U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

NUREG/CR-7278

GUIDANCE ON PERFORMING AND DOCUMENTING PROBABILISTIC FRACTURE MECHANICS ANALYSES

[Company Address]

AVAILABILITY OF REFERENCE MATERIALS IN NRC PUBLICATIONS

NRC Reference Material

As of November 1999, you may electronically access NUREG-series publications and other NRC records at the NRC's Public Electronic Reading Room at http://www.nrc.gov/reading-rm.html. Publicly released records include, to name a few, NUREG-series publications; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigative reports; licensee event reports; and Commission papers and their attachments.

NRC publications in the NUREG series, NRC regulations, and Title 10, "Energy," in the *Code of Federal Regulations* may also be purchased from one of these two sources.

1. **The Superintendent of Documents**
   U.S. Government Publishing Office
   Mail Stop SSOP
   Washington, DC 20402-0001
   Internet: http://bookstore.gpo.gov
   Telephone: 1-866-512-1800
   Fax: (202) 512-2104

2. **The National Technical Information Service**
   5301 Shawnee Road
   Alexandria, VA 22161-0002
   http://www.ntis.gov
   1-800-553-6847 or, locally, (703) 605-6000

A single copy of each NRC draft report for comment is available free, to the extent of supply, upon written request as follows:

U.S. Nuclear Regulatory Commission
Office of Administration
Publications Branch
Washington, DC 20555-0001
E-mail: distribution.resource@nrc.gov
Facsimile: (301) 415-2289

Some publications in the NUREG series that are posted at the NRC's Web site address http://www.nrc.gov/reading-rm/doc-collections/nuregs are updated periodically and may differ from the last printed version. Although references to material found on a Web site bear the date the material was accessed, the material available on the date cited may subsequently be removed from the site.

Non-NRC Reference Material

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, transactions, *Federal Register* notices, Federal and State legislation, and congressional reports. Such documents as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings may be purchased from their sponsoring organization.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at—

The NRC Technical Library
Two White Flint North
11545 Rockville Pike
Rockville, MD 20852-2738

These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—

American National Standards Institute
11 West 42nd Street
New York, NY 10036-8002
http://www.ansi.org
(212) 642-4900

Legally binding regulatory requirements are stated only in laws; NRC regulations; licenses, including technical specifications; or orders, not in NUREG-series publications. The views expressed in contractor-prepared publications in this series are not necessarily those of the NRC.

The NUREG series comprises (1) technical and administrative reports and books prepared by the staff (NUREG-XXXX) or agency contractors (NUREG/CR-XXXX), (2) proceedings of conferences (NUREG/CP-XXXX), (3) reports resulting from international agreements (NUREG/IA-XXXX), (4) brochures (NUREG/BR-XXXX), and (5) compilations of legal decisions and orders of the Commission and Atomic and Safety Licensing Boards and of Directors' decisions under Section 2.206 of NRC's regulations (NUREG-0750).

DISCLAIMER: This report was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any employee, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product, or process disclosed in this publication, or represents that its use by such third party would not infringe privately owned rights.

GUIDANCE ON PERFORMING AND DOCUMENTING PROBABILISTIC FRACTURE MECHANICS ANALYSES

[Comments]

Manuscript Completed: January 2021
Date Published: TBD

Prepared by:
L. Hund
J. Lewis
N. Martin
M. Starr
D. Brooks
A. Zhang
R. Dingreville
A. Eckert
J. Mullins

Sandia National Laboratories
P.O. Box 8500
Albuquerque, NM 87185

P. Raynaud
D. Rudland
D. Dijamco
S. Cumblidge

U.S. Nuclear Regulatory Commission

Patrick Raynaud, NRC Project Manager

COMMENTS ON DRAFT REPORT

Any interested party may submit comments on this report for consideration by the NRC staff. Comments may be accompanied by additional relevant information or supporting data. Please specify the report number **NUREG/CR-7278** in your comments, and send them by the end of the comment period specified in the *Federal Register* notice announcing the availability of this report.

Addresses: You may submit comments by any one of the following methods. Please include Docket ID **NRC-2021-** in the subject line of your comments. Comments submitted in writing or in electronic form will be posted on the NRC website and on the Federal rulemaking website http://www.regulations.gov.

Federal Rulemaking Website: Go to http://www.regulations.gov and search for documents filed under Docket ID **NRC-xxxx-xxxx**.

Mail comments to: Office of Administration, Mail Stop: TWFN-7-A60M, U.S. Nuclear Regulatory Commission, Washington, DC 20555-0001, ATTN: Program Management, Announcements and Editing Staff.

For any questions about the material in this report, please contact: Patrick Raynaud, Senior Materials Engineer, (301) 415-1987 or by e-mail at Patrick.Raynaud@nrc.gov

Please be aware that any comments that you submit to the NRC will be considered a public record and entered into the Agencywide Documents Access and Management System (ADAMS). Do not provide information you would not want to be publicly available.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

ABSTRACT

This NUREG, on probabilistic fracture mechanics (PFM), is a companion document to Draft Regulatory Guide 1382 (DG-1382) / Regulatory Guide 1.245 (RG-1.245), “Preparing Probabilistic Fracture Mechanics (PFM) Submittals.” This document provides guidance on a graded approach to developing PFM submittal documentation and a generalized technical basis for conducting PFM analyses.

The guidance provided for PFM submittal documentation represents a balance between the efficiencies gained by clear, consistent, and comprehensive submittals and the need to maintain flexibility for PFM analyses that by their nature will include many situation-specific aspects. The resulting guidance outlines a procedure that includes a suggested graded approach for PFM analyses and submittals. The unique characteristics of the underlying regulatory application dictate the breadth and depth of content included in the submission.

This document also describes a hypothetical process for conducting a PFM analysis. This process is aligned with the position on documentation elements given previously in the U.S. Nuclear Regulatory Commission’s (NRC’s) technical letter report, “Important Aspects of Probabilistic Fracture Mechanics Analyses,” issued in 2018. The NUREG provides fundamental background for the concepts and methods introduced in the analysis process. Its examples give details for analysts on (nonprescriptive) approaches for PFM analyses. It provides general guidance on PFM analysis submittals. However, specific applications or submittals may deviate from this guidance to address specific features, with acceptable justifications for the deviations from these guidelines. PFM submittals that explicitly identify deviations from these frameworks will assist the NRC staff in efficient reviews of those submittals.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

This technical basis NUREG document, and the regulatory guide it is associated with, were developed from the guidance introduced in the technical letter report, “Important Aspects of Probabilistic Fracture Mechanics Analyses,” issued in 2018. In conjunction with the release of the technical letter report, the U.S. Nuclear Regulatory Commission held a series of public meetings to present a general framework of the expected content of a probabilistic fracture mechanics (PFM) analysis. This report further develops the concept of a PFM analysis methodology and outlines important considerations for a high-quality and high-confidence PFM analysis. Realizing that PFM information and results only make up a portion of the information needed to make risk-informed decisions, and guided by the agency’s desire to ensure that submittals that include PFM information (hence called PFM submittals) are sufficiently clear and complete, this NUREG explicitly describes the minimum expected documentation.

This NUREG contains three technical sections: Section 2 presents the contents of a PFM submittal following a graded approach, Section 3 presents the analytical steps in a PFM submittal, and Section 4 presents the methods used in PFM analysis. These three sections are linked together through the development structure, but the guidance provided in each section is geared toward different audiences. The first technical section is intended for applicants of all experience levels. The second technical section could be used by applicants who are familiar with PFM submittals but are seeking some guidance on the development of an analysis structure or formalism. The third technical section could be used by applicants who are seeking explicit guidance on the theoretical underpinnings of the processes that are used to establish the credibility of a PFM analysis.

The guidance provided for PFM submittal documentation represents a balance between the efficiencies gained by clear, consistent, and comprehensive submittals and the need to maintain flexibility for PFM analyses that by their nature will include many situation-specific aspects. The resulting guidance outlines a procedure in which a suggested minimum set of documented evidence may be augmented by additional details. The unique characteristics of the underlying regulatory application dictate the breadth and depth of content included in the submission. Expected documentation elements are explicitly linked to the analysis framework that is described.

This report presents a general framework to describe, perform, and evaluate a PFM analysis. The important pieces of a PFM analysis that should be considered include models, inputs, uncertainty characterization, probabilistic framework, and PFM outputs:

- Models can be categorized into different types, but in all cases, model verification, validation, and uncertainty quantification are key steps to gain confidence in the adequacy of the models used.

- Treatment of random inputs may consist of constructing probability distributions; determining input bounds if applicable; and quantifying any assumptions, conservatisms, or dependencies among inputs.

- Uncertainty characterization and treatment are at the core of a PFM analysis. In many PFM analyses, separation of epistemic and aleatory uncertainty may be useful. Uncertainty identification, quantification, and propagation are essential elements in describing a PFM methodology or analysis. The proper choice of sampling techniques is also an important step that needs justification. The report discusses concepts and methods to verify and validate a probabilistic framework.

- Ways to demonstrate PFM convergence include varying sample size and sampling strategy, as well as performing stability analysis. Output uncertainty analysis can take various forms depending on the problem being analyzed. Sensitivity analyses can help to identify the drivers of uncertainty for a given problem or output. Sensitivity studies are useful to understand which parameters drive the issue being investigated, and to show that some expected trends are indeed reflected in the analysis results. The report presents methods to perform such studies.

ABBREVIATIONS AND ACRONYMS

AIC	Akaike Information Criterion
ANS	American Nuclear Society
ASME	American Society of Mechanical Engineers
BIC	Bayesian Information Criterion CDF cumulative distribution function
CFR	*Code of Federal Regulations*
CV	coefficient of variation
EPRI	Electric Power Research Institute
FAVOR	Fracture Analysis of Vessels—Oak Ridge
FORM	first-order reliability method
GP	Gaussian process
IAEA	International Atomic Energy Agency
LHS	Latin hypercube sample/sampling
MARS	multivariate adaptive regression splines
ML	machine learning
MPP	most probable point
MRP	Materials Reliability Program
NRC	U.S. Nuclear Regulatory Commission
NUREG	NRC technical report designation
PDF	probability density function
PFM	probabilistic fracture mechanics
PRA	probabilistic risk assessment
QA	quality assurance
QoI	quantity of interest
SA	sensitivity analysis
SORM	second-order reliability method
SQA	software quality assurance
SRS	simple random sampling
V&V	verification and validation
xLPR	extremely low probability of rupture

1 INTRODUCTION

The purpose of this NUREG is to provide a generalized technical basis for conducting probabilistic fracture mechanics (PFM) analyses and to describe a graded approach for developing submittal documentation. PFM is a subset of fracture mechanics that complements deterministic fracture analysis. Specifically, PFM is based on a deterministic fracture mechanics framework that quantifies crack propagation or damage accumulation while accounting for uncertainty in aspects such as the physical models, physical parameters, geometry, loading, deformation mechanisms, and environmental exposure. Analysis of a PFM framework allows for assessments of the structural integrity of components to enable risk-informed decisions in a regulatory application. PFM allows the direct representation of uncertainties using best estimate models and distributed inputs.

1.1 Fracture Mechanics Approach to Structural Integrity Analysis

Any fracture mechanics approach (deterministic or probabilistic) to structural integrity analysis quantifies the combination of at least three key elements: (1) the applied stress produced by structural loading, (2) the flaw size, and (3) the fracture toughness. The stress and flaw size provide the driving force for fracture, while the fracture toughness provides a measure of the material's resistance to crack propagation and failure. Techniques for computing fracture driving force range from simple to complex, and the most appropriate methodology depends on the geometry, loading, and materials properties. Flaw size may be determined by nondestructive evaluation of an indication found to exist in the structure. It may represent the size of a flaw that nondestructive evaluation could miss, or it may represent a nominal flaw size agreed to as appropriate for certain types of assessments. The driving force and fracture toughness are compared to assess the likelihood of failure. Environment and time generally complete the list of other elements included in most fracture mechanics analyses. All these variables may or may not evolve with time and spatial location within a component or structure. Fracture mechanics provides mathematical relationships among these quantities.

There are two general options for performing a fracture analysis (although they can be equivalent in certain circumstances)—the energy criterion approach and the stress-intensity factor approach:

- In the energy balance approach, a fracture mechanics-based failure criterion is considered when the strain energy release rate associated with crack advance matches or exceeds the energy needed to create new crack surfaces, to account for plastic flow, and to account for other types of energy dissipation associated with the degradation mechanisms considered. In this interpretation of fracture mechanics, the crack will grow when the critical energy release rate is exceeded.

- In the stress-intensity factor approach, a fracture mechanics-based failure criterion considers that the material fails locally at some critical combination of stress and strain for given crack-tip conditions. In the case of a linear elastic body, the classic stress-intensity factor is used; in the case of a nonlinear body (or equivalently an elastic-plastic body under monotonic loading), the J-integral is used.

For both the energy balance and stress-intensity factor approaches, the applied load is typically determined either through a finite-element analysis of the actual structure or by a closed-form analysis of a simplified representation of the structure. In the case of linear elastic fracture mechanics, one considers materials under quasistatic conditions, while elasto-plastic fracture mechanics involve consideration of plastic deformation under quasistatic conditions. Dynamic, viscoelastic, and viscoplastic fracture mechanics include time as a variable.

1.2 Historical Perspective on Probabilistic Fracture Mechanics Analysis of Nuclear Structures

Historically, most assessments of structural integrity have been performed deterministically; for example, a single value of fracture toughness is used to estimate the failure stress or critical flaw size. This is true for many U.S. Nuclear Regulatory Commission (NRC) regulations. In the past, the NRC has typically regulated the use of nuclear reactor structural materials on a deterministic basis. Consensus codes and standards used for the design and analysis of such structures, such as the American Society of Mechanical Engineers (ASME) Boiler and Pressure Vessel Code, typically rely on conservative fracture models with applied safety factors and conservative bounding inputs to account for the numerous uncertainties that may be present. Improving the reliability of such models by quantifying the impacts of the assumptions and uncertainties becomes difficult because of the conservative nature of the models and inputs and the lack of historical documentation of the basis for safety factors.

Observations of the character of the three key fracture mechanics elements introduced previously show that (1) loads exerted on a structure may include random noise, (2) structures contain many flaws with various sizes, orientations, and locations, and (3) fracture toughness data in the ductile-brittle transition region are widely scattered. As such, the reliance on a deterministic basis for engineering designs and regulations has given way to increased use of probabilistic techniques. Many factors support and motivate this evolution:

- **NRC policy decision.** In the mid-1990s, the NRC issued a policy statement (Reference 1-1) that encouraged the use of probabilistic risk assessments (PRAs) to improve safety decisionmaking and improve regulatory efficiency. This policy statement formalized the Commission’s commitment to the expanded use of PRA, stating in part that “the use of PRA technology should be increased in all regulatory matters to the extent supported by the state-of-the-art in PRA methods and data and in a manner that complements the NRC’s deterministic approach and supports the NRC’s traditional defense-in-depth philosophy.” Since that time, the NRC has made progress in its efforts to implement risk-informed and performance-based approaches into its regulation and continues to revisit and update the approaches on a regular basis. Two notable efforts in PFM include the FAVOR (Fracture Analysis of Vessels—Oak Ridge) (References 1-2, 1-3) and xLPR (extremely low probability of rupture) projects (Reference 1-4).

- **Factors unanticipated in the design phase or not addressed by codes and standards.** There is a fundamental difference between how deficiencies, or potential deficiencies, are addressed when they are discovered during the design and construction of a structure versus when they are revealed later, often after many years or decades of safe service. During design and construction, deficiencies that do not meet specifications are often addressed by repair, replacement, or reconstruction, because the effort to demonstrate the acceptability of the deficiency often exceeds the effort associated with correcting the deficiency. However, once operation begins, repairs that were considered feasible during construction can become cost prohibitive (“cost” in terms of dollars, time, or dose). While the NRC’s primary mission is safety, it is obligated (see Title 10 of the *Code of Federal Regulations* (10 CFR) 50.109(c)(5) and (7) (Reference 1-5)) to assess whether safety benefits justify the attendant cost. PFM assessments are ideally suited to such situations because PFM metrics relate directly and clearly to systems that can challenge safety (i.e., probability of structural failure). Indeed, the Backfit Rule (see 10 CFR 50.109(c)(3) (Reference 1-5)) explicitly requires an assessment of risk. PFM also provides more flexible methods to account for factors that occur during service (e.g., new damage mechanisms, unanticipated loadings, aging) that were not considered during design. Especially when such factors are encountered for the first time, the performance of deterministic analyses following the guidelines of codes, standards, and regulations can be difficult because these established procedures may not account for the new factors. Historically, unanticipated material degradation mechanisms have regularly arisen in nuclear power plants (Reference 1-6). Examples include the primary water stress-corrosion cracking aging issue in Alloy 600 and 182/82 welds in pressurized-water reactors (which led in part to the development of the xLPR code), cold head cracking, and control rod drive mechanism thermal sleeve wear.

- **Need to understand conservatisms.** One of the factors in the evolution of PFM is that bases are needed to understand the level of conservatism in typical deterministic evaluation. PFM is a means to calculate best estimate values and the associated uncertainties and margins, and in turn it is a means to quantify conservatisms. By understanding these conservatisms, analysts can refine the safety requirements.

Over the years, the NRC has received numerous submittals that contain PFM results, with varying levels of quality. The inconsistency in the contents of the submittals has often led to low efficiency in the reviews and a lack of predictable regulatory outcomes. For example, the Electric Power Research Institute's (EPRI's) Materials Reliability Program (MRP) has submitted to the NRC several reports containing PFM analyses, either for information or for review and approval. Such efforts include the following:

- "Materials Reliability Program: Probabilistic Fracture Mechanics Analysis of PWR Reactor Pressure Vessel Top Head Nozzle Cracking (MRP-105)," Report 1007834, issued 2004 (Reference 1-7)

- "Materials Reliability Program: Alloy 82/182 Pipe Butt Weld Safety Assessment for U.S. PWR Plant Designs (MRP-113)," Report 1009549, issued 2006 (Reference 1-8)

- "Materials Reliability Program: Probabilistic Risk Assessment of Alloy 82/182 Piping Butt Welds (MRP-116)," Report 1009806, issued 2004 (Reference 1-9)

- "Materials Reliability Program: Inspection and Evaluation Guidelines for Reactor Vessel Bottom-Mounted Nozzles in U.S. PWR Plants (MRP-206)," Report 1016594, issued 2009 (Reference 1-10)

- "Materials Reliability Program: Topical Report for Primary Water Stress Corrosion Cracking Mitigation by Surface Stress Improvement (MRP-335 Revision 3)," Report 3002007392, issued 2016 (Reference 1-11)

- "Materials Reliability Program: Reevaluation of Technical Basis for Inspection of Alloy 600 PWR Reactor Vessel Top Head Nozzles (MRP-395)," Report 3002003099, issued 2014 (Reference 1-12)

- "BWRVIP-05: BWR Reactor Pressure Vessel Shell Weld Inspection Recommendations (BWRVIP-05)," TR-105697, issued 1995 (Reference 1-13)

- “BWRVIP-241-A: BWR Vessel and Internals Project: Probabilistic Fracture Mechanics Evaluation for the Boiling Water Reactor Nozzle-to-Vessel Shell Welds and Nozzle Blend Radii,” Report 3002013093, issued 2018 (Reference 1-14)
- “BWRVIP-108-A: BWR Vessel and Internals Project: Technical Basis for the Reduction of Inspection Requirements for the Boiling Water Reactor Nozzle-to-Vessel Shell Welds and Nozzle Blend Radii,” Report 3002013092, issued 2018 (Reference 1-15)

1.3 Objective

The NRC intends this document to provide a generalized technical basis for the following:

- validating and verifying a PFM capability
- developing input distributions that feed into the PFM framework
- characterizing and propagating input and model uncertainties
- understanding the impacts of problem assumptions on the adequacy of the results
- choosing a methodology with the appropriate complexity for the intended application
- properly conducting a PFM analysis
- correctly interpreting the results of a PFM analysis in a regulatory context
- documenting the important steps and information relevant to the PFM code and analysis at hand

This NUREG provides some thoughts on how to improve confidence in structural analyses performed using PFM by focusing on topics such as problem definition, PFM model development, input definition, uncertainty analyses, probabilistic framework development, and output analysis, including sensitivity analyses (SAs) (to determine impact of uncertainties on result) and sensitivity studies (to determine impact of plausible changes to analysis assumptions). For each of these topics, this NUREG proposes a graded approach for PFM analyses and submittals (see Section 2).

1.4 Structure of This Document

This NUREG has three technical sections. The content provided in all three sections is linked, but an applicant’s experience and familiarity with PFM analyses will determine whether it needs to refer to that content.

Section 2 provides a tiered framework for a submittal that contains PFM analyses and results and could be used by applicants of all experience levels. This section provides a graded approach that may be used in PFM analyses and submittals.

Section 3 provides a framework for performing a PFM analysis. This section could be used by applicants who have used PFM in prior submittals but who are seeking some guidance on the development of an analysis structure or formalism. This section is not intended to prescribe a linear analysis, since PFM analyses are typically iterative in nature. Furthermore, not every application needs all steps and actions, and the analyst can evaluate the necessity to perform each step and action on a case-by-case basis. Table 2-1 in Section 2 provides a mapping between the analysis actions given in Section 3 and the associated documentation.

Section 4 details analysis methodologies, including notional examples for context. This section could be used by applicants who are seeking explicit guidance on the theoretical underpinnings of the processes that are used to establish the credibility of a PFM analysis. Each subsection is linked to an action that was introduced in Section 3.

1.5 References

1-1. "Use of Probabilistic Risk Assessment Methods in Nuclear Regulatory Activities; Final Policy Statement," *Federal Register*, Vol. 60, No. 158, page 42622, August 16, 1995.

1-2. Williams, P.T., Dickson, T.L., Bass, B.R., and Klasky, H.B., "Fracture Analysis of Vessels—Oak Ridge—FAVOR v16.1, Computer Code: Theory and Implementation Algorithms, Methods, and Correlations," ORNL/LTR-2016/309, Oak Ridge National Laboratory, 2016.

1-3. Dickson, T.L., Williams, P.T., Bass, B.R., and Klasky, H.B., "Fracture Analysis of Vessels—Oak Ridge—FAVOR v16.1, Computer Code: User's Guide," ORNL/LTR-2016/310, Oak Ridge National Laboratory, 2016.

1-4. Rudland, D.L., and Harrington, C., "xLPR Pilot Study Report," NUREG-2110, U.S. Nuclear Regulatory Commission, May 2012.

1-5. *U.S. Code of Federal Regulations*, "Domestic licensing of production and utilization facilities," Part 50, Chapter I, Title 10, "Energy."

1-6. Wilkowski, G., Tregoning, R., Scott, P., and Rudland, D., "Status of Efforts to Evaluate LOCA Frequency Estimates Using Combined PRA and PFM Approaches," in *Proceedings of 28th MPA Seminar*, Stuttgart, Germany, October 2002.

1-7. Electric Power Research Institute, "Materials Reliability Program: Probabilistic Fracture Mechanics Analysis of PWR Reactor Pressure Vessel Top Head Nozzle Cracking (MRP-105)," Report 1007834, 2004.

1-8. Electric Power Research Institute, "Materials Reliability Program: Alloy 82/182 Pipe Butt Weld Safety Assessment for US PWR Plant Designs (MRP-113)," Report 1009549, 2006.

1-9. Electric Power Research Institute, "Materials Reliability Program: Probabilistic Risk Assessment of Alloy 82/182 Piping Butt Welds (MRP-116)," Report 1009806, 2004.

1-10. Electric Power Research Institute, "Materials Reliability Program: Inspection and Evaluation Guidelines for Reactor Vessel Bottom-Mounted Nozzles in U.S. PWR Plants (MRP-206)," Report 1016594, 2009.

1-11. Electric Power Research Institute, "Materials Reliability Program: Topical Report for Primary Water Stress Corrosion Cracking Mitigation by Surface Stress Improvement (MRP-335 Revision 3)," Report 3002007392, 2016.

1-12. Electric Power Research Institute, “Materials Reliability Program: Reevaluation of Technical Basis for Inspection of Alloy 600 PWR Reactor Vessel Top Head Nozzles (MRP-395),” Report 3002003099, 2014.

1-13. Electric Power Research Institute, “BWRVIP-05: BWR Reactor Pressure Vessel Shell Weld Inspection Recommendations (BWRVIP-05),” TR-105697, 1995.

1-14. Electric Power Research Institute, “BWRVIP-241-A: BWR Vessel and Internals Project: Probabilistic Fracture Mechanics Evaluation for the Boiling Water Reactor Nozzle-to-Vessel Shell Welds and Nozzle Blend Radii,” Report 3002013093, 2018.

1-15. Electric Power Research Institute, “BWRVIP-108-A: BWR Vessel and Internals Project: Technical Basis for the Reduction of Inspection Requirements for the Boiling Water Reactor Nozzle-to-Vessel Shell Welds and Nozzle Blend Radii,” Report 3002013092, 2018.

2 CONTENT OF A PROBABILISTIC FRACTURE MECHANICS SUBMITTAL

Building toward the release of DG-1382/RG-1.245, “Preparing Probabilistic Fracture Mechanics (PFM) Submittals,” (Reference 2-1) , the NRC held a series of public meetings and released a technical letter report, “Important Aspects of Probabilistic Fracture Mechanics Analyses,” in 2018 (Reference 2-2), to present a general framework of the expected content of a PFM analysis. The NRC’s desire to ensure that submittals containing PFM information are sufficiently clear and complete guided the development of this NUREG. The submittal guidance in this section should not be considered as a set of mandatory requirements.

2.1 Background on a Probabilistic Fracture Mechanics Graded Approach

In the past, the NRC has typically regulated the use of nuclear structural materials on a deterministic basis. Safety factors, margins, and conservatisms were used to account for model and input uncertainty. However, as described in Section 1, the NRC has made progress in its efforts to implement risk-informed approaches into its regulation. In one such effort, the NRC developed guidance on a risk-informed decisionmaking process that is acceptable to use for design-basis changes.

In any regulatory submittal, the level of effort associated with analysis and documentation activities is dependent upon the goals of the analysis, and there is no universal set of guidelines. Instead, each analysis is considered uniquely within its own specific context to make determinations about the expected level of rigor. This guidance is particularly true as the safety significance of the analysis application increases, and the consequences of an incorrect decision are more severe. The availability of supplemental evidence to support the decision is also part of the consideration. For example, if inspection data or operational measurements are available in addition to analysis results, the analysis may be viewed as one piece of evidence in a larger context, and the level of rigor may be adjusted accordingly. The guiding principle is that the content of any PFM submittal should contain detail that is commensurate with the safety significance of the subject and the complexity of the problem.

In October 2018, the NRC held a public meeting to discuss a graded approach for PFM codes and analyses for regulatory applications. At the meeting, EPRI presented suggestions for expected content in a PFM submittal. EPRI also submitted a white paper containing additional details and guidelines. The NRC staff concurred that EPRI’s approach constituted a quality basis from which to build further guidance. Consequently, based on a submitted proposal from industry (Reference 2-3), Section 2.2 defines a practical framework for the content of PFM submittals to maintain the effectiveness of NRC reviews of such submittals while improving review efficiency.

2.2 Submittal Content Summary

This section outlines a detailed framework for PFM submittals that integrates Section 2 of EPRI’s white paper (Reference 2-3) with Section 3 in this document. Each subsection relates to an item expected in a submittal. The content in each subsection comes from, in large part, the suggested minimum content for PFM submittals that was developed in EPRI’s white paper. Tables in each subsection provide guidance for different documentation expectations. Each table contains circumstances under which specific information should be provided for a complete submittal. Each subsection is also mapped to the NRC’s analytical steps in Section 3 and the item number of the suggested minimum content and considerations of additional content given in Tables 1 and 2, respectively, of EPRI's white paper (Reference 2-3). Table 2-1 gives the complete mapping.

Table 2-1 Submittal Content Mapping to NUREG Section and EPRI White Paper (Reference 2-3)

NUREG Section	Content	NUREG PFM Analytical Steps Section 3	EPRI White Paper (Reference 2-3)	
			Suggested Content, Table 1	Additional Considerations, Table 2
0	Regulatory Context	3.1.1	-	-
2.2.2	Information Made Available to NRC Staff		1	-
2.2.2.1	PFM Software	3.1.3	1.1	1, 4, 11, 12, 13
2.2.2.2	Supporting Documents	3.1.3	1.2	-
2.2.3	Quantities of Interest and Acceptance Criteria	3.1.2	8	-
2.2.4	Software quality assurance and verification and validation	3.1.3	6	1
2.2.5	Models	3.1.3	2	1, 2, 5, 6, 9, 10
2.2.6	Inputs	3.2.1 3.2.2 3.3.1 3.4.1	3, 5	3, 4, 5, 6
2.2.7	Uncertainty Propagation	3.3.1	7	3, 10
2.2.8	Convergence	3.3.2	4	3
2.2.9	Sensitivity Analyses	3.3.3	-	-
2.2.10	Output Uncertainty Characterization	3.3.4	-	-
2.2.11	Sensitivity Studies	3.4.1 3.4.2	5	1, 2, 11

It is important to note that submittals should be dictated by the specific details and elements of each analysis and need not include all of the listed elements, though careful consideration should be applied to arrive at that conclusion.

2.2.1 Regulatory Context

When using PFM in support of an application to the NRC, it is important to understand the need for using a probabilistic approach and how PFM informs whether regulatory requirements have been met (Section 3.1.1), specifically why a probabilistic approach is appropriate for the problem at hand and how the probabilistic approach is used to demonstrate compliance with the regulatory criteria. It is particularly important to explain how the probabilistic approach informs the regulatory action when no specific acceptance criteria exist for demonstrating compliance for the problem at hand.

2.2.2 Information Made Available to the NRC Staff with a Probabilistic Fracture Mechanics Submittal

The applicant should have a plan addressing supporting information that may be necessary to review the submittal. This may include information made available with the submittal, which may be provided upon request, or which may not be directly transmittable but might be reviewed under specific agreed-upon circumstances.

2.2.2.1 Probabilistic Fracture Mechanics Software

A key factor in the process of the NRC developing confidence in PFM software is its availability to the NRC staff, the opportunity to perform benchmarking studies against similar existing NRC codes, or both. If a sufficiently similar PFM code is not available for the NRC to perform a meaningful benchmarking comparison, an alternate approach such as the following should be considered:

- The NRC staff could participate with the applicant in an informal review meeting during which the PFM submittal developers run analysis cases as requested by the NRC staff.

- The NRC staff could submit some analysis requests in advance of the meeting, or two separate meetings could be held to allow time between meetings for the PFM submittal developers to run cases.

- To address runtime concerns, developers could optimize runtime or consider a fast run mode that does not include all code features. Having some capability to perform runs during a review meeting would be advantageous.

More complex codes and new codes may warrant a more thorough review (e.g., more meetings or more cases run by request) than do codes more familiar to the NRC staff. The extent of the differences between the new code and the codes previously approved by the NRC should also be considered. Similar considerations should take place when a code previously reviewed by the NRC is applied in a new way (i.e., outside the previously reviewed range of use for the code). Certain specific applications of the code, such as those involving a high safety significance or if the code is plant specific (versus an intended generic application), may also warrant deeper and more thorough investigations.

2.2.2.2 Probabilistic Fracture Mechanics Software Quality Assurance and Verification and Validation Documents

The quality assurance (QA) program or procedures under which the PFM analysis code is being developed (and thus also the standards with which that QA program or procedures comply) will define what additional supporting QA and verification and validation (V&V) documents will need to be generated (see Section 2.2.4). It is not necessary or appropriate to transmit all such supporting documentation to the NRC. However, the organization(s) that developed the PFM analysis code will need to retain that supporting documentation. Depending on the application of the PFM code, different QA programs may apply, which in turn may impact the level of documentation required. Regardless of the QA program or procedures applied, the applicant might consider ways to facilitate making such supporting documents available for examination by the NRC staff during an in-person audit.

2.2.3 Quantities of Interest and Acceptance Criteria

The submittal to the NRC should document the model output quantities of interest (QoIs) and the probabilistic acceptance criteria that are being applied for the PFM analysis. The basis for those acceptance criteria should also be provided, such as a previous precedent established by the NRC. Appropriate care must be taken when invoking previously approved acceptance criteria from a similar analytical process or evaluation framework to ensure that inherent assumptions and requirements of the source activity are respected, and any apparent differences are reconciled.

The NRC typically approves the acceptance criteria, which may be relative or absolute. Relative acceptance criteria refer to a relative comparison of probabilistic results under the proposed approach versus an already acceptable approach. In general, the rigor required in demonstrating that a relative acceptance criterion is met is lower than that required in demonstrating that an absolute acceptance criterion is met.

Acceptance criteria for any application are beyond the scope of DG-1382/RG-1.245 (Reference 2-1) and this NUREG, but they should be derived based on risk-informed decisionmaking principles. Regulatory Guide 1.200 (Reference 2-4), "Acceptability of Probabilistic Risk Assessment Results for Risk-Informed Activities," and Regulatory Guide 1.174 (Reference 2-5), "An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis," discuss the topic further.

Table 2-2 relates to the documentation of QoI and acceptance criteria. If the analysis includes more than one QoI, then these elements should be documented for each QoI.

Table 2-2 Submittal Guidelines for QoI and Acceptance Criteria

Submittal Guidelines	Reference
The QoI definition, including both the units of measurement and time period	Section 3.1.2
The relationship between the QoI and model output	Section 3.1.2
The acceptance criteria	Section 3.1.2
If the QoI is a rare probability, a description of how this affected analysis choices	Section 3.1.4

2.2.4 Software Quality Assurance and Verification and Validation

In any analysis, the level of effort associated with software quality assurance (SQA) and V&V activities is dependent upon the goals of the analysis, and there is no universal set of requirements. For different analyses, the level of experience in using the tools may vary. For a code that the NRC has previously approved, the technical basis for using the code is likely well understood, such that supplemental SQA and V&V efforts are unnecessary to understand the credibility of the results. For a code that the NRC has previously approved but that has been modified for the analysis being performed, understanding the technical basis for the modifications is important. For a code that is new and has not been previously approved in any form, understanding the entire technical basis informs the credibility of the results. With this in mind, the set of different analysis codes can be divided into three categories, defined in Table 2-3.

V&V may be performed on individual submodels and the unifying framework, or it can be performed directly on the overall code. Some QA programs also allow for checks using alternate calculation methods (e.g., spreadsheets or alternate implementations). The applicable QA program, plan, or procedures define the supporting documents created in conjunction with PFM analysis code development. A graded approach to QA for software development, with different minimum requirements depending on the software application, such as that outlined in International Atomic Energy Agency (IAEA) Technical Report Series No. 397, “Quality Assurance for Software Important to Safety,” issued in 2000 (Reference 2-6), may be considered. Furthermore, as the applicable QA program may depend on the safety significance of the component or system being evaluated, the corresponding rigor of V&V may also vary.

If a code is used for an application that is different than the one for which it was developed, the existing verification may still be valid, but the validation may need to be extended or redone if the previous validation was specific to a different range of use.

2.2.5 Models

The goal of any engineering assessment methodology is to determine the response of a system to a variety of inputs. The assessment of the system in question should be designed and developed using a set of models that best represents the physical behavior of the system of interest. Model selection may involve balancing the accuracy and practicality of various mathematical approaches. Whenever a model is constructed, simplifications are injected into the representation to make model evaluation feasible. The rationale for each decision or simplification is communicated, and then, where possible, the effect of this simplification on the analysis results is determined. It may not always be possible to construct multiple models of varying fidelity to consider the numerical sensitivity to the analysis choices, but the influence of these choices can be considered qualitatively.

Engineering judgment in the model development process helps assess the credibility of the analysis results. Determining the relevant physics and material behavior to capture in a model of interest is a critical first step in determining the applicability of a given code and modeling approach. The process of developing a model begins with a conceptual model, which defines the physics to include. This decision is often aided by a process that defines the most critical physics to capture in the analysis. Then, for each relevant physics, a mathematical model is chosen to represent that physics, and a code is selected or developed to solve the chosen mathematical model. Over the course of an analysis, the model and code may be updated, revised, or calibrated with available data to improve predictive capability and understand how similar the conditions of validation tests are to the application space of interest.

Another factor for consideration is computational resources. While a particular approach may be considered the “best estimate,” it may not be practical for a PFM analysis given the time and resource constraints imposed on the analyst. The occasional need to choose a model that has less fidelity but is easier to solve due to the solution speed requirements of PFM may affect results. In such cases, biases or uncertainties relative to the “best estimate” model should be quantified and accounted for by propagating the associated uncertainty through the probabilistic model. Model choice can be complicated further by the fact that PFM requires the use of the most accurate deterministic models rather than conservative models. These more accurate models may require longer solution times but yet still contain systematic model biases and uncertainties.

The inclusion of increased detail in model development may be appropriate if the applied model is relatively new (e.g., first-of-a-kind applications), the failure mode represents a new phenomenon, the failure phenomenon is emergent or ongoing, the extent of the plant experience and operational experience with the phenomenon is small, and the implications of the unknowns are not well characterized or yield significantly different outcomes. In all cases, the degree of detail should be commensurate with the safety significance of the components being evaluated.

Table 2-4 provides information on model documentation.

2.2.6 Inputs

Both deterministic and uncertain[1] inputs should be documented in detail to justify the choice of inputs as appropriate for the PFM application. The rationale for input choices should be described clearly and supported by relevant data, references, sensitivity analyses/studies, expert judgment, or a combination of these. For inputs that are highly important or that have uncertainty in their characterization, the NRC may need additional information (e.g., the data to which a probability distribution was fit).

Table 2-5 and Table 2-6 relate to the documentation of inputs. In Table 2-5, “knowledge” refers to the depth of information available to prescribe either the deterministic inputs or the distributions on the uncertain inputs. “Importance” refers to the relative effect of input on the QoI. If certain inputs are found to have a significant effect on the QoI (e.g., through SAs), a more detailed description of the basis for those inputs may be needed. Additionally, sensitivity studies (Section 3.4.1) may be necessary to demonstrate the effect of input classification or distribution choices on the QoI.

A more in depth description of the basis for inputs may also be needed if the failure mode is poorly understood or has a large impact on other systems or safety or if there is an emergent issue or first-of-a-kind application. If there is little margin between the QoI and the acceptance criteria (e.g., less than one order of magnitude in the case of probabilities or frequencies), more scrutiny of highly important inputs is warranted. Additionally, if the submittal is for a generic application, additional proof may be needed to ensure the inputs cover a wide enough range for the application.

2.2.7 Uncertainty Propagation

Propagation of uncertainty in the model inputs is a key component of estimating uncertainty in the QoI. Documentation should be provided that explicitly describes the methods used for uncertainty propagation and allows for the reproduction of analysis results. If the code is computationally expensive, details should be provided on any additional measures that were taken to adequately propagate uncertainty under the computational constraints. In particular, if importance sampling is used to oversample important regions of the input space, a justification for the choice of importance distribution should be provided. If applicable, documentation should include details on the surrogate model used for uncertainty propagation, including the surrogate model form, any approximations or assumptions, the method used for fitting the surrogate, and a measure of the error associated with the surrogate model approximation.

[1] Throughout this document, “uncertain” refers to input variability because of the randomness of the data or lack of knowledge; the term may be used interchangeably with “random.”

Table 2-7 provides information on the documentation of uncertainty propagation.

2.2.8 Convergence

To assess the convergence of the QoI estimate, documentation should be provided that demonstrates the convergence for any discretization used in the analysis (e.g., time step, spatial discretization), as well as statistical convergence based on the sample size and sampling method used in the probabilistic analysis. The primary goal should be to show that the conclusions of the analysis would not change if a more refined discretization or a larger sample size were used.

If significant margin exists between the QoI and the acceptance criteria, less stringent convergence levels may be adequate. In this case, a basis for defining that margin should be provided. For new or modified codes (categories QV-1C, QV-2, and QV-3), more indepth discretization convergence analyses may be needed. When maintaining a separation of aleatory and epistemic uncertainties, a sample size convergence analysis should be performed for both the aleatory and epistemic sample sizes.

PFM codes in categories QV-1A and QV-1B are exempted from documenting discretization convergence, but analysts should nonetheless verify that discretization convergence is achieved. For all other PFM codes, the applicant should document the approach used for assessing discretization convergence, as well as demonstrate and document that a more refined discretization does not significantly affect the outcome of the analysis.

Table 2-8 relates to the documentation of statistical convergence.

2.2.9 Sensitivity Analyses

Sensitivity analysis (SA) is a useful tool for identifying important uncertain model inputs that explain a high proportion of the uncertainty in the QoI. There are many approaches to performing SA, and, therefore, it is important to document the method(s) used, any relevant assumptions, and the interpretation of the results. Inputs that are important may warrant additional scrutiny.

More indepth documentation of SAs and important model inputs may be needed if there is a high safety significance, the failure mode is poorly understood, there is a low margin between the QoI and the acceptance criteria, or the application is the first of a kind. Additional documentation may also be necessary if the submittal is requesting a change to the plant licensing basis.

Table 2-9 covers the documentation of SAs.

2.2.10 Quantity of Interest (Output) Uncertainty Characterization

It is important to characterize the QoI uncertainty clearly and accurately such that the results from the analysis are easy to interpret. When characterizing QoI uncertainty, information about the scope of the analysis, its limitations, and any conservatisms should be documented. A conclusive description of the results of the analysis based on this characterization should also be provided.

If little margin exists between the QoI and the acceptance criteria (e.g., less than one order of magnitude in the case of probabilities or frequencies), more indepth documentation of model assumptions and simplifications may be needed. Additionally, if potential unknowns may affect analysis conclusions, they should be noted when discussing output uncertainty results.

Table 2-10 relates to the documentation of QoI uncertainty characterization.

2.2.11 Sensitivity Studies

Sensitivity studies can be used to assess how uncertain analysis assumptions may change analysis results. Results of the sensitivity studies can be used to justify or prompt the refinement of analysis choices. A detailed description of the sensitivity studies performed and their conclusions should be provided.

More extensive sensitivity studies may be needed when an approved code has been modified, for new and highly complex codes (as the impact of different modeling choices may not be fully understood), when there is a high safety significance or poorly understood failure mode, when SAs demonstrate that particular variables drive uncertainty, or if there are large perceived uncharacterized uncertainties.

Table 2-11 covers the documentation of sensitivity studies.

2.2.12 Submittal Guideline Tables

The tables in this section contain all of the submittal guidelines for Sections 2.2.4 through 2.2.11. In general, when using these tables, the applicant should in categorize each input independently. However, the implications of input correlation or input dependencies on the input categorization should be considered.

Table 2-3 SQA and V&V Code Categories

Category	Description	Submittal Guidelines	Reference
QV-1	NRC-approved code [a]		
QV-1A	Exercised within validated range	Demonstrate code applicability within the validated range.	
		Describe features of the specific application where the code is validated and applicable (i.e., areas of known code capability).	Section 3.1.3
QV-1B	Exercised outside of validated range	Provide evidence for the applicability of the code to the specific application with respect to the areas of unknown code capability.	Section 3.1.3
		Describe features of the specific application where the code has not been previously validated and applied (i.e., areas of unknown code capability).	Section 3.1.3
QV-1C	Modified	Give an SQA summary and V&V description for modified portions of the code.	
		Demonstrate that the code was not “broken” as a result of changes.	
		Make detailed documentation available for further review upon request (audit).	
QV-2	Commercial off-the-shelf software designed for the specific purpose of the application [b]	Demonstrate code applicability.	
		Describe the software and its pedigree.	
		Make software and documentation available for review upon request (audit).	
QV-3	Custom code	Summarize the SQA program and its implementation.	Section 3.1.3
		Provide a basic description of the measures for QA, including V&V of the PFM analysis code as applied in the subject report.	Section 3.1.3
		For very simple applications, possibly provide the source code instead of standardized SQA and V&V.	
		Include separate deterministic fracture mechanics analyses to support other validation results, as appropriate for a given application.	Section 3.1.3

[a] As of the publication of this NUREG/CR, NRC-approved PFM codes include the latest version of the FAVOR and xLPR codes, as well as the SRRA code approved in the Safety Evaluation Report related to Topical Report WCAP-14572, Revision 1 (Reference 2-7).

[b] Examples would include publicly available (for purchase or free) commercial software specifically to perform PFM analyses. Combinations of commercial off-the-shelf software may be acceptable (e.g., a finite-element software such as ABAQUS or ANSYS coupled with a probabilistic framework such as GoldSim or DAKOTA).

Table 2-4 Submittal Guidelines for Models

Category	Description	Submittal Guidelines	Reference
M-1	Model from a code in category QV-1A or QV-1B within the same validated range	Reference existing documentation for that model in the NRC-approved code, demonstrate that the current range of the model is within the previously approved and validated range, and demonstrate that the model functions as intended in the new software.	
M-2	Model from a code in category QV-1A or QV-1B outside the validated range	See the submittal guidelines for M-1, except demonstrate validity of the model for the <u>new</u> applicability range (document a comparison of model predictions for the entire new range to applicable supporting data, including quantitative goodness-of-fit analyses).	
M-3	Model derived from a category M-1 or M-2 model	See the submittal guidelines for M-2, and include a detailed description of changes to the M-1 or M-2 model, with justification for the validity of the new model.	
M-4	Well-established model not previously part of an NRC-approved code	Describe gaps and limitations in the code capabilities for the analysis, combined with a strategy for mitigating identified gaps and communicating any remaining issues or risks.	Section 3.1.3
		Describe the model(s) applied in the PFM analysis code in sufficient detail so a competent analyst familiar with the relevant subject area could independently implement the model(s) from the documentation alone. Model forms can either be theoretical, semiempirical, or empirical.	Section 3.1.3
		Establish a basis for all significant aspects of the model(s). This may consist of raw data or published references. Document or reference any algorithms or numerical methods (e.g., root-finding, optimization) needed to implement the model(s). Discuss any significant assumptions, approximations, and simplifications made, including their potential impacts on the analysis.	Section 3.1.3
		Identify important uncertainties or conservatisms.	Section 3.1.3
		Describe the computational expense of the model and how that might affect analysis choices.	Section 3.1.4
M-5	First-of-a-kind model not yet published in a peer-reviewed journal	See the submittal guidelines for M-4, and perform and document model sensitivity studies to understand trends in the model, as compared to expected model behavior and to the data used to develop the model, and describe model maturity and the status of the technical basis.	

Table 2-5 Categorization Based on Knowledge and the Importance of Inputs Used in the Analysis

Input Category	Low Knowledge of Input Characteristics		High Knowledge of Input Characteristics	
	Deterministic	Uncertain	Deterministic	Uncertain
High Importance	I-4D	I-4R	I-3D	I-3R
Low Importance	I-2D	I-2R	I-1D	I-1R

Table 2-6 Submittal Guidelines for Inputs

Category	Submittal Guidelines	Reference
I-1D	List input value.	
I-1R	List input distribution type and parameters.	
	If applicable, list uncertainty classification (aleatory or epistemic).	Section 3.2.1
I-2D	List input value.	
	If there is a lack of data, justify the use of expert judgment.	Section 3.2.2
I-2R	List input distribution type and parameters.	
	If applicable, list uncertainty classification (aleatory or epistemic).	
	If there is a lack of data, justify the use of expert judgment.	Section 3.2.2
I-3D	List input value.	
	State the rationale for setting the input to a deterministic value.	
	State the rationale for setting the input to a deterministic value.	Section 3.2.1
	For each deterministic input, give the rationale (method and data) for the selection of its numerical value, along with any known conservatisms in that numerical value and the rationale for such conservatisms.	Section 3.2.1
	Reference documents that contain the foundation for input choices.	Section 3.2.2
	Explain the correlations between inputs and how they are modeled, and verify that correlated inputs remain consistent and physically valid.	Section 3.2.2
	Describe any sensitivity analyses/studies performed to show that the input or its classification does not have a significant effect on the QoI.	Section 3.2.2
I-3R	List input distribution type and parameters.	
	If applicable, list uncertainty classification (aleatory or epistemic).	
	If relevant, classify uncertain inputs as aleatory or epistemic and give the corresponding rationale.	Section 3.2.1
	For each uncertain input, describe both its distribution parameter values and its distributional form. Give the rationale (method and data) for selecting each distribution, including any known conservatisms in the specified input distributions and the rationale for the conservatism. Detail the distributional fitting method, including interpolation, extrapolation, distribution truncation, and curve fitting.	Section 3.2.2
	Reference documents that contain the foundation for input choices.	Section 3.2.2
	Explain the correlations between inputs and how they are modeled, and verify that correlated inputs remain consistent and physically valid.	Section 3.2.2
	Describe any sensitivity analyses/studies performed to show that the input or its classification does not have a significant effect on the QoI.	Section 3.2.2
I-4D	See the submittal guidelines for I-3D.	
	If there is a lack of data, justify the use of expert judgment.	Section 3.2.2
I-4R	See the submittal guidelines for I-3R.	
	If there is a lack of data, justify the use of expert judgment.	Section 3.2.2

Table 2-7 Submittal Guidelines for Uncertainty Propagation

Category	Description	Submittal Guidelines	Reference
UP-1	Analysis does not employ a surrogate model	Give the method for uncertainty propagation and describe the simulation framework.	Section 3.3.1
		If Monte Carlo sampling is used, describe the finalized sampling scheme and rationale for the sampling scheme, including the sampling method, sample size, the pseudo-random number generation method, and the random seeds used.	Section 3.3.1
		Describe the approach for maintaining separation of aleatory and epistemic uncertainties, if applicable.	Section 3.3.1
		If importance sampling is used to oversample important regions of the input space, justify the choice of importance distribution.	Section 3.3.1
UP-2	Analysis does employ a surrogate model	See the submittal guidelines for UP-1, and describe the form of the surrogate model(s), any approximations or assumptions, the method used for fitting the surrogate, and the validation process for the surrogate model.	
UP-2A	Surrogate model is used for SA	See the submittal guidelines for UP-2, and describe the features of the different surrogate models used.	
UP-2B	Surrogate model is used for uncertainty propagation	See the submittal guidelines for UP-2, and quantify the magnitude of error associated with the surrogate model approximation and include as additional uncertainty in the estimation of the QoI.	

Table 2-8 Submittal Guidelines for Statistical Convergence

Category	Description	Submittal Guidelines	Reference
SC-1 [a]	[Acceptance criteria met with at least one order of magnitude margin] AND [no importance sampling AND no surrogate models used]	No sampling uncertainty characterization recommended, as long as the uncertainty is sufficiently small relative to the margin. [b]	
SC-2A	[Acceptance criteria met with at least one order of magnitude margin] AND [use of importance sampling OR surrogate models OR both]	Describe the approach used for assessing statistical convergence, with one method needed for sampling uncertainty characterization.	Section 3.3.2
		Explain the approach used for characterizing sampling uncertainty.	Section 3.3.2
		Justify why the sampling uncertainty is small enough for the intended purpose (i.e., why statistical convergence is sufficient for the intended purpose).	Section 3.3.2
		Describe how sampling uncertainty is used in the interpretation of the results.	Section 3.3.2
SC-2B	[[Acceptance criteria met with at least one order of magnitude margin] AND [use of importance sampling OR surrogate models OR both]] AND [separation of aleatory and epistemic uncertainties is implemented in the PFM code]	See the submittal guidelines for SC-2A, and distinguish between epistemic and aleatory means and standard deviations.	
SC-3A	[Acceptance criteria met with less than one order of magnitude margin]	See the submittal guidelines for SC-2A, and provide two different methods for sampling uncertainty characterization.	
SC-3B	[Acceptance criteria met with less than one order of magnitude margin] AND [separation of aleatory and epistemic uncertainties is implemented in the PFM code]	See the submittal guidelines for SC-3A, and give a sample size convergence analysis for both the aleatory and epistemic sample sizes.	Section 3.3.2

[a] Data type may have an impact on the convergence category. Continuous outputs can be category SC-1, but binary outputs inherently must be category SC-2 or SC-3 unless epistemic and aleatory uncertainties are separated.

[b] Some assessment of uncertainty is necessary, even if qualitative, as long as the uncertainty itself is understood to be small.

Table 2-9 Submittal Guidelines for SAs

Category	Description	SA Needed? [a]	Submittal Guidelines	Reference
SA-1	Previously approved code (QV-1, QV-1A) with same QoI and same inputs [b]	No	Describe important input and measure of input importance from previous use.	
SA-2	Previously approved code (QV-1, QV-1A) with different QoI	Yes	Explain the methods used for SA, including any initial screening and model approximations and assumptions.	Section 3.3.3
			State whether a local or global SA approach is used.	Section 3.3.3
			Give the QoI used for the SA.	Section 3.3.3
			For a global SA, describe the sampling scheme along with the rationale for selection, including the sampling technique, number of model realizations, and random seed for the model realizations.	Section 3.3.3
			Provide the results of the SA, including the most important model inputs identified; a measure of the input importance, such as the variance explained by the most important inputs; and relevant graphical summaries of the SA results.	Section 3.3.3
SA-3	Modified approved code with limited independent variables (e.g., <5, determined on a case-by-case basis)	Yes	Describe analyses, important input, and measure of input importance.	
SA-4	Modified approved code with many independent variables (e.g., >5, determined on a case-by-case basis)	Yes	See the submittal guidelines for SA-2.	
SA-5	First-of-a-kind code with limited independent variables (e.g., <5, determined on a case-by-case basis)	Yes	Describe the analyses, important input, and measure of input importance and include additional documentation.	
SA-6	First-of-a-kind code with many independent variables (e.g., >5, determined on a case-by-case basis)	Yes, with submodel SA as appropriate	See the submittal guidelines for SA-2.	
			Indicate how the SA results informed future uncertainty propagation for estimation of the QoI and associated uncertainty.	Section 3.3.3
			State whether the results of the SA are consistent with the expected important inputs based on expert judgment.	Section 3.3.3

[a] Local sensitivity analysis may be used as a screening step if completing a global sensitivity analysis with all inputs is not computationally feasible (as the cost of performing a global sensitivity analysis increases with the number of inputs). The results from local sensitivity analysis can help reduce the input space for a global sensitivity analysis, but local sensitivity analysis does have its risks in that it can miss important inputs if the input/output relationship is nonlinear. Sensitivity analysis should be performed unless there is a strong basis for what inputs are important (e.g., previous analyses, expert judgment, or it is obvious what inputs are important since it is a simple code).

[b] Inputs must remain the same because sensitivity is dependent on the input distributions.

Table 2-10 Submittal Guidelines for QoI (Output) Uncertainty Characterization

Category	Description	Submittal Guidelines	Reference
O-1	Acceptance criteria met with at least one order of magnitude margin	Give a measure of the best estimate and uncertainty in the QoI.	Section 3.3.4
		Include a graphical display of the output uncertainty.	Section 3.3.4
		Describe how the best estimate and its uncertainty were calculated, including a clear description of the types of uncertainty (e.g., input, sampling, epistemic) being summarized.	Section 3.3.4
		Summarize key uncertainties considered in the analysis and any major assumptions, conservatisms, or simplifications that were included and assess (qualitative or quantitative) their effect on the analysis conclusions.	Section 3.3.4
O-2A	Acceptance criteria met with less than one order of magnitude margin and a strong basis for input distributions and uncertainty classification	See the submittal guidelines for O-1, and provide the reasoning behind a strong basis.	
O-2B	Acceptance criteria met with less than one order of magnitude margin and no strong basis for input distributions or uncertainty classification, or both	See the submittal guidelines for O-1.	
		Include an SA (if important inputs are unknown) and sensitivity studies for any inputs that do not have a strong basis.	Section 3.4.1 Section 3.4.2
O-3	O-1, O-2A, or O-2B and potential unknowns	See the submittal guidelines for O-1, and provide the reasoning behind a strong basis. Describe potential unknowns and their possible effect on analysis results.	
		OR	
		Include an SA (if important inputs are unknown) and sensitivity studies for any inputs that do not have a strong basis.	Section 3.4.1 Section 3.4.2

Table 2-11 Submittal Guidelines for Sensitivity Studies

Category	Description	Sensitivity Study Needed?	Submittal Guidelines	Reference
SS-1	Category QV-1A code with same QoI [a]	No	Summarize sensitivity studies conducted in prior approval.	
SS-2	Category QV-1A code with different QoI	Limited, focused on inputs related to QoI	Summarize past sensitivity studies conducted in prior approval and current sensitivity studies.	
SS-3	Category QV-1B or QV-1C code with limited independent variables (e.g., <5, determined on a case-by-case basis)	Limited, focused on impact of modification	Summarize past and current sensitivity studies.	
SS-4	Category QV-1B or QV-1C code with many independent variables (e.g., >5, determined on a case-by-case basis)	Yes, focused on inputs related to QoI	Summarize past and current sensitivity studies.	
			List the uncertain assumptions that are considered for sensitivity studies.	Section 3.4.1
			State the impact and conclusion of each sensitivity study.	Section 3.4.1
			Give the rationale for why certain assumptions were or were not considered for sensitivity studies.	Section 3.4.1
			Provide the specific question(s) each sensitivity study is attempting to answer.	Section 3.4.2
			Describe a reference realization.	Section 3.4.2
			Describe how each sensitivity study is translated into model realizations, and compare the study and the reference realization.	Section 3.4.2
			List changes to the code and the QA procedure used.	Section 3.4.2
SS-5	Category QV-2 or QV-3 code with limited independent variables (e.g., <5, determined on a case-by-case basis)	Yes	See the submittal guidelines for SS-4.	
SS-6	Category QV-2 or QV-3 code with many independent variables (e.g., >5, determined on a case-by-case basis)	Yes, model and input studies	See the submittal guidelines for SS-4.	

[a] Inputs must remain the same because sensitivity is dependent on the input distributions.

2.3 References

2-1. U.S. Nuclear Regulatory Commission, DG-1382/RG-1.245: Preparing Probabilistic Fracture Mechanics (PFM) Submittals, Washington, DC, USA: U.S. NRC.

2-2. Raynaud, P., Kirk, M., Benson, M., and Homiack, M., “Important Aspects of Probabilistic Fracture Mechanics Analyses,” U.S. Nuclear Regulatory Commission, 2018.

2-3. Palm, N., “White Paper on Suggested Content for PFM Submittals to the NRC,” BWRVIP 2019-016, Electrical Power Research Institute, 2019.

2-4. U.S. Nuclear Regulatory Commission, RG-1.200: An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities, Washington, DC, USA: U.S. NRC.

2-5. U.S. Nuclear Regulatory Commission, RG-1.174: An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis, Washington, DC, USA: U.S. NRC.

2-6. International Atomic Energy Agency, “Quality Assurance for Software Important to Safety,” Technical Report Series No. 397 (TRS-397), Vienna, Austria, 2000.

2-7. U.S. Nuclear Regulatory Commission, Safety Evaluation Report related to “Westinghouse Owners Group application of Risk-Informed Methods to Piping Inservice Inspection” (Topical Report WCAP-14572, Revision 1), December 1998, Washington, DC: U.S. NRC.

3 ANALYTICAL STEPS IN A PROBABILISTIC FRACTURE MECHANICS ANALYSIS

This section describes a process for conducting a PFM analysis. It is generally assumed that an analysis process is implemented after PFM code quality and credibility have been established through SQA processes and V&V. The process followed in performing analyses for a PFM submittal is not required to be the same as the process outlined here, but it should be structured to address the specific features of the application under investigation.

A generalized PFM analysis process is structured according to five key steps:

(1) Translate regulatory requirements into an analysis plan.
(2) Characterize input uncertainty.
(3) Estimate QoIs and their associated uncertainty.
(4) Conduct sensitivity studies to assess credibility of modeling assumptions.
(5) Draw conclusions from analysis results.

This section describes each step in the PFM analysis process and its corresponding analyst actions, along with the following information:

- **Purpose.** Motivation for including this step in a PFM analysis.
- **Description.** High-level description of the concept.

These steps and actions are intended to provide a conceptual framework for conducting and presenting the results of a PFM analysis that can be used in a risk-informed regulatory assessment, but they are not intended to be performed in a strictly linear fashion. PFM analyses are typically iterative in nature. Furthermore, not all steps and actions are needed in every application, and the analyst should evaluate the necessity to perform each step and action on a case-by-case basis. Different applications will warrant different levels of analysis complexity and documentation. If separate PFM analyses are conducted for different regulatory contexts or QoIs, then these analyses should be documented separately.

Figure 3-1 summarizes the steps and actions and their relationship to one another. This figure also shows the organization of this section and the iterative nature of PFM analyses.

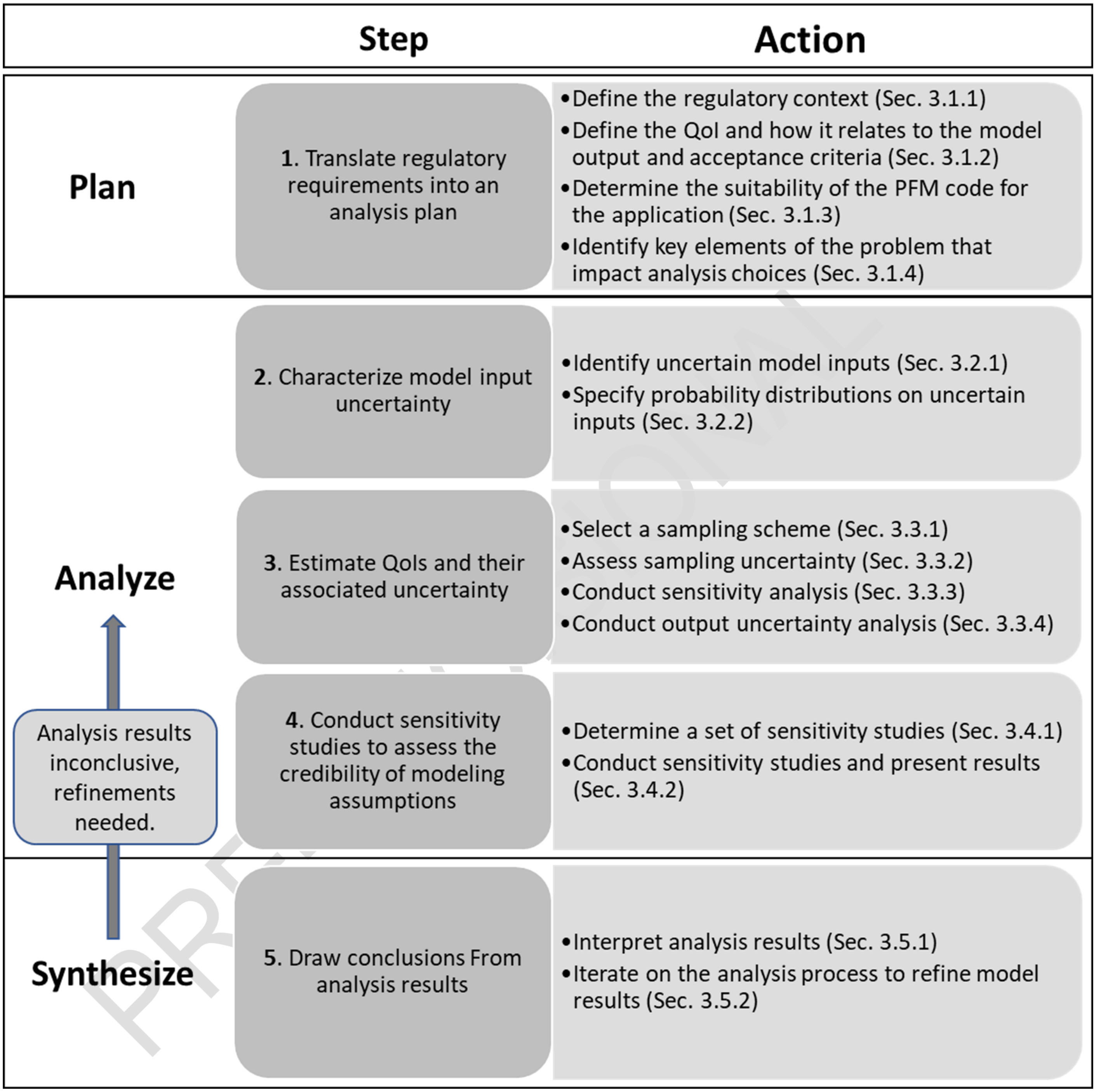

Figure 3-1 Flowchart Describing the Steps and Actions of a PFM Analysis

A key element of risk-informed decisionmaking is identifying uncertainties that impact the analysis results and subsequent regulatory decision. The proposed steps and actions are intended to reflect sources of uncertainty that are common to all PFM applications, including the following:

- **Step 2: Input uncertainty.** The specific values of model inputs are typically unknown; this input uncertainty results in uncertainty in the model output, such as the likelihood of an adverse event. Accounting for this uncertainty in model inputs is what distinguishes deterministic and probabilistic fracture mechanics applications.
- **Step 3: QoI approximation uncertainty.** PFM analyses are based on a finite number of model realizations, resulting in sampling uncertainty. This sampling uncertainty can impact the accuracy of the analysis results.
- **Step 4: Modeling assumption uncertainties.** PFM analyses may rely on assumptions and approximations that introduce additional uncertainty into the analysis. The impact of uncertain assumptions can be addressed using sensitivity studies.

The discussion in this section refers to Section 4, which includes specific technical details about elements of PFM analyses. Section 2.2 suggests documentation for different steps in this process.

3.1 Step 1: Translation of Regulatory Requirements into an Analysis Plan

The first step in a PFM analysis is translating regulatory requirements into a PFM analysis plan. This step involves four key actions:

(1) Define the regulatory context.
(2) Define the QoI and how it relates to the PFM model output.
(3) Determine suitability of the PFM code for the application.
(4) Identify key elements of the problem that impact analysis choices.

3.1.1 Step 1: Action 1—Define the Regulatory Context

Purpose: The purpose of this step is to define how PFM analyses will be used as a technical basis for a regulatory action, including the criteria to be used to support a proposed regulatory action.

Description: When using PFM in support of an application to the NRC, it is important to understand how PFM informs whether regulatory requirements have been met, specifically why a probabilistic approach is appropriate for the problem at hand, and how the probabilistic approach is used to demonstrate compliance with the regulatory criteria. It is particularly important to explain how the probabilistic approach informs the regulatory action when no specific acceptance criteria exist to demonstrate compliance for the problem at hand.

3.1.2 Step 1: Action 2—Define the Quantity of Interest and How it Relates to the Model Output and Acceptance Criteria

Purpose: The purpose of this step is to directly map regulatory requirements onto specific model outputs, ensuring that the model is predicting appropriate and relevant quantities.

Description: The model output is directly linked to one or more QoIs and the acceptance criteria.

A QoI is a quantity that is directly tied to a regulatory decision. The QoI is a model output or a function of outputs; for a PFM model to be useful, understanding the relationship between the model output and the QoI is critical. For example, suppose the QoI is the probability of rupture by year for a single pipe. For each set of inputs, the PFM model may output the year in which rupture occurs. The QoI is then estimated by calculating the frequency of rupture by year across many realizations of this single pipe's performance.

In PFM analyses, the QoI will frequently be a probability of an adverse event; however, using a proxy for an adverse event may be necessary when its probability is too small to accurately estimate using computer simulation. For example, probability of rupture could be related to crack length or crack depth, and one or both of these quantities could potentially be used as surrogates for rupture.

The QoI is typically tied to the acceptance criteria. Often, an acceptance criterion is expressed as a point in the QoI space at which decisions are determined based on whether the QoI exceeds the threshold. An example acceptance criterion is "the 95th percentile of the predicted leak rate must remain below the makeup capacity of the system."

Both the QoI and acceptance criteria are defined relative to the unit of measurement and the time period over which the QoI is calculated.

The unit of measurement specifies the target population for inference, defined as the entire set of objects to which the analyst is trying to generalize the results of the analysis. The QoI is interpreted relative to the units of measurement, such as a fleet of power plants, a single plant, a line within the plant, or a single weld within a plant. The units of measurement can also be defined spatially, such as per kilometer of pipe.

The time period is the interval of time over which the QoI is calculated, such as per year, per decade, or over the life of the plant.

As an example, consider an analysis intended to show that the likelihood of a single pipe leaking is small over the life of a plant. The QoI is the probability of pipe leakage, the acceptance criterion is the acceptable upper limit on the probability of leakage, the time period is the plant life duration, and the units are the single pipe of interest. All quantities are dependent on the modeling assumptions. For example, no mitigation, 10-year inspection intervals, and the leak detection system all impact the assessments.

3.1.3 Step 1: Action 3—Determine the Suitability of the Probabilistic Fracture Mechanics Code for the Specific Application

Purpose: The purpose of this step is to determine whether a specific PFM code is suitable for the application of interest and to identify any potential limitations of the code with regard to the application.

Description: The SQA process should follow the graded approach suggested in Section 2.2.4. It is intended to provide assurance that the software was developed in a deliberate and controlled manner, such that every aspect of the software is known and understood. Furthermore, the SQA process ensures source and version control, so as to prevent inadvertent changes to the software that could have unintended consequences on the software predictions. For nuclear regulatory applications, Title 10 of the *Code of Regulations*, Part 50, “Domestic licensing of production and utilization facilities,” (Reference 3-1) Appendix B, “Quality Assurance Criteria for Nuclear Power Plants and Fuel Reprocessing Plants,” requires that the applicants have an approved QA process in place.

The V&V process is intended to provide the critical evidence for the credibility of a code and a set of analysis tools, and it is composed of two primary activities, known as verification and validation. In general, verification seeks to determine whether a given mathematical model has been solved correctly within the analysis framework. This process has two components, referred to as code verification and solution verification. Code verification specifically focuses on the implementation of software to solve a given set of governing equations (i.e., the mathematical model). Solution verification focuses on approximations to the governing equations that are needed in order to solve them on a computer. These approximations may be made in space, time, or stochastic dimensions. Solution verification has the goal of quantifying the error incurred by these approximations and determining that these effects converge toward zero as resolution is increased (e.g., time steps are reduced or spatial approximations are refined).

Validation seeks to determine whether a chosen mathematical model is an accurate description of reality. Traditional validation involves comparing outcomes of a simulation to experimental data taken from a representative real-world scenario to determine the accuracy of the overall model representation. An alternative validation approach in the absence of experimental data includes benchmarking the software with comparable software that has been verified and validated previously. The model fidelity has several components, including the physics-based models, the material models, and the geometric description of a system of interest.

Researchers have detailed these elements in a variety of references (e.g., References 3-2, 3-3, 3-4) and a set of standards produced by ASME (References 3-5 and 3-6). While nominally discipline specific, the methods described in these guides and the references therein are very general in nature and provide a good basis for foundational V&V activities in support of model credibility.

Section 2.2.4 provides SQA and V&V documentation guidance for all PFM analysis codes used in analysis. Individual analyses will apply the code in a specific manner; an important aspect of the credibility of the overall analysis is the degree of confidence in the code for the intended application. The intent of this action is to identify and resolve any important gaps in the code capabilities for the intended application.

Code capabilities. Code capabilities refer to all scenarios for which a code has been through an appropriate set of V&V activities. Examples of code capabilities include (1) the range of inputs that were included in verification tests and validation test data, (2) the set of material models or geometries that have an established pedigree, (3) the underlying physics models and the assumptions underlying their range of applicability, and (4) the numerical approximation schemes (e.g., grid size, spatial and temporal resolution) with appropriate solution verification.

Examples of questions to consider with regard to code capabilities include the following:

- How well does the chosen model represent the application?
- Is there a rationale for defining certain model assumptions as conservative?

- Is the coding for the physics-based models available for review?
- Are the physics-based models well understood and established?
- Are code limitations that may impact the regulatory question/issue identified?
- Is mathematical justification for the model representation of the physics well established?
- Are limitations of the methodology identified with respect to interpolation or extrapolation?

Analysis features and code capabilities. An important first step in an analysis is to compare features of the intended application to the code capabilities to determine whether the code is suitable for the application. This process identifies any features that are incompatible with the code capabilities. Further, it identifies any features of the analysis for which the code does not have sufficient V&V evidence. As an example, if a PFM code was validated and calibrated for a specific range of weld residual stresses, then considering the implications of applying the code outside of this input range is critical for interpreting the model credibility.

The following are some of the key considerations for code capability:

- Does the range of inputs for which the code has been calibrated and validated include the range of inputs that are required for the specific application? Have the physics models been changed for the specific application? Are the numerical approximations sufficient for the application?
- If application-specific changes have been made, is the phenomenological behavior of the code expected to be similar for this application relative to the applications for which validation occurred (i.e., are the same physics models still relevant and adequate?)
- Are there any additional test data to support the applicability of the code for the current application?

Addressing code limitations. Potential limitations of the code for the application can be addressed in two ways:

(1) Risk can be mitigated by collecting additional information to improve the vetting of the code in the identified risk areas.

(2) When it is not possible to collect additional information, justification for the credibility of the code capabilities for the application can often be based on appropriate engineering arguments. When sufficient evidence cannot be collected to address certain gaps, understanding the associated risk to the analysis credibility is critical to interpreting the final results.

3.1.4 Step 1: Action 4—Identify Key Elements of the Problem that Impact Analysis Choices

Purpose: The purpose of this step is to identify key elements of the PFM application that will determine how to conduct the analysis. Simplifying assumptions and approximations may be necessary based on the complexity of the problem or, conversely, may be justified because the problem at hand is inherently not complex.

Description: Specific aspects of the application drive the methods used in a PFM analysis. In an ideal situation, simple analysis techniques can be applied. More sophisticated analysis methodologies are useful when the following is true:

- *The model is computationally expensive.* When models are computationally inexpensive to run, sampling uncertainty due to limited model realizations is a secondary issue, because the sample size can often be made arbitrarily large such that sampling uncertainty is negligible. On the other hand, computationally expensive models require more forethought about how to select model realizations and how to design model sampling schemes to achieve converged results.

- *The QoI is a rare event probability*. Estimating rare event likelihoods typically requires more realizations, more sophisticated sampling schemes, or both. Rare event probabilities (e.g., adverse event or failure probabilities) are defined as probabilities that are close enough to zero that the number of samples needed to estimate the probability is large with respect to computational budget. For example, to estimate a $1x10^{-6}$ probability using simple Monte Carlo sampling (Section 4.3.1), more than $1x10^{6}$ model realizations are required.

- *There are many model inputs.* When the number of model inputs is large, then there are more input uncertainties to characterize. Also, identifying important/sensitive model inputs is more difficult because there are more candidate inputs.

- *Separation of aleatory and epistemic uncertainty is maintained*. Uncertainty can arise from different causes; the most commonly considered types of uncertainty are aleatory and epistemic uncertainty (Section 4.1.1). For a specific adverse event, the quantification of aleatory uncertainties targets the question, “How likely is the event to happen?” while the quantification of the epistemic uncertainties targets the question, “How confident are we in this estimate of the event likelihood?” PFM analyses can treat aleatory and epistemic uncertainties separately to distinguish the frequency of event occurrence from the confidence in the frequency estimate. Separating uncertainty introduces additional complexity and computational burden into an analysis, because of the double-looping algorithm for separation described in Section 4.1.1. Section 3.2.1 and Section 4.1.1 provide more details about classifying and separating aleatory and epistemic uncertainty.

For each of these attributes, it is generally more challenging to conduct SAs to identify important inputs (Section 3.3.3) and design sampling algorithms to achieve statistical model convergence (Sections 3.3.1 and 3.3.2). The points at which these elements can impact the analysis decisions are highlighted throughout the PFM analysis process.

3.2 Step 2: Model Input Uncertainty Characterization

The second step in a PFM analysis is characterizing input uncertainty. This step involves two key actions:

(1) Identify uncertain model inputs.
(2) Specify probability distributions on uncertain inputs.

The end goal of this step is to determine probability distributions to represent input uncertainty.

3.2.1 Step 2: Action 1—Identify Uncertain Model Inputs

Purpose: The purpose of this step is to determine which model inputs are treated with uncertainty and, if relevant, the type of uncertainty (aleatory or epistemic) for each input.

Description: This action includes classifying deterministic versus uncertain inputs and classifying aleatory versus epistemic uncertain inputs (if relevant).

Deterministic versus uncertain inputs. Inputs to a PFM analysis can be represented in two ways:

- Deterministic inputs take on a single value.
- Uncertain inputs can take on a range of potential values.

Deterministic inputs are fixed to a single value across all model realizations. Such inputs can be fixed for several reasons: (1) they have known physical values (e.g., a known yield strength of a material), (2) the chosen fixed value is determined to be a value of interest (e.g., a conservative value used for a specific reason or a value of relevance for sensitivity studies (see Section 3.4)), or (3) including uncertainty would not affect decisionmaking. Uncertain inputs determine the amount of variability in the model output, conditional on the values of the deterministic inputs. This uncertainty in model inputs is what distinguishes a purely deterministic analysis from a PFM analysis. If the QoI is a failure probability, this probability is determined based on the uncertainty in the model's uncertain inputs, conditional on the values of the deterministic inputs. Data, expert judgment, and SA (Section 3.3.3) inform whether an input is modeled as deterministic or uncertain.

Understanding the rationale for classifying inputs as deterministic or uncertain is important when interpreting the analysis results. If there is uncertainty as to whether an input is deterministic or uncertain, then modeling the input as uncertain is preferable.

Avoiding excessive conservatism in model inputs. Deterministic fracture mechanics models have historically relied on conservatisms; introducing conservatism into a PFM analysis makes the results difficult to interpret. Conservatisms in inputs may propagate to produce an unrealistically conservative output. For example, the probability that 10 independent variables all take values at or above their respective 90th percentile is 1×10^{-10}, or 1 chance in 10 billion. Hence, taking a conservative approach and setting each of these inputs to their 90th percentile in a deterministic model realization results in a highly unlikely output. Even setting a single input to a conservative value can substantively change the interpretation of the model results; if the model output is highly sensitive to this input, then subsequent modeling results will on average be conservative. Additionally, conservative assumptions in submodels may be anticonservative in full system models. For example, increases in leak rate may be considered conservative at a submodel level. However, when combined with leak rate detection, this conservatism could lead to the suppression of failures due to increased leak rate detection.

Understanding when and why conservative inputs are used is important to interpreting the final model results. The influence of conservative choices can be addressed using sensitivity studies. These sensitivity studies are especially important when specifying a best estimate or conservative value is difficult due to limited information.

The best estimate is defined as an approximation based on the best available information. Using a best estimate does not imply the chosen deterministic value or input distribution has no uncertainty.

Aleatory versus epistemic uncertain inputs. If an analysis maintains separation between aleatory and epistemic uncertainty, then uncertain inputs are classified as epistemic or aleatory. Section 4.1.1 provides more details on aleatory versus epistemic uncertainty. This classification is not necessarily straightforward, because the uncertainty type often depends on the context and granularity of the problem. As an example, in a conventional linear elastic fracture mechanics model, the uncertainty in the linear elastic fracture toughness (K_{Ic}) may be regarded as aleatory (irreducible or inherent). Conversely, in a micromechanics model that accounts for features such as grain size, inclusions, and dislocations (i.e., the factors that create the uncertainty in K_{Ic}), this uncertainty may be regarded as epistemic. Mixed situations (part aleatory, part epistemic) are also possible. The categorization of uncertainty is therefore not totally objective and may change depending on the context of the problem.

To interpret modeling results, it is important to understand how aleatory and epistemic uncertainty are defined in the context of the application and to understand the rationale for classifying inputs as epistemic or aleatory. If it is uncertain whether an important input is aleatory or epistemic, sensitivity studies (Section 3.4) can be conducted to determine the impact of changing the classification.

3.2.2 Step 2: Action 2—Specify Probability Distributions on Uncertain Inputs

Purpose: In PFM analyses, uncertainty in model inputs is represented through probability distributions. This uncertainty is propagated forward to the model outputs to estimate and quantify uncertainty in QoIs.

Description: This action includes considering attributes of input distribution specification, including the following:

- iterative nature of input distribution specification
- importance of analysis context in characterizing input uncertainty
- nonprobabilistic representations of input uncertainty
- expert judgment
- distribution specification methods
- bounding input distributions
- accounting for correlation in model inputs

Iterative nature of input distribution specification. A PFM analysis focuses on those inputs that have the most influence on the model output. These influential inputs are typically identified using SA (Section 3.3.3). If an input's uncertainty has little impact on the output uncertainty, a strong technical basis for the input distribution may not be necessary and a deterministic value could be used. If these results indicate a large impact, additional data, more refined statistical techniques, or further expert elicitation may be needed to further refine the input's probability distribution. In this way, the development of inputs for PFM analysis is an iterative process, and the distributions specified in this step may be iteratively refined in the analysis process.

Importance of analysis context in characterizing input uncertainty. The context of the analysis impacts the input uncertainty. Specific analyses will often have narrower uncertainty ranges than more general analyses. For example, if an analysis is specific to a certain pipe in a specific plant, then the geometry and other characteristics of the system are likely to be defined precisely and the uncertainty range may be relatively small. In contrast, for an analysis meant to represent a series of welds or generic configurations across the U.S. reactor fleet, the variability in geometry, operating conditions, materials, and possible flaw mitigation is likely to be larger.

Nonprobabilistic representations of input uncertainty. In PFM applications, it is common practice to represent input uncertainty by specifying probability distributions on the inputs. In some analyses, it may be appropriate to use other nonprobabilistic representations of uncertainty to characterize an unknown input. Specifically, for epistemic uncertainties, if the lack of knowledge is too great to specify a probability distribution on an input, then nonprobabilistic, interval-based bounding methods can be considered (References 3-2,3-7). Probabilistic representation of uncertainty is often sufficient in PFM applications; understanding the rationale for deviating from a fully probabilistic analysis is important to interpreting the analysis results.

Expert judgment. In PFM applications, relevant data needed to define input distributions are often sparse or unavailable. In these cases, literature and expert opinion can be leveraged. The NRC has provided specific guidance on expert elicitation, with applications to uncertain model inputs (Reference 3-9).

Distribution specification methods. Proper selection of a probability distribution for an uncertain input requires detailed knowledge of the available data as well as qualitative judgments. Expert judgment and the amount and pedigree of the data, as well as the importance of the particular input on the analysis results, are relevant considerations when justifying a distribution. Distribution specification can be highly subjective and uncertain when data are limited.

Inputs with substantial uncertainty about the probability distribution or uncertainty representation may be candidates for future sensitivity studies to understand the impact of the chosen distribution on analysis results.

Section 4.2.1 contains more information about fitting probability distributions to data.

Bounding input distributions. Input bounds are the upper and lower truncation points defining the physical range of the input. In PFM applications, uncertain inputs are often bounded within a known range. Probability distributions that place nonzero likelihood only within this range can be used to prevent the sampling algorithm from selecting input values that are undesirable, nonphysical, or both. Section 4.2.1 discusses methods for specifying bounded probability distributions.

Inputs with substantial uncertainty about the ranges may be candidates for future sensitivity studies.

Accounting for correlation in model inputs. In a PFM analysis, some uncertain input variables may be statistically dependent (i.e., correlated). Accounting for the dependence between inputs often ensures a physically possible input set (i.e., ensures that physical laws are preserved).

Section 4.2.2 contains more information on dependent inputs.

3.3 <u>Step 3: Estimation of Quantity of Interest and Associated Uncertainty</u>

The third step of a PFM analysis is propagating input uncertainty established in Step 2 through the model to provide a converged estimate of the QoI and characterize its uncertainty. The QoI uncertainty characterized in this step includes uncertainty induced by input uncertainty and sampling uncertainty.

The goal is to estimate the QoI and its uncertainty with sufficient sampling precision (i.e., achieve converged model results). This step includes four key actions:

(1) Select a sampling scheme for sampling uncertain model inputs.
(2) Assess sampling uncertainty.
(3) Conduct SA to determine input uncertainty importance.
(4) Conduct output uncertainty analysis.

These actions are iterative. First, a sampling scheme is selected and used to estimate the QoI. The second action uses the sampling scheme to estimate the sampling uncertainty in the QoI and determines whether the estimate has converged. The third action uses SAs to identify the input uncertainties that drive the problem. SAs help to better understand the input-output relationship. Results from the second and third actions can be used as a basis to update a sampling scheme to improve convergence. Once a converged solution is found, the fourth action provides a final estimate of the QoI and associated uncertainty.

3.3.1 Step 3: Action 1—Select a Sampling Scheme for Sampling Uncertain Model Inputs

Purpose: The purpose of this step is to select a method for propagating uncertainty in the model inputs through the model to estimate the QoI and the associated uncertainty.

Description: This action involves selecting a sampling scheme and using it to estimate the QoI and its uncertainty. While many PFM analyses will rely on Monte Carlo sampling methods to estimate a QoI, nonsampling based methods are also available and may be appropriate in some applications.

Nonsampling approaches. Reliability methods, such as the first-order reliability methods (FORM) and second-order reliability methods (SORM), use gradient-based methods to calculate failure probabilities (Section 4.3.4). These methods work best when the model output is sufficiently smooth and differentiable. In such conditions, they can estimate low probabilities (i.e., $1x10^{-4}$ probability or less) with greater accuracy and fewer realizations than with Monte Carlo sampling methods. These derivative methods are limited by the fact that calculating second-order derivatives can quickly become impracticable as the number of uncertain inputs increases beyond 15 or 20.

Sampling approaches. PFM analyses often use Monte Carlo methods to propagate input uncertainty through the model. Selecting a sampling scheme includes specifying the following:

- sampling method
- sample size
- random seed
- method for sampling aleatory and epistemic uncertainties, if relevant

Sampling methods. Inputs can be sampled in different ways. The simplest form of Monte Carlo sampling is simple random sampling (SRS), described in Section 4.3.1. SRS is easy to implement but is often not the most statistically efficient method. Relative to SRS, other sampling schemes can produce more precise estimates of a QoI with the same number of model realizations. When models are computationally expensive or the QoI is a rare probability, or both, more targeted sampling methods can be implemented to decrease the number of realizations required for model convergence. Examples of targeted sampling methods include Latin hypercube sampling (LHS) (Section 4.3.2), importance sampling (Section 4.3.3), and adaptive sampling.

Importance sampling (Section 4.3.3) is a common sampling method for oversampling important regions of the input space to reduce the sampling uncertainty of QoI estimates. When estimating rare probabilities, the regions of the input space where failures are more likely are oversampled to estimate the probability with less sampling uncertainty (making importance sampling particularly relevant for PFM applications targeting adverse event likelihoods). To implement importance sampling, the analyst selects variables on which the technique is to be applied and their respective importance distributions. One general strategy is to first find the failure regions that contribute to the probability of the rare event and construct the importance distributions based on this information. SAs (Section 3.3.3) and subject matter expertise on important inputs can inform this process. The choice of importance distributions is paramount, since poor choices can lead to higher variance estimates with higher sampling uncertainty. Inefficiency in importance sampling often occurs in high-dimensional problems where many variables are importance sampled (Reference 3-10).

Sample size. The sample size is the number of realizations at different input settings (i.e., the number of sets of inputs that are propagated through the model). There is a natural relationship between the computational burden of the model, the sample size, and the sampling scheme. Specifically, computationally inexpensive models can be run many times, resulting in large sample sizes. In such cases, simple sampling schemes such as SRS are likely sufficient. If the model is computationally expensive, sample sizes will be lower and more efficient sampling schemes are required. Relatedly, if the QoI is a probability, the sample scheme and sample size are related to the magnitude of the probabilities. As the probability gets closer to 0 or 1, more samples or more efficient sampling schemes, or both, are required.

Random seed. Sampling-based approaches rely on random number generators to select the random sample. Random seeds can be selected for a random number generator to ensure that the same random sample is selected each time a set of model realizations is run so that exact results can be reproduced.

Separation of aleatory and epistemic uncertainty. If the analysis maintains separation of aleatory and epistemic uncertainty, then the input uncertainties are typically sampled using a double-loop method, described in Section 4.1.1. This method first samples epistemic inputs. Then, for each set of epistemic inputs, aleatory inputs are sampled numerous times to obtain a distribution of model outputs over aleatory uncertainty. The double-loop structure is computationally expensive, because the QoI is estimated for each set of epistemic samples. Surrogate modeling (Section 4.3.10) is often used to increase the computational efficiency of the double-loop method, relying on a computationally inexpensive statistical model approximation to post hoc separate aleatory and epistemic uncertainty (as described in Section 4.1.1).

3.3.2 Step 3: Action 2—Assess Sampling Uncertainty: Statistical Convergence Analysis

Purpose: The purpose of this step is to assess the statistical convergence of QoI estimates from model outputs given a sampling scheme.

Description: PFM analyses are based on a finite number of realizations. Since the model cannot be run at all points in the input space, sampling uncertainty is associated with estimating a QoI. Quantifying the sampling uncertainty of QoI estimates is important to determine whether the analysis conclusions might change with an improved sampling scheme. Methods to assess sampling uncertainty convergence include the following:

- assessing stability of an estimate as the sample size increases
- calculating statistical sampling uncertainty metrics
- comparing replicates and assessing variation in the QoI estimates
- using surrogate modeling to estimate sampling uncertainty
- updating the sampling scheme

Section 4.3.5 discusses these methods in more detail.

Assessing stability in an estimate as the sample size increases. For a given sampling scheme, the sample size can be increased iteratively until QoI estimate is sufficiently stable, suggesting statistical convergence. For example, with an SRS sampling scheme, more input samples can be selected to increase the sample size. Augmented LHS designs can be used to add input samples to an initial LHS design. Stability in the QoI can again be measured as the sample size increases. The major advantages of this approach are that it can be applied to any sampling scheme and that it does not require multiple independent model realizations. However, the approach does not provide a direct measure of sampling uncertainty in the QoI estimate and can be rather computationally expensive.

Calculating statistical sampling uncertainty metrics. Statistical sampling uncertainty metrics quantify the sampling uncertainty in the QoI estimate using statistical sampling theory. Methods for calculating statistical sampling uncertainty metrics are specific to the sampling scheme. Using an SRS sampling scheme, the standard deviation, coefficient of variation (CV), or confidence interval for a QoI can be calculated directly from the sample (Section 4.3.6). If the QoI for a PFM analysis is a rare probability and zero events are observed, then an upper bound on the probability can be calculated using statistical metrics under an SRS scheme. However, SRS is not the most efficient sampling scheme when the QoI is a rare probability.

Under sampling schemes other than SRS, statistical resampling methods, such as bootstrapping, can be used to calculate statistical sampling uncertainty metrics (Section 4.3.7). Resampling methods are easy to implement but can be more computationally expensive; further, resampling methods can produce inaccurate estimates when the QoI is a rare probability.

LHS schemes do not offer a simple analytic form for an unbiased estimate of sampling uncertainties from a single sample (References 3-11, 3-12). Furthermore, bootstrapping an LHS is not possible. Since LHS is more efficient (Reference 3-11), SRS uncertainty metrics applied to an LHS scheme will be conservative.

Comparing replicates and assessing variation in the QoI estimates. Another method to assess QoI convergence is to run the sampling scheme several independent times with unique random number seeds. The QoI is estimated for each independent realization, and the variation across the realizations is measured. Example metrics to assess convergence based on these realizations include the standard deviation of the QoI across realizations, the CV (ratio of the standard deviation to the mean), or a confidence interval on the QoI. These metrics can be compared to the desired level of convergence for the application. The major advantages of this approach are that it can be applied to any sampling scheme and that it gives a direct measurement of estimate variability; however, the approach is computationally expensive. In general, the number of replicates is selected to be large enough that the conclusion would not change significantly if more replicates were provided.

Using surrogate modeling to estimate sampling uncertainty. When the model is computationally expensive to run and only a small number of input samples can be propagated through the model, surrogate models (Section 4.3.10) can be used to provide a computationally efficient alternative to the full model. A surrogate model is a statistical approximation to the full, computationally expensive model and is estimated from a set of model realizations. The sampling uncertainty in the surrogate model can be propagated to sampling uncertainty estimates for the QoI.

Updating the Sampling Scheme. If the selected sampling scheme does not provide converged results, then this scheme can be updated by increasing the sample size, changing the sampling method, or both.

3.3.3 Step 3: Action 3—Conduct Sensitivity Analyses to Determine Input Uncertainty Importance

Purpose: SAs help identify problem drivers, defined as uncertain model inputs that explain substantial uncertainty in the model output. Understanding problem drivers allows the analyst to do the following:

- Confirm that the model is behaving as expected.
- Identify inputs whose uncertainty distribution is itself uncertain and that may need refinement before final estimation of the QoI.
- Identify assumptions that are uncertain and thus may be candidates for sensitivity studies (Step 4).
- Improve the accuracy of the output uncertainty analysis by reducing the dimension of the input space and identifying important inputs that can be used in more targeted sampling methods such as importance sampling.

SA plays a critical role in improving output uncertainty analysis. A common goal of a PFM analysis is to accurately estimate a QoI along with its associated uncertainty. By informing the final sampling scheme, SAs can improve QoI estimation. For example, SA can identify inputs with a large impact on the model output; these inputs may be candidates for importance sampling (Section 4.3.3) to increase the precision of QoI estimates. This action is closely tied to Step 3, Action 1, which provides more detail on selecting an appropriate sampling scheme for the estimation of a QoI and its uncertainty.

Description: In broad terms, SA focuses on identifying how the input uncertainties contribute to the uncertainty in the outputs of interest. References 3-13, 3-14, 3-15, 3-16, 3-17, 3-18, and 3-19 are some of the sources that describe SA techniques and examples. The discussion below addresses the following:

- the types of SA
- forward propagation of uncertainty for SA

- the stages of SA
- modeling nonlinearities and interactions in SA
- SA for submodels
- uncertainties in SA

Types of SA. There are two general types of SA:

(1) *Global SA* is the process of decomposing variance in the model output according to the model inputs (see Section 4.3.8).

(2) *Local SA* is the process of determining how changes to uncertain inputs affect outputs with respect to a reference point in the input domain (see Section 4.3.9).

Forward propagation of uncertainty for SA. SA is performed after an initial set of uncertain inputs has been propagated through the model, resulting in a distribution of model outputs. SA is often conducted on an initial set of model realizations, with uncertain inputs sampled using a standard Monte Carlo-based sampling scheme with broad coverage of the input space, such that model input-output relationships can be discerned from the sample. The number of model realizations needed depends on the goals of the SA and the computational burden of the model. For example, if the goal of global SA is to understand how inputs vary with the output to select the number of model realizations, the analyst can consider the complexity of the input-output relationship and the number of uncertain model inputs. Local SA typically requires fewer model realizations. After model results are obtained from forward propagation of uncertainty, the analyst can proceed with the two stages of SA described below.

Stages of SA. Typically, SAs have two stages:

(1) *Exploratory data analysis* involves graphically exploring input-output relationships using scatter plots and calculating local SA metrics, as needed. The SA results can present scatter plots for important inputs. Reference 3-20 describes formal procedures for the analysis of scatterplots.

(2) *Global sensitivity metrics estimation* involves the estimation of the proportion of variance in the model output explained by each model input (first-order sensitivity index) and its interactions with other inputs (total-order sensitivity index).

In practice, SA is an iterative process, and these two stages may repeat multiple times. For example, given a large number of inputs and complexities in the input-output relationships, selecting the correct visualizations and interpreting them can be difficult. Estimation of the global sensitivity metrics in the second stage can help to identify the important inputs to visualize. These visualizations can inform results of the global SA.

Model nonlinearities and interactions in SA. PFM applications often involve systems of linked models with complex relationships. SAs allow for the identification and quantification of the input-output relationship, including nonlinearities in the input-output relationship and interactions between model inputs.

Global SA is a commonly used tool for summarizing input importance in PFM studies and identifying the effects of nonlinearities and interactions. Local SAs identify sensitivities within a small neighborhood around a point of interest and therefore do not identify nonlinearities and interactions; local SA is informative if the goal is to understand local variations (see

Section 4.3.9 for more detail). Global and local SA are often used together in the same PFM analysis at different iterations of the SA.

SA for submodels. Since PFM applications often involve systems of linked models, it may be appropriate to conduct SA on specific submodels (in addition to the full PFM model) as some dominant submodels may hide the results and impacts of other submodels. For example, to investigate the impact of active degradation mechanisms on the probability of leakage or rupture, it may be appropriate for the analysis to exclude fatigue damage. As another example, it may be prudent to identify inputs impacting crack growth before a full SA determining important inputs for rupture.

Uncertainties in SA. Understanding aspects of the model and input uncertainty characterization informs how to conduct SA, as shown in the following examples:

- Model approximations for computationally expensive models. Estimating sensitivity metrics is computationally expensive, often requiring many model realizations. As a solution, model approximations or surrogates (Section 4.3.10) are often used in SA as a computationally practical approximation to the full model. Sufficiently flexible model surrogates allow for nonlinearities and interactions between inputs. If the model approximation contains substantial uncertainty, then multiple different model approximation methods can be compared to assess robustness of the SA results to the model approximation method.

- High-dimensional inputs. Building an accurate model approximation requires more model realizations when the input space is high dimensional. Without enough realizations, true input-output relationships may not be identified.

- Continuous versus binary or discrete outputs for SA. Binary or discrete outputs (such as failure events) inherently contain less statistical information than continuous outputs. More realizations will be needed to identify important model inputs impacting a binary indicator variable than for a continuous model output. An alternative is to identify continuous responses associated with the binary event for SAs, insofar as there is a clear, justifiable connection between the binary event and the continuous variable. For example, instead of conducting SA on the binary indicator for rupture, the analyst could use crack length as the output for SA.

- Separation of aleatory and epistemic uncertainty. If the separation of uncertainty types is maintained, SA is conducted for both uncertain aleatory and epistemic inputs. The SA can be run over all uncertainties to determine which inputs have the largest impact on the outputs of interest. Additional SAs can be conducted for aleatory and epistemic inputs separately to identify the impacts of irreducible and reducible uncertainties, respectively.

3.3.4 Step 3: Action 4—Conduct Output Uncertainty Analysis

Purpose: The purpose of this step is to provide a final estimate, with associated uncertainty, of the QoI and to visualize results.

Description: A summary of the QoI results may include the following:

- a best estimate of the QoI

- an estimate of uncertainty in the QoI
- a graphical display of the QoI estimate and uncertainty

Best estimate of a QoI. The definition of the best estimate of a QoI will depend on the application. When the QoI is uncertain, the best estimate is often quantified using either the mean or median of the QoI distribution. The mean is the arithmetic average over the QoI distribution, and the median is the 50th percentile of this distribution.

Estimate of uncertainty in the QoI. When estimating and visualizing uncertainty in a QoI estimate, it is critical to be clear about the type of uncertainty being summarized. QoI uncertainty can refer to different types of uncertainty, depending on the relationship between the model output and the QoI. Example types of uncertainty include the following:

- <u>Input uncertainty</u>. If the QoI is a model output, QoI uncertainty may refer to uncertainty in the QoI due to uncertain inputs. A best estimate of the QoI is the mean or median of the QoI over the input space, and the uncertainty in the QoI refers to the distribution of the QoI over uncertain inputs.

- <u>Sampling uncertainty (also called aleatory uncertainty)</u>. QoI uncertainty may also arise due to a limited number of model realizations resulting in uncertain QoI estimates. When convergence analyses (Section 3.3.2) suggest sampling uncertainty is negligible, then visualizing sampling uncertainty will not be necessary. If the sampling uncertainty is not sufficiently small based on convergence analysis results, then this sampling uncertainty can be measured and presented as a source of QoI uncertainty.

- <u>Epistemic (lack of knowledge) uncertainty</u>. When aleatory and epistemic uncertainty are separated, the QoI is typically calculated for each epistemic sample. Epistemic uncertainty in the QoI measures how the QoI varies due to knowledge uncertainty. A best estimate of the QoI is the mean or median QoI estimate over all epistemic samples.

- <u>Uncertainty in the QoI results in a distribution of QoI estimates</u>. This uncertainty can be summarized using percentiles of the uncertainty distribution; measures such as variance and standard deviation can also provide useful summaries of QoI uncertainty.

If the QoI is a failure probability calculated from uncertain model outputs, then the QoI already incorporates uncertainty in the model inputs. In this case, if aleatory and epistemic uncertainty are not separated and sampling uncertainty is negligible (i.e., a high degree of statistical convergence has been achieved), then there may be no need to present a measure of uncertainty about the failure probability estimate.

Graphical display of the QoI estimate and uncertainty. Graphical displays of the best estimate and uncertainty in the QoI can be used to communicate the results of an uncertainty analysis. The form of the graphical display will depend on the types of uncertainty being visualized and whether the QoI is a function of time or a single scalar. The best approach to visualizing results is application specific. Section 4.3.11 provides more details on output uncertainty analysis.

3.4 <u>Step 4: Sensitivity Studies to Assess the Credibility of Modeling Assumptions</u>

The fourth step in a PFM analysis is conducting sensitivity studies, defined as additional analyses conducted under different, yet plausible, assumptions. The purpose of sensitivity studies is to challenge uncertain analysis assumptions that could substantively change the analysis results. Sensitivity studies involve two key actions:

(1) Determine a set of sensitivity studies.
(2) Conduct sensitivity studies and present results.

3.4.1 Step 4: Action 1—Determine a Set of Sensitivity Studies

Purpose: The purpose of this action is to identify important assumptions that merit further scrutiny to understand what might happen if these assumptions were changed. For example, in the study of a plant, the distribution of a specific input could have been calibrated using information from a global set of similar but different plants. This calibration raises the question of what might be different about the distribution for the individual plant and how that would change the conclusions of the analysis.

Description: Given the complexity of PFM analyses, it is not possible to enumerate all plausible changes in the assumptions. Instead, to evaluate whether a sensitivity study is needed for a specific assumption, two criteria are evaluated:

(1) Plausible alternate assumptions can be identified.
(2) Changes to the assumption in question can substantively impact the calculated QoI.

The specific number of sensitivity studies will depend on the application, but the goal is to conduct enough studies such that there is a sufficiently low chance that the results of the analysis depend heavily on unverifiable or uncertain assumptions.

Uncertain analysis assumptions can often be classified as either modeling assumptions or input parameter specification assumptions. Modeling assumptions include any assumptions in the computational modeling framework, while input parameter specification assumptions refer to any assumptions made when specifying the values of the input parameters to the PFM model. Common types of sensitivity studies include considering changes in the results if the following occurs:

- A plausible alternative model is used.
- A different probability distribution for an uncertain input (or several uncertain inputs) is used.
- The categorization of an input as aleatory or epistemic is changed.

Reference 3-21 provides guidance on selecting sensitivity studies, which this report reviews in Section 4.4.

3.4.2 Step 4: Action 2—Conduct Sensitivity Studies and Present Results

Purpose: The purpose of this action is to perform the sensitivity studies.

Description: Sensitivity studies can take on many different forms, and there is no prescriptive method for conducting sensitivity studies. However, they will all include some common elements:

- a reference realization (or baseline case) with a documentation of the QoI
- one or several modified realizations illustrating the concept that needs to be represented
- a comparison between the reference realization and the modified realization(s)
- a comparison criterion to decide whether the change is significant
- a conclusion, including potential consequences

3.5 Step 5: Draw Conclusions from Analysis Results

The fifth step in a PFM analysis is to draw conclusions using the results of Steps 1–4. This step includes two key actions:

(1) Interpret analysis results.
(2) Iterate on the analysis process to refine model results.

3.5.1 Step 5: Action 1—Interpret Analysis Results

Purpose. The purpose of this action is to synthesize the information gathered in Steps 1–4 and draw conclusions from this information.

Description. In an ideal situation, PFM analysis results can be compared directly to acceptance criteria to make a regulatory decision. In practice, determinations about whether acceptance criteria are met are typically not made based on a single PFM calculation or analysis but rather based on a set of analyses that are compiled into an overall evidence package. Information about the analysis results, scope, and limitations must be considered when drawing final conclusions, considering all elements of the PFM analyses described above in Steps 1–4. Subsequently, drawing final conclusions based on the analysis requires substantial expert judgment to synthesize all information together to make actionable guidelines.

3.5.2 Step 5: Action 2—Iterate on the Analysis Process to Refine Model Results

Purpose. The purpose of this action is to determine whether additional analyses are required to draw informative conclusions from the modeling.

Description. If analysis results are inconclusive concerning whether the acceptance criteria are met, then the analyst can consider additional refinements to the analysis to provide the required additional information. For example, the analyst can consider the following:

- changing or clarifying aspects of the PFM code (Section 3.1.3)
- refining the input uncertainty distributions (Section 3.2.2)
- choosing a different sampling scheme or increasing the number of model realizations (Section 3.3.1)
- adding more sensitivity studies to address existing limitations (Section 3.3.3)

PFM analyses are typically iterative in nature, such that initial modeling results inform future analyses. The iterative process continues until the analyst has sufficient information to draw clear conclusions about whether the acceptance criteria are met for the application.

3.6 References

3-1. *U.S. Code of Federal Regulations*, "Domestic licensing of production and utilization facilities," Part 50, Chapter I, Title 10, "Energy."

3-2. Oberkampf, W.L., Pilch, M., and Trucano, T.G., "Predictive Capability Maturity Model for Computational Modeling and Simulation," SAND2007-5948, Sandia National Laboratories, 2007.

3-3. Oberkampf, W.L., and Roy, C.J., *Verification and Validation in Scientific Computing*, Cambridge University Press, United Kingdom, 2010.

3-4. Roache, P.J., Verification and Validation in Computational Science and Engineering, Hermosa Publishers, Albuquerque, NM, 1998.

3-5. American Society of Mechanical Engineers, "Guide for Verification and Validation in Computational Solid Mechanics," ASME V&V 10, 2006.

3-6. American Society of Mechanical Engineers, "Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer," ASME V&V 20, 2009.

3-7. Helton, J.C., Johnson, J.D., and Oberkampf, W.L., "An exploration of alternative approaches to the representation of uncertainty in model predictions," *Reliability Engineering and System Safety*, 85, 39–71, 2004.

3-8. Helton, J.C., Johnson, J.D., Oberkampf, W.L., and Sallaberry, C.J., "Representation of analysis results involving aleatory and epistemic uncertainty," *International Journal of General Systems*, 39, 605–646, 2010.

3-9. Xing, J., and Morrow, S., "White Paper: Practical Insights and Lessons Learned on Implementing Expert Elicitation," U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, 2016.

3-10. Au, S.K., and Beck, J.L., "Important sampling in high dimensions," *Structural Safety*, 25(2), 139–163, 2003.

3-11. Stein, M., "Large Sample Properties using Latin Hypercube Sampling," *Technometrics*, 239–245, 1987.

3-12. Helton, J.C., and Davis, F.J., "Latin Hypercube Sampling and the propagation of uncertainty in analyses of computer systems," *Reliability Engineering and System Safety*, 23–69, 2003.

3-13. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., and Tarantola, S., *Global Sensitivity Analysis. The Primer*, John Wiley & Sons, Ltd., 2008.

3-14. Helton, J.C., "Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal," *Reliability Engineering and System Safety*, 42, 327–367, 1993.

3-15. Helton, J.C., Johnson, J.D., Sallaberry, C.J., and Storlie, C.B., “Survey of sampling-based methods for uncertainty and sensitivity analysis,” *Reliability Engineering and System Safety*, 91, 1175–1209, 2006.

3-16. Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F., “Sensitivity analysis for chemical models,” *Chemical Reviews*, 105, 2811–2828, 2005.

3-17. Storlie, C.B., Swiler, L.P., Helton, J.C., and Sallaberry, C.J., “Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models,” *Reliability Engineering and System Safety*, 94, 1735–1763, 2009.

3-18. Wei, P., Lu, Z., and Song, S., “Variable importance analysis: A comprehensive review,” *Reliability Engineering and System Safety*, 142, 399–432, 2015.

3-19. Borgonovo, E., and Plischke, E., “Sensitivity analysis: A review of recent advances,” *European Journal of Operational Research*, 248, 869–887, 2016.

3-20. Kleignen, J.P.C., and Helton, J.C., “Statistical analyses of scatterplots to identify important factors in large-scale simulations,” “1: Review and comparison of techniques,” *Reliability Engineering and System Safety*, 65, 147–185, 1999.

3-21. EricksonKirk, M.T., Dickson, T., Mintz, T., and Simonen, F., “Sensitivity Studies of the Probabilistic Fracture Mechanics Model Used in FAVOR,” NUREG-1808, U.S. Nuclear Regulatory Commission, 2010..

4 USEFUL METHODS FOR ESTABLISHING CONFIDENCE IN PROBABILISTIC FRACTURE MECHANICS ANALYSIS

This section details a concise review of analysis methodologies, including notional examples for context that are linked directly to an action introduced in Section 3. While this is not a comprehensive list of acceptable methodologies, this section can be used by applicants who are seeking explicit guidance on the theoretical underpinnings of the processes that are used to establish the credibility of a PFM analysis. For example, Section 4.1 provides technical detail that could be used to develop the technical basis for the action defined in Section 3.1. Each section introduces a concept/method and provides the following information about it:

- **What is it?**—Gives a high-level description of the concept/method.

- **How to use?**—Provides general details on how the concept/method is used, including specific steps or an algorithm where appropriate.

- **When/Why?**—Discusses the PFM context in which the concept/method is used and maps this use to the process described in Section 3.

- **Technical details—**Describes technical details and complexities that are important to the use/implementation/interpretation of the method in the PFM context.

- **References—**Lists references that provide further technical details.

4.1 Useful Methods for Translating Regulatory Requirements into an Analysis Plan

4.1.1 Separation of Aleatory and Epistemic Uncertainty

When constructing an analysis plan, one aspect to consider is the treatment of uncertainty—namely, will uncertainty be treated probabilistically, and, if so, will different types of uncertainty be distinguished? Separating types of uncertainty can be necessary when there is a need to quantify the uncertainty on a statistical QoI (a frequency or probability) or to separate inherent variability from lack-of-knowledge uncertainty. Such separation generally provides additional insights on the magnitude of the uncertainties and on whether they can be reduced. However, separation of uncertainties also comes with increased computational cost and analysis effort, and the decision to maintain the separation influences many steps of the subsequent analysis workflow. As a result, this tradeoff decision needs to be considered at an early stage of the analysis planning.

Specifically, the strategy for handling uncertainty may vary for different types of analysis questions. If the analysis objective is to compute a single best estimate event probability, it is likely sufficient to consider all sources of uncertainty together to arrive at this probability. However, this approach can obscure information. Separating types of uncertainty instead of considering all sources of uncertainty together can lead to more interpretable analysis. For example, rather than computing a single best estimate probability, the analyst may want to understand the confidence in the computed frequency of an event given the current state of knowledge. Some elements of this knowledge uncertainty may be reducible, potentially improving confidence in the frequency estimate and increasing the precision of the analysis results. As described in this section, these reducible sources of uncertainty (referred to as epistemic uncertainty) can be treated separately to maintain this information for the communication of results and decisionmaking about additional activities to conduct.

4.1.1.1 What Is It?

Two primary types of uncertainty sources are often considered in risk analysis (References 4-1, 4-2, 4-3, 4-4, 4-5, 4-6, 4-7, 4-8):

(1) **Aleatory uncertainty** is defined as “uncertainty based on the randomness of the nature of the events or phenomena that cannot be reduced by increasing the analyst’s knowledge of the systems being modeled” (Reference 4-9). Aleatory uncertainty represents the (perceived) randomness in the modeled system that cannot be reduced. Aleatory uncertainties reflect natural, intrinsic, or stochastic variability.

(2) **Epistemic uncertainty** is defined as “the uncertainty related to the lack of knowledge or confidence about the system or model and is also known as state-of-knowledge uncertainty” (Reference 4-9). Epistemic uncertainty represents the lack-of-knowledge uncertainty in the modeled system that can be reduced.

Historical PFM analyses of nuclear power plant structures either (1) do not distinguish between types of uncertainty (References 4-10, 4-11, 4-12, 4-13) or (2) treat the uncertainty as either aleatory or epistemic (References 4-14, 4-15).

4.1.1.2 How to Use?

If an analysis separates aleatory and epistemic uncertainty, then it requires additional effort to separate uncertainty types and iterate over epistemic samples in a “double-loop” sampling algorithm, as described below. This involves the following three steps:

(1) **Classify types of uncertainty for an application.** The first step in separating types of uncertainty is classifying input variables as aleatory or epistemic. The specific PFM application typically drives classification choices. If it is unclear how to classify an uncertainty, it may be worth considering a sensitivity study (Section 4.4) to understand the impact of the classification.

(2) **Determine how to represent uncertainty.** After classifying uncertainty types, the next step is to determine how to represent the different types of uncertainty. PFM analysis typically represents uncertainties using probability distributions, though other options are possible (see Section 4.1.1.4).

(3) **Propagate uncertainty while maintaining separation of types.** Given a model output, a QoI, and uncertainties related to the model input parameters, the next step is propagating both types of uncertainty. For sampling-based uncertainty propagation, separation of aleatory and epistemic uncertainty is maintained using a double-loop (i.e., nested loop) framework. The following steps can be applied to propagate input uncertainty:

   – Epistemic variables are sampled in an outer loop.

   – For each epistemic sample, aleatory variables are sampled in an inner loop.

- – QoIs are calculated for each epistemic sample (calculated over all aleatory samples), generating an epistemic distribution of QoIs.

The sections below provide more information about the double-loop procedure.

4.1.1.3 When/Why?

The risk community has made the distinction between the inherent risk (aleatory) and the uncertainty due to lack of knowledge (epistemic). The purpose of this distinction is to acknowledge there will always be a risk given a specific situation, and the consequences can lead to different interpretations in terms of decisionmaking. The analyst can choose whether to separate types of uncertainty. The decision to separate uncertainty types typically depends on several factors:

- **Computational feasibility.** In practice, maintaining separation during uncertainty propagation (for example, through Monte Carlo sampling) can be computationally challenging, due to the need to construct a “double-loop” sampling scheme. Section 4.1.1.4 contains more information and suggestions for efficiently implementing the double-loop scheme.

- **Conceptual interpretation of results.** The interpretation of the results of a PFM analysis changes depending on whether the separation of uncertainty types is maintained (Reference 0). Section 4.1.1.4 contains more information.

- **Strength of technical basis.** Ultimately, the separation of uncertainties can help to make a stronger, more comprehensive case and help the analyst understand what needs to be done to improve the accuracy of the answer.

4.1.1.4 Technical Details

Representing epistemic uncertainty. In many risk analysis applications, it can be difficult to specify probability distributions on epistemic uncertainties because, by definition, these uncertainties arise due to lack of knowledge. While probabilistic representation of epistemic uncertainty will be sufficient for most PFM applications, nonprobabilistic representations may be appropriate in certain instances (Reference 4-3, 4-16, 4-17). For example, sensitivity studies (Section 4.4) conducted at deterministic (i.e., fixed) values of the epistemic inputs can inform about a “worst case scenario.”

Computational burden of separating uncertainty. The double-loop framework for sampling typically requires a large sample size. For each epistemic sample, the aleatory sample is selected to be sufficiently large for the accurate estimation of the QoI (e.g., failure frequency). More sophisticated sampling schemes (Sections 4.3.2 and 4.3.3) may be needed to make double-looping computationally feasible. If the model is too computationally expensive to directly implement the double-loop sampling, there are two options: (1) do not separate uncertainty types, or (2) build a computationally efficient surrogate model to approximate the full model. Surrogate models are data-driven approximations of the physics model output across the input space, as discussed in Section 4.3.10. Surrogate models introduce additional uncertainty into the problem because the surrogate is itself a model approximation.

Interpretation of results. Maintaining separation of the two types of uncertainty facilitates making statements about confidence in the frequency of an event or the probability of frequency. Specifically, probability of frequency refers to an analysis that models aleatory and epistemic uncertainties probabilistically and separates them in presenting the results (References 4-2, 4-18). As an example, in a PFM analysis aiming to characterize the likelihood of an adverse event, aleatory probabilities represent the frequency of an adverse event (e.g., crack, rupture) given a set of epistemic inputs/assumptions. These frequencies will vary with the set of epistemic inputs/assumptions. This variation represents the epistemic uncertainty/confidence in the frequency.

If an analysis does not distinguish between aleatory and epistemic uncertainties, frequencies of an adverse event are computed over all uncertainties, and the analysis cannot quantify the impact of uncertainties that arise due to lack of knowledge. The implications of such a choice are explained with an example of the double-loop procedure below.

To illustrate the double-looping procedure, consider estimating the frequency of pipe rupture. The model output is a binary indicator taking the value 1 if the pipe ruptured and 0 otherwise. Each input is categorized as either epistemic or aleatory and is assigned a probability distribution to represent its uncertainty. Then, the model is run, each time with different inputs, using the double-loop algorithm to separate uncertainty:

(1) A set of epistemic variables is sampled randomly from the variables' probability distributions.

(2) Fixing this set, many samples (e.g., 1×10^4) of the aleatory variables are sampled randomly and the model is run, collecting the binary output for each realization.

(3) Steps 1 and 2 are repeated many times (e.g., 1×10^3). The separation of the results by epistemic variable is maintained.

Example results appear in Figure 4-1, which shows the proportion of the 1×10^4 aleatory samples that resulted in pipe rupture for the first 50 epistemic samples. For each epistemic realization, this proportion is the estimated frequency of pipe rupture, given the set of epistemic variables.

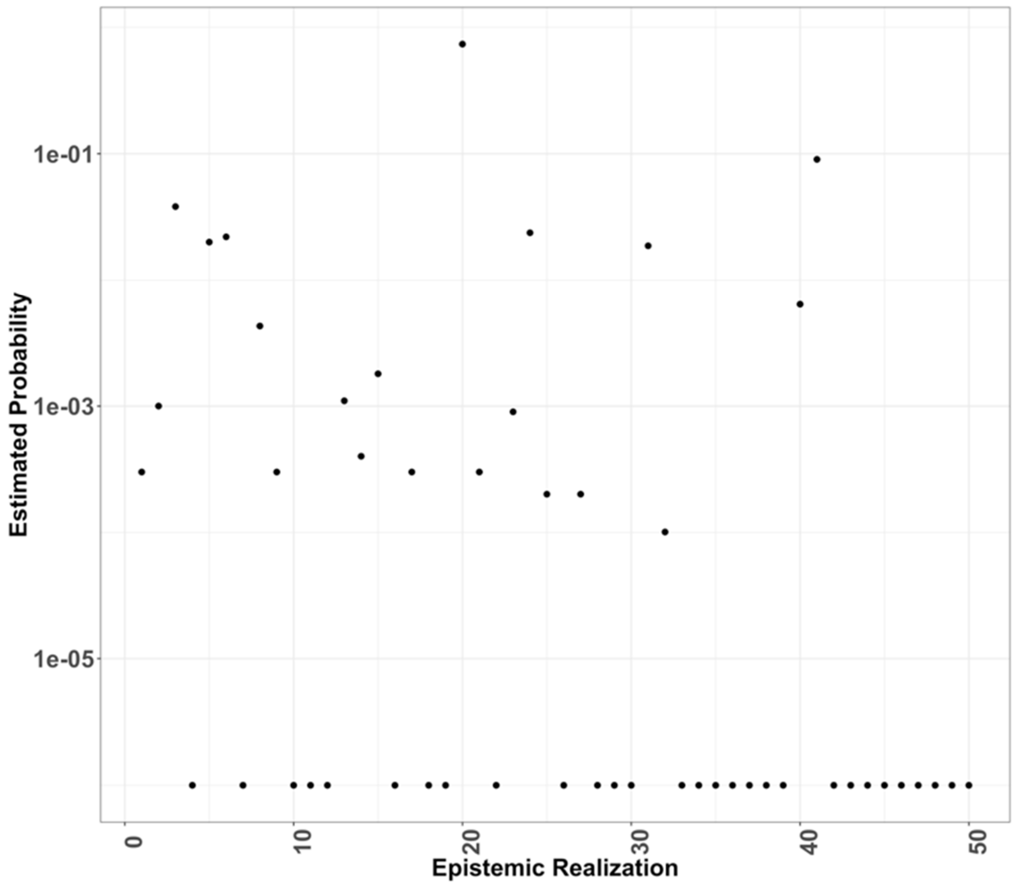

Figure 4-1 The Estimated Probability of Pipe Rupture for the First 50 Epistemic Samples

For many of the epistemic realizations, the estimated failure frequency is zero, meaning none of the 1×10^4 random realizations resulted in a pipe rupture. However, several of the estimates are nonzero. Across the 1×10^3 epistemic realizations, the estimated failure frequency ranges from 0 to 0.99, with roughly 83 percent falling below 0.05 (the vertical line). The histogram of the estimated failure frequencies in Figure 4-2 shows this.

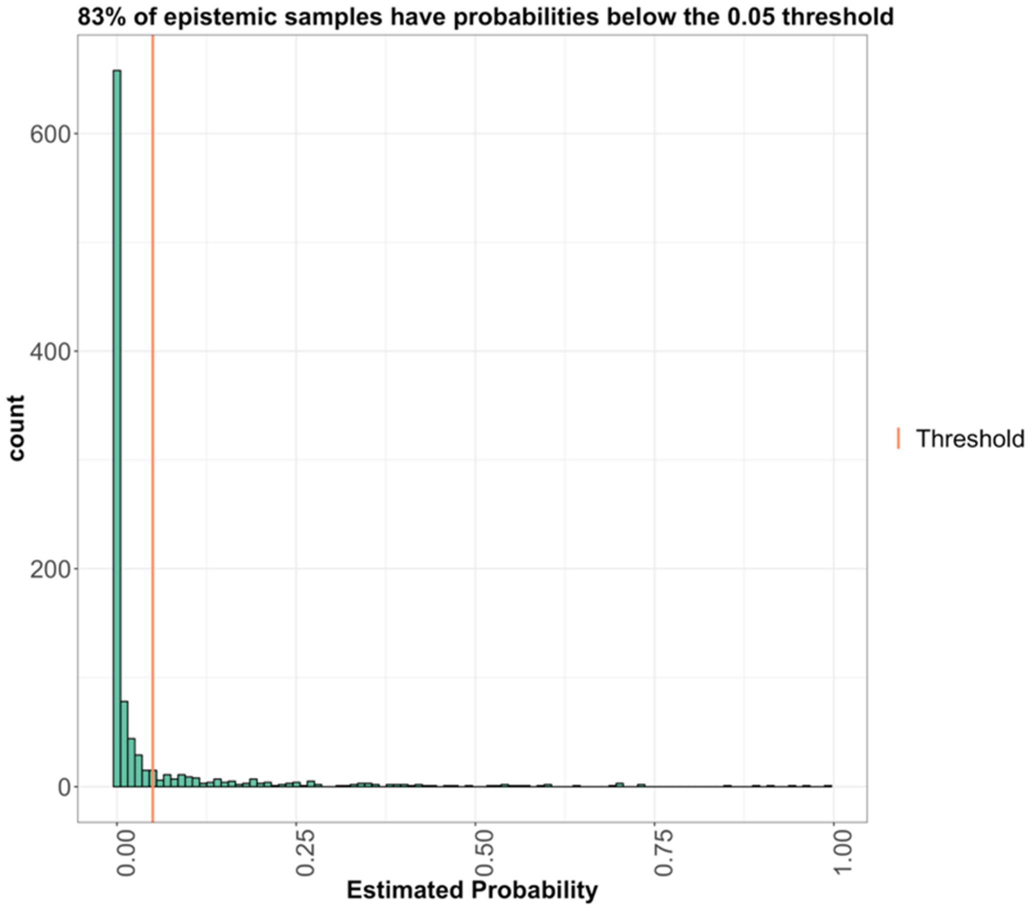

Figure 4-2 Histogram of Estimated Probabilities Across 1,000 Epistemic Realizations

With maintaining the separation of aleatory and epistemic uncertainties, the results can be interpreted as follows: there is roughly 83-percent confidence that the rupture probability is below 0.05. This is likely an optimistic estimate of confidence in the sense that the sampling uncertainty (i.e., finite sample size uncertainty) for each of the estimated probabilities has not been considered.

Without maintaining separation of aleatory and epistemic uncertainties, the estimate of pipe rupture probability would be the proportion over all samples. This proportion is 0.046, which is below the 0.05 threshold used above. However, such an approach mixes the likelihood of rupture (i.e., aleatory uncertainty in rupture) and the confidence associated with rupture (epistemic uncertainty of rupture). When separating, the conclusion is that roughly 17 percent of epistemic values result in rupture probabilities above 0.05. When not separating, the conclusion is the estimated probability of pipe rupture is 0.046. These are two very different conclusions.

4.2 Methods for Model Input Uncertainty Characterization

4.2.1 Statistical Distribution Fitting

4.2.1.1 What Is It?

Given a set of representative data about the input parameter, statistical distribution fitting is the process of estimating the probability distribution of the input parameter using the available data.

4.2.1.2 How to Use?

Statistical distribution fitting has five steps:

(1) Determine relevant data.
(2) Select candidate probability distributions.
(3) Fit distributions to the data.
(4) Evaluate the fit of the distributions to the data.
(5) Select a final input distribution model.

Given a candidate distribution and ample data, most statistical software programs can produce estimates of input distribution parameters, uncertainty in these parameters, and evaluations of model fit. Important considerations for statistical distribution fitting include the following:

- How many data are available and what is the pedigree of those data?
- How much subject matter knowledge is available about the range and shape of the input parameter distribution?
- How much accuracy is needed in the input distribution? (More important inputs require more accuracy.)
- After distribution fitting, how much uncertainty is there in the final estimate of the input distribution?

Reference 4-19 provides specific guidance on fitting models to input distributions. The sections below provide more technical details on distribution fitting.

4.2.1.3 When/Why?

Probability distributions are often used to represent uncertainty in model inputs. Statistical distribution fitting is used when data are available to learn about the form of the input distribution.

The chosen input distribution can impact the PFM results. Expert judgment can inform the distribution, especially when limited or inexact data are available. Additionally, sensitivity studies (Section 4.4) on important input distributions may be needed to assess the impact of assumptions made in the distribution fitting process. When data are not available to estimate probability distributions, expert elicitation can be used (Reference 4-20).

4.2.1.4 Technical Details

This section discusses the five steps in statistical distribution fitting in more detail.

(1) Determine relevant data.

Data quality. The amount and pedigree of the source data are important considerations when determining an input distribution. In practice, cost and time limit data quality. Data quality considerations include the following:

- limited data/small sample size (i.e., the sample is too small to estimate the input distribution with sufficient accuracy)
- data relevance (i.e., not all data points are direct measures of the outcome of interest)
- data uncertainty (i.e., individual data points can contain uncertainty due to measurement error)

The minimum number of data points needed for a suitable fit is subjective and context specific, but smaller sample sizes lead to larger uncertainty in the best fitting input probability distribution. Additionally, very small sample sizes do not allow for data-driven statistical distribution fitting. Expert judgment about the input and its impact on the results can provide additional insight into the process of choosing the input probability distribution.

(2) Select candidate probability distributions.

Distribution models. There is a large set of possible distribution models for input parameters, but, in most cases, simple parametric forms for inputs are used. Common choices include the normal, truncated normal, lognormal, uniform, triangular, and Weibull distributions. Other distributions can be selected, based on their appropriateness for the application at hand.

Considerations in choosing probability distributions include the following:

- range of values the input takes
- tail behavior and overall shape of the distribution

Input Ranges. To specify an input distribution, it is important to consider the range of inputs. Specifically, the range of a distribution should be broad enough to include all possibilities but narrow enough to exclude unrealistic or nonphysical values.

There are two options for bounding the range of an input: (1) select a probability distribution whose range is consistent with the known range of the data, or (2) use a truncated form of a probability distribution. For example, suppose we know an input parameter, such as material strength, is always greater than 0. Then, we can use a distributional model that puts 0 probability mass on values less than 0, such as the lognormal, uniform, or Weibull model. Alternatively, we could use a truncated normal model that truncates the normal distribution such that the input is always greater than 0. Figure 4-3 depicts these two options.

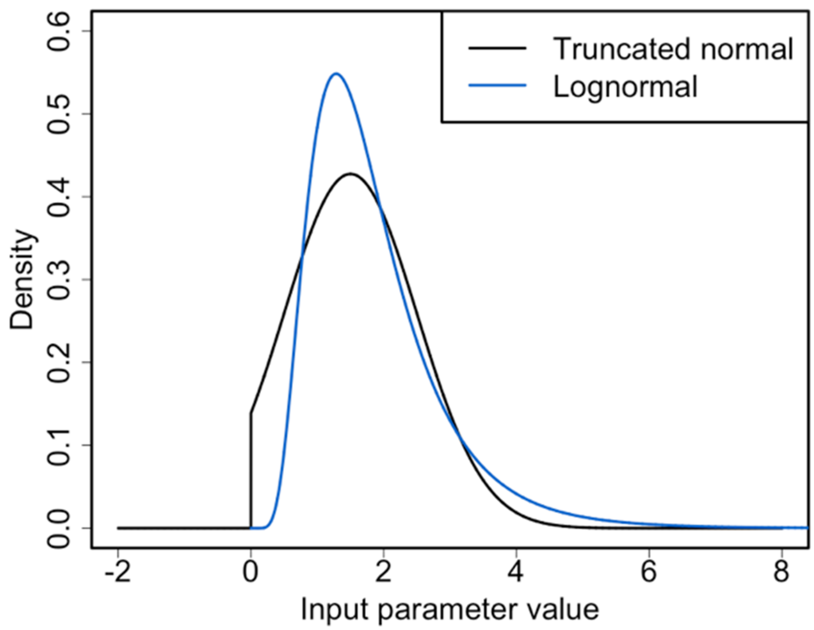

Figure 4-3 Distributions with Nonnegative Input Parameters

Tail behavior and shape. Examples of heavy-tailed and skewed distributions appear in Figure 4-4. Determining the tail behavior and shape requires large sample sizes or expert judgment. Given that the tails of distributions often drive structural failures, it is important to investigate the confidence in the underlying probability distributional form and whether the specified distribution fits the underlying data well in the tails.

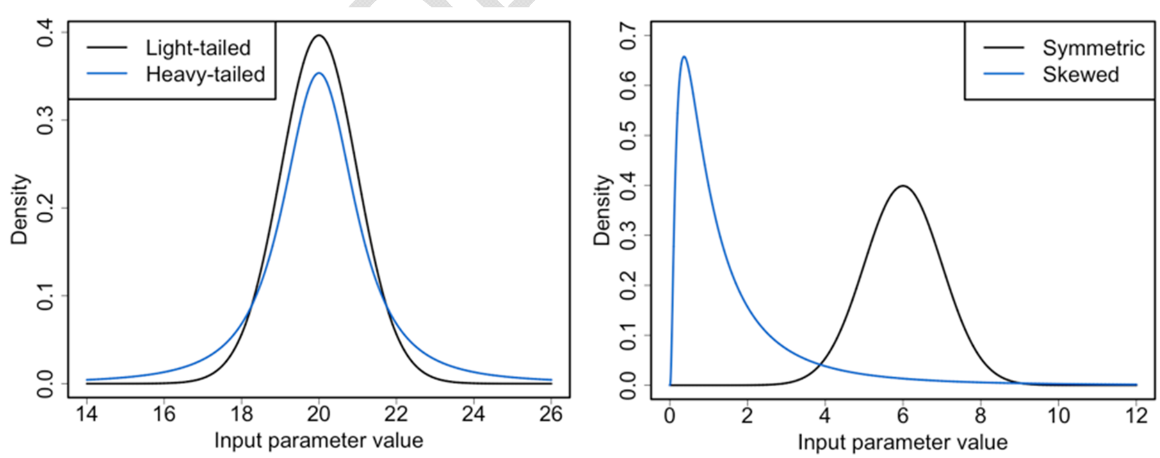

Figure 4-4 Heavy-Tailed and Skewed Distributions

Data transformations. Inputs can be modeled on different scales. A common data transformation is the natural logarithm, where inputs are modeled on the natural log scale, rather than the absolute scale, of the data. This transformation is particularly useful for skewed, positive inputs.

(3) Fit the distributions.

Parameter estimation. Given a candidate probability distribution and a set of data, statistical inference can be used to estimate the parameters of that distribution. Most statistical software programs (e.g., R, Python, Minitab, MATLAB with the appropriate toolbox, Easyfit©) can estimate distribution parameters, along with uncertainty in those parameters. These parameters are typically estimated using statistical inference techniques such as maximum likelihood estimation, Bayesian inference, or method of moments. If the data cannot be modeled well using a known probability distribution, then nonparametric approaches can be applied.

Input uncertainty. The estimated distribution parameters contain sampling uncertainty, because they were estimated based on a finite sample of data (Figure 4-5). In the figure, the grey bars are a histogram of the data, with best fit normal distribution shown as the black line. The blue lines are sampling uncertainty in the distributional fit due to the limited sample size when n=10 (left) and n=100 (right).

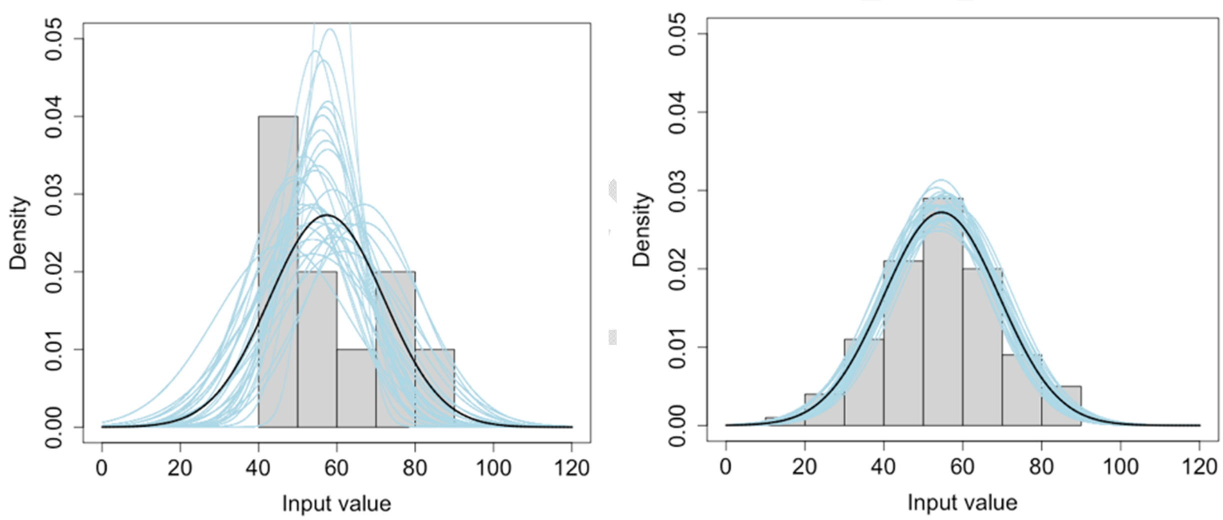

Figure 4-5 Sampling Uncertainty in Input Distribution Fits

(4) Evaluate distributional fit.

After fitting a distribution to data, it is important to evaluate how well the distribution matches the observed data. From Reference 4-19, the basic principle behind evaluating distributional fits is to compare the parametric estimates from the model fit to nonparametric quantities that are not based on a fitted model. Representations of some of the graphical tools, described below, appear in Figure 4-6:

- Overlay a parametric fit of the probability density function onto a histogram of the data. The left plot shows n=50 data points fit to a Weibull distribution. Large differences between the histogram and parametric probability density function (PDF) estimate would signal poor model fit.

- Overlay a parametric fit of the cumulative distribution function (CDF) onto the empirical CDF of the data (middle plot). Large differences between the empirical CDF and parametric CDF estimate would signal poor model fit.

- Construct a probability plot, also called a Quantile-Quantile plot (right plot). Probability plots compare model-estimated versus empirical quantiles of the data. A departure from the reference line indicates a region where the model poorly fits the data. Because there is sampling uncertainty in the quantile estimates, confidence intervals can help assess whether there is statistical evidence of a lack of model fit. Points falling outside the bounds indicate a lack of fit of the probability distribution.

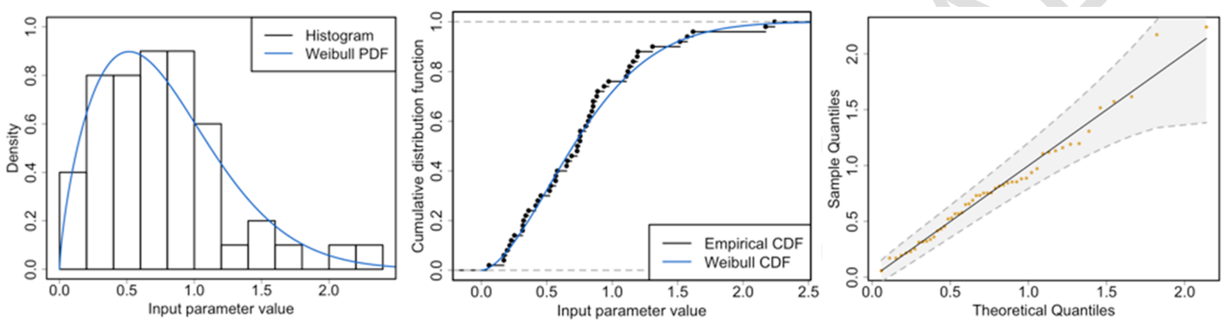

Figure 4-6 Graphical Diagnostics for Parametric Model Fit

Additionally, statistical goodness-of-fit hypothesis tests can also be used to detect evidence of a poor model fit. These tests have strong, known limitations that limit their applicability in practice and must be supplemented with graphical tools and expert judgment to determine whether a model is a reasonable fit to data (Reference 4-19).

(5) Select a final input distribution model.

The final input distribution is an estimate of input uncertainty that reflects both data-driven evidence and expert judgment (particularly in limited-data scenarios). To select an input distribution, the analyst selects the following:

- **Distribution model.** It is best practice to consider several different candidate probability distributions (e.g., normal, lognormal, Weibull), and select the final distribution based on which is the best fit to the data, a process known as model selection.

- **Values for the distribution parameters.** Recall that input parameters are uncertain, due to estimation based on a finite sample size.

In many instances, there will be uncertainty in the choice of the distribution model and distribution parameter values. Input distribution uncertainty can have a large impact on the final estimate in a PFM analysis. This source of uncertainty is most important when one of the following is true:

- The output is sensitive to the input.

- There are insufficient data to accurately estimate an input distribution.
- The acceptance criterion relates to bounding a probability below a very low threshold (e.g., $p < 1\times10^{-6}$).

This uncertainty can be incorporated into the final PFM analysis in different ways:

- Treat uncertain probability distributions and their parameters as additional sources of epistemic uncertainty.
- Choose values of the distribution parameters resulting in conservative values for the inputs with respect to the application at hand.
- Examine the robustness to changes in the input distribution using a sensitivity study (Section 4.4).

4.2.2 Preserving Physical Relationships between Inputs

4.2.2.1 What Is It?

In a PFM analysis, most uncertain inputs are assumed to be statistically independent; that is, changing the value of one input does not impact the value of other inputs. However, a subset of input variables is often statistically dependent. For the input set to be physically realistic, these dependencies should be preserved.

4.2.2.2 How to Use?

Before modeling, expert judgment is applied and exploratory data analyses are conducted to understand the relationship between inputs. Relationships can manifest themselves as correlations or as more general dependencies such as nonlinear relationships or ordering relationships (i.e., input 1 must be larger than input 2).

Some approaches to specifying statistically dependent inputs include the following:

- inducing correlation in random samples through the transformation of independent samples
- constructing a joint probability distribution for the inputs that includes the dependencies
- specifying a conditional probability distribution for one input as a function of the other input
- constraining the parameter space

Section 4.2.2.4 provides technical details on these approaches.

4.2.2.3 When/Why?

Preserving dependence between variables is often needed to ensure a physically realistic input set or to maintain the physical laws that drive the problem. For example, the inner diameter of a pipe must be smaller than its outer diameter, so a relationship between these variables may be imposed to ensure the physicality of the inputs.

4.2.2.4 Technical Details

The technical details below describe the four approaches to specifying statistically dependent inputs given above.

Transforming Independent Samples*.* Correlated inputs can be generated by transforming independent random samples. If inputs follow a multivariate normal distribution, then we can directly transform the inputs to induce correlation. Specifically, consider two inputs x and y and a random, independent sample of size n for each. Center and scale the samples of x and y to have mean 0 and standard deviation 1. Let $\boldsymbol{X}$ be the nx2 matrix whose columns are formed by the centered and scaled samples of x and y, respectively. Let $\boldsymbol{S}$ be the 2x2 matrix specifying the correlation:

$$\boldsymbol{S} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

where ρ is the desired correlation between the inputs. Let $\boldsymbol{S} = \boldsymbol{C}^{\top}\boldsymbol{C}$ be the Cholesky decomposition of $\boldsymbol{S}$ and set $\boldsymbol{X}^{*} = \boldsymbol{X}\boldsymbol{C}$. The correlation between the columns of $\boldsymbol{X}^{*}$ is $Corr(\boldsymbol{X}^{*}) = (\boldsymbol{X}\boldsymbol{C})^{\top}(\boldsymbol{X}\boldsymbol{C}) = \boldsymbol{C}^{\top}\,\boldsymbol{X}^{\top}\,\boldsymbol{X}\,\boldsymbol{C} = \boldsymbol{C}^{\top}\,\boldsymbol{I}\,\boldsymbol{C} = \boldsymbol{S}$. The desired standard deviation for each input can be achieved by scaling each column by its desired standard deviation. Next, the desired mean can be added to each column. If inputs are not normally distributed, then this transformation method will not preserve the probability distributions of the individual inputs (i.e., marginal distributions).

If inputs are not normally distributed, then an alternative approach is to induce correlation on the ranks of the inputs (Reference 4-21). This approach has the advantages of being distribution free and preserving the marginal distribution of the inputs. To implement this approach, the analyst specifies the correlations between the ranks of the inputs, which ideally can be estimated using experimental data or expert judgment. This estimate can be applied with SRS and LHS. Reference 4-21 gives details on implementing the rank correlation method. Figure 4-7 provides an example of the rank approach for two input variables. On the left is a scatterplot of a random sample of two variables. Transforming these points using the rank method results in the scatterplot on the right where a strong negative correlation now exists.

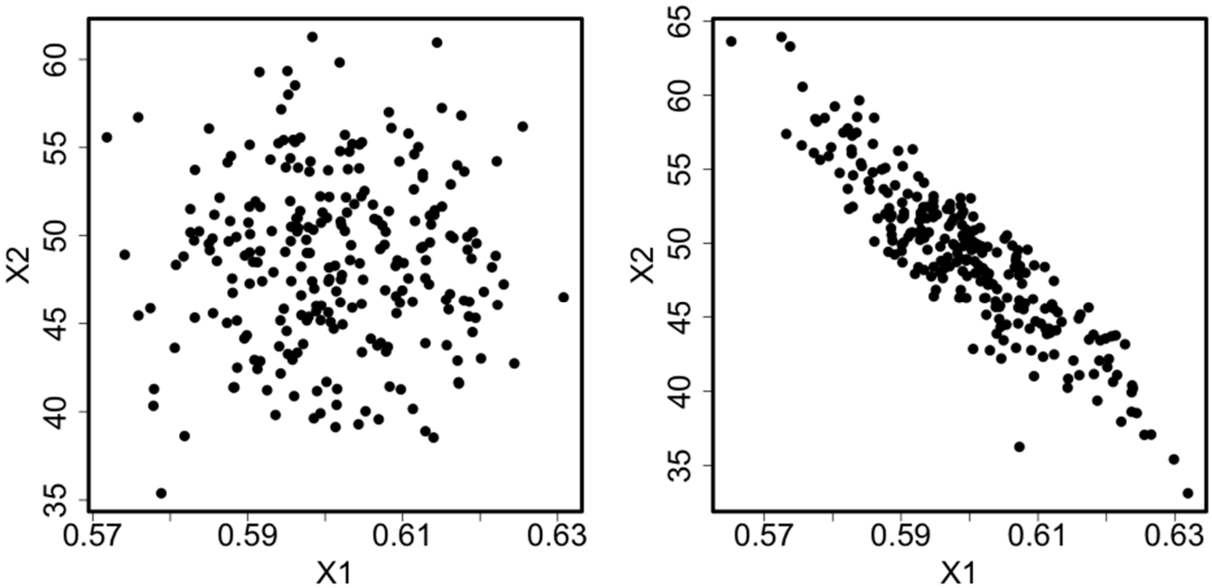

Figure 4-7 Randomly Sampled Inputs (Left) and Transformed Inputs (Right)

Joint distribution modeling. Input parameters can be directly sampled from a joint distribution for the parameters that include a correlation structure. The multivariate normal distribution is a straightforward model for correlated inputs but is only appropriate when a normal distribution can reasonably represent the inputs. The multivariate normal distribution is parameterized by the mean and variance of each variable, along with the statistical (Pearson) correlation between pairs of variables.

If a multivariate normal distribution cannot reasonably represent the joint distribution, more sophisticated statistical models can be applied to specify a joint distribution. Specifically, copula methods are a popular statistical approach for specifying joint distributions of correlated variables (Reference 4-22).

Conditional probability. If one input is dependent on another input, the relationship can be modeled to induce correlation between inputs. Specifically, the joint distribution of inputs X and Y can be factorized as the product between the marginal distribution of X and the conditional distribution of Y given X: $f(X,Y) = f(X)f(Y|X)$. As an example, suppose the marginal distribution of X is Weibull:

$$X \sim Weibull(1,1), \tag{1}$$

and the conditional distribution of Y given $X = x$ is normal with a mean dependent on x:

$$Y = Normal(15 + 2\log(x), .3). \tag{2}$$

Figure 4-8 displays samples from the joint distribution of X and Y. To sample from the joint distribution of X and Y, first sample a value of X from the Weibull distribution (Eq. 1), and then sample the value of Y from a normal distribution with mean $15 + 2\log(x)$ and variance 0.3 (Eq. 2).

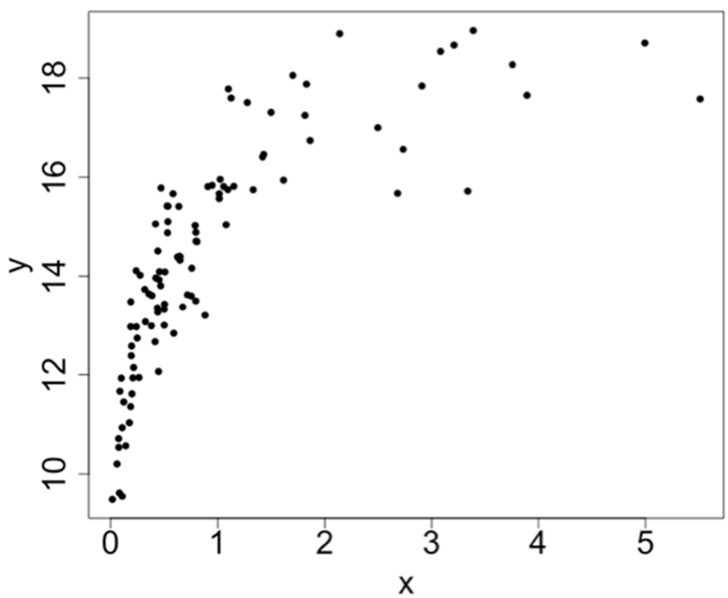

Figure 4-8 Sampled Inputs from the Joint Distribution of *X* and *Y*

Constrain the parameter space. The input set can be constrained to ensure consistency and physicality. For example, yield strength of a material is lower than its ultimate tensile strength (Figure 4-9), and any samples not satisfying this constraint can simply be discarded. Note that constraining the input space changes the uncertainty distribution on the inputs and can induce correlation between inputs. Therefore, this approach should be used with caution to ensure that imposed constraints accurately represent uncertainty in the inputs.

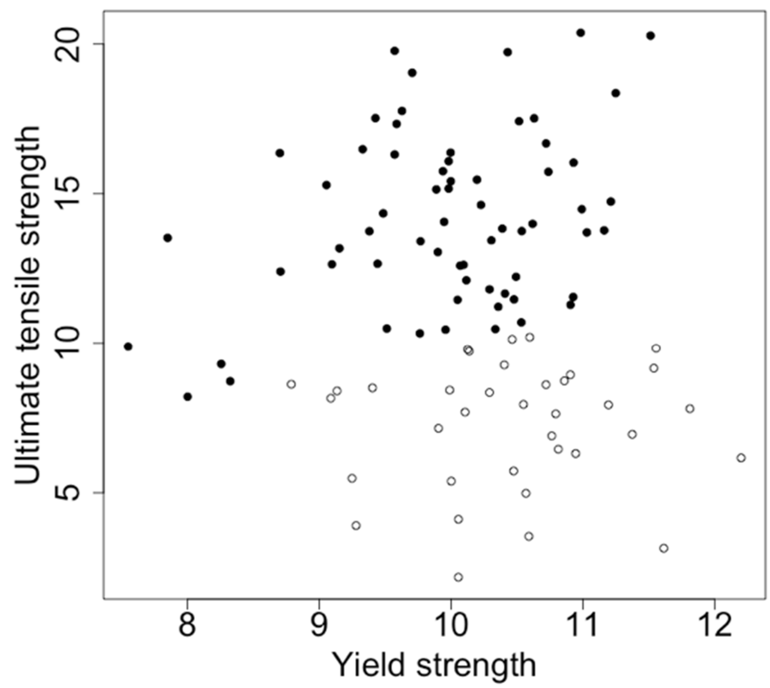

Figure 4-9 Constraining Input Distributions to Ensure that Yield Strength Is Less than Ultimate Tensile Strength (Open Circles Are Not Admissible)

All four of the above approaches are viable options for modeling dependencies in inputs. An advantage of the first approach (transforming independent samples) is that it is distribution free and requires only knowledge of the correlation between inputs (Reference 4-21). An advantage of the second approach (joint distribution modeling) is that correlation is directly built into the input parameter distribution. The third approach (conditional probability) gives flexibility with respect to the functional form of the dependency between variables. The fourth approach (constrain the parameter space) offers simplicity in implementation.

4.3 <u>Useful Methods for Forward Propagation of Input Uncertainty</u>

4.3.1 Simple Random Sampling

4.3.1.1 What Is It?

Simple random sampling (SRS) is a Monte Carlo sampling technique in which each uncertain input is sampled randomly from its corresponding probability distribution.

4.3.1.2 How to Use?

The SRS approach follows four steps:

(1) Specify probability distributions for the uncertain inputs.

(2) Choose the sample size.

(3) Implement SRS by randomly sampling the inputs from their probability distributions.

(4) Evaluate the computer model at each of the sampled inputs. The sampled outputs represent a random sample of outputs corresponding to the probability distribution implied by the distributions on the inputs.

4.3.1.3 When/Why?

SRS is easy to implement and therefore serves as a good "first pass" sampling scheme for understanding output variability. However, SRS is often less efficient than alternative sampling schemes (see Sections 4.3.2, 4.3.3, and 4.3.4 for alternatives); that is, more realizations are needed using SRS than alternative sampling schemes to estimate a QoI with the same precision. Section 4.3.1.4 gives information on choosing a sample size for SRS.

4.3.1.4 Technical Details

Specifying probability distributions for the uncertain inputs. Section 4.2 discusses methods to identify and specify probability distributions for uncertain inputs.

Choosing a sample size n. It is usually feasible to directly estimate the amount of sampling uncertainty in a QoI associated with SRS. For example, the law of large numbers indicates that the magnitude of the sampling error associated with many QoIs estimated using SRS will be proportional to $1/\sqrt{n}$ where n represents the sample size.

To use SRS to estimate a probability, the number of samples should be large relative to the probability of the event occurring. As a rule of thumb, the sample size should be at least 10 to 20 times larger than $1/p,$ where $p < 0.5$ is the probability of interest, to generate stable results (Reference 4-23).

Implementing SRS. Most software programs can directly implement SRS for many common probability distributions. Alternatively, a simple random sample can be generated for general distributions by transforming uniform random samples on the interval 0 to 1 using the probability integral transform (Reference 4-24). The uniform samples represent the quantiles of the distribution from which the sample is desired. These quantiles are transformed by applying the inverse CDF for the desired distribution. All that is needed to implement the probability integral transform is the ability to randomly sample uniform variables and evaluate the inverse CDF. As an example, a two-dimensional sample of uniform variables appears in Figure 4-10. Each dimension is transformed using the probability integral transform to obtain the simple random sample of variables shown in Figure 4-11, where the first dimension is distributed uniformly on the interval -1, 1 ($U(-1,1)$) and the second dimension is normally distributed with mean 0 and variance 1 ($N(0,1)$).

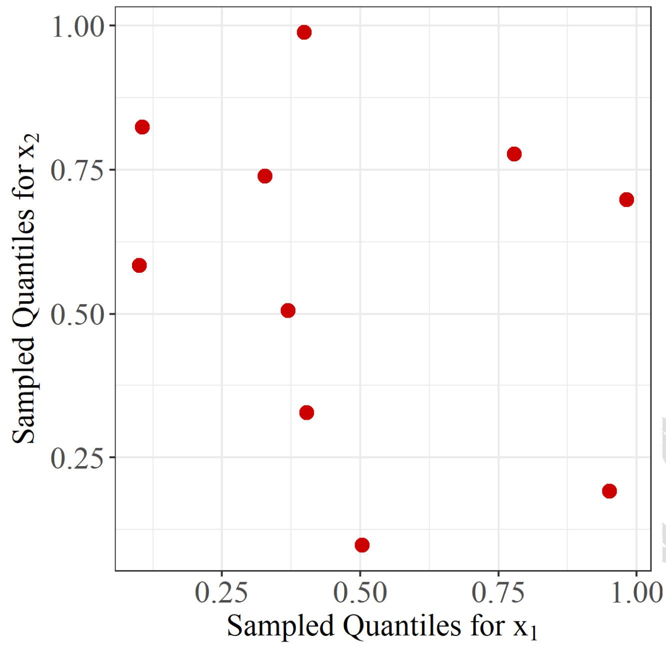

Figure 4-10 SRS Sample in the Quantile Space for Two Input Variables (n=10)

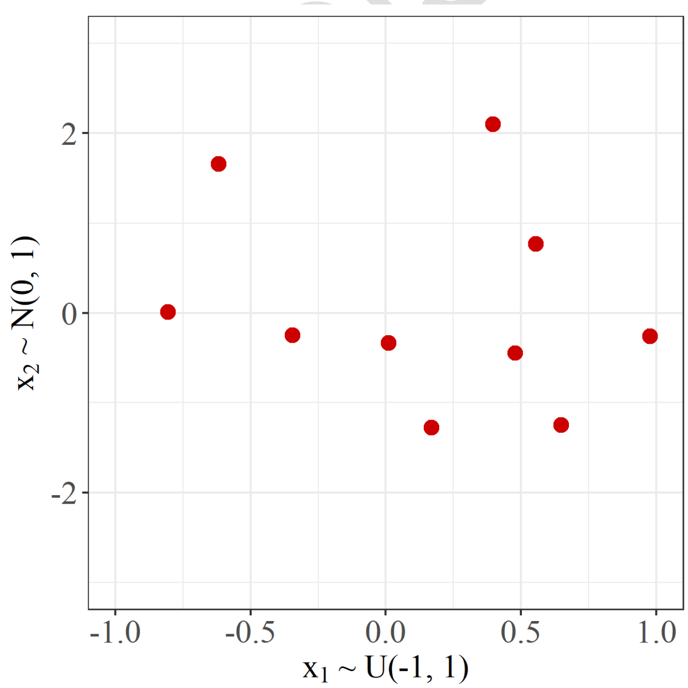

Figure 4-11 SRS Sample Transformed into the Input Space for Two Input Variables (n=10)

4.3.2 Latin Hypercube Sampling

4.3.2.1 What Is It?

Latin hypercube sampling (LHS) is a Monte Carlo sampling technique. LHS is a method to obtain a sample that is more spread out across the input space than a typical SRS sample, producing estimates with more statistical precision on average.

4.3.2.2 How to Use?

Many statistical software programs can implement LHS. The following steps describe a method for generating an LHS of size n from independent input distributions associated with p uncertain inputs (Reference 4-25):

(1) Stratify the input space by dividing the range of each input, x_j, into n disjoint intervals of equal probability.

(2) For each input, randomly sample a single value from each interval, resulting in n sampled values for each input. For a given input and interval, the sample is taken from the conditional distribution of the input on the interval.

(3) Randomly combine samples without replacement:

- a. Randomly pair, without replacement, the n values sampled from the first input, x_1, with the n values from the second input, x_2, to produce n pairs.
- b. Randomly combine these pairs, without replacement, with the n values sampled from the third input, x_3.
- c. Continue this process iteratively on $x_4, x_5, \dots, x_p$ resulting in a set of n p-tuples.

The correlation between inputs can be incorporated using the Iman-Conover procedure that induces correlation based on the ranks of inputs (Reference 4.3; see Section 4.2.2).

To summarize, LHS stratifies each input dimension into equally probable strata. In each dimension, each stratum is sampled once (the regions formed by the sampled strata create a pattern akin to a Latin hypercube in experimental design, such as described in Reference 4-26). Within each of the sampled regions, a single sample is randomly sampled according to the distribution within the region (Reference 4-25, 4-27).

4.3.2.3 When/Why?

LHS is designed to cover the range of the input space more efficiently than SRS (Section 4.3.1). For this reason, it is a common technique for forward propagation of uncertainty and for building surrogate models (Section 4.3.10). Compared to SRS, LHS will typically result in more statistically precise estimates of a QoI; however, the increase in precision diminishes as the sample size increases. Quantifying statistical uncertainties on QoIs calculated using LHS is more challenging.

4.3.2.4 Technical Details

The following simple example demonstrates the steps of the LHS algorithm outlined above. It shows an LHS for two input variables, x_1 and x_2, with n=10 samples. In this example, x_1 is uniformly distributed from -1 to 1, and x_2 is normally distributed with a mean of 0 and a standard deviation of 1:

(1) **Stratify the input space.** First, each input distribution is divided into $n = 10$ intervals (strata) of equal probability according to their respective distributions. This stratification can be done first in the quantile space defined as the two-dimensional hypercube on (0,1). The intervals (strata) in the quantile space are evenly spaced, as displayed in Figure 4-12.

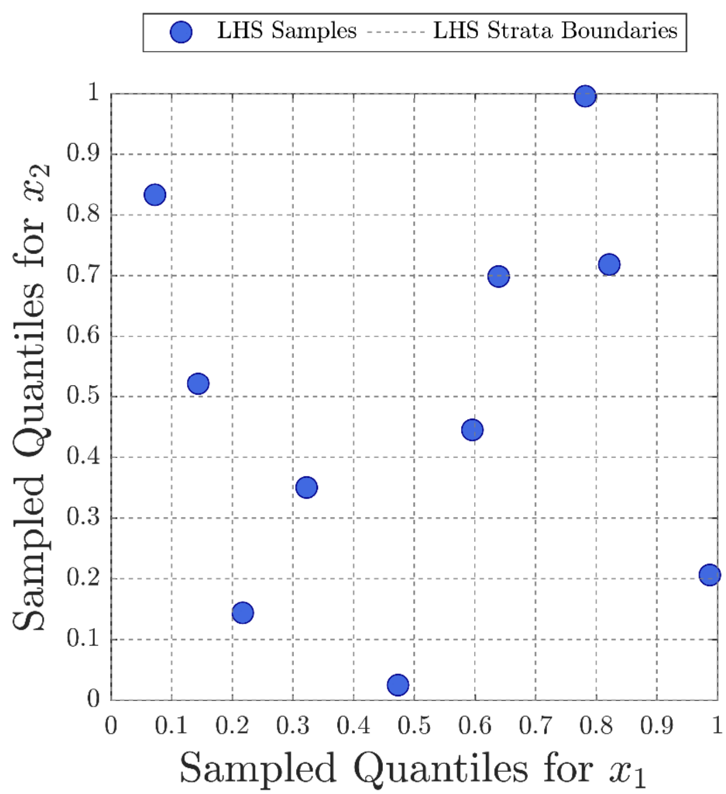

Figure 4-12 Example of an LHS in the Quantile Space for Two Input Variables (n=10)

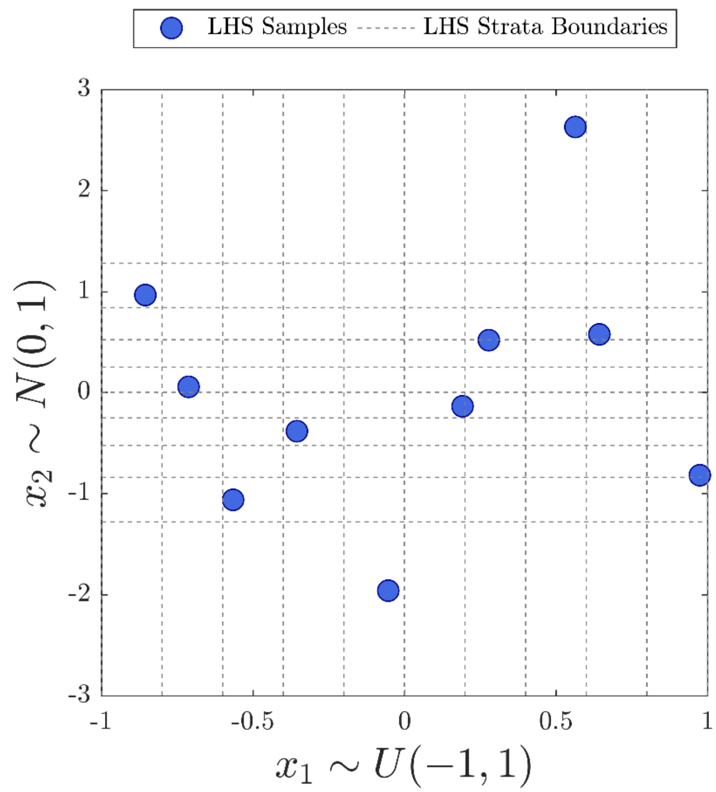

Figure 4-13 LHS Transformed into the Input Space for Two Input Variables (n=10)

(2) **Randomly sample from each interval.** Next, random uniform samples are taken on each interval in each dimension. These samples are transformed to a sample from their specified input distributions using the probability integral transform (Reference 4-28). Figure 4-12 shows uniform samples for each stratum plotted in the two-dimensional quantile space. The transformed samples (and strata) appear in Figure 4-13. The next step describes the displayed pairing of each x_1 sample with an x_2 sample.

(3) **Randomly combine samples without replacement.** Finally, the values sampled from the first input, x_1, are randomly paired without replacement with the values sampled from the second input, x_2. Once a sampled value of x_2 is randomly paired with a sampled value of x_1, this sampled value of x_2 cannot be paired with a different value of x_1. In the example shown in Figure 4-12, the sample from the first strata of x_1 is randomly paired with the sample from the ninth strata of x_2. The sample from the first strata of x_1 is not paired with any other samples of x_2. Similarly, the sample from the ninth strata of x_2 is not paired with any other samples of x_1. Next, the sample from the second strata of x_1 is randomly paired with the sample from the sixth strata of x_2, and so on, until all n=10 pairs are selected.

Augmenting LHS designs. The design of an LHS depends on a known sample size. If, after a sample is selected, the analyst wants to decrease sampling uncertainty in a QoI by increasing the sample size, care must be taken to preserve the properties of an LHS. Reference 4-29 outlines one method for augmenting an initial LHS while preserving the LHS structure.

Discrete Probability Distributions. A discrete probability distribution is a related method that produces a discrete approximation to a continuous distribution (Reference 4-30). Like LHS, strata are created in each dimension. Rather than sampling within each stratum, the conditional mean within the stratum is chosen as the sample point.

4.3.3 Importance Sampling

4.3.3.1 What Is It?

Importance sampling is a Monte Carlo technique that can be used to more efficiently sample from the model input space than equal-probability sampling methods such as SRS and LHS (Reference 4-31). Importance sampling concentrates samples in a specific area of interest in the input space to improve estimation of QoIs (e.g., failure probabilities).

4.3.3.2 How to Use?

The implementation of importance sampling follows three steps:

(1) **Choose an importance distribution for the model inputs.** The importance sampling distribution concentrates samples in regions of the input range that have a strong influence on the estimate of the QoI. The importance distribution depends on the relationship between the QoI and the inputs and therefore can be informed by SA. The distribution should be selected such that the estimated QoI has smaller variance than a QoI estimate from an equal-probability sample.

(2) **Sample inputs from the importance distribution** and run the model at these inputs.

(3) **Estimate the QoI.** Importance samples are weighted to obtain an unbiased estimate of the QoI. The importance weights are derived from the density functions of the original input distribution and the importance distribution (Reference 4-32).

4.3.3.3 When/Why?

Importance sampling is used to reduce the sampling uncertainty in the estimate of a QoI. If a good importance sampling distribution has been selected, then the estimate of the QoI will be more precise with fewer samples relative to SRS or LHS. Importance sampling reduces sampling uncertainty by concentrating samples in the regions of importance (i.e., those regions that contribute most to the QoI). While importance sampling is theoretically used for variance reduction, poor choice of an importance distribution will increase the variance of a QoI estimate (Reference 4-31).

Importance sampling can be particularly beneficial in PFM applications when the QoI is a rare probability. That is, to estimate a 1×10^{-6} probability, it is more computationally efficient to concentrate more samples around the area where events are more likely to occur. Without importance sampling, an event will only be observed, on average, every one million samples. Importance sampling algorithms can be designed to dramatically increase the number of observed events and, subsequently, decrease the variance of the probability estimate for a fixed number of samples.

4.3.3.4 Technical Details

Many QoI estimation problems can be formulated as the estimation of an expectation:

$$E_\pi[f(\boldsymbol{x})] = \int f(\boldsymbol{x})\pi(\boldsymbol{x})d\boldsymbol{x}$$

A common PFM example is when $\pi(\boldsymbol{x})$ represents the probability distribution on a multivariate input $\boldsymbol{x}$ and $f(\boldsymbol{x})$ is a model output; that is, the indicator of an adverse event (e.g., pipe rupture) at input x (i.e., $f(\boldsymbol{x}) = 1$ if the event occurs and 0 otherwise). In this case, the expectation reduces to the probability of the adverse event.

By the law of large numbers, the average of f over a random sample $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(n)}$ from $\pi(\boldsymbol{x})$ will converge to $E_\pi[f(\boldsymbol{x})]$ as n grows. Hence, it is straightforward to estimate the integral with the average:

$$E_\pi[f(\boldsymbol{x})] \approx \frac{1}{n}\sum_{i=1}^{n} f\left(\boldsymbol{x}^{(i)}\right)$$

In the case of an indicator of a rare event (and other cases), this average is inefficient since very few of the random samples will result in $f(\boldsymbol{x}) = 1$ (i.e., it is difficult to randomly sample an input that results in the adverse event). Instead of sampling from $\pi(\boldsymbol{x})$, importance sampling takes its sample from an importance distribution $h(\boldsymbol{x})$. Rewriting the above integral as

$$E_\pi[f(\boldsymbol{x})] = \int f(\boldsymbol{x})\frac{\pi(\boldsymbol{x})}{h(\boldsymbol{x})}h(\boldsymbol{x})d\boldsymbol{x} = E_h\left[f(\boldsymbol{x})\frac{\pi(\boldsymbol{x})}{h(\boldsymbol{x})}\right],$$

we notice that it can be estimated from a sample $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(n)}$ from $h(\boldsymbol{x})$ using a weighted average:

$$E_\pi[f(\boldsymbol{x})] \approx \frac{1}{n}\sum_{i=1}^{n} f\left(\boldsymbol{x}^{(i)}\right)\frac{\pi\left(\boldsymbol{x}^{(i)}\right)}{h\left(\boldsymbol{x}^{(i)}\right)}.$$

The values $w^{(i)} = \frac{\pi\left(\boldsymbol{x}^{(i)}\right)}{h\left(\boldsymbol{x}^{(i)}\right)}$ are importance weights on the $\boldsymbol{x}^{(i)}$. As we demonstrate below, the careful choice of importance distribution h can dramatically reduce sampling uncertainty of the estimate. Reference 4-32 gives the technical conditions on h needed for the weighted average to converge to $E_\pi[f(\boldsymbol{x})]$ as n grows.

Choose an importance distribution for the model inputs. In practice, choosing the importance distribution h usually involves selecting individual importance distributions for a few of the inputs. An incorrect choice of inputs for importance sampling or a poor selection of the importance distribution may lead to increasing the sampling uncertainty in the estimate of the QoI when compared to an estimate generated without the use of importance sampling. As a result, a careful and thorough analysis is necessary before selecting the importance distribution. Effective implementation of importance sampling requires (1) understanding what regions of the input space are important to the QoI and (2) selecting the importance distribution correctly given the relationship between the important input and the QoI (References 4-31, 4-33, 4-34, 4-35). This aspect of importance sampling is often not straightforward in PFM studies and requires SAs to support the selected inputs, which are often confirmed with expert elicitation.

Inputs that have a strong relationship with the output are good candidates for importance sampling. SA methods (Section 4.3.8) can be used to quantify the input-output relationship and rank variables in terms of their influence on the QoI. The variables that are found to have the strongest relationship with the QoI are considered for importance sampling.

Inefficiency in importance sampling often occurs in high-dimensional problems where too many variables are importance sampled (Reference 4-36). To avoid this inefficiency, importance sampling should be limited to only a few important variables.

Importance sampling and estimation of the QoI. After the nontrivial task of choosing a good importance distribution, the implementation of the importance sampling methodology is straightforward. Inputs are randomly sampled from their importance distribution and propagated through the model. The final QoI is then estimated as a weighted average, weighted by the importance weights (References 4-31, 4-33, 4-35).

Illustration of importance sampling. Importance sampling can be demonstrated as follows. Suppose there is one normally distributed input $x \sim N(0,1)$, and the goal is to estimate $E[f(x)] = P(X > 2.5)$. The true probability is known to be 0.0062. Figure 4-14 shows estimates using repeated simple random samples of size $n = 1{,}000$ from $N(0,1)$ and repeated importance samples from a Student-t distribution centered at 2.5 with 3 degrees of freedom. The importance distribution was chosen to ensure more samples fell inside the failure region (> 2.5), and 3 degrees of freedom were used to produce a heavy-tailed Student-t distribution in that region, as smaller degrees of freedom increase the tail weight of the distribution. The histograms in Figure 4-14 represent sampling uncertainty in the QoI estimate. While both estimates are unbiased around the true probability indicated by the vertical dashed line, the standard deviation of the estimates under importance sampling is 0.0003 compared to 0.003 under random sampling. This change represents a reduction in the sampling uncertainty by an order of magnitude. Reference 4-32 describes a similar example.

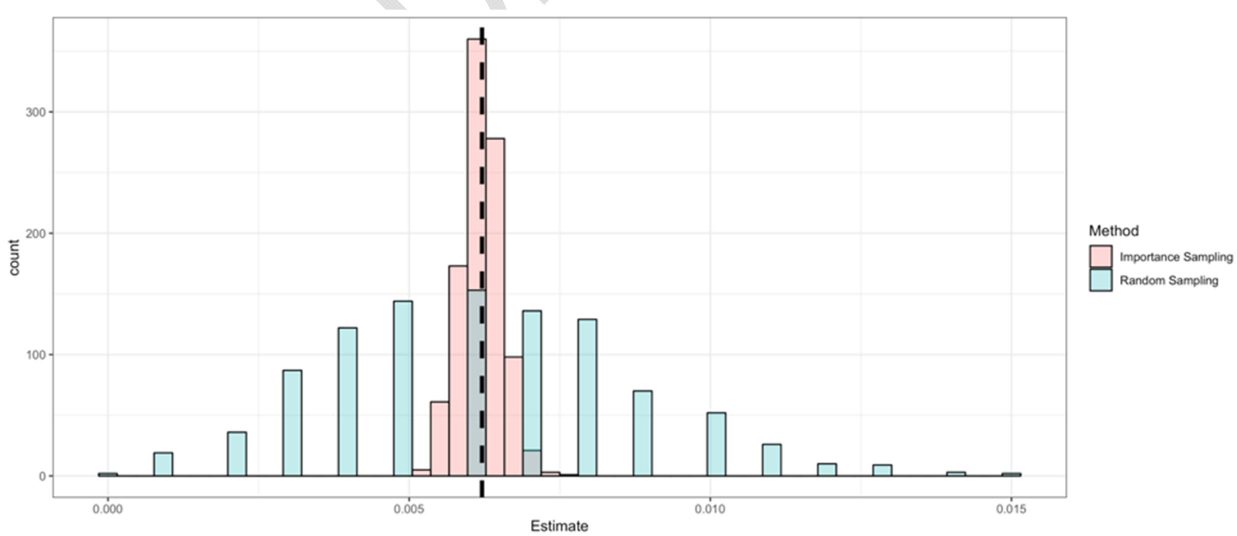

Figure 4-14 Example of Estimating a Probability Using Random Sampling and Importance Sampling

Adaptive importance sampling. In reliability analysis (see Section 4.3.4), whose goal is to estimate a failure probability, adaptive importance sampling techniques are often used to help refine the importance distribution. The goal in these applications is to detect the failure boundary, defined as the boundary of the region separating failures and nonfailures (Reference 4-37), in order to improve failure probability estimates. Adaptive methods iteratively update the importance distribution to better estimate this boundary by considering the outputs of previously sampled points in the domain (References 4-38, 4-39).

To implement adaptive importance sampling, first, an optimization problem is solved to find a particular point on the failure boundary known as the most probable point (MPP). The MPP can be used as a starting point to define an initial importance distribution, which through adaptive sampling is updated as model evaluations are obtained. For example, multimodal sampling (Reference 4-38) and curvature-based sampling (Reference 4-39) begin centered at the MPP and then update the sampling distribution by assigning weights to various candidate density functions. Many software packages such as DAKOTA (Reference 4-31) have the capability to implement adaptive sampling methods.

Instead of using additional evaluations of the computational model, adaptive sampling can also use information from surrogate models (Section 4.3.10). This feature is most useful when there is a need to reduce the computational expense of model evaluations. For example, efficient global reliability analysis (Reference 4-40) aims to create a Gaussian process (GP) surrogate model for the function of interest and then adaptively select sample points in the domain near the failure region to improve the quality of the surrogate model.

4.3.4 First- and Second-Order Reliability Methods

4.3.4.1 What Is It?

Reliability methods estimate a failure probability by approximating the probability of violating a certain threshold criterion for a probabilistic analysis of continuous random variables. For example, these methods can estimate the probability that a particular material stress is greater than the yield stress. Often, the estimate requires many fewer samples of the computer model than using Monte Carlo sampling methods. The methods described here are known as the first-order reliability method (FORM) and the second order reliability method (SORM).

4.3.4.2 How to Use?

The use of FORM and SORM follows three steps:

(1) Define the failure region in terms of a continuous output and a threshold value. Specifically, a failure occurs when the output exceeds the threshold. The failure probability is the integral of the input probability distributions over the failure region.

(2) Approximate the failure region using FORM or SORM Taylor series approximation around the MPP. The MPP is the point on the failure region boundary with highest input probability density. Determining the location of the MPP requires evaluating the computational model within an optimization algorithm.

(3) Estimate the failure probability using the integral over the approximate failure region.

4.3.4.3 When/Why?

Reliability methods are particularly useful because of the computational efficiency of the algorithms. As described in Sections 4.3.1, 4.3.2, and 4.3.3, sampling-based algorithms can also be used to approximate the failure probability of a system. However, these methods often require thousands or tens of thousands of samples to provide good estimates. In contrast, FORM and SORM are more efficient reliability methods because they seek to directly understand the location and probabilistic distance to the limit state (i.e., the boundary of the failure region). Often, for some low-dimensional input spaces, the MPP can be located accurately with a small number of model evaluations (on the order of 10 points). This difference provides substantial computational savings especially when the analysis model is computationally expensive to evaluate. However, the failure probability estimate obtained from FORM or SORM relies on a Taylor series approximation to the shape of the limit state (first-order series in FORM and second-order series in SORM) and can result in a poor estimate if the approximation is not good.

4.3.4.4 Technical Details

Defining the failure region: Following Reference 4-41, consider a model that predicts an output Y as a function g of some set of input random variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$:

$$Y = g(\boldsymbol{X})$$

Suppose a failure event of interest is defined when $y < 0$. Note that any problem can be formulated in this way by considering y to be a margin against failure. For example, X_1 might be the predicted stress in a material, and X_2 might be the yield stress of the material. If we define failure when the material yields, then the margin Y against failure is defined as $Y = X_2 - X_1$, and failure occurs when $y < 0$ for some values (x_1, x_2).

In any such scenario, the goal of reliability methods is to compute the failure probability $p_f = P(Y < 0) = \int_{FR} \pi(\boldsymbol{x}) d\boldsymbol{x}$, where π is the joint probability density function of the inputs $\boldsymbol{X}$ and $FR = \{\boldsymbol{X} : Y = g(X) < 0\}$ is the failure region. The unknown and potentially complex failure region makes computing the integral difficult. To simplify the computation, reliability methods like FORM and SORM make simplifying assumptions on $g(\boldsymbol{X})$.

Approximating the failure region using FORM with normally distributed inputs. After the failure region has been defined, the failure probability can be estimated with FORM using the following steps, described in detail below:

(1) Transform each input variable into the standard normal space.

(2) Find the MPP, $\boldsymbol{x}'^{*}$.

(3) Calculate the distance, β, from 0 to the MPP.

(4) Use this distance to estimate the failure probability $p_f \approx 1 - \Phi(\beta)$, where Φ is the CDF of the standard normal distribution.

Figure 4-15 depicts this process visually (following Reference 4-42).

Transformation of inputs. A problem first addressed by the Hasofer-Lind method (Reference 4-43) assumes each input is independently normally distributed. The first step in this method (and others that relax the normality assumption) is to transform each random variable into the standard normal space (i.e., $N(0,1)$) so that all variables in the input domain have a common scale:

$$X_i' = \frac{X_i - \mu_{X_i}}{\sigma_{X_i}}, \qquad \forall X_i \in i = 1,2,\dots,n$$

Here, $X_i{}'$ denotes the standard normal transformation of input random variable X_i. Figure 4-15 depicts this transformation.

Find the MPP. Once this transformation has been performed for all the input random variables, determining the location of the MPP ($\boldsymbol{x}'^{*}$) involves solving the following inverse problem:

$$\boldsymbol{x}^{'*} = \operatorname{argmin} \sqrt{\boldsymbol{x}'^{T}\boldsymbol{x}'}$$
$$s.t.\, g(\boldsymbol{x}') = 0$$

The point $\boldsymbol{x}'^{*}$ is estimated by finding the values of $\boldsymbol{x}'$ that fall on the failure region boundary $(g(\boldsymbol{x}') = 0)$ and are the closest (minimum) distance to 0 (the mean of the transformed random variables). The inverse equation above can be solved by optimization methods such as the Rackwitz algorithm (Reference 4-44) and the Newton-Raphson recursive algorithm (Reference 4-45).

Calculate the distance from the MPP to 0. Once $\boldsymbol{x}'^{*}$ is found, its distance from 0 can be calculated as $\beta = \sqrt{\boldsymbol{x}'^{*T}\boldsymbol{x}'^{*}}$. A visualization of $\boldsymbol{x}'^{*}$ and β appears in Figure 4-15. The parameter β is the distance from **0** to the MPP $\boldsymbol{x}'^{*}$ and is known as the safety index for the reliability problem.

Estimate the failure probability. The failure probability p_f can then be approximated directly by making an assumption about the shape of the failure envelope. The simplest assumption, known as the FORM, is to assume the failure envelope is linear, as shown by the blue dashed line in Figure 4-15. In this case, the approximation of p_f becomes

$$p_f \approx P(l(\boldsymbol{x}) < 0) = \Phi(-\beta) = 1 - \Phi(\beta),$$

where $l(\boldsymbol{x})$ is the first-order Taylor series approximation of $g(\boldsymbol{x})$ about $\boldsymbol{x}'^{*}$ and Φ is the standard normal CDF. The first equality follows from the linear (and normal) assumption on $l(\boldsymbol{x})$ with parameters governed by the Taylor series about $\boldsymbol{x}'^{*}$. The second equality follows from the symmetry of the standard normal distribution. Figure 4-15 shows the linear approximation of the failure envelope using the FORM method.

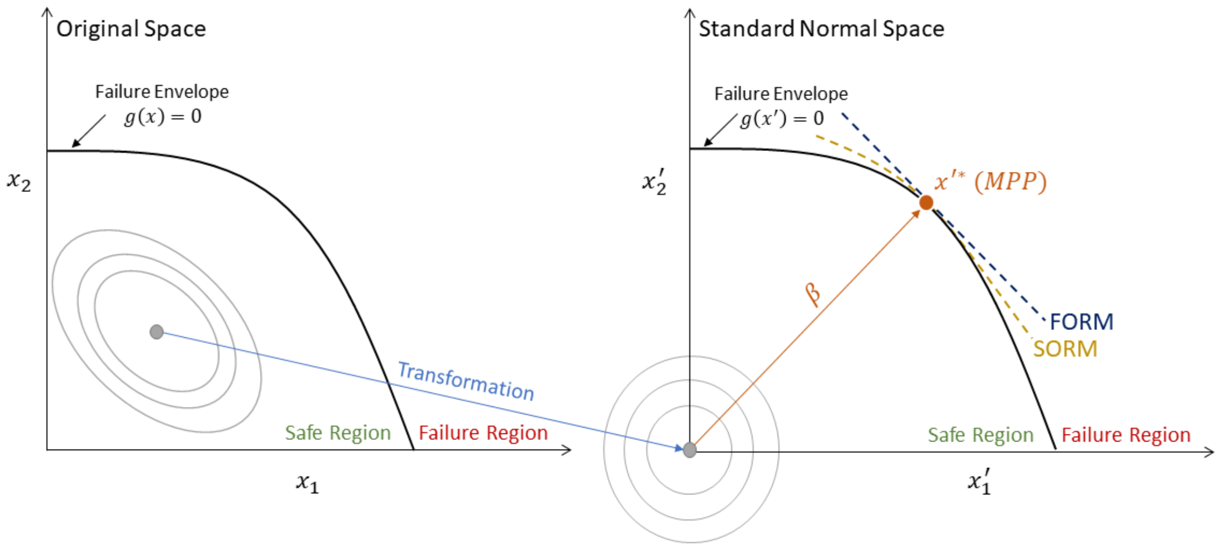

Figure 4-15 Example of the FORM and SORM Methods in the Standard Normal Space (Following Reference 4-42)

Nonnormally distributed inputs. One limitation of the approach above is that all input variables must be independent normal random variables in order for the approach to be valid. When they are instead independent nonnormal random variables, they can be transformed into approximate standard normal random variables by estimating equivalent normal distribution parameters $\mu_{X_i}^N$ and $\sigma_{X_i}^N$ for each nonnormal random variable X_i. The Rackwitz-Fiessler two-parameter equivalent normal transformation (Reference 4-46) achieves this by equating the PDF and CDF of variable X_i to the PDF and CDF of an equivalent standard normal distribution. Once this additional transformation is performed, the same inverse problem can be solved by the previously mentioned optimization methods to arrive at the MPP in equivalent standard normal space.

Approximating the failure region using SORM. Another limitation of the FORM approximation is that it may be overly conservative when the actual failure envelope is highly nonlinear. To improve upon this limitation, curvature of the limit state can be considered by also including the partial derivatives $\frac{\partial g}{\partial X_i}(\boldsymbol{x}^*)$ of the function g with respect to each X_i in the Taylor series expansion of the function g. These improvements to the approximation of the limit state function lead to an improvement to the approximation of p_f. Since the approximation is now a second-order Taylor series, the method is called the second-order reflexibility method (SORM). The yellow dashed line in Figure 4-15 provides a notional example of the curved approximation to the failure envelope using SORM. References 4-47, 4-48, and 4-49 provide further details.

4.3.5 Convergence Analysis

4.3.5.1 What Is It?

When propagating uncertainty forward through a model, there will be uncertainty in the estimate of the QoI due to the limited number of model realizations. The purpose of a convergence analysis is to assess the magnitude of sampling uncertainty associated with the QoI estimates obtained from Monte Carlo forward propagation of uncertainty (e.g., Sections 4.3.1, 4.3.2, and

4.3.3). Ultimately, an estimate has converged if the conclusions of the analysis do not change solely due to sampling uncertainty.

4.3.5.2 How to Use?

To conduct a convergence analysis, the analyst will take the following steps:

(1) Quantify sampling uncertainty with a metric.
(2) Compare the metric to a threshold value.

The threshold defines the maximum level of uncertainty acceptable for the application.

When Monte Carlo sampling (e.g., Sections 4.3.1, 4.3.2, and 4.3.3) is used to estimate a QoI, the following are three general methods for quantifying sampling uncertainty:

- **Calculate sampling uncertainty metrics for an estimate.** Section 4.3.6 covers closed-form sampling uncertainty metrics under SRS for probability estimates. Section 4.3.7 discusses statistical bootstrapping as an alternative to closed-form metrics. The metrics are calculated on a single simulation but require statistical assumptions that must be evaluated in practice.

- **Assess stability of a QoI estimate as the sample size increases.** The estimate of the QoI is monitored as the sample size grows to determine the appropriate sample size.

- **Compare QoI estimates over replicate simulations.** Several independent replicates of the model simulations are needed, which may not be feasible to implement in practice. The variation between these replicate simulations is assessed.

Section 4.3.5.4 discusses these methods in more depth. In general, the best method for convergence analysis depends on the computational complexity of the model as well as the type of sampling scheme.

4.3.5.3 When/Why?

In PFM, sampling uncertainty exists in estimates of QoIs. Rigorous assessment of the sampling uncertainty is conducted to ensure that the conclusions of the PFM analysis would not change solely due to random variations of estimates in different simulations.

4.3.5.4 Technical Details

Sampling uncertainty. Sampling uncertainty arises because the model can only be run for a finite number of realizations; a set of model realizations used to estimate a QoI such as a failure probability is called a model simulation. Replicate model simulations at different random seeds will produce different results. Consider the problem of estimating a rare probability. The histogram in Figure 4-16 displays estimates of this probability from many independent simulations; each simulation is based on n=10,000 model realizations sampled using SRS. This histogram represents the sampling distribution of the probability estimate, defined as the distribution of estimates obtained from repeated simulations. The true probability is 0.001, indicated by the red vertical dashed line, with estimates ranging from 0 to approximately 0.003. This range represents the sampling uncertainty. For a PFM analysis, this range could be acceptable or unacceptable, depending on the requirements of the analysis.

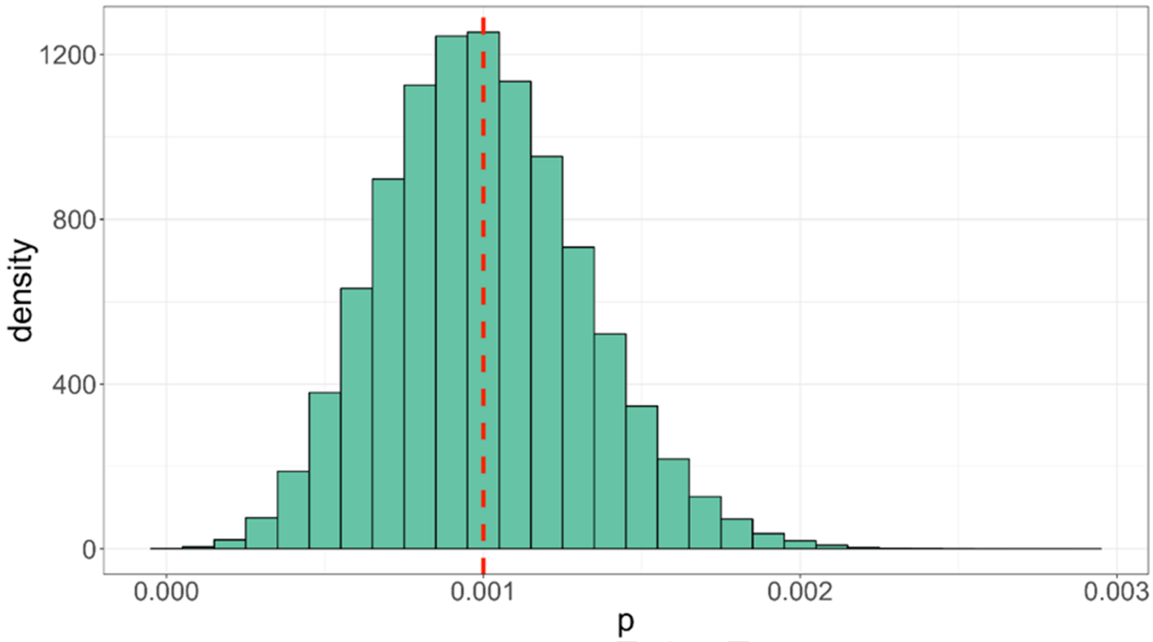

Figure 4-16 Histogram of Probability Estimates from a Simple Random Sample

Quantifying sampling uncertainty with a statistical metric. A convergence metric quantifies the sampling uncertainty in the estimate of a QoI, calculated using the output realizations. Convergence metrics can be compared to a prespecified threshold to determine whether the sample size is sufficiently large. Examples of statistical metrics to quantify sampling uncertainty in a convergence analysis include the following:

- *Standard error* is the standard deviation of the sampling distribution of the QoI. It is a measure of the variation in the estimate across repeated simulations.
- *Coefficient of variation (CV)* is the ratio of the estimated standard error of the QoI to the mean estimate of the QoI. The CV should only be used for a positive QoI, and it is not recommended if the mean estimate of the QoI is close to zero because the estimate of the CV can become very volatile.
- *Confidence interval* is an interval estimate of a QoI, providing a range of values for which we have high confidence that the true value of the QoI lies in the interval.

Sections 4.3.6 and 4.3.7 discuss methods for calculating these metrics based on a model simulation. These metrics are compared against predetermined thresholds to determine whether sampling uncertainty is sufficiently low. For example, the standard error or CV can be compared to a threshold defining the maximum acceptable value. The maximum acceptable width of a confidence interval is another possible threshold (Reference 4-50).

Assessing the stability of an estimate as the sample size increases. One common method for assessing model convergence is incrementally increasing the sample size and examining the stability of the QoI estimate as a function of sample size. As the sample size increases and sampling uncertainty decreases, the estimate of the QoI will stabilize. Metrics for QoI stability include the standard error, CV, and confidence intervals, all of which are calculated from the sample.

Figure 4-17 depicts an example demonstrating the convergence of estimating a small probability ($1x10^{-3}$) using SRS. The simulation was run for $1x10^{7}$ iterations, and the x-axis is plotted on the log scale. The estimated probability is plotted as the black line in the figure. This estimate is 0 until a sampling size of about $1x10^{3}$. Then it is volatile until a sample size of around $1x10^{6}$, where it begins to converge to the true value. A two-sided 95-percent confidence interval, represented by the red dashed lines, provides a convergence metric. This bound was constructed using the Clopper-Pearson confidence interval (References 4-50, 4-51). Suppose that the threshold for model convergence is met when the 95-percent confidence interval has width less than $1x10^{-4}$. It takes 1,516,000 samples to satisfy the metric in this case.

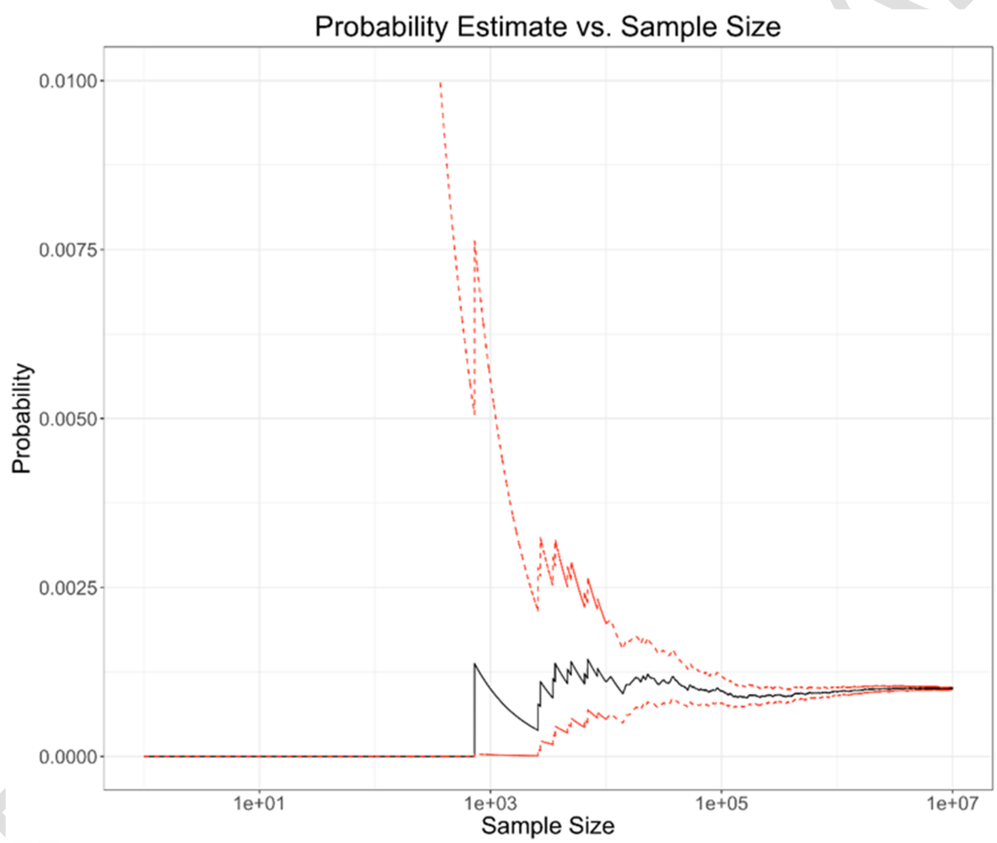

Figure 4-17 Confidence Interval used to Assess the Convergence of a Probability Estimate

Comparing estimates over replicate simulations. A more computationally expensive approach to assessing convergence of a QoI is to conduct many replicate simulations to repeatedly estimate a QoI and then directly estimate variability in the QoI across replicates. The variability in the QoI estimate across simulations provides information for sampling uncertainty. The advantage of this method is that it is easy to apply to any sampling scheme. The disadvantage is that conducting replicate simulations is computationally expensive. To determine whether a sample of size n is sufficient, a total of nr realizations is computed, where r is the number of replications of the simulation. The specific sample size and reasonable number of resamples depend on the application. The sample size of each replicate set should be close to that of the empirical data. Further, these samples can later be combined to produce a more precise final estimate of the QoI.

This sampling uncertainty can be quantified using different metrics, such as the standard deviation of the QoI estimates, the CV, or a statistical prediction interval for future QoI estimates. A prediction interval is similar to a confidence interval and provides interval bounds such that there is a high level of confidence that a new QoI estimate would lie in this range. An approximate $100(1-\alpha)\%$ confident prediction interval for a normally distributed random variable is

$$\bar{x} \pm t_{\alpha/2,r-1}\sqrt{s^2\left(1+\frac{1}{r}\right)},$$

where $\bar{x}$ is the average of the QoI estimates, s is the standard deviation of the QoI estimates, and $t_{\frac{\alpha}{2},r-1}$ is the $(\alpha/2)th$ percentile of the Student-t distribution with $r-1$ degrees of freedom. Note that, to compute a prediction interval, the distribution of the QoI estimate must be known. Often, the QoI is an average of many model realizations such that QoI estimates will be approximately normally distributed (based on the central limit theorem). When the QoI is a rare probability, this normal approximation can perform poorly, and normal prediction intervals should be interpreted with caution.

As an example, suppose we are measuring sampling variability in an estimate of the probability of an axial crack in a pipe over 60 years. Figure 4-18 plots replicate QoI estimates as a function of time based on $r = 5$ replications. Figure 4-19 shows the two-sided 95-percent prediction interval for this example. The width of the prediction interval as a function of time can be compared to a predetermined threshold on the acceptable maximal width to assess convergence. The choice of $r = 5$ should be justified. The more replicates the better, and one can assess the stability of the estimated standard deviation as r increases. Second, the chosen threshold (width of the prediction interval) is important. In more typical PFM analyses where the probability of the event is much lower, the threshold will be much more difficult to achieve than in this simple example.

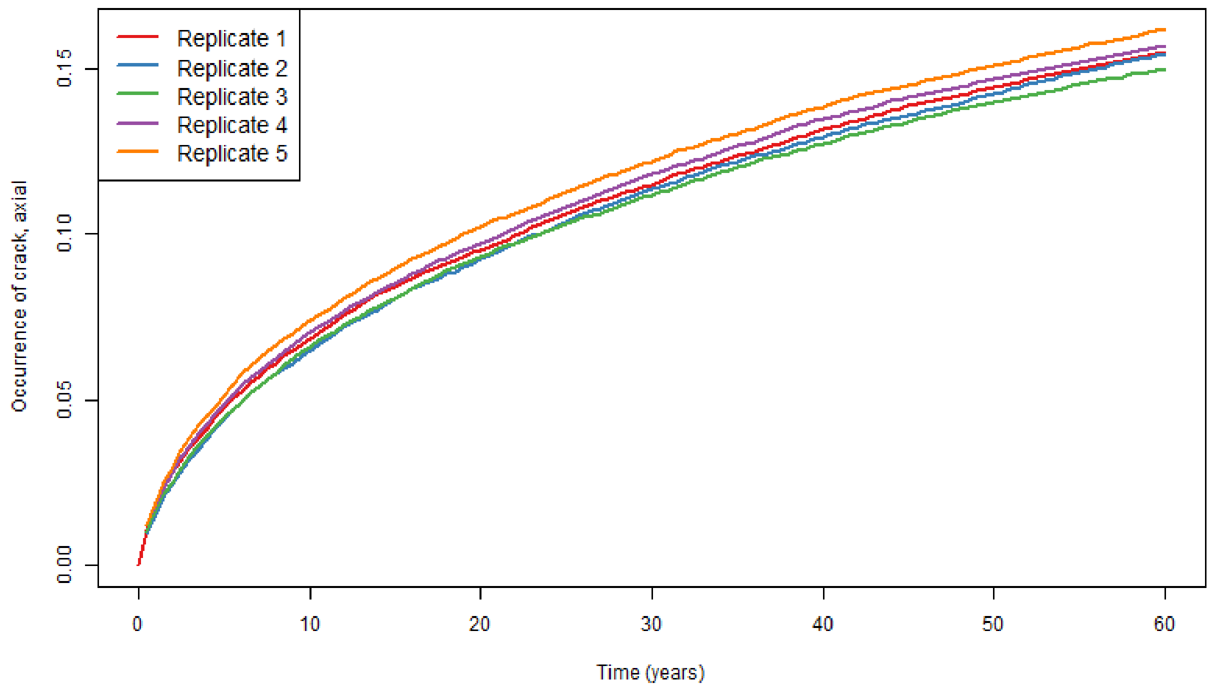

Figure 4-18 Estimates of the Probability of Axial Crack for r=5 Independent Replications Using the Same Sampling Scheme

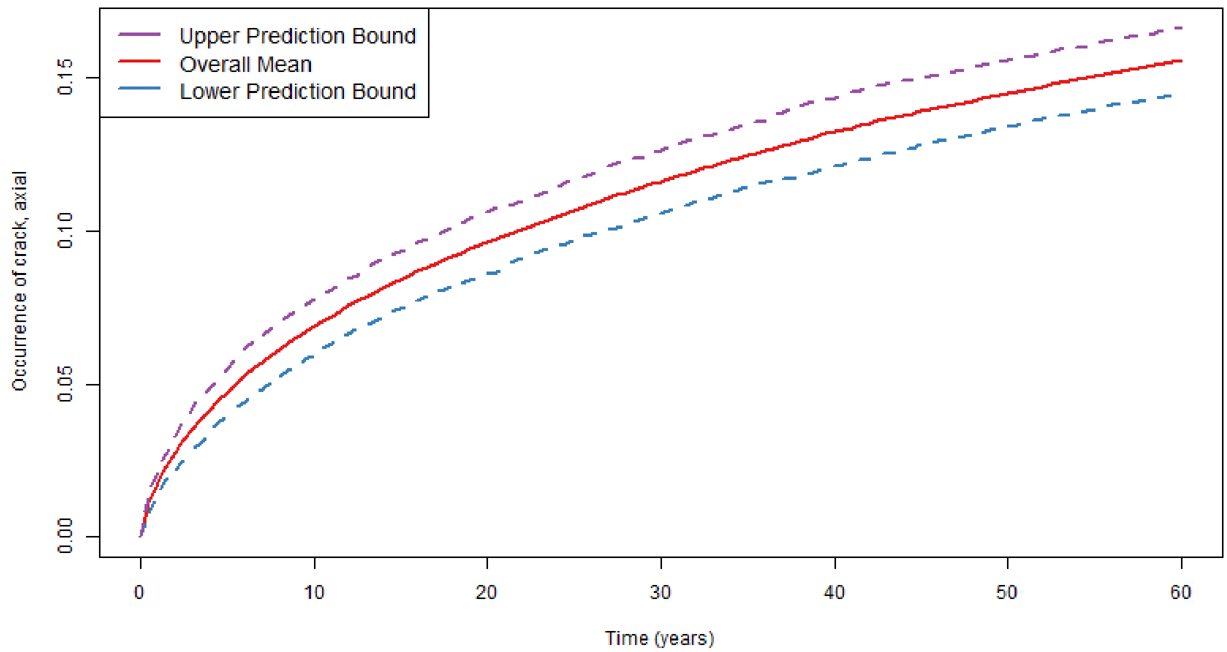

Figure 4-19 Prediction Interval Computed from the Five Independent Simulations

4.3.6 Closed-Form Metric for Simple Random Sampling Uncertainty in a Probability Estimate

4.3.6.1 What Is It?

When estimating the probability of an event (e.g., some failure scenario of interest), the sampling uncertainty in the estimate should be well understood to determine model convergence (Section 4.3.5). Both the sample size and the rarity of the event under consideration influence the accuracy of the estimate. This section provides a sampling uncertainty metric for a probability estimate when SRS is applied in uncertainty propagation.

4.3.6.2 How to Use?

Computing this sampling uncertainty metric involves the following three steps:

(1) Propagate an SRS of size n from the inputs through the model. Record the number of events and nonevents.

(2) Estimate the probability of the event using the total number of recorded events divided by the sample size.

(3) Compute a sampling uncertainty metric, such as the standard error, the CV, or confidence interval (see Section 4.3.5).

When LHS or adaptive sampling methods are used, care should be taken when estimating closed-form metrics for sampling uncertainty since assumptions may be violated. Sampling uncertainty is still present when these sampling algorithms are used and should be assessed using alternative approaches (see Sections 4.3.5 and 4.3.7).

4.3.6.3 When/Why?

When SRS is used for forward propagation of uncertainty, these metrics can be computed to quantify sampling uncertainty in the probability estimate, providing useful insight about the precision of the estimate. Results may suggest that a larger sample or other variance reduction techniques (see Sections 4.3.2 and 4.3.3) are needed if the precision is insufficient.

4.3.6.4 Technical Details

Estimate the event probability. After propagating inputs sampled using SRS through the model and recording whether the event occurred, the probability of failure p_f can be estimated by the ratio of number of failures (n_f) to the number of trials (n), known as a binomial proportion

$$p_f \approx \hat{p}_f = \frac{n_f}{n}.$$

Compute the standard error and CV. The sampling uncertainty of $\hat{p}_f$ relative to p_f decreases as $n \to \infty$ and $p_f \to 1$. That is, increasing n will decrease the sampling uncertainty in the estimate, but the relative decrease depends on the failure probability, with smaller p_f resulting in larger relative sampling uncertainty. References 4-52 and 4-53 explain SRS and its associated uncertainty in estimation in detail.

Assuming each 0/1 outcome is independent, the number of failures n_f can be assumed to follow a binomial distribution. Based on the binomial distribution, the estimated standard error of $\hat{p}_f$ is

$$\hat{\sigma}_{\hat{p}_f} \approx \sqrt{\frac{(1-\hat{p}_f)\hat{p}_f}{n}}$$

The accuracy of this approximation increases as $n/p(1-p)$ gets large. Given the estimated standard error, the CV is

$$CV(\hat{p}_f) = \frac{\sigma_{\hat{p}_f}}{\mu_{\hat{p}_f}} = \frac{\sqrt{\frac{(1-p_f)p_f}{n}}}{p_f}$$

where $\mu_{\hat{p}_f}$ and $\sigma_{\hat{p}_f}$ are the mean and standard error of the $\hat{p}_f$, respectively.

The CV highlights the fact that the relative uncertainty in a probability estimate $\hat{p}_f$ can be quite large, especially when the target p_f is small. For example, for 10,000 simulations of an event with $p_f = 0.01$, $\hat{\sigma}_{\hat{p}_f}$ is about 0.001 (or 10 percent of the desired estimate), but if $p_f = 0.001$, with 10,000 simulations $\hat{\sigma}_{\hat{p}_f}$ is about 0.00032 (32 percent of the desired estimate).

Compute a confidence Interval. Statistical confidence intervals provide a plausible range in which a parameter is likely to fall based on the observed data. There are many methods for computing confidence intervals for p_f; Reference 4-54 discusses several in detail. A commonly used approximate $100(1-\alpha)\%$ confidence interval for a binomial proportion is

$$\hat{p}_f \pm z_{\alpha/2}\, \hat{\sigma}_{\hat{p}_f}$$

where $z_{\alpha/2}$ is the $(\alpha/2)th$ percentile of the standard normal distribution. This confidence interval relies on approximate normality of $\hat{p}_f$, which is valid only if $\hat{p}_f$ is not too close to 0 or 1. A rule of thumb is to use this interval only if $n\hat{p}_f > 5$ and $n\left(1-\hat{p}_f\right) > 5$; that is, at least five failures and nonfailures are observed. In PFM applications where the true probability of failure is very small, this confidence interval is unlikely to perform well since the number of observed failures under SRS will often be very small. In addition, when np_f is small, zero failures may be observed, and the interval above is meaningless. Reference 4-54 outlines several alternative confidence intervals for binomial proportions. The next paragraph outlines one method for bounding the probability of failure when no failures are observed.

Confidence interval when no failures are observed. If $n_f = 0$ failures are observed in n realizations, then we can use the fact that n_f follows a binomial distribution to place a one-sided confidence interval on the probability of failure. Specifically, there is 100(1-α)% confidence that $p_f < p_u$, where

$$p_u = 1 - \alpha^{\frac{1}{n}}$$

For example, if it must be established that $p_f < 10^{-6}$ with 95-percent confidence, a simple random sample of size n=$\log(.05)\,/\,\log(1-10^{-6}) \approx 3 \times 10^6$ with no observed failures is needed. References 4-55 and 4-56 provide more details.

4.3.7 Statistical Bootstrapping

4.3.7.1 What Is It?

Statistical bootstrapping is a flexible statistical method for calculating sampling uncertainty in a QoI estimate. Bootstrapping relies on resampling from the observed data to calculate QoI uncertainty and is particularly useful when closed-form metrics (as described in Section 4.3.6) are difficult or impossible to derive.

While there are many versions of bootstrapping, the general idea is to repeatedly resample from the observed data, each time estimating the QoI. The variability in the QoI estimates across bootstrap resamples provides a measure of sampling uncertainty.

4.3.7.2 How to Use?

The most common bootstrap method is to resample directly from the observed data. This form of bootstrapping has three steps:

(1) Take a sample from the observed data. The sample size is the same size as the observed data. The sample is taken with replacement, where single observations in the

data can be included multiple times in a single bootstrap resample. The sampling at this step should be consistent with the way the data were generated.

(2) Calculate the QoI from the sampled data.

(3) Repeat steps 1 and 2 many times. Use the collection of calculated QoIs to approximate the sampling uncertainty in the QoI. For example, the standard deviation of the collection of QoIs is an estimate of the standard error.

4.3.7.3 When/Why?

Bootstrapping offers a flexible method for estimating sampling uncertainty. The following are the main reasons to use bootstrapping:

- The algorithm is generic, so it can be applied to most QoIs and many sampling schemes.
- The algorithm is simple to implement, requiring only the ability to resample from the data and repeatedly calculate the QoI.
- Closed-form metrics for sampling uncertainty (e.g., Section 4.3.6 for a probability estimate) are difficult to estimate without violating assumptions in many cases.

However, it is important to understand when not to use bootstrapping. The bootstrap will result in inaccurate measures of sampling uncertainty when either of the following is true:

- The sample size is small (i.e., sparse data).
- The original sample was drawn using a complex sampling scheme that cannot be resampled (e.g., LHS).

Section 4.3.7.4 contains more information about conditions for bootstrap failure.

4.3.7.4 Technical Details

Figure 4-20 depicts the steps of the bootstrap. The left plot is a histogram of the original data. The middle histograms displayed vertically represent B different bootstrap samples of the original data. Each of these has the same sample size as the original data. However, these samples are taken with replacement, meaning that some values may be observed more than once. The variation among these histograms is an estimate of the sampling variation of the observed sample. For each of the bootstrap samples, an estimate of the QoI is computed and is aggregated in the histogram on the right. This right-most histogram is an estimate of the sampling uncertainty in the estimate of the QoI. The vertical dashed lines are the 0.025 and 0.975 quantiles of the bootstrap QoI estimates and correspond to an estimate of a 95-percent confidence interval.

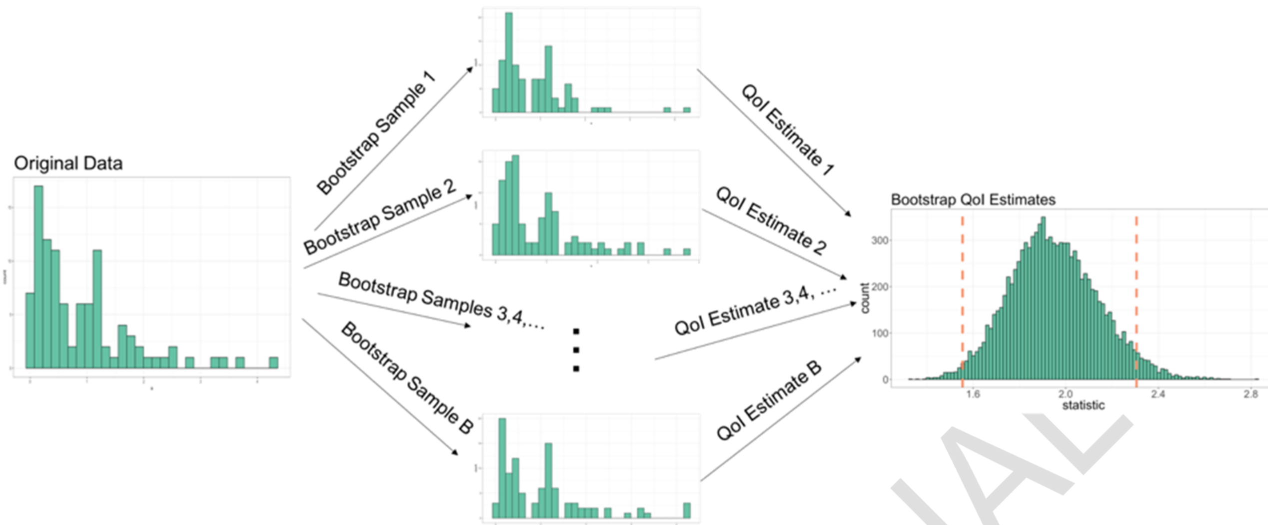

Figure 4-20 Visualization of the Steps Taken for the Standard Statistical Bootstrap

The following are the most common ways the bootstrap fails:

- Data are too sparse.
- The resampling does not reflect how the data were generated.

The bootstrap can underestimate uncertainty when data are sparse. Specifically, standard errors will be too small, and confidence intervals will be too narrow. In PFM applications, sparse data are likely to occur when the number of computer realizations is small relative to the QoI. More samples are needed for estimating rare event probabilities and extreme percentiles, because the sparsity of the data is not judged based on the overall number of model realizations but on the overall number of events of interest that occur. Therefore, we may need many more than 1×10^6 simulated realizations to accurately quantify uncertainty in a 1×10^{-6} probability.

Further, the bootstrap will not accurately estimate sampling uncertainty unless, in the resampling step, the resampling reflects how the data were generated. In PFM applications, the bootstrap can be used with both simple random samples and importance sampling. The bootstrap cannot provide accurate uncertainty quantification in complex sampling schemes such as LHS because there is no way to resample from the observed data in a way that approximates the original LHS scheme.

More technically, the major assumption of the bootstrap is that, by resampling from the data, we are constructing samples that approximate the empirical distribution of the data. When data are sparse, we cannot approximate this distribution well. When data are generated from a complex sampling scheme such as LHS, we cannot resample from the data in a way that approximates the empirical distribution of the original sample obtained using LHS.

Bootstrap confidence intervals. Confidence intervals are often desired to provide a plausible range in which a parameter is likely to fall based on the sampled data (see Section 4.3.5). Commonly, a probability distribution for the observed data is assumed (either through fitting to data or by expert judgment), and confidence intervals can be derived directly from this assumption. If the choice of probability distribution does not have a strong basis, the bootstrap is an alternative approach as it bypasses the need to analytically derive confidence intervals using an assumed probability distribution.

For example, suppose the QoI is the mean μ from a population from which a sample of data of size n is collected: $x^{(1)}, \ldots, x^{(n)}$. If it is assumed the population is normally distributed, then the analytically derived $100(1-\alpha)\%$ confidence interval is

$$\bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}},$$

where $\bar{x}$ and s are the sample mean and standard deviation and $t_{n-1,\alpha/2}$ is the $\alpha/2$ quantile of a t-distribution with $n-1$ degrees of freedom.

The nonparametric bootstrap approach to the above problem takes a sample of the data of size n with replacement B times (commonly 1,000 or more), each time computing the sample mean. This procedure results in a collection of sample means from which confidence intervals can be constructed. The simplest, but often least accurate, approach to constructing bootstrap confidence interval for a QoI θ is using empirical quantiles of the bootstrap distribution

$$({\theta^*}_{\alpha/2}, {\theta^*}_{1-\alpha/2}),$$

where ${\theta^*}_{\alpha/2}$ is the $\alpha/2$th percentile of the bootstrap distribution of θ. Another approach is the basic method, defined as

$$(2\hat{\theta} - \theta^*_{1-\alpha/2}, 2\hat{\theta} - \theta^*_{\alpha/2}),$$

where $\hat{\theta}$ is the estimate of θ from the original sample and $\theta^*_{1-\alpha/2}$ is the $1-(\alpha/2)$th percentile of the bootstrap distribution for θ. Alternatively, the Studentized method for estimating a confidence interval can be calculated as

$$(\theta - t^*_{1-\alpha/2}\widehat{se}_\theta, \theta - t^*_{\alpha/2}\widehat{se}_\theta),$$

where $t^*_{1-\alpha/2}$ is the $1-(\alpha/2)$th percentile of the bootstrapped Student's t-test $t^* = (\hat{\theta}^* - \hat{\theta})/\widehat{se}_\theta$. Here, $\hat{\theta}^*$ and $\widehat{se}_\theta$ are the estimate and the standard error, respectively, of the bootstrap distribution of θ. There are many other ways to construct bootstrap intervals, each with their own advantages and disadvantages. References 4-57 and 4-58 provide more information.

When no closed-form expression is available for a confidence interval, bootstrapping is often a simple solution for obtaining a confidence interval. As an example, suppose the QoI was some function of the population mean and standard deviation, such as $(\mu^2 + \mu)/\sigma$. For each bootstrap sample, the estimated QoI $(\bar{x}^2 + \bar{x})/s$ is computed.

As an example, Figure 4-21 shows the bootstrap distribution of this statistic from a sample of size 100 of data from a normal distribution with mean and variance 1. Using the bootstrap distribution, a confidence interval for $(\mu^2 + \mu)/\sigma$ can be easily calculated (see References 4-57 and 4-58). The vertical dashed lines in the figure show a 95-percent bootstrap interval. In this example, the true value of $(\mu^2 + \mu)/\sigma$ is known to be 2, which is clearly within the bootstrap confidence interval.

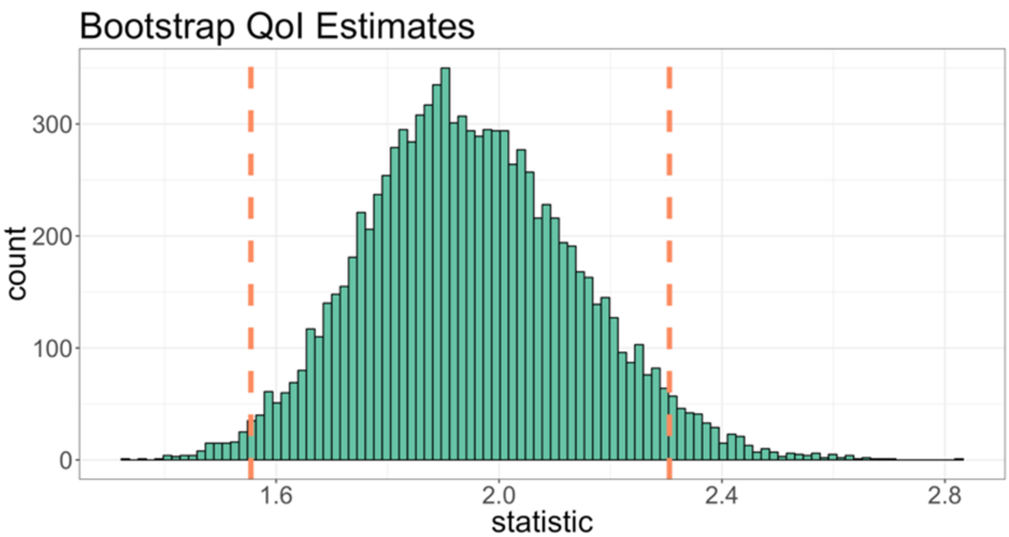

Figure 4-21 Bootstrap Sampling Distribution along with a 95-Percent Confidence Interval for the Complex Estimator Example

4.3.8 Global Sensitivity Analysis

4.3.8.1 What Is It?

Sensitivity analysis (SA) seeks to answer a fundamental question: how sensitive is a model to its input parameters and which inputs are most important (Reference 4-59)? SA can be used to identify the inputs that have the strongest impact on the outputs (i.e., most sensitive or important inputs). Further, SA can help understand the nature of the input-output relationship. Global SA is used to quantify the amount of output uncertainty that can be attributed to uncertainty in the input variables (Reference 4-59).

4.3.8.2 How to Use?

Before performing a SA, it is important to choose a relevant output to analyze. The output should be closely related to the QoI. Further, binary/categorical outputs inherently contain less statistical information than continuous outputs. Frequently, the binary output is a function of continuous outputs, and these continuous outputs can often provide better information on input sensitivity with fewer samples. Because of this, it is generally beneficial to use continuous outputs for SA when possible.

After an output has been chosen, an SA can be performed using exploratory data analysis and global sensitivity metrics estimation:

- **Exploratory data analysis.** Exploratory data analysis summarizes characteristics of the input-output relationships using summary statistics and visualizations (Reference 4-60). Perhaps the most useful visualization to understand the relationship between PFM inputs and outputs is a scatterplot. If the number of input and output variables is small, scatterplots can be produced for each output with each input. With many inputs and outputs, relevant visualizations may be chosen based on subject matter knowledge. Alternatively, one can estimate the global sensitivity metrics first and use these to choose the visualizations.

- **Global sensitivity metrics estimation.** Sensitivity metrics provide a quantitative value that characterizes the relationship between inputs and outputs. The following two metrics can be used to quantify the input/output relationship (References 4-59, 4-61, 4-62):
  - First-order sensitivity indices refer to the proportion of the variance in the output that is explained by the variance in a single input.
  - Total-order sensitivity indices refer to the proportion of the variance in the output that is explained by the variance in an input and its interactions with other inputs.

Section 4.3.8.4 includes details about estimating these sensitivity metrics.

4.3.8.3 When/Why?

SA can be performed to achieve the following:

- Understand the problem drivers and rank inputs based on the magnitude of their effect on the output(s).
- Improve the precision and accuracy of uncertainty propagation by doing the following:
  - identifying important inputs whose uncertainty distributions may need further refinement
  - determining candidate inputs for importance sampling

4.3.8.4 Technical Details

Exploratory data analysis. Scatterplots can be used to visually assess the nature and magnitude of the relationship between an input and output.

Figure 4-22 shows an example of an input without (left) and with (right) a strong relationship with the response. For important inputs, scatterplots can also be used to determine whether the relationship is roughly linear, monotonic (i.e., entirely increasing or decreasing), or more complex.

Figure 4-23 shows examples of linear (left), nonlinear/monotonic (middle), and nonlinear/nonmonotonic (right) input/output relationships. Reference 4-63 gives formal procedures for analyzing scatter plots. In practice, it can be difficult to visually inspect a large number of scatter plots, and more complex relationships involving interactions can often be missed. Estimating sensitivity metrics can help identify the most important relationships to visualize.

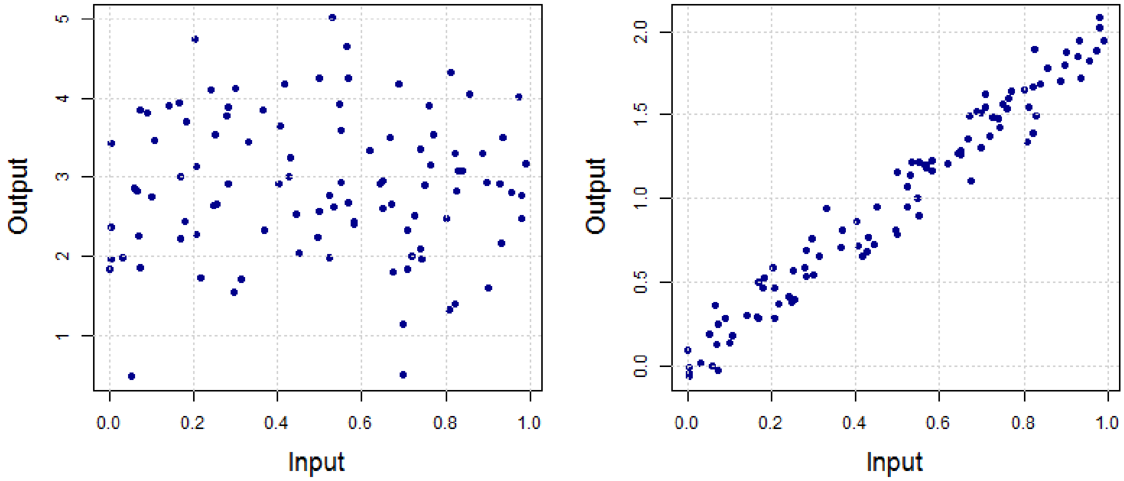

Figure 4-22 Scatterplots Showing an Input Without (Left) and with (Right) a Significant Relationship with the Output Variable

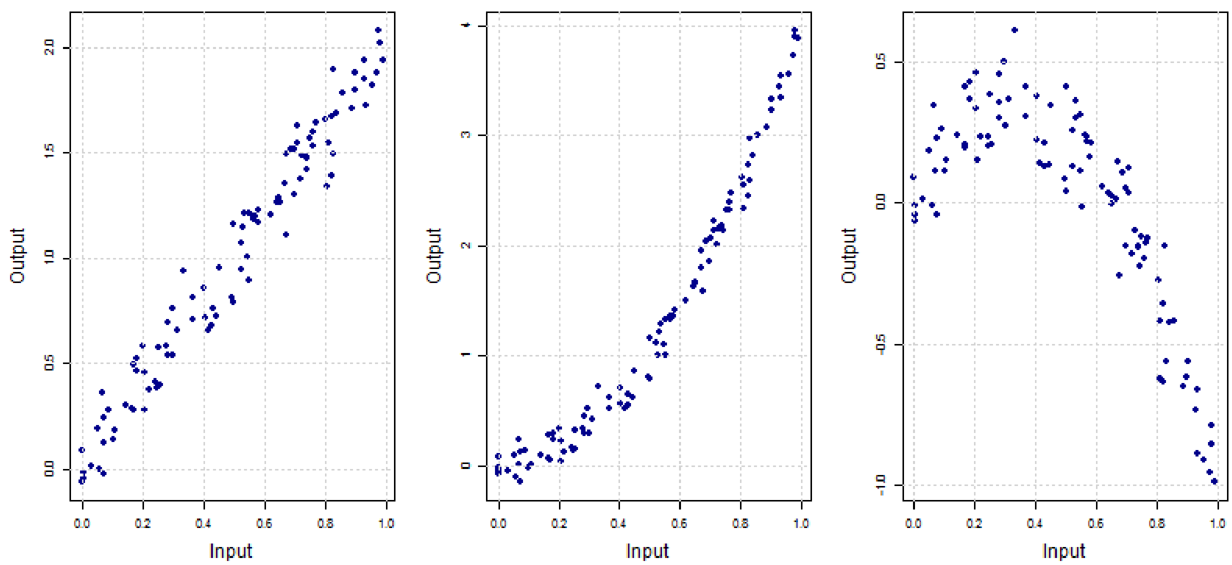

Figure 4-23 Scatterplots Showing Linear (Left), Nonlinear/Monotonic (Middle), and Nonlinear/Nonmonotonic (Right) Relationships between the Input and Output Variables

Global sensitivity metrics estimation. Input sensitivity can be measured in a variety of different ways (References 4-64 and 4-65). Variance-based indices are common sensitivity metrics that decompose the output variance and attribute this variance to certain inputs. Heuristically, the first-order sensitivity index reflects the proportion of the total output uncertainty that is explained by the uncertainty in the input x_j alone. The total effects sensitivity index reflects the fraction of the output uncertainty that is explained by x_j by itself and together with its interaction with other variables.

Mathematically, the first-order and total effects sensitivity indices can be described as follows. Suppose the output of the computer model is

$$y = f(\boldsymbol{x}), \tag{3}$$

where y is the model output, $\boldsymbol{x} = [x_{1,} \dots, x_p]$ is a vector of p input variables, and f is the model. The first- and total-order sensitivity indices for an input x_j (denoted S_j and T_j, respectively) are defined as

$$S_j = \frac{Var(E[f(\boldsymbol{x})|x_j])}{Var(f(\boldsymbol{x}))}, \tag{4}$$

$$T_j = \frac{E(Var[f(\boldsymbol{x})|\boldsymbol{x}_{(-j)}])}{Var(f(\boldsymbol{x}))} = 1 - \frac{Var(E[f(\boldsymbol{x})|\boldsymbol{x}_{(-j)}])}{Var(f(\boldsymbol{x}))}, \tag{5}$$

where $\boldsymbol{x}_{(-j)}$ is a vector of all p input variables, excluding the j^{th} input (Reference 4-59). The numerator of the first-order sensitivity metric in Eq. 4 is the variance of $E[f(\boldsymbol{x})|x_j]$, the average value of the output $f(\boldsymbol{x})$, conditional on the input of interest x_j. This variance is taken with respect to the distribution on x_j; therefore, the numerator measures how much the average output varies as x_j varies. A large variance indicates x_j affects the output $f(\boldsymbol{x})$ and a small variance indicates it does not. The meaning of "large" and "small" is relative to the total variation of the output, the denominator of Eq. 4.

S_j reflects the proportion of the total output uncertainty that is explained by the uncertainty in x_j alone, though a similar metric, T_j, is used to assess the proportion of uncertainty in the output explained by x_j and its interactions with other variables. The numerator in the total-order sensitivity metric in Eq. 5 is the expectation (average) of $Var[f(\boldsymbol{x})|\boldsymbol{x}_{(-j)}]$, the variance of the output $f(\boldsymbol{x})$ given all but the j^{th} input. If there is high variation for a wide range of $\boldsymbol{x}_{(-j)}$, then the outer expectation will be large, resulting in a large value for T_j. If T_j is much larger than S_j, it implies that there are significant interactions between x_j and the other inputs.

Estimating sensitivity metrics using surrogate models. The calculation of first- and total-order sensitivity indices involves the estimation of high-dimensional integrals representing the expectations and variances in Eq. 4 and Eq. 5. The Monte Carlo integration approaches detailed in References 4-59 and 4-66 can be used to estimate these indices. However, this can be computationally prohibitive for many applications because it requires a large number of realizations when there are a large number of inputs. To address this problem, surrogate models can be used to estimate the indices (References 4-67 and 4-68). Section 4.3.10 contains more information on surrogate models.

Active subspaces. At times, the dimensionality of the input space may be prohibitively large for performing SAs. If this is the case, dimensionality reduction methods such as active subspaces may be useful. Active subspaces are a way to identify important directions of the input space that affect the QOI. Directions that are not important can be ignored, resulting in reduced dimensionality. References 4-69 and 4-70 provide more information on active subspaces.

4.3.9 Local Sensitivity Analysis

4.3.9.1 What Is It?

Local SA specifically focuses on how changes to each input at or near a specific reference point in the input domain, like a mean or median, affect outputs of interest (Reference 4-71). Alternatively, global SA attempts to quantify the effects of the uncertain inputs on the output relative to the entire input space (Section 4.3.8).

4.3.9.2 How to Use?

Local SA determines the rate of change of a specified output with respect to a given model input. The aim is to compute the partial derivative with respect to the input at a specific point in the domain (input space). One method for computing this partial derivative for a single input involves the following steps:

(1) Run the model at the specified value of the input.
(2) Perturb the input and run the model again.
(3) Measure the change in the output by estimating the partial derivative.

Typically, the other inputs remain fixed during this process, and the measured change in the output is attributed to a single input, conditional on the values of the other inputs.

4.3.9.3 When/Why?

Local SA is a relatively efficient first step toward learning about the important parameters in a model. With only two evaluations of a model, the linear effect on an output of changing a single parameter can be estimated. This step can help down-select to a smaller set of parameters to study in a full uncertainty analysis. It also provides some physical intuition for how certain parameters affect the output. However, the local nature of this analysis should always be kept in mind because (1) a parameter with low local sensitivity can still have a major effect on an output of interest if its associated uncertainty is large, and (2) the local sensitivity of a parameter can sometimes change significantly over the domain of interest. Global SA informs the effect of the parameter over its full uncertainty range and across the entire domain.

4.3.9.4 Technical Details

Calculating local sensitivity metrics. Local SA only requires a small number of model evaluations. First, a nominal input value is chosen and that value is perturbed by some amount in one direction (i.e., perturb one dimension of the input space). The amount chosen should be large enough so that a significant change in the output can be observed, but it should be small enough to stay within the region of the input space of concern.

The sensitivity is measured by the partial derivative, which is estimated by a finite difference. For example, for some output of interest h, the sensitivity of a single input x_i at $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)$ is approximated by perturbing x_i by an amount δx_i and approximating the partial derivative with the finite difference:

$$\frac{\partial h(\boldsymbol{x})}{\partial x_i} \approx \frac{h\left(x_1, \dots, x_i + \delta x_i, \dots, x_p\right) - h(\boldsymbol{x})}{\delta x_i} \tag{6}$$

Partial derivatives can be compared across a set of inputs by repeating Steps 1–3 in Section 4.3.9.2, perturbing a single input each time. Approximating the partial derivative with a finite difference is effectively a polynomial approximation using a Taylor series expansion around the reference point (Reference 4-71, 4-72). Note that the reference point (the model realization at the nominal input values) can be reused for the computation for each input. From the results of the local SA, the inputs can be ranked in terms of their contribution to an output of interest.

4.3.10 Surrogate Models

4.3.10.1 What Is It?

Surrogate models (also known as emulators, metamodels, and response surfaces) are relatively fast statistical models that approximate more complex computer models. Surrogates are less computationally expensive to evaluate than the computer model and can be useful for SA and uncertainty propagation when conducting a sufficient number of computer model realizations is computationally prohibitive.

4.3.10.2 How to Use?

Surrogate models are constructed using the following steps (with more information in Section 4.3.10.4):

(1) Generate training data by running the computer model using several sets of input values and obtaining the corresponding output values.

(2) Use the training data to construct the surrogate model.

(3) Validate the surrogate on a new set of computer model realizations (testing data) to check its quality.

A surrogate can be used to approximate the full computer model for SA and uncertainty propagation. The choice of surrogate is dependent on the assumptions the user is willing to make, which relate to the type of output and the complexity of the input/output relationship. Two different output types are commonly seen in PFM models:

(1) **Continuous data** can take on an infinite number of possible (physical) values (e.g., crack length). Common surrogates for continuous data include linear regression, multivariate adaptive regression splines, and GP regression.

(2) **Binary data** take on only two levels for the output. Typically, the binary variable is an indicator for an event, taking on 0 if the event did not occur and 1 if it did (e.g., rupture or no rupture). Surrogates for binary data model the probability of the event occurring. An example surrogate for binary data is a generalized linear model.

4.3.10.3 When/Why?

Surrogate models can be used to decrease computation time through building a computationally efficient computer model approximation. In PFM applications, surrogates can be used in SA and uncertainty propagation:

- In SA, surrogates can be used to determine how uncertainty in the inputs affects uncertainty in the outputs (Sections 4.3.8 and 4.3.9).
- In uncertainty propagation, surrogates can be used for propagating uncertainty in the inputs through the computer model (Section 4.3). Input samples are propagated through the surrogate model rather than the full computer model to allow for many more evaluations.

Surrogate models approximate the computer model, and this approximation adds uncertainty in the PFM analysis. Surrogate uncertainty should be considered in the interpretation of the results under the following conditions:

- If surrogates are used for SA, several different surrogates can be tried to explore the sensitivity of the SA results to the selected surrogate model.
- If surrogates are used for uncertainty propagation, the magnitude of error associated with the surrogate model approximation can be quantified and included as additional uncertainty in the estimation of the QoI.

4.3.10.4 Technical Details

Generate training data. An output y of a physical process or computer model can be represented as a function of the input $\boldsymbol{x}$:

$$y = f(\boldsymbol{x}).$$

The representation here is deterministic; given the same value of $\boldsymbol{x}$, the same value of the output y will result. A surrogate estimates the true process function f statistically using a set of training data $(y^{(i)}, \boldsymbol{x}^{(i)})$, $i = 1, 2, \ldots, n$ where $\boldsymbol{x}^{(i)}$ is the i^{th} set of inputs on which a computer model of the process is evaluated resulting in the output $y^{(i)}$.

The accuracy of the surrogate increases with the size of the training set. For continuous outputs, a general rule of thumb for the number of data points n is approximately 10*p*, where *p* is the number of input variables (Reference 4-73). Reference 4-74 gives an overview of options for choosing the input combinations that will be used in constructing the surrogate model. A useful and common choice is an LHS (Section 4.3.2). This section outlines several options for constructing a surrogate.

Construct the surrogate model. The training data are used to fit a statistical model approximating $f(\boldsymbol{x})$ for any $\boldsymbol{x}$**.** The choice of surrogate model to use will depend on several aspects of the problem, such as the type of output variable (e.g., continuous or binary), continuity or discontinuity of f, the size of the training data set, and the domain on which a surrogate is required. This section discusses several types of surrogate models. Ideally, uncertainty in the surrogate model predictions are measured. An example surrogate appears in Figure 4-24 where the black points represent the training data, the blue curve represents the true computer model $f(\boldsymbol{x})$ across the entire input space, the red curve is the surrogate estimate

$\hat{f}(x)$, and the gray curves represent statistical uncertainty in the surrogate estimate. This particular surrogate, a GP, interpolates the training data and has the intuitive property that statistical uncertainty is larger for locations farther away from the training data (e.g., Reference 4-75).

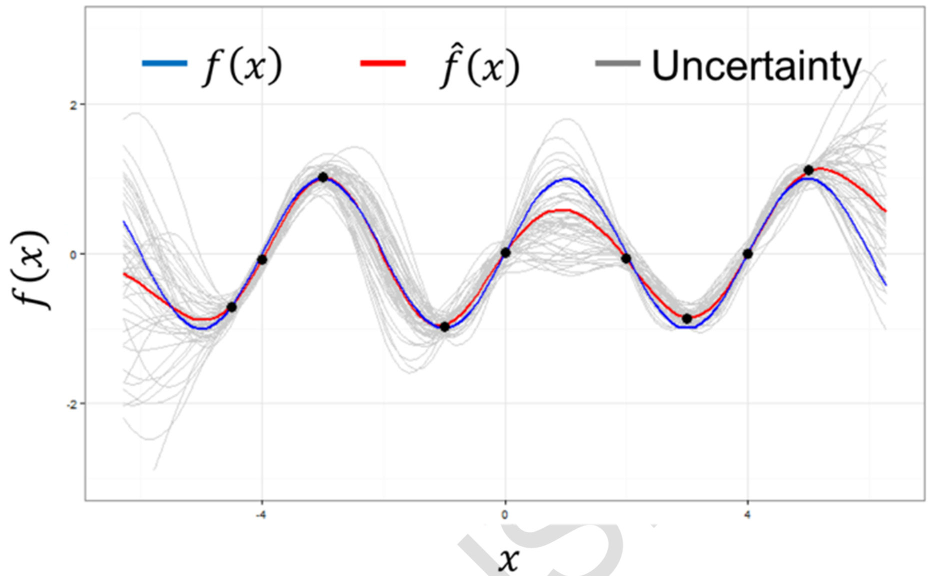

Figure 4-24 GP Surrogate Fit to Training Data (Black Points) from the True but Unknown Function $f(x)$

Surrogate validation. Validation of the surrogate can be done using the following steps:

(1) Use the surrogate to predict the response at a set of new input values x not used in construction of the surrogate.

(2) Run the full computer model at x.

(3) Compare the predicted response using the surrogate to the response using the computer model.

(4) Determine whether the surrogate is sufficiently accurate. If not, then more realizations from the computer model are needed to improve the surrogate, or a different surrogate model is needed.

(5) If the surrogate is used for uncertainty propagation, the error associated with the surrogate's approximation should be considered when quantifying uncertainty in the QoI.

Reference 4-74 provides more information on surrogate validation.

It is important to ensure that the surrogate model is properly approximating the computer model by checking for potential over- or under-fitting of the surrogate, and multicollinearity. Overfitting refers to a surrogate representing the training data set so well that the surrogate does not generalize to new datasets and has low prediction capabilities. Underfitting refers to a surrogate that does not represent the training data well and therefore also does not generalize well. Multicollinearity may arise when two or more independent input variables in a surrogate model are correlated. This is potentially concerning because multicollinearity can result in unstable and unreliable output results.

Iterating between the surrogate model construction and validation steps is necessary to develop the most appropriate surrogate model for approximating the computer model.

Surrogate models for continuous data. There are many types of surrogates for continuous data. Examples of surrogate models include linear regression (Reference 4-76), multivariate adaptive regression splines (MARS) (Reference 4-77), and GPs (Reference 4-75). References 4-78 and 4-79 provide detailed overviews of these and other techniques, as well as details on how to use these surrogates for SA. The following gives a brief description of them:

- **Linear regression** is a statistical surrogate that models the output as a linear function of the inputs and tends to be one of the more interpretable models. It includes uncertainty in the coefficients and allows for uncertainty estimates in the outputs. Linear regression is often used as an initial screening tool in SA to identify the most important variables and can be used as a surrogate for the computer model.

  Despite the name, it is possible to model interactions and nonlinearities within a linear model. To sort through the many potential model candidates, fit criteria can be used to find the best model. For example, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) can be used to quantify model fit to the data, as well as automated methods to find the optimal AIC/BIC, such as stepwise selection.

- **Multivariate adaptive regression splines (MARS)** is a machine learning (ML) method that is used for flexible nonparametric regression modeling of high-dimensional data. Separate splines are fit to different intervals of the predictor variables. Variables, knots and interactions are evaluated simultaneously to produce an optimal fit. MARS allows for automatic variable selection and transformations and for variable interactions.

- **Gaussian process (GP) regression** is an ML method that assumes the input-output relationship can be modeled as a GP, which is a specific type of multivariate normal distribution. Specifically, correlation between the outputs is induced using a correlation structure that is a function of the inputs. The correlation structure is constructed such that inputs close together produce more similar outputs. GP is a flexible tool for interpolating outputs throughout the parameter input space. A primary disadvantage of GP is that it can become computationally expensive and unstable with large training sets or many inputs. Dimension reduction approximation techniques can be applied to make GPs more computationally feasible (Reference 4-78).

Surrogates for binary data. Binary data can arise in PFM applications when the model output is the occurrence of an adverse event (such as crack or rupture). As with continuous data, there are many different options for fitting surrogates to binary data. Because binary data contain less information than continuous data do, more initial computer model realizations (i.e., a larger training sample) are required to accurately model the relationship between inputs and outputs. In particular, to create a surrogate for rare events, more initial computer model realizations are commonly required, along with a strategic sampling plan, such as importance sampling (Section 4.3.3).

For example, a generalized linear model is a flexible extension of linear regression that can be used when the response does not satisfy the assumption of having a normal error distribution (e.g., when the response variable is binary). Common examples of generalized linear models for binary data are logistic regression and probit regression (References 4-80, 4-81).

Additional methods. Additional methods (for both continuous and binary data) include the following:

- **Machine learning** (ML) covers a broad group of flexible techniques that fit complex relationships in the data, with the goal of predicting an unobserved output as accurately as possible. ML methods can be used for both continuous and binary outputs. Examples of ML techniques include the already mentioned MARS and GP models, as well as neural networks (deep learning), regression trees, and support vector machines. Many texts (e.g., Reference 4-82, 4-83, 4-84, 4-85) provide technical details for a wide range of ML and statistical learning methods. Note that whether a model is considered to be an ML model varies from group to group. Some texts also consider select Bayesian models to be ML models. While some ML models may be limited in interpretability and uncertainty quantification, research is underway to improve interpretability and uncertainty estimates for ML models.

- **Bayesian models** integrate information based on probability theory. These models take into consideration prior knowledge with data to produce an output using the Bayes Theorem. The posterior distribution (output) is proportional to the product of the likelihood distribution (probability distribution that represents the observed test data from the computer experiment) and the prior distribution (probability distribution that represents the knowledge before observing the test data). All statistical inferences on the QoI are done on the posterior distribution.

Figure 4-25 shows an example of a surrogate prediction for binary data, where the probability of failure is predicted as a function of a single input variable—age. The surrogate models the probability of failure based on observed pass/fail (y=0 or y=1) outputs; the surrogate model is then compared to predictions of the failure probability from the computer model as validation of the surrogate. The points represent responses for components of varying ages, with a 0 meaning the component did not fail and a 1 meaning the component did fail. The orange line shows the true probability of failure (estimated from the computer model). The teal line shows the estimated probability of failure using logistic regression. The gray band represents 95-percent confidence bands on the probability.

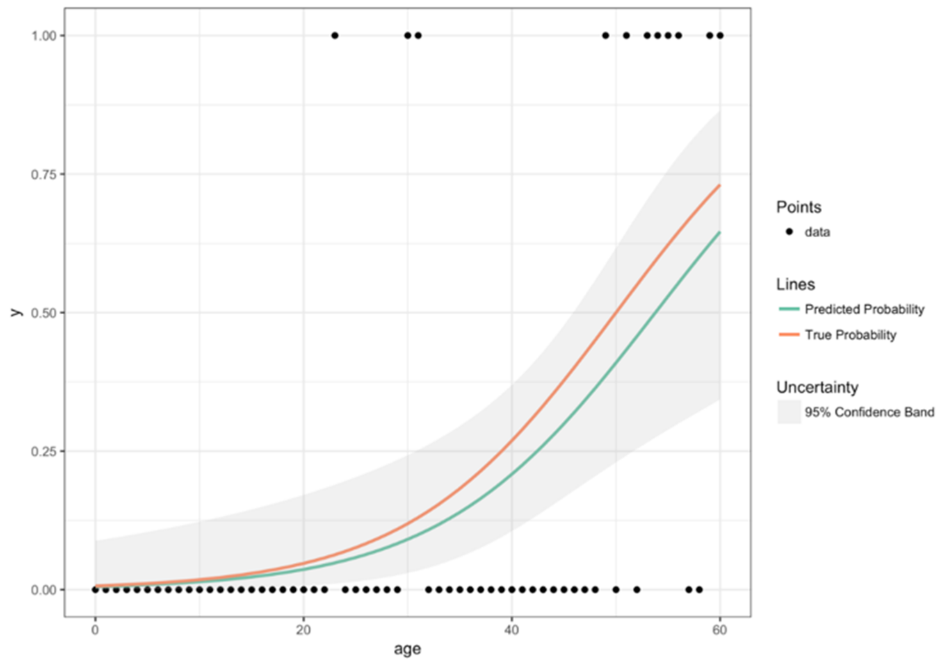

Figure 4-25 Example of a Generalized Linear Model for Binary Data—Component Failure as a Function of Age

4.3.11 Visualizing Output Uncertainty Due to Input Uncertainty

4.3.11.1 What Is It?

PFM analyses are conducted to estimate a specific QoI, though QoIs are never estimated as a single exact value due to uncertainty. Uncertainty analysis is the process of understanding and documenting uncertainty in a QoI estimate across model realizations. The uncertainty analysis approach depends on the QoI and the sampling design for the model realizations. Communication of the uncertainty analysis involves visualization. This section outlines common techniques for visualizing QoI estimates along with the quantifiable uncertainty in those estimates.

4.3.11.2 How to Use?

The appropriate visualization technique depends on three considerations:

(1) whether the analysis separates aleatory and epistemic uncertainties

(2) the type of the QoI (i.e., whether the probability QoI is represented as a function of time (continuous performance measure) or for a single point in time)

(3) whether the model realizations have equal weight

Section 4.3.11.4 describes the appropriate visualization techniques based on these three dependencies.

4.3.11.3 When/Why

Analysts visualize uncertainty and variability throughout a study. Thoughtful visualizations enhance the final results and help communicate the results effectively.

4.3.11.4 Technical Details

This section distinguishes the cases with and without separation of uncertainty types and discusses the type of the QoI. It also gives an overview of visualizing results with unequal weighting.

No separation of aleatory and epistemic uncertainty. Without separation of uncertainty, the model results will consist of a set of n outputs, where the outputs may be measured over time.

QoI is a continuous performance measure. Often, the QoI is continuous, such as crack length or leak rate. When the output is continuous and is not measured over time, the empirical CDF of the output samples should be plotted. When the output is measured over time, uncertainty in a continuous output can be visualized by plotting the output over time for each of the n realizations and overlaying the best estimate (e.g., a mean or median) and measure of uncertainty (e.g., quantiles of the output) at each time point. Figure 4-26 provides an example for two scenarios: (1) the output is not measured over time, and (2) the output is measured over time. The left plot shows an empirical CDF of a continuous output at a single time point over 1,000 realizations, with a solid vertical line at the median and dashed vertical lines at the 5th and 95th percentiles of the output. The right plot shows a continuous output over time over 1,000 realizations, with a solid line at the median and dashed lines at the 5th and 95th percentiles of the output.

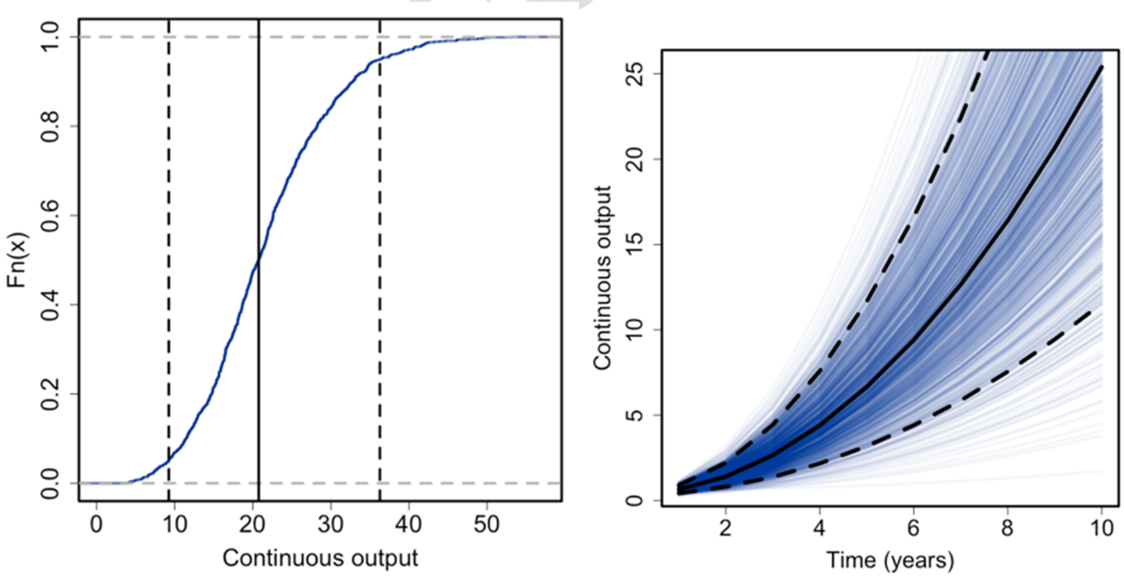

Figure 4-26 Continuous Output at a Single Time Point (Left) and Over Time (Right)

QoI is a probability. If the QoI is a probability (e.g., probability of failure), the model often outputs a binary 0/1 variable indicating that an event did or did not occur for that realization. The number of the 0/1 outputs divided by n number of samples is then used to estimate the probability of the event. Sampling uncertainty in the probability estimate can be computed using confidence intervals or other methods for computing sampling uncertainty.

When the QoI is measured over time, a graph of the estimate over time provides insight into how the estimate changes with time (Figure 4-27). When estimating rare probabilities, it is often more informative to plot the estimates on the log scale, so that the order of magnitude of the probability can be easily ascertained from the plot.

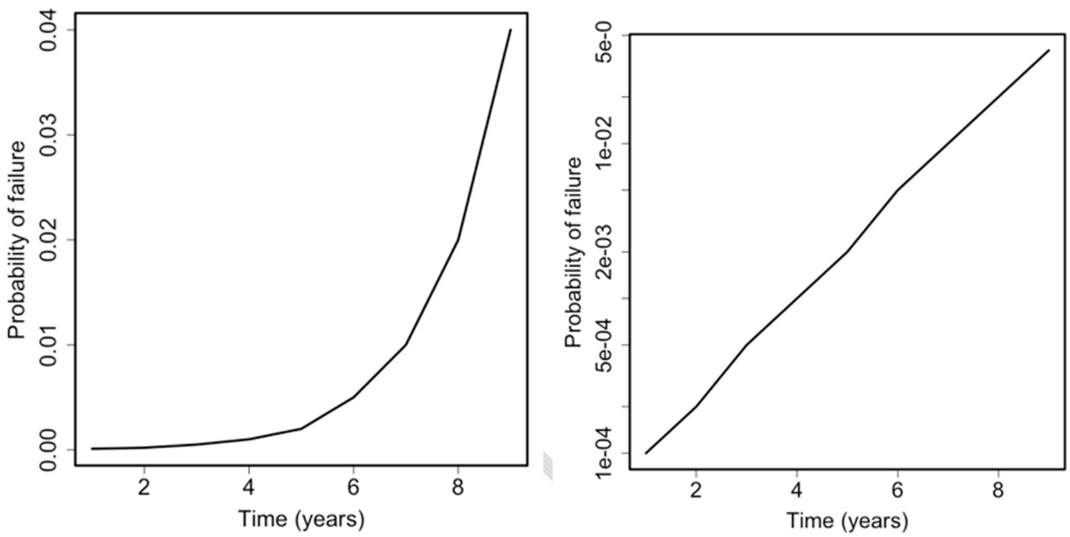

Figure 4-27 Failure Probability Over Time when Aleatory and Epistemic Uncertainty are not Separated; Linear Scale (Left) and Log Scale (Right)

Separation of aleatory and epistemic uncertainty. Separating uncertainties allows for the direct quantification of the impact of epistemic uncertainty (Section 4.1.1). Specifically, when uncertainties are separated using a double-loop algorithm, the set of model realizations will consist of n_e unique epistemic samples and n_a aleatory samples within each epistemic sample. The final sample size is then $n = n_e * n_a$.

QoI is a continuous performance measure. An estimate of the QoI, say Q_i, is computed from the output across the aleatory samples for each unique epistemic sample i = 1,2,..., n_e. A best estimate of the QoI is a measure of centrality (e.g., the mean or median) of the set of Q_is. The epistemic uncertainty in the QoI can be represented using percentiles of the Q_is. Specifically, the median, 5th, and 95th percentile of the Q_is can be presented as a best estimate and uncertainty for the QoI.

QoI is a probability. When the QoI is a probability, the average of the 0/1 output across the aleatory samples for each unique epistemic sample is computed to estimate the probability of the event, conditioned on the value of the epistemic input. A best estimate of the probability is the mean or median of these estimates. The epistemic uncertainty in the probability can be represented using percentiles of the estimates. Specifically, the median, 5th, and 95th percentiles can be presented as a best estimate with uncertainty.

In general, an estimate Q_i is provided for each epistemic input. These estimates contain sampling uncertainty due to a finite aleatory sample size. The precision of the individual Q_is should be considered. The number of samples n_a and sampling scheme determine how accurately each Q_i can be estimated. For example, if the probability of failure is on the order of 1×10^{-3}, then more than 1×10^{3} samples ($n_a > 10^3$) will be required to accurately estimate each Q_i using equal-probability weighted samples of 0/1 outputs (see Section 4.1.1).

The left plot in Figure 4-28 displays visualizations for the case when a probability is estimated at a single time point (or not a function of time) from 0/1 output. The figure plots the CDF of the estimated probabilities for each epistemic sample out of 1,000 samples. The solid vertical line is the median, and the dashed vertical lines are the 5th and 95th percentiles of the output. The estimated probabilities represent frequencies of the event over the aleatory samples. While it looks similar to the plot in Figure 4-26, its interpretation is different. If the aleatory sample size is large enough to make the sampling uncertainty in each estimate negligible, the spread in this CDF represents the spread due to epistemic uncertainty. Likewise, the plot on the right visualizes probability estimates as a function of time. Each blue curve represents an estimate of the probability given a fixed epistemic parameter. The solid line is the median, and the dashed lines are the 5th and 95th percentiles of the epistemic output. If the aleatory sample size is large enough, the spread in these blue curves represents uncertainty due to epistemic uncertainty in the inputs.

It is important to understand that if the aleatory uncertainty is a significant contributor to the uncertainty in each Q_i, the variability observed in these plots is due to both aleatory and epistemic uncertainty (Reference 4-86).

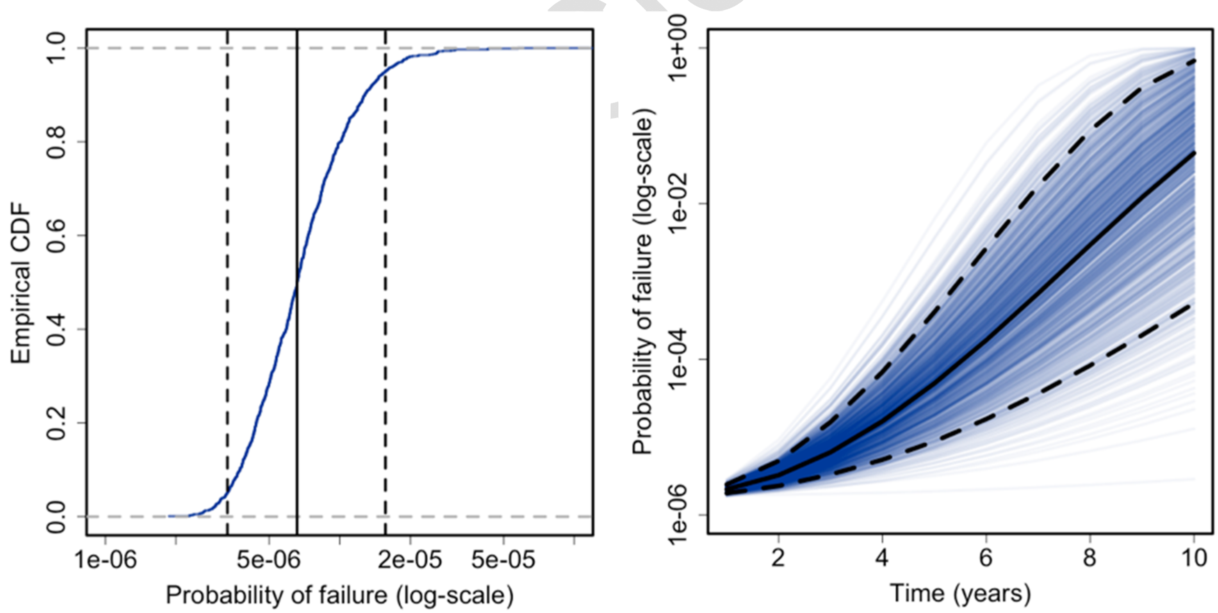

Figure 4-28 Frequency over Aleatory Samples at a Single Time Point (Left) and as a Function of Time (Right)

Weighting model realizations. When inputs are sampled using SRS or LHS, the generated inputs all have an equal probability of selection and the outputs are weighted equally when calculating a QoI. When inputs are sampled using importance sampling or another weighted sampling method, the model outputs are weighted differently. When calculating the best estimate and uncertainty in a QoI, the relevant weights should be applied. When visualizing uncertainty, realizations from the true output distribution should be plotted, rather than the observed output distribution with unequally weighted outputs. Sampling from the true output distribution can be achieved by sampling with replacement from the observed, weighted data using the following algorithm:

(1) Each output i has a corresponding weight w_i based on the selected sampling method (under SRS or LHS with no importance sampling, $w_i = 1$ for each i).

(2) The analyst should resample with replacement from the n outputs, where each output has a probability of being sampled proportional to its weight.

(3) The resampled data can be considered an unweighted, simple random sample of outputs.

Figure 4-29 shows an example of reweighting an importance-sampled distribution. The figure displays empirical CDFs of a probability of failure. The blue CDF is that of the output from the importance sampling. Since the importance sampling oversamples regions where failures are likely to occur, the CDF calculated directly from these outputs results in much larger estimated probabilities than those computed under SRS (black CDF). The green CDF is created by resampling the importance-sampled distribution using the importance sampling weights as described in the algorithm above. As desired, this reconstructed CDF is much closer to the one observed under SRS and actually estimates the distribution of failure probability.

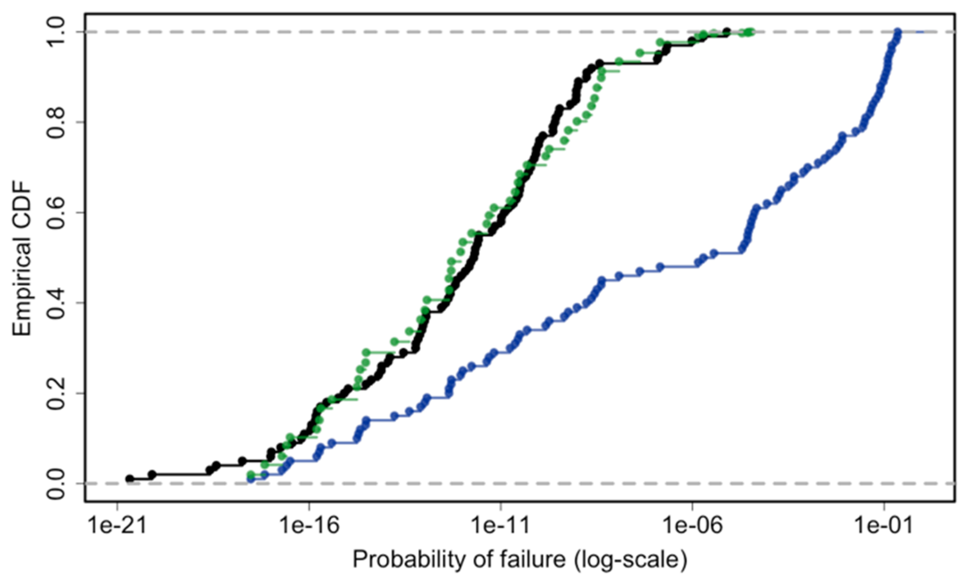

Figure 4-29 Importance-Sampled Distribution (Blue), Simple Random Sample (Black) Distribution, and Reconstructed Unweighted Distribution (Green) for a Probability of Failure

4.4 Useful Methods for Sensitivity Studies

4.4.1 Sensitivity Studies

4.4.1.1 What Is It?

Sensitivity studies are case studies that exercise the PFM computational framework under different assumptions. The goal of sensitivity studies is to determine whether uncertain assumptions impact the conclusions of the PFM analysis.

4.4.1.2 How to Use?

Key aspects of conducting sensitivity studies include the following:

- determining the set of uncertain assumptions that will be evaluated using sensitivity studies
- designing and running sensitivity studies

Determining the set of uncertain assumptions. The complexity of PFM computational frameworks results in a large set of assumptions, some of which may be uncertain and thus candidates for sensitivity studies. Assumption uncertainties can often be categorized as model uncertainty or input uncertainty. Further, uncertain assumptions can often be categorized by the degree of uncertainty in the assumption. Section 4.4.1.4 contains more information on model versus input uncertainty and classifying the degree of uncertainty to determine whether a sensitivity study is needed.

Designing and running sensitivity studies. When setting up a sensitivity study, the settings in the model and inputs are changed to reflect the plausible alternative assumption(s) under study. The analyst has a choice to conduct deterministic model realizations at a single value of the model inputs or to conduct probabilistic analyses over the range of the model inputs. The analyst will select either probabilistic or deterministic analysis for the sensitivity studies based on the change in the assumptions and the specific question being asked.

The sensitivity studies are designed to evaluate how changing an assumption impacts the results of the analysis. This requires knowledge gained throughout the PFM process as well as subject matter expertise. For example, the inputs whose assumptions are natural candidates for sensitivity studies are those considered important in SAs. Subject matter experts can help determine the credibility of the models and input parameters and identify plausible alternatives.

4.4.1.3 When/Why?

A typical PFM analysis relies on a complex model that consists of many submodels with many inputs and outputs joined together in an overall model framework. The complexity of these models results in a large set of assumptions, some of which may be uncertain. Each submodel and parameter input relies on assumptions that represents a decision to set up the problem and model in a specific way. Since the results of the PFM analysis depend on these uncertain assumptions, the effect of the assumptions should be studied with the goal of understanding whether plausible alternative assumptions will significantly change the results.

4.4.1.4 Technical Details

Determining a set of uncertain assumptions to study. PFM analyses contain many uncertain assumptions, but sensitivity studies should not be conducted for all such assumptions. Two primary factors should drive whether sensitivity studies are conducted (Reference 4-87):

(1) plausibility of assumption violation
(2) impact on analysis results

In general, sensitivity studies should be considered for more plausible assumptions that can impact the QoI. If subject matter experts or SAs cannot determine the plausibility or impact of a particular assumption a priori, a sensitivity study should generally be considered.

Types of uncertain assumptions. To determine a set of plausible alternative assumptions, Reference 4-87 distinguishes between two types of assumptions:

(1) **Modeling assumptions** refer to the types of submodels used in the PFM code, the assumptions made to develop each of the submodels, and any approximations made during calculations performed within each of the submodels. Modeling assumptions also include context assumptions that pertain to the context of the PFM analysis. Changes in analysis context are related to completeness uncertainty, defined as "uncertainty caused by the limitations in the scope of the model, such as whether all applicable physical phenomena have been adequately represented, and all accident scenarios that could significantly affect the determination of risk have been identified" (Reference 4-88). Examples of context assumptions for PFM applications include alternate scenarios, such as worst case scenarios and different intervention scenarios (discussed more below).

(2) **Input parameter specification assumptions** refer to any assumptions made when specifying the values of the input parameters to propagate through the PFM code. These include the choice to explicitly separate aleatory and epistemic uncertainties and the classification of each variable into these categories, the choice of fixed probability distributions for the inputs, the choice of correlation structure between the inputs, and the choice to treat certain inputs as deterministic (i.e., fixed).

Degree of assumption uncertainty. Reference 4-87 provides a useful set of categories for models and input parameter assumptions that helps to identify and rank sensitivity studies in terms of their plausibility and impact, as summarized below. The summary describes a list of categories for both models and input parameters, with a corresponding suggestion for whether a sensitivity study is needed:

- **Model categories:**[2] Models can be categorized according to the uncertainty in the modeling assumptions. Potential categorizations include the following:

  – *The model/submodel is a correct and credible representation of the underlying physical process.* Sensitivity studies are typically not needed. There is little benefit in subjecting a correct model to a sensitivity study. This category implies

[2] "Model" and "submodel" are used interchangeably here and should not cause confusion. Typically, it is the individual submodels that are categorized before the categorization of the overall model.

that there are either no other plausible models or any other plausible model is similar to the current model and would have low impact on the QoI.

- *The applicability of this model to all conditions of interest cannot be assessed reliably with the current state of knowledge.* Sensitivity studies should be considered. The correctness of the model is unknown. It is possible that there is no known plausible alternative model on which to develop a sensitivity study. In such cases, sensitivity studies scrutinizing the engineering decisions made in developing the model can help determine whether these decisions have unforeseen significant effects on the results. That is, there are potentially other plausible engineering decisions that could have been made and that would impact the QoI.

- *Plausible alternatives to the model adopted exist for a given physical process, and these alternatives have roughly equal justification to the model adopted.* Sensitivity studies should be considered. The alternative plausible models with roughly equal justification are usually candidates for sensitivity studies, especially if the model affects the QoI. The alternatives may include context assumptions such as worst case scenarios and intervention scenarios.

- *A model provides a conservative representation of the underlying physical process.* Sensitivity studies might be conducted. Conservative models are often adopted because of a lack of information. It may be necessary to set up studies to quantify the impact of the conservative choices.

- **Input parameter categories:** Inputs can also be categorized according to uncertainty in their assumptions. Potential categorizations include the following:

  - *The uncertainty distribution for the input parameter accurately represents the input for the conditions of interest. Additionally, the choice to classify the input as aleatory or epistemic is unambiguous.* A sensitivity study is not needed. This category implies there are no alternatives worth considering for the input parameter specification.

  - *The value or the uncertainty distribution was developed using limited prior information or data. Alternatively (or in addition), the choice to classify the parameter as aleatory or epistemic is ambiguous.* A sensitivity study should be considered. Given the limited information used to specify the input, plausible alternatives likely exist and are candidates for sensitivity studies. The analyst considers the impact the input has on the QoI when determining whether a study is needed. If the distribution/value is highly uncertain, but SA results and expert judgment agree that the input does not drive variability in the QoI, then a sensitivity study is typically not necessary. Alternatively, if the variable does drive variability in the QoI, a sensitivity study should be conducted.

  - *The distribution for the input parameter is considered a conservative representation of the parameter for the conditions of interest.* A sensitivity study might be conducted. Conservative input parameters are often used out of necessity due to a lack of information. As plausible alternative and potentially less conservative input specifications exist, quantifying the impact of these conservatisms could be helpful in building credibility.

Designing sensitivity studies*.* Sensitivity studies are designed based on the question of interest. Typical questions asked in sensitivity studies include the following:

- Do the results change significantly if a plausible alternative model is used? *(Step 1: Action 3)*

  If assumptions about the underlying code or physics model may be violated for the specific application, then sensitivity studies can address how the QoI changes under different model form assumptions (e.g., geometric fidelity, material model selection, new submodels). Sensitivity studies can demonstrate that the overall behavior of the PFM code is consistent with expert understanding of the expected system behavior, including demonstrating expected trends and correlations between inputs and outputs of interest. Benchmarking against other comparable codes may be used to increase confidence in the PFM code by demonstrating that the results produced by the PFM code are reasonable and can be predicted by similar codes (Reference 4-89).

- Do the results change significantly if a different distribution is used for an important input variable? *(Step 2: Action 2)*

  Sensitivity studies that vary the type of input distribution or distribution parameters can be conducted to determine the impact of the chosen distribution. The analyst should consider changing the characteristics of input distributions (e.g., shifting the mean, variance, or other distribution moments, such as skewness and kurtosis) as well as changing the distribution itself to highlight the uncertainty in specifying the distribution correctly.

- Do the results change significantly if a variable is considered aleatory rather than epistemic or vice versa? (*Step 2: Action 1)*

  If the analysis maintains the separation of aleatory and epistemic uncertainty and the uncertainty of an input cannot be clearly defined as aleatory or epistemic, then sensitivity studies can address how the analysis results change depending on the classification of this uncertainty type.

- Do the results change significantly under different context assumptions used to set up the problem *(Step 1: Action 1)*

  Examples of alternate scenarios could include the following:

  – *Worst case scenarios* or any adverse condition (such as accidents) are often considered for sensitivity studies in support of defense in depth. They are either designed by experts or found in benchmarking studies.

  – *Intervention scenarios* study the impact of some (usually positive) changes in the system, such as inspection or mitigation.

  – *Defense-in-depth scenarios* involve changes to nominal assumptions to represent adverse conditions or beyond-design-basis conditions. Such studies can be combined with different intervention scenarios to assess the benefit of the interventions under extreme conditions.

Sensitivity studies are important for both deterministic and probabilistic fracture analyses. For example, Reference 4-90 describes sensitivity studies conducted to understand the effects of potential changes to selected inputs and mechanisms in calculating failure probabilities under different inservice inspections programs. Sensitivity studies have also been used in support of defense in depth by considering beyond-design-basis accidents (see Reference 4-88 for some examples). Sensitivity studies seek to assess the credibility of the PFM model and analysis within the domain of the application. This is different from credibility in the context of V&V or statistical and numerical stability, where the goal is to build trust that the results are computed correctly and accurately enough to represent the phenomenon under study. Rather, sensitivity studies seek to inform what may happen under alternative assumptions made when defining the problem under consideration by quantifying the effects of the alternative assumptions.

4.5 References

4-1. Kaplan, S. and Garrick, B.J., "On the quantitative definition of risk," *Risk Analysis*, 1(1), 11–27, 1981.

4-2. Kaplan, S., "Formalisms for handling phenomenological uncertainties: the concepts of probability, frequency, variability, and probability of frequency," *Nuclear Technology*, 102(1), 137–142, 1993.

4-3. Hoffman, F.O., and Hammonds, J.S., "Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability," *Risk Analysis*, *14*(5), 707–712, 1994.

4-4. Helton, J.C., "Treatment of uncertainty in performance assessments for complex systems," *Risk Analysis*, 14(4), 483–511, 1994.

4-5. Der Kiureghian, A., and Ditlevsen, O., "Aleatory or epistemic? Does it matter?" *Structural Safety*, 31(2), 105–112, 2009.

4-6. Apostolakis, G., "The distinction between aleatory and epistemic uncertainties is important: an example from the inclusion of aging effects into PSA," in *Proceedings of PSA '99, International Topical Meeting on Probabilistic Safety Assessment*, 135–142, 1999.

4-7. Helton, J.C., Johnson, J., and Oberkampf, W., "An exploration of alternative approaches to the representation of uncertainty in model predictions," *Reliability Engineering and System Safety*, 89, 39–71, 2004.

4-8. Helton, J.C., and Burmaster, D.E., "Guest editorial: treatment of aleatory and epistemic uncertainty in performance assessments for complex systems," *Reliability Engineering and System Safety*, 54(2–3), 91–94, 1996.

4-9. Drouin, M., Gonzalez, M., Herrick, S., Hyslop, J.S., Stroup, D., Lehner, J., Pratt, T., Dennis, M., LaChance, J., and Wheeler, T., "Glossary of Risk-Related Terms in Support of Risk-Informed Decisionmaking," NUREG-2122, U.S. Nuclear Regulatory Commission, 2013.

4-10. Chapman, O.J., “RR-PRODIGAL—A Model for Estimating the Probabilities of Defects in Reactor Pressure Vessel Welds,” NUREG/CR-5505, U.S. Nuclear Regulatory Commission, 1998.

4-11. Duan, X., Wang, M., and Kozluk, M.J., “Benchmarking PRAISE-CANDU 1.0 With NURBIM Project Fatigue Cases,” PVP2013-97785, *ASME 2013 Pressure Vessels and Piping Conference Volume 2: Computer Technology and Bolted Joints,* Paris, France, American Society of Mechanical Engineers, 2013.

4-12. Harris, D., and Lim, E., Dedhia, D., “Probability of Pipe Fracture in the Primary Coolant Loop of a PWR Plant,” Volume 5, “Probabilistic Fracture Mechanics Analysis,” NUREG/CR-2189 Vol. 5, UCID-18967 Vol.5, U.S. Nuclear Regulatory Commision, 1981.

4-13. Harris, D.O., “A Probabilistic Fracture Mechanics Code for Piping Reliability Analysis (pcPRAISE code),” NUREG/CR-5864, U.S. Nuclear Regulatory Commission, 1992.

4-14. Rudland, D., Harrington, C., and Dingreville, R., “Development of the Extremely Low Probability of Rupture (xLPR) Version 2.0 Code,” PVP2015-45134, *Proceedings of the ASME 2015 Pressure Vessels and Piping Conference*, American Society of Mechanical Engineers, 2015.

4-15. Kurth, R., Sallaberry, C., Young, B., Scott, P., Brust, F., and Kurth, E., “Benchmarking Probabilistic Codes for LBB Analysis for Circumferential Cracks,” *Proceedings of the ASME 2017 Pressure Vessels and Piping Conference,* Waikola, HI, American Society of Mechanical Engineers, 2017.

4-16. Helton, J.C., and Johnson, D.J., “Quantification of margins and uncertainties: Alternative representations of epistemic uncertainty,” *Reliability Engineering and System Safety*, 96(9), 1034–1052, 2011.

4-17. Helton, J.C., Johnson, J.D., Oberkampf, W.L., and Sallaberry, C.J., “Representation of analysis results involving aleatory and epistemic uncertainty,” *International Journal of General Systems*, 39(6), 605–646, 2010.

4-18. Helton, J.C., “Quantification of margins and uncertainties: Conceptual and computational basis,” *Reliability Engineering and System Safety*, 96(9), 976–1013, 2011.

4-19. Atwood, C.L., LaChance, J.L., Martz, H.F., Anderson, D.J., Englehardt, M., Whitehead, D., and Wheeler, T., “Handbook of Parameter Estimation for Probabilistic Risk Assessment,” NUREG/CR-6823, U.S. Nuclear Regulatory Commission, 2003.

4-20. Xing, J., and Morrow, S., “White Paper: Practical Insights and Lessons Learned on Implementing Expert Elicitation, U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, 2016.

4-21. Iman, R.L., and Conover, W.J., “A distribution-free approach to inducing rank correlation among input variables,” *Communication in Statistics—Simulation and Computation,* 11(3), 311–334, 1982.

4-22. Jaworski, P., Durante, F., Härdle, W.K., and Rychlik, T. (Eds.), *Copula theory and its applications: proceedings of the workshop held in Warsaw, 25–26 September 2009*, Volume 198, Springer Science and Business Media, 2010.

4-23. Haldar, A., and Mahadevan, S., Probability, Reliability, and Statistical Methods in Engineering Design, John Wiley and Sons, Inc., 2000.

4-24. Robert, C.P., *Monte Carlo Statistical Methods*, John Wiley and Sons, Ltd., 2004.

4-25. Helton, J.C., and Davis, F.J., "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," Reliability Engineering and Systems Safety, 81(1), 23–69, 2003.

4-26. Mead, R., *The Design of Experiments*, Cambridge University Press, United Kingdom, 1988.

4-27. McKay, M., Beckman, R., and Conover, W., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, 21(2), 239–245, 1979.

4-28. Robert, C.P., *Monte Carlo Statistical Methods,* John Wiley and Sons, Ltd., 2004.

4-29. Sallaberry, C.J., Helton, J.C., and Hora, S.C., "Extension of Latin hypercube samples with correlated variables," *Reliability Engineering and System Safety*, 93(7), 1047–1059, 2008.

4-30. Kaplan, S., "On the method of discrete probability distributions in risk and reliability calculations—application to seismic risk assessment," *Risk Analysis*, 1: 189–196, 1981.

4-31. Adams, B.M., Bauman, L.E., Bohnhoff, W.J., Dalbey, K.R., Eddy, J.P., Ebeida, M.S., Eldred, M.S., Hough, P.D., Hu, K.T., Jakeman, J.D., Swiler, L.P., Stephens, J.A., Vigil, D.M., and Wildey, T.M, "DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.0 User's Manual," Sandia Technical Report SAND2014-4633, July 2014, updated May 2017.

4-32. Robert, C.P. *Monte Carlo Statistical Methods*, John Wiley and Sons, Ltd., 2004.

4-33. Helton, J.C., and Davis, F.J., "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering and Systems Safety*, 81(1), 23–69, 2003.

4-34. Swiler, L. and West, N., "Importance Sampling: Promises and Limitations," 51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2010.

4-35. Glynn, P.W., and Iglehart, D.L., "Importance Sampling for Stochastic Simulations," *Management Science*, 35(11), 1367–1392, 1989.

4-36. Au, S.K., and Beck, J.L., "Important sampling in high dimensions," *Structural Safety*, 25(2), 139–163, 2003.

4-37. Haldar, A., and Mahadevan, S., Probability, Reliability, and Statistical Methods in Engineering Design, John Wiley and Sons, Inc., 2000.

4-38. Karamchandani, A., Bjerager, P., and Cornell, A.C., "Adaptive Importance Sampling," *Proceedings, International Conference on Structural Safety and Reliability (ICOSSAR),* San Francisco, CA, 855–862, 1989.

4-39. Wu, Y.T., "An Adaptive Importance Sampling Method for Structural System Reliability Analysis, Reliability Technology 1992," in T.A. Cruse (Editor), *ASME Winter Annual Meeting,* Volume AD-28, Anaheim, CA, 217–231, 1992.

4-40. Bichon, B.J., Eldred, M.S., Swiler, L.P., Mahadevan, S., and McFarland, J.M., "Efficient Global Reliability Analysis for Nonlinear Implicit Performance Functions," *AIAA Journal*, 46(10), 2459–2468, 2008.

4-41. Haldar, A., and Mahadevan, S., Probability, Reliability, and Statistical Methods in Engineering Design, John Wiley and Sons, Inc., 2000.

4-42. Lopez, R., and Beck, A., "Reliability-Based Design Optimization Strategies Based on FORM: A Review," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 34, 506–514, 2012.

4-43. Hasofer, A.M., and Lind, N.C., "Exact and Invariant Second Moment Code Format," *Journal of the Engineering Mechanics Division, ASCE, 100(EM1),* 111–121, 1974.

4-44. Rackwitz, R., "Practical Probabilistic Approach to Design," *Bulletin No. 112*, Comité European du Béton, Paris, France, 1976.

4-45. Rackwitz, R., and Fiessler, B., "Structural Reliability Under Combined Random Load Sequences," *Computers and Structures*, 9(5), 484–494, 1978.

4-46. Rackwitz, R., and Fiessler, B., "Note on Discrete Safety Checking When Using Non-Normal Stochastic Models for Basic Variables," *Load Project Working Session*, Massachusetts Institute of Technology, 1976.

4-47. Fiessler, B., Neumann, H.J., and Rackwitz, R., "Quadratic Limit States in Structural Reliability," *Journal of Engineering Mechanics,* American Society of Civil Engineers, 1095(4), 661–676, 1979.

4-48. Breitung, K., "Asymptotic Approximations for Multinormal Integrals," *Journal of Engineering Mechanics*, American Society of Civil Engineers, 110(3), 357–366, 1984.

4-49. Hohenbichler, M., Gollwitzer, S., Kruse, W., and Rackwitz, R., "New Light on First- and Second-Order Reliability Methods," *Structural Safety*, 4, 267–284, 1987.

4-50. Brown, L.D., Cai, T.T., and DasGupta, A., "Interval estimation for a binomial proportion," *Statistical science*, 101–117, 2001.

4-51. Clopper, C.J., and Pearson, E.S., "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, 26, 404–413, 1934.

4-52. Lohr, S., *Sampling: design and analysis*, Nelson Education, 2009.

4-53. Haldar, A., and Mahadevan, S., Probability, Reliability, and Statistical Methods in Engineering Design, John Wiley and Sons, Inc., 2000.

4-54. Brown, L.D., Cai, T.T., and DasGupta, A., “Interval estimation for a binomial proportion,” *Statistical science,* 101–117, 2001.

4-55. Louis, T., “Confidence Intervals for a Binomial Parameter after Observing No Successes,” *The American Statistician*, 35(3), 154, 1981.

4-56. Bickel, P.J., and Doksum, K.A., Mathematical Statistics: Basic Ideas and Selected Topics, Holden-Day, 1977.

4-57. Davison, A.C., and Hinkley, D.V., *Bootstrap Methods and their Application*, Volume 1, Cambridge University Press, United Kingdom, 1997.

4-58. Davison, A.C., Hinkley, D.V., and Young, G.A., “Recent developments in bootstrap methodology,” *Statistical Science*, 141–157, 2003.

4-59. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., and Tarantola, S., *Global Sensitivity Analysis. The Primer*, John Wiley and Sons, Ltd., 2008.

4-60. Tukey, J.W., *Exploratory Data Analysis*, Volume 2, 1977.

4-61. Cukier, R.I., Fortuin, C.M., and Shuler, K.E., “Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients,” “I: Theory,” *The Journal of Chemical Physics*, 59, 2003.

4-62. Sobol’, I.M., “On sensitivity estimation for nonlinear mathematical models,” *Matematicheskoe Modelirovanie*, 2(1), 112–118, 1990. (In Russian)

4-63. Kleignen, J.P.C., and Helton, J.C., “Statistical analyses of scatterplots to identify important factors in large-scale simulations,” “1: Review and comparison of techniques,” *Reliability Engineering and System Safety*, 65, 147–185, 1999.

4-64. Helton, J.C., Johnson, J.D., Sallaberry, C.J., and Storlie, C.B., “Survey of sampling-based methods for uncertainty and sensitivity analysis,” *Reliability Engineering and System Safety*, 91(10–11), 1175–1209, 2006.

4-65. Borgonovo, E., and Plischke, E., “Sensitivity analysis: a review of recent advances,” *European Journal of Operational Research*, 248(3), 869–887, 2016.

4-66. Robert, C.P., *Monte Carlo Statistical Methods*, John Wiley and Sons, Ltd., 2004

4-67. Storlie, C.B., and Helton, J.C., “Multiple predictor smoothing methods for sensitivity analysis: Description of techniques,” *Reliability Engineering and System Safety,* 93, 28–54, 2008.

4-68. Storlie, C.B., Swiler, L.P., Helton, J.C., and Sallaberry, C.J., “Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models," *Reliability Engineering and System Safety*, 94, 1735–1763, 2009.

4-69. Sullivan, T.J., *Introduction to Uncertainty Quantification*, Volume 63, Springer, 2015.

4-70. Constantine, P.G., "Active subspaces: Emerging ideas for dimension reduction in parameter studies," Volume 2, *SIAM Spotlights*, Society for Industrial and Applied Mathematics, 2015.

4-71. Smith, R.C., *Uncertainty Quantification Theory, Implementation, and Applications*, Society for Industrial and Applied Mathematics, 2014.

4-72. Helton, J.C., "Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal," *Reliability Engineering and System Safety*, 42(2–3), 327–367, 1993.

4-73. Loeppky, J.L., Sacks, J., and Welch, W.J., "Choosing the sample size of a computer experiment: A practical guide," *Technometrics*, 51(4), 366–376, 2009.

4-74. Queipo, N.V., Haftka, R.T., Shyy, W., Goel, T., Vaidyanathan, R., Tucker, P.K., "Surrogate-based analysis and optimization," *Progress in Aerospace Sciences*, 41, 1–28, 2005.

4-75. Rasmussen, C.E., and Williams, C., *Gaussian Processes for Machine Learning*, MIT Press, 2006.

4-76. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., and Tarantola, S., *Global Sensitivity Analysis. The Primer*, John Wiley and Sons, Ltd., 2008.

4-77. Friedman, J., "Multivariate adaptive regression splines" (with discussion), *Annals of Statistics*, 1–141, 1991.

4-78. Storlie, C.B., and Helton, J.C., "Multiple predictor smoothing methods for sensitivity analysis: Description of techniques," *Reliability Engineering and System Safety*, 93, 28–54, 2008.

4-79. Storlie, C.B., Swiler, L.P., Helton, J.C., and Sallaberry, C.J., "Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models," *Reliability Engineering and System Safety*, 94, 1735–1763, 2009.

4-80. Casella, G., and Berger, R.L., *Statistical Inference*, Volume 2, Duxbury, 2002.

4-81. McCullagh, P., and Nelder, J., *Generalized Linear Models*, *Second Edition*, Chapman and Hall/CRC, 1989.

4-82. Breiman, L., Friedman, J., Olshen, R., and Stone, C., *Classification and Regression Trees*, Wadsworth Intl., 1984.

4-83. Hastie, T., Tibshirani, T., and Friedman, J., *The Elements of Statistical Learning*, Springer-Verlag, 2009.

4-84. Casarez, C., “Global Sensitivity Analysis of xLPR using Metamodeling,” *CNSC-IAEA 2nd International Seminar on Probabilistic Methodologies for Nuclear Applications*, Ottawa, Ontario, Canada, 2017.

4-85. Sutton, C.D., “11—Classification and Regression Trees, Bagging, and Boosting,” *Handbook of Statistics*, 24, 303–329, 2005.

4-86. Jyrkama, M.I., and Pandey, M.D., “On the separation of aleatory and epistemic uncertainties in probabilistic assessments,” *Nuclear Engineering and Design*, 303, 68–74, 2016.

4-87. EricksonKirk, M.T., Dickson, T., Mintz, T., and Simonen, F., “Sensitivity Studies of the Probabilistic Fracture Mechanics Model Used in FAVOR,” NUREG-1808, U.S. Nuclear Regulatory Commission, 2010.

4-88. Drouin, M., Gonzalez, M., Herrick, S., Hyslop, J.S., Stroup, D., Lehner, J., Pratt, T., Dennis, M., LaChance, J., and Wheeler, T., “Glossary of Risk-Related Terms in Support of Risk-Informed Decisionmaking,” NUREG-2122, U.S. Nuclear Regulatory Commission, 2013.

4-89. Brickstad, B., Schimpfke, T., Chapman, O.J.V., Dillstrom, P., Cueto-Felgueroso, C., and Bell, C.D., “Project NURBIM (Nuclear RI-ISI Methodology for Passive Components): Benchmarking of Structural Reliability Models and Associated Software,” in *ASME/JSME 2004 Pressure Vessels and Piping Conference*, American Society of Mechanical Engineers, 109–119, 2004.

4-90. Khaleel, M.A., and Simonen, F.A., “Evaluation of Structural Failure Probabilities and Candidate Inservice Inspection Programs,” NUREG/CR-6986, U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, 2009.

5 SUMMARY AND CONCLUSIONS

The guidance provided in this document outlines a framework for conducting PFM analyses that recognizes the fact that each regulatory application may have unique characteristics. To address the unique characteristics of diverse PFM applications, this NUREG presents the elements of a graded approach for PFM analyses and the corresponding expectations for supporting documentation. These elements are aligned with the documentation elements previously given in the NRC's technical letter report, "Important Aspects of Probabilistic Fracture Mechanics Analyses," and the outcomes from the NRC public meeting to discuss a graded approach for PFM codes and analyses for regulatory applications.

The three technical sections (Sections 2, 3, and 4) develop the concept of PFM analysis methodology and outline important considerations for a high-quality and high-confidence PFM analysis. The sections are linked together and progressively dive into more detailed elements of PFM applications, but each may have different audiences:

- Section 2 is intended for applicants of all experience levels. Each subsection introduces an element of content expected in a PFM submittal. It identifies representative circumstances for a submittal to guide the applicant through a graded approach for the specific information to provide to the NRC.

- Section 3 could be used by applicants who are familiar with PFM submittals but are seeking some guidance on the development of an analysis structure or on formalism. Each subsection presents an analytical step that may exist in a PFM submittal. The analysis steps are linked to the expected documentation presented in Section 2.

- Section 4 could be used by applicants who are seeking explicit guidance on the theoretical underpinnings of the processes used to establish the credibility of a PFM analysis. Each subsection presents the fundamental background for the concepts and methods used in a PFM analysis. Examples give details for analysts on (nonprescriptive) approaches for PFM analyses. The concepts and methods are linked directly to the analysis steps presented in Section 3.

The NRC does not require PFM submittals to follow the process outlined in this NUREG, but rather the submittals should be structured to address the specific features of the application under investigation. The staff recommends that applicants identify deviations from the framework presented in this document to enhance the efficiency of NRC reviews of PFM submittals.

GLOSSARY

Acceptance Criteria

Set of conditions that must be met to achieve success for the desired application.

Accuracy and Precision

"Accuracy" is the degree to which the result of a measurement, calculation, or specification conforms to the correct value (i.e., reality or a standard accepted value). "Precision" is a description of random errors and a measure of statistical variability for a given quantity. In other words, "accuracy" is the proximity of measurement results to the true value; "precision" is the repeatability or reproducibility of the measurement.

Aleatory Uncertainty

Uncertainty based on the randomness of the nature of the events or phenomena that cannot be reduced by increasing the analyst's knowledge of the systems being modeled (Reference 0-1).

Assumption

A decision or judgment made in the development of a model or analysis (Reference 0-1).

Bayesian Inference

Type of data analysis in which an initial estimate about a parameter value is combined with evidence to arrive at a more informed estimate (Reference 0-1).

Benchmark (in the context of PFM computational analyses)

An established point of reference against which computers or programs can be measured in tests comparing their performance, reliability, output, etc. A standard against which similar analyses must be measured or judged. Benchmarks are often a part of validation for scientific analysis software.

Best Estimate

Approximation of a quantity based on the best available information (Reference 0-1). Models that attempt to fit data or phenomena as best as possible. That is, models that do not intentionally bound data for a given phenomenon or are not intentionally conservative or optimistic.

Calibration

The process of adjusting physical modeling parameters in the computational model to improve agreement with experimental data (Reference 0-2).

Code

The computer implementation of algorithms developed to facilitate the formulation and approximation solution of a class of problems. (Reference 0-2).

Code Verification

The process of determining and documenting the extent to which a computer program ("code") correctly solves the equations of the mathematical model (Reference 0-3).

Completeness Uncertainty

Caused by the limitations in the scope of the model, such as whether all applicable physical phenomena have been adequately represented and all accident scenarios that could significantly affect the determination of risk have been identified (Reference 0-1).

Component

A part of a system in a nuclear power plant (Reference 0-1).

Conditional Probability

Probability of occurrence of an event, given that a prior event has occurred (Reference 0-1).

Confidence Interval

A range of values that has a specified likelihood of including the true value of a random variable (Reference 0-1).

Consequence

In the context of nuclear regulatory submittals, the health effects or the economic costs resulting from a nuclear power plant accident (Reference 0-1).

Conservative Analysis

An analysis that uses assumptions such that the assessed outcome is meant or found to be less favorable than the expected outcome (Reference 0-1).

Convergence Analysis

An analysis with the purpose of assessing the approximation error in the quantity of interest estimates to establish that conclusions of the analysis would not change solely due to sampling uncertainty.

Correlation

A general term for interdependence between pairs of variables (Reference 0-4).

Credibility

The quality to elicit belief or trust in modeling and simulation results (Reference 0-5).

Cumulative Distribution Function

A function that provides the probability that a parameter is less than or equal to a given value (Reference 0-1).

Continuous variable

See “Discrete versus Continuous Variables.”

Dependent

Not independent.

Deterministic

A characteristic of decisionmaking in which results from engineering analyses not involving probabilistic considerations are used to support a decision (Reference 0-1). Consistent with the principles of determinism, which hold that specific causes completely and certainly determine effects of all sorts (Reference 0-6). Also refers to fixed model inputs.

Deterministic Fracture Mechanics

An analysis that uses fixed values of input parameters to a fracture mechanics model to estimate a fixed model output or quantity of interest computed from the output.

Discrete versus Continuous Variables

A discrete random variable is a variable that has a nonzero probability for only a finite, or countably infinite, set of values. A continuous random variable is a variable that has an absolutely continuous cumulative distribution function (Reference 0-3).

Distribution

A function specifying the values that the random variable can take and the likelihood they will occur.

Engineering Judgment

The scientific process by which a design, installation, operation/maintenance, or safety problem is systematically evaluated. The decision made by an engineer based on the available data to propose a design or a line of action.

Epistemic Uncertainty

The uncertainty related to the lack of knowledge or confidence about the system or model; also known as "state-of-knowledge uncertainty." As defined by the American Society of Mechanical Engineers (ASME)/American Nuclear Society (ANS) probabilistic risk assessment (PRA) standard (Reference 0-1), "the uncertainty attributable to incomplete knowledge about a phenomenon that affects our ability to model it. Epistemic uncertainty is reflected in ranges of values for parameters, a range of viable models, the level of model detail, multiple expert interpretations, and statistical confidence. In principle, epistemic uncertainty can be reduced by the accumulation of additional information. (Epistemic uncertainty is sometimes also called 'modeling uncertainty.')" (Reference 0-1)

Expert Elicitation

A formal, structured, and documented process to obtain judgments from expert(s). May be used to obtain information from technical experts on topics that are uncertain. A process in which experts are assembled and their judgment is sought and aggregated in a formal way. (Reference 0-1)

Expert Judgment

Information (or opinion) provided by one or more technical experts based on their experience and knowledge. Used when there is a lack of information, for example, if certain parameter values are unknown, or there are questions about phenomenology in accident progression. May be part of a structured approach, such as expert elicitation, but is not necessarily as formal. May be the opinion of one or more experts, whereas expert elicitation is a highly structured process in which the opinions of several experts are sought, collected, and aggregated in a very formal way. (Reference 0-1)

Failure Probability

As defined in the ASME/ANS PRA standard (Reference 0-1), “the likelihood that a system or component will fail to operate upon demand or fail to operate for a specific mission time” (Reference 0-1). For components, can also be the likelihood of a component being in a defective, unacceptable condition (adverse condition or event) (e.g., leakage from reactor coolant pressure boundary).

Frequency

The expected number of occurrences of an event or accident condition expressed per unit of time. Normally expressed in events per plant (or reactor) operating year or events per plant (or reactor) calendar year (Reference 0-1).

Global Sensitivity Analysis

The study of how the uncertainty in the output or quantity of interest of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input. The term “global” ensures that the analysis considers more than just local or one-factor-at-a-time effects. Hence, interactions and nonlinearities are important components of a global statistical sensitivity analysis (Reference 0-3).

Important Variable

A variable whose uncertainty contributes substantially to the uncertainty in the response (Reference 0-7).

Independence

Two events are said to be independent if knowing the outcome of one tells us nothing about the other (Reference 0-4).

Input

Data or parameters that users can specify for a model; the output of the model varies as a function of the inputs, which can consist of physical values (e.g., material properties, tolerances) and model specifications (e.g., spatial resolution).

Input Uncertainty

The uncertainty in the values of the inputs to the model represented by probabilistic distributions (Reference 0-1).

Interaction Effect

A term applied when two (or more) explanatory variables do not act independently on a response variable.

Level of Detail

The degree of resolution or specificity in the analyses performed. Generally refers to the level to which a system is modeled; dictated by (1) the level of detail to which information is available, (2) the level of detail required so that dependencies are included, (3) the level of detail so that the risk contributors are included, and (4) the level of detail sufficient to support the application (Reference 0-1).

Local Sensitivity Analysis

A sensitivity analysis that is relative to the location in the input space chosen and not for the entire input space (Reference 0-7).

Margin

The distance between the quantity of interest and the acceptance criteria.

Mean

The average of a set of numerical values; more technically, the expected value of a random variable (Reference 0-1).

Median

The value that a random variable is equally likely to be above and below. Also known as the 50th percentile of the distribution of a random variable (Reference 0-1).

Model

A representation of a physical process that allows for prediction of the process' behavior (Reference 0-1).

Model Uncertainty

Related to an issue for which no consensus approach or model exists and where the choice of approach or model is known to have an effect on the decision made (Reference 0-1).

Output

A value calculated by the model given a set of inputs.

Parameter

A numerical characteristic of a population or probability distribution. More technically, the variables used to calculate and describe frequencies and probabilities (Reference 0-1).

Percentile

The set of divisions that produce exactly 100 equal parts in a series of continuous values (Reference 0-4).

Point Estimate

An estimate of a parameter in the form of a single value (Reference 0-1).

Precision

See "Accuracy and Precision."

Prediction

The use of a model to make statements about quantities of interest in settings (initial conditions, physical regimes, parameter values, etc.) that are inside (interpolative) or outside (extrapolative) the conditions for which the model validation effort occurred (Reference 0-3).

Probabilistic

A characteristic of an evaluation that considers the likelihood of events (Reference 0-1).

Probabilistic Fracture Mechanics

An analysis that uses probabilistic representations of uncertain input parameters to a fracture mechanics model to estimate uncertainty in the model outputs or quantities of interest computed from the outputs (Reference 0-8).

Probabilistic Risk Assessment

A systematic method for assessing the likelihood of accidents and their potential consequences (Reference 0-1).

Probability

A number between 0 and 1 describing the likelihood or chance of an event occurring. There are two main interpretations of probability:

(1) *Frequency interpretation.* The probability of an event is the relative frequency of the occurrence of the event in a long sequence of trials in which the event does or does not occur. In other words, the likelihood that an event will occur is expressed by the ratio of the number of actual occurrences to the total number of possible occurrences (Reference 0-1).

(2) *Subjective interpretation.* The probability of an event comes from expert judgment about uncertain events or quantities, in the form of probability statements about future events. It is not based on any precise computation but is often a reasonable assessment by a knowledgeable person (Reference 0-3).

Probability Density Function

A function of a continuous random variable whose integral over an interval gives the probability that its value will fall within the interval (Reference 0-1). Analogous to probability distribution for continuous random variables.

Probability Distribution

A function specifying the values that the random variable can take and the likelihood they will occur (Reference 0-1).

Quantiles

Divisions of a probability distribution or frequency distribution into equal, ordered subgroups (Reference 0-4).

Quantity of Interest

A numerical characteristic of the system being modeled, the value of which is of interest to stakeholders, typically because it informs a decision (Reference 0-3). Can refer to either a physical quantity that is an output from a model or a given feature of the probability distribution function of the output of a deterministic model with uncertain inputs. (Reference 0-9)

Random Uncertainty

See “Aleatory Uncertainty.”

Random Variable

A variable, the values of which occur according to some specified probability distribution (Reference 0-4).

Rank

The relative position of the members of a sample with respect to some characteristic (Reference 0-4).

Rare

Events that are unlikely to occur. Rare event probabilities are defined as probabilities that are close enough to 0 that the number of samples needed to estimate the probability is large relative to the computational budget.

Realization

The execution of a model for a single set of input parameter values (Reference 0-8).

Regression

A form of statistical analysis in which observational data are used to statistically fit a mathematical function that presents the data (i.e., dependent variables) as a function of a set of parameters and one or more independent variables (Reference 0-3).

Reliability

The likelihood that a system, structure, or component performs its required function(s) for a specific period of time (Reference 0-1).

Risk

The combined answer to the three questions that consider (1) what can go wrong, (2) how likely it is, and (3) what its consequences might be (Reference 0-1).

Risk-Informed

A characteristic of decisionmaking in which risk results or insights are used together with other factors to support a decision (Reference 0-1).

Robustness

The degree to which deviations from a “best” decision provide suboptimal values of the desired criterion. These deviations can be due to uncertainty in model formulation, assumed parameter values, etc. (Reference 0-3).

Sampling

The process of selecting some part of a population to observe, so as to estimate something of interest about the whole population (Reference 0-4).

Sampling Uncertainty

The uncertainty in an estimate of a quantity of interest that arises due to finite sampling. Different sets of model realizations will result in different estimates. This type of uncertainty contributes to uncertainty in the true value of the quantity of interest and is often summarized using the sampling variance.

Sampling Variance

The variance of an estimate of a quantity of interest that arises due to sampling uncertainty (i.e., finite sampling). An estimate of this variance is often used to summarize sampling uncertainty.

Sensitive Variable

A variable that has a significant influence on the response (Reference 0-10).

Sensitivity Analysis

The study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input (Reference 0-10).

Sensitivity Metrics

Quantitative values that characterize the relationship between input and output variables. The following two metrics can be used:

(1) First-order sensitivity indices measure the proportion of the uncertainty in the output that is explained by the uncertainty in a single input.

(2) Total-order sensitivity indices measure the proportion of the uncertainty in the output that is explained by the uncertainty in an input and its interactions with other inputs (Reference 0-10).

Sensitivity Studies

Probabilistic fracture mechanics analyses that are conducted under credible alternative assumptions (Reference 0-11).

Significant

A factor that can have a major or notable influence on the results of a risk analysis (Reference 0-1).

Simulation

The execution of a computer code to mimic an actual system (Reference 0-3). Typically comprises a set of model realizations.

Software Quality Assurance

A planned and systematic pattern of all actions necessary to provide adequate confidence that a software item or product conforms to established technical requirements; a set of activities designed to evaluate the process by which the software products are developed or manufactured (Reference 0-12).

Solution Verification

The process of determining as completely as possible the accuracy with which the algorithms solve the mathematical-model equations for a specified quantity of interest (Reference 0-3).

State-of-Knowledge Uncertainty

See “Epistemic Uncertainty” (Reference 0-1).

Statistic

A numerical characteristic of a sample, such as the sample mean and sample variance (Reference 0-4).

Statistical Model

A description of the assumed structure of a set of observations that can range from a fairly imprecise verbal account to, more usually, a formalized mathematical expression of the process assumed to have generated the data (Reference 0-4).

Stochastic Uncertainty

See “Aleatory Uncertainty” (Reference 0-1).

Subjective Probability

Expert judgment about uncertain events or quantities, in the form of probability statements about future events. Not based on any precise computation but often a reasonable assessment by a knowledgeable person (Reference 0-3).

Surrogate

A function that predicts outputs from a model as a function of the model inputs (Reference 0-3). Also known as response surface, metamodel, or emulator.

Uncertainty

Variability in an estimate because of the randomness of the data or the lack of knowledge (Reference 0-1).

Uncertainty Analysis

A process for determining the level of imprecision in the results of the probabilistic analysis and its parameters (Reference 0-1).

Uncertainty Distribution

See “Probability Distribution” (Reference 0-1).

Uncertainty Interval/Range

A range that bounds the uncertainty value(s) of a parameter or analysis result by establishing upper and lower limits (see “Confidence Interval,” “Probability Distribution”) (Reference 0-1).

Uncertainty Propagation

Characterizing the uncertainty of a model’s responses that results from the propagation through the model of the uncertainty in the model’s inputs (Reference 0-3).

Uncertainty Quantification

The process of characterizing all relevant uncertainties in a model and quantifying their effect on a quantity of interest (Reference 0-3).

Validation

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 (Reference 0-3).

Variable

Some characteristic that differs from subject to subject or from time to time (Reference 0-4).

Variance

The second moment of a probability distribution, defined as $E(X-\mu)^2$, where μ is the first moment of the random variable $X.$ A common measure of variability around the mean of a distribution (Reference 0-3).

Verification

The process of determining whether a computer program (“code”) correctly solves the mathematical-model equations. This includes code verification (determining whether the code correctly implements the intended algorithms) and solution verification (determining the accuracy with which the algorithms solve the mathematical-model equations for specified quantities of interest) (Reference 0-3).

Glossary References

0-1. Drouin, M., Gonzalez, M., Herrick, S., Hyslop, J.S., Stroup, D., Lehner, J., Pratt, T., Dennis, M., LaChance, J., and Wheeler, T., “Glossary of Risk-Related Terms in Support of Risk-Informed Decisionmaking,” NUREG-2122, U.S. Nuclear Regulatory Commission, 2013.

0-2. Kumar, M., “Guide for Verification and Validation in Computational Solid Mechanics,” ASME V&V 10, American Society of Mechanical Engineers, 2006.

0-3. National Research Council, *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*, The National Academies Press, 2012.

0-4. Everitt, B.S., and Skrondal, A., *The Cambridge Dictionary of Statistics, 4th Edition*, Cambridge University Press, United Kingdom, 2010.

0-5. National Aeronautics and Space Administration, “Standard for Models and Simulations,” NASA-STD-7009A, 2016.

0-6. U.S. Nuclear Regulatory Commission, “Glossary,” Web site last updated March 19, 2020, http://www.nrc.gov/reading-rm/basic-ref/glossary.html.

0-7. Hamby, D.M., “A review of techniques for parameter sensitivity analysis of environmental models,” *Environmental Monitoring and Assessment*, 32.2: 135–154, 1994.

0-8. Raynaud, P., Kirk, M., Benson, M., and Homiack, M., “Important Aspects of Probabilistic Fracture Mechanics Analyses,” U.S. Nuclear Regulatory Commission, 2018.

0-9. Pasanisi, A., Keller, M., and Parent, E., “Estimation of a quantity of interest in uncertainty analysis: Some help from Bayesian decision theory,” *Reliability Engineering & System Safety*, 100, 93–101, 2012.

0-10. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., and Tarantola, S., Global Sensitivity Analysis. The Primer, John Wiley and Sons, Ltd., 2008.

0-11. Erickson Kirk, M., Dickson, T., Mintz, T., and Simonen, F., “Sensitivity Studies of the Probabilistic Fracture Mechanics Model Used in FAVOR,” NUREG-1808, U.S. Nuclear Regulatory Commission, 2010.

0-12. IEEE Standards Board, *IEEE Standard Glossary of Software Engineering Terminology*, Institute of Electrical and Electronics Engineers, 1990.