Random Fields as An Alternative to Monte Carlo and Analytical Methods for Uncertainty and Reliability Analyses

LANL UQWG

2/21/02 (+)

David Robinson

Risk and Reliability Analysis Dept

http://reliability.sandia.gov/crax/robin.html

505-844-5883

Sandia National Laboratories

Outline

- Structural Reliability Analysis Methods
- Field Analysis Methodology
  - Spatial Estimation of Response Function
  - Failure Probability Estimation
- Generation of Samples
  - Pseudo-MC
  - Quasi-MC
  - Comparison
- Test Cases
- Stockpile applications

Structural Reliability Methods

- Analytical Methods
  - general procedure:
    - » iterative selection of samples from unknown performance function
    - » regression methods to approximate performance function
    - » search technique to find most probable point(s)
  - no info regarding sample points (success/failure) is used
- Sampling Methods
  - general procedure:
    - » sample of observations from unknown performance function
    - » success/failure evaluation at each point
  - no info on performance function is used in analysis
- Importance sampling lies between these extremes and therefore utilizes more information in making probability estimates offering one explanation as to why IS methods are so efficient
- Proposed Approach - Field Analysis Method
  - also lies between the extremes and can be used with any sampling method, including importance sampling, to improve efficiency
  - utilizes spatial statistics to probabilistically characterize the likelihood of any point being the success or failure region

Spatial Estimation of Response Function

- Assumptions:
  - Underlying response model fixed but unknown function of random variables

    $$Z(\mathbf{s}) = Z(x_1, x_2, \ldots, x_d)$$

  - $Z(\mathbf{s})$ can be locally characterized by a linear combination of known functions:

    $$Z(\mathbf{s}) = \sum_{i=1}^{p+1} f_{i-1}(\mathbf{s})\beta_{i-1} + \delta(\mathbf{s})$$

  - where: $\delta(\mathbf{s})$ is an intrinsically stationary random process, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ is a vector of unknown parameters and $f_i(\mathbf{s})$ are an independent set of known functions, i.e $x_1^0, x_1^1, x_1 x_2, x_1^2, \ldots$

- Expected value and covariance of $Z(\mathbf{s})$ can be estimated:

$$E[Z(\mathbf{s})] = m$$

$$E\{[Z(\mathbf{s}) - m][Z(\mathbf{s}+\mathbf{h}) - m]\} = C(\mathbf{s}, \mathbf{s}+\mathbf{h}) = C(|\,\mathbf{h}\,|)$$

- A specific form must be chosen for $\mathbf{C}(\mathbf{h})$, however the analysis is very robust to the particular form chosen

Spatial Estimation of Response Function (cont.)

- An estimate the mean of the response at any particular point in the design space is composed of a linear combination of neighboring observations

$$\hat{Z}(\mathbf{s}_o) = \sum_{i=1}^{N} \lambda_i Z(\mathbf{s}_i)$$

where the weights are found from:

$$\mathbf{w} = \left(\lambda_1, \lambda_2, \ldots, \lambda_N, -\beta_0, -\beta_1, \ldots, -\beta_p\right)^T = \mathrm{K}^{-1}\mathbf{c}$$

given:

$$\mathbf{c} = \left(C(\mathbf{s}_o - \mathbf{s}_1), \ldots, C(\mathbf{s}_o - \mathbf{s}_N), f_0(\mathbf{s}_o), \ldots, f_p(\mathbf{s}_o)\right)^T$$

$$\mathbf{K} = \begin{cases} C(\mathbf{s}_i - \mathbf{s}_j), & i = 1, \ldots, N; \quad j = 1, \ldots, N \\ f_{j-1-N}(\mathbf{s}_i), & i = 1, .., N; \quad j = N+1, \ldots, N+p+1 \\ 0, & i = N+1, \ldots, N+p+1; \quad j = N+1, \ldots, N+p+1 \end{cases}$$

An estimate of the variance of the response at any point can then be estimated: $\sigma_e^2(\mathbf{s}_o) = C(0) - \mathbf{w}^T\mathbf{c}$

Failure Probability Estimation

- At each point on the response surface an estimate of the mean and variance of the estimation error is now available
- Under the assumption of a Gaussian error process, it is possible to estimate the probability that any selected point is a member of either the success or failure region

$$\Pr\{Z(\mathbf{s}_i) < z_{crit} \mid \mathbf{S} = \mathbf{s}_i\} = \Phi\left(\frac{Z_{crit} - \mu(\mathbf{s}_i)}{\sigma(\mathbf{s}_i)} \right)$$

- If N points are selected then an estimate of the probability of failure is given by:

$$p_f = \frac{\sum_{i=1}^{N} \Phi\left(\frac{Z_{crit} - \mu(\mathbf{s}_i)}{\sigma(\mathbf{s}_i)} \right)}{N}$$

Pseudo-random Sampling

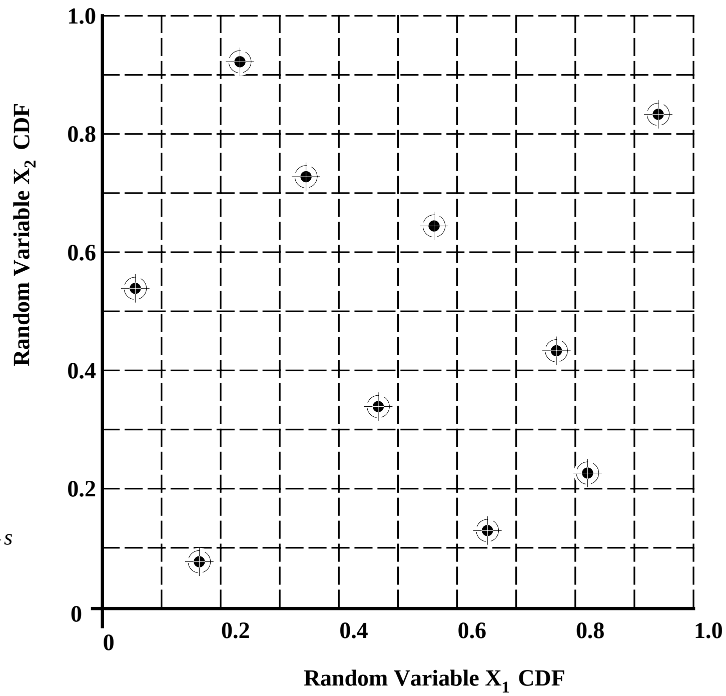
- Monte Carlo
  - developed in nuclear weapons programs in the 1940's
  - let $I^s = [0,1]^s$ be a s -dimensional cube and let $f(\mathbf{t})$ be defined on I^s
  - let $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ be a pseudo-random sample of N points from I^s
  - Given these samples Monte Carlo analysis provides an approximation of a continuous average with discrete average

$$\int_{I^s} f(\mathbf{t}) d\mathbf{t} \approx \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i)$$

  - **PLUS**:
    - » sampling can be conducted sequentially (easy to add new samples)
    - » error bounds not dependent on dimension s $\quad O\left(N^{-1/2}\right)$
  - **MINUS**:
    - » Probabilistic error bounds depends on equidistribution of sample points in I^s
    - » no methodical method of constructing sample to achieve error bound, therefore
    - » rate of convergence is very slow

Pseudo-random Sampling

- Latin Hypercube Sampling
  - also based on pseudo-random sampling
  - form of stratified sampling in which
    the samples are 'forced' to be dispersed across the support space
  - number of samples dictates the number of regions
  - **PLUS**:
    - » significant reduction in number of samples compared to traditional MC
  - **MINUS**:
    - » samples do not provide good uniformity across I^s
    - » samples can not be generated sequentially



Quasi-random Sampling

- Quasi-random sample is commonly referred to as a low-discrepancy sequence
- Low discrepancy sequence is one that places sample points nearly uniformly in the sample space of interest
- Low-discrepancy $\rightarrow$ low integration error
- Deterministic error bounds - $O\left(N^{-1}(\log N)^{k-1}\right)$
- Variety of sequences
  - Halton (simple, leaped, RR2)
  - Hammersley
  - Fauer
  - Sobol

Simple Halton Sequence

- Defined in *s*-dimensional space by using *s* prime bases to generate a sequence of N quasi-random vectors

$$x_n = \left(\Phi_{b_1}(n), \Phi_{b_2}(n), \ldots, \Phi_{b_s}(n)\right), \qquad n = 1, 2, \ldots, N$$

where the radix inverse function is defined:

$$\Phi_{b_j}(n) = 0.n_0 n_1 \cdots n_s = n_0 b_j^{-1} + n_1 b_j^{-2} + \cdots + n_s b_j^{-s-1}$$

$$= \sum_{i=0}^{s} n_i b_j^{-i-1}$$

- Integer coefficients n_i ($0 \leq n_i \leq b_j$) result from expansion of integer n in base b_j :

$$n = n_s n_{s-1} \cdots n_2 n_1 n_0 = n_0 + n_1 b_j + n_2 b_j^2 + \cdots + n_s b_j^s$$

$$= \sum_{i=0}^{s} n_i b_j^i$$

Example

Halton sequence for N=6, s=3

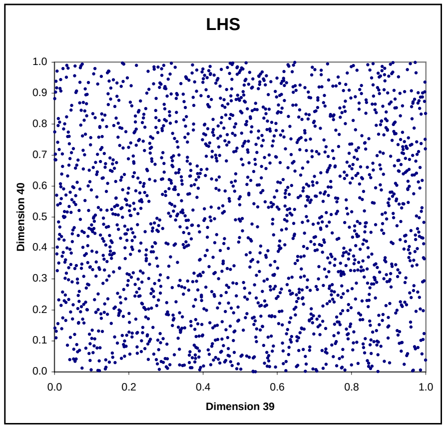
n	x_1 $b_1 = 2$	x_2 $b_2 = 3$	x_3 $b_3 = 5$
1	0.5	0.333333	0.2
2	0.25	0.666667	0.4
3	0.75	0.111111	0.6
4	0.125	0.444444	0.8
5	0.625	0.777778	0.04
6	0.375	0.222222	0.24

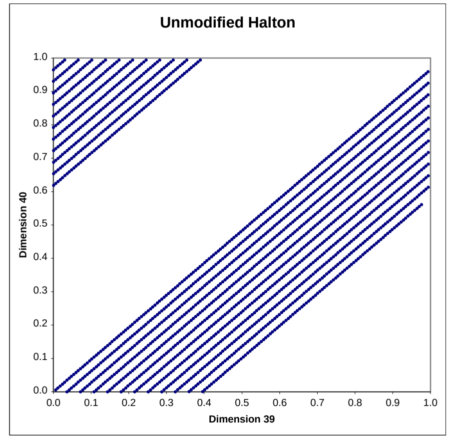
Example

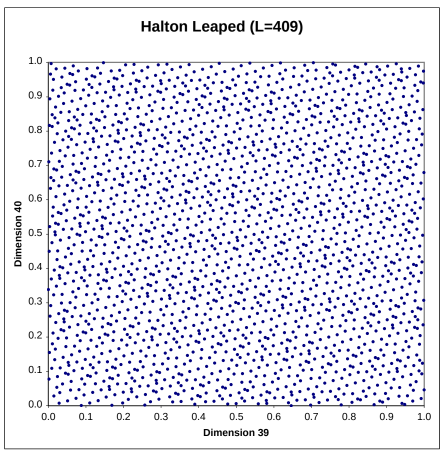
Halton sequence for N=12, s=3

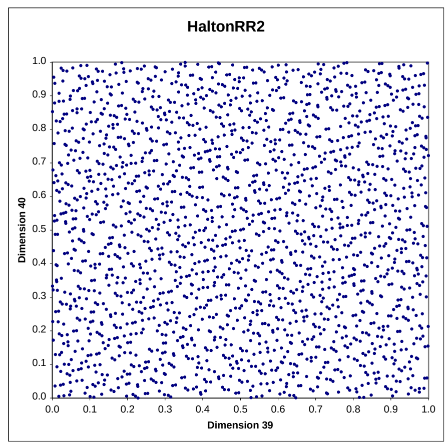
n	x_1 $b_1=2$	x_2 $b_2=3$	x_3 $b_3=5$	n	x_1 $b_1=2$	x_2 $b_2=3$	x_3 $b_3=5$
1	0.5	0.333333	0.2	7	0.875	0.555556	0.44
2	0.25	0.666667	0.4	8	0.0625	0.888889	0.64
3	0.75	0.111111	0.6	9	0.5625	0.037037	0.84
4	0.125	0.444444	0.8	10	0.3125	0.37037	0.08
5	0.625	0.777778	0.04	11	0.8125	0.703704	0.28
6	0.375	0.222222	0.24	12	0.1875	0.148148	0.48

Discrepancy



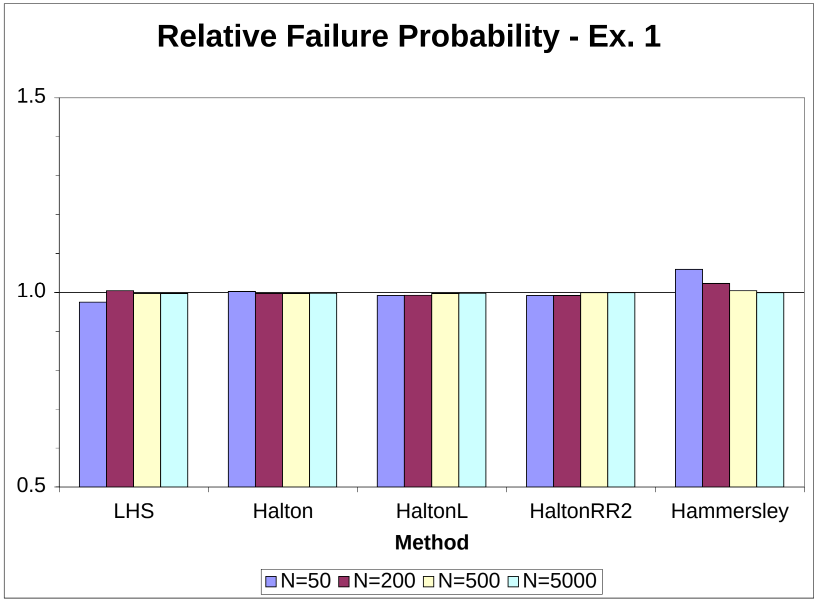


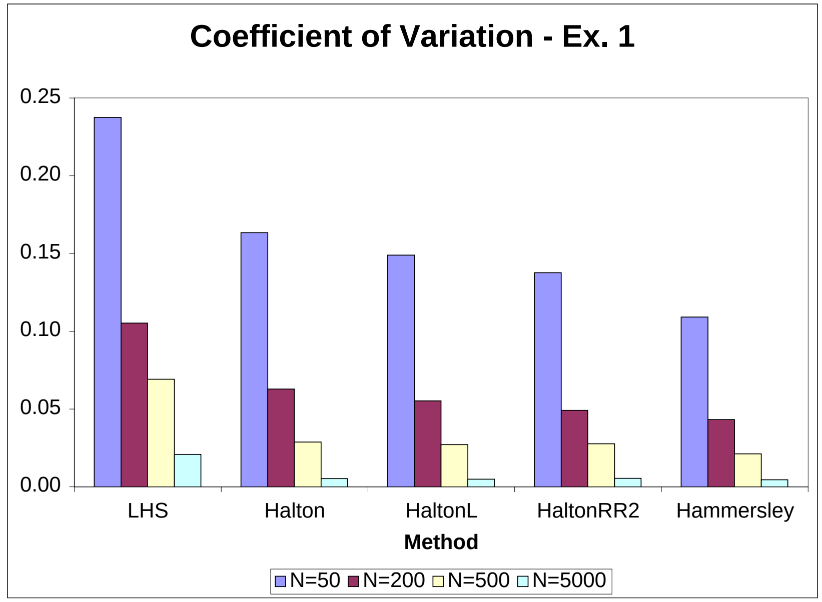




Comparisons

Example 1: $g_1 = \beta\sqrt{n} - \sum_{i=1}^{n} U_i$

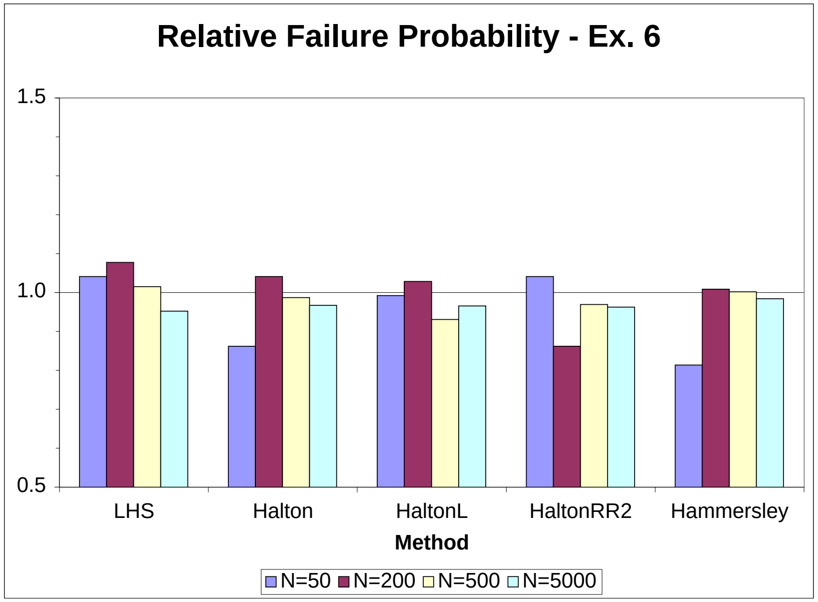


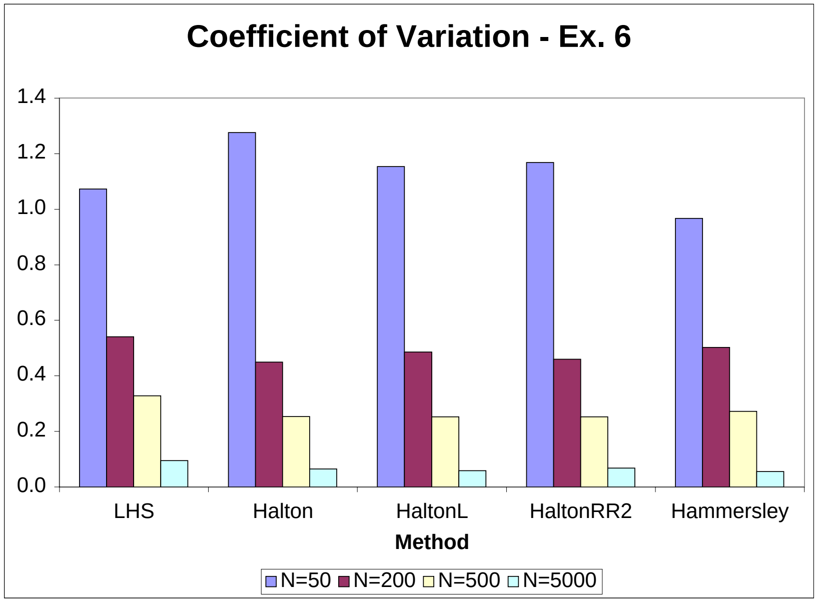


Comparisons

Example 6: $g_6 = X_1 + 2X_2 + 2X_3 + X_4 - 5X_5 - 5X_6$

$$+0.001\sum_{i=1}^{6}\sin(100X_i)$$





q-MC Discussion

- Overall the Halton Leaped quasi-Monte Carlo sampling proved to have lower mean estimation error and/or have faster convergence
- There were unique cases where LHS was better however:
  - primarily for very small samples and
  - repeated samples were inconclusive (sometimes better/worse)
- Major benefit of Halton-type sequences was the ability to sequentially add samples as results converged - this is not possible with LHS
- Performance function evaluation is becoming computationally burdensome (100K processor hours for single evaluation)
- Ability to reduce the number of samples and sequentially sample is becoming critical

Test Cases

- Test Cases (all treated as 'black boxes - no derivative info used)
  - Case 1: $Z_1(x_1, x_2) = 2x_2 + x_1 x_2 - (x_1 - 1)^2 - 2$

    with:

    $X_1 \sim Ln(\mu = 3, \sigma = 1.5)$

    $X_2 \sim Ln(\mu = 3, \sigma = 2.25)$

  - Case 2:

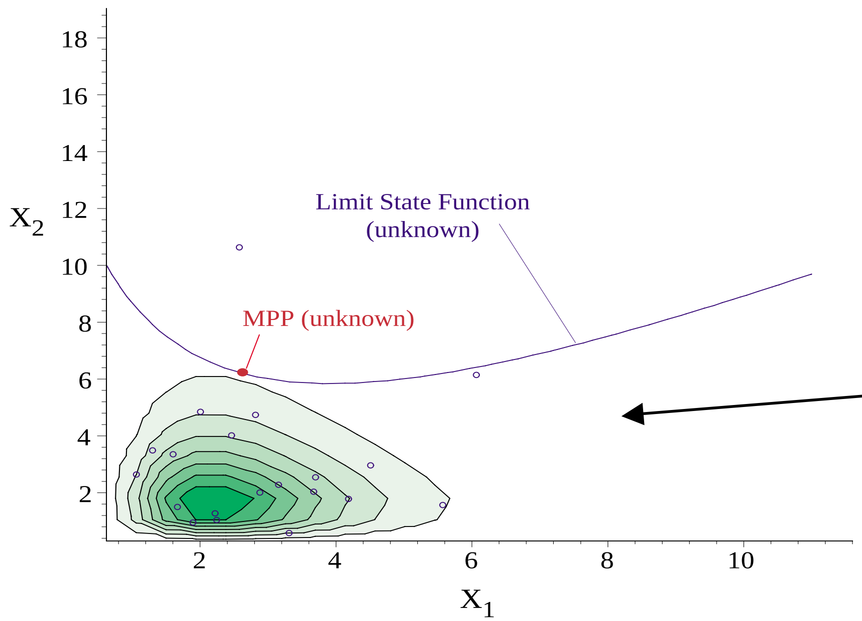
$$Z_2(x_1, x_2, x_3, x_4) = k_1 \sqrt{\frac{12 x_1 x_3^2}{x_2 x_4^{k_2}}}$$

where: $k_1 = 3.52, k_2 = 4.0$

Random Variable	Mean	Coefficient of Variation
x_1	10000000	0.03
x_2	0.00025	0.05
x_3	0.980	0.05
x_4	20.0	0.05

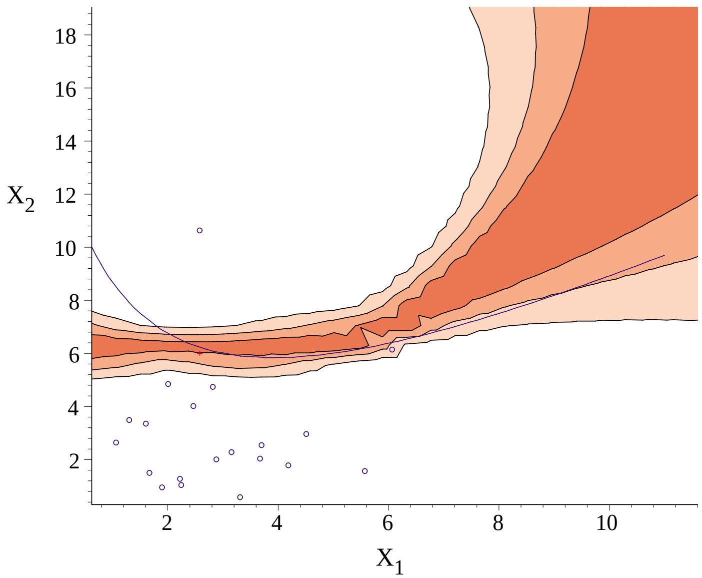
Sample Generation

- Sample Generation
  - quasi-Monte Carlo
  - use of Halton sequence permits iterative generation of samples
  - could use importance sampling methods to focus new samples, but was not beneficial in these two test cases



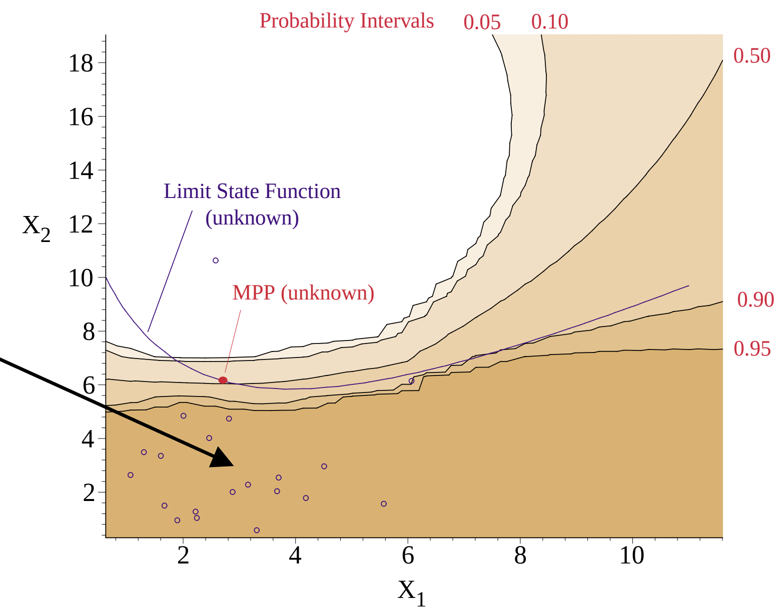
Typical: Sample for Case 1 based on likelihood function

Intermediate Results - Case 1
(based on only first 20 samples)



PDF representation-
The darker the color the higher the probability that the region contains the limit state function

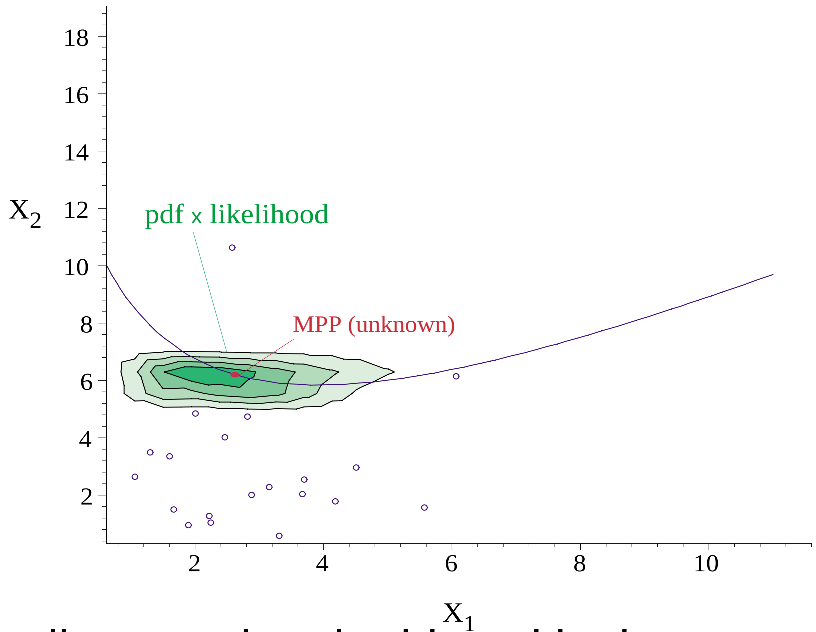
CDF representation-
Colors represent lines of equal probability



Sandia National Laboratories

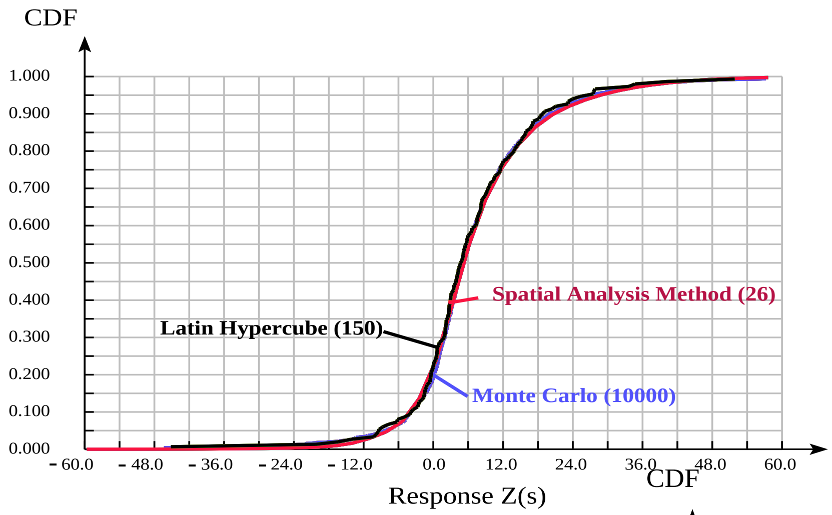
MPP Location and Resampling

- Probability-based identification of the location of MPP(s) can be accomplished by combining the likelihood function and the spatial PDF of the random field



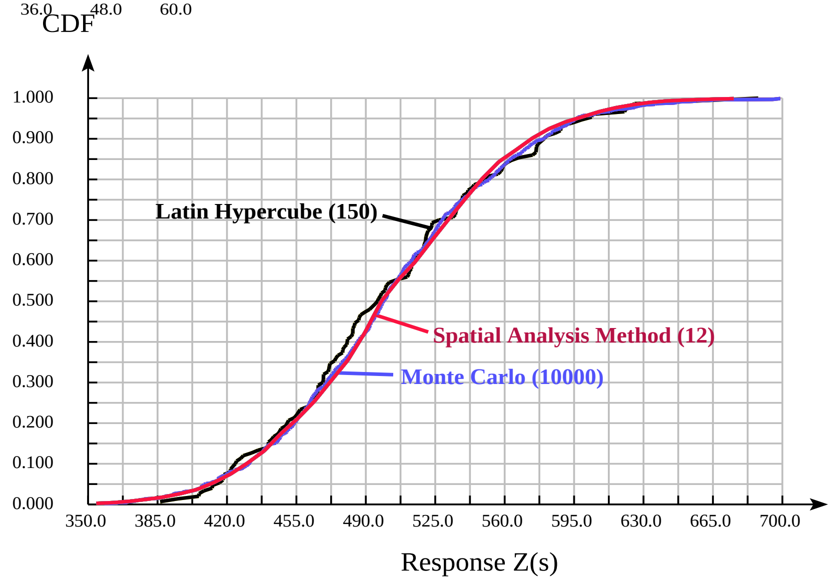
- Resampling can then be biased in the area of the MPP locations similar to the approach used in various importance sampling methods

Results - Case 1 and Case 2



Case 1

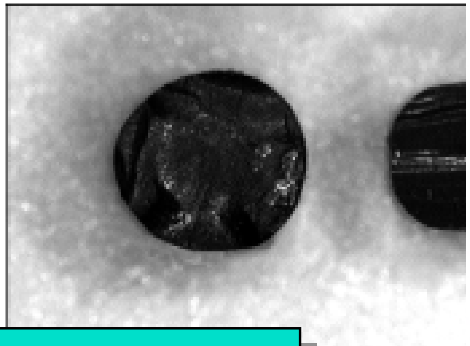
Case 2



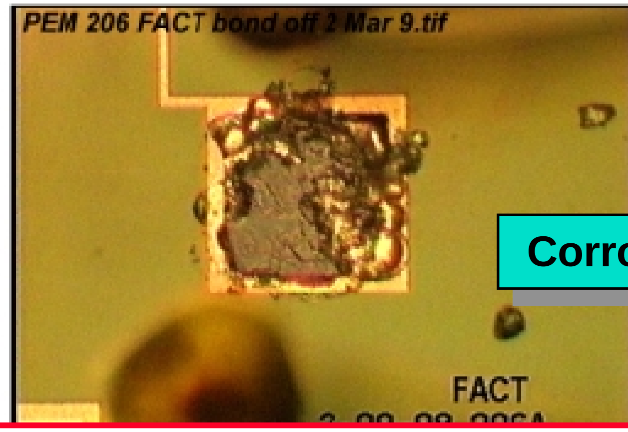
Conclusions

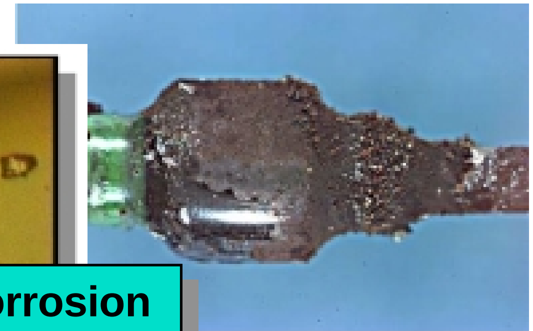
- Good things:
  - For all test cases investigated thus far, the new method requires significantly fewer function evaluations in comparison with traditional analytically-based FORM/SORM methods
  - The number of function evaluations is not dependent on the dimension of the design space
  - Without sampling bias, the number of function evaluations will always be less than or equal to the number required for a full quasi-Monte Carlo evaluation and pseudo-MC methods such as LHS
- Bad things:
  - New method can run into numerical problems in the far extremes of the design space

Cassandra Project



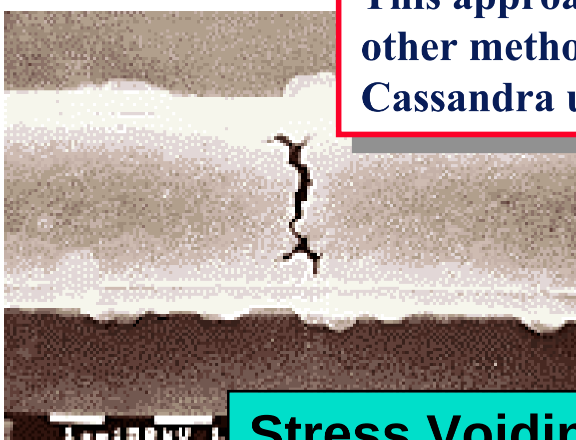
Polymer Aging





Corrosion

This approach along with a large number of other methods, is an integral part of the Cassandra uncertainty analysis library



Stress Voiding

Small Lot Manufacturing