SAND2013-4046C

Exceptional service in the national interest

Sandia National Laboratories

Validation Based on a Probabilistic Measure of Response

Dan Segalman, Sandia National Laboratories, Livermore, CA

Tom Paez, Thomas Paez Consulting, Sedona, AZ

Lara Bauman, Sandia National Laboratories, Livermore, CA

U.S. DEPARTMENT OF ENERGY | NNSA

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

Outline

- A new approach to quantification of margin and uncertainty
- Employment of that that approach in defining a validation metric

- What do we mean by Margin?

An Unfortunate Irony Of Life

Sandia National Laboratories

1. The less reliable a design is, the easier it is to make meaningful estimates for probability of failure.

   If there is a history of failure 25% of the time, we can estimate the probability of failure.

2. The more reliable it is, the less meaningful are efforts to quantify that reliability.

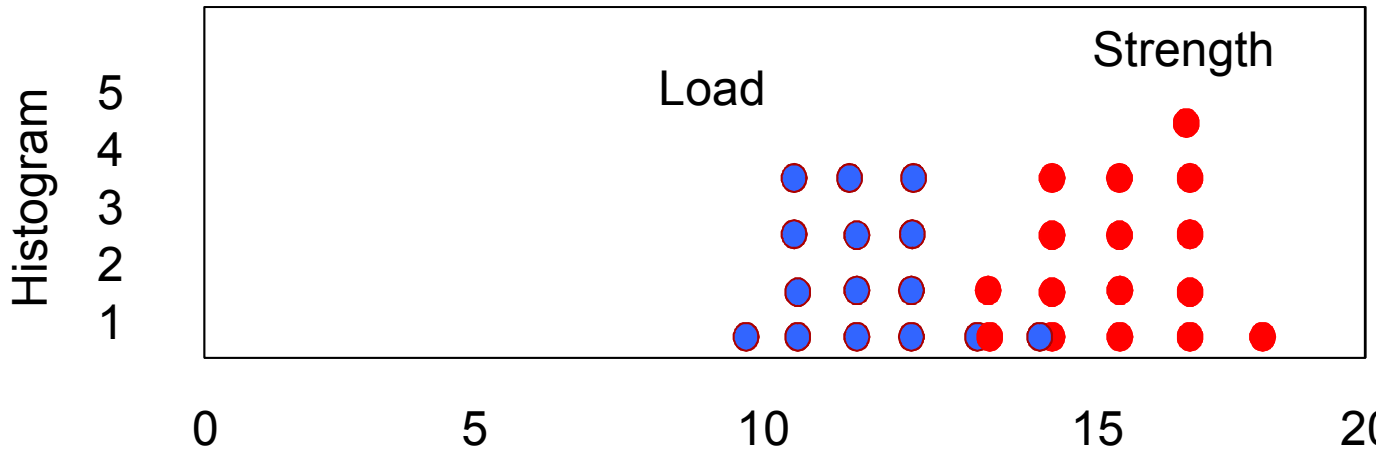
   If there has never been a failure, how do we estimate the probability of failure?

T2

Slide 4

T2 In 2, instead of likelihood I would use chance or probability.
Tom, 5/12/2013

Predicting Probability of Failure of Un-conservative Designs is Easy



Probability of failure $P_{F,d}$ approximated from statistics of experience T3

$$P_{F,d} = \frac{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y} H(x_i - y_j)}{N_x\, N_y}$$

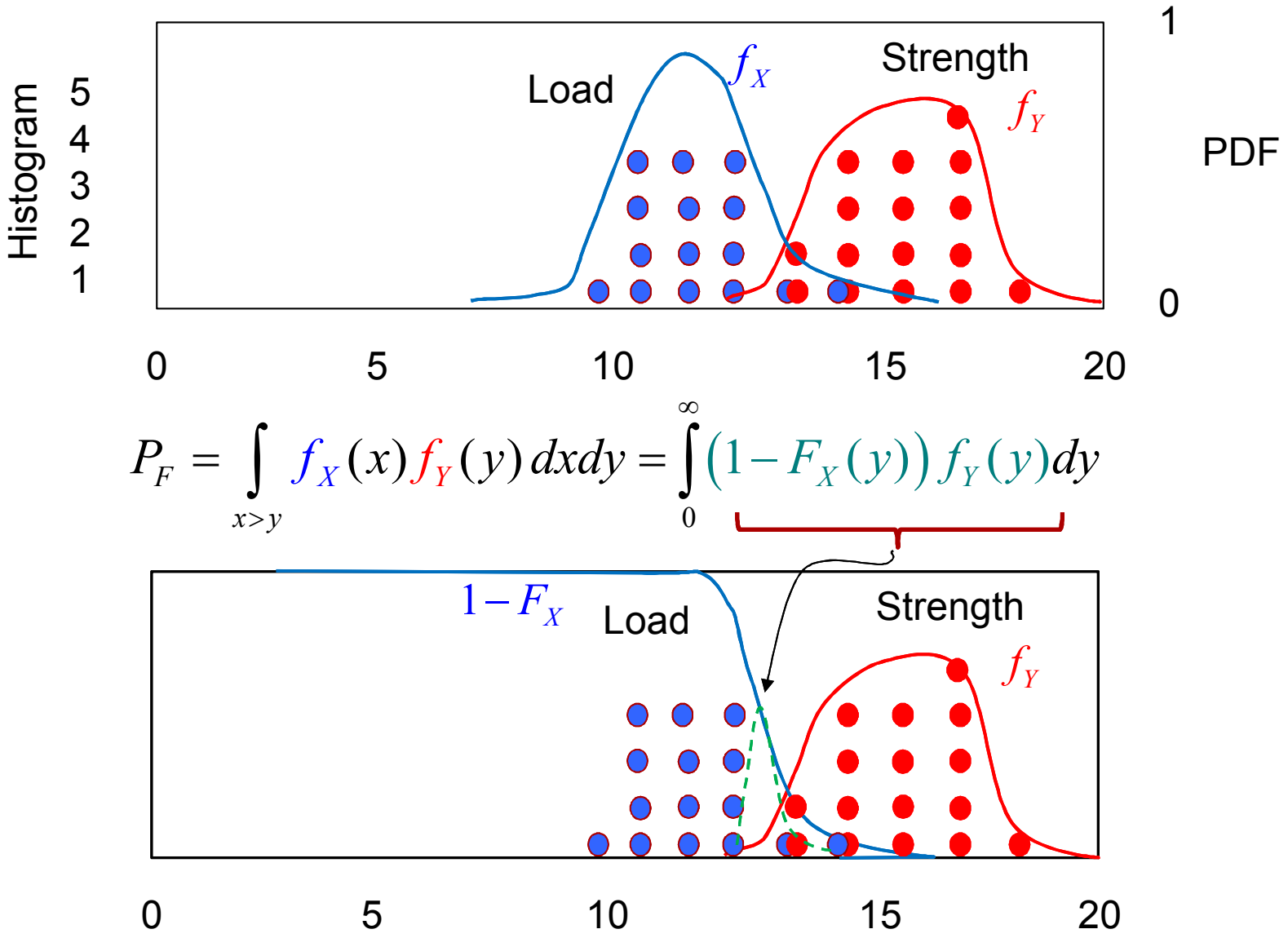
where

$$H(s) = \begin{cases} 0 & \text{if } s < 0 \\ 1/2 & \text{if } s = 0 \\ 1 & \text{if } s > 0 \end{cases}$$

Slide 5

T3 Leave the formulas in, but refer to the graph only.
Tom, 5/12/2013

Approximate Data by Continuous Probability Density Functions (PDFs) to Express Probability of Failure (PoF) in Integral Form



Slide 6

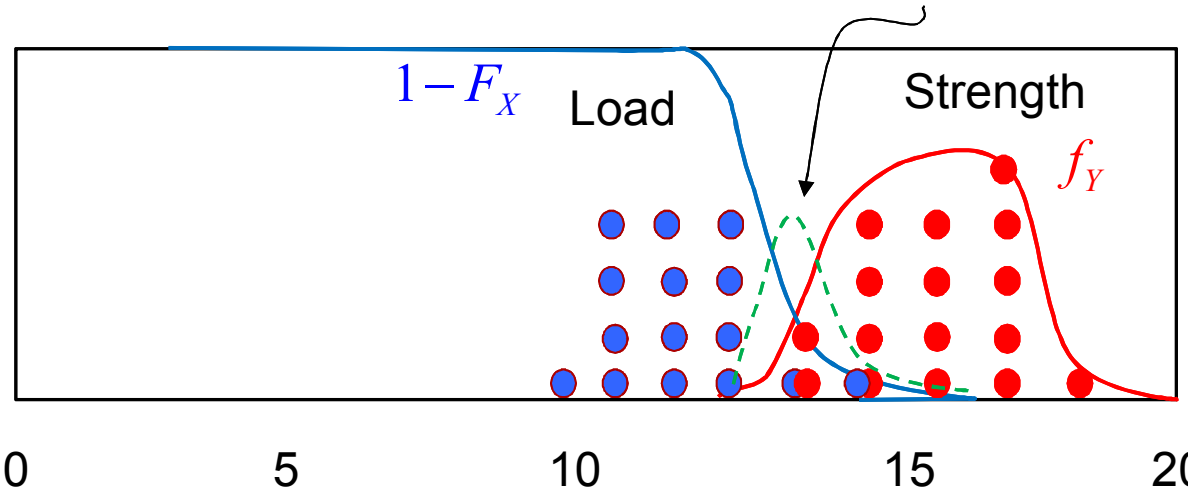
T4 Do not explain the entire equation - or any of it. Refer to final element, only, plus the lower graphic.
Tom, 5/12/2013

Predicting Probability of Failure of Un-conservative Designs is Easy

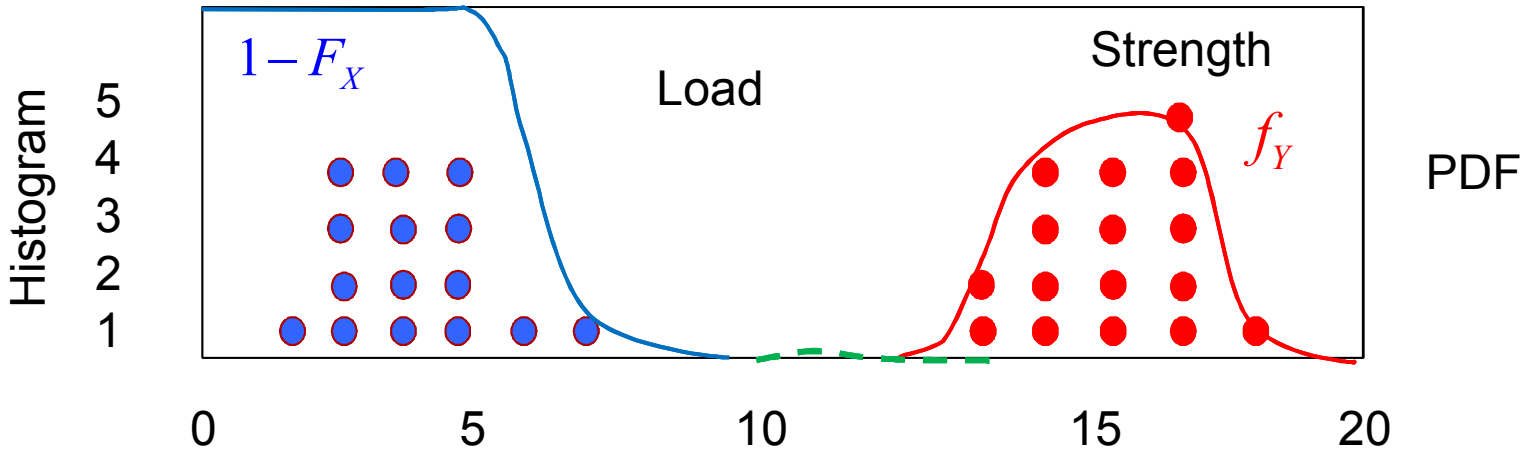
Sandia National Laboratories

1. Calculations of PoF occur where there is failure experience – it can be quantified directly from data.
2. The body of the integrand occurs in a “sweet spot” where there is an overlap of data.
3. The fitted PDFs are reasonable interpolations of the data there
4. Where there is data, the two approaches yield similar approaches.

$$P_F = \int_{x>y} f_X(x) f_Y(y)\, dxdy = \int_0^\infty \left(1 - F_X(y)\right) f_Y(y) dy$$



What Happens for Cases of Conservative Design



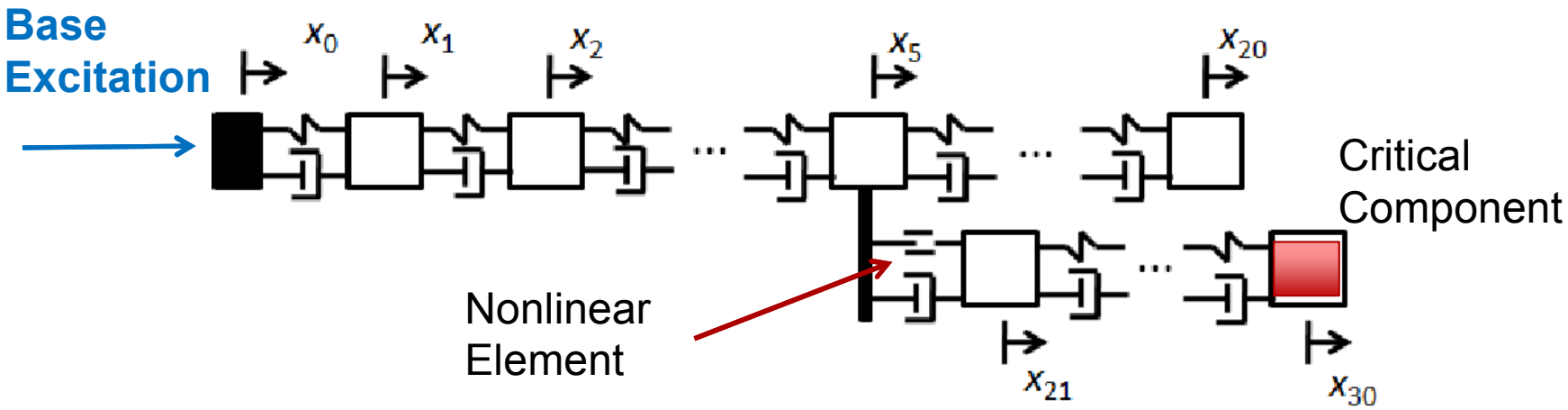
1. There is no overlap of Load and Strength data.
2. PoF cannot be calculated from experience
3. The PDFs are ***extrapolation*** of the data
   a) All the action is in the tails of the distributions – where there is no data
   b) Predicted PoFs are very sensitive to the form of distribution postulated
   c) The predicted PoFs are very sensitive to small changes in the data used to estimate the PDFs

Slide 8

T5 Make the word "extrapolation" italic, bold, and red.
Tom, 5/12/2013

Sandia National Laboratories

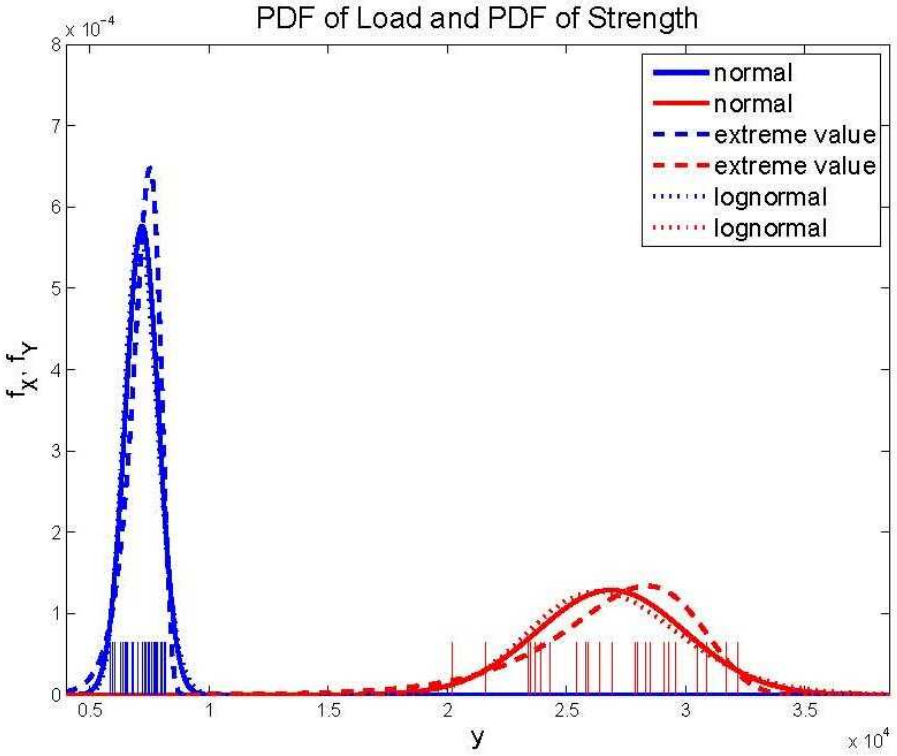
Consider a Nonlinear System



Slide 9

T6 Remove "of."
Tom, 5/12/2013

What Happens for Cases of Conservative Design: An Example



30 loads from Nonlinear (truth) model and 25 strengths sampled from simulations of Tom Paez. Data points are shown as ticks.

Three common forms of PDF are fitted by maximum likelihood for each of the load and strength data.

Each of these distributions appears a reasonable approximation of the data. All distribution forms had adequate Kolgomorov-Smirnov goodness-of-fit to the data.

What Happens for Cases of Conservative Design: An Example

We calculate the PoF for each combination of Load and Strength distribution form

$$P_F = \int_0^{\infty} \left(1 - F_X(y)\right) f_Y(y) dy$$

extreme value-lognormal	3.690e-24
normal-lognormal	1.248e-20
lognormal-lognormal	5.836e-18
extreme value-normal	2.259e-10
normal-normal	2.822e-10
lognormal-normal	3.472e-10
normal-extreme value	4.234e-04
extreme value-extreme value	4.242e-04
lognormal-extreme value	4.247e-04

T7

The predicted PoFs may vary by tens of orders of magnitude, depending on the distribution forms assumed.

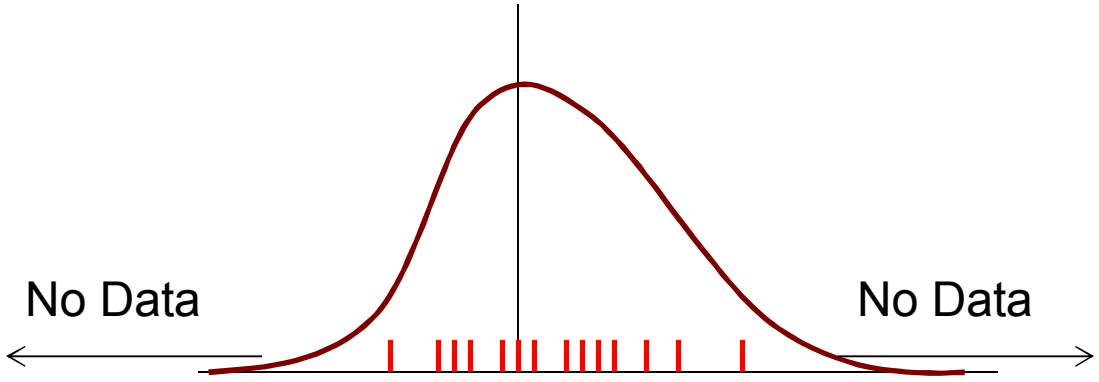
The problem is in the tails. When we postulate a distribution form, we are postulating the asymptotic nature of the tails.

Slide 11

T7 Cursory coverage only. Emphasize range of results.
Tom, 5/12/2013

Except Where Central Limit Theorem Applies (and that is not often in our business):

- Statistical fits to data should be considered interpolation.
- There is no reason to expect that a distribution form that fits data near the center of the distribution would fit data associated with the tails – even if that data were available. T8

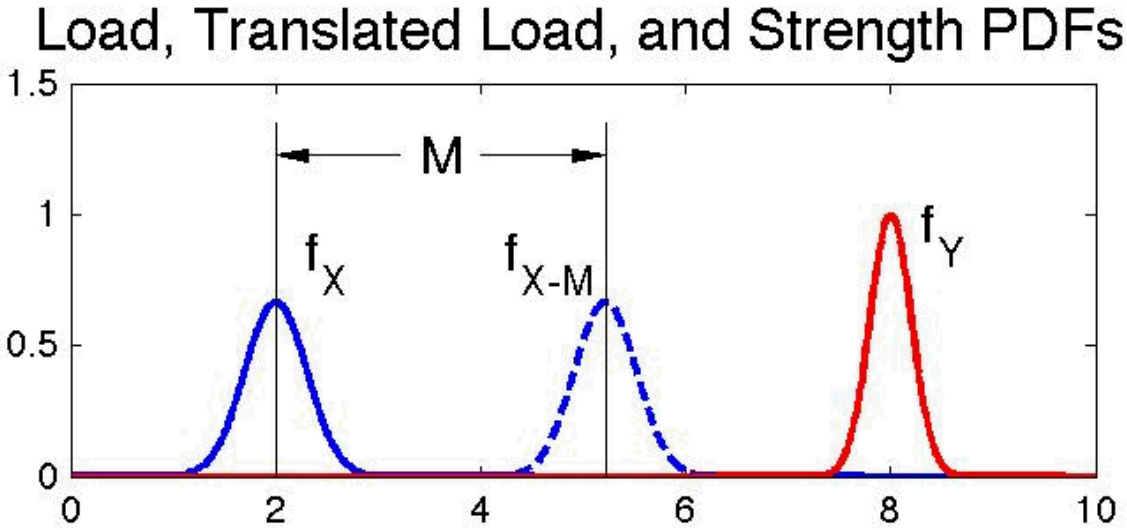


Slide 12

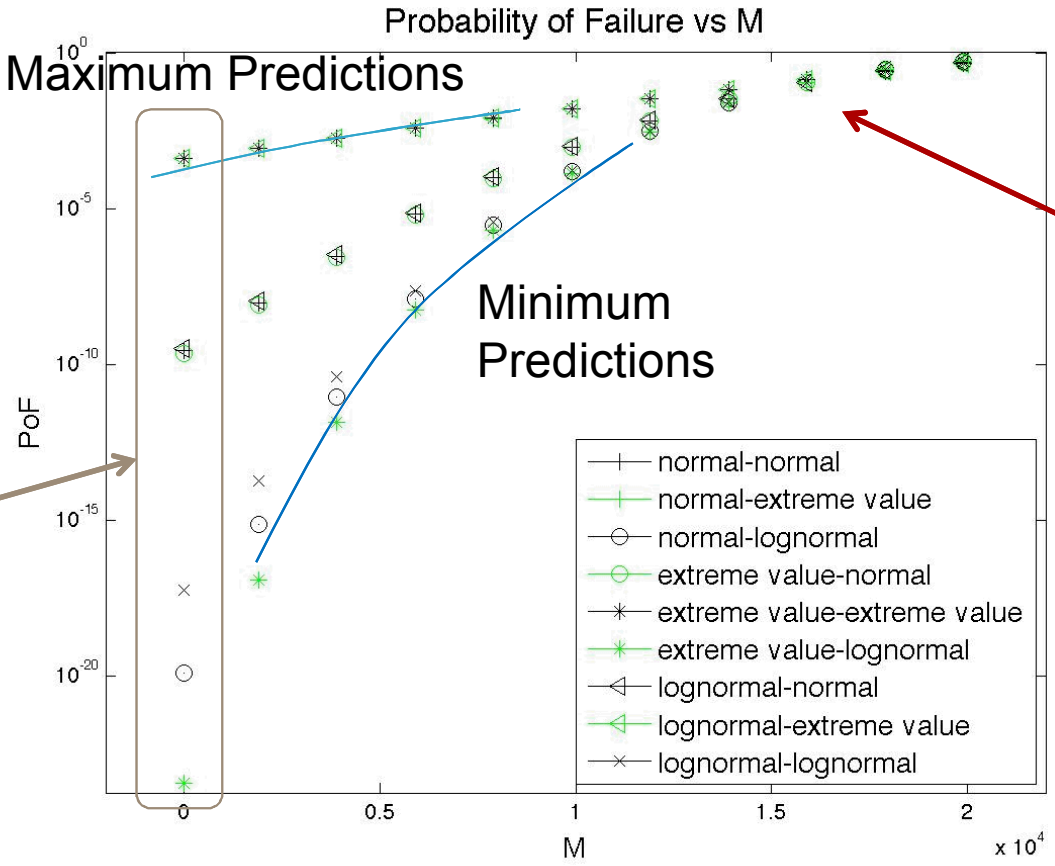
T8 Should reach this point within about five minutes. Is it possible?
Tom, 5/12/2013

A Thought Experiment

- Consider the calculation of Probably of Failure (PoF)
  1. Using the Load + M instead of original Load data
  2. Instead Fit PDFs to this new load distribution
  3. Perform failure integration



Plot Probability of Failure as a Function of M for the Discussed PDFs



Original Range of PoF

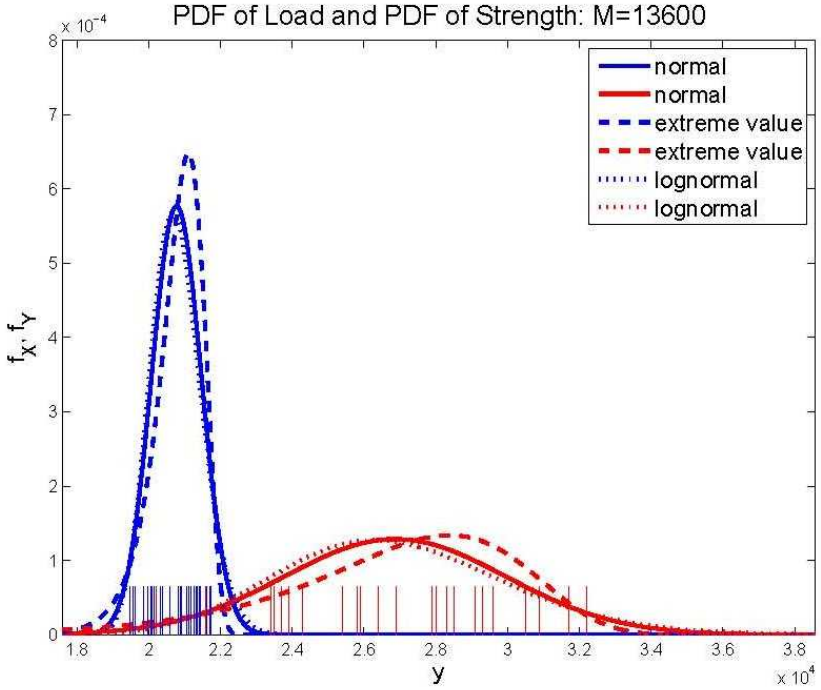
Convergence of Upper and Lower Estimates of PoF T9

As M increases the calculated PoFs begin to converge.
We can use this to choose margin.

Slide 14

T9 Difficult to convey - emphasize this point.
Tom, 5/12/2013

Reduction to a Simpler Problem



Calculated Probability of Failure for Load + 13600

extreme value-lognormal	2.054e-02
normal-lognormal	2.059e-02
lognormal-lognormal	2.091e-02
extreme value-normal	2.758e-02
normal-normal	2.755e-02
lognormal-normal	2.781e-02
normal-extreme value	6.259e-02
extreme value-extreme value	6.270e-02
lognormal-extreme value	6.274e-02

For Appropriate M,

- The problem reduces to the easier case where load and strength overlap.
- The predicted probability is much less dependent on the forms of distribution assumed and much less dependent on small changes in the data.
- Provides a probability-based definition for margin:

  Margin is the amount that must be added to load so that the calculated Probability of Failure is minimally dependent on assumed distribution forms.

Slide 15

T10 Show the M-value and then emphasize range.
Tom, 5/12/2013

Probability of Exceeding Margin

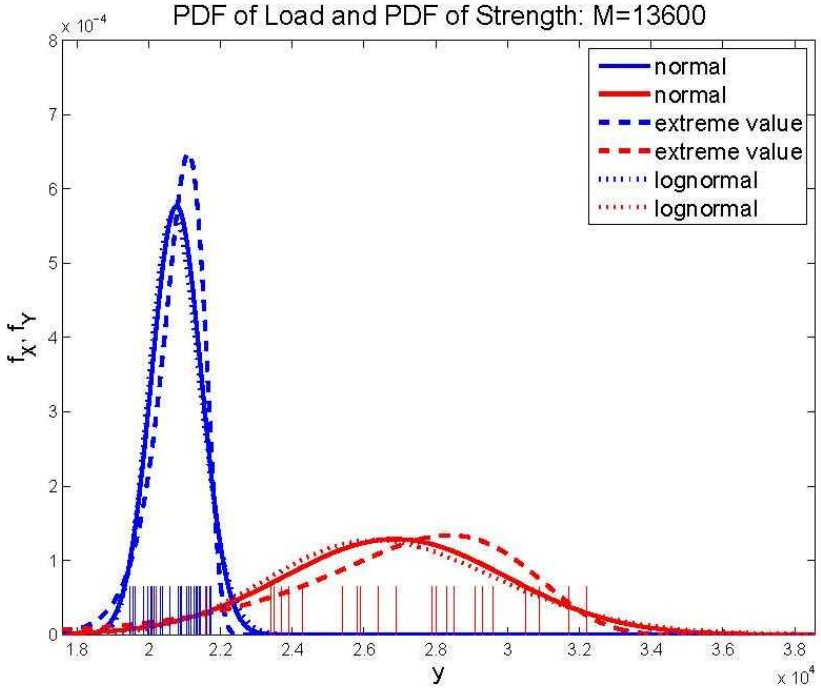
T11

$$P_F(M) = \int_{x+M>y} f_X(x) f_Y(y)\, dxdy = \int_0^\infty \left(1 - F_X(y-M)\right) f_Y(y)\, dy$$

Slide 16

T11 No extended explanation. Simply state "We can get the formula."
Tom, 5/12/2013

Integrated Statements about Margin and Probability



Calculated Probability of Failure for Load + M

extreme value-lognormal	2.054e-02
normal-lognormal	2.059e-02
lognormal-lognormal	2.091e-02
extreme value-normal	2.758e-02
normal-normal	2.755e-02
lognormal-normal	2.781e-02
normal-extreme value	6.259e-02
extreme value-extreme value	6.270e-02
lognormal-extreme value	6.274e-02

In this case:

The probability of Load exceeding Margin of 13600 is on the order of 6% T12

Slide 17

T12 Make this point only - no reiteration of range of values.
Tom, 5/12/2013

Key Features of this Approach

- Reasonably independent of
  - Forms postulated for PDFs
  - Character of tails
  - Small variations in available data
- Motivates estimate for Margin from probabilistic considerations of existent data
- Mathematically Defensible

T16

Slide 18

T16 Need to finish this slide within ten minutes without talking fast. 30 seconds per slide.
Tom, 5/12/2013

We Can Attach a Confidence to PEM

T17

We shall use a re-sampling method: bootstrap.

This permits us to estimate the empirical distribution of margins without specifying margin form.

Slide 19

T17 I would make the theme of this slide "We can evaluate confidence of PEM." Then skip slides 23 - 27.
Tom, 5/12/2013

For Example:

T18

- Using discrete (Step/Delta) forms for distributions with 80% confidence, $P(X+13500>Y)<9.3\%$
- Using KDE expansion for distributions with 80% confidence, $P(X+13500>Y)<11\%$

These are strong statements reasonably independent of the forms of distribution employed in the quadratures.

Slide 20

T18 Here, I would say "For example ...."
Tom, 5/12/2013

About Validation

- Definition: The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.
- Involves Experimental Data and Model:
  - Identify quantities critical to performance/failure
  - Perform validation experiments on system of interest; compute T19 measure(s) of system response.
    - Calculate PEM T20 using experimental data: $P_F^E(M)$
    - Use Bootstra T21 o calculate a pool of such values: $P_{F,k}^E(M)$
  - Create model and use it to predict response realized during validation experiments.
    - Calculate PEM using modeled values: $P_F^M(M)$
    - Use Bootstrap to calculate a pool of such values: $P_{F,k}^M(M)$
  - Compare

Slide 21

T19 Add: We will use PEM, (insert symbol from slide 36), as measure of interest.
Tom, 5/12/2013

T20 Add: Use responses to predict PEM, (Insert symbol from slide 36).
New bullet: If predicted PEM is satisfactorily accurate then model is validated.
Tom, 5/12/2013

T21 Insert slide 36 next.
then show slides 30 and 31.
Then state "We can perform analysis of confidence."
Tom, 5/12/2013

Define Model Goodness

- Define

$$v_M = \log\left(\frac{\overline{P_M^M}}{\overline{P_M^E}}\right)$$

- Note

$$\begin{cases} v_M > 0 \rightarrow & \text{conservative} \\ v_M = 0 \rightarrow & \text{accurate} \\ v_M < 0 \rightarrow & \text{nonconservative} \end{cases}$$

- How acceptable is this particular value of v_M ?
  - Compare to possible values

Goodness – Validity – of the Model is Assessed

T37

Sandia National Laboratories

- Remember, we had calculated sets $\left\{P_{F,k}^{E}(M)\right\}$ and $\left\{P_{F,k}^{M}(M)\right\}$
- Randomly sample to generate $\left\{v_{M,k}\right\} = \left\{\log\left(\frac{P_{M,n(k)}^{M}}{P_{M,m(k)}^{E}}\right)\right\}$
- Interpolate $v_{M,k}$ by KDE to obtain continuous distribution $f_V(v)$
- Examine where v_M fits in the distribution. If validation metric is satisfactorily near zero, then model is valid!

T36

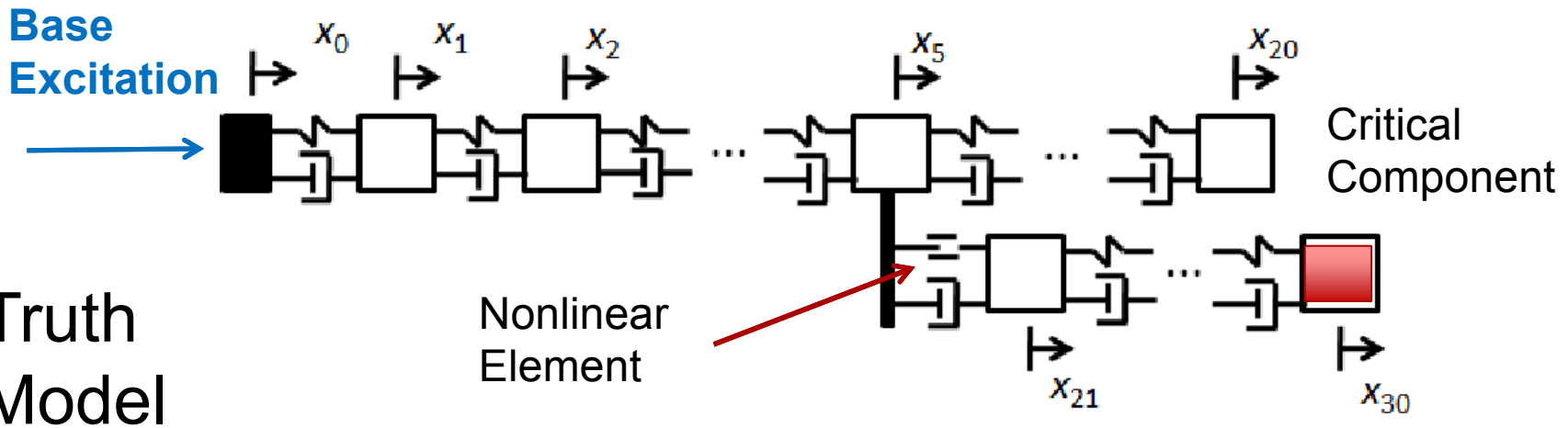
Slide 23

T36 Reiterate here "If validation metric is satisfactorily near zero, then model is valid!"
Tom, 5/12/2013

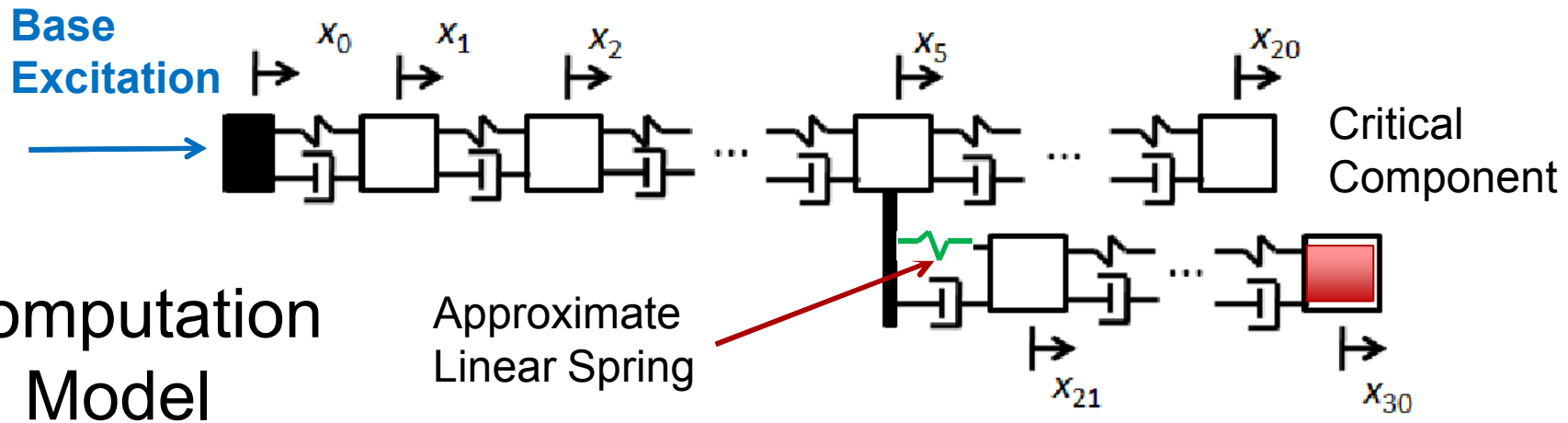
T37 Because details skipped, you need to emphasize "This can be done."
Tom, 5/12/2013

Return to Example Problem

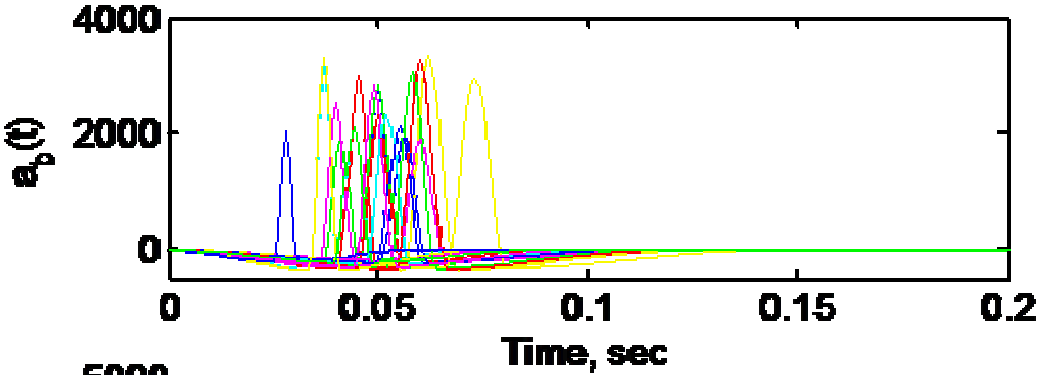
Sandia National Laboratories



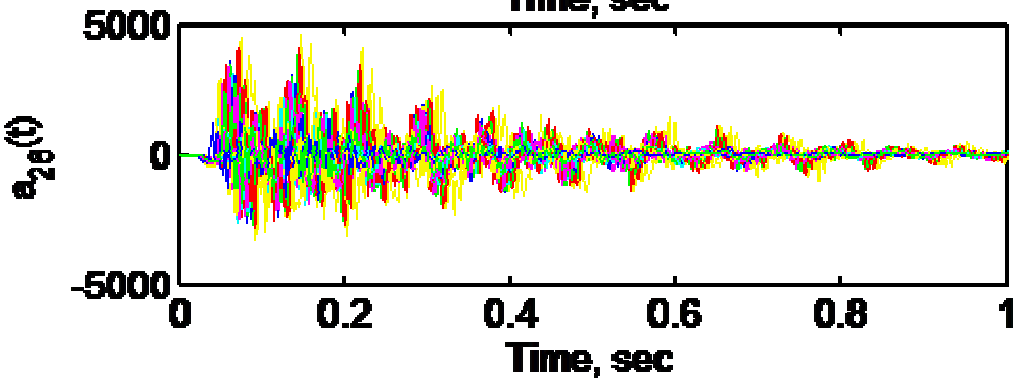
Truth Model



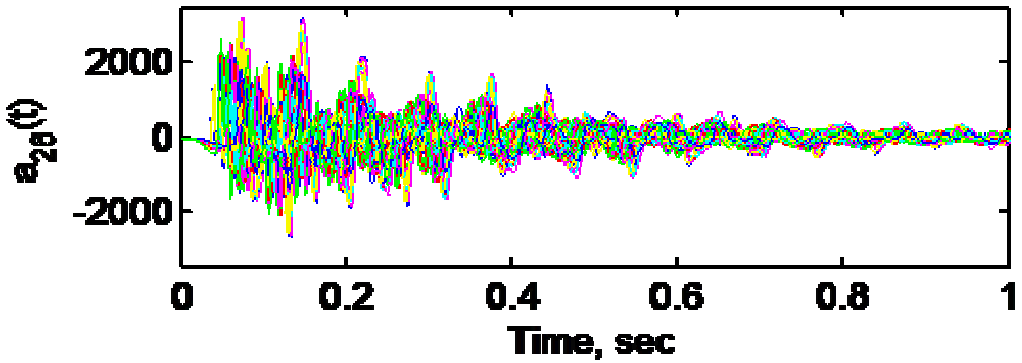
Computational Model



Thirty imposed base accelerations with randomly chosen parameters



Thirty realizations of critical component acceleration of Truth model

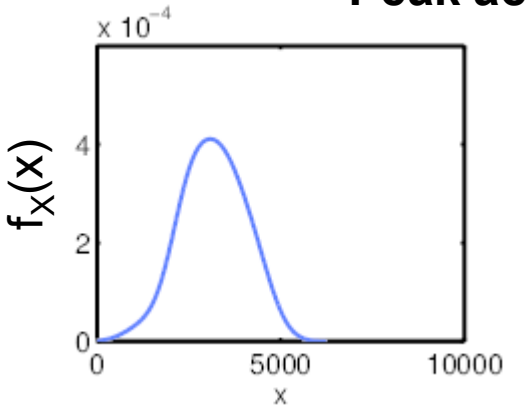


100 realizations of critical component acceleration of computational model

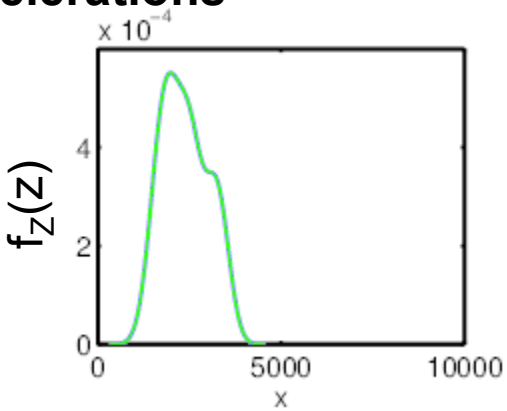
Relevant PDFs

T38

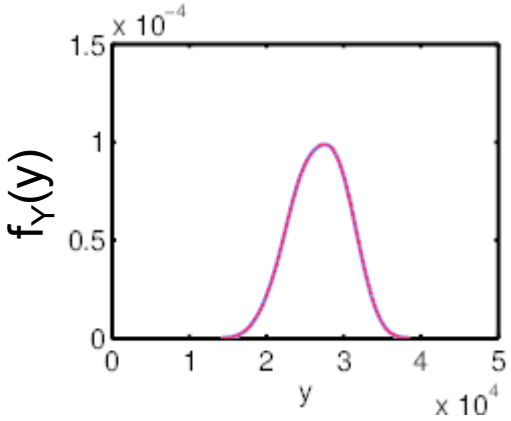
Peak accelerations



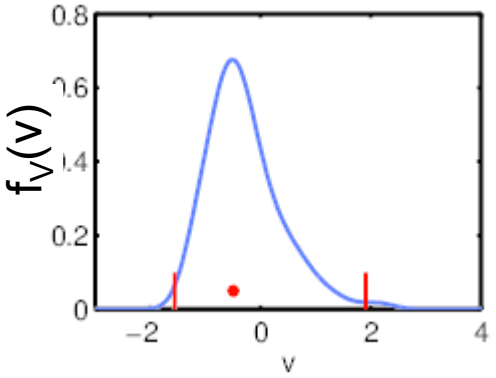
Truth Model



Computational Model



Component strengths



Distribution of v

Slide 26

T38 I would eliminate.
Tom, 5/12/2013

Conclusions

- Definition of Margin, Uncertainty, and Confidence requires some approach which is reasonably independent of assumed distributions.
- One such approach, PEM, is suggested here.
- Validation can be performed in terms of probabilistic measures of system behavior, including PEM.

T28

A fuller discussion can be found in
SAND2013-2823: A Robust Approach to QMU, Validation, and Conservative Prediction. Segalman, Paez & Bauman

Slide 27

T28 I would change to: Validation can be performed in terms of probabilistic measures of system behavior, including PEM."
Tom, 5/12/2013

Backup Slides

Sandia National Laboratories