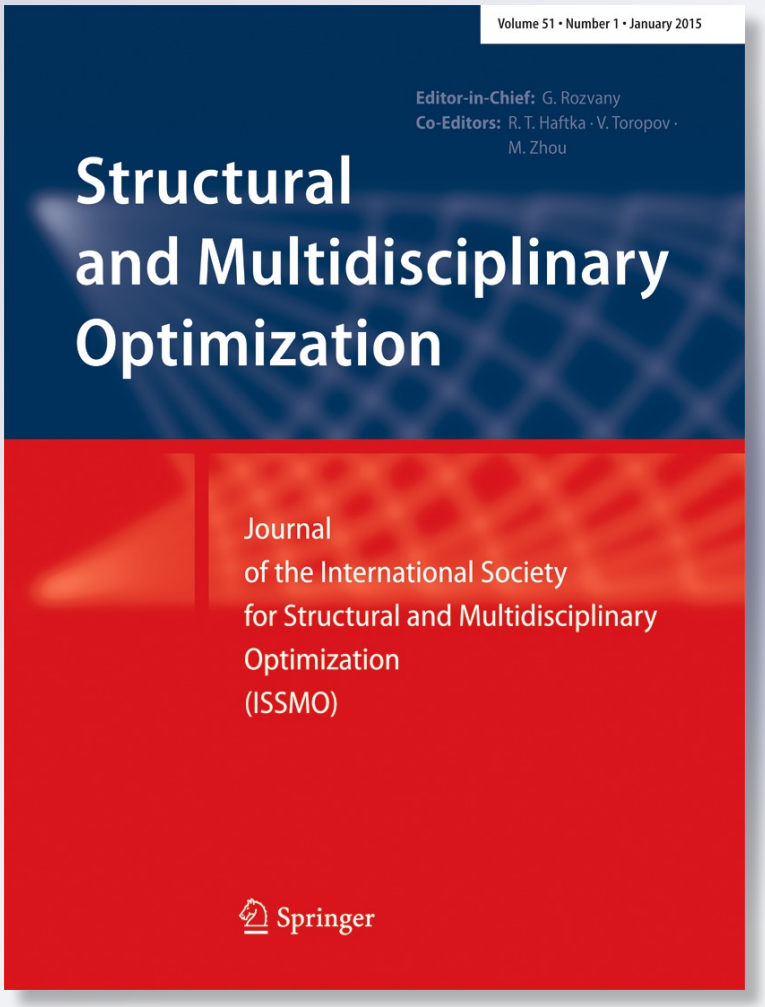
Accurate predictions from noisy data: replication versus exploration with applications to structural failure

Taiki Matsumura, Raphael T. Haftka & Nam H. Kim

Structural and Multidisciplinary Optimization

ISSN 1615-147X
Volume 51
Number 1

Struct Multidisc Optim (2015) 51:23-40
DOI 10.1007/s00158-014-1115-x



Springer

Struct Multidisc Optim (2015) 51:23–40
DOI 10.1007/s00158-014-1115-x

RESEARCH PAPER

Accurate predictions from noisy data: replication versus exploration with applications to structural failure

Taiki Matsumura · Raphael T. Haftka · Nam H. Kim

Received: 24 September 2013 / Revised: 18 March 2014 / Accepted: 7 May 2014 / Published online: 21 June 2014

Abstract This paper investigates an effective test strategy for structural failure criterion characterization. The goals include identification of potential failure modes and a better approximation of failure boundary, e.g., failure load mapping with respect to geometry and load conditions. We typically replicate and test the same structural configuration several times in order to deal with noisy observation. However, our study shows that replication is not necessarily needed, because of the smoothing effect of surrogate models, and we show that exploring with as many different configurations as possible is more important. We illustrate the failure criterion characterization with two structural examples with various surrogate models, including polynomial response surface (PRS), support vector regression (SVR) and Gaussian process regression (GPR). We also examine the treatment of replicated test data for surrogate fitting. While fitting to all replicated test data works well for GPR, fitting only to the mean values of the replicated data helps a not-well-tuned surrogate (SVR in this paper) by compensating for the proneness to overfitting. When the noise level is significant as compared to the error due to surrogate modeling, a denser matrix might be prone to overfitting for GPR and SVR.

Keywords Failure load approximation · Test · Surrogate models · Resource allocation · Noisy observation

T. Matsumura · R. T. Haftka · N. H. Kim (✉)
Department of Mechanical & Aerospace Engineering, University of Florida, Gainesville, FL, USA
e-mail: nkim@ufl.edu

T. Matsumura
e-mail: taikimatsumura@gmail.com

R. T. Haftka
e-mail: haftka@ufl.edu

1 Introduction

For structural design, appropriate characterization of failure criteria is critical. The main objectives of failure criterion characterization are (1) identifying underlying failure modes and (2) constructing an accurate design allowable chart for each failure mode, e.g., failure load map with respect to geometry and load conditions. Inaccurate characterization may lead to structural designs that would experience unexpected failure modes and suffer large errors.

Using analytical failure theories, such as Tsai-Wu and von-Mises, is a reasonable approach for well-known materials and structures. However, analytical theories may not be reliable enough for newly introduced materials, e.g., composite materials (Ilcewicz and Murphy 2005), and for structural elements due to lack of knowledge. Therefore, failure criterion characterization often relies on experimental approaches.

In experimental approaches, we typically conduct a series of tests for a structural element for a particular use. To discover potential failure modes, it is important to explore within the design space with as many different structural configurations as possible. At the same time, failure boundary mapping, e.g., failure load, is carried out by fitting a surrogate to the observed test data (Department of Defense 2002a, b). Because the test results are noisy due to variability in material properties, test conditions, and errors in measurement devices, we often replicate the same structural configurations for statistical analysis of the observations.

Since tests are costly, these processes need to be accomplished under a budgetary constraint, i.e., a limited number of tests. Then there arises a resource allocation problem between exploration and replication. This paper studies effective resource allocation of tests for

failure criteria characterization. We are particularly interested in the question of whether we can take advantage of the smoothing effect of surrogate techniques, equivalent to a noise filter by replication. The same problem of exploration versus replication is also encountered in fitting surrogates to noisy simulations, such as the probability calculated from Monte Carlo simulations. So in the next section we review the pertinent literature in this field too.

In this paper, we illustrate the failure criterion characterization using two example structural elements. Each structural element has two potential failure modes, one of which dominates the design space. The less dominant failure mode is considered as an un-modeled failure mode when it is missed by the test matrix. The failure load map of the dominant mode is assumed to be approximated by using test data and surrogate models.

With the help of the examples, we will discuss effective test matrix strategies from the following perspectives: (1) exploration or replication, (2) the use of replicated data; that is, using the mean value for fitting or not, (3) the effect of nonlinearity of the failure load surface, and (4) the effect of noise level. To do that, one of the structural examples has a simple failure load surface, and the other has a highly nonlinear surface. For both examples, the observed data is subjected to uncertainty due to material properties and geometry. We test different types of surrogate models, including polynomial response surface, Gaussian process regression, and support vector regression.

This paper is organized as follows. Section 2 discusses the use of surrogate models for noisy data including literature review and test strategy for failure criterion characterization. Section 3 briefly describes the formulations of the surrogate models. Section 4 illustrates the example structural problems, followed by the results discussed in Section 5. Finally, concluding remarks are addressed in Section 6.

2 Fitting noisy observations and test strategy

2.1 Surrogate models for noisy observations

Surrogate models have been widely used as an alternative of expensive computer simulations and experiments in engineering design. Their usefulness has extensively been studied and proven by a variety of applications, such as design optimization (Queipo et al. 2005; Jin et al. 2003; Simpson et al. 2004; Forrester and Keane 2009) and uncertainty quantification (Giunta et al. 2006; Eldred et al. 2002; Kim and Choi 2008). They have also been used for experimental optimization (Chaudhuri et al. 2013; Viana et al. 2011). In fact, response surface techniques were originally developed in order to optimize crop yields by changing inputs such as water and nutrients, based on measured crop yields in subplots with different inputs.

Fitting surrogate models to noisy data is important for engineering applications because of their capability on reducing underlying uncertainty Computational models may suffer from noise due to discretization errors and incomplete convergence of iterative simulation such as optimization and Monte Carlo simulation. Experimental observations vary due to variability in material properties and geometry of test specimens and variability in test conditions, e.g., boundary conditions. There are studies on surrogate models dealing with noisy data from various aspects, such as performance of approximation and design optimization (Giunta et al. 1994; Papila and Haftka 2000). However, limited research has been done on sampling strategy, also known as design of experiments (DOE), for noisy data especially on the effects of repeated observations, which is the scope of this paper.

Studies on optimal DOE for PRS to noisy observations date back to the middle of the 20th century (Elfving 1952; Kiefer 1961; St. John and Draper 1975). Early studies focused only on simple polynomial regression and sought for an appropriate allocation of samples considering repetitions according to optimality criteria; e.g., D-optimality. Later, an empirical investigation into optimal allocation with Kriging is conducted (Picheny 2009). Picheny's study found that, for a two-dimensional function with uniformly distributed noise over the space, having a large number of observations from high-noise simulations provided a better result in terms of the prediction variance (Myers and Montgomery 1995) than a small number of observations from a low-noise simulation. More recently, an adaptive design strategy for Kriging-based design optimization against noisy data has been investigated (Picheny et al. 2013a, b).

Note that the performance metrics used for these studies on DOE are associated with the level of confidence of the prediction models, such as D-optimality and prediction variance. The confidence level of prediction does not necessarily represent the error against the true function, in which the designers working on failure criterion characterization are primarily interested. In fact, a DOE that is optimized by a particular criterion might poorly perform for other performance measures (Goel et al. 2008). For example, Goel et al. showed that D-optimal design, which minimizes the maximum of variance of coefficients, may have a large bias error.

In this paper, we extend the previous work and empirically investigate the effective test matrix for a practical and specific engineering application, i.e., structural failure boundary characterization using surrogate models. In practical engineering applications, the magnitude of noise in

observation may not be constant over the range of design space because it may depend on geometry etc. In terms of the performance of surrogate models, the error in a failure load prediction directly corresponds to the error in design allowable of a structure. Therefore, we focus on the accuracy of surrogate models to the true functions rather than the confidence level of surrogate prediction. Furthermore, we investigate the treatment of replicated data for fitting; that is whether we fit a surrogate model to all replicated data or fit to the mean values of repeated data. Finally, for surrogate models to be examined we include support vector regression (SVR) as well as PRS and Gaussian process regression (GPR).

2.2 Test strategy for failure criterion characterization

A major role of characterizing tests is to refine failure criteria of structural elements, which will be used to determine design allowables. The process is conducted at an early stage of design process as part of so-called buildingblock testing approach (Marshall Rouse et al. 2005). The test specimens limit to lower complexity structures, such as structural joints and composite laminate plates. Lower level of complexity allows the designers to easily analyze observed data, understand failure mechanisms, and construct prediction models.

The expected variations in design parameters, such as external conditions and geometry of a structure need to be taken into account to determine the test configuration. A matrix of tests is usually deployed in order to cover the entire design space and evaluate the effects of all possible combinations of the design parameters. A matrix of tests is also called a full factorial design (Forrester et al. 2008).

It is obvious that increasing the density of matrix (level of factorial design) is more likely to locate underlying failure modes, allowing the designers to investigate and obtain insights into the failure mechanisms. For example, a design handbook for composite materials issued by the U.S. Department of Defense (DoD) addresses the importance of selection of structural configurations, quoted, "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" (Department of Defense 2002b). In addition, the handbook recommends replicating tests for statistical data analysis.

A key question is how we should allocate the limited number of tests whether to replication or exploration. Figure 1 illustrates a tradeoff for up to 50 tests. If the designer more cares about the noise rather than about missing failure modes, 4×4 matrix with three replications would be an option. 7×7 matrix without any replication is an alternative in search of underlying potential failure modes. Since it is obvious that exploration is important for failure mode identification, our main interest is how replication contributes to reducing the error in failure load approximation.

3 Surrogate models

In this section, we summarize the formulations of surrogate models, including polynomial response surface (PRS), Gaussian process regression (GPR), and support vector regression (SVR).

3.1 Polynomial response surface (PRS)

Polynomial response surface uses polynomial functions and the least squares fit to approximate a true function. Let a prediction of output y be $\hat{y}$, and x_* be a location where we predict y. $\hat{y}$ is expressed as a linear combination of polynomial functions as

$$\hat{y}(x_*) = \sum_{i=1}^{k} \beta_i \xi_i(x_*) \tag{1}$$

where $\xi_i(x)$ are basis functions, typically monomials β_i represent coefficients, and k is the number of coefficients (Myers and Montgomery 1995).

As the true function is represented by $y = \hat{y} + \varepsilon$, where ε is the error, PRS smoothens the noise (Giunta and Watson 1998; Jin et al. 2001). PRS is also known for its computational tractability. Because of polynomial functions being applied, it may cause a problem when being fitted to functions not approximated well by polynomials.

Since the basis function is typically unknown and we need to choose an appropriate one, we will also discuss the selection of the basis function using some metrics, such as the standard error and cross validation error, i.e., leave-one-out cross validation called PRESS (prediction of residual error sum of square). We used the Surrogate Tool Box (Viana 2010) which refers to the Matlab 'regress' routine for the fitting (MATLAB and Statistics Toolbox Release 2012).

3.2 Gaussian process regression (GPR)

Gaussian process regression is originally developed as a method of the spatial statistics (Rasmussen and Williams 2005). A special type of GPR is also known as Kriging (Sacks et al. 1989). GPR views a set of data points as a collection of random variables that follow some rule of correlation, called random process, defined by (2). The name of Gaussian process originates from the form of random process using multivariate normal (Gaussian) distribution:

$$y \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}_i, \boldsymbol{x}_j)) \tag{2}$$

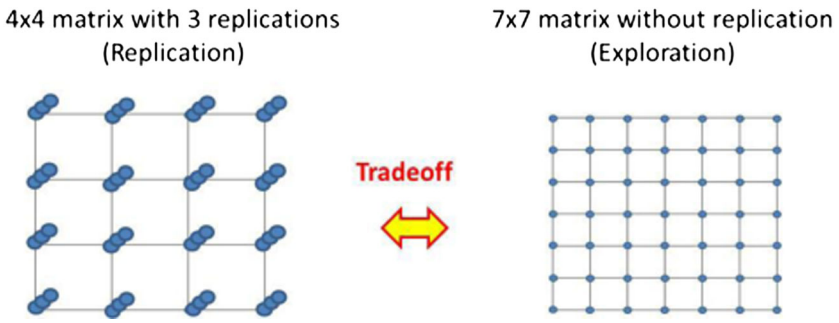


Fig. 1 Tradeoff between replication and exploration given up to 50 tests

where m (x) is the mean function of Gaussian process,, and $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ represents the correlation between points. In this paper, we use a linear model for the mean function and the Gaussian correlation model, which is the most commonly used for engineering applications, expressed as

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sigma^2 \exp\left(-\sum_{l=1}^{d} \frac{\left|x_i^l - x_j^l\right|^2}{2\theta_l^2}\right) \tag{3}$$

where σ^2 is the process variance with zero mean and θ_l is the scaling parameter of the l-th component of $\boldsymbol{x}$ in d dimension which determines the correlation between two points.

The prediction $\hat{y}$ at a new location $\boldsymbol{x_*}$ is assumed to be a realization of the random process that is identified by given N observations, $\boldsymbol{y} = (y_1, y_2, ..., y_N)^{\boldsymbol{T}}$ and $\boldsymbol{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\}$, and is obtained by the following function.

$$\hat{y} = m(\boldsymbol{x_*})|\boldsymbol{y}, \boldsymbol{X} = m(\boldsymbol{x}_*) + \boldsymbol{K}(\boldsymbol{x}_*, \boldsymbol{X})$$

$$\times \left(\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \boldsymbol{\Delta}_{\boldsymbol{\sigma_n^2}}\right)^{-1} (\boldsymbol{y} - m(\boldsymbol{x}_*)\boldsymbol{1}) \tag{4}$$

with $\boldsymbol{K}(\boldsymbol{x}_*, \boldsymbol{X}) = [k(\boldsymbol{x}_*, \boldsymbol{x_1})\ k(\boldsymbol{x}_*, \boldsymbol{x_2})\ ...\ k(\boldsymbol{x}_*, \boldsymbol{x_N})]$, and (i,j) component of $\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X})$ is $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. $\boldsymbol{1}$ is an $N \times 1$ vector of ones. $\Delta_{\boldsymbol{\sigma_n^2}}$ is a diagonal matrix of diagonal terms σ_n^2. Noise variance σ_n^2 with zero mean is assumed to be independent of σ and enables us to deal with replicated data. The hyperparameters, $(\boldsymbol{x})$, θ_l, σ, and σ_n are selected such that the likelihood of observing $\boldsymbol{y}$ is maximum. In analogy with PRS, this process corresponds to the selection of monomials.

Equation (4) is equivalent to determine N coefficients of the radial basis functions $\boldsymbol{K}(\boldsymbol{x}_*, \boldsymbol{X})$. In case there exists replication, the total number of coefficients is reduced by a factor of the number of replications because the radial basis functions corresponding to the replicated points are the same. The advantage of GPR is the flexibility of fitting to nonlinear functions. However, the fitting process of GPR is time consuming due to the optimization of the hyperparameters.

We use the Gaussian Process Regression and Classification Toolbox version 3.2. (Rasmussen and Williams 2005) for the implementation. Since the toolbox deploys a line-search method for the optimization of the hyperparameters, the optimal solution tends to depend on starting points of the search. To avoid ending up with a local optimum, we use 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 $exp\left(-\sum_{l=1}^{d} \frac{\left|x_i^{(l)} - x_j^{(l)}\right|^2}{2\theta_l^2}\right) = 0.9$ for the closest two points among the training points assuming that the nearby points should be highly correlated. After fitting with all the combinations of starting point, we select the best model based on the maximum likelihood.

3.3 Support vector regression (SVR)

Support vector regression evolved from a machine learning algorithm (Smola and Scholkopf 2004; Clarke et al. 2005; Vapnik 1998). SVR balances the flatness of the approximated function and the residual error by a user-define parameter called regularization parameter C. A unique aspect of SVR is an explicit treatment of noise by a user supplied error tolerance ε. That is, only differences between the fit and the data that are larger than ε are minimized. Figure 2 illustrates one of the most common models for the error tolerance, so-called ε-insensitive loss function. When the error $(= \hat{y}_i - y_i)$ is within the tolerance $\pm\varepsilon$, the loss (ξ) is zero, otherwise the loss is proportionally increased with the error. We use the ε-insensitive loss function for the study.

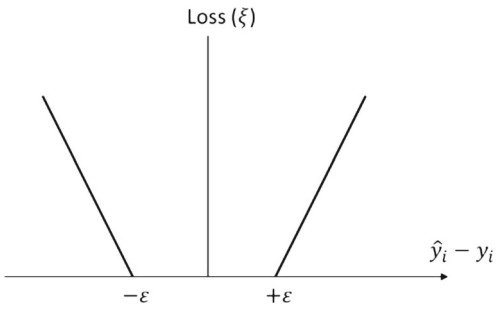


Fig. 2 ε-sensitive loss function for SVR

The prediction model at a new location $\boldsymbol{x}_*$ can be expressed by using the Lagrange multipliers, α_i and α_i^* of the two constraints in (5) as

$$\hat{y}(\boldsymbol{x}_*) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \boldsymbol{x}_i^T \boldsymbol{x}_* + b \tag{5}$$

where $\boldsymbol{x_i}$ represents the i-th training point.

There potentially are N parameters to be optimized (N sets of Lagrange multipliers). In case we have replication, since all the replicated points have the same term $\boldsymbol{x}_i^T \boldsymbol{x}_*$, the number of parameters to be optimized essentially reduced by a factor of the number of replications, like GPR. Furthermore, the prediction model is only determined by the training points corresponding to non-zero Lagrange multipliers, called support vectors. When ε-insensitive loss function is used, support vectors correspond to the training points being located outside of the error tolerance.

For the implementation, we use the Surrogate Tool Box (Viana 2010), which uses the MATLAB code offered by Gunn (1998). One of the challenges of SVR is to select an appropriate set of parameters. In general, for regularization parameter, substantially large C is suggested (Jordaan and Smits 2002) and the exact selection of C "is not overly critical" (Forrester et al. 2008) and "has only negligible effect on the generalization performance" (Cherkassky and Ma 2004). Based on some experiments seeking accurate fitting for the example problems, we use infinity for C. For the kernel function, we use Gaussian model as shown in (6), which is commonly used, with θ being also a user-defined parameter.

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

In the same manner discussed for GPR, because nearby points should be smoothly connected, and after some experimentation, we selected θ such that $k = 0.9$ for the closest two points. For ε, which is suggested to be close to the level of noise (Cherkassky and Mulier 2006), we use the average standard deviation of the observed data from 7×7 matrix with 7 replications. We consider it a practical assumption because the designer can get some idea about the noise level from the observation.

For the selection of hyperparameters, i.e., C, ε, and θ, using cross validation is suggested (Forrester et al. 2008; Smola and Scholkopf 2004; Basudhar 2012). However, identifying the best hyperparameters is out of the scope of this paper and testing a number of combinations of the hyperparameters is computationally intractable. Instead we will discuss how the parameter selection affects the performance of the approximation.

4 Example problems

In order to illustrate practical failure criterion characterization, we chose two simple examples for clarity and to allow exhaustive study of a large number of strategies. The examples are a 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. The composite laminate plate has a high order of nonlinearity of the failure load surface, while the support bracket has a smooth and almost linear surface. The following subsections describe the example structures, test matrix, treatment of the replicated data for approximation, and error evaluation for analyzing the results.

4.1 Support bracket

A simple support bracket mounted on a base structure is shown in Fig. 3. The load is applied to 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 l and length a of the bracket are fixed due to space constraints. The diameter d of the cylindrical part is considered as a design parameter. Table 1 shows the geometry and its variabillity 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 4 illustrates the critical failure modes of the structure. Because of the additive effect of the torsion and transverse shear stresses or bending and axial stresses, the stress at point D is likely to exceed the strength.

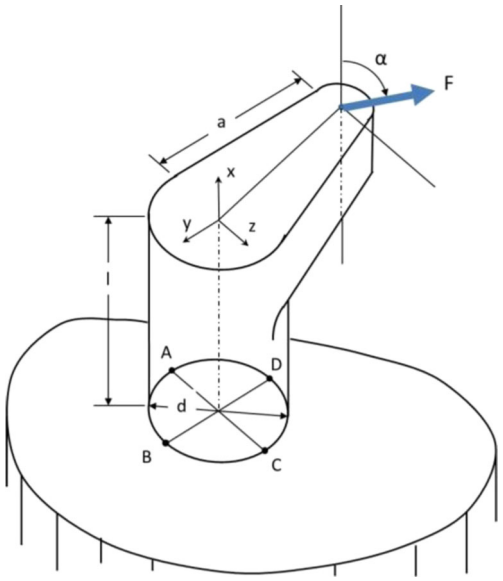


Fig. 3 Support bracket

Table 1 Properties of support bracket

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

However, point A can be a critical point under some conditions as shown in Fig. 4. If the designer fails to locate the failure mode initiated at 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 the noise in test observations. The 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 point D. Figure 5 depicts the failure load surface corresponding to point D, which will be approximated by the surrogate models.

4.2 Composite laminate plate

For the second example, intended to have a more complex failure load surface, a symmetric composite laminate with three ply angles $[0°/-\theta/+\theta]$s is considered (Fig. 6). The laminate is subject to mechanical loading along x and

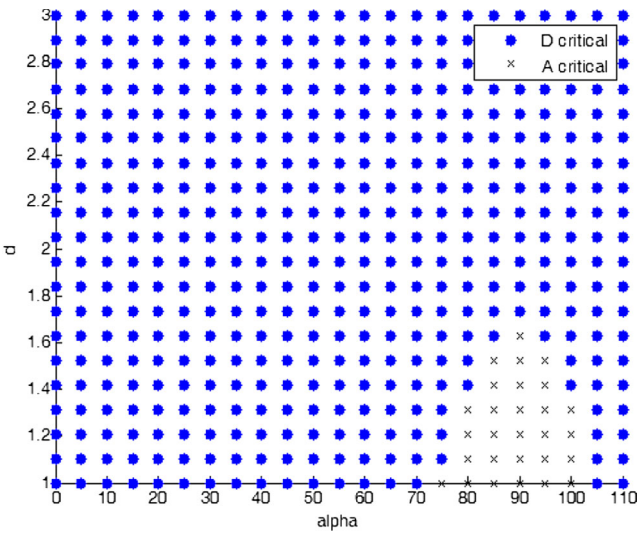


Fig. 4 Critical failure modes of support bracket. A *blue circle* indicates that failure initiates at point D, while a *cross* indicates initial failure at point A

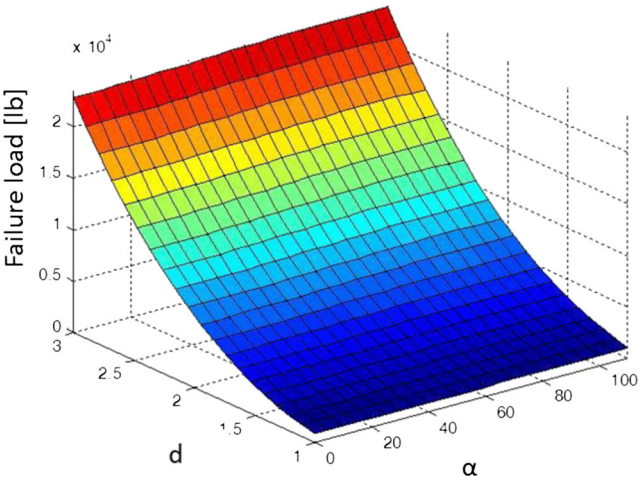


Fig. 5 Failure load surface of support bracket initiated at point D

y directions defined by the load ratio α, such that $N_x = (1-\alpha)F$ and $N_y = \alpha F$. 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 $\varepsilon_{1\text{allow}}$, transverse direction $\varepsilon_{2\text{allow}}$, and shear $\gamma_{12\text{allow}}$. All the properties are assumed to be normally distributed and the sources of noise in test observations. The strains are predicted by the Classical Lamination Theory (CLT).

Figure 7 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 8 is the failure load surface due to ply axial strain.

4.3 Test matrix and fitting strategy

Our test matrices range from 4 × 4 to 7 × 7 with evenly spaced test points in order to investigate the effect of the density of matrix on the accuracy of approximation. For

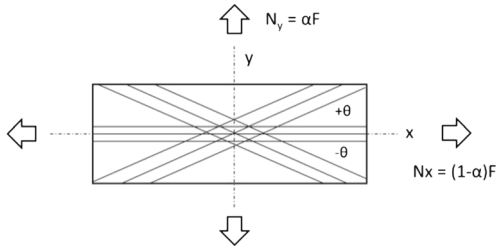


Fig. 6 Composite laminate plate

Table 2 Properties of composite laminate plate

Property	Quantity	
	Mean	CV
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
$\varepsilon_{1\text{allow}}$	$\pm$0.01	6 %
$\varepsilon_{2\text{allow}}$	$\pm$0.01	6 %
$\gamma_{12\text{allow}}$	$\pm$0.015	6 %

each test matrix, we replicate the same test configuration up to seven times. Table 3 shows the total number of tests for the matrices. For both structural examples, 5×5 test matrix or denser ones will detect the less dominant failure modes; therefore obviously 4×4 matrix is not a desirable option. We compare the following two strategies for fitting the surrogate models. Note that both strategies provide the same result for PRS.

(1) All-at-once fitting strategy
The surrogate models are fitted to all test data including the replicated ones.

(2) Mean fitting strategy
The mean values of the replicated data are taken first at each location in the design space. Then, the surrogate models are fitted to the means

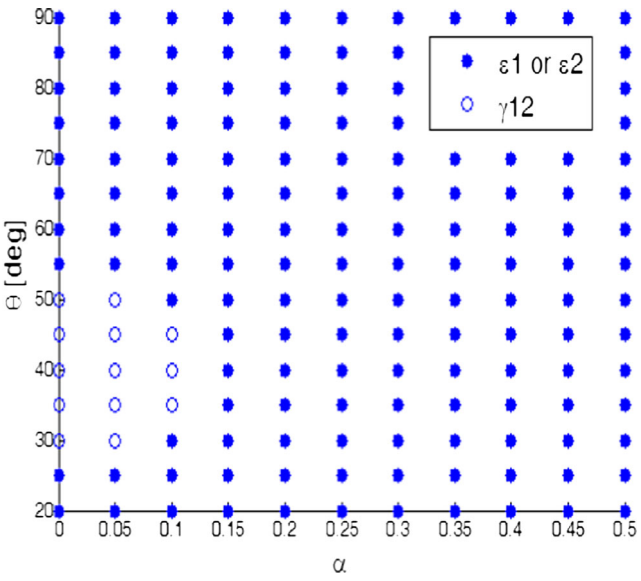


Fig. 7 Critical failure modes of composite laminate plate. A *solid circle* indicates the failure due to strain along fiber direction ε_1 or strain transverse along fiber direction ε_2, while an *empty circle* indicates the failure due to shear strain γ_{12}

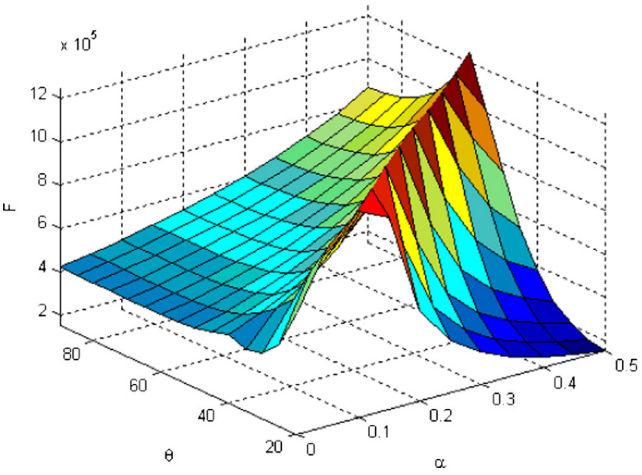


Fig. 8 Failure load surface of composite laminate plate due to axial strain

4.4 Error evaluation

In order to evaluate the accuracy of the failure load mapping, we compare the surrogate predictions with the true values from 20×20 matrix of test points (in total 400 points). We define the true values of the failure load as the mean of infinite number of test observations. For the support bracket, this corresponds to the mean value of the material yield strength (only uncertainty in the problem and linear relationship of the failure load). For the composite laminate plate, since multiple sources of uncertainty are involved, we estimate the true mean values from 10,000 random samples. The standard errors of the sample means is less than 0.5 % of the range of the failure load, which is small compared to the surrogate errors.

To measure overall accuracy, we use the root mean square error normalized by the range of failure loads (NRMSE) calculated by (7). We also evaluate the normalized maximum absolute error calculated by (8) in order to examine the robustness of the surrogate models.

(a) Normalized root mean square error (NRMSE)

$$NRMSE = \frac{1}{\text{range of } y}\sqrt{\frac{1}{400}\sum_{i=1}^{400}(\hat{y}_i - y_i)^2} \tag{7}$$

(b) Normalized maximum absolute error (NMAE)

$$NMAE = \frac{1}{\text{range of } y}\max(|\hat{y}_1 - y_1|, |\hat{y}_2 - y_2|, ..., |\hat{y}_{400} - y_{400}|) \tag{8}$$

We first produce the failure loads corresponding to a set of randomly generated input structural and geometrical properties for a particular matrix of experiments Then,

Table 3 Test matrix and total number of tests

Matrix	Number of replications						
	1	2	3	4	5	6	7
4 × 4	16	32	48	64	80	96	112
5 × 5	25	50	75	100	125	150	175
6 × 6	36	72	108	144	180	216	252
7 × 7	49	98	147	196	245	294	343

we fit the surrogate models to the failure loads, and evaluate NRMSE and NMAE. We repeat this fitting process 100 times, each with a different set of the random inputs, failure loads and failure load approximation.

The errors discussed in the following section are the mean values over 100 fits. The standard errors of the mean values over 100 runs are, on average, less than 0.1 % for NRMSE and less than 0.4 % for NMAE, which means that only differences between surrogate models that are substantially larger than these standard errors are statistically significant.

5 Results

5.1 Support bracket

We first discuss the results of surrogate models fitted to the almost linear surface of failure load of support bracket (Fig. 5) For PRS, 1st, 2nd, and 3rd order polynomial functions were fitted, and then, 2nd order PRS was selected as the best ones based on the leave-one-out cross validation, PRESS. PRESS predicted well the best polynomial functions that offer smallest NRMSE except for the cases of 7 × 7 matrix with 6 and 7 replications. In other words, PRESS properly warned that 3rd and 4th order PRSs overfitted noise. For SVR, the error tolerance ε is selected at 936 lb as the average of the noise level (one standard deviation) from 87 lb to 2421 lb Note that 936 lb corresponds to 4.1 % of the range of failure load.

To examine the resource allocation (replication vs. exploration), Figs. 9 and 10 show NRMSE and NMAE of the three surrogate models with respect to the total number of tests when the all-at-once fitting strategy is applied (for the details of test matrix, see Table 3). For PRS and GPR, all four curves corresponding to the densities of matrix (from 4 × 4 matrix to 7 × 7 matrix) form a single curve in NRMSE. This means that replication and exploration contribute equally to improving the accuracy of approximation.

For PRS, this trend is supported by the behavior of the standard error predicted by PRS, which represents the unbiased estimator of noise variance Fig. 11 shows the boxplot of standard errors of the various test strategies with given about 50 tests, including 7 × 7 matrix without replication, 5 × 5 matrix with 2 replications and 4 × 4 with 3 replications. It can be seen that the medians and variations of the standard

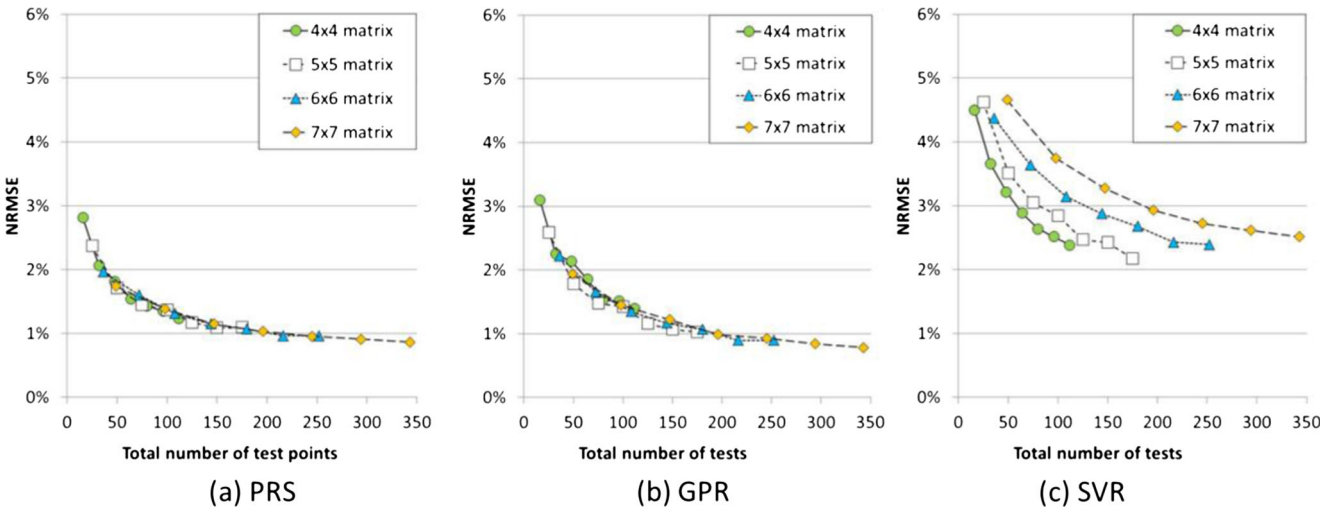


Fig. 9 Error comparison for support bracket: NRMSE for all-at-once fitting strategy. Markers of each line correspond to one replication through seven replications from left to right (see Table 3 for details on the numbers of tests). The average standard errors of the means of NRMSE are 0.05 % (PRS), 0.06 % (GPR) and 0.05 % (SVR)

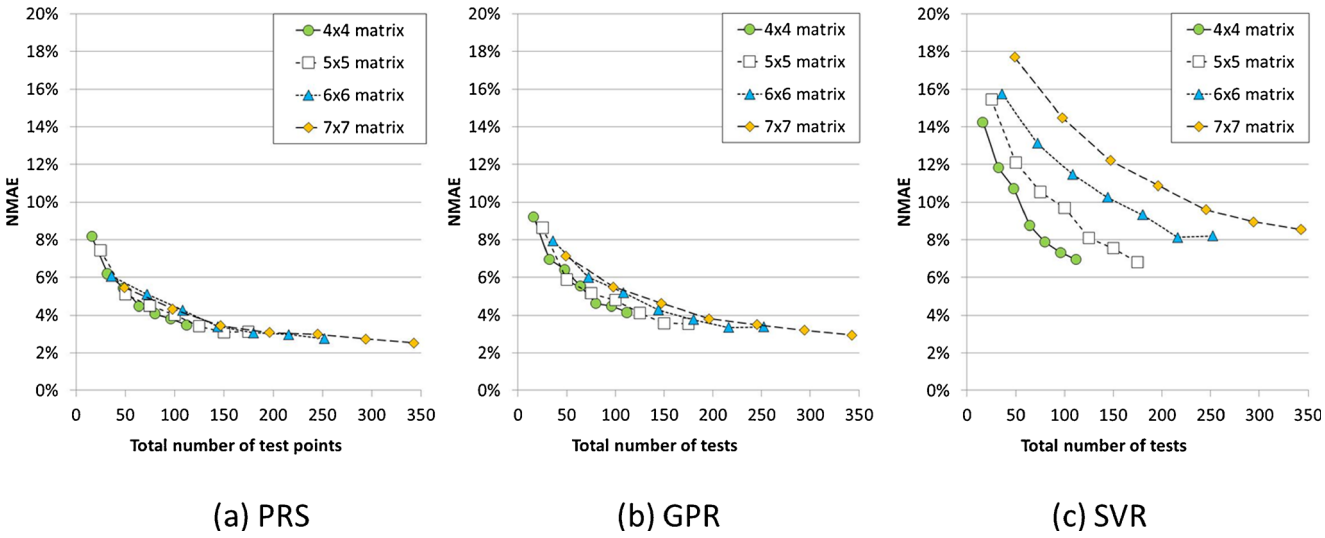


Fig. 10 Error comparison for support bracket: NMAE for all-at-once fitting strategy. Markers of each line correspond to one replication through seven replications from left to right. The average standard errors of the means of NRMSE are 0.18 % (PRS), 0.25 % (GPR) and 0.30 % (SVR)

errors were almost the same, indicating that whether it is replication or exploration did not matter from the standpoint of noise prediction.

In order to identify the causes of error, Table 4 compares the errors fitted to noise-free data and that to noisy data with 7×7 matrix. NRMSE for noise-free data purely represents the modeling error of surrogate. For example, PRS has a 0.6 % error for noise-free data, but the error increased to 1.7 % when the noise is introduced. Similarly, for other surrogate models, most of the errors are due to noise rather than the modeling error.

From these observations, for PRS and GPR there was no significant advantage of replication over exploration in terms of the accuracy of approximation. This leads us to conclude that exploration is more important than replication for this example in the context of failure criterion characterization where we are subjected to search for potential failure modes.

SVR, on the other hand, shows different trends and underperforms PRS and GPR. This may reflects the fact that SVR particularly applied to this problem is less sensitive than PRS to very large errors. While PRS minimizes L2 error norm (i.e., NRMSE) by the least square method, SVR with ε-insensitive loss function minimizes L1 error norm. This loss function does not penalize at all small errors, and does not emphasize the effect of the largest errors by squaring them In fact, as seen in Table 4 L1 norm considering the error tolerance, defined by the following equation, of SVR (0.6 %) is substantially smaller than those of others (more than 4 %):

$$\varepsilon \text{ sensitive error} = \frac{1}{(\text{range of } y)} \frac{1}{N} \sum_{\hat{y}_i - y_i^{obs} > \varepsilon} \times \left| \left(\hat{y}_i - y_i^{obs} \right) - \varepsilon \right| \tag{9}$$

where N is the number of test points in which $\hat{y}_i - y_i^{obs} > \varepsilon$. Note that ε of SVR is set to be zero when SVR was fitted to noise-free data.

It is also observed that the accuracy of SVR deteriorates as the test matrix becomes denser both in NRMSE (Fig. 9) and NMAE (Fig. 10). GPR showed a similar trend in NMAE

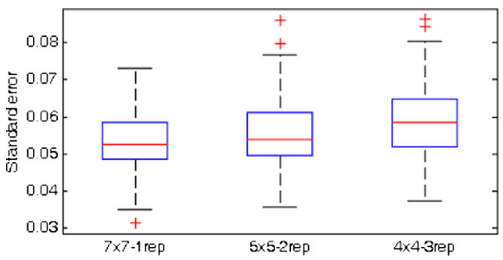


Fig. 11 Standard errors predicted by PRS for support bracket for given about 50 tests

Table 4 Errors fitted to noise-free data and sensitive errors (support bracket)

		Noise-free data 7×7 matrix	Noisy data 7×7 matrix with no replication
NRMSE	PRS	0.6 %	1.7 %
	GPR	0.0 %	1.9 %
	SVR	0.1 %	4.8 %
ε sensitive error	PRS	–	4.3 %
	GPR	–	4.0 %
	SVR	–	0.6 %

(Fig. 10), but it was not as significant as that of SVR. This is explained mainly by the ratio of the number of training points and the number of parameters that determine the prediction model—it is called 'parameter ratio' in this paper. If the parameter ratio is small, i.e. the number of parameters to be tuned is larger than the number of training points, the regression model has a danger of overfitting (Hawkins 2004).

Table 5 shows the numbers of parameters and the parameter ratios of all the surrogate models fitted to for given about 50 tests. For example, 2nd order PRS for two input variables has six coefficients regardless of the test matrix If it is fitted to 4×4 matrix with 3 replications (48 points), the ratio is 8 ($= 48/6$). For GPR, as discussed earlier, the number of parameters is the same as the number of training points, leading to the parameter to be 3, which is substantially smaller than that of PRS. In case of replication, where the total number of parameters is reduced by a factor of the number of replications, the parameter ratio becomes even smaller and is 1. Similarly, SVR, for which the number of support vectors determines the parameters, has the smaller parameters. These ratios account for the accuracy deterioration of GPR and SVR with respect to matrix density.

The poor performance of SVR may also be attributed to not-well-tuned parameters, i.e., ε, C, and θ, unlike GPR which optimizes all the hyperparameters by the maximum likelihood estimator. Our numerical experiments showed that tuning the parameters of SVR improved the accuracy and alleviated the accuracy deterioration with a denser matrix. Figure 12 illustrates that increasing the correlation coefficient for the nearest two points from 0.9 (original set) to 0.99 reduced the error. This makes sense because a higher correlation governs a larger area of the space and makes the fitting curve flatter and less sensitive to noise In fact, the optimized correlation coefficients by GPR turned out to be higher than 0.9. Figure 12 also shows that tuning the regularization parameter C, which is originally set as infinity, improved the performance.

Next we investigated the treatment of replicated data for fitting. Figure 13 compares the all-at-once fitting and the mean fitting for GPR in terms of NRMSE. It is shown that there was no significant difference between the two fitting strategies. From the stand point of the parameter ratio, the mean fitting has a higher risk of overfitting. As shown in Table 5 the parameter ratio of the mean fitting is always 1 and smaller than the ratios of the all-at-once fitting. On the other hand, taking the mean values essentially reduces the noise level, thereby less threat of overfitting. It seems that these two phenomena cancel out each other.

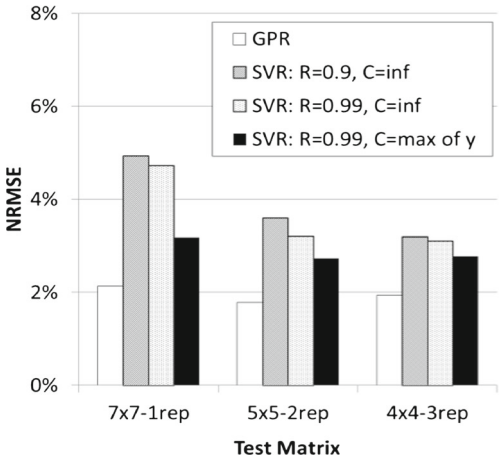


Fig. 12 Performance of SVR with various combinations of C and R. C is the regularization parameter, and R is the correlation coefficient for the closest two points. 'Max of y' represents the maximum value of the failure load

Figure 14 shows that the mean fitting performed substantially better than the all-at-once fitting for SVR. On top of the smaller parameter ratio, the hyperparameters of SVR are not well-tuned. In such a case, reducing the noise by taking the means was more influential than the proneness to overfitting. Note that we adjusted the error tolerance ε when the mean fitting is applied in order to account for the accuracy of the estimator of sample mean. For the adjustment, the original ε was divided by the square root of the number of replications ($\varepsilon_{adjusted} = \varepsilon/\sqrt{N_{rep}}$). It is also noteworthy that using the original ε for the mean fitting offered a poorer performance than the all-at-once fitting.

Table 5 Ratio between the number of training data points and the number of parameters for given about 50 tests (support bracket)

Fitting strategy	Test matrix	PRS	GPR	SVR
All-at-once fitting	4×4-3 reps	8.0 (6)	3 (16)	3.4 (13.9)
	5×5-2 reps	8.3 (6)	2 (25)	2.5 (20.3)
	7×7-1 rep	8.2 (6)	1 (49)	1.7 (29.1)
Mean fitting	4×4-3 reps	–	1 (16)	1.5 (8.3)
	5×5-2 reps	–	1 (25)	1.7 (11.4)

The numbers in parentheses represent the number of parameters to be optimized for fitting. All numbers are the mean over 100 fits

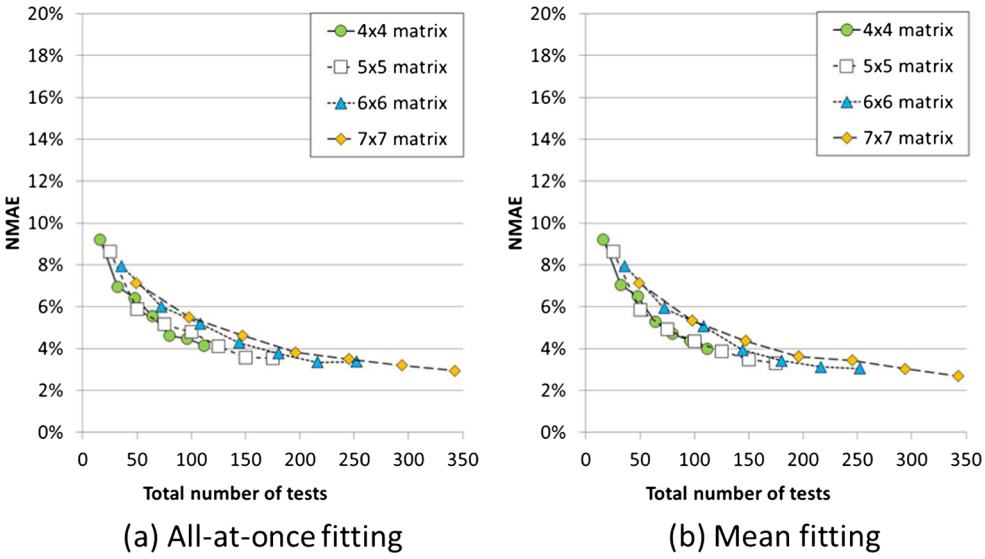


Fig. 13 Comparison of fitting strategy: NRMSE of GPR for support bracket. Markers of each line correspond to one replication through seven replications from left to right. The average standard error of the means of NRMSE is 0.07 %

5.2 Composite laminate plate

The fitting performance to a complex failure load surface of the composite laminate plate is discussed in this section For PRS, various orders of polynomial functions ranging from 2nd order to 8th order were tested and the best models listed in Table 6 are selected based on NRMSE. For the selection, we limited the number of coefficients of the polynomial functions to be smaller than the number of training points as the least square fit typically assumes (Myers and Montgomery 1995). Since NRMSE is unknown in reality, we will discuss the selection of best PRS in the next subsection.

Figures 15 and 16 show NRMSE and NMAE of all the surrogate models with the all-at-once fitting. It can be seen that the 4×4 matrix did not capture well the failure load surface for all the surrogate models. Once 5×5 or denser matrix is available, the accuracy was substantially improved. An important observation from these two figures is that the contribution of replication to reducing the error is miniscule compared to that of increasing the density of matrix For example, with PRS, the error of 7×7 matrix without replication (total 49 tests) was smaller than that of 6×6 matrix with 7 replications (total 252 tests).

In terms of the error sources Table 7 compares the errors of the surrogate models fitted to the noise-free data and those to the noisy data. Obviously, most of the errors come from the modeling error rather than the noise for all surrogate models. Even for SVR, the error for noise-free data

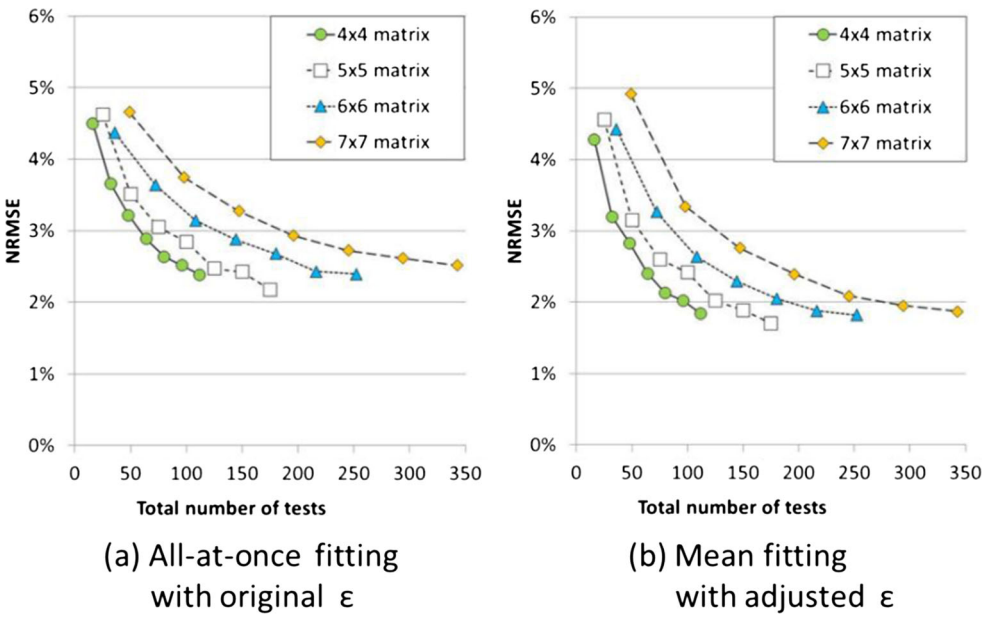


Fig. 14 Comparison of fitting strategy: NRMSE of SVR for support bracket. Markers of each line correspond to one replication through seven replications from left to right. The average standard error of the means of NRMSE is 0.29 %

Table 6 Best polynomial order for PRS based on NRMSE for composite laminate plate

Test matrix		Best function	Test matrix		Best function
4 × 4	1 rep	4th	6 × 6	1 rep	7th
	2 rep	4th		2 rep	7th
	3 rep	4th		3 rep	7th
	4 rep	4th		4 rep	7th
	5 rep	4th		5 rep	7th
	6 rep	4th		6 rep	7th
	7 rep	4th		7 rep	7th
5 × 5	1 rep	5th	7 × 7	1 rep	8th
	2 rep	5th		2 rep	8th
	3 rep	5th		3 rep	8th
	4 rep	5th		4 rep	8th
	5 rep	5th		5 rep	8th
	6 rep	5th		6 rep	8th
	7 rep	5th		7 rep	8th

was worse than the average error over 100 fits to noisy data. Table 8 shows the parameter ratios for given about 50 tests. SVR has support vectors, which are located out of the error tolerance, at almost all test points. For example, with 7 × 7 matrix without replication, there are support vectors at 46.6 locations on average out of 49 locations. This also indicates that the modeling error is significant.

As a conclusion, it can be said that exploring different locations was found to be more important than replicating tests at the same location when modeling error was dominant For PRS, this was also supported by the standard error which followed the same trend.

As for the treatment of replicated data, Fig. 17 compares different fitting strategies for GPR. Unlike the support bracket problem, the all-at-once fitting clearly outperformed the mean fitting. It seems that informing as many observations including replicated ones (though they suffer from noise) improved the accuracy As discussed in the Section 2.1, this result is consistent with the previous work conducted by Picheny (2009) on Kriging by assessing the prediction variance. He experimentally demonstrated that having a larger number of different observations from high-noise simulation provided a smaller prediction variance than a smaller number of observations from a lower-noise

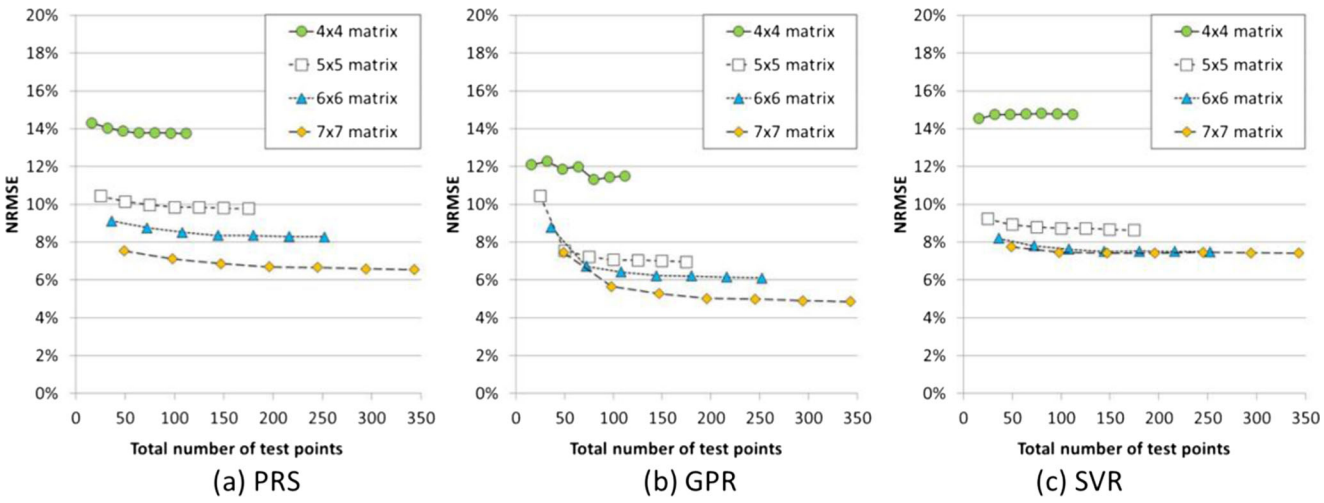


Fig. 15 Error comparison for composite laminate plate: NRMSE for all-at-once fitting strategy. Markers of each line correspond to one replication through seven replications from left to right. The average standard errors of the mean of NRMSE are 0.05 % (PRS), 0.08 % (GPR) and 0.07 % (SVR)

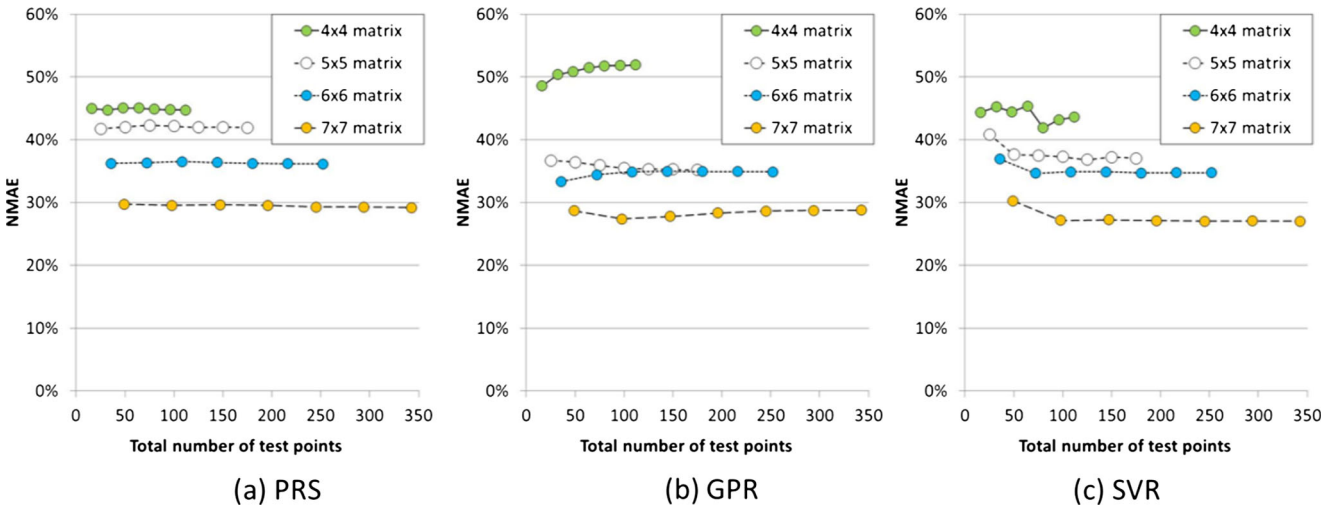


Fig. 16 Error comparison for composite laminate plate: NMAE for all-at-once fitting strategy. Markers of each line correspond to one replication through seven replications from left to right. The average standard errors of the mean of NRMSE are 0.29 % (PRS), 0.34 % (GPR) and 0.27 % (SVR)

simulation For SVR, Fig. 18 illustrates that the mean fitting still helped to compensate for the not-well-tuned model as discussed for the support bracket problem.

Since the errors discussed this section are the average out of 100 fits, we were interested in whether exploration always outperforms over replication for each of fitting iterations. Test strategies for given about 100 tests are examined, including 7×7 matrix with two replications (98 tests), 6×6 matrix with three replications (108 tests), and 5×5 matrix with four replications (100 tests). Table 9 depicts how many times which test strategy performed better than others. Except for the comparison of SVR between 7×7 matrix and 6×6 matrix, whose performances were very comparable (Fig. 15), a denser test matrix steadily provided a more accurate approximation at almost more than 90 %.

Finally, one might wonder which surrogate model should be chosen among others. We examined whether the PRESS predicted the best surrogate for each of the fitting iterations of 7×7 matrix with 2 replications. Table 10 shows the selections of PRESS and the correct rate PRESS well predicted that GPR was better than PRS at the correct rate of 98 %. As for the comparison to SVR, PRESS seemed to work well. However, we found that PRESS of SVR was substantially higher than those of other surrogate models, and that is why PRESS always predicted that SVR was inferior to others. This reflects the fact that the SVR used in this paper is very sensitive to noise.

5.3 Selection of best PRS for composite laminate plate

Finally we discuss the selection of the best polynomial order of PRS. Table 11 compares the best polynomial order based on NRMSE, the prediction of the best function by PRESS and by the standard error (SE). With 4×4 matrix, neither PRESS nor SE predicted properly the best order. With the denser matrices, they predicted correctly. Nonetheless PRESS always failed to predict in the case of no replication and tended to pick up a smaller order of polynomial function compared to the actual best. As shown in Table 8, PRS fitted to no-replicated data was prone to overfitting because of the smaller parameter ratio. This made PRESS of a higher

Table 7 Errors of surrogate models fitted to noise-free data (composite laminate plate)

	Error		
	Noise-free data (7×7 matrix)	Noisy data (7×7 matrix with no replication)	Difference
PRS	7.30 %	7.54 %	0.24 %
GPR	7.02 %	7.45 %	0.35 %
SVR	8.00 %	7.75 %	0.25 %

Table 8 Ratio between the number of training data points and the number of parameters for given about 50 tests (composite laminate plate)

Fitting strategy	Test matrix	PRS	GPR	SVR
All-at-once	4 × 4-3 reps	3.2 (15)	3 (16)	3.0 (16.0)
	5 × 5-2 reps	2.4 (21)	2 (25)	2.0 (24.9)
	7 × 7-1 rep	1.1 (45)	1 (49)	1.1 (46.6)
Mean	4 × 4-3reps	–	1 (16)	1 (16)
	5 × 5-2 reps	–	1 (25)	1 (25)

The numbers in parentheses represent the number of parameters to be optimized for fitting. All numbers are the mean over 100 fits

order of polynomial order larger, and in turn, PRESS preferred a lower order of polynomial function. SE, on the other hand, steadily predicted the best orders or at least a reasonably higher order of polynomial function even when there was no replication.

Interestingly, in many cases, we observed that the best PRS models violated the requirement that the number of coefficients be smaller than the number of observations. For example, for 5 × 5 matrix (25 different observations in the input space), 6th order PRS (28 coefficients, hence undetermined system) offered the minimum NRMSE. Matlab 'regress' function (MATLAB and Statistics Toolbox Release 2012) handles such undetermined problems by ignoring linearly dependent column vectors of $\boldsymbol{X}$ by applying the Householder QR decomposition. The QR decomposition tends to choose column vectors of $\boldsymbol{X}$ that have a higher norm as linearly independent columns. As a result, for the 6th order PRS, after the dependent column vectors being ignored, the highest monomial x^6 appeared to remain, giving some flexibility to the fitting curve and offering a better accuracy. It should be noted that the selection of monomials highly depends on how we normalize the input space. Note that in this paper, we normalized the input variables from 0 to 1. Another important warning about this behavior is that it is very hard to identify such best function for the undetermined problem either by PRESS, standard error, or R-square In other words, PRESS for undetermined system might be misleading.

5.4 Effects of dimensionality and noise level on effective test strategy

An important consideration is the dimensionality of input space; that is, if the conclusions in the previous section hold for higher dimension problems. As the dimension increases, the sampling data become relatively sparse. When the failure load surface is simple, as discussed in the first example, there is no clear advantage of investing in replication over exploration for a sparse data matrix. When the failure load surface is highly nonlinear, as demonstrated in the second example, it is obvious that increasing the density of matrix is the choice. Thus, exploration is more effective than replication regardless of dimensionality.

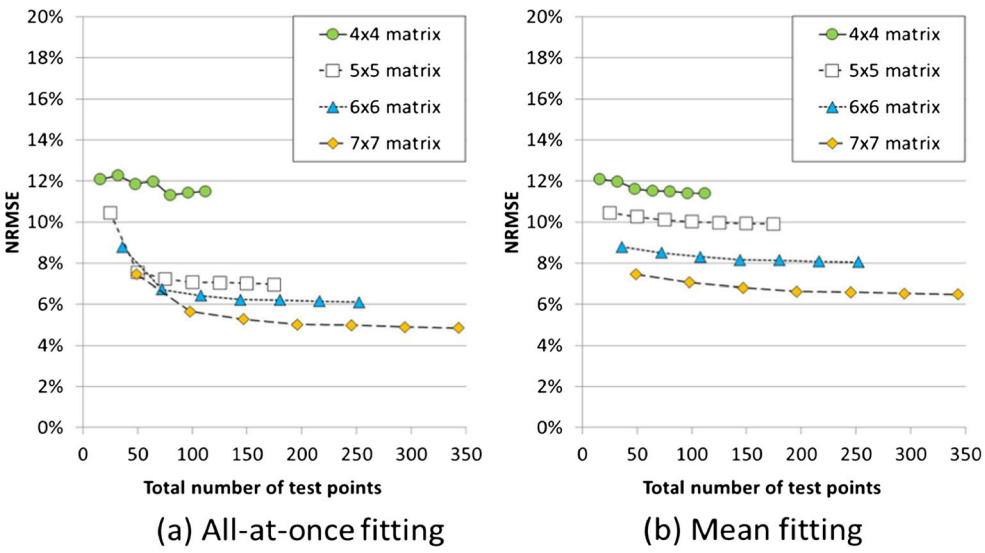


Fig. 17 Comparison of fitting strategy: NRMSE of GPR for composite laminate plate. Markers of each line correspond to one replication through seven replications from left to right. The average standard error of the means of NRMSE is 0.07 %

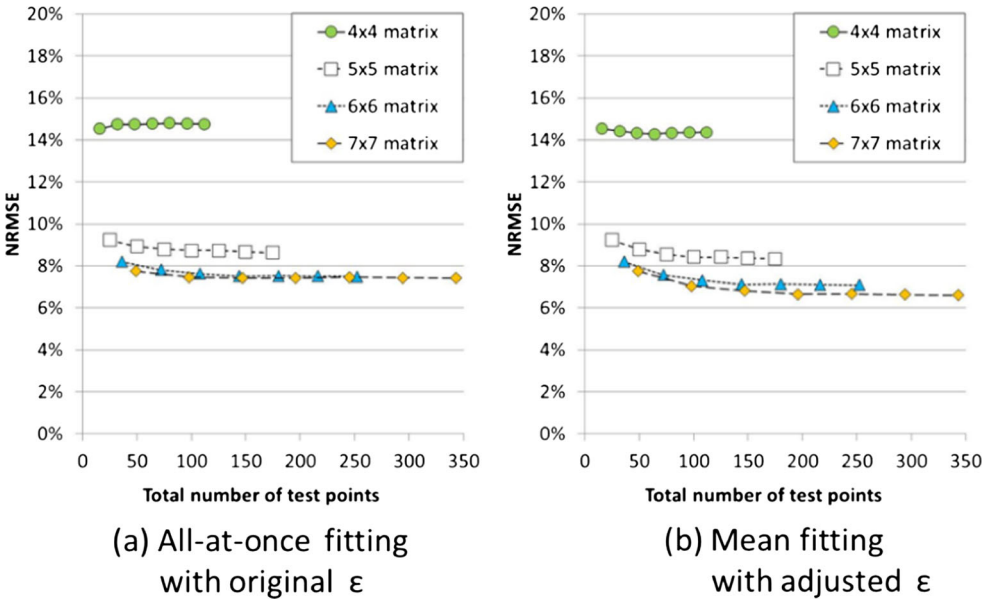


Fig. 18 Comparison of fitting strategy: NRMSE of SVR for composite laminate plate. Markers of each line correspond to one replication through seven replications from left to right. The average standard error of the means of NRMSE is 0.07 %

Another important issue is whether the noise level affects the selection of test strategy. As shown in the second example, when the error due to surrogate modeling is dominant as compared to the size of noise, the noise plays a very small role, and exploration is important to improve the accuracy. On the other hand, when the noise is a dominant source of the error, as in the first example, GPR and SVR with a denser test matrix is prone to overfitting because of their high parameter ratio. So polynomial response surfaces may be preferable. If SVR or GPR are used, the density of test matrix should be sufficient enough to capture the surface but not too high to avoid overfitting. PRS is less sensitive to noise as long as an appropriate order of polynomial function is selected, and for that, SE seemed to be better than PRESS.

6 Concluding remarks

We investigated an effective test strategy for failure criterion characterization, focusing on allocation of tests to

Table 9 Performance comparison between test matrices for each fitting iteration for given about 100 tests (Composite laminate plate)

	Probability of having better accuracy		
	7 × 7 matrix (2 replications)	6 × 6 matrix (3 replications)	5 × 5 matrix (4 replications)
PRS	93 %	7 %	–
	99 %	–	1 %
	–	99 %	1 %
GPR with all-at-once fitting	89 %	11 %	–
	90 %	–	10 %
	–	86 %	14 %
SVR with mean fitting	69 %	31 %	–
	93 %	–	7 %
	–	93 %	7 %

Surrogate models are fitted to 100 sets of random test observation. For each set of the test observation, NRMSEs of the surrogate models are compared

Table 10 Surrogate selection by PRESS (Composite laminate plate)

Test	Prediction of PRESS and actual result		Correct rate
GPR with all-at-once fitting is better than PRS	True/Positive	98	98 %
	False/Positive	2	
	False/negative	0	
	True/Negative	0	
PRS is better than SVR with mean fitting	True/Positive	72	72 %
	False/Positive	28	
	False/negative	0	
	True/Negative	0	
GPR with all-at-once fitting is better than SVR with mean fitting	True/Positive	98	98 %
	False/Positive	2	
	False/negative	0	
	True/Negative	0	

The surrogate models fitted to 7×7 matrix with 2 replications are compared

replication or exploration. For approximating the failure load surfaces, polynomial response surface (PRS), Gaussian process regression (GPR), and support vector regression (SVR) were examined. With the illustration of two structural element examples, we conclude that replication is not necessarily needed and that exploration is more important both for discovering underlying failure modes and for the accuracy of failure load approximation.

For the example with almost linear failure load surface the noise in observations was significant compared to the error in surrogate modeling, and then replication and exploration contributed equally to reducing the error for PRS and GPR. Then, the tests should be spent for exploration for the purpose of discovering unexpected failure modes.

On the other hand, for the example with complicated nonlinear failure load surface, in which the error in surrogate model is dominant, exploration was clearly more important for all surrogate models both for capturing the behavior of the surface and for identifying unexpected failure modes.

Table 11 Best polynomial functions for PRS predicted by PRESS and SE for composite laminate plate

Test matrix		Best NRMSE	Prediction		Test matrix		Best NRMSE	Prediction	
			PRESS	SE				PRESS	SE
4×4	1 rep	4th	2nd	3rd	6×6	1 rep	7th	3rd	5th
	2 rep	4th	3rd	3rd		2 rep	7th	7th	7th
	3 rep	4th	3rd	4th		3 rep	7th	7th	7th
	4 rep	4th	3rd	4th		4 rep	7th	7th	7th
	5 rep	4th	4th	4th		5 rep	7th	7th	7th
	6 rep	4th	4th	4th		6 rep	7th	7th	7th
	7 rep	4th	4th	4th		7 rep	7th	7th	7th
5×5	1 rep	5th	2nd	5th	7×7	1 rep	8th	5th	8th
	2 rep	5th	5th	5th		2 rep	8th	8th	8th
	3 rep	5th	5th	5th		3 rep	8th	8th	8th
	4 rep	5th	5th	5th		4 rep	8th	8th	8th
	5 rep	5th	5th	5th		5 rep	8th	8th	8th
	6 rep	5th	5th	5th		6 rep	8th	8th	8th
	7 rep	5th	5th	5th		7 rep	8th	8th	8th

Highlighted cells represent that the predictions are wrong

High dimensionality in input space can be considered to be equivalent to sparseness of space filling, which harms the accuracy of surrogate models and the chance of spotting unknown failure modes. When the failure load surface is highly nonlinear, exploration is a more effective approach to tackle high dimensionality than replication. An important warning here is that exploration may encounter the issue of overfitting when the noise level is high, as illustrated in the second example with GPR and SVR. We also examined two different treatments of replicated data for surrogate fitting: (1) fitting a surrogate to all replicated data simultaneously and (2) fitting only to the mean values of replicated data. While the all-at-once fitting outperformed the mean fitting for GPR The mean fitting compensated for the proneness of the not-well-tuned SVR to overfitting by reducing the noise. We also found that it was important to adjust the error tolerance ε for the mean fitting to account for the accuracy of sample mean estimator.

Finally, we addressed the issue that the least square fit provided by Matlab might offer a better solution when the number of coefficients is larger than the number of test points (undetermined problem). However, the solution varies depending on how we normalize input space and the performance metrics, such as PRESS, might not be reliable.

Acknowledgments The authors greatly appreciate the support by National Science Foundation grant CMMI-1131103.

References

Basudhar A (2012) Selection of anisotropic kernel parameters using multiple surrogate information. Paper presented at the 14th AIAA/ISSMO multidisciplinary analysis and optimization conference, Indianapolis, Indiana, 17–19 September

Chaudhuri A, Haftka RT, Ifju P, Villanueva D, Chang K, Rue J, Tyler C, Schmitz T (2013) Experimental optimization and uncertainty quantification of flapping wing of a micro air vehicle. Paper presented at the 10th World congress on structural and multidisciplinary optimization, Orlando, FL, May 20–24

Cherkassky V, Ma YQ (2004) Practical selection of SVM parameters and noise estimation for SVM regression. Neural Netw 17(1):113–126. doi:10.1016/S0893-6080(03)00169-2

Cherkassky V, Mulier F (2006) Learning from data: concepts, theory, and methods. Wiley, Hoboken

Clarke SM, Griebsch JH, Simpson TW (2005) Analysis of support vector regression for approximation of complex engineering analyses. J Mech Des 127(6):1077–1087. doi:10.1115/1.1897403

Department of Defense (2002a) Department of Defense Handbook, Composite materials handbook, vol 1, Chapter 8 MIL-HDBK-17-3F

Department of Defense (2002b) Department of Defense Handbook, Composite materials handbook, vol 3, Chapter 4 MIL-HDBK-17-3F

Eldred MS, Giunta AA, Wojtkiewicz SF, Trucano TG (2002) Formulations for surrogate-based optimization under uncertainty. Paper presented at the 9th AIAA/ISSMO symposium on multidisciplinary analysis and optimization, Atlanta, GA, 4–6 September

Elfving G (1952) Optimum allocation in linear regression theory. Ann Math Stat 23(2):255–262. doi:10.1214/aoms/1177729442

Forrester AIJ, Keane AJ (2009) Recent advances in surrogate-based optimization. Prog Aerosp Sci 45(1–3):50–79

Forrester AIJ, Sóbester A, Keane AJ (2008) Engineering design via surrogate modelling: a practical guide. Wiley

Giunta AA, Dudley JM, Narducci R, Grossman B, Haftka RT, Mason WH, Watson LT (1994) Noisy aerodynamic response and smooth approximations in HSCT design. In: Paper presented at the 5th symposium on multidisciplinary analysis and optimization, Panama City Beach, FL, September 7–9

Giunta AA, Mcfarland JM, Swiler LP, Eldred MS (2006) The promise and peril of uncertainty quantification using response surface approximations. Struct Infrastruct E 2(3–4):175–189. doi:10.1080/15732470600590507

Giunta AA, Watson LT (1998) A comparison of approximation modeling techniques: polynomial versus interpolating models. Paper presented at the 7th AIAA/USAF/NASA/ISSMO symposium on multidisciplinary analysis and optimization, St. Louis, MO, September 2–4

Goel T, Haftka RT, Shyy W, Watson LT (2008) Pitfalls of using a single criterion for selecting experimental designs. Int J Numer Methods Eng 75(2):127–155. doi:10.1002/Nme.2242

Gunn SR (1998) Support vector machines for classification and regression. University of Southampton

Hawkins DM (2004) The problem of overfitting. J Chem Inf Comput Sci 44(1):1–12. doi:10.1021/ci0342472

Ilcewicz L, Murphy B (2005) Safety & certification initiatives for composite airframe structure. Paper presented at the 46th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference, Austin, Texas, April 18–21

Jin R, Chen W, Simpson TW (2001) Comparative studies of metamodelling techniques under multiple modelling criteria. Struct Multidiscip Optim 23(1):1–13

Jin R, Du X, Chen W (2003) The use of metamodeling techniques for optimization under uncertainty. Struct Multidiscip Optim 25(2):99–116. doi:10.1007/s00158-002-0277-0

Jordaan EM, Smits GF (2002) Estimation of the regularization parameter for support vector regression. Paper presented at the World conference on computational intelligence 2002, Honolulu, Hawaii, May 12–17

Kiefer J (1961) Optimum designs in regression problems .2. Ann Math Stat 32(1):298–325. doi:10.1214/aoms/1177705160

Kim C, Choi KK (2008) Reliability-based design optimization using response surface method with prediction interval estimation. J Mech Des 130(12):121401–121412. Artn 121401. doi:10.1115/1.2988476

Marshall Rouse DCJ, McGowan DM, Bush HG, Allen Waters W (2005) Utilization of the building-block approach in structural mechanics research. Paper presented at the 46th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics & materials conference, Austin, Texas, 18–21 April

MATLAB and Statistics Toolbox Release (2012) The MathWorks, Inc., Natick, Massachusetts

Myers RH, Montgomery DC (1995) Response surface methodology: process and product optimization using designed experiments. Wiley, New York

Papila M, Haftka RT (2000) Response surface approximations: noise, error repair, and modeling errors. AIAA J 38(12):2336–2343. doi:10.2514/2.903

Picheny V (2009) Improving accuracy and compensating for uncertainty in surrogate modeling. University of Florida, Gainesville

Picheny V, Ginsbourger D, Richet Y, Caplin G (2013a) Quantile-based optimization of noisy computer experiments with tunable precision. Technometrics 55(1):2–13. doi:10.1080/00401706.2012.707580

Picheny V, Wanger T, Ginsbourger D (2013b) A benchmark of kriging-based infill criteria for noisy. Struct Multidiscip Optim (in press). doi:10.1007/s00158-013-0919-4

Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Tucker PK (2005) Surrogate-based analysis and optimization. Prog Aerosp Sci 41(1):1–28. doi:10.1016/j.paerosci.2005.02.001

Rasmussen CE, Williams CKI (2005) Gaussian processes in machine learning. MIT Press, Cambridge

Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. Stat Sci 4(4):409–423

Simpson TW, Booker AJ, Ghosh D, Giunta AA, Koch PN, Yang RJ (2004) Approximation methods in multidisciplinary analysis and optimization: a panel discussion. Struct Multidiscip Optim 27(5):302–313. doi:10.1007/s00158-004-0389-9

Smola AJ, Scholkopf B (2004) A tutorial on support vector regression. Stat Comput 14(3):199–222. doi:10.1023/B:Stco.0000035301.49549.88

St. John RC, Draper NR (1975) D-optimality for regression design—review. Technometrics 17(1):15–23

Vapnik VN (1998) Statistical learning theory. Wiley-Interscience, New York

Viana FAC (2010) Surrogates toolbox user's guide (Ver. 2.1). Gainesville

Viana FAC, Haftka RT, Hamman R, Venter G (2011) Efficient global optimization with experimental data: revisiting the paper helicopter design. Paper presented at the 52nd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference, Denver, Colorado, 4–7 April