10th World Congress on Structural and Multidisciplinary Optimization
May 19 -24, 2013, Orlando, Florida, USA

Effect of Characterization Test Matrix on Design Errors: Repetition or Exploration?

Taiki Matsumura[1], Raphael T. Haftka[2] and Nam-H. Kim[3]

[1] University of Florida, Gainesville, Florida, USA, taiki.matsumura@ufl.edu
[2] University of Florida, Gainesville, Florida, USA, haftka@ufl.edu
[3] University of Florida, Gainesville, Florida, USA, nkim@ufl.edu

1. Abstract
This paper investigates an effective strategy of experimental characterization of structural failure criteria. While repeating tests for the same structural configuration is preferable for tackling noise in test observation, exploring the design space with different structural configurations may uncover unmodeled critical failure modes. For a given number of tests, there arises a resource allocation problem: repetition or exploration? We examine the problem using interpolation techniques, known as surrogate models, to predict the failure load of a structural element as a function of problem parameters. Polynomial response surface (PRS), support vector regression (SVR) and Gaussian process regression (GPR) are tested. We compare repetition and exploration for two structural elements, a support bracket and a composite laminate. We conclude that repetition of tests is not necessarily needed since fitting surrogates has the effect of filtering out noise. This conclusion is intensified when the failure load surface is complicated. Furthermore, fitting the surrogate models with all the repeated data led to better accuracy than using only the mean values of repeated data.
2. Keywords: Test, surrogate models, Failure modes, resource allocation

3. Introduction
Commercial aviation is the safest mode of transportation, partly due to large safety factor used in aircraft design. However, factors of safety may not compensate for large errors in the analytical predictions of structural failure, especially when designers fail to identify failure modes. Such failures may cause critical delays in aircraft development and later in service. Thus, it is imperative to identify failure modes and appropriately characterize them as early during the design stage as possible.

A key element in the process of helping designers predict structural failure is the construction of design allowable charts for each failure mode, e.g., failure load map with respect to geometry and load conditions. Due to the complexity of failure mechanisms and lack of knowledge, especially for newly introduced materials and structures, analytical prediction models, are not reliable enough. Therefore, failure-criterion characterization tends to rely on tests. We typically conduct a matrix of characterization tests to cover the expected design space. To achieve the accuracy of failure load mapping, we may want to repeat several tests for the same configuration to eliminate the effect of noise in the test due to variability in material properties and test conditions. On the other hand, exploring within the design space with many different configurations is more likely to spot yet-unidentified failure modes.

The objective of this paper is to shed light on key factors of the characterizing tests and their relationship to failure prediction errors. We view this test process as a resource allocation problem. Because tests are often expensive, a key question is how we effectively allocate the limited number of tests to satisfy those two different objectives; predicting accurately a failure load map and spotting yet-unidentified failure modes (repetition vs. exploration). We illustrate the failure-criterion characterization using two example structural elements and metamodeling techniques, also known as surrogate models, for the failure load mapping.

We test different types of surrogate models, including polynomial response surface, Gaussian process regression, and support vector regression, all of which are known to be capable of smoothing equivalent to a noise filter. In addition, we examine the treatment of the repeated data. That is, whether we fit surrogates using repeated data or whether we fit to the mean values t of the repeated data. With the help of the examples, we will discuss effective strategies of the failure-criterion characterization both in the context of structural test and in the context of use of surrogate models.

4. Characterizing Test
A major role of characterizing tests is to refine failure criteria which are critical for determining design allowables. As quoted by the Department of Defense handbook for composite materials [1], "unfortunately, the capability of the state of the art analysis methods are limited." This process is usually conducted with a family of element structures intended for a particular use. A series of tests needs to consider multiple factors, such as external environments, geometry of structure, variability of material properties, defects and so on.

Characterizing failure criteria has two goals. The first goal is to identify all potential failure modes. For doing that, to quote the DOD handbook [1], "it is important to carefully select the correct test specimens that will simulate the desired failure modes. Special attention should be given to matrix sensitive failure modes." The handbook also states that "the multiplicity of potential failure modes is perhaps the main reason that the building block (test) approach is essential in the development of composite structure substantiation." The second goal is to enable accurate design allowable prediction, for example a failure load chart with respect to geometry and load conditions. This is usually implemented by a matrix of experiments and approximation techniques to interpolate the observed data.

Challenges of achieving these goals include noisy data observation due to material variability, error in test conditions and measurement, and unknown potential failure modes. While the effect of noise can be statistically quantified by repeating test observation on each point in the matrix, exploring within the matrix with many different points is more likely to capture underlying potential failure modes. A key question is how we allocate the limited number of tests: repetition or exploration? (Fig. 1)

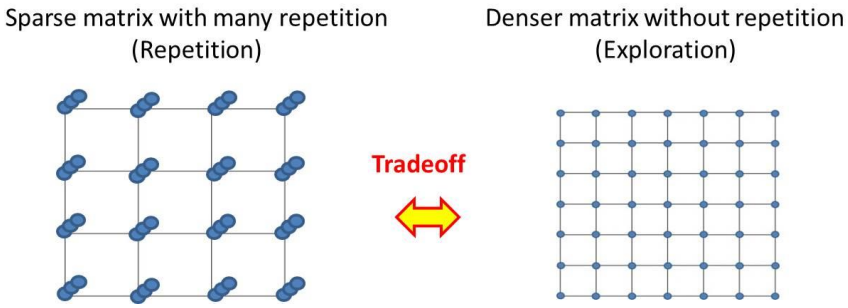


Figure 1: Resource allocation issue in determining test matrix

In this paper, we focus on failure criterion characterization with a low dimension and a small number of tests. We chose two very simple examples, for clarity and to allow exhaustive study of large number of strategies. The examples are support bracket and a composite laminate plate. Each structure has two underlying failure modes; one is dominant in the design space and the other is rare, representing an un-modeled mode that might be missed. Whereas the composite laminate plate has a high order of nonlinearity of the failure load surface, the support bracket has a smooth simple surface.

4.1 **Support bracket**

A simple support bracket mounted on a base structure is shown in Fig 2. The load is imposed on the handle and the expected operational load angle α is 0 to 110 deg in the x-z plane. It is also assumed that the height of the bracket l and length a are fixed due to space constraints. The diameter of the cylindrical part d is considered as a design parameter. Table 1 shows the properties of the structure.

The combination of loading and geometry generates multi-axial states of stress due to axial, bending, torsion, and torsional shear stresses. Figure 3 illustrates the critical failure modes of the structure. Because of the additive effect of the torsion and torsional shear stresses or bending and axial stresses, point D is likely to be a critical failure point. However, point A can be a critical point under some conditions as shown in Figure 3. If the designer fails to locate the mode initiated at the point A, the design allowable will be underestimated.

It is assumed that the yield strength of the material is normally distributed, and the geometry of test specimens varies within the tolerances of manufacturing, which are the sources of noisy test observation. Failure is predicted by the Von-Mises criterion ignoring stress concentrations. The tests seek to allow designers to predict the mean failure loads due to the dominant mode at the point D, which is determined by the mean of yield strength and the nominal values of d and α. Figure 4 depicts the failure load mapping at the point D.

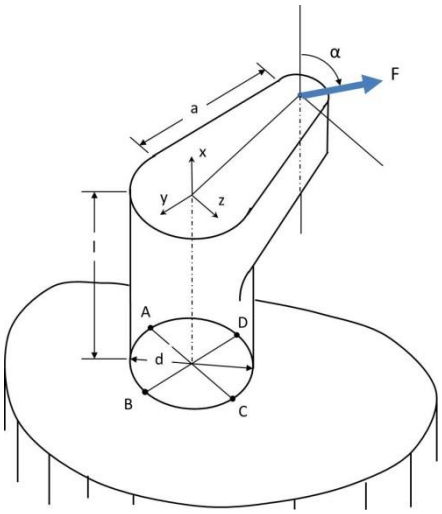


Figure 2. Support Bracket

Table 1. Properties of support bracket

Property	Quantity	Variability
l [inch]	2	Uniform ±2%
a [inch]	4.6	Uniform ±2%
d [inch]	[1, 3]	Uniform ±2%
α [deg]	[0, 110]	N/A
Yield strength [psi]	43,000	Normal 10% COV

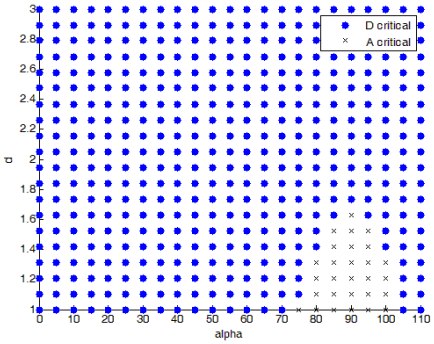


Figure 3. Critical failure modes

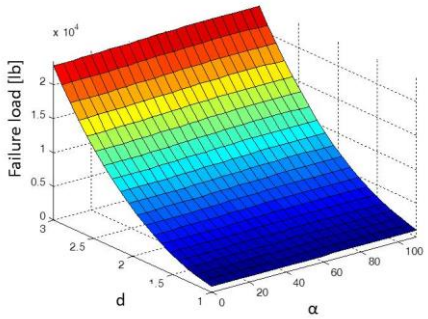


Figure 4. Failure load surface for Point-D failure

4.2 Composite laminate plate

For the second example, intended to have a more complex failure surface, a symmetric composite laminate with three ply angles [0˚/ -θ/ +θ]s is considered (Fig. 5). The laminate is subject to mechanical loading along the x and y directions defined by the load ratio α, such that Nx = (1-α)F and Ny = αF. As design parameters for the failure load identification, the ply angle θ and the loading condition α are selected. The range of the parameters are set as [0, 90] deg for θ and [0, 0.5] for α. Table 2 shows the material properties and strain allowables, including strain allowable along fiber direction ε_{1allow}, transverse along fiber direction ε_{2allow}, and shear $\gamma_{12allow}$. All the properties are assumed to be normally distributed and the source of the noise in data observation. The strains are predicted by the classical laminated plate theory.

Figure 6 shows the mapping of the critical failure modes, one due to the ply axial strain, which is dominant, and the other due to ply shear strain, which is rare. The designer is assumed to conduct a series of tests in order to construct an accurate approximation of the failure load map of the dominant mode as well as to spot the less dominant mode. Figure 7 is the failure load surface due to ply axial strain.

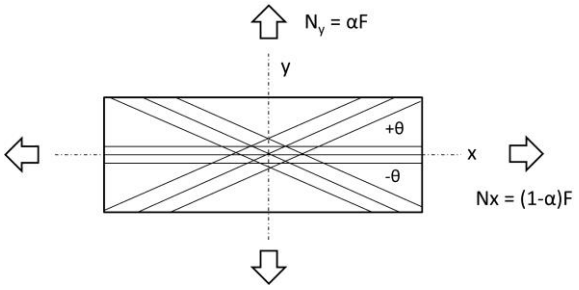


Figure 5. Composite laminate

Table 2. Properties of composite laminate plate

Property	Quantity (mean)	COV
E_1 [GPa]	150	5%
E_2 [GPa]	9	5%
ν_{12}	0.34	5%
G_{12} [GPa]	4.6	5%
Thickness of ply [μm]	125	N/A
ε_{1allow}	±0.01	6%
ε_{2allow}	±0.01	6%
$\gamma_{12allow}$	±0.015	6%

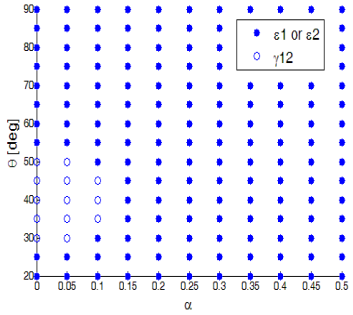


Figure 6. Critical failure modes

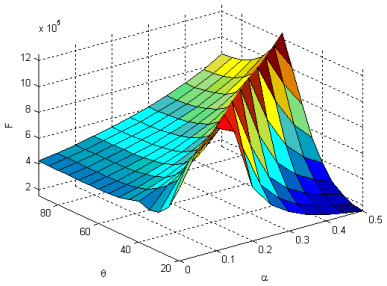


Figure 7. Failure load surface due to axial ply strain

4.3 Test matrix

In this study, we use matrices with from 4x4 to 7x7 grid in order to investigate the effect of the density of matrix. For each of the test matrices, we repeat the same test configuration (N_r) up to seven times. Table 3 shows the total number of test points for the matrices.

As mentioned earlier, it is obvious that a denser matrix is desirable to spot rare failure modes. For both structural examples, 5x5 test matrix or denser with evenly spaced test points (N_g) in the two dimensions will detect the less dominant failure modes.

Table 3. Test matrix and total number of tests

Matrix (N_gxN_g)	Number of repetitions (N_r)						
	1	2	3	4	5	6	7
4x4	16	32	48	64	80	96	112
5x5	25	50	75	100	125	150	175
6x6	36	72	108	144	180	216	252
7x7	49	98	147	196	245	294	343

5. Surrogate Models

Surrogate models have been widely used in lieu of expensive computer simulations and experiments in engineering design. Their capability and validity have been proven by a variety of applications, such as design optimization [2, 3], uncertainty quantification [4]. Since one of the challenges of the characterizing test is to tackle the noise in observation, we select three surrogate models that have a smoothing effect, including polynomial response surface (PRS), support vector regression (SVR) and Gaussian process regression (GPR). The following sections briefly describe each of the surrogate models and the treatment of the test data for surrogate fitting.

5.1 Polynomial Response Surface (PRS)

Polynomial response surface employs a polynomial function and the least square fit to approximate the true function. As an example, the second order polynomial model is shown in Eq.(1).

$$\hat{y} = \sum_{i=1}^{k} \alpha_i x_i + \sum_{i=1}^{k} \sum_{j=1}^{k} \alpha_{ij} x_i x_j \quad (1)$$

where x represents the input variable, $\hat{y}$ represents the prediction of output y. α represents the coefficient and k is the number of training data points. PRS is known for computational tractability as well as for smoothing noisy observations [5, 6]. Because of polynomial functions being applied, it may cause a problem when being fitted to functions not well approximated by polynomials. In this study, we test various orders of polynomial function (up to 9^{th} order), then choose the best one based on the accuracy of the prediction, i.e., root mean square error (RMSE).

5.2. Support Vector Regression (SVR)

Support vector regression evolved from machine learning algorithm [7]. SVR has a tolerance of error and finds an approximation model that minimizes the degree of violation beyond the error tolerance. In case of linear approximation, the prediction model is formulated as

$$\hat{y}(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{x} \rangle + b \quad (2)$$

where $\langle \boldsymbol{w}, \boldsymbol{x} \rangle$ is a dot product of the coefficient vector $\boldsymbol{w}$ and $\boldsymbol{x}$, and b is the base term. The regression is carried out by optimizing $\boldsymbol{w}$ and b by solving the optimization problem shown in Eqs.(3) and (4). Regularization parameter C is a user-defined parameter and tradeoffs between flatness of the prediction model (low coefficients), the first term of Eq.(3), and the violation of the error tolerance, the second term of Eq.(3).

$$\min_{w,b} \frac{1}{2} |\boldsymbol{w}|^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) \quad (3)$$

$$\text{Subject to} \quad \begin{aligned} y_i - \langle w, \boldsymbol{x}_i \rangle - b &\leq \varepsilon + \xi_i \\ \langle w, \boldsymbol{x}_i \rangle + b - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned} \quad (4)$$

Figure 8 illustrates one of the most common models for the error tolerance (used here), so-called ε-sensitive loss function. When the error ($= \hat{y}_i - y_i$) is within the tolerance ±ε, the loss (ξ) is zero, otherwise the loss is proportionally increased with the error.

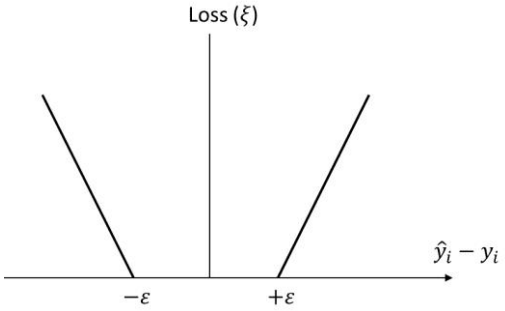


Figure 8. ε-sensitive loss function

A more general SVR model is expressed by Lagrange multipliers, α_i and α_i^*.

$$\hat{y}(\boldsymbol{x}) = \sum_{i=1}^{k} (\alpha_i - \alpha_i^*) K(\boldsymbol{x_i}, \boldsymbol{x}) + b \quad (5)$$

where $K(\boldsymbol{x_i}, \boldsymbol{x})$ is the kernel function, and $\boldsymbol{x_i}$ represents the ith training point and $\boldsymbol{x}$ is the point to be estimated by the surrogate model.

For the implementation, we deploy the Surrogate Tool Box [8], which employs the MATLAB code offered by Gunn [9]. One of the challenges of SVR is to select an appropriate regularization parameter C. In general,

substantial large C is suggested [10] and the exact selection of C "is not overly critical" [11]. Since our focus is on noise-canceling rather than flatness, and based on some empirical works to seek for an accurate fitting, we select infinity for C. For the kernel function, we use Gaussian model as shown in Eq.(6), which is often used, with θ being also a user-defined parameter, which determines the correlation between the points.

$$K(\boldsymbol{x_i}, \boldsymbol{x}) = exp\left(-\frac{|\boldsymbol{x_i} - \boldsymbol{x}|^2}{2\theta^2}\right) \tag{6}$$

Nearby points should be highly correlated for smooth interpolation, and after some experimentation, we selected θ such that K=0.9 for the closest two training data points. For ε, we use the average standard deviation of the observed data from 7x7 matrix with 7 repetitions, assuming that the designer has some idea about the noise in observation. Table 4 shows the selected ε comparing to the noise level. However, since this assumption might not be realistic, the selection of ε will be further investigated for the future work whether we can take advantage of statistic estimated from the observed data.

Table 4. ε selection and noise level. Noise level varies depending on the location in the design space. All numbers are normalized by the range of the failure load).

	Noise level		ε
	Max	Min	
Support bracket	0.004	0.122	0.041
Composite laminate	0.009	0.085	0.041

5.3. Gaussian Process Regression (GPR)

Gaussian process regression is originally developed as a method of the spatial statistics [12]. GPR is also known as Kriging [13]. Instead of assuming an approximated function like PRS, GPR views a set of data points as a collection of random variables that follow some rule of correlation, called random process. The name of Gaussian process originates from the form of random process using multivariate normal (Gaussian) distribution. The prediction is formulated as

$$\hat{y} = \sum_{i=1}^{k} \beta_i \xi_i(\boldsymbol{x}) + Z(\boldsymbol{x}) \tag{7}$$

The first terms of Eq.(7) is called the trend function and $Z(x)$ represents a departure from the trend function. The departure is assumed to be a realization of a random process which is expressed by a spatial correlation function. For example, Gaussian model is expressed as

$$\text{cov}\left(Z(\boldsymbol{x_i}), Z(\boldsymbol{x_j})\right) = \sigma^2 \exp\left(-\sum_{l=1}^{m} \frac{\left|x_l^{(i)} - x_l^{(j)}\right|^2}{2{\theta_l}^2}\right) + {\sigma_n}^2 \tag{8}$$

where σ^2 is process variance with zero mean. where θ_l is scaling parameter of the l-th component of $\boldsymbol{x}$, which determines the correlation between the points, and σ_n is the noise variance independent of σ in order to handle the repeated data. Throughout the regression process, the trend function and hyperparameters of the correlation function are optimized to maximize the likelihood of having the training points. The advantage of GPR is the flexibility of fitting to nonlinear functions. However, fitting process of GPR is time consuming due to the optimization process.

In this paper, we use the Gaussian Process Regression and Classification Toolbox version 3.2 [12] for the implementation. We select a linear model for the trend function and Gaussian model for the correlation function, shown in Eq.(8), which is one of common correlation functions. Since the toolbox deploys a line-search method for the optimization of the hyperparameters θ and the variances, the optimal solution tends to depend on the starting points of the hyperparameters. To avoid the danger of resulting in a local optimum, we apply multiple starting points [1, 0.1, 0.01, 0.001, 0.0001, 0.00001] for both σ and σ_n in the normalized output space (36 combinations of the starting points). We also select the starting point of θ such that $\exp\left(-\sum_{l=1}^{m} \frac{\left|x_l^{(i)} - x_l^{(j)}\right|^2}{2{\theta_l}^2}\right) = 0.9$

for the closest two points among the training points, as discussed for the kernel function of SVR. After fitting with all the starting points, we select the best model based on their maximum likelihood.

5.3. Treatment of test data

We test the following two strategies of test data for fitting the surrogate models. Note that both strategies provide the same result for PRS.

(1) All-at-once strategy

We fit the surrogate models to all the test data including the repeated data.

(2) Mean strategy

We first take the mean values of the repeated data at each location in the design space. Then, the surrogate models are fitted to the mean values.

6. Results of Failure Load Mapping

6.1 Error evaluation

In order to evaluate the accuracy of the failure load mapping, we compare the predictions from the surrogate models with the true values at a 20x20 matrix of testing points (in total 400 points). The true values are the failure loads corresponding to the mean values of all the input random variables, i.e., the material properties and geometry. To measure overall accuracy, we use root mean square error normalized by the range of the failure loads (NRMSE) calculated by Eq.(9). We also evaluate normalized maximum absolute error calculated by Eq.(10) in order to examine the robustness of the surrogate models.

(a) Normalized root mean square error (NRMSE)

$$NRMSE = \frac{1}{\text{range of } y}\sqrt{\frac{1}{400}\sum_{i=1}^{400}(\hat{y}_i - y_i)^2} \tag{9}$$

(b) Normalized maximum absolute error (NMAE)

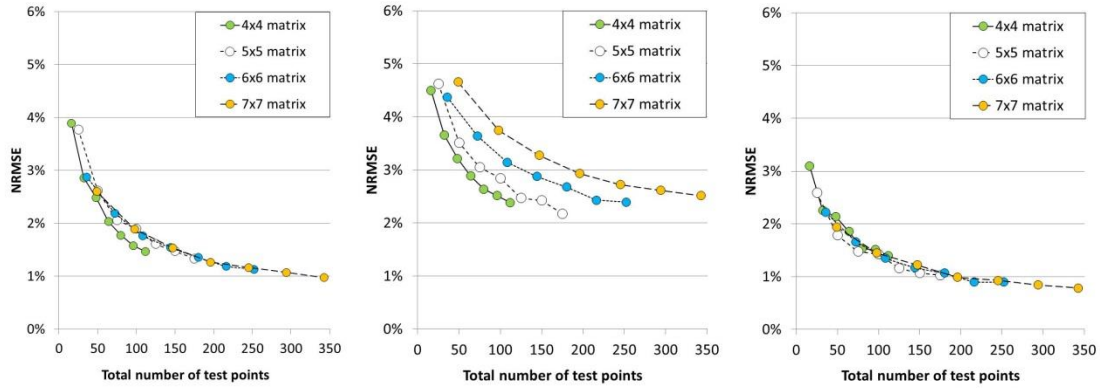
$$NMAE = \frac{1}{\text{range of } y}\max(|\hat{y}_1 - y_1|, |\hat{y}_2 - y_2|, \dots, |\hat{y}_{400} - y_{400}|) \tag{10}$$

First, we calculate the failure loads based on randomly generated test data at the matrix of experiments (ranging from 4x4 to 7x7 grid). Then, we fit the surrogate models to the failure loads. We repeat the fitting process 100 times, each of which has a different set of the random test data and the failure load surface. Finally, we evaluate the errors, i.e., NRMSE and NMAE, for each of the fittings. The errors discussed in this section are the mean values of NRMSE and NMAE over 100 runs. The standard errors of the mean values over 100 runs are, in average, 0.2% for NRMSE and less than 0.5% for NMAE, meaning that they are small enough for the comparison discussed in the next section.

6.2 Support bracket

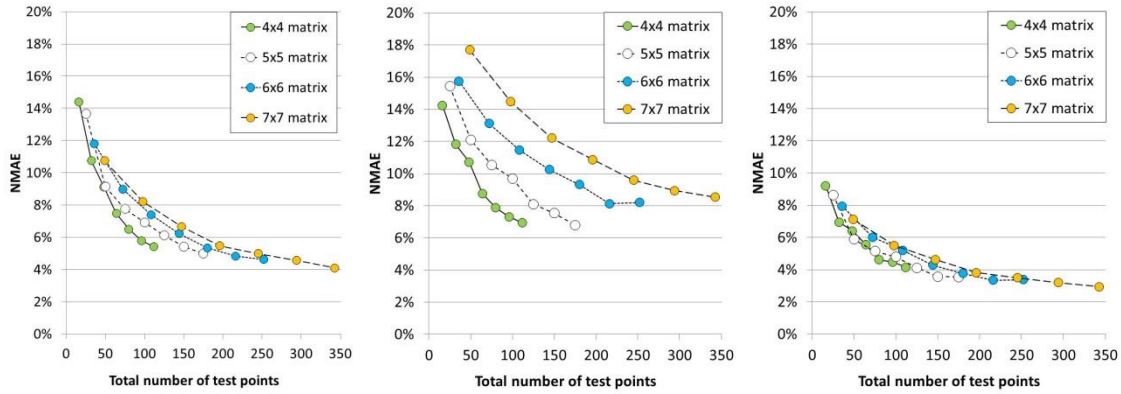
Figure 9 compares the NRMSE with respect to the total number of test points ($N_g^2 \times N_r$) when the all-at-once strategy is applied to PRS, SVR, and GPR. The markers on each of the lines correspond to, from left to right, one repetition through seven repetitions. For PRS, 4^{th} order polynomial function performs best for all the matrices.

First of all, it can be seen that PRS performs better than SVR. This may reflect the fact that PRS is more error sensitive than SVR because while PRS minimizes L2 error norm (i.e., RMSE) by the least square method, SVR with ε-sensitive loss function minimizes L1 error norm or that the selection of ε might be inappropriate. In terms of NMAE, both in PRS and SVR, as the test matrix becomes denser, the accuracy deteriorates. This is because the more test data observed, the higher the chance of observing outliers which harm the accuracy of the fitting. However, this does not apply to GPR, indicating that GPR is more robust against outliers.



(a) PRS (b) SVR (c) GPR

Figure 9: Error comparison for support bracket: **NRMSE for all-at-once strategy**
Markers of each line correspond to one repetition through seven repetitions from left to right (see Table 3 for details on the numbers of data). The error values are the means out of 100 runs.



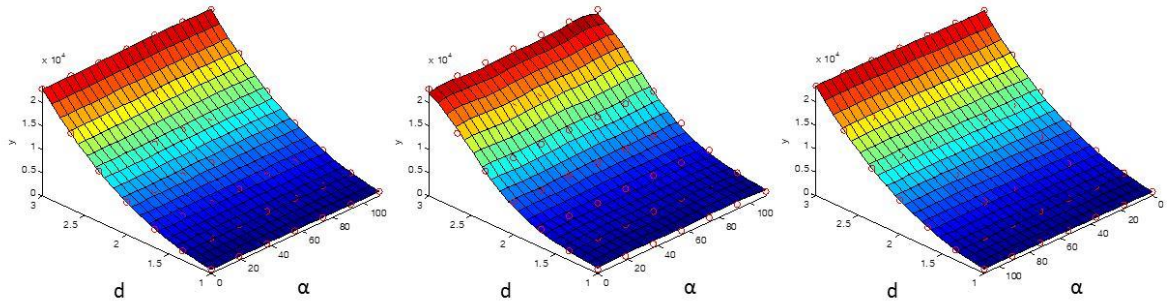
(a) PRS (b) SVR (c) GPR

Figure 10: Error comparison for support bracket: **NMAE for all-at-once strategy**.
Markers of each line correspond to one repetition through seven repetitions from left to right. The error values are the means over 100 runs.

Furthermore, we measured the errors associated with the surrogate models fitted to the noise-free test observation by 7x7 matrix. Table 5 shows the errors of the surrogate models fitted to the noise-free data and the remaining errors. The predictions of the surrogates are shown in Fig.11. Almost zero errors of PRS and GPR means that the errors in these surrogates are mainly due to the noise in observation rather than the modeling error. The large error in SVR (2.71%) reflects the wavy behavior of SVR prediction shown in Fig.11(b). Based on the remaining errors, SVR and GPR seem to work better as a noise filter than PRS.

Table 5. Means (over 100 runs) of errors of the surrogates fitted to noise-free data (support bracket). The number in parentheses represents the standard error of the mean value.

	Error		
	Noise-free data (7x7 matrix)	Noisy data (7x7 matrix with no repetition)	Difference
PRS	0.02%	2.60% (±0.07%)	2.58%
SVR	2.71%	4.66% (±0.08%)	1.94%
GPR	0.02%	1.94% (±0.09%)	1.92%

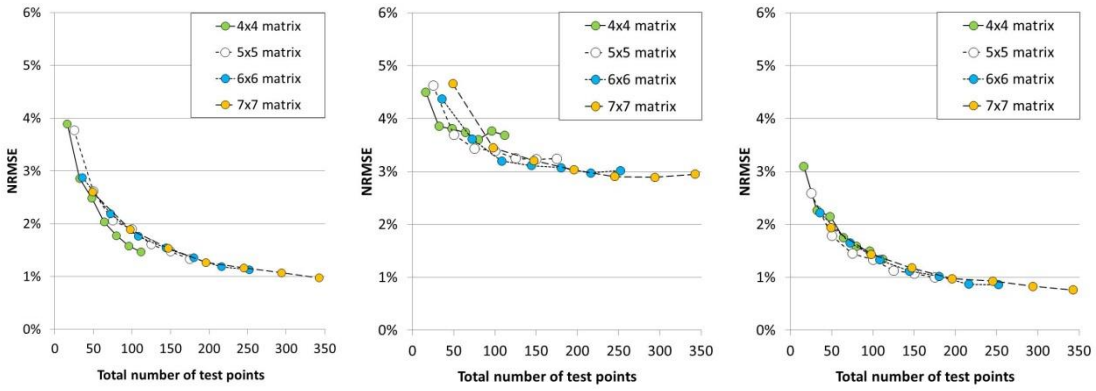


(a) PRS (b) SVR (c) GP

Figure 11: Surrogate models fitted to the noise-free data (7x7 matrix) of the support bracket

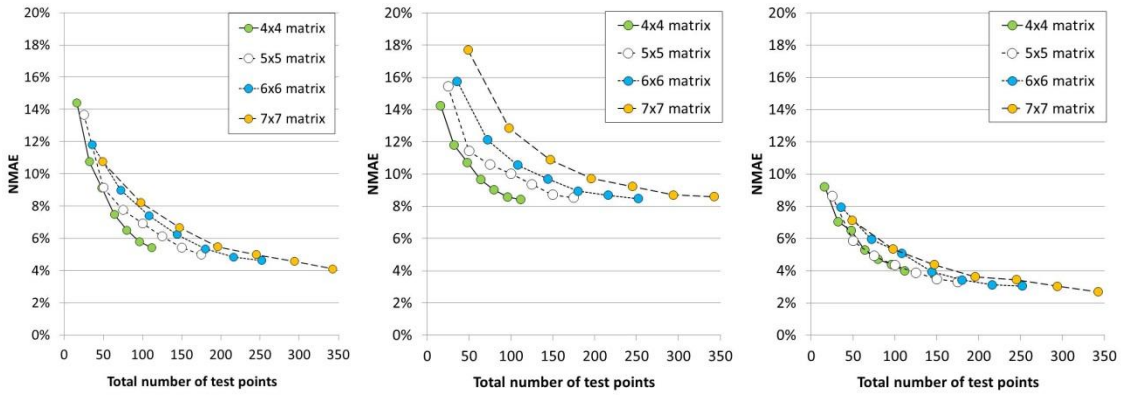
Next, we examine the resource allocation (repetition vs. exploration). For PRS and GPR in Fig.9, all the lines corresponding to the density of matrix form a single line, meaning that the resource allocation between repetition and exploration does not matter, but the total number of tests does. As discussed in the previous paragraph, the errors mainly come from the noise rather than the prediction models. For such case, increasing the number of tests contributes to reducing the error regardless of whether they are for repetition or whether they are for exploration. However, recall that PRS might have a danger of outliers from a denser matrix, which harms the accuracy.

Finally, in terms of the treatment of repeated data, Figs.12 and 13 show NRMSE and NMAE respectively when the mean strategy is applied. Compared to the all-at-once strategy shown in Figs.9 and 10, while no significant difference is observed both in NRMSE and NMAE for PRS and GPR, the all-at-once strategy offers a better performance than the mean strategy for SVR. This happens because SVR only considers the data outside of the tolerance $\pm\varepsilon$ as the errors to be reduced. Since SVR with the mean strategy does only provide one point, i.e., the mean value, at each location in the design space, it does not fully take advantage of the fitting algorithm. This may conclude that it is recommended to inform SVR about as many data as possible.



(a) PRS (b) SVR (c) GPR

Figure 12: Error comparison for support bracket: **NRMSE for mean strategy**.
Markers of each line correspond to one repetition through seven repetitions from left to right. The error values are the means out of 100 runs.



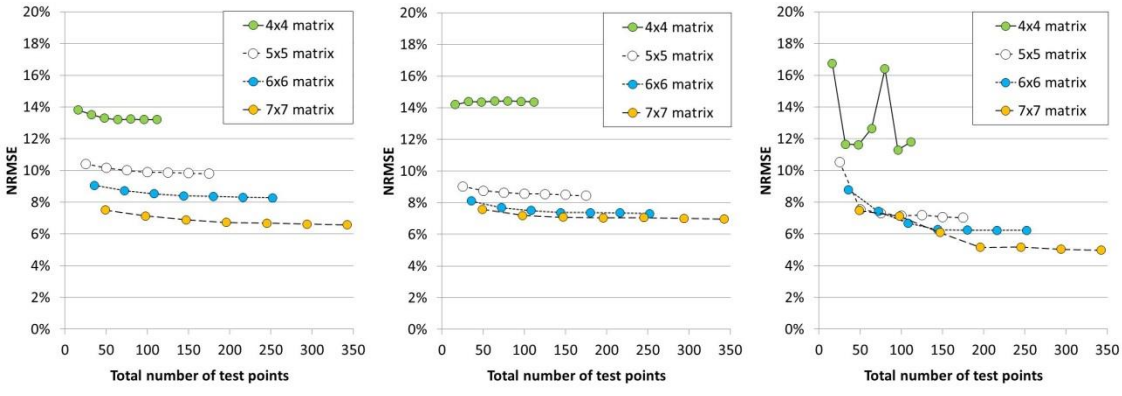
(a) PRS (b) SVR (c) GPR

Figure 13: Error comparison for support bracket: **NMAE for mean strategy.**
Markers of each line correspond to one repetition through seven repetitions from left to right. The error values are the means out of 100 runs.

6.3 Composite laminate plate

For the fitting performance to the nonlinear failure load surface of the composite laminate plate, Figure 14 shows NRMSE of all the surrogate models. For PRS, the best performed polynomial functions are listed in Table 6. Due to the non-linearity of the surface, the performance for the sparse matrices (4x4) is quite poor and unstable for all the models. Once the denser matrices (5x5 matrix or denser) are available, which are substantial enough to capture the nonlinear behavior, the accuracy of the prediction is drastically improved.

In terms of overall accuracy, GPR surpasses PRS and SVR. PRS and SVR are comparable for the 7x7 matrix, but SVR performs better for sparser matrices, i.e., 5x5 and 6x6. A remarkable observation is that the contribution of repetition to reducing error is miniscule. For example, given about 100 test points, 7x7 matrix with two repetitions is better than 6x6 with matrix with three repetitions and 5x5 matrix with four repetitions (for GPR, a significant difference can be seen when a higher number of tests are sampled). Table 7 depicts the probability that a denser matrix with a smaller number of repetitions offers a better accuracy than a sparse matrix with a larger number of repetitions. NRMSE is compared at each of the 100 fitting iterations. Except for the comparison of SVR between 7x7 matrix and 6x6 matrix, for the given numbers of tests, exploration steadily provides a more accurate approximation than repetition.



(a) PRS (b) SVR (c) GPR

Figure 14: Error comparison for composite laminate plate: **NRMSE for all-at-once strategy**
Markers of each line correspond to one repetition through seven repetitions from left to right. The error values are the means out of 100 runs.

Table 6. Best performed PRS for composite laminate plate

Matrix	Order of polynomial function
4x4	5^{th} order
5x5	5^{th} order
6x6	7^{th} order
7x7	8^{th} order

Table 7. Comparison of test strategies for a given number of tests

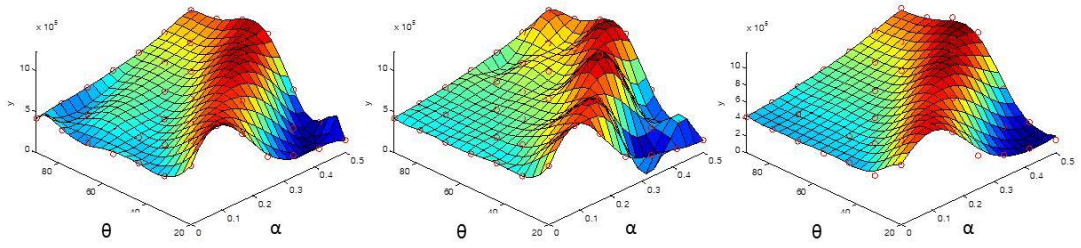
	Comparison			Chance that a denser matrix is more accurate
	7x7 matrix	6x6 matrix	5x5 matrix	
PRS (given about 100 tests)[1]	✓	✓		96%
	✓		✓	99%
		✓	✓	98%
SVR all-at-once (given about 100 tests)[1]	✓	✓		64%
	✓		✓	95%
		✓	✓	96%
GPR all-at-once (given about 200 tests)[2]	✓	✓		82%
	✓		✓	92%
		✓	✓	100%

[1] PRS and SVR: 7x7 (2 repetitions) vs. 6x6 (3 repetitions) vs. 5x5 (4 repetitions)
[2] GPR : 7x7 (4 repetitions) vs. 6x6 (6 repetitions) vs. 5x5 (7 repetitions)

As Table 8 shows that the errors of the models fitted to the noise-free data, this is because the major source of total error is due to the error in prediction models rather than the noise. In order to improve the accuracy of the prediction models, we need to capture the nonlinear behavior of the surface by as many different locations as possible in the design space. Figure 15 illustrates the predictions fitted to the noise-free data. Similar trends are also observed in NMAE shown in Fig.16, but the contribution of repetition to the accuracy can be hardly seen.

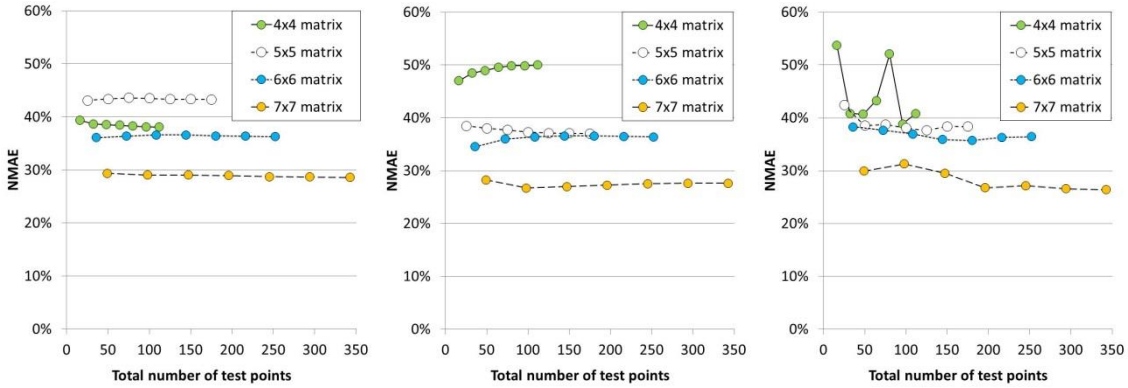
Table 8. Means (over 100 runs) of errors of the surrogates fitted to noise-free data (support bracket). The number in parentheses represents the standard error of the mean.

	Error		
	Noise-free data (7x7 matrix)	Noisy data (7x7 matrix with no repetition)	Difference
PRS	6.26%	7.51% (±0.13%)	1.25%
SVR	6.43%	7.58% (±0.12%)	1.15%
GPR	6.71%	7.48% (±0.12%)	0.77%



(a) PRS (b) SVR (c) GPR

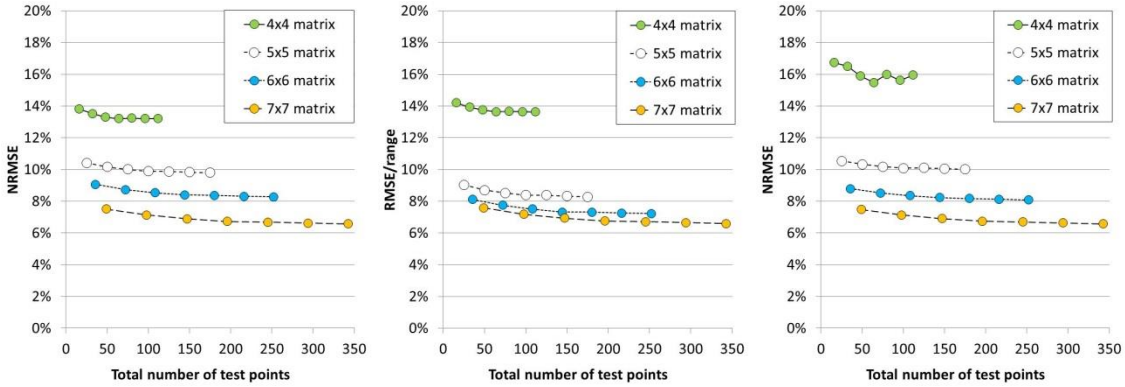
Figure 15: Surrogate models fitted to the noise-free failure load surface of the composite laminate plate



(a) PRS (b) SVR (c) GPR

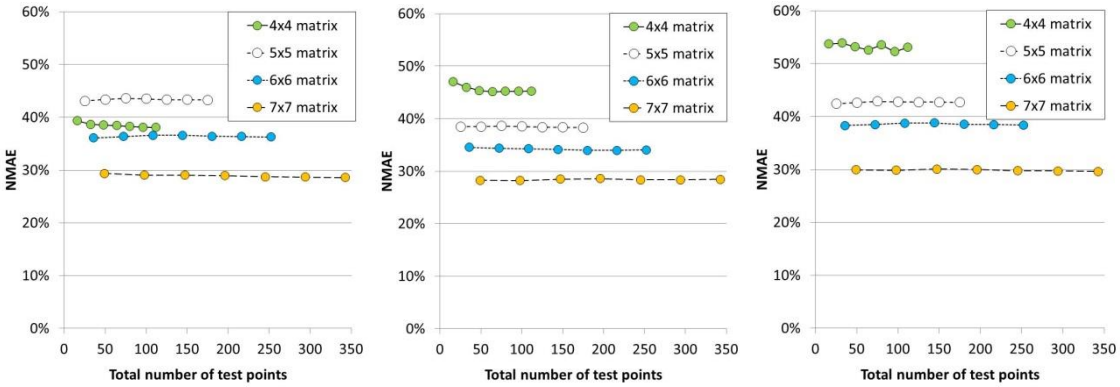
Figure 16: Error comparison for composite laminate plate: **NMAE for all-at-once strategy**.
Markers of each line correspond to one repetition through seven repetitions from left to right. The error values are the means out of 100 runs.

As for the treatment of repeated data for fitting, Figs.17 and 18 illustrate NRMSE and NMAE when the mean strategy is used. For SVR, the all-at-once strategy is better than the mean strategy to a small extent. SVR still takes advantage of as many data, as discussed previously. In case of GPR, the mean strategy also underperforms the all-at-once strategy. It seems that as many data is helpful to improve the model accuracy even for GPR.



(a) PRS (b) SVR (c) GPR

Figure 17: Error comparison for composite laminate plate: **NRMSE for mean strategy**
Markers of each line correspond to one repetition through seven repetitions from left to right. The error values are the means out of 100 runs.



(a) PRS (b) SVR (c) GPR

Figure 18: Error comparison for composite laminate plate: **NMAE for mean strategy**.
Markers of each line correspond to one repetition through seven repetitions from left to right. The error values are the means out of 100 runs.

7 Concluding remarks and future work

We investigated an effective strategy of matrix of tests for failure criterion characterization, focusing on allocation of tests between repetition and exploration. Polynomial response surface (PRS), support vector regression (SVR) and Gaussian process regression (GPR) were examined to approximate the failure load surface. With two structural elements, it is found that repetition of tests is not necessarily needed in terms of the accuracy of approximation because of the smoothing effect of surrogate models. This conclusion is stronger when the failure load surface is complex, because the error in surrogate fitting tends to be more dominant than noise in observation. Furthermore, fitting the surrogates (GPR and SVR) to the all the repeated data performed better than fitting only with the mean values of repeated data.

For future work, we will further investigate the treatment of repeated data in order to see whether we can take advantage of statistics of repeated data. In the present paper, we assume that the designer knows about the noise level and uses the average value the noise over the design space for ε. However, it is not realistic. For example, a variance estimator might be useful to select the error tolerance ε of SVR. ε might as well be adjusted according to the number of repetitions, like the standard error of mean. Furthermore, the weighted least square fitting is used to diminish the effect of outliers.

8. References

[1] "Department of Defense Handbook, Composite materials handbook, Volume 3, Chapter 4," MIL-HDBK-17-3F,

[2] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P. K., "Surrogate-based analysis and optimization," *Progress in Aerospace Sciences*, Vol. 41, No. 1, 2005, pp. 1-28.
doi: DOI 10.1016/j.paerosci.2005.02.001

[3] Jin, R., Du, X., and Chen, W., "The use of metamodeling techniques for optimization under uncertainty," *Structural and Multidisciplinary Optimization*, Vol. 25, No. 2, 2003, pp. 99-116.
doi: DOI 10.1007/s00158-002-0277-0

[4] Giunta, A. A., Mcfarland, J. M., Swiler, L. P., and Eldred, M. S., "The promise and peril of uncertainty quantification using response surface approximations," *Structure and Infrastructure Engineering*, Vol. 2, No. 3-4, 2006, pp. 175-189.
doi: Doi 10.1080/15732470600590507

[5] Giunta, A. A., and Watson, L. T., "A comparison of approximation modeling techniques: Polynomial versus interpolating models," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, St. Louis, MO, 1998, pp. 392–404.

[6] Jin, R., Chen, W., and Simpson, T. W., "Comparative studies of metamodelling techniques under multiple modelling criteria," *Structural and Multidisciplinary Optimization*, Vol. 23, No. 1, 2001, pp. 1-13.

[7] Smola, A. J., and Scholkopf, B., "A tutorial on support vector regression," *Statistics and Computing*, Vol. 14, No. 3, 2004, pp. 199-222.
doi: Doi 10.1023/B:Stco.0000035301.49549.88

[8] Viana, F. A. C., "SURROGATES TOOLBOX User's Guide (Ver. 2.1)," 2010.

[9] Gunn, S. R., "Support Vector Machines for Classification and Regression," University of Southampton, 1998.

[10] Jordaan, E. M., and Smits, G. F., "Estimation of the regularization parameter for support vector regression," *World Conference on Computational Intelligence 2002*,, Honolulu Hawaii, May 12-17, 2002.

[11] Forrester, A. I. J., Sóbester, A., and Keane, A. J., *Engineering Design via Surrogate Modelling: A Practical Guide*, Wiley, 2008.

[12] Rasmussen, C. E., and Williams, C. K. I., *Gaussian Processes in Machine Learning*. Cambridge, The MIT Press, 2005.

[13] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, 1989, pp. 409-423.