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.

SAND2015-1989C

Tutorial on Piping Probabilistic Fracture Mechanics – Approaches and Applications

Code Development of PFM software

ASME 2015 Pressure Vessels & Piping Division Conference
July 14-18, 2015, Boston, Massachusetts

What's a probabilistic complex analysis

- Every probabilistic complex analysis starts with the same 4 questions:

Q1: What can happen?

Q2: How likely is it to happen?

Q3: What are the consequences if it does happen?

Q4: How much confidence do you have in the answers to the first three questions?

Given that: We don't know exactly how a component behaves and we don't know exactly what the environment (normal, abnormal, hostile) is;

→

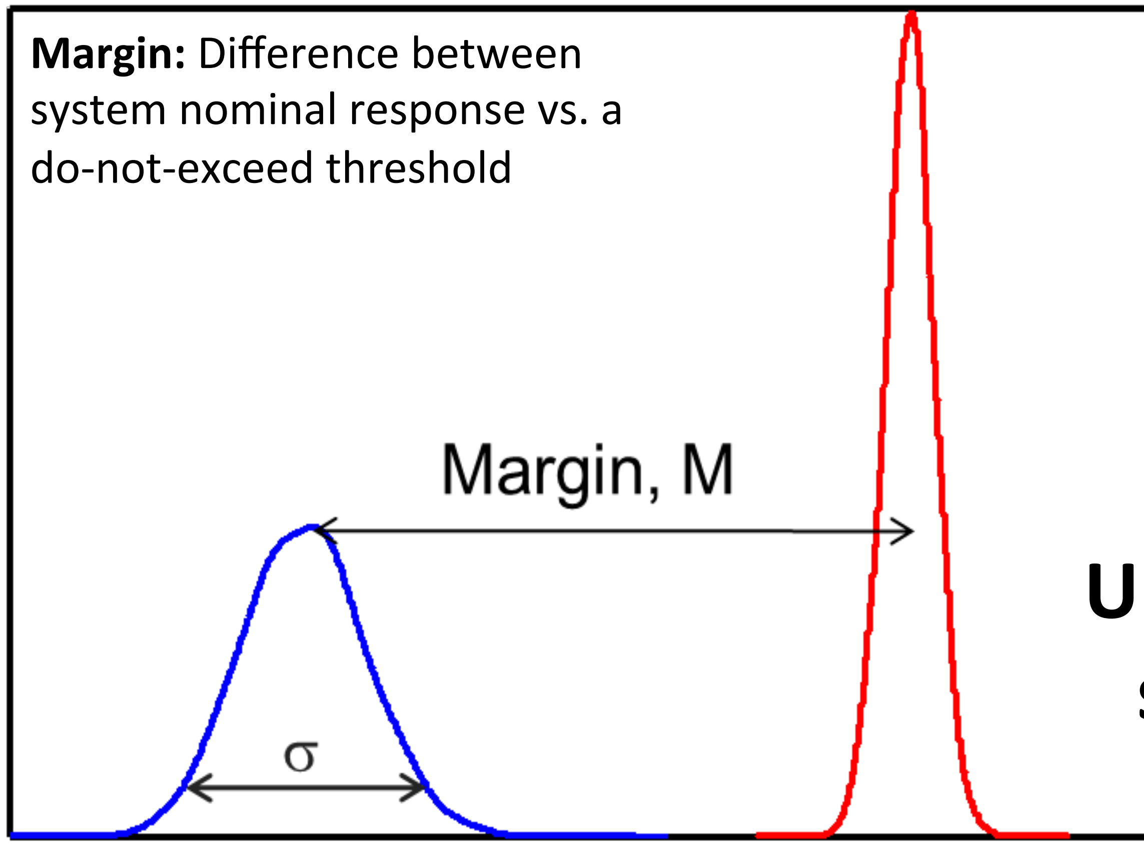
Can we determine if: The component will survive and by how much

How do we quantify uncertainty?

- ❑ **Through data from experiments on multiple pieces of hardware**
  - ➢ Pros: Best way to quantify unit-to-unit variability
  - ➢ Cons: $$
- ❑ **Through historical data from legacy system**
  - ➢ Warning: Knowledge from the past might not be relevant to the future
- ❑ **Through the use of models representing the system behavior (PFM software)**
  - ➢ Pros: In principle, model can be run many times in a stochastic way to quantify uncertainty
  - ➢ Cons: Not always an accurate representation of the real system behavior
- ❑ **A combination of both experiments and models**

Quantification of margins and uncertainty

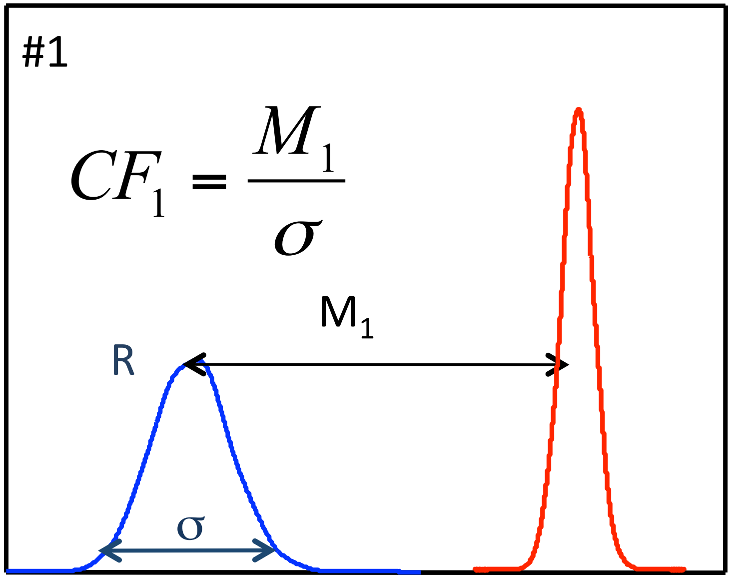
- ❑ **Quantifies the performance threshold and associated margins for the system made under conditions of uncertainty**

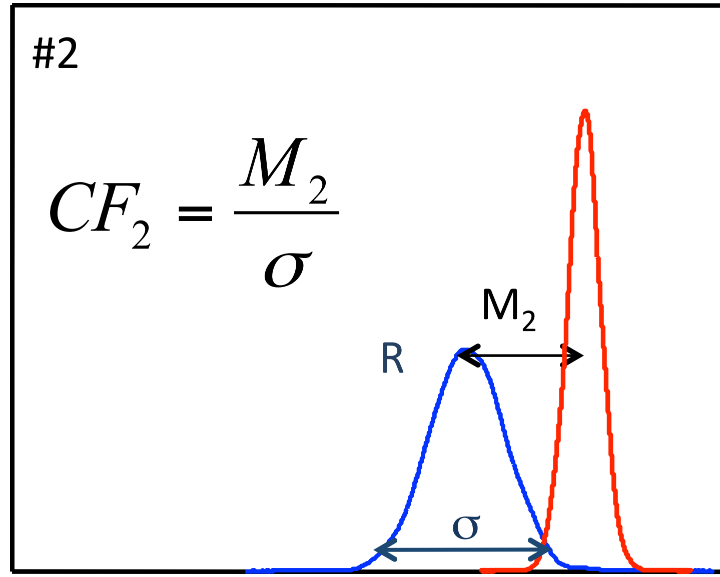


Ultimate stress

Applied stress

Quantification of margins and uncertainty





- $CF_1 > CF_2$
- For Case #1, due to large margin, a large uncertainty in the model response can be tolerated.

- Conversely, for Case #2, due to the small margin, the uncertainty estimated by the model becomes very important and thus needs to be reduced as much as possible.

- How do we assure that the accuracy and the amount of uncertainty in Response R is adequate given these different scenarios?
- Can above question be related to validation?

Why do we perform a probabilistic analysis?

- **To better understand performance margins and uncertainties**

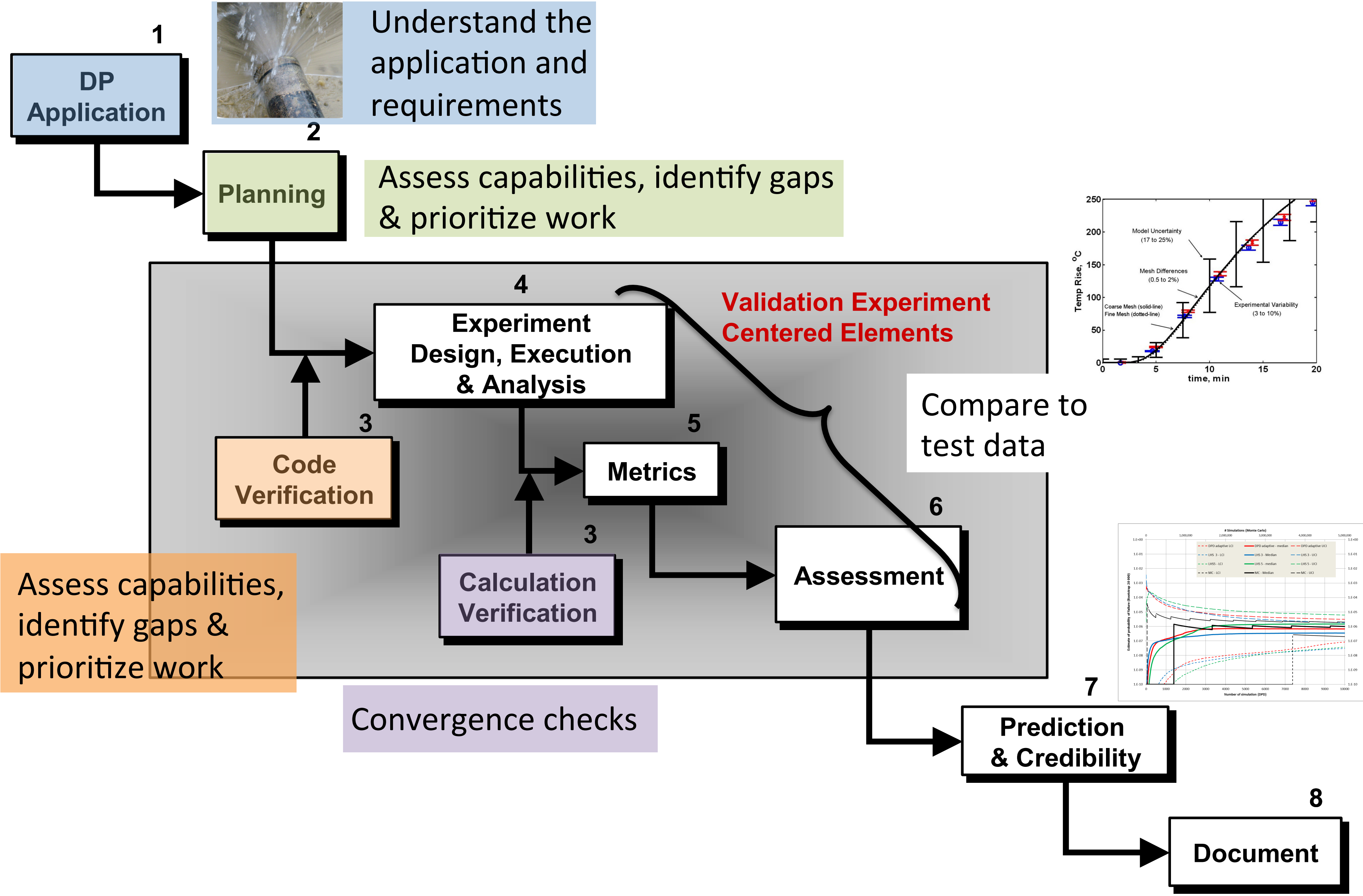
- **Most applications have statistics-based performance requirements:**
  - Probability of an undesirable event happening < 10^{-m}

- **To provide a consistent set of criteria on systems so that resources can be focused where needed most**

- **Qualification support**
  - Level of confidence in design
  - Body of evidence that the system meets its design requirements

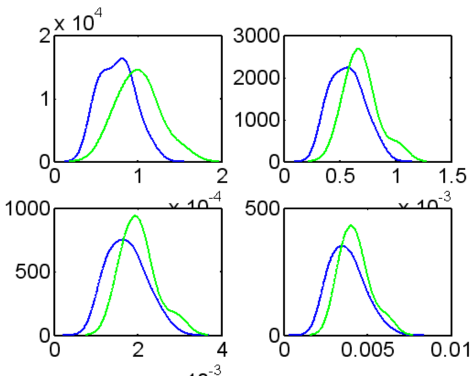
PFM code architecture

PFM software development process

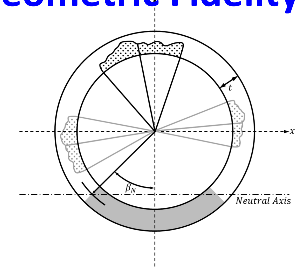


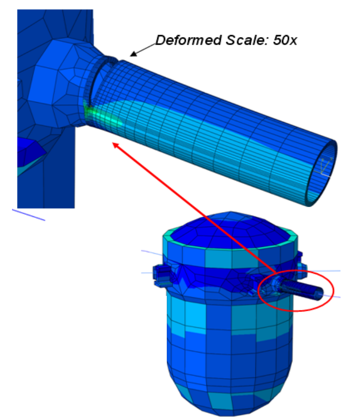
PFM software development process

Model Validation

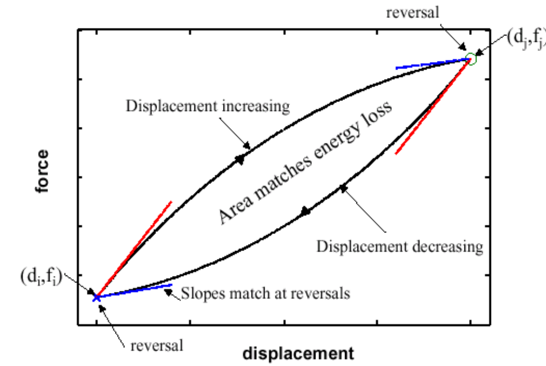


Representation and Geometric Fidelity

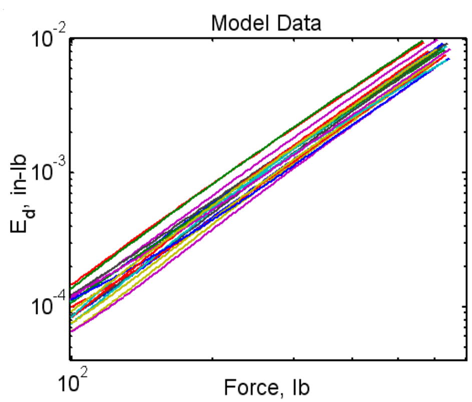




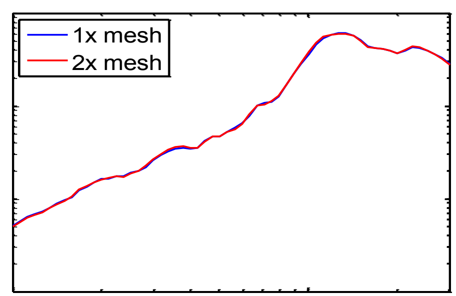
Physics and Material Model Fidelity



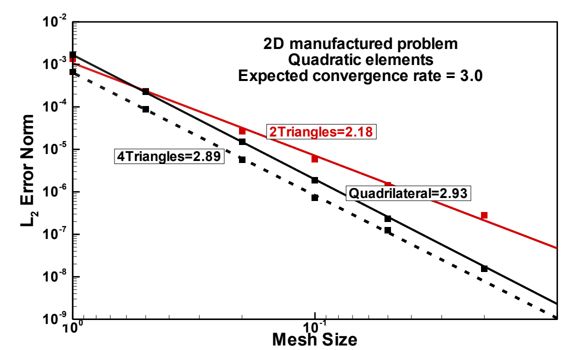
Uncertainty Quantification



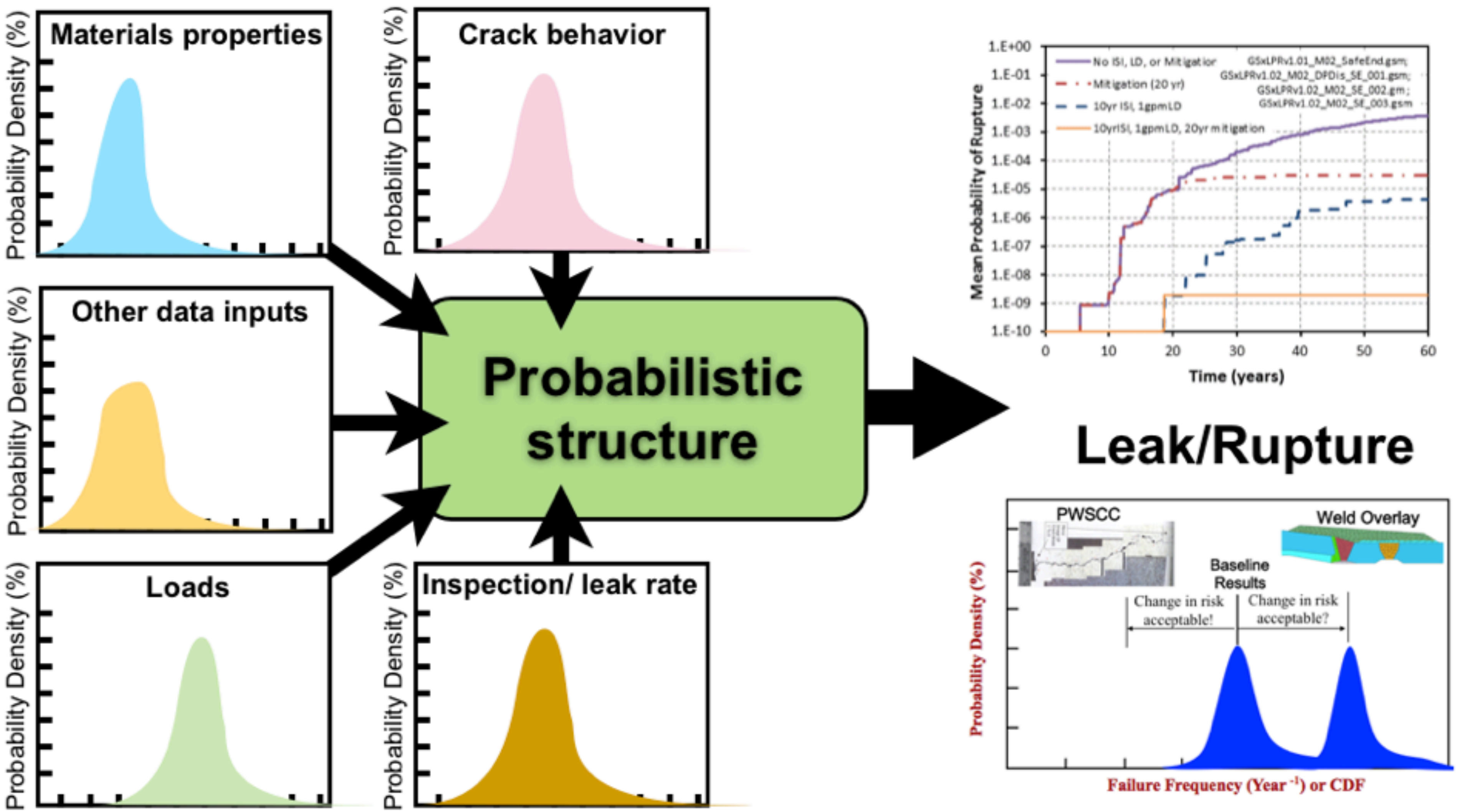
Solution Verification



Code Verification



PFM code architecture



Example of PFM code architecture

❑ Probabilistic simulation engine:

- ➢ Integrate various modules composing the overall model under one umbrella.

❑ Input interface structure:

- ➢ Interface between user and global structure.
- ➢ Uncertainty distribution associated with each input .

❑ Deterministic model:

- ➢ Linking sub-models to the probabilistic simulation engine.

❑ Sampling structure:

- ➢ Defines the number and order of realization and appropriate values to be sampled.
- ➢ Defines different sampling schemes available.

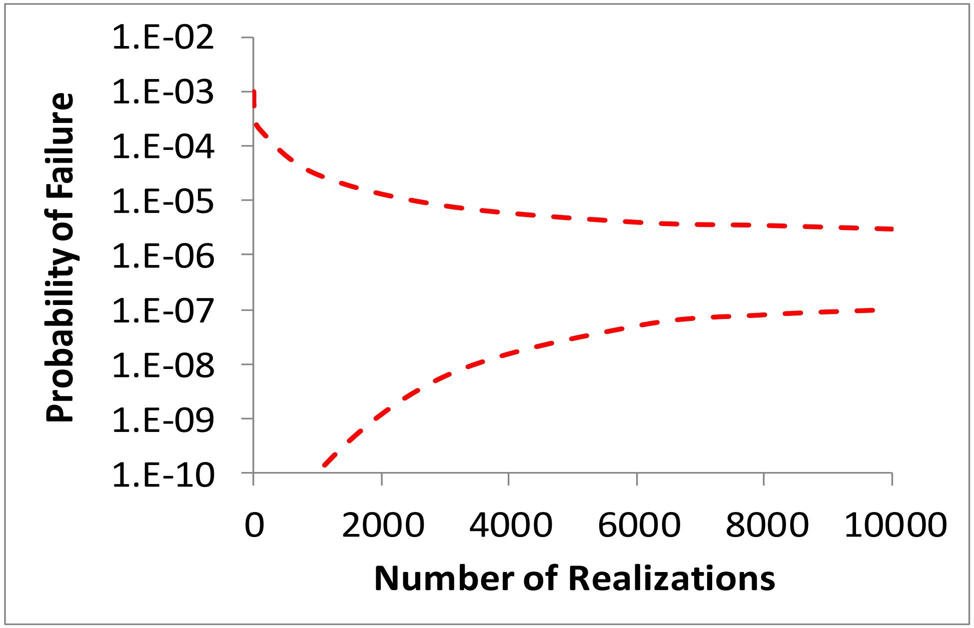
❑ Landing platform structure:

- ➢ Tie up the interface, sampling structure and deterministic model.
- ➢ List all inputs and user selected options required by the model.

Categorization and propagation of uncertainties

Objectives of uncertainty characterization in PFM

- Capture uncertainty in model predictions
- Reduce uncertainty in predicted pipe failure frequency
- Determine how likely certain outcomes are if some aspects of the system are not exactly known
- Uncertainty propagation: “Mapping” uncertainty from inputs to outputs



Difference between epistemic and aleatory uncertainty and spatial variability

- **Aleatory uncertainty:** (Perceived) randomness in the occurrence of future events.

- **Epistemic uncertainty:** Lack of knowledge w.r.t. the appropriate value to use for a quantity that has a fixed, but poorly known, value in the context of a specific analysis.

- Treat questionable uncertainties as epistemic, and then determines the ones that dominate the epistemic output uncertainty. Only for those that are dominant, additional evaluation becomes necessary to justify their treatment as epistemic. All other uncertainties can then be allocated to the aleatory category.

- **Spatial variability:** inherent variability over space of a quantity, that usually cannot be measured precisely or at the expected scale. Spatial variability is **NOT** aleatory or epistemic uncertainty. Variability is linked to uncertainty.

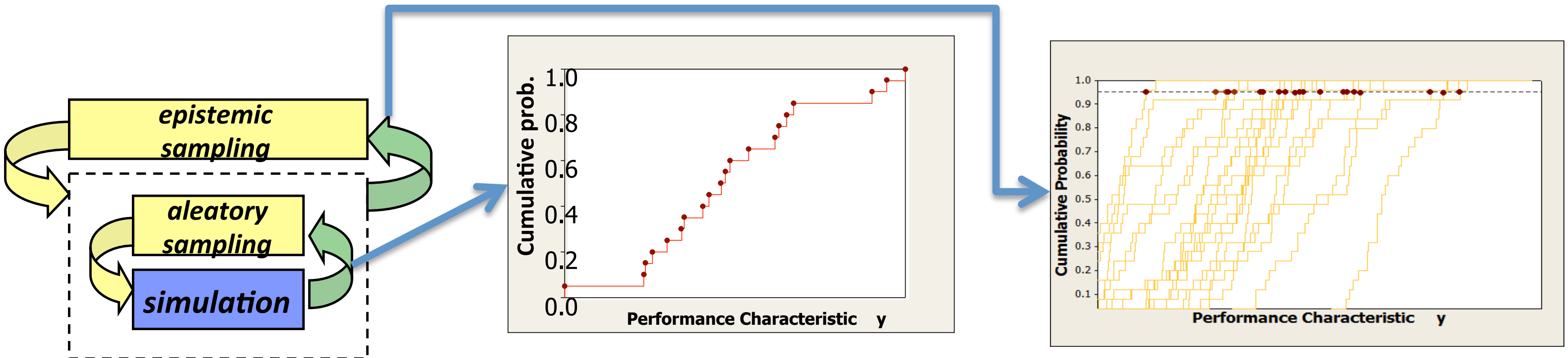
- Probability usually used to characterize both aleatory and epistemic uncertainties and spatial variability.

- Alternatives to probability to the representation of epistemic exist, such as evidence theory, possibility theory, interval analysis and others.

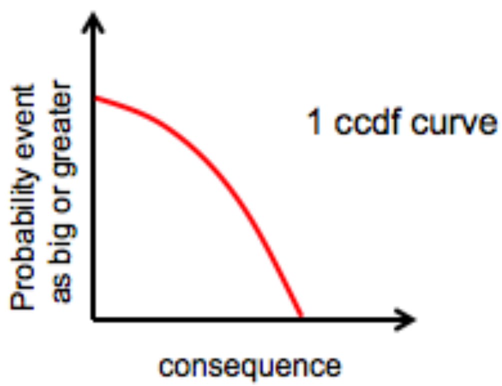
Representation and interpretation of results

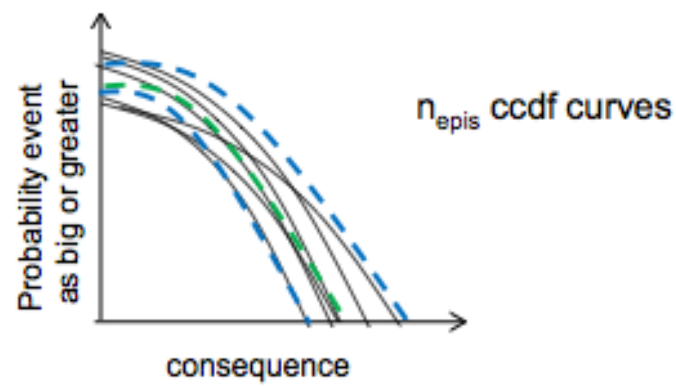
- Parameters selected either as aleatory or epistemic: Guidance needed

- Inner aleatory loop vs. outer epistemic loop



- Interpretation of the results:





Aleatory uncertainty represents the **risk**. Not simply [probability]x[consequence] but probability and consequence

Epistemic uncertainty represent **the level of knowledge** we have with respect to this risk

Characterization of uncertainty

- **Uncertainty (both aleatory and epistemic) is usually characterized using probability distributions**

- **Distribution may depend on the type of uncertainty selected (aleatory or epistemic) which in turn depends on the problem considered (study of one weld vs. collection of welds)**

- **In order to insure that the combination of inputs leads to physically acceptable set, relationships may be required amongst some parameters. When traditional MC used, it is common to represent these dependencies with correlations**

- **Traditional techniques to generate distribution include:**
  - Expert review: used when no data is available.
  - Bayesian updating: used when data becomes available, to update expert elicitation
  - Maximum entropy: used when enough data is available to fit distribution
  - Bootstrap: used when some data is available
  - Other techniques include for example evidence theory, special objective response surface

Epistemic uncertainty characterization can be an iterative process

Possible strategy:

- Perform initial exploratory analysis with "crud: characterization of the distributions characterizing epistemic uncertainty
- Use sensitivity analysis to determine the elements that dominate the uncertainty in analysis outcomes of interest
- Perform detailed uncertainty assessments for the important variables identified in the sensitivity analysis
- Carry out final decision-supporting analysis with new distribution

Desiderata in epistemic uncertainty assessment:

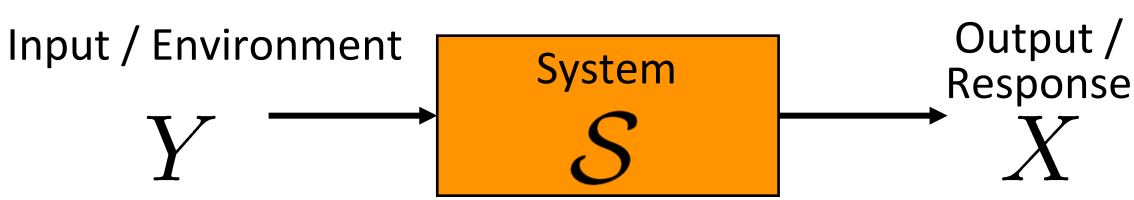
- Avoid being either deliberately optimistic (i.e. non-conservative) or deliberately pessimistic (i.e. conservative) in uncertainty assessments
- Be honest w.r.t. the uncertainty that is present

Source of uncertainty

- **The model structure, i.e., how accurately a mathematical model describes the true system for a real-life situation, may only be known approximately**

- **The numerical approximation, i.e., how appropriately a numerical method is used in approximating the operation of the system**

- **Input and/or model parameters**
  - may only be known approximately.
  - may vary between different instances of the same object for which predictions are sought.

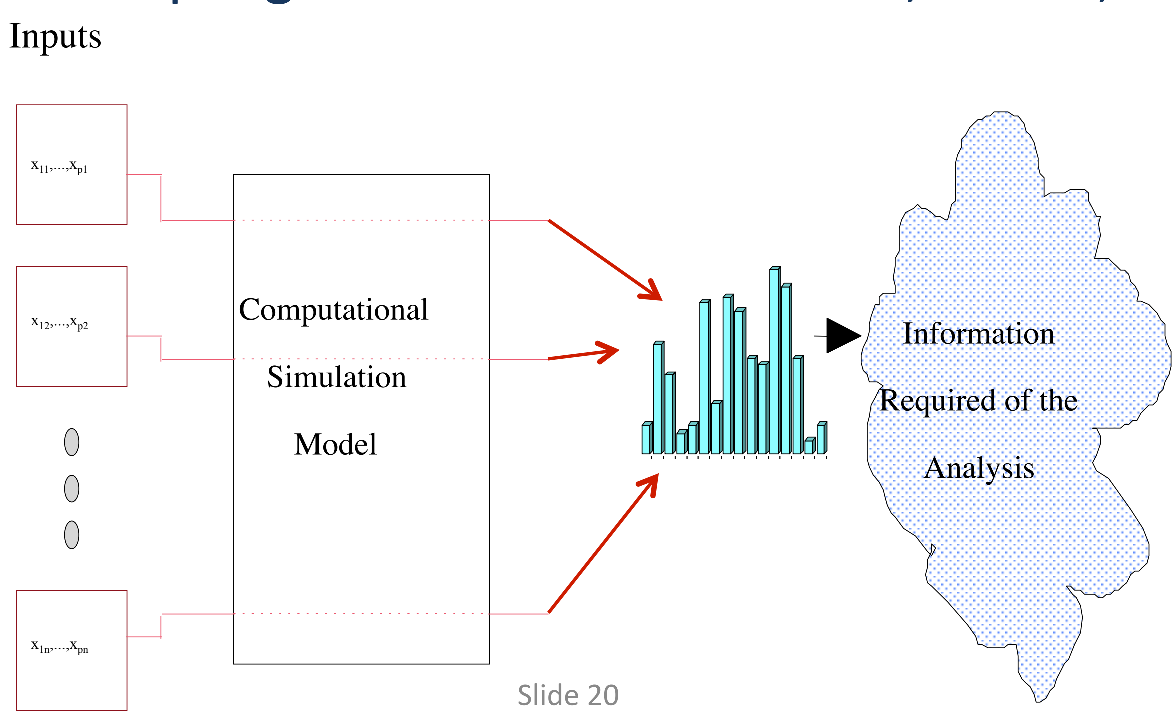


Given representations for uncertainty in *Y* and/or *S*, how do we propagate this information to *X* ?

Advanced algorithms

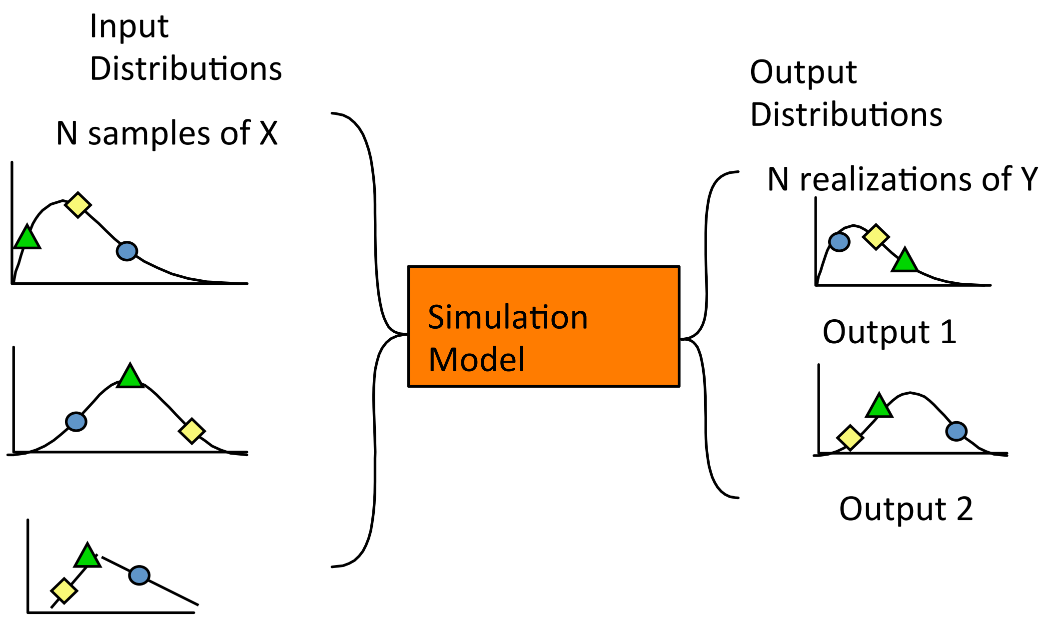
Sampling techniques

- ❑ Random sampling (Monte Carlo sampling)
- ❑ Latin Hypercube Sampling (LHS)
- ❑ Discrete Probability Distribution (DPD)
- ❑ Importance sampling
- ❑ Adaptive sampling
- ❑ Other methods exist (quasi-MC, etc).
- ❑ Alternative to sampling-based methods: FORM, SORM, AMV

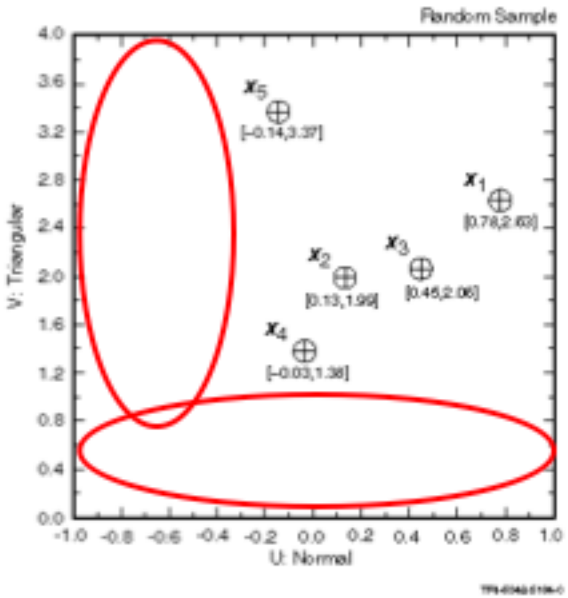


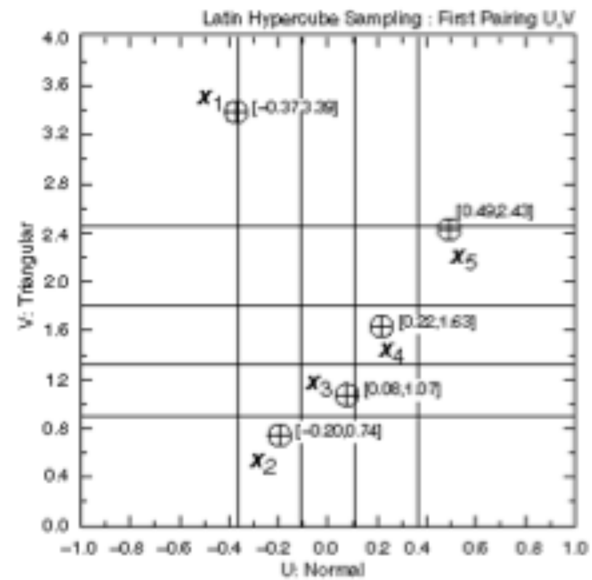
From mathematical characterization to implementation: Sampling approach

1. Characterization of distributions on the uncertain input values
2. Generation of sample from those distributions
3. Propagation of sample through analysis execution repeatedly
4. Presentation of uncertainty analysis results in the form of distributions of the outputs
5. Determination of sensitivity analysis results



LHS vs. Monte Carlo sampling





Random sampling

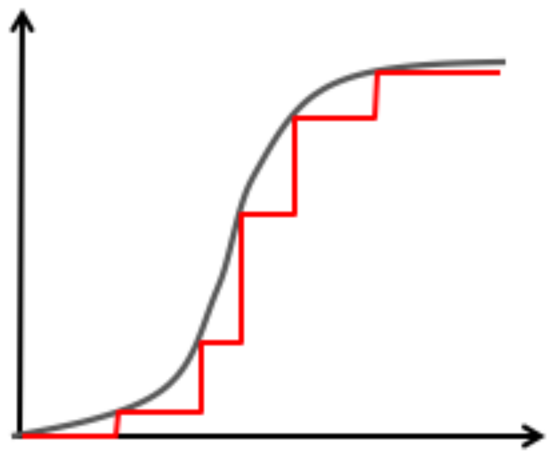
- Preferred when sufficiently large samples are possible
- Easy to implement
- Easy to explain
- Unbiased estimates for means, variances and distribution functions
- Sufficiently large samples may not be possible

LHS

- Unbiased estimates for means and distribution functions
- "Force" samples to be spread out across domain of the input distributions
- Dense stratification across range of each variable
- Used when large samples not computationally practicable and estimation of high quantiles not required
- Uncertainty/sensitivity results robust with relatively small sample sizes (e.g., nLHS = 50 to 200)

Discrete Probability Distribution (DPD)

- **DPD uses discrete values from probability distributions**
  - Each value can be equally probable or of different likelihood

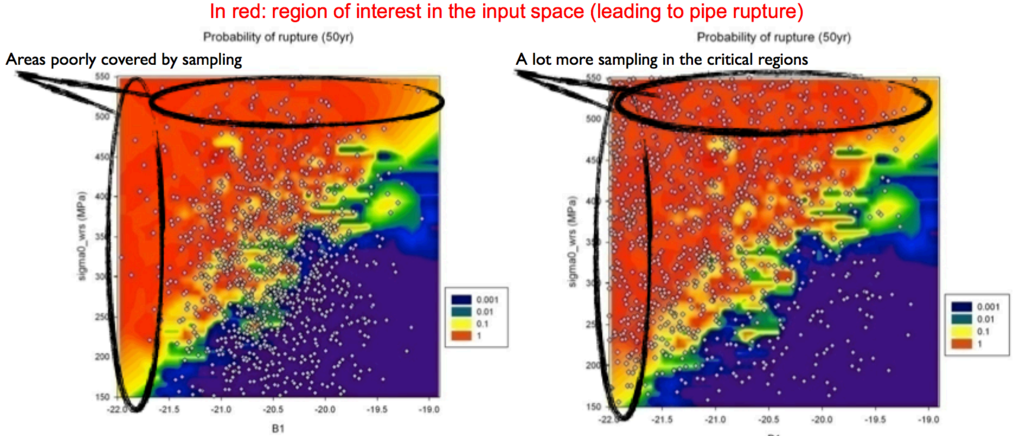


- **Difference with LHS:**
  - Less dense stratification. Worse than LHS if events of interest occur more for extreme values of inputs.
  - Higher combination (i.e. better multidimensional coverage). Better than LHS if events of interest occur more for combination of inputs.

Importance sampling

- **Importance sampling is a variance reduction technique that can be used in the Monte Carlo method**
  - Certain values of the input random variables in a simulation have more impact on the parameter being estimated than others. If these "important" values are emphasized by sampling more frequently, then the estimator variance can be reduced.
  - Cannot be applied to all variables!
- **The basic methodology in importance sampling is to choose a distribution which "encourages" the important values. The outputs are weighted to correct for the use of the biased distribution, and this ensures that the new importance sampling estimator is unbiased.**

In red: region of interest in the input space (leading to pipe rupture)



Regular sampling
More than half of the points sampled outside of the region of interest

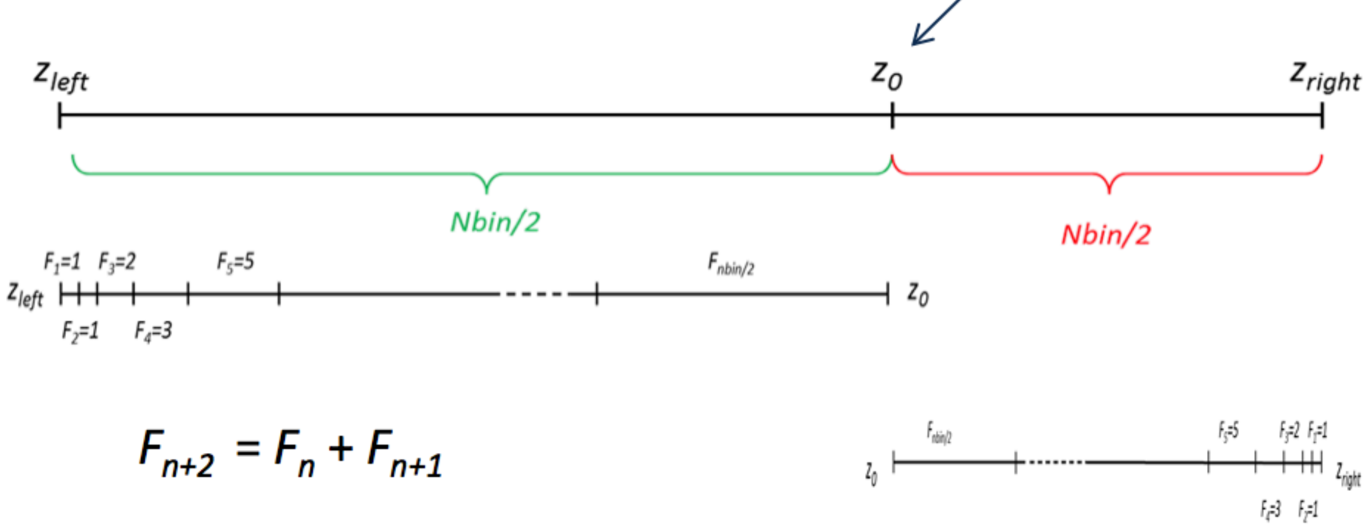
Importance sampling
Most of the points sampled in the region of interest

Example: DPD using importance sampling

- For example, a Fibonacci series works well for interrogating tails of distributions

- The strength of the Fibonacci series can be controlled by an exponent γ between 0 and 1 on the F value

A value z_0 is selected to split
The strata into two groups



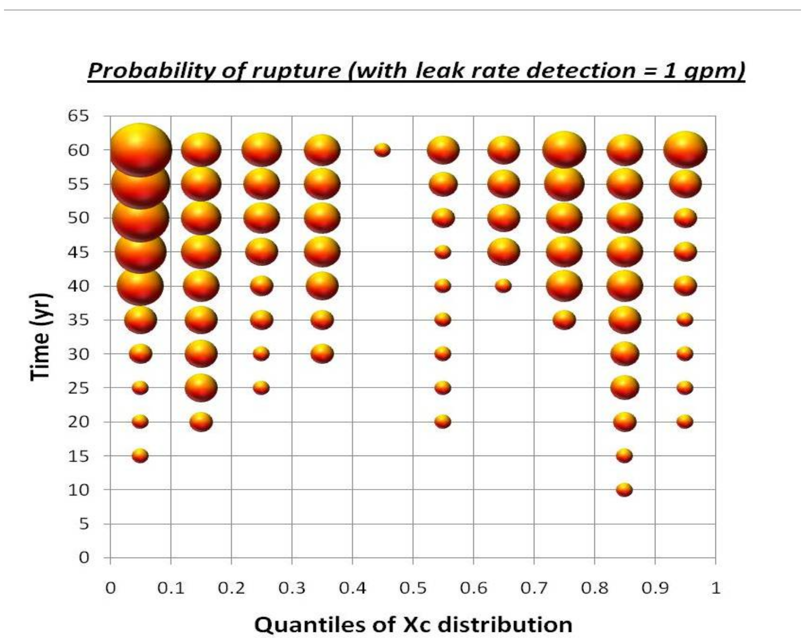
Fibonacci series is used to generate strata of unequal size with focus on the tail

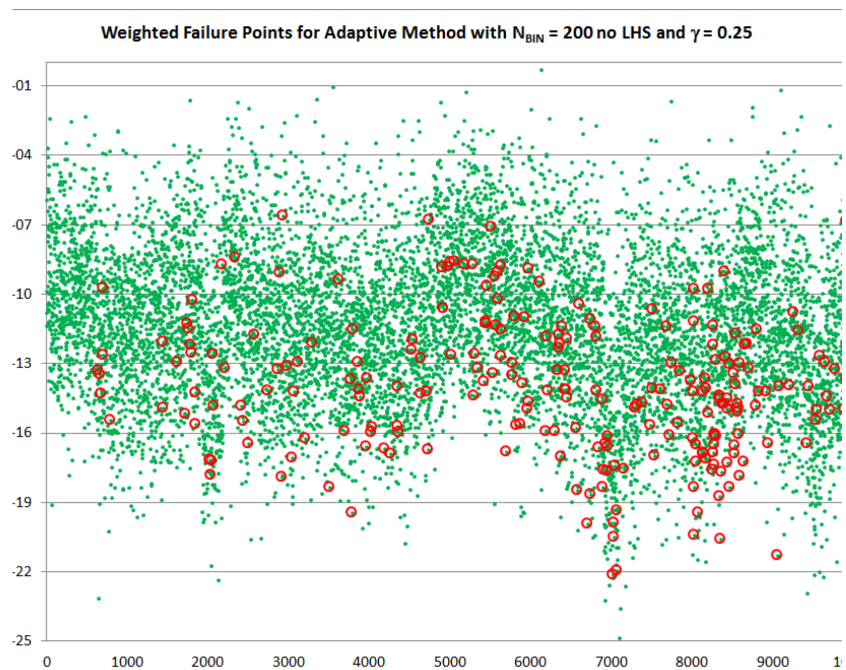
$$F_{n+2} = F_n + F_{n+1}$$

Optimization method: Adaptive sampling

- Adaptive sampling promotes importance sampling by using model results to identify and focus on space of interest

- Adaptive sampling can cover more densely disparate regions in the input space, and reduces the number of samples needed to confidently estimate low probability (~10^{-6})

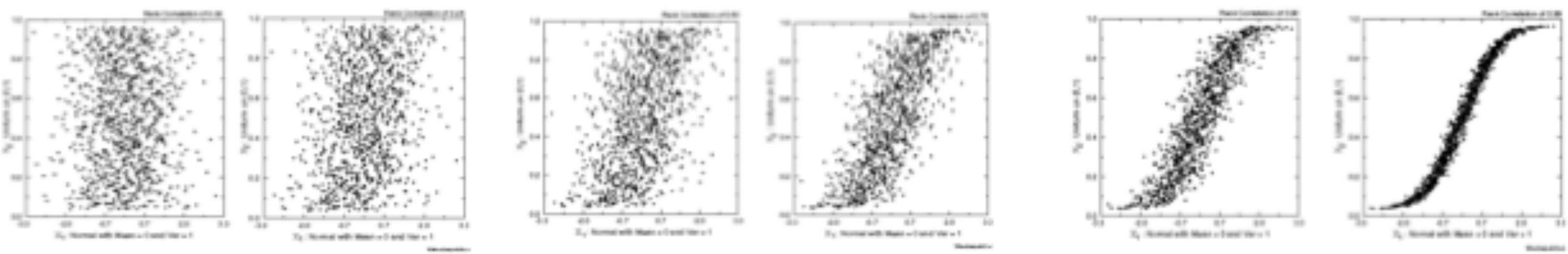




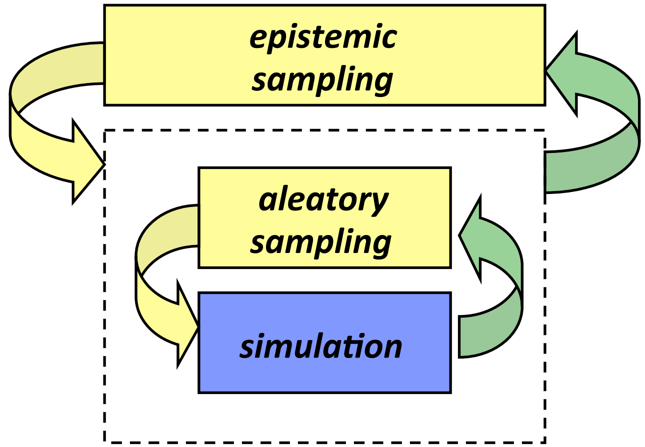
Sampling correlated inputs

- **Correlation can be used to force behavior between two or more variables in order to remain in a physically acceptable input space**
  - Individual inputs are not independent

- **Correlation control should be done for a limited selected variables based on model/input recommendations**



Sampling strategy

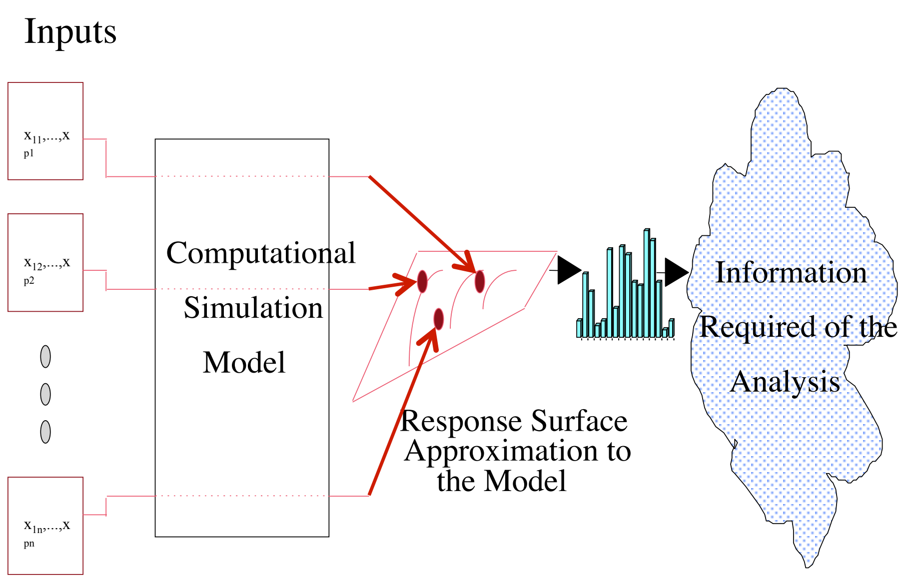


- ❑ **Two loops can be considered (one can be ignored by setting the sampling size to 1). For each loop, on can select from the following options:**
  - ➢ Simple random sampling or Latin Hypercube Sampling (LHS)
  - ➢ DPD
  - ➢ Importance sampling applied to selected values
  - ➢ Use of optimization instead of importance sampling for selected values

- **Example: Possibility of creating 12 sampling combination: [LHS vs. RS]x[DPD vs. no DPD] x[No importance vs. importance vs. adaptive] for each loop (totaling 12^2 combinations)**

Other methods: Response surface methods

- The response surface approach consists of constructing a surrogate to the computational model that is generally simpler in form and cheaper (very cheep)to access
- The response surface is constructed based on the results from a set of simulations. Example of surface response include Gaussian response, Polynomial response
- There is no clear “best” method of construction -- a preferred approach depends on the application, and even then there is rarely a clear-cut "best" response surface methodology



Verification, validation and benchmarking of PFM software

Verification vs. validation

❑ Verification:

- ➢ "Are we solving the equations correctly?"
- ➢ Code verification: Correctness of implemented mathematical algorithms (bug-free?)
- ➢ Solution verification: Convergence to the correct answer at the correct rate, as model is refined.

❑ Validation:

- ➢ "Are we solving the correct equations?"
- ➢ Correctness of physical model and sufficiency for the application.

❑ Uncertainty quantification

- ➢ Statistical propagation of uncertainty through a simulation model, and statistical interpretation of model response.

Code verification:
Are software errors or algorithms deficiencies corrupting simulation results?

- **Apply SQA processes**
  - Do we have a mature code development process?
- **Assess SQA processes**
  - Verify that codes are developed with an appropriate level SQA maturity.
- **Provide adequate test coverage**
  - Can the user be confident that the code is adequately tested for the intended application?
- **Quantify computation errors**
  - What is the impact of undetected code or algorithm deficiencies on simulation results?

Solution verification:
Are procedural errors or numerical solution errors corrupting simulation results?

- **Quantify numerical solution errors**
  - What is the impact of numerical solution accuracy on the system response quantities?

- **Verify all simulation inputs and outputs**
  - Have corrupted simulation results with incorrect inputs or post-processing errors?

- **Perform technical review**
  - Verify that the solution verification activities are relevant, adequate and executed in technically sound manner?

Uncertainty quantification

- **Uncertainty quantification:**
  - What is the impact of Monte Carlo accuracy on the system response quantities?

- **Source of uncertainty:**
  - The model structure, i.e., how accurately a mathematical model describes the true system for a real-life situation, may only be known approximately?
  - The numerical approximation, i.e., how appropriately a numerical method is used in approximating the operation of the systems
  - Input and/or model parameters

- **Perform technical review**
  - Verify that the solution verification activities are relevant, adequate and executed in technically sound manner?

PFM software validation

❑ Model validation

- ➢ Field data vs. lab data
- ➢ Validation criteria to assess agreement between data and model prediction
  - Quantitative metrics
  - Criteria should be well-defined and reproducible
- ➢ Validation with alternative software
- ➢ Validation with engineering judgment (lesser prominence)

❑ Range of applicability

- ➢ Is the validation contingent on some fixed assumption for model parameters?
- ➢ Only applies to specific cases?
- ➢ Engineering judgment may be use to extend range of applicability to input ranges that were not explicitly validated with data.

Summary

- **In this module, discussion was provided on the representation of uncertainty in PFM**
  - Parameters may be represented by numerous types of distributions
  - Separation of aleatory vs. epistemic

- **Many sampling techniques are available to investigate the performance requirement of interest**
  - Sampling strategy developed to reduce the number of realization needed to quantify performance requirement:
    - RS vs. LHS vs. DPD vs. importance sampling vs. adaptive sampling

- **V&V effort crucial to achieve desired confidence in risk-informed decision making**