbf{x}_1^G \dots \mathbf{x}_{n_G{}^p}^G\}$ over input space
with n_G grid points per input dimension (p dimensions)
minimize σ_m^2 for $\mathbf{x}_0 \in [0,1]^p$:
$y_0 = \mathcal{M}_{\text{guess}}^{\text{GPR}}(\mathbf{x}_0)$
build surrogate model $\mathcal{M}_0^{\text{GPR}}$ based on $\{(\mathbf{X}, \mathbf{X}_{\text{guess}}, \mathbf{x}_0), (\mathbf{y}, \mathbf{y}_{\text{guess}}, y_0)\}$
↳ use hyperparameters $\boldsymbol{\theta}$
$\sigma_m^2 = \frac{1}{n_G{}^p} \sum_{i=1}^{n_G{}^p} \sigma^2(\mathbf{x}_i^G)$ (mean prediction variance of $\mathcal{M}_0$)
return σ_m^2
Output: $\mathbf{x}_0$ for minimum σ_m^2

In this work, S-optimal sampling is used which aims at maximizing the S-value

$$S(\mathbf{A}) = \left(\frac{\sqrt{\det(\mathbf{A}^\top \mathbf{A})}}{\prod_{i=1}^{n_{\boldsymbol{\alpha}}} \|A_i\|_2} \right)^{\frac{1}{n_{\boldsymbol{\alpha}}}}. \tag{10}$$

While maximizing the determinant, the S-value additionally aims at maximizing the column orthogonality of $\mathbf{A}$. Here, A_i denotes the i-th column of information matrix $\mathbf{A}$. As denoted in Algorithm 1, polynomial basis $\mathcal{A}$ remains unchanged between added samples within one sample set. Therefore, the column size of $\mathbf{A}$ does not change, but one row is appended with each added sample containing corresponding polynomial basis evaluations. Hence, least-angle regression is not conducted in these steps and computation cost is relatively small.
As comparative study, new samples are selected based on a conventional distance-based measure. Here, a new sample

$$\mathbf{x}_{\text{new}} = \underset{\mathbf{x}_{\text{new}} \in [0,1]^p}{\operatorname{argmax}} \left(\min_{i \in \{1 \dots n\}} \|\mathbf{x}_i - \mathbf{x}_{\text{new}}\|_2 \right) \tag{11}$$

is added at the point in the input space that maximizes the minimum euclidean distance to existing points $\{\mathbf{x}_1 \dots \mathbf{x}_n\}$ in the design. This procedure is known as *farthest point strategy* [2].

4 TEST CASE

A modified version of the Ishigami function

$$f(\mathbf{x}) = \sin(x_1) + 3\sin^2(x_2) + 2x_2^4 \sin(x_1) \tag{12}$$

is chosen to investigate the sampling methods according to Algorithm 1. The modification is done in order to obtain a two-dimensional function so that sampling methods can be visualized graphically.
First, a maximin Latin hypercube design with $n_0 = 10$ samples is generated. Then, five sample sets are added, each with two samples ($n_i = 2, i = 1 \dots 5$). In case of Gaussian process regression, initially, the method of selecting samples at maximum prediction variance is applied. In another experiment, the proposed grid-based method according to Algorithm 2 is applied.

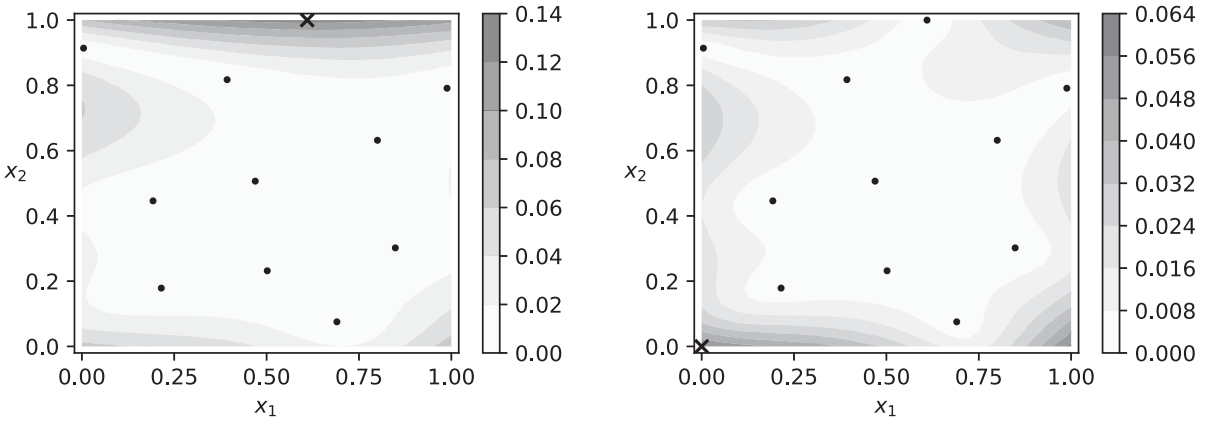

Figure 1: Gaussian process regression: prediction variance σ^2 (eq. 3) with respect to inputs x_1 and x_2 before adding the first sample (left) and the second sample (right) of the first sample set. Existing samples are marked as dots, added samples are marked as crosses.

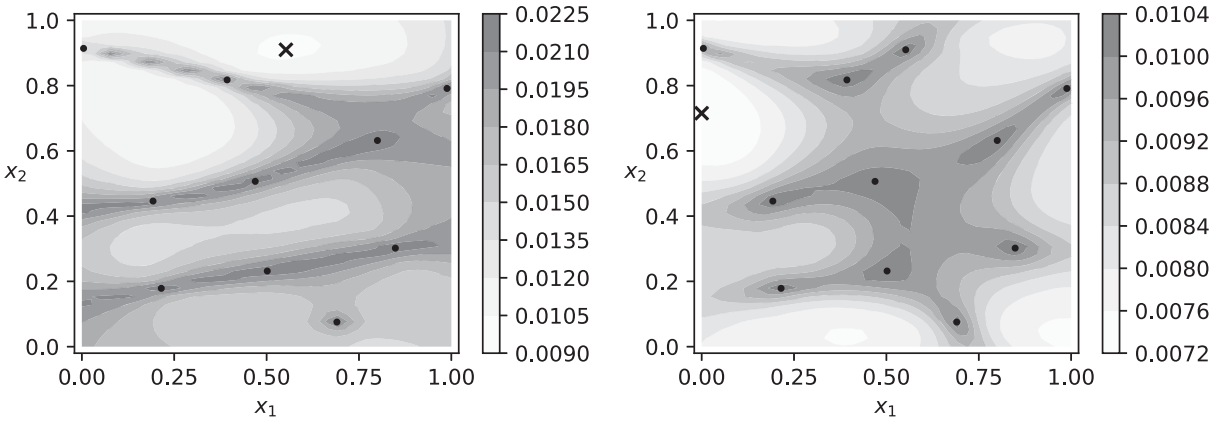

Figure 2: Gaussian process regression: mean prediction variance σ_m^2 (Algorithm 2) from the grid-based method with respect to inputs x_1 and x_2 before adding the first sample (left) and the second sample (right) of the first sample set. Existing samples are marked as dots, added samples are marked as crosses.

In case of polynomial chaos expansion, the distance-based criterion (eq. 11) and S-optimal sampling (eq. 10) are applied to select new samples. In all cases, particle swarm optimization is used to find such points in the input space. The obtained models $\mathcal{M}$ are validated after each sample set.

For comparison, new samples are generated by standard Monte Carlo sampling using the same number of samples per set $n_i = 2, i = 1 \dots 5$.
In Figures 1, 2, 3 and 4 the locations of $n_1 = 2$ added samples in the first sample set are shown for Gaussian process regression and polynomial chaos expansion for chosen sampling methods, respectively. The same space-filling design is used to allow for better comparison. In Figure 1, the contour plot indicates prediction variance σ^2 of model $\mathcal{M}_{\text{guess}}^{\text{GPR}}$ that is used for sample selection. Points with maximum prediction variance after each step are selected as new samples. In Figure 2, the contour plot shows the expected mean prediction variance σ_m^2 (see

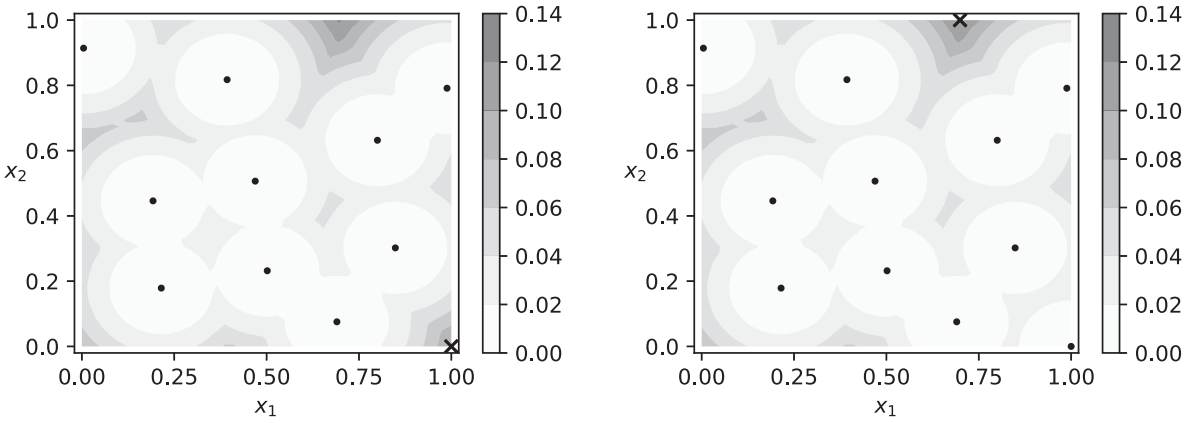

Figure 3: Polynomial chaos expansion: distance-based criterion (eq. 11) for new samples with respect to inputs x_1 and x_2 before adding the first sample (left) and the second sample (right) of the first sample set. Existing samples are marked as dots, added samples are marked as crosses.

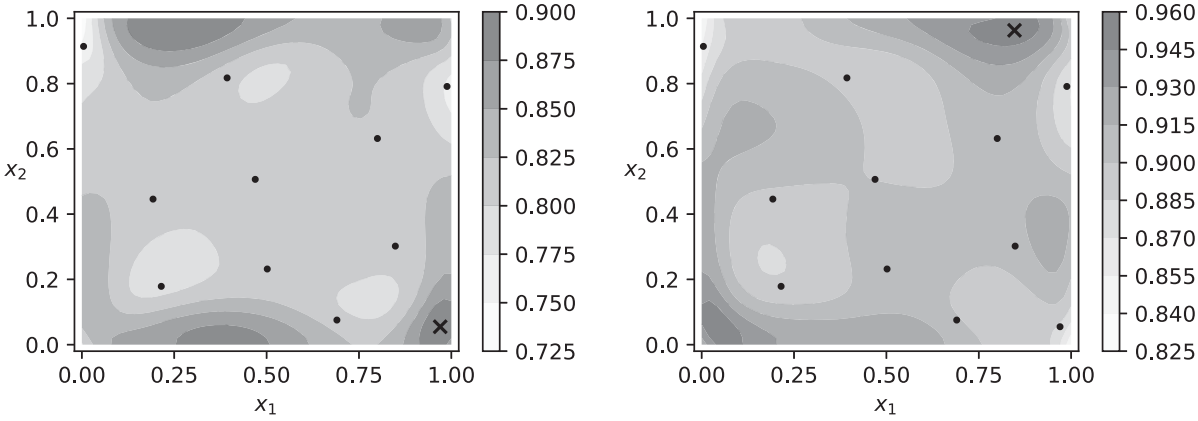

Figure 4: Polynomial chaos expansion: S-value (eq. 10) with respect to inputs x_1 and x_2 before adding the first sample (left) and the second sample (right) of the first sample set. Existing samples are marked as dots, added samples are marked as crosses.

Algorithm 2) that would be expected to result from a new sample at respective points in the input space. Points with the lowest values of σ_m^2 are selected as new samples. In Figure 3, the contour plot indicates the euclidean distance of points in the input space to the closest existing sample. New samples are chosen that maximize this distance. In Figure 4, the contour plot indicates the S-value (eq. 10) that is maximized to select new samples.
In all demonstrated cases, the selection of new samples shows to ensure good space-filling properties. For Gaussian process regression, it can be recognized that new samples are less likely to occur on the boundary, if the proposed grid-based selection method is applied.
In Figure 5 the validation error $\varepsilon_{\text{RMSE}}$ (eq. 8) is illustrated after each sample set for all investigated sampling methods. The methods are compared to the case where new samples are selected according to standard Monte Carlo sampling. It is distinct that sample selection methods (Algorithm 1) are superior compared to Monte Carlo sampling. In the considered test case, Gaussian

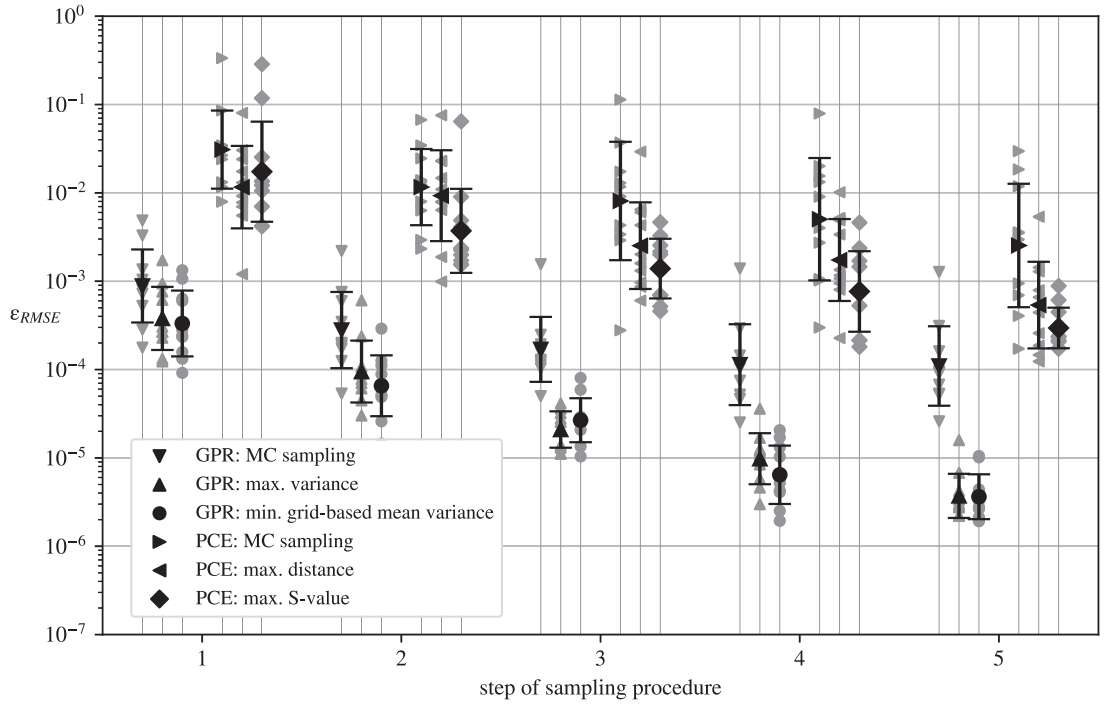

Figure 5: Validation error $\varepsilon_{\mathrm{RMSE}}$ for different sampling methods (vertical lines) after each of five sample sets (horizontal axis). Grey markers show the results for single experiments beginning with new space-filling designs, while black markers show their mean values with standard deviations.

process regression yields better results than polynomial chaos expansion. This is due to the fact that the number of polynomial basis functions is very limited regarding the small number of samples in order to avoid over-fitting. In this work, the maximum number of basis functions for least-angle regression is updated after each sample set to half of the current sample size. In addition, the maximum polynomial degree is increased during the sampling procedure.

For Gaussian process regression, the proposed grid-based method yields only vaguely smaller validation errors compared to the conventional sample selection method for Gaussian process regression. However, based on the sample selection technique it is presumed that the global prediction variance reduction over the whole input space is improved through the proposed grid-based method.

For polynomial chaos expansion, the S-optimality criterion yields slightly smaller errors compared to the distance-based measure. However, it is emphasized, that the polynomial basis is updated after each sample set. Thus, the optimality criterion may change in such a way that the chosen samples are not optimal anymore with regard to the new basis. The distance-based measure has shown to be more robust in very sparse designs (i. e. less than ten samples) and thus for a small number of basis functions.

5 CONCLUSIONS

Studies on a simple example have shown that proposed multi-point sequential sampling methods are very promising compared to standard Monte Carlo sampling and should be taken

into consideration, especially if computationally intensive models are investigated and the necessary sample size for the desired surrogate model quality is not known in advance. For sparse designs such as in the considered test case, Gaussian process regression appears to be a superior regression tool compared to polynomial chaos expansion. This is due to great limitations of polynomial chaos expansion regarding the number and order of basis functions for sparse designs.
Even though no significant improvement was obtained by using the proposed grid-based sampling technique, it may be worthy to further investigate this method if surrogate model quality demands are high. The additional computational effort may still be small if computationally expensive models are investigated. However, it becomes apparent that the conventional sample selection method for Gaussian process regression, namely to search for points in the input space with maximum prediction variance, is a sufficient and computationally affordable tool in general. Instead of using a fixed grid, other possibilities may be reasonable. For example, a Latin hypercube design may be used as grid to introduce randomness. The Latin hypercube design may then be randomly updated after each added sample, so that biases originating from fixed grid points may be prevented.
The distance-based measure has shown to be a robust and successful tool that can be applied for other surrogate models as it only considers the input space. If further improvement is desired for polynomial chaos expansion, optimality criteria such as S-optimality may be applied.

REFERENCES

[1] B. Settles. Active learning literature survey. Computer sciences technical report 1648. *University of Wisconsin-Madison*, 2009.

[2] Y. Eldar, M. Lindenbaum, M. Porat, Y. Y. Zeevi. The farthest point strategy for progressive image sampling. *IEEE Transactions on Image Processing*, **6**, 1305–1315, 1997.

[3] P. Z. G. Qian, Nested Latin hypercube designs, *Biometrika*, **96**, 957–970, 2009.

[4] S. Seo, M. Wallat, T. Graepel, K. Obermayer. Gaussian process regression: active data selection and test point rejection. *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, **3**, 241–246, 2000.

[5] D. Ginsbourger, R. L. Riche, L. Carraro. A Multi-points Criterion for Deterministic Parallel Global Optimization based on Gaussian Processes. hal-00260579, 2008.

[6] S. Marelli, B. Sudret. An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis. *Structural Safety*, **75**, 67–74, 2018.

[7] R.T. Haftka, D. Villanueva, A. Chaudhuri. Parallel surrogate-assisted global optimization with expensive functions – a survey. *Structural and Multidisciplinary Optimization*, **54**, 3–13, 2016.

[8] A. J. Booker, J. E. Dennis, P. D. Frank et al. A rigorous framework for optimization of expensive functions by surrogates. *Structural Optimization* **17**, 1–13, 1999.

[9] N. Lüthen, S. Marelli, B. Sudret. Sparse polynomial chaos expansions: Literature survey and benchmark. arXiv preprint arXiv:2002.01290, 2020.

[10] C. E. Rasmussen, C. K. I. Williams. Gaussian Processes for Machine Learning. The MIT Press. 2005.

[11] G. Blatman, Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis, Université Blaise Pascal, Clermont-Ferrand, France, 2009.

[12] G. Blatman, B. Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression, *Journal of Computational Physics*, **230**, 2345–2367, 2011.

[13] M.E. Johnson, L.M. Moore, D. Ylvisaker. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, **26**, 131–148, 1990.

[14] J. Kennedy, R. Eberhart. Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, **4**, 1942–1948, 1995.