Error Amplification in Failure Probability Estimates of Small Errors in Surrogates

Palaniappan Ramu, Nam H. Kim and Raphael T. Haftka
University of Florida

ABSTRACT

Response surface methods which approximate the actual performance function using simple algebraic equations are widely used in structural reliability studies. The response surface approximations are often used to estimate the reliability of a structure. Errors in the response surface approximation affect the results of reliability analysis. This work investigates the error in the failure probability estimated using a response surface approximation. It is observed that small errors in the response surface may amplify to large errors in the failure probability. It is observed that the amplification occurs when the failure surface is far away from the response mean and the DOE has more points near the mean. Another situation is when the failure region is a small island encompassed within the safe region, and the points in the DOE fail to capture the failure region. Analytical and engineering application examples are investigated to understand the amplification of error in the failure probability.

INTRODUCTION

In structural design, safety measures are used to gauge the safety level of the structure. The design process involves several parameters, such as types of loadings, material properties, geometry, mathematical approximations, failure modes, etc. Uncertainties in these parameters are inevitable. Traditionally, deterministic approaches used safety factors to account for the uncertainties (Elishakoff, 2004). Later, probabilistic approaches were introduced in structural optimization to account for uncertain variables.

Polynomial response surface approximations are often used to alleviate the computational expense in reliability studies(Rajashekhar and Ellingwood, 1993, Venter et al, 1998, Kutaran et al, 2002). Reliability studies require the assessment of the performance function that describes the behavior of the system, which is implicit or too complex for explicit evaluation in most real applications. Especially for complex problems where Monte Carlo Simulation (MCS) is the only feasible approach, the polynomial response surface approximations greatly lessen the computational burden by substituting calculation of a polynomial instead of an expensive physical simulation.

The quality of the response surface is judged by various error metrics. The response surface is used to predict the response values at extrapolated regions; i.e., regions outside the design of experiments (DOE). When the response surface is used to predict failure probability, the magnitude of failure probability is often low and it requires information in the tail part of the distribution rather than in the central part. Here, careful implementation of the response surface is required, especially when it is used for extrapolation. It has been observed that sometimes the response surfaces poorly approximate the actual behavior at the extrapolated regions in spite of good error metrics at the interpolated region. At times, these short comings in the response surface approximations contribute to amplification of errors in failure probability estimates.

The objective of this paper is to investigate the amplification of error in failure probability estimates and understand the circumstances that trigger this phenomenon so that use of polynomial response surfaces will be able to exercise extra caution under these circumstances. The paper is structured as follows; failure probability computation is discussed in Section 2. Section 3 discusses response surface approximations and error amplifications in failure probability estimates. The error amplification phenomenon is described with examples and followed by discussion in Section 4. Section 5 discusses future work.

FAILURE PROBABILITY COMPUTATION

In structural reliability, safety of structure depends on the value the performance function G. When $G > 0$, the structure is safe and considered failed otherwise. The safety of a structure is expressed in terms of failure probability. MCS is widely used to estimate failure probability because of the ease of its use and robust

nature. The failure probability is computed using the expression:

$$P_f \approx \frac{\text{num}(G(\hat{\mathbf{x}}) \leq 0)}{N} \tag{1}$$

where, P_f is the failure probability, $\hat{\mathbf{x}}$ is the randomly chosen sample point, $G(\hat{\mathbf{x}})$ is the performance function which describes the behavior of the structure, $\text{num}(G(\hat{\mathbf{x}}) \leq 0)$ denotes the number of samples for which $(G(\hat{\mathbf{x}}) \leq 0)$ and N is the total number of samples. $G(\hat{\mathbf{x}})$ is mostly approximated using a response surface. Essentially, MCS assigns a value of 1 or 0 to each sample depending on whether it violates the performance function or not. Summation of these values for all the samples and dividing it by the total number of samples provides the failure probability. The accuracy of the failure probability depends on the number of samples. For a fixed number of samples, the accuracy deteriorates with a decrease in the actual failure probability. Hence, estimating a very low failure probability with a good accuracy requires a huge sample size. The coefficient of variation (COV) of the computed failure probability is given by:

$$COV = \frac{\sigma_p}{P_f} = \sqrt{\frac{(1 - P_f)}{P_f N}} \tag{2}$$

Usually, researchers use the failure probability computed using MCS as a standard for comparing the failure probabilities computed using other methods and use it in optimization.

RESPONSE SURFACE APPROXIMATIONS AND ERROR AMPLIFICATIONS IN FAILURE PROBABILITY ESTIMATES

Researchers rely heavily on response surface approximations for two reasons:

1. When the performance function is not available explicitly as an analytical expression. In this case, responses for several combinations of the design variables are generated and polynomial response surface (PRS) is constructed in terms of the design variables.

2. To reduce the computational expense in evaluating a high fidelity model, especially in optimization problems in which the model need to be analyzed repetitively. Response surface approximations using low order polynomials are used to capture the nature of the problem and are used in the analyses.

Moreover, numerical noise in the parameters involved in the optimization introduces additional difficulties in computing sensitivities. Response surface approxima-

tions are used in these situations to approximate the response using a smooth function.

Although polynomial response surface approximations have been successfully used for failure probability estimation and optimization, they might reduce the accuracy. The quality of a polynomial response surface approximation is given by its error metrics. It is observed and demonstrated in this work that the measures of accuracy commonly used for response surface approximations do not necessarily correspond to accuracy in failure probability. That is even if the error metrics of the response surface are good, the prediction of failure probability estimated based on the response surface might not be accurate. The error metrics used in this work are described in Appendix.

Two numerical examples are treated in this work (i) Cantilever beam example and (ii) Branin-Hoo function. It is noted that there are two situations in which low failure probabilities can occur. First, the failure surface is far from the mean value. Second, the failure region itself is small contributing to the low failure probability. The cantilever beam example is representative of the former case. The Branin-Hoo function example has an island failure region. That is, the failure region is encompassed within the safe region. It represents the latter case with multiple failure regions. Two different DOEs are used in each example (i) Latin Hypercube Sampling (LHS) with a normal distribution (ii) Orthogonal array (OA). The first one covers a small region near the mean. If the failure region is far away from the mean, the response surface needs to extrapolate substantially and it leads to error amplification. On the other hand, for the OA DOE, the response surface approximation is not accurate because it covers a larger region and the error in failure probability is not due of amplification of errors.

A number of 10^5 samples are used to estimate the failure probability. The trend of the cumulative distribution function (CDF) of the response surface with respect to that of the actual response is monitored by counting the number of the samples (out of 10^5) that falls in the equally spaced bins (in terms of limit state values). The trend (overestimate or underestimate) of the response surface is consistent around the mean of the response but, there is a reversal in the trend in the tails of the distribution. This is clearly explained in the examples.

Since response surfaces are widely used to approximate the response of a structure and hence to compute the failure probability, this paper attempts to caution the researchers about the possible amplification of errors in failure probability estimate, though the error metrics are good. Moreover, a reversal in the trend near the tails is observed in most cases. These can lead to an entirely different estimation of the required safety measure.

NUMERICAL EXAMPLES

(I) CANTILEVERED BEAM
The widely used cantilever beam example for reliability analysis introduced by Wu *et al*. (2001) presented in Figure 1 is used for demonstration purpose. The beam is under the horizontal and vertical loads at the tip.
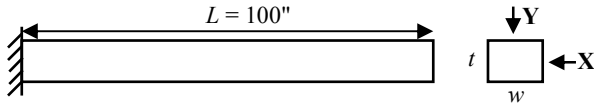

Figure 1   Cantilevered beam subject to horizontal and vertical loads

The performance function is the difference between allowable and the actual deflections at the tip. From traditional beam theory, the performance function can be defined as:

$$G_d = D_0 - D = D_0 - \frac{4L^3}{Ewt}\sqrt{\left(\left(\frac{Y}{t^2}\right)^2 + \left(\frac{X}{w^2}\right)^2\right)} \quad (3)$$

where Y is the transverse load = 500lb, D_0 is the allowable initial deflection taken as 2.5", and w = 2.6535 and t = 3.9792 are dimensions of the cross-section obtained through reliability-based optimization.

In order to illustrate the quality of the response surface in reliability analysis, it is assumed that the horizontal load X and the elastic modulus E are random variables. Even if other parameters are also uncertain, we chose only two random variables for illustration purpose. The distribution is shown in Table 1.

A quadratic polynomial response surface is used in this study to approximate the tip deflection. There are two DOE considered: (i) LHS with a normal distribution and (ii) Orthogonal Arrays. In both cases 25 data points are used to fit the response surface approximation. Results using the LHS case are provided first followed by those using the OA. Failure occurs when G_d is less than zero.

Table 1   Random variables for the beam problem

Random variable	X (lb)	E (psi)
Distribution	Normal	Normal
(mean, std. dev.)	(700, 100)	(29E6, 5E6)

CASE 1 – LHS with a normal distribution DOE

The cloud of test points that are generated based on distribution of the random variables X and E is shown in Figure 2 along with the LHS samples (data points) represented by the stars. Based on the responses at the DOE samples, the response surface is constructed. Tables 2 and 3 show statistics at data points and test

points, respectively. It is seen that the statistics of the response surface are close to that of the actual function.

The error metrics presented in Table 4 show that the R^2 error is close to 1 and since there is no large difference between the R^2 and R^2_{adj}, it is evident that there is no unnecessary terms in the response surface. The RMS errors are to be compared to the range presented in Table 2 which also shows that the fit is reasonably good. However, the PRESS_RMS error is higher than the RMS errors. The R^2 predicted using the PRESS error matches the R^2 well.

In spite of a good statistics and error metrics, Figure 3 clearly depicts the inaccuracy in the failure probability estimation. The failure probability estimated by the two models presented in Table 5 differs largely. The difference cannot be attributed to sampling errors. They are also not consistent with the error metrics presented in Table 4. In order to explain the error, 18 equal-width bins (in terms of the function values) are used and the number of samples that fall in each bin is monitored. These results are presented in Table 6. Though the quality of the response surface was good based on the error metrics presented in Table 3, the failure probability estimation at the tails suffer from large errors. This is due to the fact that the failure surface is away from the mean and the response surface is accurate in the central region of the distribution of the response and the accuracy deteriorates in the tails of the distributions. It is noted that when CDF is considered, the errors are amplified in the left tail. In the right tail, the cumulative effect of negative and positive differences cancels out. The 6% error in PDF in the fourth bin translates to 70% error in the CDF. But after that, the error in the CDF decreases very quickly.
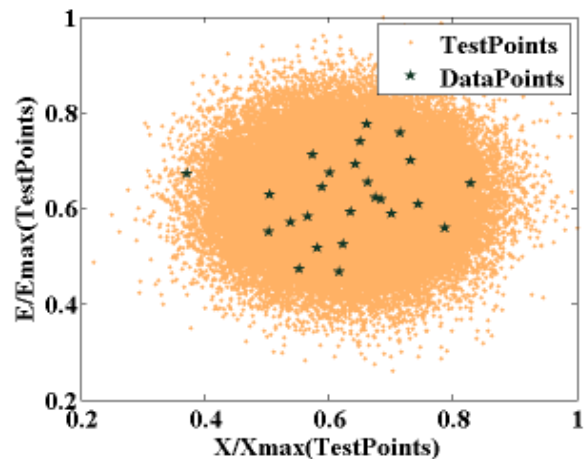

Figure 2   Cantilever beam. LHS data points and cloud of 1e5 test points.

Table 2 Cantilever beam. Statistics at LHS data points

	y_{act}	y_{rs}	$y_{act} - y_{rs}$
Mean	0.66	0.66	0
Std. Dev.	0.14	0.14	0
Range	0.63	0.63	

Table 3 Cantilever beam. Test points statistics (Case 1)

	y_{act}	y_{rs}	$y_{act} - y_{rs}$
Mean	0.58	0.65	-0.07
Std. Dev.	0.14	0.16	0.02

Table 4 Cantilever beam. Error metrics for the response surface . LHS DOE (see Appendix for metric definition)

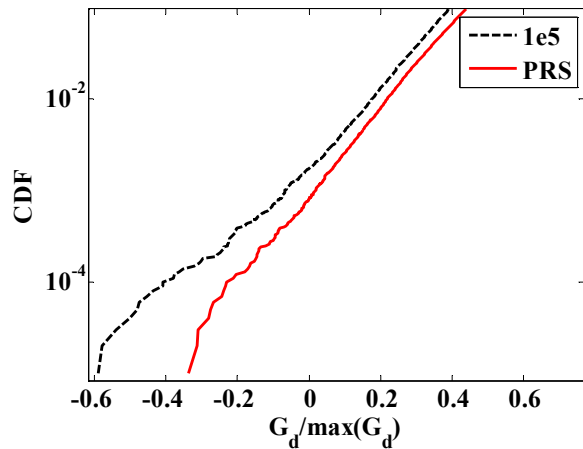| Error Metric	Value
R^2	0.99980
R^2_{adj}	0.99974
RMS	0.00201
RMS_{pred}	0.00231
PRESS_RMS	0.00335
R^2_{pred}	0.99943

Figure 3   Cantilever beam. Comparison of 1e5 and PRS deflection CDFs. LHS DOE

Table 5 Cantilever beam. LHS DOE. Probability of failure estimation

	1e5 Samples	RS
Pf	0.00174	0.00081
COV	0.0747	0.111

Table 6 Cantilever beam. LHS DOE. Samples in the equal interval bins

Bins		Samples	
LB >	UB ≤	1e5 Samples	RS
-0.59	-0.51	3	0
-0.51	-0.42	5	0
-0.42	-0.34	6	0
-0.34	-0.25	6	6
-0.25	-0.17	24	8
-0.17	-0.09	36	22
-0.09	0.00	94	43
0.00	0.08	199	133
0.08	0.17	524	346
0.17	0.25	1402	881
0.25	0.34	3140	2099
0.34	0.42	6929	4458
0.42	0.50	13317	8293
0.50	0.59	21260	14069
0.59	0.67	24911	19783
0.67	0.76	18797	22150
0.76	0.84	7811	17585
0.84	0.92	1453	8291
0.92	1.01	83	1719
1.01	1.09	0	114
1.09	1.18	0	1

CASE 2 – Orthogonal Arrays (OA) DOE

The orthogonal arrays data points and the cloud of test points are presented in Figure 4. All the conditions are same with the earlier case other than the DOE. The comparison of the statistics of the actual response and approximation in Tables 7 and 8 is good. The error metrics presented in Table 9 show that the quality of the approximation is satisfactory, but not as good as that from LHS samples. This is expected because the OA covers wider range of input space than the LHS.

In spite of acceptable response surface approximation, as observed in the earlier case, the failure probability is grossly misestimated, as shown in Table 10. The corresponding plot is presented in Figure 5. A reversal in the trend of the CDF is also observed in this case. The response surface predicts a failure probability that is 3 times larger than the actual failure probability. However, the error amplification is less extreme because the quality of the response surface approximation is not as good as that from LHS samples. The samples that fall in different bins are presented in Table 11. It can be concluded that the error is not triggered by the wrong choice in the DOE and cannot be rectified by just covering the entire design space.
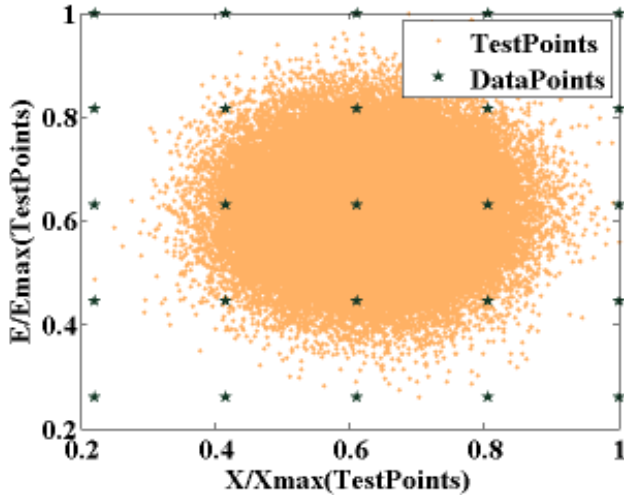
Figure 4   Cantilever beam. OA data points and cloud of 1e5 test points. Case 2
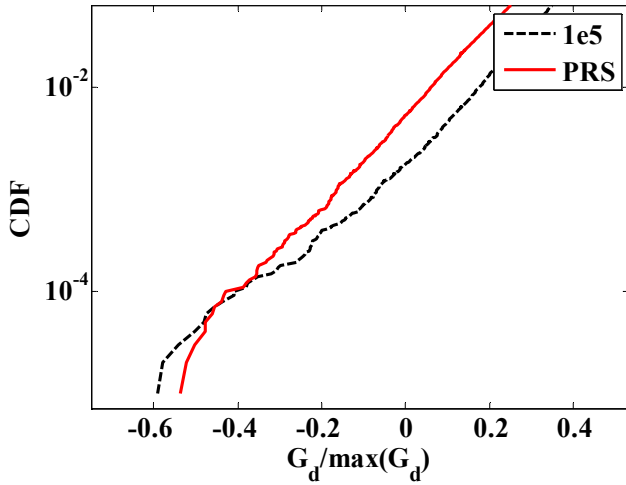

Figure 5   Cantilever beam. Comparison of 1e5 and PRS CDFs. Case 2

Table 7   Cantilever beam. Data points statistics (Case 2)

	y_{act}	y_{rs}	$y_{act} - y_{rs}$
Mean	0.40	0.40	0.00
Std. Dev.	0.54	0.54	0.08
Range	2.19	2.06	

Table 8   Cantilever beam. Test point statistics (Case 2)

	y_{act}	y_{rs}	$y_{act} - y_{rs}$
Mean	0.58	0.53	0.05
Std. Dev.	0.14	0.17	0.03

Table 9   Cantilever beam. Error metrics for the response surface (Case 2)

Error Metric	Value
R^2	0.98008
R^2_{adj}	0.97484
RMS	0.07485
RMS_{pred}	0.08586
PRESS_RMS	0.11156
R^2_{pred}	0.95576

Table 10 Cantilever beam. Probability of failure estimation (Case 2)

	1e5 Samples	RS
Pf	0.00174	0.00529
COV	0.0727	0.0435

Table 11 Cantilever beam. Samples in the equal interval bins (Case 2)

Bins		Samples	
LB >	UB ≤	1e5 Samples	RS
-0.59	-0.51	3	2
-0.51	-0.43	5	7
-0.43	-0.35	6	8
-0.35	-0.27	4	19
-0.27	-0.19	21	27
-0.19	-0.11	23	96
-0.11	-0.03	68	206
-0.03	0.05	126	464
0.05	0.13	328	1090
0.13	0.20	795	2203
0.20	0.28	1811	4267
0.28	0.36	4063	7274
0.36	0.44	8128	11631
0.44	0.52	14579	16530
0.52	0.60	21284	19577
0.60	0.68	23316	18558
0.68	0.76	16790	12521
0.76	0.84	7125	4742
0.84	0.92	1431	749
0.92	1.00	94	29

(II)BRANIN-HOO FUNCTION

The Branin-Hoo function is a two variable analytical function given as

$$f(x_1, x_2) = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6 \right)^2 + 10\left(1 - \frac{1}{8\pi} \right)\cos(x_1) + 10 \quad (4)$$

The plot of the function is presented in Figure 6. A cubic polynomial response surface is used to approximate the function. The variables x_1 and x_2 are considered to be random and follow prescribed statistical distributions. Four different cases that are presented in Table 12 are considered. A number of 25 data points are used in the DOE. Cases 1 and 3 use the LHS method, while cases 2 and 4 use the OA method to generate samples. The performance function is defined as

$$y(x_1, x_2) = f(x_1, x_2) - f_{\text{limit}} \quad (5)$$

If the function value is less than f_{limit}, it is considered failed. The f_{limit} for the first two cases is 4.0, while for the last two cases, it is 0.42. The corresponding failure regions are presented in Figures 7 and 8. It can be observed that the failure regions are encompassed within the safe region. This is a case of multiple island failure regions. A number of 10^5 test points are used to estimate the failure probability. The quantity of comparison is the failure probability estimated by the actual function and that by the response surface.

Table 12 Distributions of the variables and DOE used

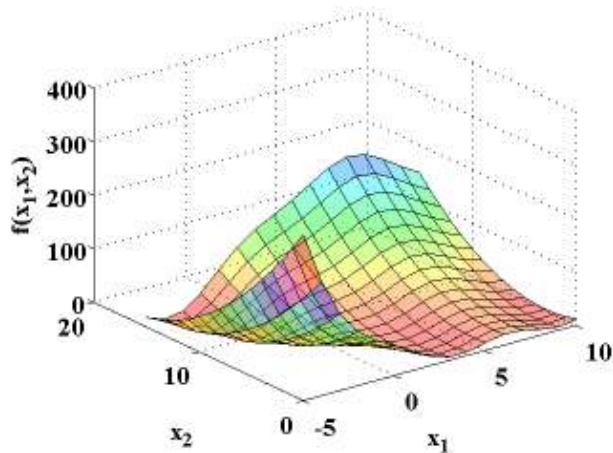| Normal Distribution		DOE	
		LHS(Normal)	OA
$x_1 \sim N(0,1)$	$x_2 \sim N(0,1)$	Case 1	Case 2
$x_1 \sim N(3,1)$	$x_2 \sim N(4,2)$	Case 3	Case 4

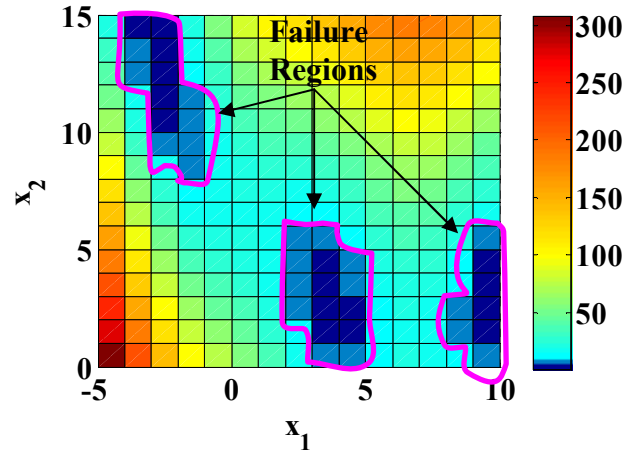Figure 6   Branin-Hoo function.

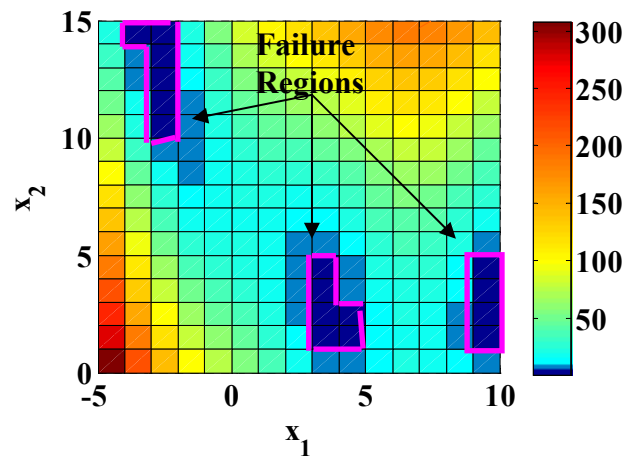Figure 7   Branin-Hoo function – Failure regions. $f_{\text{limit}} = 4$

Figure 8   Branin-Hoo function – Failure regions. $f_{\text{limit}} = 0.42$

CASE1 –LHS DOE:

The DOE data points from LHS and the cloud of 10^5 test points are presented in Figure 9. The statistics at the design point show that the order of the response surface is good and there is no alarming reason to consider the response surface is inaccurate (Tables 13 and 14). The error metrics in Table 15 show that the response surface is a good fit and there are no unnecessary terms. The PRESS errors are higher for this case, but are still small compared to the range. In spite of good error metrics, the response surface predicts the failure probability with large error as shown in Table 16. The estimated failure probability varies by about 5 times with respect to the actual failure probability. The failure regions in the response surface are presented in Figure 11. It is observed that the response surface is unable to capture the island trend, instead approximates the failure region to be a half plane and does not predict failure at the upper left corner of the design space.
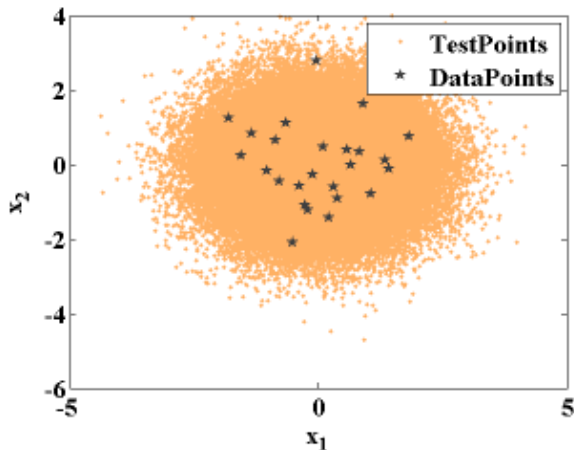
Figure 9   Branin-Hoo. DOE data points from LHS and cloud of 1e5 test points.

Table 13 Branin-Hoo. Statistics at data points. Case 1

	y_{act}	y_{rs}	$y_{\text{act}} - y_{\text{rs}}$
Mean	56.05	56.05	0.00
Std. Dev.	21.55	21.55	0.00
Range	83.41	83.62	

Table 14 Branin-Hoo . Statistics at test points. Case 1

	y_{act}	y_{rs}	$y_{\text{act}} - y_{\text{rs}}$
Mean	56.90	56.42	0.48
Std. Dev.	24.03	23.77	0.26

Table 15 Branin-Hoo. Error metrics for the response surface

Error Metric	Value
R^2	0.99992
R^2_{adj}	0.99987
RMS	0.18716
RMS$_{\text{pred}}$	0.24162
PRESS_RMS	0.82435
R^2_{pred}	0.99848

Table 16 Branin-Hoo. Probability of failure estimation (Case 1)

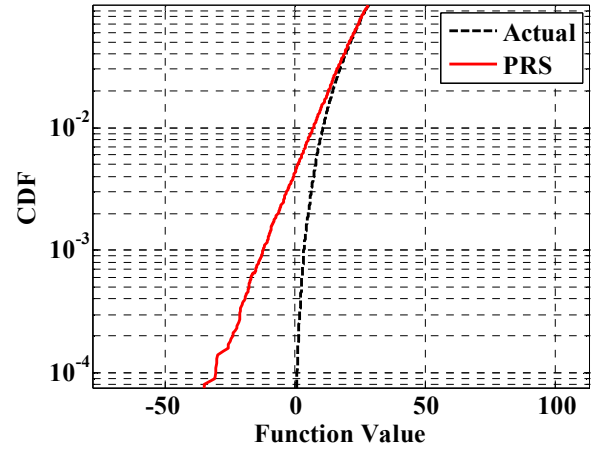|	1e5 Samples	RS
Pf	0.0014	0.0072
COV	0.0857	0.0375

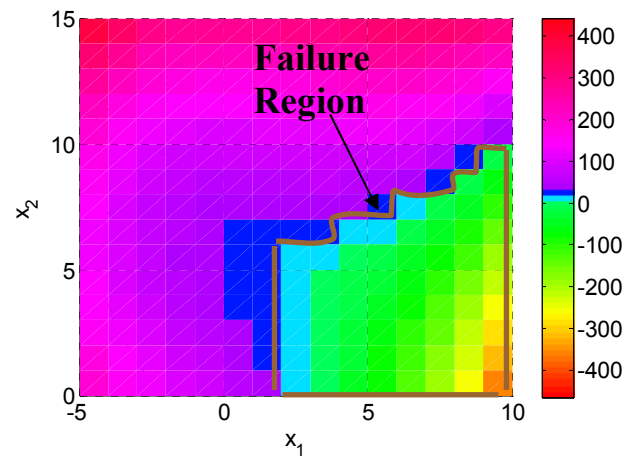Figure 10 Branin-Hoo response surface (case1). Failure region

Figure 11 Branin-Hoo response surface (case1). Failure region

CASE 2 – OA DOE:
The DOE and the test points used are presented in Figure 12. The error metrics presented in Table 19 show that the response surface fit is good though the PRESS error is high. The failure probability estimate obtained through the response surface is close to the actual value. But it is to be noted with the help of Figure 13 that the error in the failure probability depends on the choice of f_{limit} . For example, if it is chosen to be 11.05, with the help of bin data in Table 21, one can see that the error in the failure probability will be about two times. The actual function cannot take values less than 0. However, it is clear from Figure 13 that the response surface can take negative values in the extrapolated regions. This DOE was able to capture the island in the top left of the design space though it was not able to identify the other failure regions individually (Figure 14), instead approximating both regions by an approximate half plane.
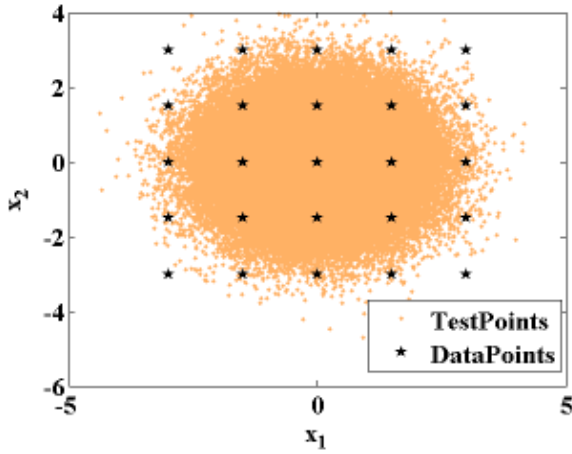
Figure 12 Branin-Hoo. OA data points and cloud of 1e5 test points. Case 2

Table 17 Branin-Hoo. Statistics at data points (Case 2)

	y_{act}	y_{rs}	$y_{\text{act}} - y_{\text{rs}}$
Mean	67.84	67.84	0.00
Std. Dev.	58.81	58.78	1.92
Range	222.75	222.75	

Table 18 Branin-Hoo. Statistics at test points (Case 2)

	y_{act}	y_{rs}	$y_{\text{act}} - y_{\text{rs}}$
Mean	56.90	55.91	0.99
Std. Dev.	24.03	24.15	0.12

Table 19 Branin-Hoo. Error metrics for the response surface (Case 2)

Error Metric	Value
R^2	0.99893
R^2_{adj}	0.99829
RMS	1.88122
RMS_{pred}	2.42865
PRESS_RMS	2.84214
R^2_{pred}	0.99757

Table 20 Branin-Hoo. Probability of failure estimation (Case 2)

	1e5 Samples	RS
Pf	0.0014	0.0013
COV	0.0857	0.0846

Table 21 Branin-Hoo. Samples in the equal interval bins (Case 2)

Bins		Samples	
LB >	UB ≤	1e5 Samples	RS
-17.28	-3.12	0	19
-3.12	11.05	1076	551
11.05	25.21	6525	6261
25.21	39.37	15877	19235
39.37	53.53	24242	25597
53.53	67.70	22975	21251
67.70	81.86	15394	13598
81.86	96.02	7884	7157
96.02	110.19	3501	3491
110.19	124.35	1417	1561
124.35	138.51	647	741
138.51	152.68	264	317
152.68	166.84	103	125
166.84	181.00	49	55
181.00	195.17	26	28
195.17	209.33	16	11
209.33	223.49	2	1
223.49	237.65	1	0
237.65	251.82	0	1
251.82	265.98	1	0

Figure 13 Branin-Hoo. Comparison of actual and response surface CDFs. Case2
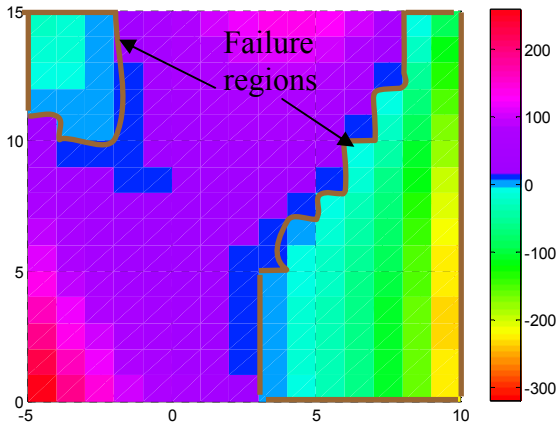
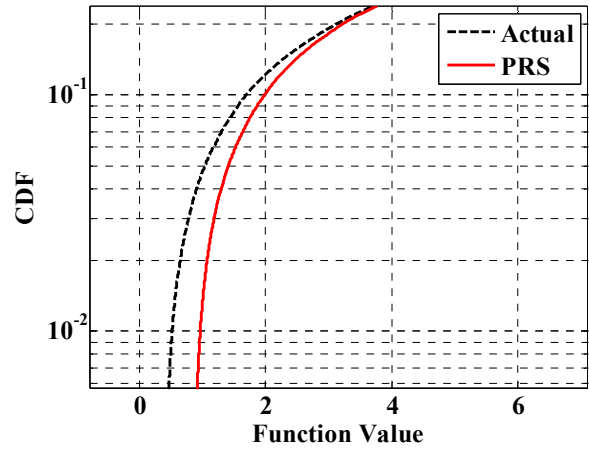Figure 14 Branin-Hoo response surface (case2). Failure region

Figure 16 Branin-Hoo. Comparison of actual and response surface CDFs. Case3

Case 3 – LHS DOE:

The error metrics and the statistics at data points (Tables 22-24) show that the response surface fit is good. The DOE and the cloud of test points are presented in Figure 15. It can be observed from Figure 16 that the response surface constantly underestimates the failure probability. Figure 17 shows the failure region approximated by the response surface. This DOE is able to capture the failure region in the lower central portion of the design space. However, the other two failure regions are not captured and hence contribute to the error in the estimated failure probability. If the function limit was 1, the response surface predicts a fully safe condition whereas the actual function corresponds to a finite failure probability (Table 25).

Table 22 Branin-Hoo. Statistics at data points (Case 3)

	y_{act}	y_{rs}	$y_{act} - y_{rs}$
Mean	11.59	11.59	0.00
Std. Dev.	11.21	11.20	0.53
Range	41.05	40.65	

Table 23 Branin-Hoo. Statistics at test points (Case 3)

	y_{act}	y_{rs}	$y_{act} - y_{rs}$
Mean	11.14	11.24	-0.10
Std. Dev.	10.08	10.37	1.84

Table 24 Branin-Hoo. Error metrics for the response surface (Case 3)

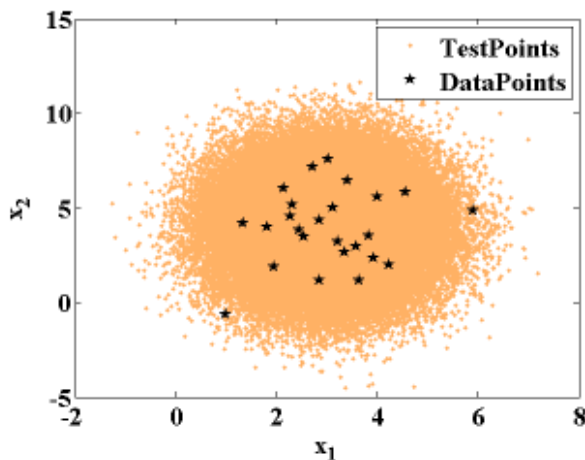| Error Metric	Value
R^2	0.99780
R^2_{adj}	0.99648
RMS	0.51496
RMS_{pred}	0.66481
PRESS_RMS	4.05256
R^2_{pred}	0.86386

Figure 15 Branin-Hoo. OA data points from and cloud of 1e5 test points. Case 3
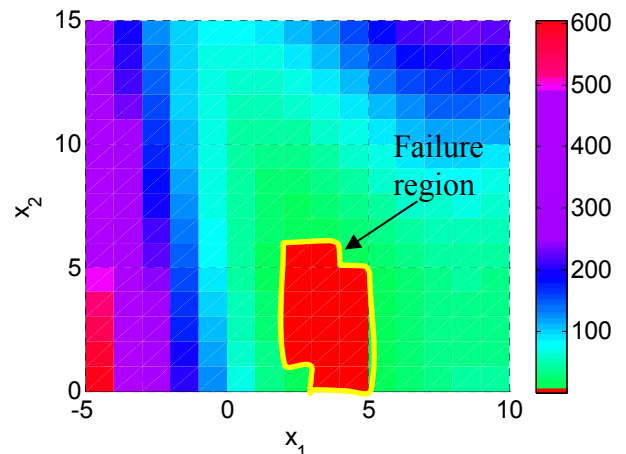

Figure 17 Branin-Hoo. Response surface (case3). Failure region

Table 25 Branin-Hoo. Probability of failure estimation (Case 3)

	1e5 Samples	RS
Pf	0.002	0
COV	0.07	N/A

Table 26 Branin-Hoo. Samples in the equal interval bins (Case 3)

| Bins		Samples	
LB >	UB ≤	1e5 Samples	RS
0.40	6.12	38226	39163
6.12	11.84	26141	26145
11.84	17.56	16106	14965
17.56	23.28	8877	8475
23.28	29.00	4628	4775
29.00	34.72	2574	2680
34.72	40.44	1479	1586
40.44	46.16	801	897
46.16	51.88	498	522
51.88	57.60	282	297
57.60	63.32	173	222
63.32	69.04	99	109
69.04	74.77	60	68
74.77	80.49	26	42
80.49	86.21	12	24
86.21	91.93	7	9
91.93	97.65	6	8
97.65	103.37	5	7
103.37	109.09	0	5
109.09	114.81	0	1
114.81	120.53	0	0
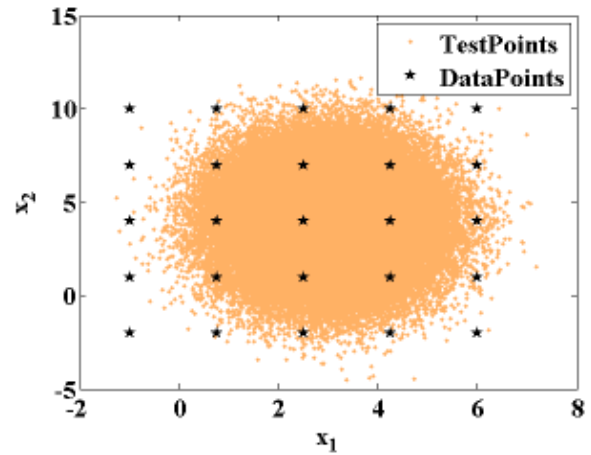
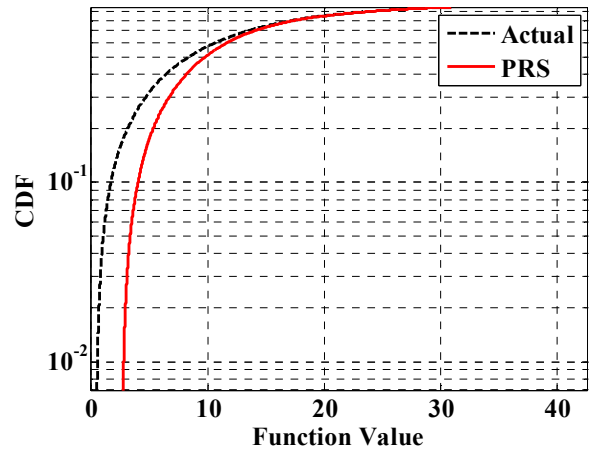Figure 18 Branin-Hoo. OA data points from and cloud of 1e5 test points. Case 4

Figure 19 Branin-Hoo. Comparison of actual and response surface CDFs. Case4

CASE 4- OA DOE:

The DOE and the cloud of test points are presented in Figure 18. The error metrics show that the fit is good (Tables 28-30). But there is large error in the failure probability estimate (Tables 31-31 and Figure 19). Similar to the earlier case, the response surface consistently predicts a lesser failure probability. The failure regions approximated by the response surface is presented in Figure 20. Here, the island in the top right of the design space is not captured.

Table 27 Branin-Hoo. Statistics at data points (Case 4)

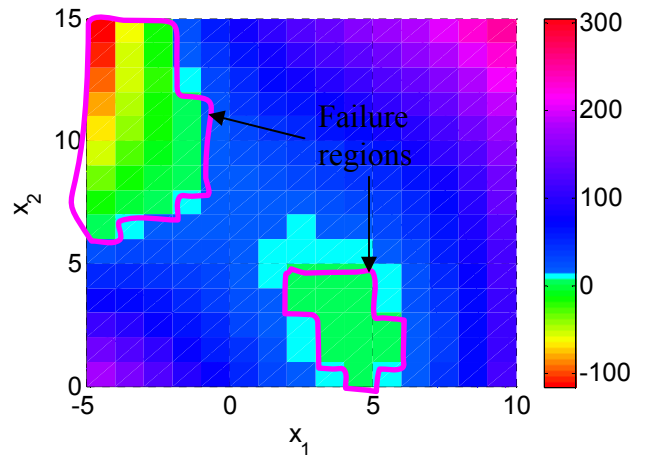|	y_{act}	y_{rs}	$y_{act} - y_{rs}$
Mean	35.95	35.95	0.00
Std. Dev.	28.05	27.97	2.13
Range	106.00	106.42	

Figure 20 Branin-Hoo response surface (case 4). Failure region

Table 28 Branin-Hoo .Statistics at test points (Case 4)

	y_{act}	y_{rs}	$y_{\text{act}} - y_{\text{rs}}$
Mean	11.14	12.40	-1.26
Std. Dev.	10.08	9.30	2.35

Table 29 Branin-Hoo. Error metrics for the response surface (Case 4)

Error Metric	Value
R^2	0.99426
R^2_{adj}	0.99081
RMS	2.08257
RMS_{pred}	2.68859
PRESS_RMS	3.14634
R^2_{pred}	0.98689

Table 30 Branin-Hoo. Probability of failure estimation (Case 4)

	1e5 Samples	RS
Pf	0.002	0
COV	0.07	N/A

Table 31 Branin-Hoo. Samples in the equal interval bins (Case 4)

Bins		Samples	
LB	UB	1e5 Samples	RS
0.40	5.72	36097	22869
5.72	11.05	25360	33953
11.05	16.38	16395	20277
16.38	21.71	9617	10113
21.71	27.03	5218	5382
27.03	32.36	2971	3108
32.36	37.69	1775	1794
37.69	43.01	1009	990
43.01	48.34	615	632
48.34	53.67	366	350
53.67	59.00	235	236
59.00	64.32	145	120
64.32	69.65	87	73
69.65	74.98	55	50
74.98	80.30	25	29
80.30	85.63	11	4
85.63	90.96	8	12
90.96	96.28	6	3
96.28	101.61	4	4
101.61	106.94	1	0
106.94	112.27	0	1

CONCLUSIONS AND FUTURE WORK

This paper attempted to investigate the amplification of errors in response surface approximations in failure probability calculation. It demonstrated the situations in which the error amplifies with two numerical examples with different DOEs. It is observed that the commonly used error metrics for the response surface approximations do not necessarily reflect the accuracy in failure probability estimation. When the samples are located near the mean and the failure region is located far away from the mean (the cantilever beam with LHS), the failure probability estimate using the response surface approximation shows large error amplification. When the samples are evenly distributed in the input space (the cantilever beam with OA), the the error metrics are less misleading but there is still substantial amplification. When the failure region is close to the mean but its size is small (Branin-Hoo function), error metrics of LHS is better than that of OA. However, the accuracy in the failure probability turns out opposite. In addition, both sampling methods fail to identify one or more of the small failure regions.

ACKNOWLEDGEMENT

REFERENCES

[1] Elishakoff I., Safety factors and reliability: friends or foes? Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004

[2] Khuri, A.I., and Cornell, J.A., Response Surfaces, Second Edition. Dekker, Inc., New York, NY, 1996.

[3] Kutaran, H., A. Eskandarian, D. Marzougui, and N. E. Bedewi., 2002, "Crashworthiness design optimization using successive response surface approximations. Computational Mechanics 29, 409-421.

[4] Melchers, R.E. 1999: Structural reliability analysis and prediction, Wiley, New York.

[5] Rajashekhar R.M. and Ellingwood. B.R., 1993, "A new look at the response surface approach for reliability analysis," Structural Safety, pp 205-220

[6] Gupta. S., and Manohar C.S., 2004, "An improved response surface method for the determination of failure probability and importance measures," Structural Safety, Vol. 26, pp. 123-139.

[7] Venter, G., Haftka, R.T., and Starnes, J.H., Jr. 1998, "Construction of response surface approximation for design optimization", AIAA J., 36(12), pp. 2242-2249

[8] Youn, B.D. and Choi, K.K., "Selecting probabilistic approaches for reliability-based design optimization", AIAA Journal, 42(1), 2004.

[9] Chen, S., Nikolaidis, E. and Cudney, H. H., Comparison of Probabilistic and Fuzzy Set Methods for Designing under Uncertainty, AIAA/ASME/ASCE/ AHS/ASC Structures, Structural dynamics, and Materials Conference and Exhibit, 2860-2874, 1999.

Appendix. Error Metrics

1. R^2:

The coefficient of multiple determinations is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \qquad (6)$$

where y_i is the actual value at the ith design point, $\hat{y}_i$ the predicted value at the ith design point, and $\overline{y}$ the mean of the actual response.

R^2 is a measure of the amount of reduction in the variability of y obtained by using the response surface. $0 \le R^2 \le 1$. A larger value of R^2 is desirable for a good response surface. But, a larger R^2 does not necessarily guarantee a good response surface. Thus, this estimate should be used in conjunction with other error estimates to gauge the quality of the response surface. R^2 continuously increases with addition of terms irrespective of whether the additional term is statistically significant.

2. ADJUSTED R^2:

The adjusted coefficient of multiple determinations is defined as

$$R^2_{adj} = 1 - \frac{n-1}{n-p}(1 - R^2) \qquad (7)$$

where n is the number of design points, and p is the number of regression coefficients.

Unlike R^2, R^2_{adj} decreases when unnecessary terms are added. Hence, R^2_{adj} along with R^2 can be used to comment on the quality of response surface and the presence of unnecessary terms in the response surface.

3. ROOT-MEAN-SQUARE (RMS) ERROR:

The root-mean-square error, RMS, and the predicted RMS errors are defined, respectively, as

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad (8)$$

$$\text{RMS}_{\text{pred}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n - p}} \qquad (9)$$

The predicted RMS is a measure of the variance (standard deviation) of the error.

4. PRESS:

The prediction error sum of squares provides error scaling. To estimate the PRESS, an observation is removed at a time and a new response surface is fitted to the remaining observations. The new response surface is used to predict the withheld observation. The difference between the withheld observation and the computed response value gives the PRESS residual for that observation. This process is repeated for all the observations and the PRESS statistic is defined as the sum of the squares of the *n* PRESS residuals. When polynomial response surfaces are used, the repetitive estimate of PRESS residuals can be obviated by using the following expression:

$$\text{PRESS} = \sum_{i=1}^{n}\left(\frac{e_i}{1 - E_{ii}}\right)^2 \qquad (10)$$

where $\mathbf{E} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ and $\mathbf{X}$ is the Grammian matrix $(\hat{\mathbf{y}} = \mathbf{X}\mathbf{b})$, and $\mathbf{b}$ is the coefficient vector. Data points at which E_{ii} are large will have large PRESS residuals. These observations are considered high influence points. That is, a large difference between the ordinary residual and the PRESS residual will indicate a point where the model fits the data well, but the model built without that point has a poor prediction. A RMS version of PRESS allows us to compare the PRESS_RMS with the RMS errors. This permits us to explore the influence that few points might have on the entire fit. The PRESS_RMS is expressed as:

$$\text{PRESS\_RMS} = \sqrt{\frac{\text{PRESS}}{n}} \qquad (11)$$

PRESS can be used to estimate an approximate R^2 for prediction as:

$$R^2_{\text{pred}} = 1 - \frac{\text{PRESS}}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \qquad (12)$$

The denominator in Eq. (12) is referred to as total sum of the squares.