Effective Tests for Discovering Unexpected Structural Failure Modes

Taiki Matsumura[1], Raphael T. Haftka[2] and Nam H, Kim[3]
University of Florida, Gainesville, FL, 32611

Many of airplane accidents are due to unexpected failure modes. In structural design, appropriate failure-criterion characterization including identification of potential failure modes is key to reducing such accidents. However, since analytical predictions often have substantial errors due to complexity of failure mechanism, this process tends to depend much on tests. In this study, we demonstrate failure-criterion characterization with simple structural elements in order to shed light on effective testing. We first show that ignoring the existence of critical failure modes could lead to a substantially overestimated probability of failure by two orders of magnitude for a composite laminate example. Tests suffer from noise and so they are often repeated to reduce that noise. Resource allocation between the number of different tests in search of unexpected modes and the number of repetitions of each test is investigated. It is shown that increasing the number of tests for exploration is vital to spot unexpected failure modes, at the same time, it is almost as effective as repetitions for reducing the effect of noise on accuracy. Furthermore, the importance of the number of different tests for exploration is amplified when the failure load surface becomes complicated.

I. Introduction

The safety of airplanes has been improved over decades, and the risk (fatal accidents per passenger mile) of commercial air travel is substantially lower than other modes of transportation such as private automobiles and buses [1, 2]. However, even further safety has been continuously demanded. For example, the Commercial Aviation Safety Team (CAST) is a recent nation-wide joint activity aiming at reducing the U.S. commercial fatality risk by 50 percent from 2010 to 2025 [3]. One of strong incentives to further safety is sentimental and catastrophic nature of aviation accident; single failure is likely to lead to a severe consequence involving a number of fatalities. Another consideration is the fact that aviation has been playing a pioneering role of evolving engineering technology, such as introducing new design methodologies and new materials[4], resulting in being forced to tackle uncertainties in development as well as service phase. It is also evidenced by the fact that many of fatal aviation accidents involve unexpected and un-modeled failure scenarios. For example, the Aloha Airlines accident in 1988 revealed that multi-site fatigue cracks significantly threatened airplane safety[5]. The disasters of the Space Shuttle Columbia in 2001 were caused by unexpectedly damaged thermal protection system. Therefore, effective management of unexpected failure modes is key to the further safety improvement.

While the current safety level of airplane structure is mainly due to the safety factors used in a deterministic design approach, a factor of safety may not suffice to compensate for unexpected failure modes. Unexpected failure modes may threaten safety before planned inspection and maintenance in service. Even when a technical problem due to an unexpected mode is found at a late development stage, e.g., system certification test, redesign across the system accompanied by schedule delays and additional cost will be inevitable. Thus, discovering unexpected failure modes as early as possible is vital.

The building block test approach is commonly employed in development of complex airplane structures[6, 7], and detailed guidelines are provided by the Department of Defense[8]. The building block approach is a synergetic approach between analytical predictions and experiments. Design errors are eliminated step-by-step by evaluating the discrepancy between analytical predictions and test observations as the structural complexity increases

---

throughout the development. With growing interest in replacing the deterministic design approach with probabilistic design approach [9], several studies have quantified the contribution of development tests to improving reliability [10-14]. However, past research was based on the assumption that uncertainties, such as design error and variability, behave as designers expect, ignoring the chance for the unexpected. This does not reflect the reality that high percentage of dangerous failures occurs due to unexpected failure modes.

One of important roles of the building block approach is to improve the characterization of failure criteria, including understanding potential failure modes. Because of complexity of failure mechanisms and multiplicity of failure modes, analytical models are not reliable enough. Therefore, the failure-criterion characterization tends to depend much on tests. If the experiments fail to spot a potential critical failure mode and then a designer designs a structure ignoring its existence, then failure mode may manifest itself at later development stages or, more undesirably, in flight. Besides uncovering all relevant failure modes, tests help produce accurate design allowable charts, e.g., failure load map with respect to geometry and load conditions. Because tests are often very expensive, it is vital to satisfy both objectives with limited number of tests

The objective of this paper is to shed light on key factors of effective tests for failure-criterion characterization, including discovering underlying unexpected failure modes. We demonstrate failure-criterion characterizations with two simple structural elements: a support bracket and a composite laminate. We first examine the impact of missing a failure mode on safety by evaluating the probability of failure. We then, investigate a tradeoff between the chance of detecting unexpected mode and accuracy of failure load prediction.

The paper is organized as follows. In Chapter II, we discuss details of the building block test. Then, Chapters III and IV demonstrate failure-criterion identification with simple example structures, followed by conclusions in Chapter V.

II.   Building Block tests and Unexpected Failure Modes

The building block test approach is conducted by a battery of tests as well as designer's predictions in order to ensure that a structure behaves as expected without having any defects that may threaten safety. The approach often starts with small specimens (e.g., material coupons) and progresses through structural elements (e.g., joint and composite laminates) and components (e.g., fuselage and wing), and finally a full scale complete airplane structure for certification (Fig. 1). The key philosophy of the building block approach is that discrepancy between analytical predictions and actual structural behavior is to be found as early as possible in order to minimize cost and program delays. However, because tests are expensive, it is challenging to make an effective test plan with a limited number of tests.

A major role of building block tests is to refine failure criteria which are critical for determining design allowables. This process is usually conducted before design with a family of element structures representing a structure for a particular use. We refer to these tests as generic element tests, because they seek to provide information on generic elements of the type tested. Failure prediction needs to consider multiple factors such as external environments, geometry of structure, variability of material properties, defects and so on.

Characterizing failure criteria has two goals. The first goal is to identify all potential failure modes. As quoted by the DOD handbook [8], "The multiplicity of potential failure modes is perhaps the main reason that the building block approach is essential in the development of composite structure substantiation." It is vital to understand all underlying failure modes and prevent them from happening anytime in lifecycle of the structure. The second goal is to construct design allowables, for example a failure load chart with respect to geometry and load conditions. This is usually implemented by a matrix of experiments. Then, approximation techniques, e.g., linear regression and polynomial regression, are used to interpolate the observed data. A challenge is to construct an accurate prediction while tackling noisy data observation due to material variability, error in test conditions and measurement, etc.

A key question is how we allocate the limited number of tests to satisfy the two different objectives, i.e., identifying all potential failure modes and prediction accuracy against noisy data and. While the accuracy can be achieved by repeating test observation for each point in the matrix to eliminate the effect of noise, exploring within the matrix with many different points more likely captures underlying potential failure modes.
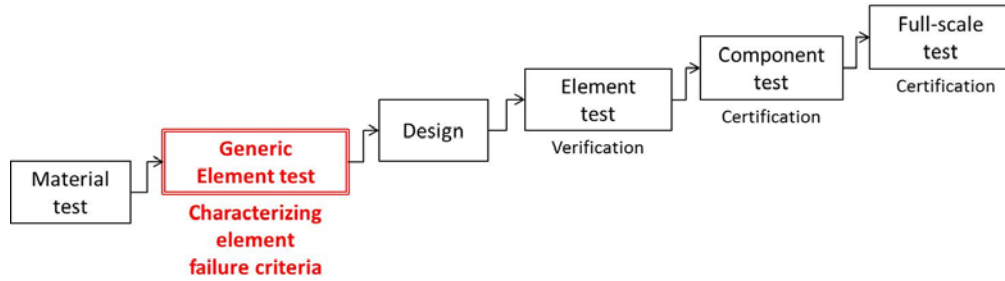
Figure1. Building block tests

III. Example I: Support Bracket

A. Description of structure

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

   The combination of loading and geometry generates multi-axial states of stress due to axial, bending, torsion, and torsional shear stresses. Because of the additive effect of the torsion and torsional shear stresses or bending and axial stresses, point D is likely to be the critical failure point (Fig 3), but it is not always the case. When the diameter d decreases, the axial stress at point A due to bending becomes dominant for some loading conditions at around 90deg. Figure 4 describes the critical failure modes on the structure with respect to diameter d and load angleα.The failure mode initiating from point A is considered as the unexpected failure mode for this problem.

   The yield strength of the material is normally distributed, which is the only modeled cause of noisy observations. Failure is predicted by the Von-Mises criterion ignoring stress concentrations. The tests seek to allow designers to predict the mean failure loads determined by the mean of yield strength with respect to d and α by performing experiments for different combinations of these two variables.
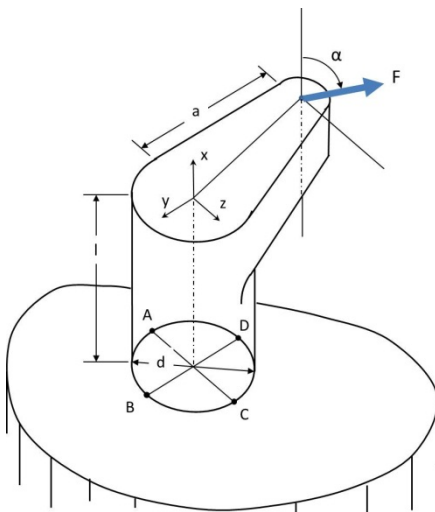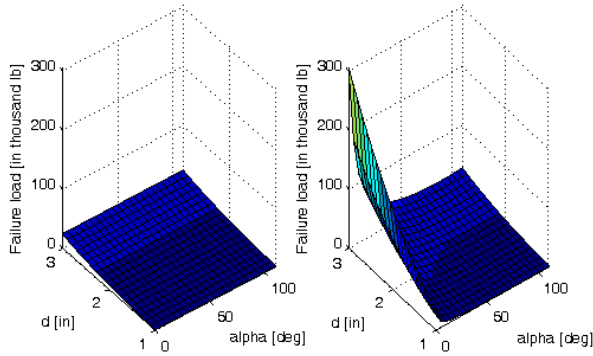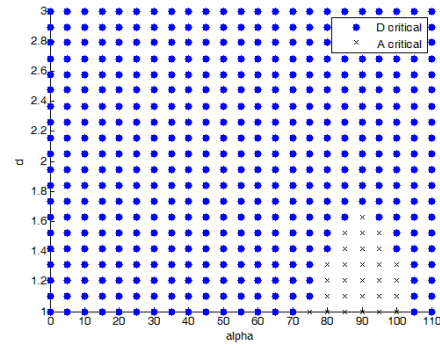

Figure2. Support Bracket

Table 1. Properties of bracket

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

**Figure 3. Comparison of failure loads
(Left: Point D, Right: Point A)**

**Figure 4. Critical failure modes
(Critical point on the structure)**

B. Impact of missing critical failure mode on probability of failure

We examine the impact of missing the critical failure on reliability estimation. The largest discrepancy in the failure loads between points A and D is observed when $d = 1$ inch and$\alpha = 90$ deg, resulting in the failure loads of 988lb at point D and 947lb at point A. Suppose that designers design the structure under the condition that the load angle is 90 deg and $d = 1$.Since it is assumed that they set the operational load based only on the failure prediction at point D, it is set at 659 lb (=988/1.5) accounting for a safety factor of 1.5. Note that no error in the failure load prediction is considered. The probability of failure is predicted to the operational load being normally distributed with coefficient of variation (COV) of 10%. All material and geometry variability is shown in Table 1. Table 2 shows the difference between the designer's predicted probability of failure and the true probability of failure driven by point A. The prediction overestimates the safety by a factor of two.

Table 2. Comparison of probabilities of failure for bracket when shear failure is missed

	Probability of failure	Remarks
Prediction	3.6×10^{-3}	Prediction missing critical mode
True	7.8×10^{-3}	Due to critical mode

C. Design of experiments

For a matrix of experiments with two variables, d and α, we employ various numbers of grid points and repetitions at each point in the matrix. Table 3shows the sets of the matrices examined and the corresponding total number of tests. Due to the variability in the yield strength, the test observation suffers from noise. We fit regression models to the mean values out of the repetitive observations of each sampling point. We apply polynomial response surface (PRS: 2nd, 3rd and 4th order) [15]and Kriging [16], and then choose the best regression model based on the prediction accuracy, i.e., cross validation error (PRESS in particular) [17]. To average out the randomness of the test results, we generate100 sets of test results and construct the regression model for each, and then take the mean value of the cross validation error as a representative for each regression model.

Table 3. Total number of tests for design of experiments

Number of grids		Number of repetitions						
α	d	1	2	3	4	5	6	7
3	3	9	18	27	36	45	54	63
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

D. Failure-criterion characterization

After comparing the PRS and Kriging, it appeared that 2nd order PRS provides the smallest cross validation error for all cases. Figure 5 shows the cross validation error of 2nd order PRS, cross validation error divided by the mean of failure loads of entire the matrix. Note that most of the error is due to the noise in test results, rather than to the error in the regression model. This is evidenced by the fact that PRS fitted to no-noise data (the mean failure loads) has very good accuracy, less than 1% in cross validation error. Figure 6 shows the relative root mean square error (RRMSE) to the mean failure loads for point D evaluated at 620 (20x31) points in the matrix. As can be seen, the error decreases as the number of repetitions increases. On the other hand, even though the number of repetitions is small, increasing the number of grid points can compensate the accuracy. For example, 7 grid points with one repetition provides 10% of error, which is comparable with the accuracy of 3 grid points with 5 repetitions and 4 grid points with 2 repetitions.
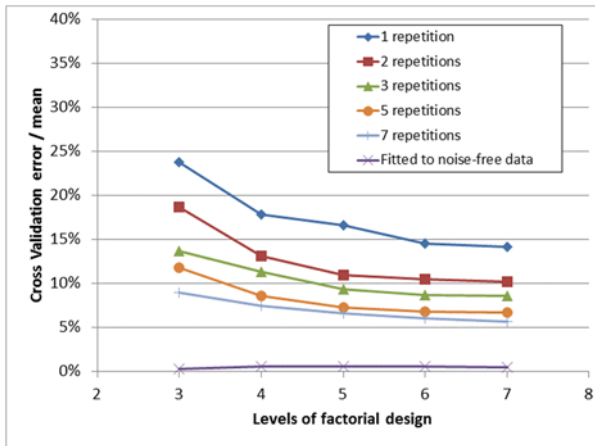

Figure 5. Cross validation error of regression model (2nd order PRS)
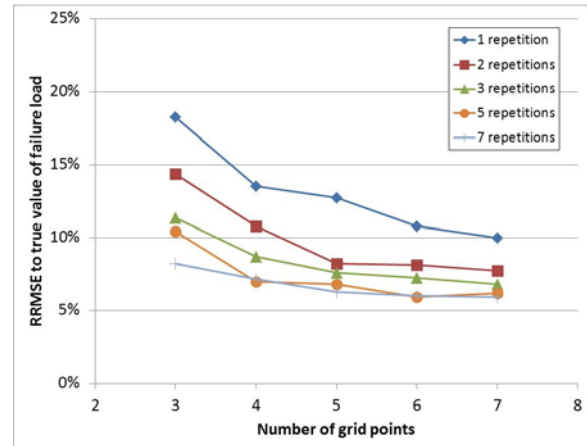
Figure 6. Relative root mean square error to true failure load (RRMSE, 2nd order PRS)

Figure 7 shows the tradeoff between the chance of discovering the unexpected failure mode and the accuracy of failure load prediction with respect to the total number of tests. In the graph, a large circle denotes that the unexpected failure mode is detected, and a small square denotes that the unexpected is not detected. In terms of discovering that failure may occur at point A, while the experiments with more than five grid points can spot it, the experiments with three and four grid points cannot. This simply reflects the fact that the denser grid is more likely to have a test in the small region of unexpected mode.

It can be also observed that there is no significant difference in the accuracy between different grid points. For example, given 100 test chances, 4 grid points with 6 repetitions (total 96 tests) provides the smallest RRMSE of 6.7%, but it missed the unexpected mode. On the other hand, 5 grid points with 4 repetitions (total 100 tests) and 7 grid points with 2 repetitions (total 98 tests), which successfully spotted the unexpected, still have comparable accuracy, 7.2% and 7.7% respectively. The reason is that the original failure load surface (left figure of Fig. 3) is fairly simple, and as previously mentioned, the error in the regression model is small. For that case, the allocation of tests whether to the grid point or to the repetition does not matter in terms of the accuracy. Thus, for a given total number of tests, increasing the number of different tests is superior to increasing the number of repetitions is a good strategy in terms both of the accuracy and discovering the unexpected failure mode. Note that, in reality, errors in tests are not only due to noise but also due to bias which the examples of this paper do not consider. So, the improvement in accuracy that is achieved by increasing the number of tests in Fig. 6 is somewhat exaggerated.
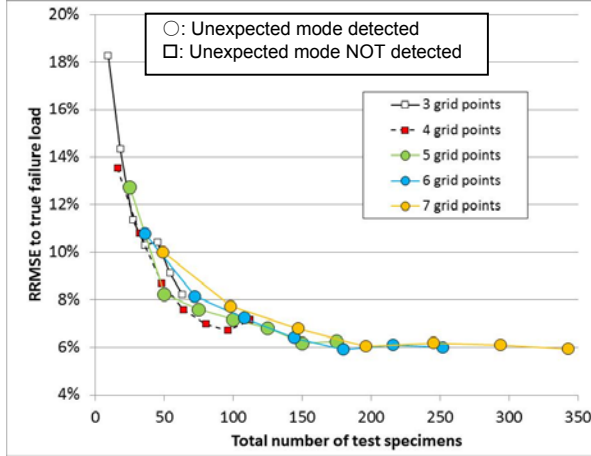
Figure 7. RRMSE vs. Total number of tests and whether or not unexpected mode being discovered (COV of yield strength =10%)
On each line, markers represent the number of repetitions, 1 thought 7from left.
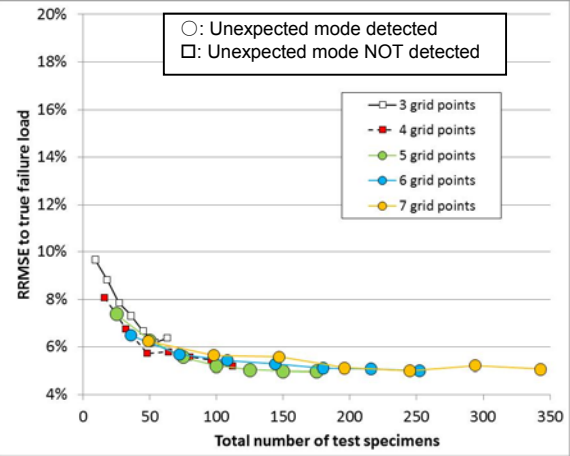
Figure 8. RRMSE vs. Total number of tests and whether or not unexpected mode being discovered (COV of yield strength =5%)
On each line, markers represent the number of repetitions, 1 thought 7from left.

Another way of dealing with noisy data is to reduce the variability by quality control. If all test specimens are manufactured from single material batch, variability of the material property can be controlled at minimum. Figure 8 shows the case that COV of the yield strength is reduced to 5%. The same trend as previously discussed for 10% COV holds, but the accuracy is improved. In this situation, for a given target accuracy, designers can more effectively allocate tests to explore within the matrix in search of potential unexpected modes. However, a shortcoming of this approach is that the material property from a batch might be biased. Designers need to be careful by checking the property of the material batch used for the element test compared to statistical data from material tests, e.g., distribution of material property, which is supposed to cover all fabricating conditions.

IV. Example II: Composite laminate

A. Description of structure and design of experiments

For the second example, intended to have a more complicated failure surface, a symmetric composite laminate with three ply angle [0° -θ +θ]s is considered (Fig.9). The laminate is subject to mechanical loading along the x and y directions defined by a ratio α, such that Nx = (1-α)F and Ny = αF. For example, if α = 0, only a tension load F in the x direction is applied. As design parameters for the failure load identification, 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 3 shows the material properties and strain allowables, including strain allowable along fiber direction ε_{1allow}, transverse along fiber direction ε_{2allow}, and shear $\gamma_{12allow}$. All the properties are assumed to be normally distributed.

We consider the ply strain failure by shear as the unexpected failure mode. Figure 10 shows the failure load surface, and Fig. 11 depicts the critical failure modes with respect to the design parameters.
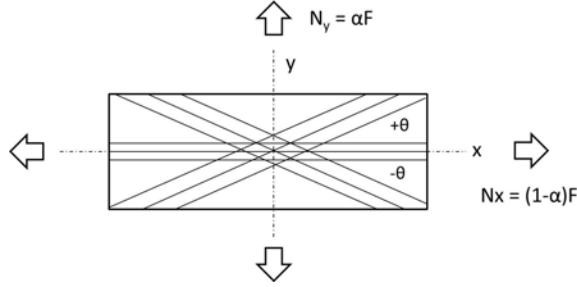
Figure9. Composite laminate

Table 3. Properties of structure

Property	Quantity (mean)	COV
E_1 [GPa]	150	5%
E_2 [GPa]	9	5%
v_{12}	0.34	5%
G_{12} [GPa]	4.6	5%
Thickness of ply [μm]	125	N/A
ε_{1allow}	±0.01	6%
ε_{2allow}	±0.01	6%
$\gamma_{12allow}$	±0.015	6%
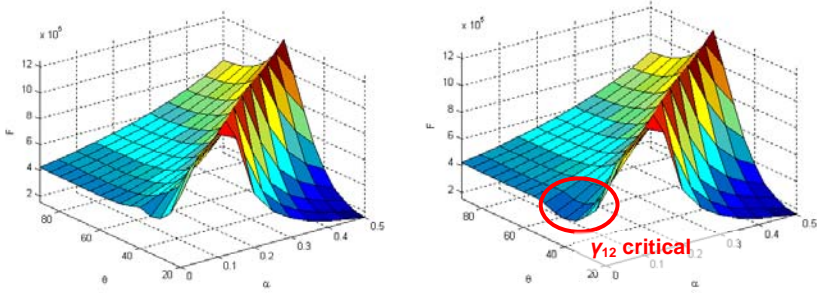
Figure 10. Comparison of failure loads
(Left: based only on ε_1 or ε_2, Right: based on $\varepsilon_1,\varepsilon_2$ or γ_{12})
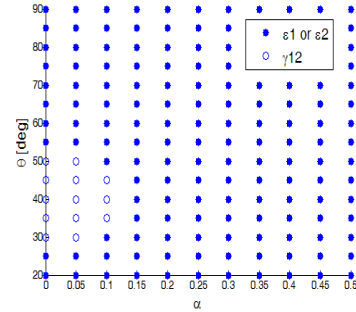
Figure 11. Critical failure modes

B. Impact of missing critical failure mode

Suppose that the composite laminate with the ply angle θ of 40deg is selected under the load condition α = 0. While the failure load is 471,000 N/m when only ply strain ε_1 and ε_2 is considered (left hand side of Fig. 10), the true failure load due to ply shear γ_{12} (right hand side of Fig. 10) is 359,900 N/m. It is assumed that designers set the upper bound of the operational load without noticing that failure may occur by play shear. The design load is set by taking into account safety factor 1.5, so that 314,000 N/m (= 471,000/1.5). Note that the load is randomly distributed according to normal distribution with COV of 10%. Table 4 shows designer's predicted probability of failure and true probability of failure. The prediction substantially overestimates safety by more than two orders of magnitude.

Table 4. Comparison of probabilities of failure for composite laminate when shear failure is missed

	Probability of failure	Remarks
Prediction	2.5×10^{-4}	Prediction missing ply shear
True	0.13	Due to ply shear

C. Failure-criterion characterization

As in the previous example, a matrix of experiments with two variables θ and α is employed following the matrix setting shown in Table 2 as well. We fit three types of PRS (2nd, 3rd and 4th order) and Kriging to the mean values out of the repetitive observations of each design point. It turned out that Kriging offers the smallest cross validation error. Figure 12 shows the cross validation error of Kriging. Note that the mean value of the cross validation error out of 100 repetitions of test results is used in order to average out the randomness of the experiments. Compared to the cross validation error of Kriging fitted to the noise-free data, it is seen that the error is mainly due to the modeling error of the regression model rather than noise in the observations. Figure 13 shows RRMSE to the mean of failure loads evaluated at 165 (11x15) points in the matrix. It is clear that the test repetition does not contribute much to improving accuracy.
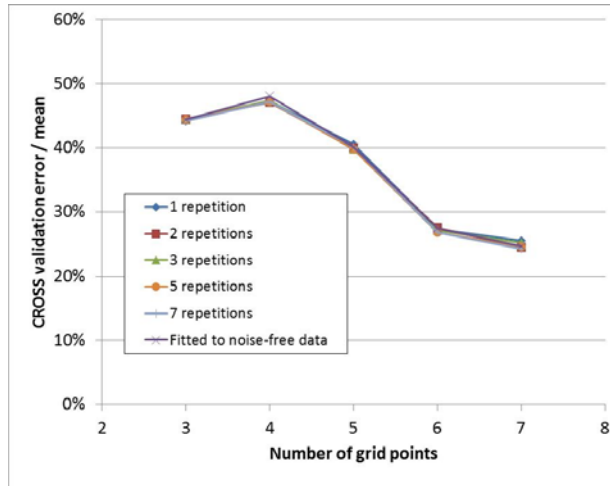
American Institute of Aeronautics and Astronautics

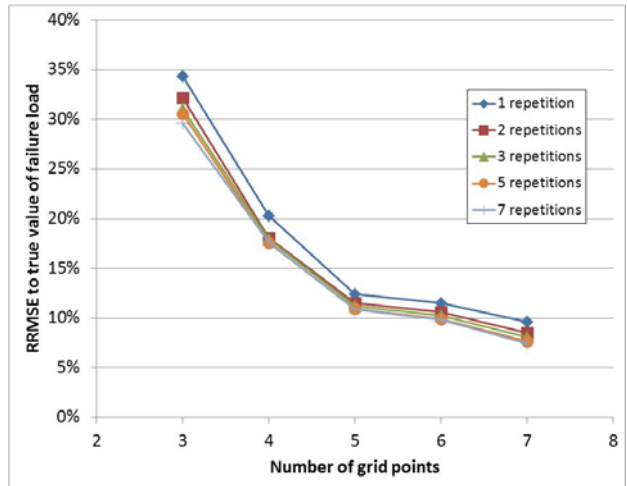**Figure 12. Cross validation error of regression model (Kriging)**

Figure 13. Relative root mean square error to true failure load (RRMSE, Kriging)

Figure 14 shows the tradeoff between the accuracy of the failure load prediction and the chance of discovering the unexpected failure mode with respect to the total number of tests. The matrix experiments with more than four grid points can discover that the laminate may fail by ply shear. Unlike the previous bracket problem, increasing the number of grids substantially surpasses the repetition both in terms of accuracy and in terms of discovering the unexpected mode. For example, 7 grid points with only one repetition has the better accuracy than any other experiments with the higher number of repetitions. It can be said that in the situation of the complicated failure surface, the importance of the higher number of grid points is intensified.
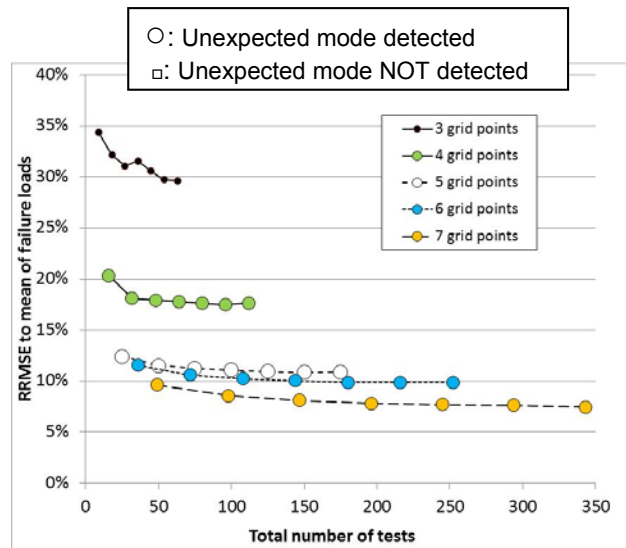

**Figure 14. RRMSE vs. Total number of tests
and whether or not unexpected mode being discovered**
On each line, markers represent the number of repetitions, 1 thought 7from left.

V.   Conclusion

Failure-criterion characterization for two structural elements is demonstrated in order to shed light on key factors of effective test management in terms of discovering unexpected failure modes and pursuing accuracy of failure load prediction. We first examine the impact of missing unexpected failure modes on safety assessment. It is shown that missing the existence of critical failure modes could lead to a substantially underestimated probability of failure by two orders of magnitude for the composite laminate example. Then, the allocation of tests

between the number of points exploring within the matrix and the number of repetitions intended to reduce noise of observation is investigated. It is observed that increasing the number of test points for exploration is vital to spot unexpected failure modes, at the same time, it is as good as repetition for reducing the effect of noise. Furthermore, the importance of the number of grid points for the matrix experiments is amplified when the failure load surface becomes complicated; concluding that the number of exploring points is the most critical factor for the tests of failure-criterion characterization.

Acknowledgments

References

[1]   "Aviation Statistical Reports," National Transportation Safety Board, [online database], http://www.ntsb.gov/data/table5.html [retrieved 28 March, 2012].

[2]   "Fatality Analysis Reporting System," National Highway Traffic Safety Administration, [online database], http://www-fars.nhtsa.dot.gov [retrieved 28 March, 2012].

[3]   "Fact Sheet - Commercial Aviation Safety team," Federal Aviation Administration, 1 December 2011.

[4]   Mohaghegh, M., "Evolution of structures design philosophy and criteria," *Journal of Aircraft,* Vol. 42, No. 4, 2005, pp. 814-831.
        doi: 10.2514/1.11717

[5]   Pitt, S., and Jones, R., "Multiple-site and widespread fatigue damage in aging aircraft," *Engineering Failure Analysis,* Vol. 4, No. 4, 1997, pp. 237-257.
        doi: doi:10.1016/S1350-6307(97)00020-4

[6]   A. Fawcettl., J. T., S. Ward., "777 EMPENNAGE CERTIFICATION APPROACH," *11th International Conference on Composite Materials*, Australia, 1997.

[7]   Marshall Rouse, D. C. J., David M. McGowan, Harold G. Bush, W. Allen Waters. "Utilization Of The Building-Block Approach In Structural Mechanics Research," *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference*, Austin, Texas, 2005.

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

[9]   Frangopol, D. M., and Maute, K., "Life-cycle reliability-based optimization of civil and aerospace structures," *Computers & Structures,* Vol. 81, No. 7, 2003, pp. 397-410.
        doi: Doi 10.1016/S0045-7949(03)00020-8

[10]   Acar, E., Haftka, R. T., and Kim, N. H., "Effects of Structural Tests on Aircraft Safety," *Aiaa Journal,* Vol. 48, No. 10, 2010, pp. 2235-2248.

[11]   An, J., Acar, E., Haftka, R. T., Kim, N. H., Ifju, P. G., and Johnson, T. F., "Being Conservative with a Limited Number of Test Results," *Journal of Aircraft,* Vol. 45, No. 6, 2008, pp. 1969-1975.
        doi: Doi 10.2514/1.35551

[12]   Acar, E., Kale, A., and Haftka, R. T., "Comparing effectiveness of measures that improve aircraft structural safety," *Journal of Aerospace Engineering,* Vol. 20, No. 3, 2007, pp. 186-199.

[13]   Villanueva, D., Haftka, R. T., and Sankar, B. V., "Including the Effect of a Future Test and Redesign in Reliability Calculations," *Aiaa Journal,* Vol. 49, No. 12, 2011, pp. 2760-2769.
doi: Doi 10.2514/1.J051150

[14]   James J. Herbert, L. H. T., "Development of Factors of Safety for Structural Analysis and Verification of Cryogenic Structures Using Probabilistic Methods," *47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Newport, Rhode Island, 2006.

[15]   Myers, R. H., Montgomery, D.C. . *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York, Wiley, 1995.

[16]   Simpson, T. W., and Mistree, F., "Kriging models for global approximation in simulation-based multidisciplinary design optimization," *Aiaa Journal,* Vol. 39, No. 12, 2001, pp. 2233-2241.

[17]   Viana, F. A. C., Haftka, R. T., and Steffen, V., "Multiple surrogates: how cross-validation errors can help us to obtain the best predictor," *Structural and Multidisciplinary Optimization,* Vol. 39, No. 4, 2009, pp. 439-457.
doi: DOI 10.1007/s00158-008-0338-0