EMAS Engineering Modelling, Analysis & Simulation THE NAFEMS JOURNAL

NAFEMS

Vol. 3, Issue 1, 2026

# Impact of Sampling Strategy on the Accuracy of Surrogate Models for Structural Analysis

Manish Barlingay[a]

[a]Goengineer, 739 East Fort Union Blvd., Midvale, UT 84047, USA

## Abstract

In modelling and simulation, approximations also known as metamodels, are commonly employed to approximate outcomes of complex simulations that are computationally intensive to evaluate. Various approximation techniques like response surface models, Kriging, radial basis functions, neural networks, etc. are utilized for surrogate modelling. These methods inherently rely on using a predefined dataset that forms the basis of training, validation and predictive capabilities of these surrogate models. Thus, the accuracy of these surrogate models is directly related to the quality of the underlying dataset used. Design of Experiments (DOE) methods are typically used to generate these datasets. The quality of the datasets is influenced by the choice of the DOE methods that are employed, as each DOE method uses a different strategy to sample the design space. Various sampling strategies like full factorial, fractional factorial, Latin hypercube sampling (LHS), Sobol sampling, etc. are available.

This study examines the impact of three sampling techniques- full factorial, fractional factorial, and Sobol sequencing on the accuracy and efficiency of surrogate models created using radial basis functions (RBF) for a structural analysis use-case.

## Keywords

Surrogate modeling, machine learning, DOE, approximations, sampling



# 1 Introduction

Surrogate models are often used in computational analysis to predict outcomes of computationally intensive problems. Surrogate models are used to map the design variables (inputs) to response variables (outputs) when the actual relationship between the two is unknown, not well understood, or computationally expensive to evaluate [1]. Full vehicle crash simulation is one such example. As the simulation itself is computationally very expensive; surrogate modeling provides an elegant and a much quicker approach to approximate simulation outcomes in such cases. The speed and accuracy of the surrogate models depend on several different factors such as the choice of approximation technique, sampling technique, sample size, etc.

DOE sampling techniques are available to be employed for generating the datasets that form the basis of these surrogate models with each technique having its own strengths and weaknesses. Some of the common techniques used are Latin Hypercube, Full factorial, Fractional Factorial, Random Sampling, Sobol Sequencing, etc. [2] provides a state-of-the-art review on different approximation and DOE techniques available.

While it seems logical that a sampling technique that explores the entire design space like the full factorial technique would provide better accuracy, but it would prove to be computationally very expensive especially for complex simulations. So, it is important to compare different sampling techniques to determine which method provides a good balance between accuracy and computational

[1]Corresponding author.
E-mail address: mbarlingay@goengineer.com (M. Barlingay)
https://doi.org/10.59972/1rf88gpv

expense. Three commonly used techniques are studied and compared in this paper and the following section provides a brief introduction of each of these three sampling techniques.

## 1.1 Full Factorial

In a full-factorial design, all combinations of all factors at all levels are evaluated. Typically, the standard engineering practice is to systematically evaluate a grid of points requiring n1×n2×n3×…ni (i = # factors, ni = # levels for factor i) design point evaluations. This practice provides extensive information for accurate estimation of factor and interaction effects. However, it is often deemed cost-prohibitive because of the number of analyses required [3]

## 1.2 Fractional Factorial

A fractional factorial experiment is a certain fractional subset (1/2, 1/4, 1/8, etc. for two-level factors and 1/3, 1/9, 1/27, etc. for three-level factors) of the full factorial experiment that is carefully selected to minimize aberrations in the experiment. Fractional factorial designs are available only when all factors have either two or three levels. Fractional factorial experiments are also useful when some factors are independent of each other or when certain interactions can be neglected [3].

## 1.3 Sobol Sequences

The Sobol sequence provides a space-filling collection of points that are highly uniform in their spacing, as defined by measures of discrepancy. The uniformity of this technique generally improves on that of the Latin Hypercube technique, with similar cost in generation time [3].

Figure 1 shows a comparison of the above three sampling techniques for 2 factors with $factor_1$ sampled at 3 levels and $factor_2$ sampled at 4 levels.

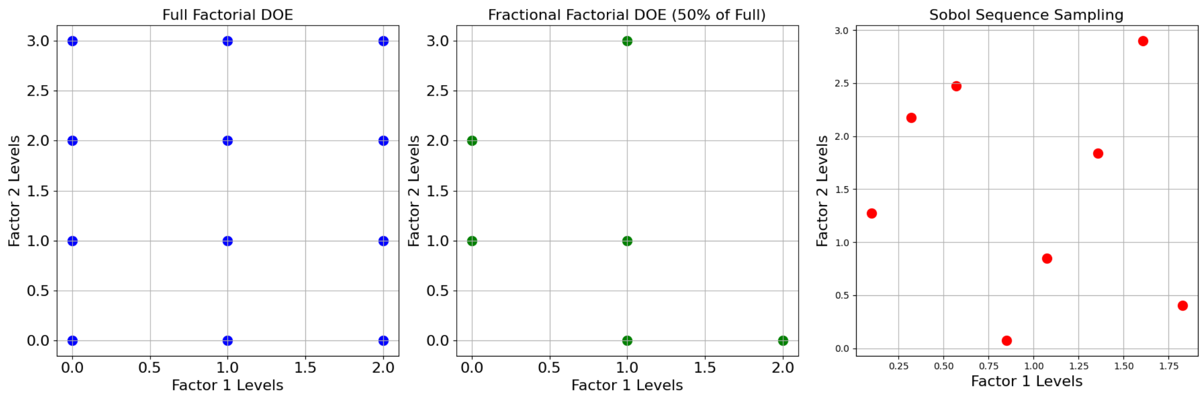


Figure 1. Design space for DOE sampled with different techniques.

# 2 Materials and Methods

The accuracy of surrogate models improves with sampling methods that provide better space coverage of the design space. This means that the accuracy will generally improve with increase in number of sampling points [4]. However, the number of sampling points considered is constrained by the maximum number of simulations that can be feasibly performed, given the available time and computational resources. So, the need for more sampling points and available computational resources need to be carefully balanced [5]. This balance needs to ensure that the surrogate model is both comprehensive enough to capture key system behaviors and efficient to avoid exceeding practical limits of the compute resources.

To understand the impact of different sampling strategies and sampling sizes on the accuracy of the surrogate models, a simple I-beam example as shown in Figure 2. is considered. A baseline finite element analysis (FEA) of an I-Beam is setup using Dassault Systemes' 3DExperience software. A steel I-Beam (3000 mm) in length is fixed at one end and acted upon by a pressure load of 0.01 MPa on the top face as shown in Figure 3. Table 1-3 show the details of the Finite Element Model.

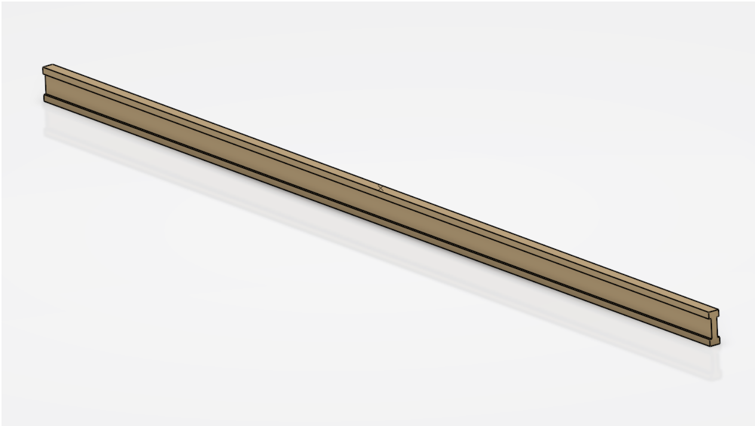

Figure 2. Cantilever I-Beam.

Table 1. Details of the Finite Element Model.

| Parameter | Value |
|---|---|
| Element Type | C3D10HS* |
| Number of Nodes | 70204 |
| Number of Elements | 37679 |
| Material Model | Isotropic Elasticity |

*10 node general purpose quadratic tetrahedron elements with improved surface stress visualization.

Table 2. Material Properties.

| Parameter | Value |
|---|---|
| Density | 7.8e-9 tonne/mm3 |
| Young’s modulus | 2e5 MPa |
| Poisson’s ratio | 0.27 |

Table 3. Mesh Quality Statistics

| Parameter | Value |
|---|---|
| Aspect Ratio | 1.696 |
| Nodes Jacobian | 0.947 |
| Maximal Angle | 97.964 |
| Minimal Angle | 45.306 |
| Modified Jacobian | 0.219 |
| Skewness | 0.622 |
| Stretch | 0.665 |

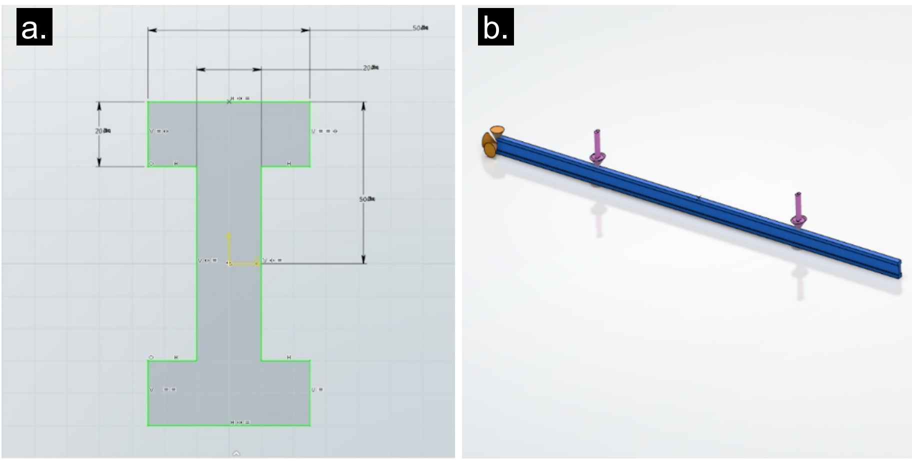


Figure 3. FEA Setup of I-Beam a. I-section dimensions and b. Boundary conditions.

Vertical beam deflection and Von Mises stress at the midspan ($x = 1500\ mm$) are monitored as the main responses. Figure 4 shows the outputs for the baseline simulation. The baseline simulation results are validated by comparing them to the analytical solution as discussed in Section 3.

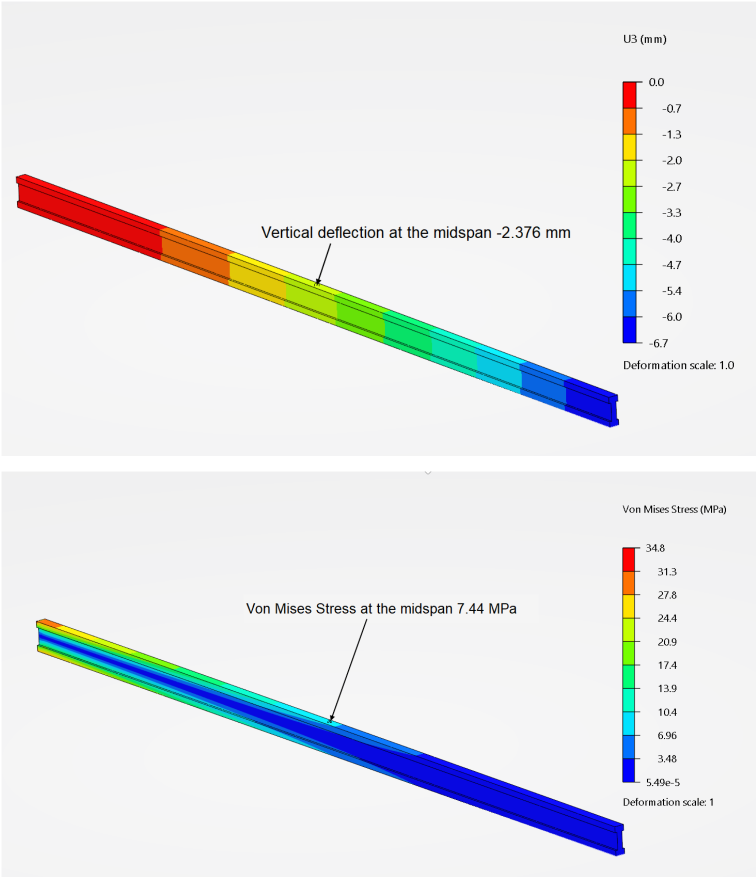


Figure 4. Vertical Deflection and Von Mises Stress at the midspan for baseline simulation.

The baseline simulation is parametrized by creating geometric parameters for the I-Beam cross section and these geometric parameters are used as the DOE inputs. Simulation result sensors are created to monitor key simulation outputs which are used as the DOE responses. Using the parametrized baseline model, multiple DOE studies are developed. Figure 5 shows a schematic for a DOE running multiple I-Beam simulation experiments. Each DOE uses a different sampling technique. Geometric parameters of the beam are iteratively varied, and corresponding outputs are monitored. For an effective comparison of the three sampling techniques, it is essential that all three algorithms sample the same design space. This can be achieved by constraining the design space using appropriate bounds on the input parameters. Each parameter is varied within these bounds of the design space.

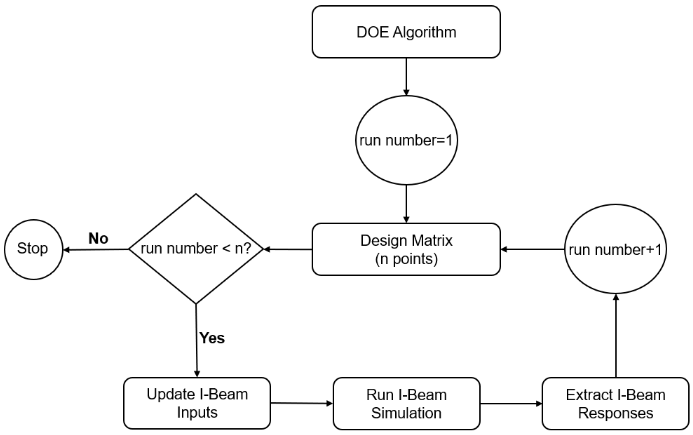


Figure 5. Schematic of DOE running iterative I-Beam simulations.

Table 4 summarizes the input parameters with their corresponding bounds and the output parameters for the DOE.

Table 4. DOE Parameter table.

| Parameter | Type | Lower Bound (mm) | Upper Bound (mm) |
|---|---|---|---|
| Beam_Height | Input | 70 | 100 |
| Beam_Width | Input | 35 | 50 |
| Flange_Thickness | Input | 14 | 20 |
| Web_Thickness | Input | 14 | 20 |
| U3 | Output | - | - |
| VM | Output | - | - |

Three DOEs are run with each DOE using four inputs and two responses and a different sampling technique. The first DOE is run using a full factorial technique in which the four inputs are evaluated at three levels. This creates a total of $3^4 = 81$ simulations. Although this technique samples the design space thoroughly, it is computationally very intensive. Table 5 shows the DOE configuration using full factorial sampling.

Table 5. DOE configuration using full factorial sampling.

| Parameter | Lower Bound (mm) | Upper Bound (mm) | Levels (mm) |
|---|---|---|---|
| Beam_Height | 70 | 100 | 70, 85, 100 |
| Beam_Width | 35 | 50 | 35,42.5, 50 |
| Flange_Thickness | 14 | 20 | 14,17, 20 |
| Web_Thickness | 14 | 20 | 14,17,20 |

The second DOE is run using a fractional factorial technique in which the four inputs are evaluated at three levels each. This creates a total of $\frac{1}{3}(3^4) = 27$ simulations. The design space bounds are maintained so that the same design space is sampled with 27 experiment points. Table 6 shows the DOE configuration using fractional factorial sampling.

Table 6. DOE configuration using Fractional Factorial sampling,

| Parameter | Lower Bound (mm) | Upper Bound (mm) | Levels (mm) |
|---|---|---|---|
| Beam_Height | 70 | 100 | 70, 85, 100 |
| Beam_Width | 35 | 50 | 35,42.5, 50 |
| Flange_Thickness | 14 | 20 | 14,17, 20 |
| Web_Thickness | 14 | 20 | 14,17, 20 |

A third DOE is run using a space filling Sobol sequence technique with the 4 inputs varied within the same bounds of the design space such that it creates a total of 27 simulations; exactly equal to the number of simulations run using the fractional factorial technique. Sobol technique provides a sequence of factor levels based on a deterministic, quasi-random algorithm [3]. In this paper, it is forced to generate 27 points to ensure a direct comparison of its performance with the fractional factorial technique.

Sobol sampling technique needs a user to define the number of corner points and interior points. Interior Points fall within the bounds of the design space whereas corner points lie on corners of the design space. This provides a uniform distribution of points within the design space and also considers the boundaries of the design space. The effect of changing the distribution of corner points and interior points is out of the scope of this paper, but it seems logical to assume that this distribution will affect the efficiency of sampling the design space. Table 7 shows the DOE configuration using Sobol sampling.

Table 7. DOE configuration using Sobol Sequencing sampling.

| Parameter | Lower Bound (mm) | Upper Bound (mm) |
|---|---|---|
| Beam_Height | 70 | 100 |
| Beam_Width | 35 | 50 |
| Flange_Thickness | 14 | 20 |
| Web_Thickness | 14 | 20 |
| Number of Corner points | 8 | |
| Number of Interior points | 19 | |

Each DOE generates a dataset with beam height, beam width, flange thickness and web thickness as inputs and corresponding deflection and Von Mises stress at the midspan as outputs. Each row in this dataset is called a data point or an experiment point.

Table 8 summarizes the normalized computational times for data generation for all three DOEs. It can be clearly seen that data generation for full factorial DOE is at least three times more expensive as compared to fractional factorial or Sobol sequence techniques.

Table 8. Comparison of compute times for data generation for all three DOE techniques

| DOE Technique | Normalized computational time |
|---|---|
| Full Factorial | 1 |
| Fractional Factorial | 0.33 |
| Sobol Sequence | 0.33 |

Each of the three datasets is then used to create a surrogate model through training and cross validation. Radial Basis Function Neural Network (RBF-NN) is used for surrogate modeling. RBF-NN is a feed-forward neural network consisting of three layers: one input layer, one hidden layer of radial units, and one output layer of linear units. As RBF-NNs use only three layers, they are reasonably compact and are characterized by reasonably fast training [3]. A detailed explanation of RBF-NN is provided in [6]. Cross validation and error measures for the RBF-NN surrogate model are discussed in Section 3.

# 3 Verification and Validation

The baseline simulation is verified by comparing it to the analytical solution. For this, the vertical deflection and Von Mises stress from the baseline simulation are plotted as a function of distance from the fixed end of the cantilever beam. These quantities are then compared to the corresponding analytical values given by the following equations [7]:

$$\delta_x = \frac{Pbx^2}{24EI}(x^2 + 6L^2 - 4Lx), \tag{1}$$

$$\sigma_x = \frac{Pbh(L-x)^2}{4I}, \tag{2}$$

where, $\delta_x$ is the vertical deflection of the beam at a distance $x$ from the fixed end, $\sigma_x$ is the bending stress at a distance $x$ from the fixed end, $P$ is the pressure acting on the beam, $b$ is the width of the cantilever beam, $L$ is the length of the cantilever beam, $h$ is the beam height, $E$ is the Young's modulus, and $I$ is the moment of Inertia of the beam cross section.

Figure 6 shows that the baseline simulation predicts the response of the cantilever beam accurately except for the stress response at the fixed end. This is due to the constrained boundary conditions at the fixed end. This outlier result value at the fixed end does not affect the validity of the DOEs or the surrogate models. This is because the DOE and the surrogate models use only the deflection and stresses at the midspan of the cantilever beam. As this simulation is considered as the baseline for DOEs, no further validation is performed.

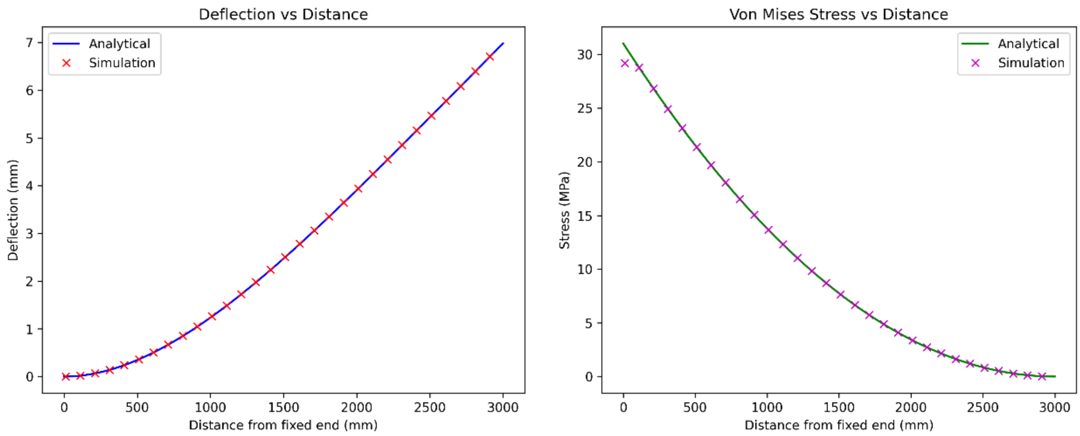


Figure 6. Comparison of baseline simulation results with analytical results.

## 3.1 Cross-validation of RBF-NN surrogate models

For validating the accuracy of the RBF-NN surrogate model, the Leave One Out Cross Validation (LOO-CV) is used. LOO-CV is an exhaustive cross-validation method in which the original sample is divided into a training and a validation set. The user specifies the size of the validation set. With each iteration, one point out of the validation set is left out and the remaining points are added back to the training set. Training is run using all, but the one left out point which is used for validation after the training. This is repeated until all the data points in the validation set have been validated [8]. In this specific paper, all the experiment points are used for cross-validation which means full factorial DOE uses 81 points and fractional factorial and Sobol each use 27 points for cross-validation. Error metrics are computed and compared for each machine learning model. The following error metrics as explained in [9] are used.

### 3.1.1 Average Error (Normalized)

For every point in the validation set, the differences between the actual and predicted values are averaged and then normalized by the range of the actual values for each response.

$$ {}_{Avg}^{Norm}Error = \frac{Avg[ABS(Actual - Predicted)]}{Actual_{max} - Actual_{min}} \tag{3} $$

## 3.2 $R^2$ value

The coefficient of determination is calculated based on the error samples. The coefficient of determination always ranges between 0 and 1, where 1 represents a perfect fit (or no prediction error):

$$ R^2 = 1 - \frac{SSR}{SSTOTAL} \tag{4} $$

where

$$ SSR = \sum_{k=1}^{n} (Actual - Predicted)^2, \tag{5} $$

and

$$ SSTOTAL = \sum_{k=1}^{n} (Actual - AVG(Actual))^2, \tag{6} $$

where n is the number of points, SSR is the sum of the squared residues, and SSTOTAL is the total sum of squares.

Table 9 summarizes the normalized computational times for training and cross-validation of all three RBF-NN surrogate models. It can be clearly seen that compute times for training and cross validation for full factorial DOE is at least twelve times more expensive as compared to fractional factorial or Sobol sequence techniques.

Table 9. Normalized computational times for training and cross-validation.

| RBF-NN model sampling | Normalized computational time |
|---|---|
| Full Factorial | 1 |
| Fractional Factorial | 0.0833 |
| Sobol Sequence | 0.0833 |

Figures 7, 8 and 9 show the validation error plots for vertical deflection and Von Mises stress for the surrogate models trained using data from full factorial, fractional factorial and Sobol sequencing sampling techniques respectively.

Table 10 summarizes the validation errors in terms of R2 fits of all three surrogate models for both the responses. It is evident that the model for dataset sampled with full factorial method gives the best fit in terms of R2 as well as a very low error. While both the models for datasets sampled using fractional factorial technique and Sobol sequencing technique provide a very good fit in terms of R2 values, the model for dataset sampled using Sobol sequencing technique provides better accuracy than that sampled using the fractional factorial technique with a mean error of 0.0204 for the Von Mises stress and 0.0248 for deflection.

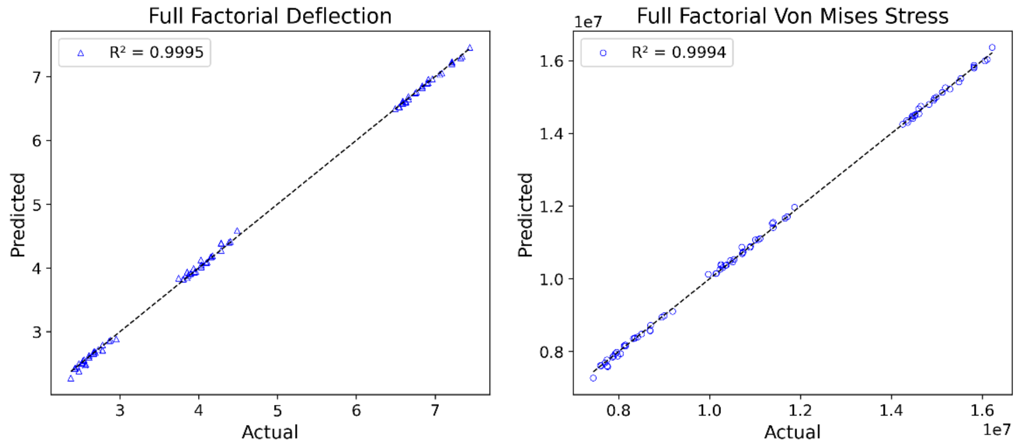


Figure 7. Actual vs Predicted plots using Full Factorial sampling.

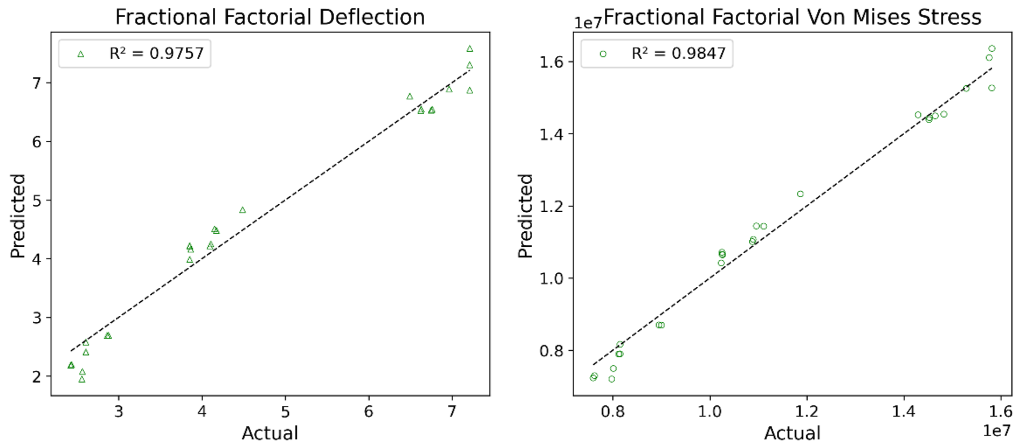


Figure 8. Actual vs Predicted plots using Fractional Factorial sampling.

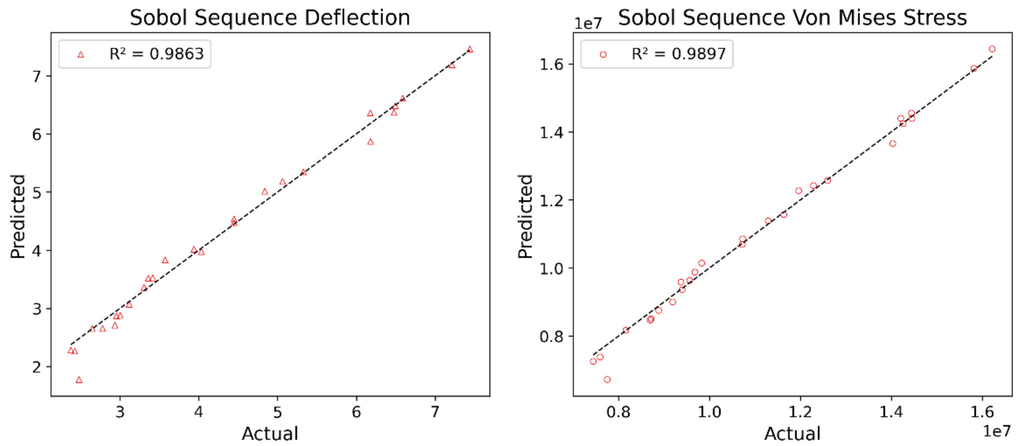


Figure 9. Actual vs Predicted plots using Sobol Sequencing sampling.

Table 10. Summary table comparing fit quality and error metrics across all datasets.

| | Von Mises Stress | | Deflection | |
|---|---|---|---|---|
| Model | $R^2$ | Avg Error | $R^2$ | Avg Error |
| RBF with Full Factorial sampling | 1 | 0.0053 | 1 | 0.0058 |
| RBF with Fractional Factorial sampling | 0.98 | 0.0378 | 0.98 | 0.0509 |
| RBF with Sobol Sequence sampling | 0.99 | 0.0204 | 0.99 | 0.0248 |

## 3.3 Validation of RBF-NN models using unseen data points

Further validation on the accuracy of all three surrogate models is performed using 25, randomly generated, unseen points within the bounds of the design space. Unseen data points are those which have not been used for training of the RBF-NN models. Using unseen data points provides qualitative indication on the predictive capabilities of the surrogate models. Table 11 summarizes the error metrics for the unseen data points for all three surrogate models.

Table 11. Error metrics for unseen data points

| | Von Mises Stress | | Deflection | |
|---|---|---|---|---|
| Model | $R^2$ | Avg Error | $R^2$ | Avg Error |
| RBF with Full Factorial sampling | 1 | 0.004 | 0.998 | 0.011 |
| RBF with Fractional Factorial sampling | 0.997 | 0.013 | 0.994 | 0.020 |
| RBF with Sobol Sequence sampling | 0.999 | 0.007 | 0.97 | 0.011 |

Trends from error metrics for unseen data points are very similar to the trends observed from the cross-validation errors. While the surrogate model from full factorial sampling provides the most accurate prediction, surrogate model from Sobol sequence provides marginally better accuracy compared to that of the fractional factorial surrogate model.

# 4 Conclusions

Surrogate models provide an advantage in terms of computational speed for use-cases which would otherwise be computationally time intensive. The approximate nature of the surrogate models warrants a strategic choice of the underlying training dataset. Although a dataset that covers the sample space better will typically provide better accuracy but larger datasets are computationally expensive to generate, train and validate. So, the computational expense associated with dataset generation, training and validation and the accuracy that the model provides need to be well balanced. This balance can depend on the choice of sampling strategy used in the dataset generation. This paper compares three different sampling methods for predicting the structural responses of a simple cantilever I-Beam use case. It can be readily seen that the full factorial method, which provides a very good coverage of the design space also provides a very good accuracy but needs 81 simulations to be run for generating the training dataset. Fractional factorial sampling only needs 27 simulations to provide sufficient accuracy in predicting the structural responses. Sobol sequencing method with equal number of data points provides better accuracy in predicting the structural responses when compared to the fractional factorial method. Therefore, out of the three techniques chosen to be investigated as a part of this paper, Sobol sequencing seems to provide for the delicate balance between accuracy and computational expense. This can be attributed to the uniform coverage of the design space achieved using the Sobol sequencing method which reduces the chances of gaps and clusters in the sampled points resulting in a comprehensive sampling.

Although this paper uses a simple cantilever I-Beam example to compare three sampling techniques, the conclusions are broadly scalable to more complex structures. Even for complex structures with more inputs and responses, the accuracy of the surrogate models will be driven by how effectively the design space has been sampled. An exhaustive sampling like full factorial will involve more experiment points and will add to the computational expense for building the dataset, training and validation. A space filling method like Sobol sequencing will provide a collection of experiment points that are highly uniform in their spacing, effectively providing a good coverage of the design space at a fractional computational expense for dataset generation, training and validation.

Further, as the problem size and complexity increases, it would become important to study the effect of variation of important parameters in Sobol sequencing technique itself. For example, the effect of distribution of corner points and interior points, the effect of reducing or increasing the number of Sobol points and the step size used in Sobol sequence generation needs to be studied.

Although the conclusion on sampling technique is broadly scalable for complex structures, further research is needed to verify the effectiveness of the RBF-NN as a choice of surrogate model to accurately predict structural responses addressing material, geometric and boundary nonlinearities in more complex use cases.

# 5 References


[1] Bianca Williams, Selen Cremaschi (2021). Novel Tool for Selecting Surrogate Modeling Techniques for Surface Approximation. Computer Aided Chemical Engineering. Elsevier. Volume 50, 451-456.

[2] Das, Sourav & Tesfamariam, Solomon. (2022). State-of-the-Art Review of Design of Experiments for Physics-Informed Deep Learning. 10.48550/arXiv.2202.06416.

[3] Dassault Systèmes 3DEXPERIENCE User Assistance R2025x on the Cloud [Online] https://help.3ds.com/2025x/english/DSDoc/FrontmatterMap/DSDocHome.htm?contextscope=cloud

[4] Huanwei Xu, Liangwen Liu, Miao Zhang (2020). Adaptive surrogate model-based optimization framework applied to battery pack design. Materials & Design. Volume 195.

[5] Jones, D.R., Schonlau, M., & Welch, W.J. (1998). Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization, 13, 455-492.

[6] Sahin, Ferat (1997). A Radial Basis Function Approach to a Color Image Classification Problem in a Real Time Industrial Application (M.Sc.). Virginia Tech. p. 26. hdl:10919/36847.

[7] Gere, J. M.; Timoshenko, S. P. (1997). Mechanics of Materials. PWS.

[8] Lumumba, Victor Wandera and Kiprotich, Dennis and Lemasulani Mpaine, Mary and Grace Makena, Njoka and Daniel Kavita, Musyimi, Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models (June 01, 2024).

[9] SIMULIA User Assistance 2025. Isight User's Guide [Online] https://help.3ds.com/2025/english/DSSIMULIA_Established/IhrUserMap/ihr-c-ov.htm?contextscope=all&id=0c92efce6ea24a1dac2205e3046019b8