Efficient Approaches to the Treatment of Uncertainty in Satisfying Regulatory Limits

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

David Grabaskas

Graduate Program in Nuclear Engineering

The Ohio State University

2012

Dissertation Committee:

Professor Tunc Aldemir, Advisor

Professor Richard Denning

Professor Marvin Nakayama

Professor Alper Yilmaz

Abstract

Utilities operating nuclear power plants in the United States are required to demonstrate that their plants comply with the safety requirements set by the U.S. Nuclear Regulatory Commission (NRC). How to show adherence to these limits through the use of computer code surrogates is not always straightforward, and different techniques have been proposed and approved by the regulator. The issue of compliance with regulatory limits is examined by rephrasing the problem in terms of hypothesis testing. By using this more rigorous framework, guidance is proposed to choose techniques to increase the probability of arriving at the correct conclusion of the analysis. The findings of this study show that the most straightforward way to achieve this goal is to reduce the variance of the output result of the computer code experiments.

By analyzing different variance reduction techniques, and different methods of satisfying the NRC's requirements, recommendations can be made about the best-practices, that would result in a more accurate and precise result. This study began with an investigation into the point estimate of the 0.95-quantile using traditional sampling methods, and new orthogonal designs. From there, new work on how to establish confidence intervals for the outputs of experiments designed using variance reduction techniques was compared to current, regulator-approved methods. Lastly, a more direct interpretation of the regulator's probability requirement was used, and confidence intervals were established for the probability of exceeding a safety limit. From there,

efforts were made at combining methods, in order to take advantage of positive aspects of different techniques.

The results of this analysis show that these variance reduction techniques can provide a more accurate and precise result compared to current methods. This means an increased probability of arriving at the correct conclusion, and a more accurate characterization of the risk associated with events. While several of these methods are asymptotic in nature, which presents potential drawbacks, issues of convergence appear to be outweighed by the reduction in variance, and improvement of the information contained in the results. Using this knowledge, recommendations were made about the applicability of these methods in the field of reactor safety, and about future regulatory limits and their implications.

Dedication

This document is dedicated to my family and wife Pamela.

Without their support, this work would not have been possible.

Acknowledgments

I would like to thank Professor Richard Denning and Professor Tunc Aldemir for their advice and guidance during my graduate work. I also greatly appreciate the freedom which they gave me to follow the path that my research led, and the flexibility to work at the pace I desired, and to pursue the goals I set forth.

I would like to thank Professor Marvin Nakayama for his substantial assistance in the fields of mathematics and statistics, without which this work could not have been completed.

I would also like to thank Professor Alper Yilmaz for taking on the additional responsibilities that were associated with assisting with my work, and offering another perspective on techniques and methods.

I also thank the Nuclear Regulatory Commission for their support of this research through an NRC Fellowship.

Thank you to my parents, Stan and Jean Grabaskas, for their unwavering love and support throughout my years at the Ohio State University.

Finally, I would like to thank my wife Pamela, the best other half a man could have.

Vita

2004	Hubbard High School
2008	B.S. Mechanical Engineering, Ohio State University
2010	M.S. Nuclear Engineering, Ohio State University
2010-Present	Graduate Fellow, Nuclear Engineering Program, Department of Mechanical and Aerospace Engineering, Ohio State University

Publications

1. D. Grabaskas, R. Denning, T. Aldemir, M. Nakayama, "The Use of Latin Hypercube Sampling for the Efficient Estimation of Confidence Intervals," Proceedings of the 2012 International Congress on Advances in Power Plants (ICAPP'12), June 2012

2. D. Grabaskas, R. Denning, T. Aldemir, "Techniques for the Efficient Assessment of the 0.95-Quantile in Safety Analyses," Proceedings of the Nuclear Thermal Hydraulics, Operations, and Safety Conference (NUTHOS-9), September 2012

3. D. Grabaskas, R. Denning, T. Aldemir, "Efficient Assessment of Epistemic Uncertainty Margins," Proceedings of the Verification and Validation for Nuclear Systems Analysis Workshop II, May 2010

4. R. Denning, A. Brunett, M. Umbel, T. Aldemir, D. Grabaskas, "Toward More Realistic Source Terms for Metallic-Fueled Sodium Fast Reactors," Proceedings of 2011 International Congress on Advances in Power Plants (ICAPP'11), June 2011

Fields of Study

Major Field: Nuclear Engineering

Table of Contents

List of Tables

List of Figures

Chapter 2

Chapter 3

List of Acronyms

AEC – Atomic Energy Commission

AV – Antithetic Variates

BFD – Backward Finite-Difference

CCDF – Complementary Cumulative Distribution Function

CDF – Cumulative Distribution Function

CFD – Central Finite-Difference

CI – Confidence Interval

CLT – Central Limit Theorem

CMC – Crude Monte Carlo

CMC-OS – Crude Monte Carlo Order Statistics

DOE – Design of Experiment

ECCS – Emergency Core Cooling System

LBE – Licensing Basis Event

LHS – Latin Hypercube Sampling

LOCA – Loss of Coolant Accident

MC – Monte Carlo

NRC – Nuclear Regulatory Commission

OA – Orthogonal Array

OLHC – Orthogonal Latin Hypercube

OSCI – One-sided Confidence Interval

PCT – Peak Clad Temperature

PDF – Probability Distribution Function

P-Method – Confidence for Probability Method

PRA – Probabilistic Risk Assessment

PWR – Pressurized Water Reactor

Q-Method – Confidence for Quantile Method

RISMC – Risk-Informed Safety Margin Characterization

rLHS – Replicated Latin Hypercube Sampling

SA – Sensitivity Analysis

SAN – Stochastic Activity Network

S.D. – Standard Deviation

TMI – Three Mile Island

UA – Uncertainty Analysis

VRT – Variance Reduction Technique

ZNPP – Zion Nuclear Power Plant

Chapter 1: Introduction

1.1 Problem Description

Since the inception of commercial nuclear power, utilities operating nuclear power plants have been required to meet safety objectives set forth by the U.S. Nuclear Regulatory Commission (NRC), and its predecessor, the Atomic Energy Commission (AEC). These safety guidelines have made it necessary for the power plant operators to demonstrate that their plants comply with the requirements set in place to protect public health. Due to the large cost, complexity, and potential hazards of nuclear power, demonstration of compliance cannot be done through integral experiments with actual operating plants. Instead, complex computer codes were developed that simulate the plant's response to a variety of situations. The codes are validated by comparison with experiments that typically involve some degree of scaling or simulation. The codes then act as a surrogate for the real nuclear systems they represent. It is the responsibility of the nuclear safety analyst to use these computer codes in an effort to determine whether the performance of a nuclear power plant would satisfy safety requirements under a given set of conditions. The analyst does this by performing a series of code calculations that predict how the plant would perform over the range of anticipated accident scenarios. The results of the analysis are then compared with a safety limit to determine acceptability. The applicability of these computer code results to the assurance of satisfaction of safety

requirements is not always clear, and there has been great debate over the interpretation of parameters and limits. In recognition that there are uncertainties in the ability of the computer code to represent the actual plant behavior, the historical safety approach taken by the NRC was to incorporate non-mechanistic conservatism in the analysis models.

More recently, the NRC has allowed the licensee to perform best-estimate plus uncertainty analyses for comparison with safety limits. However, questions remain about the interpretation of these requirements, and how to best demonstrate adherence to them. This shift to risk-informed safety analysis has provided utilities with flexibility in their analysis methods, but deterministic conservatism is still prevalent in the requirements of the NRC. Techniques that increase the accuracy of these best-estimate safety analysis methods, while reducing unnecessary conservatism, are of value to both the regulator and utility.

1.2 Objective

The objective of this work is to identify statistical methods that can most efficiently increase the probability of reaching the correct conclusion during a safety analysis comparison to regulatory limits by increasing the accuracy and precision of the results of these computer code experiments. This analysis will focus less on the suitability of these computer codes to act as a surrogate for the actual systems, but will emphasize the examination of the output results of the computer code experiments and how they pertain to set limits and constraints. The analysis will involve several tasks which will be performed using numerical experiments:

1) Compare techniques for the estimation of quantiles of the output distributions of numerical experiments. This includes newer experiment designs, such as orthogonal Latin hypercubes. The goal is to determine the most efficient techniques, and to demonstrate whether it is possible to achieve the same level of accuracy when using fixed input values, as compared to a form of random sampling.
2) Explore alternatives to the NRC-approved method of crude Monte Carlo using order statistics for the establishment of confidence intervals for the quantiles of an output distribution of a computer code analysis. This includes determining the applicability of recent work on variance reduction techniques to the goals set by the NRC.
3) Explore alternative methods of satisfying the NRC's statistical requirements. The exploration includes an examination of the use of confidence intervals for probability estimations rather than for estimated quantiles of the output distributions. The examination consists of a comparison between methods and attempts at combining the methods.
4) Comment on and discuss the results of these analyses, and the implications towards future regulatory guidelines.

1.3 Scope

This work will expand on previous research in the statistical and computer science fields. It will focus on the demonstration of recently published statistical techniques and comparisons to current, regulator-approved methods. This will be done using systems

representative of those encountered during a nuclear power plant safety analysis. The goal is to assess the applicability of these methods in the field of nuclear safety analysis, and to provide guidance about the best-practices to achieve the highest probability of correct conclusion when comparing results to a limit value.

Accomplishing this goal will entail formulating a more rigorous approach to the limit value comparison, which will allow for a uniform assessment of different methods and techniques. The expectation is that a more detailed examination of the process of comparisons to a limit value will allow the strengths and potential weakness of current methods to be seen. The more detailed examination will also provide direction to the areas where there is the biggest room for improvement. From there, through the use of representative experiments, the prospective improvement from newer techniques can be evaluated.

1.4 Dissertation Overview

Chapter 2 of this work will provide background and historical context for the problem analyzed. This includes the regulatory motivations, a comparison of various uncertainty and sensitivity analysis techniques, and a more formal phrasing of the problem using hypothesis testing. Chapter 3 focuses on the techniques used to estimate quantiles of output distributions. New work in the field has provided techniques which may offer a step-forward in sampling methodology. This chapter also describes, in detail, several systems which were used throughout this work to test statistical methods. The systems are chosen to mimic situations encountered in nuclear safety analysis. Chapter 4 focuses on the methods to establish confidence intervals for the quantiles of output

distributions. Once again, recent developments in the field have provided new options, which need to be benchmarked against current techniques. Chapter 5 goes beyond current regulatory practice, examining the meaning of certain regulatory requirements, and proposing alternative methods not based on quantiles of output distributions, but on probability. New combinatory techniques are also examined as a way of improving the chances of analyses achieving the correct result. Chapter 6 offers a discussion of the results and the applicability of the methods analyzed to actual nuclear power plant safety analyses. This also includes a segment on recommendations for future work and research.

Chapter 2: Background

This section begins with a history of U.S. nuclear power plant governmental regulatory bodies, and the evolution of their guidelines (Section 2.1). Secondly, an overview of hypothesis testing is presented (Section 2.2). The purpose of this overview is to provide a more rigorous framework for the process of comparisons of computer code outputs to regulatory limits. This framework will be used throughout this work. Lastly, a brief overview of uncertainty and sensitivity analysis methods is presented in order to provide background on the current techniques available and to offer a point of comparison to the methods documented in Sections 3, 4, and 5 (Section 2.3).

2.1 Regulatory Background

This section provides a history of regulatory limits imposed by the NRC and other regulatory bodies, and their interpretation. It also includes a discussion about proposed future regulatory guidelines and possible restrictions the limits may impose.

2.1.1. History of Regulatory Bodies

While the Atomic Energy Act of 1946 marked the development of the Atomic Energy Commission (AEC) to oversee nuclear power in the U.S., it wasn't until the subsequent Atomic Energy Act of 1954 that commercial nuclear power began to appear as a reality. With this revision to the original law, the federal government made it possible for private companies to gain access to restricted data about the production of

nuclear power. This law also assigned the AEC with the dual mandate of both "encouraging widespread participation in the development and utilization of atomic energy", and the role to "protect the health and safety of the public [1]."

While the main focus of the AEC was the protection of the public's health and safety, many in the commission were aware that overly restrictive regulation could endanger the industry's future. As AEC Commissioner Willard F. Libby remarked, "Our great hazard is that this great benefit to mankind will be killed aborning by unnecessary regulation [2]." However, the AEC realized that assurance of reactor safety was a must, as a single accident could deal a death-blow to the industry as a whole. What was not as clear was which requirements the AEC should mandate in order to demonstrate reactor safety.

The formulation of guidelines was also hindered by the fact that the AEC was assigned the onus of constructing at least six pilot plants of different designs. This made universal standards difficult and the licensing of new plants began on a case-by-case basis. These varied reactor designs, coupled with limited operating experience and material property knowledge, meant that most safety questions were a matter of engineering judgment and safety analysis was not constrained by concrete or quantifiable goals [2]. The development of a more structured regulatory process coincided with the formation of the Advisory Committee on Reactor Safeguards, which was a panel of outside experts who would conduct their own independent review of plant applications and regulatory structure.

In the 1960s, the nuclear industry grew rapidly from small demonstration reactors to orders for substantially larger plants. At that point, concern rose as to the adequacy of protective features in the event of loss of coolant accidents (LOCAs). The Ergen Study, commissioned by the AEC, indicated that emergency core cooling systems would be required for those plants [3]. The AEC then initiated a substantial research program to develop a computational capability to analyze the plant response to a LOCA and an experimental program to assist in model development and validation. The initial computer codes were extremely crude relative to modern capabilities. In addition, loss of coolant experiments indicated that the two-phase flow phenomena associated with reflooding the reactor and quenching an over-heated core were complex. In response to these considerations, regulatory guidelines were designed in order to account for these potential deficiencies.

2.1.2. Evolution of Design-Basis Safety Requirements

The initial approach to safety analysis is referred to as deterministic, in that uncertainties were not considered in a statistical manner. Under a deterministic approach, "Regulators … simply tried to imagine "credible" mishaps and their consequences at a nuclear facility and then required the defense-in-depth approach—layers of redundant safety features—to guard against them [2]." These deterministic measures focused on the use of conservative assumptions, large safety margins, and layers of redundant and diverse safety systems. Based partly on the results of the Ergen study and Loss of Flow Tests (LOFT) experiments, an emergency core cooling rulemaking led to the development of acceptance criteria for accident response during a LOCA. Prescriptive

acceptance criteria for emergency core cooling system performance were provided in Appendix K to Part 50 of the Code of Federal Regulations [4] in early 1974. More detailed information about the criteria established in Appendix K, along with methods used to demonstrate adherence to these limits, is provided in Section 4.

By late 1974, President Nixon asked congress to create a new agency with the sole focus of industry regulation. This marked the end of the AEC, and the Nuclear Regulatory Commission (NRC) began operations in 1975. The NRC was now the final arbiter of regulatory issues, and was not hampered with the developmental issues of the AEC. The following year, the final version of a major reactor risk study, WASH-1400 [5] or Rasmussen Report, on the probability of severe accidents at nuclear power plants was issued. While the report represented a major step forward in safety analysis through the use of Probabilistic Risk Assessment (PRA), criticism over the data used in report and the projected pathways to a major accident resulted in the NRC withdrawing its endorsement of its conclusions [2]. However, the accident at Three Mile Island (TMI) in 1979 led to a reevaluation of the NRC's safety requirements, and its view of PRA.

After TMI, applications for new reactors stopped, and the NRC turned its attention towards decommissioning and plant renewals. As the NRC began issuing guidelines on applications for life extensions, the industry began to push back against, what they thought, were onerous regulatory measures. A report by the Tower Perrin consulting firm in the 1990's criticized the NRC for a regulatory approach which it viewed as "negative and punitive" and not focused on the prioritization of risk [2]. Among these complaints was the argument that NRC guidelines focused too heavily on

deterministic regulations that left industry little flexibility in carrying out safety analyses. The report also recommended the use of performance-based regulations, and the ability for plants to perform risk analyses, like PRAs.

The use of PRA had been debated at the NRC since WASH-1400, and more studies were conducted in the late 1980's. Initially, the NRC considered PRA to be only a safety research activity. However, many of the aspects of the accident at TMI were effectively predicted by WASH-1400. PRAs also offered the ability to prioritize events based on risk, and the opportunity to lessen possibly overly-conservative deterministic approaches. One of the conclusions of WASH-1400 was that reactor risk is dominated by Beyond-Design-Basis Accidents. For some specific scenarios, such as anticipated transients without scram and station blackout accidents, the question of adequate protection was raised and special requirements have been established, which have effectively extended the design-basis and have become incorporated into the licensing basis. Thus, for those events, NRC provides regulatory oversight to assure compliance.

2.1.3. PRA in Regulatory Decisionmaking

The publishing of NUREG-1150, *Severe Accident Risks: An Assessment of Five U.S. Nuclear Power Plants* [6], in 1990 marked a turning point in the NRC towards risk-based concepts. Not only did it expand on the use of PRAs, but it incorporated uncertainty into the analysis, unlike WASH-1400 which addressed parameter uncertainty post-process. By 1995, the NRC issued a policy statement encouraging the use of PRA in all regulatory matters [7], and further guidance on the expectations and best practices when conducting a PRA followed in [8],[9],[10],[11].

While the NRC has accepted the move to PRA, it has not given up many characteristics of the deterministic analyses of the past. Instead, the NRC encourages the use of PRA "in a manner that compliments the NRC's deterministic approach and supports the NRC's traditional defense-in-depth philosophy [12]." Regulatory Guide (RG) 1.174 [9] was one of the first NRC documents to outline how deterministic and PRA methods could be used in combination for integrated regulatory decisionmaking. Recently, the NRC has sought to provide a clearer picture of how these two approaches can be used in unison. Figure 2. 1 shows an outline from the recently published NUREG-2150, *A Proposed Risk Management Regulatory Framework* [13], which details the process of balancing risk assessment and deterministic techniques (called the "traditional approach") in what it calls a *technical analysis*. This proposed approach by the NRC considers the uncertainty analyses and best estimate models of a risk assessment, but uses the upper bounds of these results in comparison to safety limits with built-in margin and conservatism. The safety margin is a combination of both regulatory and design margin, as shown in Figure 2. 2, and historical upper bound assumptions will be discussed in Section 3 and Section 4.

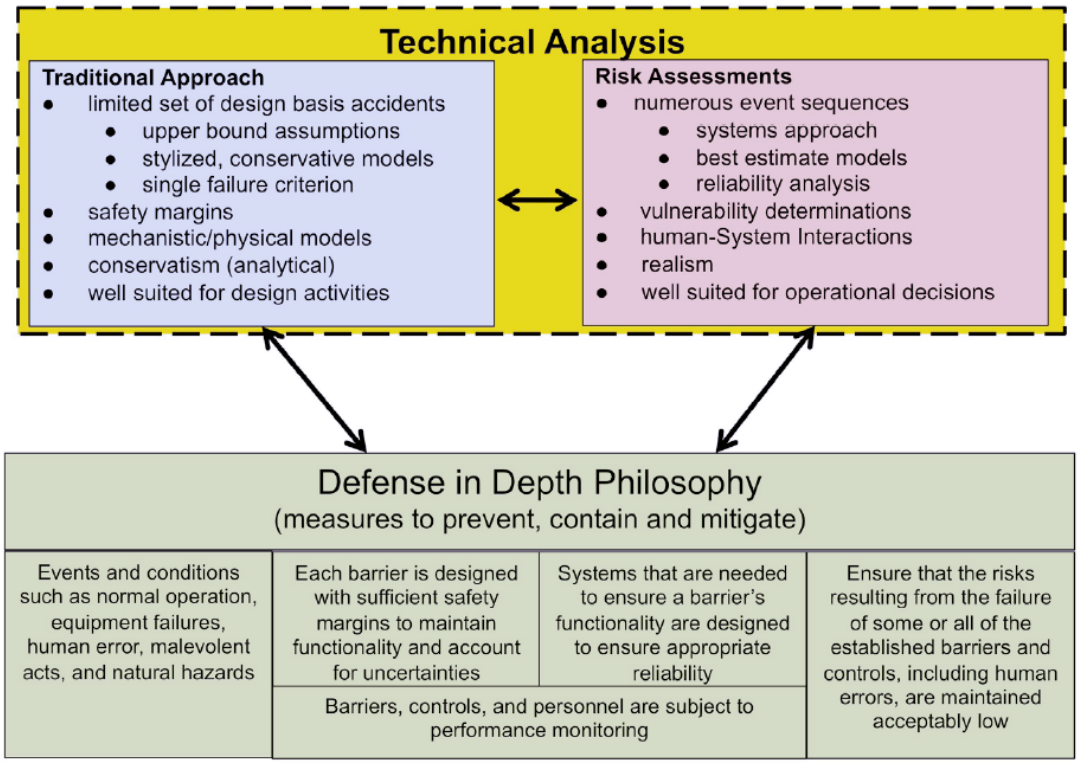

Figure 2. 1: Balancing Risk Assessments and Deterministic Techniques [13]

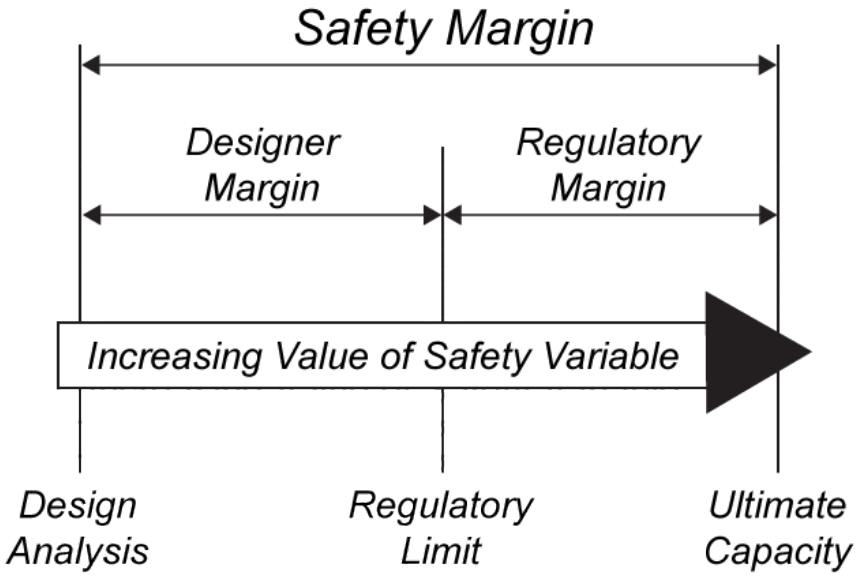

Figure 2. 2: Safety Margin Definition [14]

In this safety margin framework, there is conservatism not only in the value reported by the designer, but in the regulatory limit, which is placed well below the assumed ultimate capacity of the system. Recent work has sought to recharacterize this margin in a probabilistic, risk-informed manner as the distance between two uncertain parameters, the system load, and the system capacity [15]. As seen in Figure 2. 3, the potential hazard arises from the possible overlap between the two distributions, and the degree of overlap is constrained by the regulator.

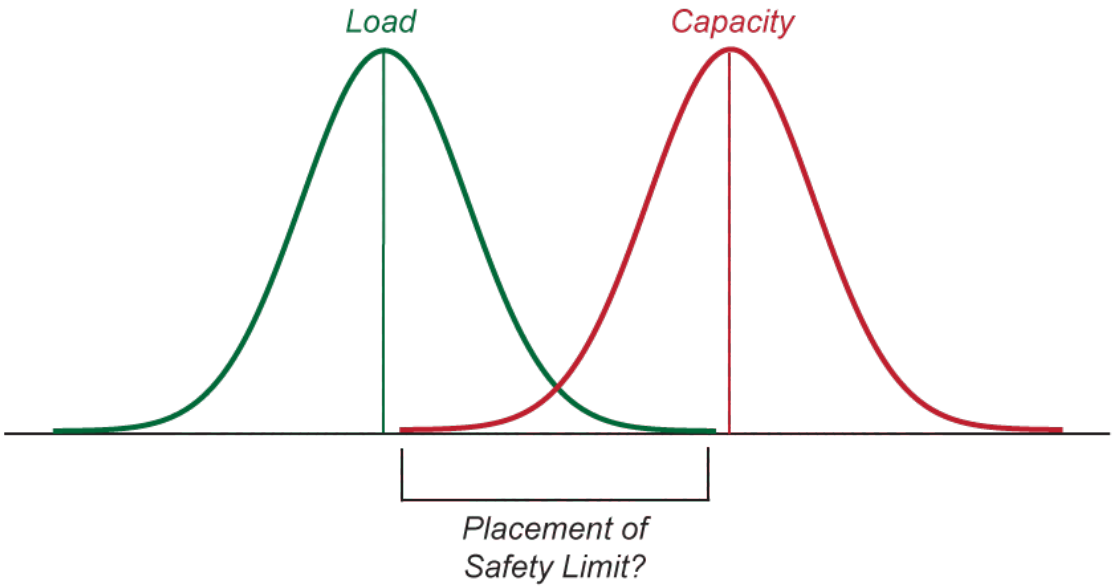

Figure 2. 3: Incorporation of Probabilistic Safety Margin

As will be discussed in Section 4, the location of the safety limits when using these distributions is still being considered.

It appears that the NRC's use of decision criteria and limits will continue for the foreseeable future. More detail on the specifics of limits currently in place, those limits

proposed for the future, and the techniques used to demonstrate adherence to these limits will be given in Section 4.

2.2 Hypothesis Testing

The process of testing the output of an analysis against a certain goal or limit can be explained more thoroughly using hypothesis testing. Hypothesis testing is a method to make decisions based on the resulting data of an analysis. Hypothesis testing usually begins with an assumption about a *parameter* of the analysis output. Here, a parameter is the true value, or a property of the full population distribution. On the other hand, the analysis or experiment provides a *statistic*. This statistic is based on the sample population used in the analysis, as explained in Table 2. 1.

Table 2. 1: Parameter and Statistic Definition

Parameter	True property of full population Ex: 0.95-quantile of the true distribution
Statistic	Property of sample Ex: 0.95-quantile estimation using *n* samples

In this process, an initial guess about a parameter will be made, then the test statistic will be found using experimentation (physical or computer modeling), and will be used as a point of comparison. The initial assumption is called the *Null Hypothesis* H_0, and usually refers to a default or general position. An example would be if an experiment was undertaken to test the effectiveness of a new drug. The null hypothesis would be the assumption that the drug has no effect. Conversely, there is the *Alternative Hypothesis*

H_1, which in the example above would be the statement that the new drug does have some non-random influence or effect.

It is important to note that the burden of proof is on the analyst to make the case for the alternative H_1. To make this clearer, it can be thought of as a court case in the judicial system. The burden of proof is on the prosecutor to prove that a defendant is guilty (H_1). If the evidence is insufficient of conviction, then the conclusion is the null hypothesis H_0, not guilty. The defendant is "innocent until proven guilty," in much the same way that H_0 is assumed true unless proven otherwise.

2.2.1. Types of Hypothesis Testing and Associated Errors

The actual process of testing can be done in various ways. Table 2. 2 shows a comparison of three possible procedures. In general, the three methods follow the same approach, but use different test statistics for comparison.

Table 2. 2: Three Approaches of Hypothesis Testing [16]

Step	Test Statistic Approach	P-Value Approach	Confidence Interval Approach
1	State H_0 and H_1	State H_0 and H_1	State H_0 and H_1
2	Determine test size α and find the critical value (CV)	Determine test size α	Determine test size α or 1-α, and a hypothesized value
3	Compute a test statistic (TS)	Compute a test statistic and its p-value	Construct the (1-α)100% confidence interval (CI)
4	Reject H_0 if TS > CV	Reject H_0 if p-value < α	Reject H_0 if a hypothesized value does not exist in CI
5	Substantive interpretation	Substantive interpretation	Substantive interpretation

The *test statistic approach* calculates a test statistic from the empirical data found during the analysis. This test statistic is then compared to a critical value, usually from a standardized normal distribution. The *p-value approach* calculates a probability, using

the test statistic, that reflects the measure of evidence against H_0 [17], referred to as the *p*-value. Whether the *p*-value is greater or less than the metric α (discussed below) determines if the results are "statistically significant." The *confidence interval approach* calculates a confidence interval around a statistic. It is then used to see if the hypothesized value falls into that interval or not. This approach will be examined in detail in Section 4. Usually, all three types of hypothesis testing can be reworded or reformulated to form a test that would fall under one of the other categories, as will be seen in Section 4.

With any hypothesis test, there are four possible outcomes, shown in Table 2. 3.

Table 2. 3: Hypothesis Testing Outcome Possibilities

	H_0 is true	H_1 is true
Accept H_0	Correct Conclusion 1-α	Incorrect Conclusion Type-II Error β
Accept H_1	Incorrect Conclusion Type-I Error α: Size of test	Correct Conclusion 1-β: Power of test

The first possible outcome is the correct conclusion to accept H_0 when indeed H_0 is true. The second outcome is the incorrect conclusion to reject H_0 when it is in fact true. This is referred to as a Type-I error, or a false positive. Its probability is given the value α, which is called the *size of the test* or significance level. It is important to note that this is not the actual size of the test (i.e. how many samples were taken or how many computer code runs were conducted), but the probability of committing a Type-I error. Conversely, the first outcome, accept H_0 when H_0 is true, occurs with probability $1 - \alpha$. Many times, the

value for α is assigned *a priori* as a measure of the willingness to accept false positives, and the test is designed to satisfy that requirement.

The third possible outcome is the correct conclusion to accept the alternative hypothesis H_1 when it is indeed true. The last outcome is the incorrect conclusion to reject H_1 when it is true. This is called a Type-II error, or a false negative, and occurs with probability β. On the other hand, the correct conclusion to accept H_1 when H_1 is true, occurs with probability $1 - \beta$, which is called the *power of a test*. Unlike α, β is usually not defined beforehand, but its value is dependent on the experiment design that was constructed for a specific α, and several other factors, as will be described in Section 2.2.2.

The following example is presented in order to make this testing process more comprehendible. For example, assume there is a regulatory safety limit with value G, that represents a prescribed limit that the true 0.95-quantile $\xi_{0.95}$ of the output of a safety analysis cannot exceed. In this case, the hypothesis test is defined with null hypothesis H_0: $\xi_{0.95} > G$ and alternative hypothesis H_1: $\xi_{0.95} \leq G$. This framework puts the burden of proof on H_1, which hypothesizes that the true 0.95-quantile value of the output falls below the prescribed limit.

In this example, the analyst carries out a certain number of computer simulations to estimate the 0.95-quantile of the system's output (more information on how to estimate a quantile value is presented in Section 3). This estimation of $\xi_{0.95}$, $\tilde{\xi}_{0.95}$ (the ~ will be used throughout this work to denote an estimated value), is then compared to the limit G using either a critical value, p-value, or confidence interval. Based on this result, H_1 will

either be accepted or rejected. Both errors, Type-I: the system appearing to satisfy the safety limit when the true quantile does not, or Type-II: the system appearing to fail the safety limit when the true quantile satisfies the limit, are possible depending on the location of the true 0.95-quantile. Based on the type of test used, an acceptable value for α would have been decided *a priori*, and the test would have been built around it.

2.2.2. Reducing Error

Both errors have negative impacts. Type-I errors would appear to be the more serious error since a system is being approved that should not be. However, Type-II errors also have drawbacks, since a system will be viewed as failing when it should not. This type of error could mean that time and resources will be dedicated to fixing a potentially non-existent problem, when they could have been applied more productively. The goal of the safety analysis should be to reduce both Type-I and Type-II errors (reduce α and β). This not only helps reduce false positives, but helps assure that safety measures are most effectively addressing true safety issues.

Since α, the probability of committing Type-I errors, is usually fixed (common values chosen are 0.05 or 0.01), then the focus is on reducing β. This can only be done in three ways:

1. Increase the distance (or Δ) between H_0 and H_1
2. Increase the sample size n
3. Increase α

Obviously, the third option is not acceptable, which leaves only two other methods. Increasing the distance between H_0 and H_1 will help reduce Type-II errors, but it is usually outside the analyst's control. The limit is often times set by some other

organization, and typically the system cannot be changed substantially. This leaves only one available option, increasing the sample size. In our example, this would mean more computer code runs, but even with today's technology, computer runs are expensive and time consuming. There is a possible solution though. The reason increasing the sample size reduces β is because it reduces the sample variance. This increases the precision of the sample statistic, which in turn makes the probability of error less likely. Therefore, the way to reduce β is to find techniques that decrease the sample variance without the need to increase the sample size.

2.3 Overview of Uncertainty and Sensitivity Analysis

This section gives a brief overview of various sensitivity and uncertainty analysis (SA and UA) methods. This is done in order to offer a better understanding of the current methods available for these analyses, and how the techniques detailed in this work compare to other methods in the field. Here, SA is defined as in the book *Uncertainty* by Morgan and Henrion [18] as "the computation of the effect of changes in input values or assumptions (including boundaries and model functional form) on the outputs", and UA as "the computation of the total uncertainty induced in the output by quantified uncertainty in the inputs and models, and the attributes of the relative importance of the input uncertainties in terms of their contributions." To state this explanation more simply, an UA is the investigation of the output space and its properties (usually mean, variance, higher moments, etc). SA is the examination of the relationship between input parameters and their effects on the value of the output.

This section will center on the analysis of input uncertainty. While many of these techniques can be used to handle model or parameter uncertainty, the main focus of this work is on methods related to input uncertainty. However, this input uncertainty may be either aleatory (stochastic in nature), or epistemic (relating to a lack of knowledge of specific value). Also, this review will focus on SA and UA methods based on classical or Bayesian probability theory. It will not cover some of the lesser known alternatives, like fuzzy logic [19], possibility theory [20], Dempster–Shafer (evidence) theory [21] [22], and info-gap decision theory [23]. While some very exciting developments are occurring in these fields, in a matter of brevity, they will not be discussed here.

For ease of reference, the following notation will be used. A system will involve d input variables, $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)$, and output variable y. The vector $\boldsymbol{x}$ is a realization of a random vector $\boldsymbol{X}$ (capital letters will denote random variables, and lower case will be a corresponding realization). Hence, y is also a realization of random variable Y, and the relationship between the input and output can be described in Eq. 1, which shows the output has its own probability density function (PDF). The total number of runs (where a run is the creation of a sample from the system) conducted will be denoted by n.

$$Y = f(X_1, X_2, \ldots, X_d) = f(\boldsymbol{X}) \qquad \text{Eq. 1}$$

To aid in the presentation of the material, this overview will be split into subsections based on the major methods of SA/UA. Section 2.2.1 will document screening designs, Section 2.2.2 will cover local methods, Section 2.2.3 will cover global methods, and Section 2.2.4 will cover the application of Bayesian techniques. Local methods will be defined as the process to determine the sensitivity (or in local methods

the derivative) of an input variable on the output at a given location. Global methods will be defined as the process to determine the sensitivity across the total input uncertainty space. Expanded definitions will be given in the following sections. Screening methods are documented separately since they are applicable at the local and global levels. Many of the methods detailed in these sections are summarized from [24] and [25].

Lastly, certain elements of UA will not be covered in this section for several reasons. First, the figures of merit being pursued by an analyst doing an UA can vary greatly depending on the analyst's goal. This review will focus on the most common goal, the moments of the output distribution. Second, many of these techniques, such as those to determine quantiles of the output distribution, or establish confidence bounds, will be discussed in length in Section 3, Section 4, and Section 5.

2.3.1. Screening Designs

The task of screening designs is to identify the most influential input variables amongst a large number of inputs. It is commonly assumed that in a model with many input variables, only a few inputs will dominate the response. This assumption is somewhat based on Pareto's law of income distribution within nations, where there are many potential factors, but the distribution can largely be characterized by several important variables [24]. Screening designs can be either local or global depending on how they are applied. However, most screening designs return only a qualitative assessment of the input factors' importance.

2.2.1.1. Classical/Global One-at-a-Time Methods

One-at-a-time (OAT) [26] methods consist of first conducting a model run with all input variables at their nominal value. Then, a single input is perturbed a set amount (usually the two extremes of the input range are chosen as perturbation values), while keeping the other variables at their nominal values. The change in the output between the nominal and perturbed trial reveal the influence of that variable. Here, classical OAT refers to a local analysis, which is only valid if the function's response at that point can be modeled linearly. This usually involves $2d + 1$ trials, and provides no information on the interaction between variables.

Global, or Morris, OAT [27] investigates the entire input uncertainty space. It takes r local measures at different points of the input range. Each value is then changed OAT and results in a total number of $2rd$ runs. This method does produce some information about input interactions, but only a qualitative indication. It does not specify which variables have interaction, just that it exists.

Both of these methods do depend somewhat on the given range of the input variables, but are not dependent on the input variable distribution, since the sample points are either chosen at the extremes, or uniformly from the range.

2.2.1.2. Full/Fractional Factorials

Perhaps the most common screening method involves full and fractional factorials [28]. A full factorial consists of splitting each input variable range into r levels, then conducting runs of all possible combinations of these levels. A positive feature of this technique is that it's possible to determine high level interactions between input variables

along with the main effects. However, the amount of runs necessary for a full factorial design grows quickly with the number of levels and variables; r^d number of runs is necessary, where r is the number of levels, and d is the number of variables. Obviously, these designs become infeasible for a large number of inputs on a complex system.

Fractional factorials reduce the number of runs necessary by using only certain combinations of input levels [28]. While this method results in fewer runs, it also means that higher order interactions will be lost, and certain assumptions must be made about the negligibility of these interactions. The designs can be formed in order to assure certain interactions can be accounted for. The level of interaction which can be found is often called the design Resolution. A Resolution III design means only main effects can be determined, and they will be confounded with interaction terms. This is not a problem if those interactions are considered negligible when compared to the main effects. Resolution IV designs allows two-factor interactions to be found along with main effects, but also results in an increase in runs. Resolution V allows three-factor interactions to be found, and so on. Usually, fractional factorials are designed using orthogonal arrays. Orthogonality helps ensure that effects are balanced, or sum to zero, across inputs. There are many subsets of fractional factorial designs, such as Taguchi arrays [29], which are a type of Resolution III orthogonal design. More information on fractional factorial resolution and orthogonal designs will be presented in Section 3.

2.3.1.3. Cotter's Design, Iterated Fractional Factorials, and Sequential Bifurcation

As a way around the interaction assumptions required for fractional factorials, Cotter proposed Systemic Fractional Replicate Design (SFRD) [30]. SFRD accomplishes this using the following procedure.

1. Initial run with all variables at their low levels
2. *d* runs with each factor, in turn, at its upper level, while others remain at low level
3. *d* runs with each factor, in turn, at its low level, while others remain at high level
4. One runs with all factors at high level

In total, SFRD requires $2d+2$ runs. The difference in the output between these runs can be used to estimate the order of importance of the variables. However, a potential drawback is that certain variable's effects may cancel out other effects, and there's no way for the analyst to check for this situation. This method also lacks precision, since input variables are only being evaluated at their extremes. For SFRD, the importance measures have a variance equal to $\sigma^2/4$, while a fractional replicate with n runs would have a variance of σ^2/n [24].

Iterated Fractional Factorial Design (IFFD) [31] uses fewer runs than there are variables, but allows for the estimation of main, quadratic, and two-factor interaction effects of influential variables. This is done by taking three levels per input variable, low, middle, and high, then using these values in a series of two-level and three-level, Resolution IV fractional factorial designs. It also takes advantage of folded matrices, which are matrices that are the mirror image of another matrix (i.e. all low values would be switched to high and high values would be switched to low) to filter out confounded

effects. Through stepwise regression, influential variables can be discovered. This process is best equipped for models that have very few influential input variables.

Bettonvil's Sequential Bifurcation (SB) [32] is a group-screening technique. For this method, it is required that the variables have known signs that the analyst can identify. These signs are important because this method involves grouping parameters into clusters. All the variables in a cluster are assigned the same level for a run (low or high). If a run shows that a cluster is not influential, those variables are dropped. If a run shows a cluster is influential, that cluster is split in two and the process is repeated, hence the name sequential bifurcation. At the end of this process, the only remaining clusters will include a single, important variable. This gives the main effects of those variables. If two-factor interactions are desired, the process can be repeated using a foldover technique. This process is very efficient in terms of computational effort, but requires a high level of analyst effort, as it is necessary to discern influential clusters for every trial [24].

2.3.1.4. Summary of Screening Designs

Each screening design has advantages and disadvantages depending on the particular analysis. OAT, full factorials, and SFRD allow for no assumptions about the interactions, but OAT and full factorials require a very large amount of runs, and SFRD lacks precision and can result in effects cancelling each other. Fractional factorials can determine main effects and interactions in a much more efficient manner, but requires certain assumptions to be made about the system. IFFD is efficient and effective, but only

if there are a few dominating variables. SB is efficient and simple to apply, but relies heavily on analyst knowledge and cannot capture higher order interactions.

While this section has focused on the use of screening designs to identify important variables, they can also be used for UA. Many of these designs, such as full and fractional factorials, can be used to create regression fits or response surfaces. These can be used to satisfy the more typical UA goals of identifying the moments and shape of the output distribution.

2.3.2. Local Methods

Local methods provide the slope of the model output distribution at a given set of values. For local SA, this slope is the goal of the examination, but for UA, local methods can provide a quick, efficient technique for a preliminary exploration of the model. This section will cover, what are commonly called, deterministic local methods. These methods do not express the output in terms of probability, or sample from an input distribution. They are best used on systems where the output can be expressed as a fixed, direct function of the inputs. These local methods usually fall under two categories, those that numerically solve for the slope, and those that analytically find the partial derivatives. In either case, extensive prior knowledge or assumptions about the distributions of the input variables is usually not needed, since only a small interval is being explored. However, some methods require detailed knowledge of model parameters.

2.3.2.1. Brute-Force Method (Indirect Method)

The Brute-Force Method (BFM) [24] consists of slightly perturbing input variables, one-at-a-time, around some nominal value. These perturbations, and their resulting output, allow for an estimation of the slope at the nominal value. This method uses the finite-difference approximation and relies on local linearity, so it is not suited for highly nonlinear systems, or models that vary many orders of magnitude in small intervals. While the process seems straightforward, it can consist of a period of trial and error. This is due to that fact that if the perturbation interval is selected too wide, it can violate the local linearity, but if the interval is too small, it is often dominated by the round-off bias of the model. This method usually requires d+1 runs, or 2d runs if central difference approximation is used. It is essentially a subset of OAT techniques.

2.3.2.2. Differential Methods (Direct Method, Green Function Method, Miller/Frencklach, Poly. Approx.)

To explain how differential methods work, it is necessary to start with the time-dependent, differential-algebraic equation seen in Eq. 2.

$$\frac{d\boldsymbol{y}}{dt} = f(\boldsymbol{y}, \boldsymbol{x}), \qquad \boldsymbol{y}(0) = \boldsymbol{y}^0 \qquad \text{Eq. 2}$$

Here, $\boldsymbol{y}$ is the vector of output variables, and $\boldsymbol{x}$ is the vector of input variables. Any change in $\boldsymbol{x}$ will also cause a subsequent change in the solution $\boldsymbol{y}^s$. This change can be expressed by a Taylor series expansion seen in Eq. 3,

$$\boldsymbol{y}^s(t, \boldsymbol{x} + \Delta\boldsymbol{x}) = \boldsymbol{y}^s(t, \boldsymbol{x}) + \sum_{j=1}^{k} \frac{\partial y_i}{\partial x_i} \Delta x_i + \frac{1}{2} \sum_{l=1}^{k} \sum_{j=1}^{k} \frac{\partial^2 y_i}{\partial x_l \partial y_j} \Delta x_l \Delta x_j + \cdots \qquad \text{Eq. 3}$$

where, $\partial f y_i / \partial x_j$ are called the first-order sensitivities (also called matrix **S**), $\partial^2 f y_i / \partial x_l \partial x_j$ are the second-order sensitivities, and so on. A differential analysis consists of the following four steps:

1. Base values and ranges for input variables are selected
2. A Taylor series approximation for the model is developed from the base values
3. Variance propagation techniques are used to estimate uncertainty in ***y***
4. Taylor series approximations are used to estimate the importance of input variables

In general, only the first-order sensitivities are found. However, to find these sensitivities, matrix **S**, the analytical solution to Eq. 2 must be known. This is only possible in the simplest cases. These next methods have been developed to overcome this shortfall.

The Direct Method (DM) [24] differentiates Eq. 2 with respect to x_i, as seen in Eq. 4, called the sensitivity differential equations.

$$\frac{d}{dt}\frac{\partial \boldsymbol{y}}{\partial x_i} = J\frac{\partial \boldsymbol{y}}{\partial x_i} + \frac{\partial f}{\partial x_i} \qquad \text{Eq. 4}$$

The matrix form can be seen in Eq. 5.

$$S = J\dot{S} + FS \qquad \text{Eq. 5}$$

Here, $J = \{\partial f_i / \partial \boldsymbol{y}\}$ is called the Jacobian matrix, and $F = \{\partial f_i / \partial x_i\}$, the parametric Jacobian. The DM solves the ODE in Eq. 4. However, to solve this ODE in an actual model, all the system parameters need to be known. This obviously becomes impractical as the model becomes more complex. The computational effort is also linearly proportional to the number of variables.

The Decoupled Direct Method (DDM) [33] [34] allows a numerical shortcut by exploiting a relationship between Eq. 2 and Eq. 4. Both these equations have the same Jacobian; therefore, the sensitivity equation can be solved with the original equation. It's important to note that the information about the system parameters is still necessary.

The Green Function Method (GFM) [24] differentiates Eq. 2 with respect to the initial values $\boldsymbol{y}^0$. This creates an initial value sensitivity matrix, or Green function, which is then able to be solved using more easily evaluated integrals. The DDM method is much easier to implement, and the GFM is only faster when there are many more system parameters than input variables.

The Method of Miller and Frenklach [35] uses a series of simpler empirical equations as a replacement for the model. It is very difficult, and time consuming, to find suitable empirical equations which can replace the more complex model, but if they exist, differentiating them can produce the same sensitivity results. Polynomial approximation uses Lagrange interpolation polynomials to approximate the solution of the sensitivity differential equations, but this method has not been applied to real world problems.

2.3.2.3. Forward and Adjoint Sensitivity Analysis Procedure

Forward Sensitivity Analysis Procedure (FSAP) [36] [37] uses Gâteaux differentials, which is a directional derivative that maps functions from one space to another. This means FSAP can find the differential of the original equation in the direction of the perturbation of the inputs, which yields a forward sensitivity system. This system then needs to be solved. This results in a computational effort equal to that of DDM. This is not surprising since FSAP is viewed as a generalization of DDM over a

total differentiation. This also means that, like DDM, FSAP becomes impractical for large systems with many parameters.

Adjoint Sensitivity Analysis Procedure (ASAP) [36] [37] reduces the computational effort by creating an adjoint function which only needs to be solved once. This can be done independently of the solution of the original model (unlike DDM which uses the same Jacobian to solve both equations). While ASAP is far more efficient than DDM or FSAP, it does require that an adjoint sensitivity system is available, and this construction may not be a trivial task. Still, ASAP tends to be the most efficient method for large-scale systems with many parameters.

2.3.2.4. Local Uncertainty Analysis

While local methods cannot completely provide the UA's goal, which would be to create a probability density function of the output, it can provide a first estimate result or give the basic characteristics of the function near that region. Using propagation of error, a linear estimate can be given for the variance of the model output based on the individual variables' derivatives. This linear estimate is essentially the sum of the contributions of the uncertainties from each input variable.

It is also possible to use propagation of error (propagation of moments) to find other moments of the output, such as the expected value, but this is usually done through non-deterministic methods, like sampling. While these techniques will be discussed in the global section, it is possible for them to be used on a smaller, local analysis. It is also important to note that with these deterministic local methods, the SA is conducted first,

and then any UA is conducted using those results. This is not the case for global, statistical techniques, where the process is reversed.

2.3.2.4. Summary of Local Methods

Overall, the BFM is by far the easiest method to implement, and requires no extra model development or differentiation using system parameters. However, it is time consuming and requires the use of trial and error to set the perturbation range. Of the differential methods, GFM is the most computationally expensive and is rarely used. DDM requires many model evaluations and scales linearly with the number of variables. FSAP becomes essentially equal to DDM in terms of computation, but can prove advantageous if the number of outputs exceeds the number of inputs. Both are impractical for large systems. ASAP is by far the most efficient method, but if the adjoint system is not developed at the same time as the model, it can be difficult to create.

2.3.3. Global Methods

Global methods will be defined here by the following two statements. First, all input variables must be varied simultaneously. Second, the complete range of the input values must be investigated. Unfortunately, this creates two problems. First, since all input variables are changed at once, the result is actually a multidimensional average of all the variables' effects. Second, the shape of the output distribution will be directly affected by the assumed input distributions. This second problem cannot be overstated and will be discussed further in this section.

2.3.3.1. Global Adjoint Sensitivity Analysis Procedure

Before reviewing the statistical global methods, it should be mentioned that there is one proposed deterministic global design. The Global Adjoint Sensitivity Analysis Procedure (GASAP) [24] is an extension of ASAP. Since it is not possible to use Taylor series at a global level (since Taylor series are a local concept), GASAP uses a global homotopy-based method. The method is complex, to say the least, and relies on both the forward and adjoint sensitivity systems to explore, exhaustively and efficiently, the entire input and output uncertainty space in order to determine important factors of the distributions (such as critical points, turning points, etc). As of now, GASAP has yet to be tried on a large scale system [24].

2.3.3.2. Monte Carlo Methods

Monte Carlo (MC) methods consist of drawing random, or quasi-random, samples from input probability distributions. The procedure includes the following five steps:

1. Define distributions of the uncertain input variables
2. Sample values from these input distributions
3. Evaluate the model at the sample points
4. Perform an UA using the results of the evaluation
5. Perform a SA by some type of mapping of the results to input variables

This section will be split into two subsections. The first subsection will discuss various means used to sample from the input distributions. The second subsection will include some of the techniques used to perform a SA using the results of a MC method.

Sampling

The most basic form of sampling is crude Monte Carlo (CMC) random sampling. In this method, samples are chosen randomly (or pseudo-randomly if a machine performs the selection, as explained below), from the input distributions. The samples selected will occur in direct proportionality to their probability distribution, which again shows the importance of the selected input distributions. Also, each sample is selected independently of the others, as long as there is no correlation between inputs. If there is correlation between inputs, additional steps need to be taken to assure the two input samples are properly related. This is often done through the use of rank-correlated pairing techniques.

Random sampling does have several weaknesses. First, there is no guarantee a certain region of the input distribution will be sampled, and usually a large number of runs is necessary to ensure proper coverage. Conversely, if the distribution has infinite tails, there is a non-negligible chance of sampling a negative or zero value. This can be avoided by truncating the distributions however [38]. Second, random sampling can be inefficient if samples are drawn from areas too close to previous samples. This, again, means large numbers of samples will be needed.

In order to improve on the efficiency of random sampling, many variance reduction techniques (VRTs) have been created. VRTs are methods used to improve the precision of a sampling scheme by either using previous knowledge of the inputs to reduce their variability, or tractable features of the model to adjust or correct outputs [39]. The most basic VRT is stratified sampling. This is the process of dividing the input

distribution into subregions, or strata. Then samples are drawn evenly from the different strata. This process helps ensure coverage over the input distribution. It can be carried out in different ways.

A basic form of stratified sampling is importance sampling (not to be confused with importance measures, which will be discussed later). Using this procedure, the input distributions are divided in a way that ensures the more "important" regions of the distributions will be sampled more frequently. This can be accomplished using analyst opinion; an example would be if an analyst knew that the higher values of an input will have a larger effect on the output, so the higher regions are divided into more strata than the lower values. It can also be carried out using a structured method, like the relation to the expected value of a similar variable. In essence, importance sampling modifies the input distribution in a way that allows the more influential regions of the original distribution to be sampled with more precision. It is important to note, the strata in importance sampling are often not of the same probability. The strata must then be weighted, according to their probability areas, in order for the final result to be consistent with the original distributions.

Perhaps the most popular form of stratified sampling is Latin Hypercube Sampling (LHS). LHS is a conceptually simple form of stratified sampling, where the strata are created using equal probability intervals. This method helps ensure coverage over the whole input distribution space, while not changing the distribution shape or requiring probability weights to be added post-process. LHS is the preferred method of stratification when there is little knowledge about the input variables. This is because

LHS does not depend on analyst opinion, or any ranking of input importance. Each input variable is divided on probability only. In terms of UA, LHS has repeatedly been shown to be more efficient than random sampling in the determination of the output's mean and population distribution. One positive of LHS is that it is possible to reweight samples if the shape of the input distribution has changed *a posteriori* [40]. This eliminates the need for repeated computer code or system runs, and allows the results that have already been found to be modified. LHS was analyzed in detail in this work, and more information about implementing the technique is provided in the following sections.

Another, relatively simply, VRT is the method of antithetic variates (AV). In this technique, a random input sample is selected, x_i (normalized $0 \leq x_i \leq 1$), then an additional sample is formulated that is negatively correlated to x_i, which in this case would be $x_{i2} = 1 - x_i$. Both samples are used to evaluate the model; then the two outputs, y_1 and y_2, are averaged. If the two samples were selected independently, the variance of the final output, $\bar{y}$, would be of the form seen in Eq. 6.

$$var(\bar{y}) = \frac{var(y_1)}{2} = \frac{var(y_2)}{2} \qquad \text{Eq. 6}$$

However, using antithetic variates, the variance is reduced by the formula seen in Eq. 7.

$$var(\bar{y}) = \frac{var(y_1) + var(y_2) + 2cov(y_1, y_2)}{4} \qquad \text{Eq. 7}$$

There are several weaknesses with this approach though. First, if the input variable has a skewed or non-symmetric distribution, the negative correlation metric will be reduced. The second issue arises if input variables are correlated. If proper steps aren't taken to handle this correlation, antithetic variates can actually result in a greater variance than

random sampling. More detail about the actual implementation of AV can be found in the Section 4.2.

The control variates method [39] is another VRT that, like importance sampling, that could use some prior knowledge about the system. There are many ways to apply control variates, but the most common technique is to use a known statistic similar to the output of interest. In this way, that known statistic is used to modify the output result. If the output of interest is $E[f(x)]$, and $\alpha = E[w(x)]$ is known. Then the output can be modified using Eq. 8,

$$f^*(x) = f(x) - c(w(x) - \alpha) \qquad \text{Eq. 8}$$

where c is a coefficient that can be optimized by using the covariance between $f(x)$ and $w(x)$, as seen in Eq. 9.

$$c = -\frac{cov(f,w)}{var(f)} \qquad \text{Eq. 9}$$

Using this approach, the variance of $f^*(x)$ is reduced by the order seen in Eq. 10.

$$var\left(f^*(x)\right) = \left(1 - \frac{[cov(f,w)]^2}{var(f)}\right) var(f(x)) \qquad \text{Eq. 10}$$

The obvious difficulty when using control variates is choosing and optimizing the control function $w(x)$.

It should also be noted that there are methods of quasi-random sampling. Perhaps the most well-known of these techniques are the Sobol' LP_τ sequences. These are referred to as low-discrepancy sequences, where discrepancy is a measure of the equidistribution of the points. By using these sequences to sample input variables, it is possible to converge to the solution at a faster rate than random sampling. It has also been proven

that quasi-random sampling can outperform random sampling and LHS for non-monotonic systems.

It should also be added here that there are possible problems when using a pseudo-random number generator. Almost every real analysis using a type Monte Carlo sampling will involve the use of a machine programmed with a pseudo-random number generator. While these samples may seem random, they are actually deterministic and will begin to repeat after some period. Usually this period is beyond the number of samples used in any experiment, but it is a possibility. Also, these pseudo-random number generators may return samples that are correlated in some fashion. In some cases, shuffling, where the seed value of the random number generator is changed between samples, is recommended to break up sequential correlations [41].

Sensitivity Analysis

As stated previously, there are many techniques to perform SA with MC methods, but they rely on the results of the UA. The simplest, and most intuitive, form of SA for MC methods is the examination of the scatterplots. This involves simply plotting the output against an input variable, then trying to distinguish trends, relationships, and thresholds. If there are only a few dominating variables, scatterplots can usually identify them with almost no additional work. Scatterplots are particularly useful with LHS since there is full stratification along the input variable range. However, scatterplots are only a qualitative tool, and it can be difficult to rank variables without a process of normalization or simple regression fits. A simple way to expand on a scatterplot is to use the Pearson product moment correlation coefficient (PEAR) [25]. This is just the linear

correlation coefficient between an input variable and the output. If a model is non-linear, it's possible to use the variables' ranks instead of the raw data; this is called the Spearman coefficient (SPEA).

It is possible to take this model fitting to the next step, in order to obtain a quantitative result, by doing a full regression analysis. Since this is a very popular method, and other techniques depend on its results, it will be reviewed in more detail. A simple regression fits develops a model, of the form seen in Eq. 11, by mapping between the input variables and the output,

$$\hat{y} = b_0 + \sum_{j=1}^{d} b_j x_j + \varepsilon \qquad \text{Eq. 11}$$

where b_j $(j = 1, \ldots, d)$ are the coefficients to be determined, and ε is the error term defined by the difference between the predicted output $\hat{y}$ and the actual output y. The most common method to determine the coefficients of Eq. 11is by using the method of least-squares. This technique is widely documented elsewhere, and will not be detailed here [42]. It is important to determine how well the regression model fits the actual data. A common metric to judge the fit is the sum of squares, which is defined in Eq. 12,

$$\begin{aligned} SS_{tot} &= SS_{reg} + SS_{res} \\ SS_{tot} &= \sum_{k=1}^{n} (y_k - \bar{y})^2 \\ SS_{reg} &= \sum_{k=1}^{n} (\hat{y}_k - \bar{y})^2 \\ SS_{res} &= \sum_{k=1}^{n} (\hat{y}_k - y_k)^2 \end{aligned} \qquad \text{Eq. 12}$$

where n is the total number of runs conducted, $\hat{y}_k$ is the regression model estimate, $y_k$is the actual run result, and $\bar{y}$ is the mean of y_k. These are referred to as the total sum of squares, regression sum of squares, and residual sum of squares. They can be used to find the ratio R^2, found in Eq. 13.

$$R^2 = SS_{reg}/SS_{tot} \quad \text{Eq. 13}$$

When R^2 is close to 1, it can indicate that the regression model is accounting for most of the uncertainty in the system. When using SS or R^2, it is important to check to make certain the model is not overfit to the data. This means the model is not fitting to the overall trend of the data, but including variations between sample point results. While this is usually not the case with linear regression fits, it is a possibility with response surfaces, which will be detailed later. Another method to determine the adequacy of a regression model uses the predicted error sum of squares (PRESS), seen in Eq. 13.

$$PRESS_k = \sum_{i=1}^{n} [y_i - \hat{y}_i(k)]^2 \quad \text{Eq. 14}$$

PRESS works by excluding a system observation from the regression fit, and then tests the fit's prediction against the actual value. The smaller the PRESS value, the better the regression fit predicted the data point.

Once a regression fit is created and determined to be adequate, a SA can commence. The first (and most obvious) way to rank the input variables is to normalize the coefficients of the regression fit. This will give a basic ranking of the input variables now that they are all on the same scale. One way to do this is to normalize the input and output variables to a mean of zero and standard deviation of one. These new coefficients

are called the standardized regression coefficients (SRCs). A formal definition is given in Eq. 15, where Eq. 11 has been algebraically reformulated to yield,

$$\frac{\hat{y}-\bar{y}}{\hat{s}}=\sum_{j=1}^{d}\frac{b_j\hat{s}_j}{\hat{s}}\frac{x_j-\bar{x}_j}{\hat{s}_j}$$
$$\bar{y}=\sum_{k=1}^{n}\frac{y_k}{n},\qquad \hat{s}=\left[\sum_{k=1}^{n}\frac{(y_k-\bar{y})^2}{n-1}\right]^{1/2} \qquad \text{Eq. 15}$$
$$\bar{x}_j=\sum_{k=1}^{n}\frac{x_{kj}}{n},\qquad \hat{s}_j=\left[\sum_{k=1}^{n}\frac{\left(x_{kj}-\bar{x}_j\right)^2}{n-1}\right]^{1/2}$$

where $b_j\hat{s}_j/\hat{s}$ are the SRCs, and can provide the variable importance.

It is also possible to determine importance by using the partial correlation coefficients (PCCs). This is done by developing two regression models; one model that includes the influence of all variables, and a model where the influence of all variables, other than the one of interest, has been removed. By a comparison of these results, the linear relationship between the input variables and the output can be determined, which has the other input variables' linear effects removed. In essence, the PCC is a measure of the strength of the linear relationship between an input and output after a correction has been made to account for the other variables, while the SRCs measure the effect on the output by a perturbation of an input value. They are similar, but can provide different importance measures.

There are drawbacks to using regression fits. Since it is based on a linear relationship between the inputs and output, it can perform poorly when the system in non-linear. This would yield a very low value for R^2 (below one and near zero, meaning a poor fit). However, as long as the relationships are monotonic, rank transformations can

be used to deal with nonlinearity. This is done by replacing either the input or output data with its rank order $(1, 2, 3, \ldots, n)$. Then a regression fit is made to this rank data, rather than the original data structure. If R^2 is larger using the rank transformation, the model has been improved. From there, the standardized rank regression coefficients (SRRCs) and partial rank correlation coefficients (PRCCs) can be found as before. It is important to note that the SRRCs and PRCCs now give data about the new regression model. Care must be taken translating these results back to the original system.

It should be noted here, that if the original design matrix were orthogonal (not randomly sampled, but using a design of experiment like a fractional factorial), determining the regression coefficients and correlation coefficients greatly simplifies to a single equation for each. This simplification has been part of the motivation to create other orthogonal or near orthogonal designs, such as orthogonal Latin hypercubes.

One last point on regression modeling; new regression techniques have been developed recently and should be mentioned. Argonne National Lab (ANL) has been developing a polynomial regression technique (using polynomial chaos expansion) which not only fits to the system input and output data, but uses the sensitivities to improve the regression surrogate [43]. This is done by automatic differentiation while the system code is conducting trials. These variable derivatives greatly increase the accuracy of the regression fit surrogate. While this process is still new, it is an interesting intersection of different methods of UA and SA. ANL claims that the calculation of these sensitivities while the code is running is possible, and has been demonstrated, but access to the system

codes and a large amount of effort would be needed to introduce this technique to an existing code. There is also Bayesian regression, which will be discussed in Section 4.

It is possible to use formal test statistics on a regression analyses, but one should be cautious, since many of these statistical techniques are based on assumptions that are not applicable to deterministic codes (codes that *always* produce the same result for a given input). Instead, two-sample tests, such as the Smirnov test [44], Cramer-von Mises test [45], the Mann-Whitney test [46], and the two sample t-test are sometimes used. These can be applied to the regression model or the original system. Most of these tests work by partitioning input variables based on quantiles of the output. If the input variable differs for the two quantile regions of the output, it can be viewed as influential. This provides only a qualitative ranking, and the result can vary greatly depending on what output quantiles are chosen and which samples are used.

There is a different approach to constructing a regression analysis, when there are many input variables, and some knowledge about their ranking is known. This method is known as stepwise regression and uses the following procedure:

1. A regression model is constructed using only the most influential input variable
2. A new regression model is constructed using the first and second most influential input variables
3. A third regression model is constructed using the first three most influential input variables

⋮

This process is continued until adding subsequent variables is no longer meaningful. Obviously, to conduct this technique, some type of ranking of the input variables must already exist. This can be done by developing a simple regression fit,

determining the SRCs, and using those in an effort to develop a better model. As before, it is especially important to keep track of R^2 and the overall model fit to ensure the analysis is ending at the right time.

2.3.3.3. Measures of Importance (Variance Based Methods)

Correlation Based

Measures of importance [25], or variance based methods, can be used in conjunction with MC methods. While there are different techniques, they all, essentially, try to provide a solution to Eq. 16, which is known as the *correlation ratio* [47].

$$\frac{Varx_j[E(Y|X_j = x_j)]}{Var(Y)} \quad \text{Eq. 16}$$

This is a ratio of the output variance based on input variable *x,* to the total output variance. The numerator of this equation is referred to as the variance correlation expectation (VCE), and the total ratio is called the correlation ratio. There are issues with this formulation though; it can be highly influenced by input variables with long-tailed distributions. This formulation tends to also lack robustness, and in an effort to increase the robustness, Iman and Hora [48] proposed using the following, modified form in Eq. 17.

$$\frac{Varx_j[E(logY|X_j = x_j)]}{Var(logY)} \quad \text{Eq. 17}$$

Here, the expectation value is estimated using regression analysis. While this increases robustness, it makes it more difficult to translate the results back to the original system. While these two equations might seem straightforward, estimating VCE can be difficult.

The detailed derivation can be found in literature [25], but often times it is necessary to conduct multiple replications of sampling schemes, like LHS, or requires resampling. Both methods can quickly become computationally expensive.

Method of Sobol'

Sobol'[49] used a different approach to determine input sensitivities. Under this method, the system *f(x)* is decomposed into summands of increasing dimensionality. This can be seen in Eq. 18.

$$f(x_1, \dots, x_d) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1 \le i < j \le d} f_{ij}(x_i, x_j) + \cdots + f_{1,2,\dots,d}(x_1, \dots, x_d) \quad \text{Eq. 18}$$

Using the fact that the integral of a summand over its own variables is zero, and that the summands end up orthogonal [24], the equation can be simplified down until a total variance, *D,* of *f(x)* can be reached, as seen in Eq. 19,

$$D = \int_{\Omega^k} f^2(x)dx - f_0^2 \quad \text{Eq. 19}$$

where Ω^k is the k-dimensional input variable space. Partial variances can be found using Eq. 20.

$$D_{i_1,\dots,i_s} = \int_0^1 \dots \int_0^1 f^2_{i_1,\dots,i_s}(x_{i_1}, \dots, x_{i_s}) dx_{i_1} \dots dx_{i_s} \quad \text{Eq. 20}$$

Then the sensitivity measures, $S_{i_1,\dots,i_s}$, are simply the ratio of the partial variances to the whole. S_i is the first-order sensitivity measure, which is the main effect of the input variable on the output. These sensitivities end up being the same as the most reduced version of the VCE analysis. S_{ij}, when $i \neq j$, is the second-order sensitivity, and measures

the interaction effect. This kind of analysis becomes impractical for systems with many factors though, since a separate sample of size *n* is required for each *S*, which would be *n* $\times 2^d$ total samples. The number of runs can be reduced by using a special sampling strategy called Winding Stairs [50].

Fourier Amplitude Sensitivity Test

The Fourier Amplitude Sensitivity Test (FAST) [51],[52], [53], [54] can arrive at the same sensitivities as Sobol', but uses a different path to get there. This is done by converting a multidimensional integral into a one-dimensional integral by using a search curve [25]. This permits sensitivities to be found for input variables independently. The mathematic explanation can be difficult, but by a transformation into *s* space, FAST converts the problem into a set of scalar variables, *s*, and angular frequencies, ω. This can be seen in Eq. 21, where G_i is the transformation function.

$$x_i = G_i(sin\omega_i s), \quad i = 1, \dots, d \qquad \text{Eq. 21}$$

An expectation value for the output can be found by the transformation out of *s* space, as seen in Eq. 22, where *f(s)* consists of the transformation functions G_i.

$$E(Y) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)ds \qquad \text{Eq. 22}$$

The variance can be approximated using Fourier series properties, as seen in Eq. 23,

$$\begin{aligned} Var(Y) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f^2(s)ds - [E(Y)]^2 \\ &\approx 2\sum_{j=1}^{\infty}(A_j^2 + B_j^2) \end{aligned} \qquad \text{Eq. 23}$$

where A and B are Fourier coefficients. In order to properly identify these coefficients, a certain number of runs must be conducted. There are various sampling strategies, but a search curve is the most frequently used. This curve changes all input variable values simultaneously, and systematically explores the input uncertainty space. There are different search curves that can be used, which explore the space in different manners. The contribution to the total variance of Y can be found using Eq. 24.

$$D_{\omega_i} \approx 2 \sum_{p=1}^{\infty} (A_{p\omega_i}^2 + B_{p\omega_i}^2) \qquad \text{Eq. 24}$$

The ratios of these individual variances to the total variance, Eq. 23, are the first-order sensitivities S_i, which are the same as the Sobol' sensitivities. The minimum sample size to compute D_i is $(2M\omega_{max}+1)$, where M is the maximum harmonic taken into consideration, and ω_{max} is the maximum frequency set. It is possible to find total sensitivity indices using, what is called, extended FAST, which gives information about the residual variance. FAST is much more efficient than Sobol' to obtain the same first-order sensitivities.

2.3.3.4. Response Surface Method

A response surface is an approximation for the model that is then used a surrogate for more detailed UA and SA [55]. This method consists of six steps:

1. Select ranges and distributions for input variables
2. Develop a DOE defining variable combinations for desired evaluations
3. Evaluate the model
4. Construct response surface
5. UA
6. SA

There are many ways to develop the design of experiment (DOE) in step two. Full and fractional factorials can be used, or MC sampling. While structured designs, like factorials, help ensure that specific interactions will be found, it is not possible to assign probabilistic weights. This is possible with MC sampling though, and can aid in the construction of the output's mean and variance. Just like regression fits, the method of least-squares is the most popular technique to complete step four. The resulting surface is used in the same way that the Taylor series is used in the differential analyses for SA. A drawback of the response surface method is that it usually takes a large number of runs to model interactions properly. Otherwise, assumptions need to be made about the magnitude and importance of these interactions.

2.3.3.5. Reliability Algorithms

Reliability algorithms [25] take a different approach than previous methods, and are best suited when the analyst is researching whether an output will exceed some failure criterion. These algorithms search the input space for the point that is most likely to lead to failure. Once this is determined, first-order (or second-order) sensitivities are found around that area. This gives information about how the different input variables are driving risk. These are known as First and Second Order Reliability Methods (FORM and SORM). These techniques do not use random sampling, but use optimization algorithms to search out these failure points. The math can be difficult, but the procedure can be summed up in three steps:

1. Formulate some performance function $g(\boldsymbol{X})$
2. Transform the problem: instead of using specific values of $\boldsymbol{X}$ to find a value of Y, a specific value of Y is used to determine the highest risk areas of $\boldsymbol{X}$
3. An optimization algorithm searches the input uncertainty space for the values which will give that specific Y value. This is done through a reliability index, which calculates the distance between an output point and the desired output point of Y. These algorithms use partial derivatives to converge on the highest risk areas of the input uncertainty space

This may not seem that difficult, but there's no guarantee of convergence, and the optimization algorithms may need tuning. This method has proven more efficient than MC sampling for some systems, but if a system is highly nonlinear, the partial derivatives will not help with convergence. This limits the use of FORM and SORM in large, highly complex, black-box type systems.

2.3.3.6. Global Methods Summary

Sampling based methods are the most common UA/SA techniques currently found in industry. They tend to be straightforward and easy to implement. However, they are *strongly* dependent on the assumed input distributions. These assumptions carry through the analysis to the UA and SA, and can greatly shape the results. They can become difficult to use if many input variables are correlated, or if the system is highly non-linear or non-monotonic. Also, randomly sampling can quickly become cost prohibitive, but VRTs or quasi-random sampling can help reduce the number of runs necessary. There are many different options to conduct a SA using sampling, but they all come after an UA, which is the opposite of the local methods in Section 2.2.2. Of these methods, studies have shown SRRCs and PRCCs to be very robust, especially when used in conjunction with LHS.

Variance based methods work off a simple principle, but can become difficult to implement. Even the most efficient method, FAST, is usually unworkable for a large system. Studies have shown they can handle non-monotonic systems though, but at a high computational cost.

Response surfaces are a very popular method for UA and SA since they allow a much simpler and faster running surrogate to be used in place of the original system. The biggest problem arises when determining the quality of the response surface. It is possible to create a very accurate response surface, but at the cost of many initial trials, which partially defeats the purpose of using a surrogate.

FORM and SORM are very innovative techniques which reverse the UA. It is a process to learn more about the input variables based on an overall goal. The methods can be efficient, but as the system becomes more complex, the optimization algorithms may begin to struggle. This can increase the number of trials necessary, or end up with a solution that cannot converge.

2.3.4. Bayesian Techniques

Bayesian techniques are based on the, now famous, formulation by Thomas Bayes, seen in Eq. 25,

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)} \qquad \text{Eq. 25}$$

where $p(\theta)$ is the prior, or marginal of θ, $p(y|\theta)$ is the likelihood, $p(\theta|y)$ is the posterior, and $p(y)$ is the normalizing constant, or marginal of y. In essence, what Bayes formula is expressing is the updating of prior knowledge, using the likelihood (new

data/knowledge), to arrive at a new probability, the posterior. The most commonly used form of Bayes theorem drops the normalizing constant, and instead uses the proportion seen in Eq. 26.

$$p(\theta|y) \propto p(y|\theta)p(\theta) \quad \text{Eq. 26}$$

2.3.4.1. Bayesian Linear Regression

While it is possible to use the formulation in Eq. 26 directly in an UA or SA, Bayes theorem is more commonly used as a way of determining the coefficients of a regression analysis, so that will be the method detailed here [56]. A regression fit can be seen as the conditional distribution of some output *y*, given *x*. A more formal expression is given in Eq. 27,

$$p(y|B,x) \quad \text{Eq. 27}$$

where, the vector *B* is the coefficients of the regression fit. The conditional mean is given in Eq. 28,

$$E(y_i|B,\sigma^2) = B_1x_{i1} + \cdots + B_kx_{id} \quad \text{Eq. 28}$$

where σ^2 is the conditional variance, and *i* is the trial number, i=1, 2,…, n. Then the likelihood can be viewed as the normal distribution in Eq. 29 [57].

$$p(y_i|B,\sigma^2) = N(B_1x_{i1} + \cdots + B_kx_{id},\sigma^2) \quad \text{Eq. 29}$$

The posterior probability, or desired result, becomes Eq. 30,

$$p(B,\sigma^2|Y,X) \propto \Pi_i p(y_i|B,\sigma^2) \times p(B,\sigma^2) \quad \text{Eq. 30}$$

where $p(B,\sigma^2)$, the prior, is usually non-informative (such as uniform$(B, log\sigma^2)$), unless the analyst is quite certain in a particular type of outcome. It is rare that this equation can

be solved analytically; instead sampling from the prior can be used. This is done by envisioning the marginal posterior as the integral in Eq. 31.

$$p(B|y) = \int p(B|\sigma^2, y)p(\sigma^2|y)d\sigma^2 \qquad \text{Eq. 31}$$

Then samples are taken from $p(\sigma^2|y)$; these are used to find $p(B|\sigma^2, y)$, and the resulting sample represents a point on the posterior. There are other ways to conduct sampling, and the most popular method is using a Gibbs sampler (Markov Chain Monte Carlo), which draws from the conditional distributions. When using sampling, it is important for the analyst to observe the output because the results usually do not begin with samples that are from the posterior distribution, but will converge to the correct distribution over a finite amount of time.

There are many different reasons why a Bayesian linear regression model would be preferred over the classical approach. In Bayesian analysis, it is relatively easy to build hierarchical models with many levels of uncertain variables. It also can be used in the case where there is missing data in the analysis. It's also possible to build "generalized" linear regression models, which can be fit to non-normal output distributions. Additional terms can also be added to the regression fit to help account for random (trial) error. There are also more detailed regression fit approaches, such as Bayesian Multivariate Adaptive Regression Spines (BMARS) [58]; these and other Gaussian process emulators can be used as a surrogate for the model itself, similar to a response surface, which allows a maximization of the UA and SA information from a limited amount of actual code trial data.

Chapter 3: Quantile Estimation and Orthogonal Arrays

If the goal of an analysis is to gain an understanding of the output of a system with uncertain inputs, the simplest way to characterize this output space is through the use of the distribution's moments. These moments include properties like the mean, variance, skewness, and kurtosis. However, finding the higher moments can involve conducting many system runs in order to discern these details. The easiest moment to estimate is the mean, but this provides very little information about the output, especially if the goal is to compare the extremes of the output distribution against some type of limit. The mean provides even less information if the output distribution is asymmetric (which is almost always the case when analyzing highly complex systems) or when the output ranges over several orders of magnitude, since it will give no indication of what percentage of the distribution falls above or below that value.

Instead of using the mean, or carrying out a large number of runs to determine the higher moments, quantiles of the output distribution can be estimated instead. A quantile is a point taken at an interval of a random variable's cumulative distribution function (CDF). For example, take the CDF shown in Figure 3. 1. Here, the quantile level p is set to 0.95, and the value of the 0.95-quantile $\xi_{0.95}$ is ≈ 706. What this means is that 95% of the distribution falls between $-\infty$ and ~ 706.

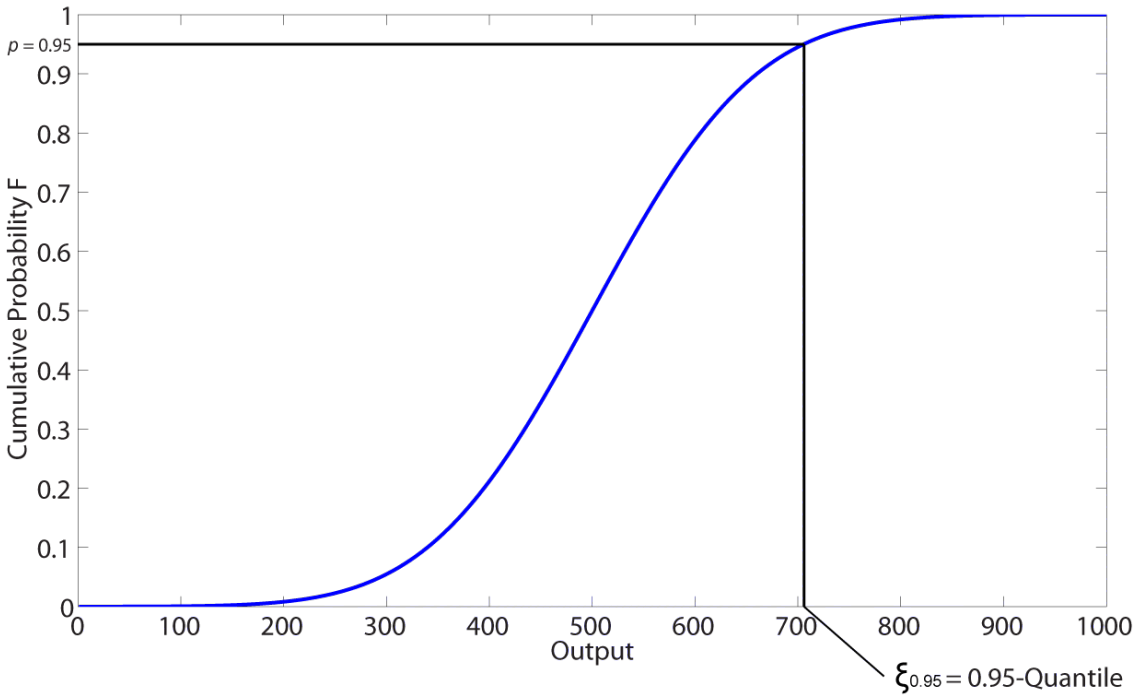

Figure 3. 1: CDF with 0.95-Quantile

This value is sometimes referred to as the 95th-percentile, where a percentile is just the quantile level p times 100. It can be seen that a quantile expresses more information than an estimation of the mean since it gives an indication of the percentage of the distribution that falls above or below that value. This is a much more useful metric than the mean if the goal of the analysis is to compare the output to a prescribed limit.

This section compares design of experiment (DOE) methods to estimate the 0.95-quantile of the output distribution of a system. The 0.95-quantile was chosen because it has historically been reported as the characterization of the upper regions of an output distribution in nuclear reactor safety analyses [6]. The goal is to find techniques that result in the most accurate and precise estimation of the quantile in as few system-runs as possible. For this analysis, *accuracy* and *precision* are defined as in Table 3. 1.

Table 3. 1: Accuracy and Precision Definitions

Accuracy	Distance from estimated quantile value to true quantile
Precision	The spread or range of possible estimated quantile values

Simply estimating the quantile of a distribution would not qualify as a hypothesis test, under the definition given in Section 2, without some comparison to a hypothesized value using a critical value, *p*-value, or confidence interval. However, as subsequent sections will show, estimating a quantile is often one of the first steps to completing a hypothesis test. Improving the quality (i.e. the precision and accuracy) of the quantile estimation may help reduce the probability of Type-I and Type-II errors (α and β in Section 2.2).

This chapter begins with an overview of quantile estimation techniques, before describing the various DOE methods analyzed here (Section 3.1). From there, systems representative of those analyzed in a nuclear reactor safety analysis will be used to compare the accuracy and precision of the quantile estimations found by using these DOE methods (Section 3.2). Using these results, recommendations are made about the applicability of the DOE methods for use in safety analyses (Section 3.3).

3.1 Techniques

3.1.1. Quantile Estimation

This section will review the process for estimating a quantile from empirical data.

To state the problem more formally, suppose there is a system that has as its input a random vector *X*, and output *Y* with cumulative distribution function (CDF) *F*. By inverting *F*, the *p*-quantile ξ_p can be found, as seen in Eq. 32,

$$\xi_p = F^{-1}(p) \qquad \text{Eq. 32}$$

meaning, if $p = 0.95$, $\xi_{0.95}$ is the true 0.95-quantile. If the goal was to find ξ_p using CMC estimation and simulation, independent and identically distributed (i.i.d.) samples $Y_1, Y_2, \ldots, Y_N$ from distribution *F* would be generated. From there, the empirical cumulative distribution function $\tilde{F}_n$ can be computed, as in Eq. 33,

$$\tilde{F}_n(y) = \frac{1}{n}\sum_{i=1}^{N} I(Y_i \leq y) \qquad \text{Eq. 33}$$

where $I(A)$ is the indicator function of a set A, which assumes value 1 on A and 0 on the complement A^C. The *p*-quantile estimator is then computed by inverting $\tilde{F}_n$, $\tilde{\xi}_{p,n} = \tilde{F}_n^{-1}(p)$.

$\tilde{\xi}_{p,n}$ is usually calculated using order statistics, where the outputs $Y_1, Y_2, \ldots, Y_n$ would be sorted in ascending order $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$, where $Y_{(i)}$ is the *i*-th smallest of the samples. At this point, there are several different methods that can be used to find the quantile estimator $\tilde{\xi}_{p,n}$. The main differences between these methods relates to the placement of the output results on the probability scale. For example, if 50 computer code runs are conducted and the output results are ordered from smallest to largest in order to create an empirical cumulative distribution function, should the lowest result be placed at $p_C = 0.0$, at $p_c = 1/50 = 0.02$, or at the midpoint of that range at $p_c = 0.1$ (where p_c is the cumulative probability)? Since the analysis here is concerned with comparing

different sampling methods, the choice of quantile interpolation method is less important as long as all sampling techniques use the same method.

For this analysis, the highest ordered result will be given $p_C = 1.0$, which will give the lowest results $p_C = 1/n$. So to estimate the quantile, the output results will be ordered as before, $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$, where $Y_{(i)}$ is the *i*th smallest of the samples. Then $\tilde{\xi}_{p,n} = Y_{(\lceil np \rceil)}$ where $\lceil \cdot \rceil$ is the round-up function. Using the round-up function circumvents the need for interpolation.

For example, if 32 runs are conducted, 32 results will be obtained and will be ordered. Each result has equal probability of $1/32 = 0.03125$. If the lowest result is placed at $p_c = 0.03125$ (rather than at $p_c = 0.0$); this means the 30th ordered result will have $p_c = 0.9375$, and the 31st ordered result will have $p_c = 0.96875$. So the 0.95-quantile will be considered the 31st ordered result. While these techniques provide an estimation of the quantile, they do not provide any indication of the level of confidence of the value, which will be detailed in Section 4.

It should be noted here, that this method to estimate quantile values has been proven for use with the techniques in Section 3.1.2, but not for those in Section 3.1.3. However, it is used for all methods in order for a consistent comparison between the methods to be made.

3.1.2. Traditional Methods - Crude Monte Carlo and Latin Hypercube Sampling

The simplest technique to estimate a quantile of the output distribution of a system is to use crude Monte-Carlo (CMC) sampling for the selection of input values.

This is the most common method used in computer experiments that contain continuous input variable distributions. However, CMC is not always the most efficient method for quantile estimation. As mentioned in Section 2, the output of a CMC analysis can vary greatly, especially at low run levels.

Latin Hypercube Sampling (LHS) is probably the most commonly used variance variance reduction technique (VRT). LHS is popular because the stratification of the input values relies only on probability, and is not based on knowledge about the system. This makes the technique easier to apply generally than other VRTs, such as control variates. It was developed by McKay [59] and induces correlation among the simulated outputs in order to increase statistical efficiency, under certain conditions. LHS is a subset of stratified sampling, and what differentiates LHS from other stratified techniques is the way the strata are chosen. With LHS, the input parameter distributions are split into a number of equal probability intervals. Figure 3. 2 and Figure 3. 3 demonstrate how this would be done on a normal distribution, if one wanted to split the input variable space into five intervals. After each input distribution is divided into equal probability strata, a value is selected randomly from the distribution inside each stratum. This method helps ensure that the input uncertainty space is covered adequately in fewer runs than CMC, but it also retains the properties of each sample having equal probability, and a stochastic element to the selection of the input values. LHS is preferred over CMC sampling for estimating a high quantile, such as the 0.90- or 0.95-quantile, when computational costs are an issue [24].

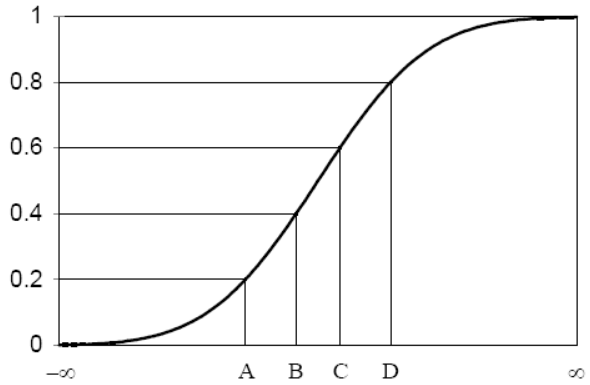

Figure 3. 2: Dividing a Normal CDF into Five Equal Probability Bins [60]

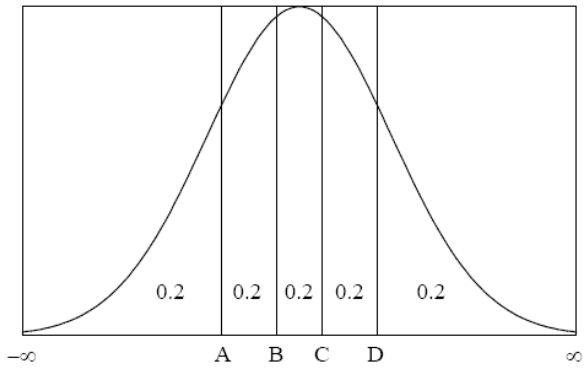

Figure 3. 3: Dividing a Normal PDF into Five Equal Probability Bins [60]

3.1.3. Orthogonal Arrays and Orthogonal Latin Hypercubes

In physical experiment design, uncertain or tunable inputs are often assigned several different levels which represent different regions of the input distribution space. This is done because using continuous input distributions with randomly sampled values in the physical world could be extremely difficult, if not impossible, due to the limitations

when controlling physical factors. For the analyst to characterize the output distribution space, they run some combination of these levels. As mentioned in Section 2, they could conduct full or fractional factorials depending on the goal of the analysis.

The most thorough way to test a system using fixed levels would be to try all possible combinations of levels in a full factorial experiment. While full factorial experiment design would provide the most data about the system, such as the input interactions which would be needed to form a response surface, the amount of runs necessary to conduct such an experiment is often times unrealistic. In order to get similar data in less runs, fractional factorials are used that are designed to return the desired characteristic of the output distribution. As stated in Section 2.3.1.2, fractional factorials are characterized by their *resolution* or *strength*, where a higher resolution implies that higher-order interaction terms can be found.

Resolution III fractional factorials are most commonly used as screening designs. Since only main effects can be found, with higher order terms considered negligible, these designs are used as the first step of an analysis where less important inputs are screened out from a large number of inputs. Then, a more in-depth analysis using a higher order fractional factorial is conducted on the remaining inputs.

While there are many ways to design fractional factorials, the most common method is to use an orthogonal array (OA). Orthogonal vectors are vectors of the same length which have an inner product, the sum of the products of their corresponding elements, of zero. OAs consist of a group of vectors which are all orthogonal to one another. Using this orthogonality in a fractional factorial allows certain input interactions

to be screened out or neutralized. This property is what makes OAs popular as a fractional factorial DOE. While OAs are frequently used for the DOE of physical experiments, they are less often found in random sampling designs.

In an attempt to use OAs in the realm of computer experiments, where input values can be sampled at random and not given prescribed levels, LHS designs based on OAs were developed [61]. These LHS designs use OAs to roughly determine which interval of each distribution should be selected for each run. This is in contrast to the usual LHS design, where the selection of which intervals to use on each run is random. The hope was that the space covering properties of the orthogonal array would help optimize the LHS design's own space covering attributes. Recent research [62] has shown that it is in fact possible to create orthogonal LHS (OLHC) designs. These are designs where the final run order is actually orthogonal, and not just loosely based on an orthogonal design. OLHCs are essentially the same as regular LHSs, but the run order of the intervals of each distribution is determined to satisfy the orthogonal properties between the variables. Here, the terminology OLHC is used because the orthogonality simply refers to the design of the run order, but not whether the values chosen to represent the strata are sampled or static (as explained in Section 3.2.1). These designs can be used with constant level values and without sampling, as in physical experiments; however, they are only Resolution II, meaning not even main effects can be resolved from the analysis since they are confounded with the main effects of other inputs.

In this work, OA designs are being tested outside of their common use, which is to determine the coefficients of a response surface, or to calculate the importance ranking

of input parameters. Instead, they are being used to estimate quantiles of the output distribution. The motivation behind this application is to determine if the orthogonal properties of these designs carry any advantage when estimating quantiles when compared to traditional techniques.

3.2 Experiments

The following experiments were designed in order to test these methods for use in nuclear safety analyses. Experiments were devised that would mimic common safety analysis situations. This included starting with a simple nonlinear equation, moving to a response surface surrogate for RELAP5, and conducting a risk assessment event tree analysis. This section begins with a description of how the methods in Section 3.1 were applied in these experiments.

3.2.1. Methods Analyzed

For this analysis, both Resolution III OAs and Resolution II OLHC were applied using two techniques. First, prescribed values were used for the levels specified in the design of experiment (DOE). This means a constant value was chosen to represent that region of the distribution. It was selected by using the midpoint of the interval in probability space. Figure 3. 4 illustrates how this representative value was found. The midpoint of the interval with bounds 0.0 and 0.2 is used to find the corresponding *x*-axis value of the CDF. This method is sometimes used in LHS designs in order to create a more uniform sampling design [18].

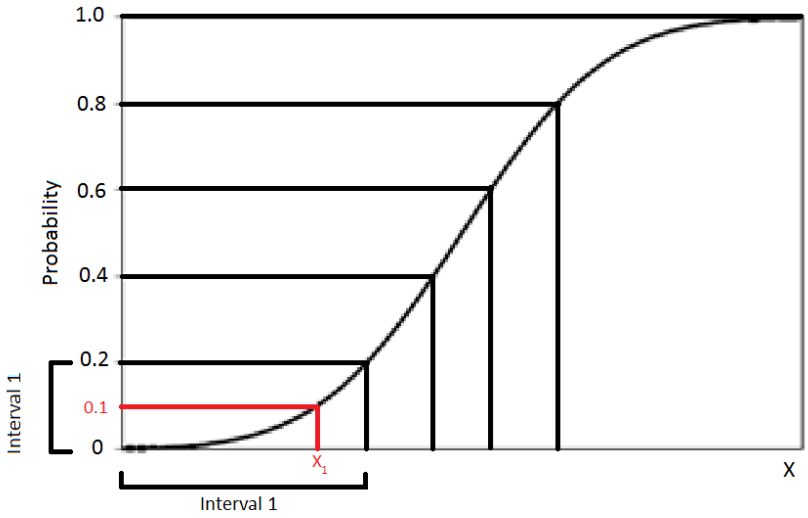

Figure 3. 4: Input CDF Split into Five Intervals with Chosen Midpoint

Second, these intervals were treated similarly to the intervals of a traditional LHS design, where values are selected randomly from the interval. Here, the random selection is made from the probability space in order to account for the distribution shape, just like a normal LHS design, as seen in Figure 3. 5. In this figure, a value is chosen randomly from the interval between 0.0 and 0.2. Once the value of 0.174 is selected, it is used to find the x-axis value of the CDF, x_1, that corresponds to that location. Both orthogonal methods, using both techniques to determine the level values, were compared to CMC sampling and traditional LHS. These methods were compared at three different run levels: 16, 32, and 64 runs. These levels were chosen because OAs and OLHCs designs can only be constructed at certain run levels, and these presented the most available options of OA and OLHC designs.

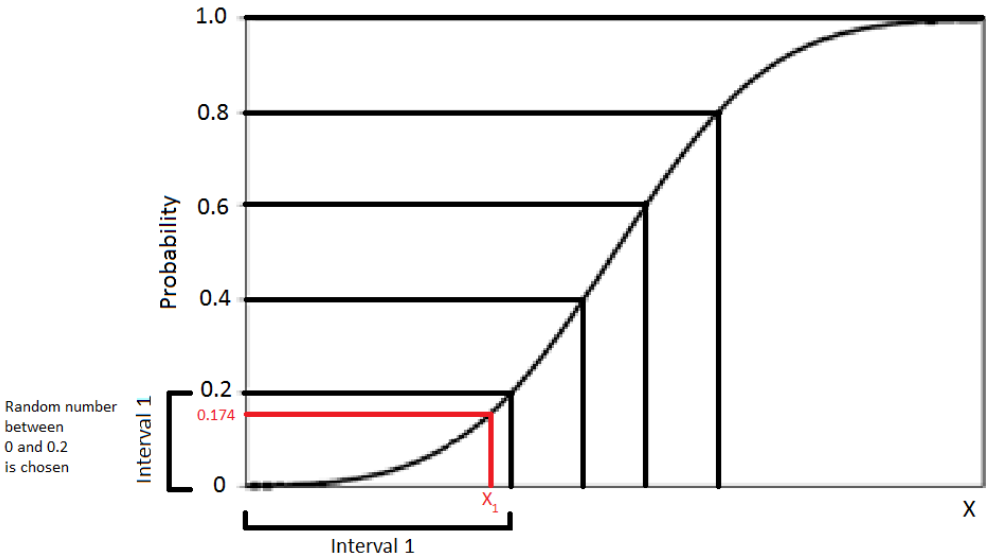

Figure 3. 5: Input CDF Split into Five Intervals with Randomly Selected Value

3.2.2. Nonlinear Equation

The first test conducted used an equation found in previous literature on sampling schemes [63]. This equation is simply a statistical test; it has no physical meaning. It is used due to its complexity and since it is difficult to model accurately with a second-order response surface. This equation was chosen as a first step to see how the proposed methods would perform with a nonlinear equation, since many such equations are found in large severe accident computer codes. It is defined in Eq. 34,

$$Y = 5 + (2 + 9X_1)^{0.7} \ln(2 + 2X_3 + X_3^2) + (1 + 2X_3)^{1.2} e^{X_4^2} + X_2^2 \qquad \text{Eq. 34}$$

where the uncertain input parameters $0 \leq X_1, X_2, X_3, X_4 \leq 2$, and Y is considered the output of interest, which would be compared to a prescribed safety goal.

3.2.2.1. Normal Inputs

For this test, all four inputs to Eq. 34 were assumed to be truncated normal distributions with mean 1.0 and standard deviation 0.22639. These distribution parameters were chosen in order for 99.99% of the non-truncated normal distribution to fall between 0 and 2. First, a CMC experiment with 10^8 runs was conducted in order to estimate the true 0.95-quantile of the system. The result was a 0.95-quantile of 40.6457, which would be considered the "true" 0.95-quantile. Figure 3. 6 shows the output distribution of the system for a 10^5-run CMC trial, which is shown simply to give the reader an idea of the range of possible outputs.

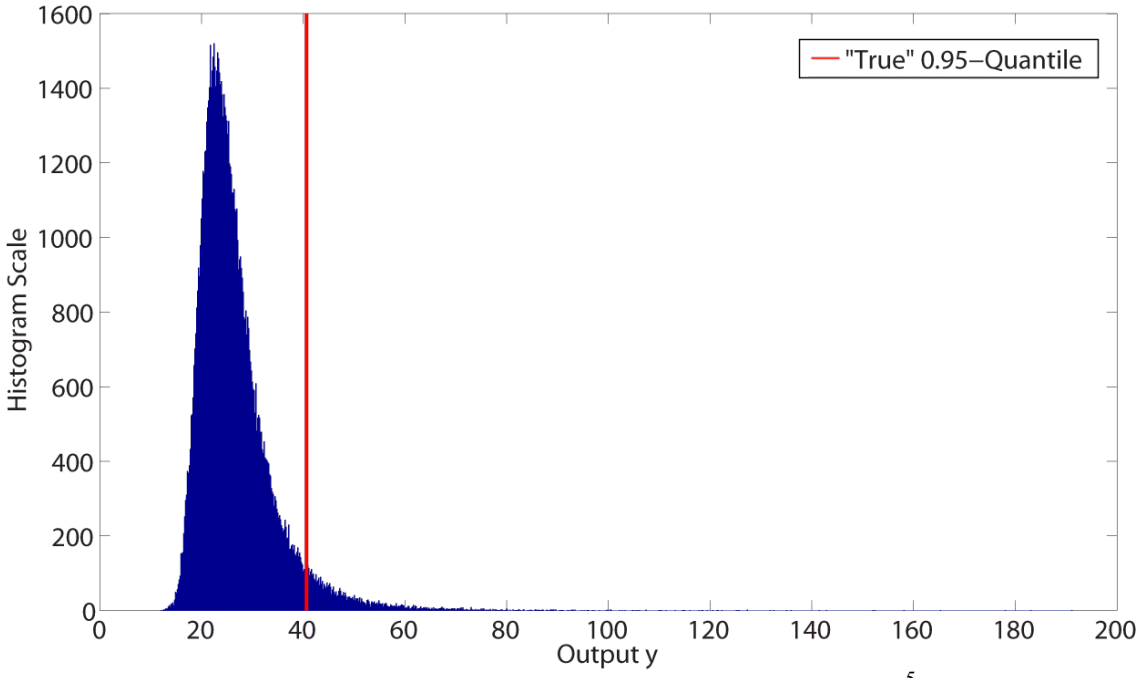

Figure 3. 6: Nonlinear Equation with Normal Input Histogram 10^5 CMC Runs

In order to test the experiment design methods listed in Section 3.1, each design was used for up to 10^5 trials, where a trial is a set of runs (a run is defined as in Section 2.3, where it is the creation of a sample from the system). So a single trial may consist of a set of 16, 32, or 64 runs, and result in a single estimated quantile value. Doing multiple trials gives information about the spread of the estimated quantile values using that method. Not all methods could be tested for 10^5 trials though. For the lower run levels, such as 16 and 32, when using OAs or OLHCs with fixed midpoints, the number of possible combinations of experiment design may be limited. For example, for the Resolution III OA with 16 runs, an OA known as L_{16}, was used. This is a four level design and can handle up to five inputs, as seen in Figure 3. 7.

Run	**Input**				
	1	2	3	4	5
1	1	1	1	1	1
2	1	2	2	2	2
3	1	3	3	3	3
4	1	4	4	4	4
5	2	1	2	3	4
6	2	2	1	4	3
7	2	3	4	1	2
8	2	4	3	2	1
9	3	1	3	4	2
10	3	2	4	3	1
11	3	3	1	2	4
12	3	4	2	1	3
13	4	1	4	2	3
14	4	2	3	1	4
15	4	3	2	4	1
16	4	4	1	3	2

Figure 3. 7: A L_{16}, Resolution II OA with Input Levels for Each Run

Since fixed midpoints are used and the system is deterministic, if the same design is repeated, it will give the same result. This experiment only had four inputs however, which meant 120 permutations of the experiment design were possible (the number of permutations is found using the formula $n!/(n-r)!$, where n is the number of things to chose from, and r is the number of things chosen). For the L_{32}, Resolution III OA used for 32 runs, there were 3024 possible permutations. The same situation occurred when using fixed midpoints with OLHC designs. Figure 3. 8 shows the OLHC used for 16 runs.

Run	Input											
	1	2	3	4	5	6	7	8	9	10	11	12
1	-15	5	9	-3	7	11	-11	7	-9	3	-15	5
2	-13	1	1	13	-7	-11	11	-7	-1	-13	-13	1
3	-11	7	-7	-11	13	-1	-1	-13	9	-3	15	-5
4	-9	3	-15	5	-13	1	1	13	1	13	13	-1
5	-7	-11	11	-7	11	-7	7	11	5	15	-3	-9
6	-5	-15	3	9	-11	7	-7	-11	13	-1	-1	-13
7	-3	-9	-5	-15	1	13	13	-1	-5	-15	3	9
8	-1	-13	-13	1	-1	-13	-13	1	-13	1	1	13
9	1	13	13	-1	-9	3	-15	5	11	-7	7	11
10	3	9	5	15	9	-3	15	-5	3	9	5	15
11	5	15	-3	-9	-3	-9	-5	-15	-11	7	-7	-11
12	7	11	-11	7	3	9	5	15	-3	-9	-5	-15
13	9	-3	15	-5	-5	-15	3	9	-7	-11	11	-7
14	11	-7	7	11	5	15	-3	-9	-15	5	9	-3
15	13	-1	-1	-13	-15	5	9	-3	7	11	-11	7
16	15	-5	-9	3	15	-5	-9	3	15	-5	-9	3

Figure 3. 8: A 16 Run OLHC Design (Resolution II OA) with 12 Inputs*

*to ensure orthogonality, the interval numbering scheme is as follows:

[-15, -13, -11, -9, -7, -5, -3, -1, 1, 3, 5, 7, 9, 11, 13, 15]

Since this design could handle up to twelve inputs, and only four are used here, this implied 11,880 possible experiment designs. Obviously, this was not a constraint when random sampling from the intervals was used since even the same experiment design would result in different values being chosen from each interval.

Table 3. 2 presents a table of the OA and OLHC designs used for this analysis. These OAs can be found in Appendix A.

Table 3. 2: List of OAs Used for Each Run Level*

Number of Runs	OA Resolution III	OLHC Resolution II
16	$L_{16'}$	OLHC.16
32	$L_{32'}$	OLHC.32
64	OA.64	OLHC.64

*The OAs can be found in Appendix A

The results of the analysis can be seen in Table 3. 3. Three values are presented for each case. The first value is a sample mean of the 0.95-quantile estimation of the 10^5 trials conducted, followed by the sample standard deviation (denoted S.D.) of each trial's 0.95-quantile estimation. The standard deviation should be viewed with caution, since as stated above, the method with fixed midpoints may have involved less than 10^5 trials, unlike the sampled methods. Finally, a percent error is given. This is the most useful metric to compare the different methods. As seen in Eq. 35, this is the mean of the percent difference of the quantile estimation of each trial and the "true" 0.95-quantile.

$$\text{Percent Difference} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\tilde{\xi}_i - \xi\right|}{\xi} x100 \qquad \text{Eq. 35}$$

where,

ξ = "true" 0.95-quantile = 40.6457
$\tilde{\xi}_i$ = Estimated $0.95-$ quantile of i^{th} trial
n = number of trials

Table 3. 3: Results for 10^5 Trials for Nonlinear Equation with Normal Inputs

Number of Runs[(a)]	Metric	CMC	LHS	OA - Res III		OLHC - Res II	
				Midpoint[(b)]	Sampling[(b)]	Midpoint[(b)]	Sampling[(b)]
16	Mean of $\tilde{\xi}$	45.13	47.35	38.02	46.12	43.97	47.43
	S.D. of $\tilde{\xi}$	13.08	12.18	1.22	12.67	5.06	12.17
	% Difference[(c)]	**21.38**	**20.74**	**6.45**	**20.17**	**11.30**	**20.20**
32	Mean of $\tilde{\xi}$	41.58	41.37	37.05	41.98	40.87	41.29
	S.D. of $\tilde{\xi}$	6.83	3.82	0.93	6.20	3.31	3.38
	% Difference[(c)]	**12.31**	**7.44**	**8.85**	**10.94**	**6.40**	**6.49**
64	Mean of $\tilde{\xi}$	40.18	40.04	36.60	40.36	40.16	40.13
	S.D. of $\tilde{\xi}$	4.18	2.43	0.65	3.65	1.86	1.93
	% Difference[(c)]	**8.15**	**4.98**	**9.96**	**7.03**	**3.90**	**4.00**

[(a)]Number of runs in a single trial, [(b)]See Section 3.2.1, [(c)]See Eq. 35

As the results show, LHS and OLHC, using sampling and fixed midpoints, outperformed CMC sampling and Resolution III OAs in relation to the percent difference metric. For example, at the 32-run level, CMC had a percent difference of 12.31 and the Resolution III OA using sampling had a percent difference of 10.94, while LHS had a value of 7.44, and OLHC using midpoints and sampling had values of 6.40 and 6.49, respectively. Resolution III OAs using sampling also outperformed CMC sampling using the percent difference value by a smaller margin. When using fixed midpoints, Resolution III OAs were the only method that underestimated the 0.95-quantile at low run levels, resulting in a mean of 38.02 compared to the true 0.95-quantile value of 40.646, and the accuracy actually got worse as the number of runs was increased, as the

percent difference metric increased from 6.45 to 9.96. The precision of the estimated quantile values increased (i.e. the S.D. was lower) when using LHS when compared to CMC, which is expected since LHS is a VRT. However, OLHC using sampling had a greater reduction in variance than normal LHS, where at the 64-run level, LHS had a S.D. of 2.43, but OLHC using sampling had a S.D. of 1.93.

3.2.2.2. Non-Normal Inputs

Next, the experiment in Section 3.2.2.1 was repeated, but the normally distributed inputs were replaced with a variety of distributions, as seen in Eq. 36.

$$\begin{aligned} &x_1 - exponential(0.173) \\ &x_2 - N(1.0, 0.22639^2) \\ &x_3 - lognormal(0.0, 0.162) \\ &x_4 - uniform(0,2) \end{aligned} \qquad \text{Eq. 36}$$

This was done in order to remove any possible influence from the use of normal distributions. Once again, the exponential and normal distributions were truncated at 0 and 2, and the parameters are chosen in order for 99.99% of the non-truncated distribution to fall within that interval. Since the lognormal distribution has no values which fall below 0, 99.995% of the non-truncated distribution falls below 2, and 100% of the uniform distribution is between those bounds. A 10^8-run CMC trial resulted in a 0.95-quantile value of 148.650. Figure 3. 9 shows the output distribution for a 10^5-run CMC trial. Once again, this is done to give the reader an idea of the output distribution shape. Compared to the previous example in Section 3.2.2.1, this output has a longer tail at high values, which results in a much higher value for the 0.95-quantile.

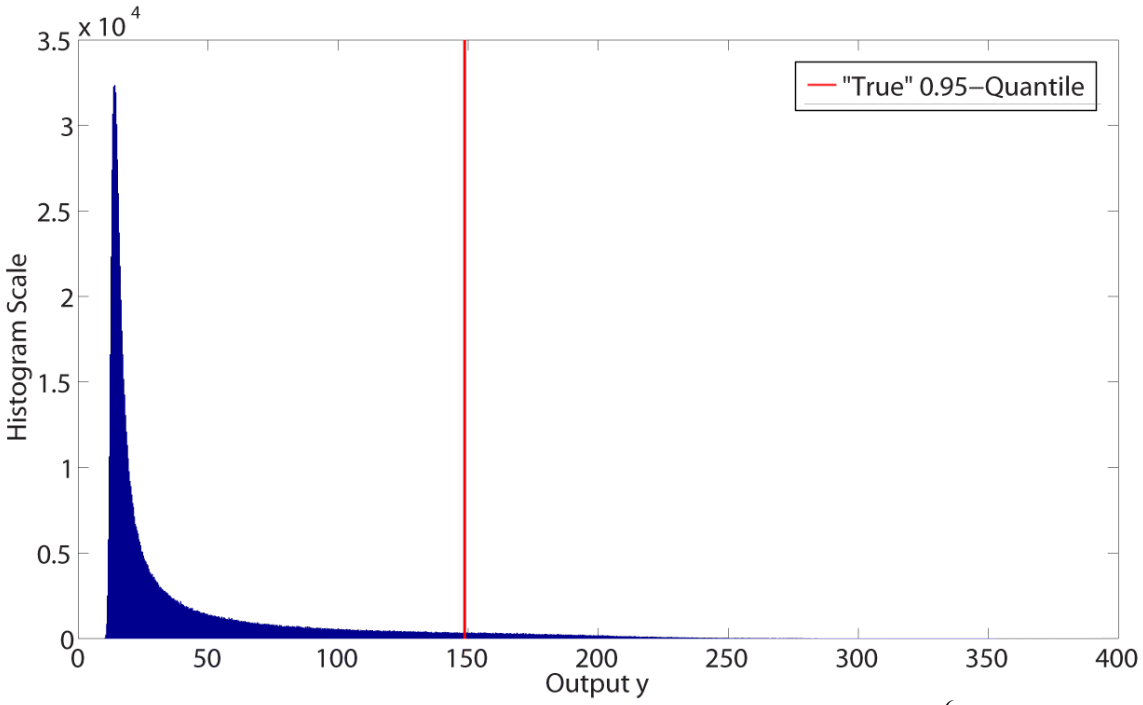

Figure 3. 9: Nonlinear Equation with Non-normal Inputs Histogram 10^6 CMC Runs

Table 3. 4 shows the results of this experiment. As with the previous experiment, OLHC and LHS had, on average, the most accurate result, in regards to the percent difference, as can be seen at the 32-run level, where LHS had a percent difference of 9.71, OLHC using sampling had a percent difference of 8.85, while CMC had a percent difference of 20.17. Resolution III OA's with sampling once again fared better than CMC sampling in regards to the percent difference, but when using fixed midpoints, it again consistently underestimated the quantile, with a mean of 103.58 at the 16-run level, compared to a true value of 148.65. Also, the results when using OLHC for both midpoints and sampling were, once again, very similar.

Table 3. 4: Results of 10^5 Trials for Nonlinear Equation with Non-normal Inputs

Number of Runs[(a)]	Metric	CMC	LHS	OA - Res III		OLHC - Res II	
				Midpoint[(b)]	Sampling[(b)]	Midpoint[(b)]	Sampling[(b)]
16	Mean of $\tilde{\xi}$	153.70	173.79	103.58	159.56	173.02	175.54
	S.D. of $\tilde{\xi}$	49.22	31.20	1.14	43.80	19.39	30.89
	% Difference[(c)]	**27.47**	**20.72**	**30.32**	**24.94**	**16.95**	**21.15**
32	Mean of $\tilde{\xi}$	145.15	153.17	102.17	147.29	151.74	152.23
	S.D. of $\tilde{\xi}$	36.71	17.83	2.13	32.64	14.43	16.31
	% Difference[(c)]	**20.17**	**9.71**	**31.27**	**18.02**	**7.84**	**8.85**
64	Mean of $\tilde{\xi}$	140.80	144.38	102.34	142.44	145.72	145.03
	S.D. of $\tilde{\xi}$	26.97	11.96	0.55	23.69	9.45	10.08
	% Difference[(c)]	**15.18**	**6.94**	**31.15**	**13.37**	**5.29**	**5.84**

[(a)]Number of runs in a single trial, [(b)]See Section 3.2.1, [(c)]See Eq. 35

3.2.3. LOCA Response Surface

The next system analyzed was a second-order response surface, developed by French [64], which models the peak clad temperature of a nuclear power reactor during a LOCA. This equation was chosen because it is designed to act as a surrogate for the RELAP5 [65] plant deck from which it was created. While it is a relatively simple equation, it is a step towards a realistic nuclear safety analysis. This response surface is shown in Eq. 37,

$$Y = a + b_1X_1 + \cdots + b_{11}X_{11} + c_1X_1^2 + \cdots + c_{11}X_{11}^2 + d_1X_1X_2 + \cdots + d_{55}X_{11}X_{10}$$
$$X_1, \ldots, X_{11} = N(0.5, 0.1^2) \qquad \text{Eq. 37}$$

where a, b_i, c_i, and d_i are constant coefficients and the eleven inputs $X_1, \ldots, X_{11}$ are independent normal random variables which are truncated at 0 and 1. The inputs are certain normalized reactor properties. The output Y is peak cladding temperature in degrees Fahrenheit. The result of a 10^8-run CMC experiment yielded a "true" 0.95-

quantile of 1683.65°F. Figure 3. 10 shows the distribution of a 10^5 CMC trial. In this case, the higher end of the distribution is fairly compact.

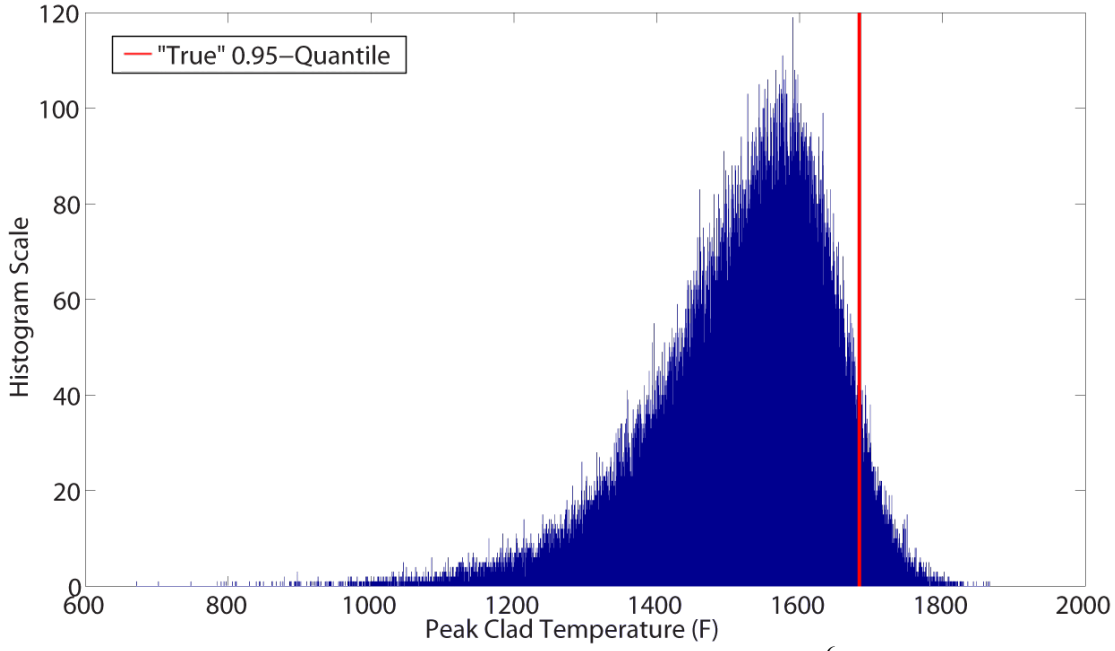

Figure 3. 10: RELAP Response Surface Histogram 10^6 CMC Runs

Since this system had eleven inputs, some OAs had to be changed to accommodate more inputs. Table 3. 5 shows the OAs and OLHCs used for this analysis.

Table 3. 5: List of OAs Used for Each Run Level*

Number of Runs	OA Resolution III	OLHC Resolution II
16	L_{16}	OLHC.16
32	L_{32}	OLHC.32
64	OA.64	OLHC.64

*The OAs can be found in Appendix A

The results of this experiment can be seen in Table 3. 6. As in previous experiments, LHS and OLHC outperform the other methods in regards to the percent difference, as can be seen at the 32-run level, for example, where LHS has a perecent difference on 1.03, OLHC using sampling has a value of 0.95, while CMC is 1.25. The OLHC designs are again very similar whether using midpoints or sampling. Also, unlike the two previous examples, the Resolution III OA using midpoints actually increases in accuracy as the number of runs grows, from a percent difference of 2.12 to 0.82, but it is still the poorest performer of all the methods.

Table 3. 6: Results of 10^5 Trials for LOCA Response Surface

Number of Runs[(a)]	Metric	CMC	LHS	OA - Res III		OLHC - Res II	
				Midpoint[(b)]	Sampling[(b)]	Midpoint[(b)]	Sampling[(b)]
16	Mean of $\tilde{\xi}$	1691.27	1697.80	1647.90	1692.18	1699.64	1701.60
	S.D. of $\tilde{\xi}$	39.08	34.20	10.11	36.28	29.38	31.92
	% Difference[(c)]	**1.85**	**1.67**	**2.12**	**1.71**	**1.50**	**1.62**
32	Mean of $\tilde{\xi}$	1683.25	1685.92	1648.37	1683.88	1687.83	1688.10
	S.D. of $\tilde{\xi}$	26.56	21.89	6.74	24.11	19.81	20.15
	% Difference[(c)]	**1.25**	**1.03**	**2.10**	**1.13**	**0.93**	**0.95**
64	Mean of $\tilde{\xi}$	1679.36	1680.62	1670.96	1679.85	1682.04	1681.98
	S.D. of $\tilde{\xi}$	18.45	14.85	8.79	13.58	11.85	11.90
	% Difference[(c)]	**0.90**	**0.72**	**0.82**	**0.67**	**0.57**	**0.57**

[(a)]Number of runs in a single trial, [(b)]See Section 3.2.1, [(c)]See Eq. 35

3.2.4. PRA Event Tree

The next example was designed to represent a probabilistic risk assessment (PRA) for a nuclear power plant. The example contains three initiating events; a small, medium,

and large break LOCA with associated system failures resulting in core damage. Each of these initiating events has its own Level-I core damage event tree, as seen in Figure 3. 11, Figure 3. 12, and Figure 3. 13. These event trees are very similar to those used in PRAs for actual plants. The events within the trees are indentified with letters that represent the success/failure of safety features, such as those associated with the ECCS. The result of each scenario in these trees indicates whether core damage has occurred, and if so, which damage-state the core is in. Once again, only a symbolic number is used to represent each of the four core damage-states.

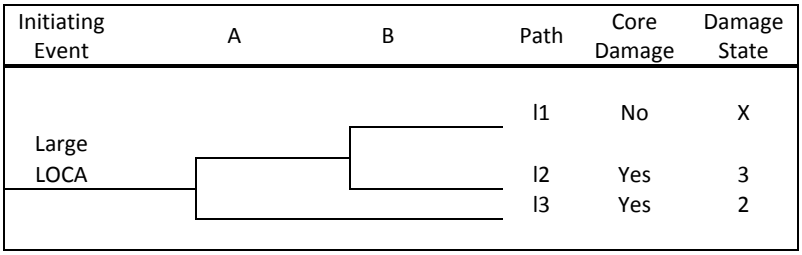

Figure 3. 11: Large Break LOCA Core Damage Event Tree

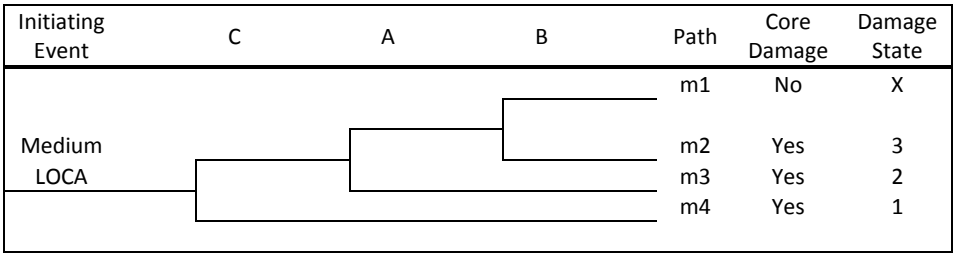

Figure 3. 12: Medium Break LOCA Core Damage Event Tree

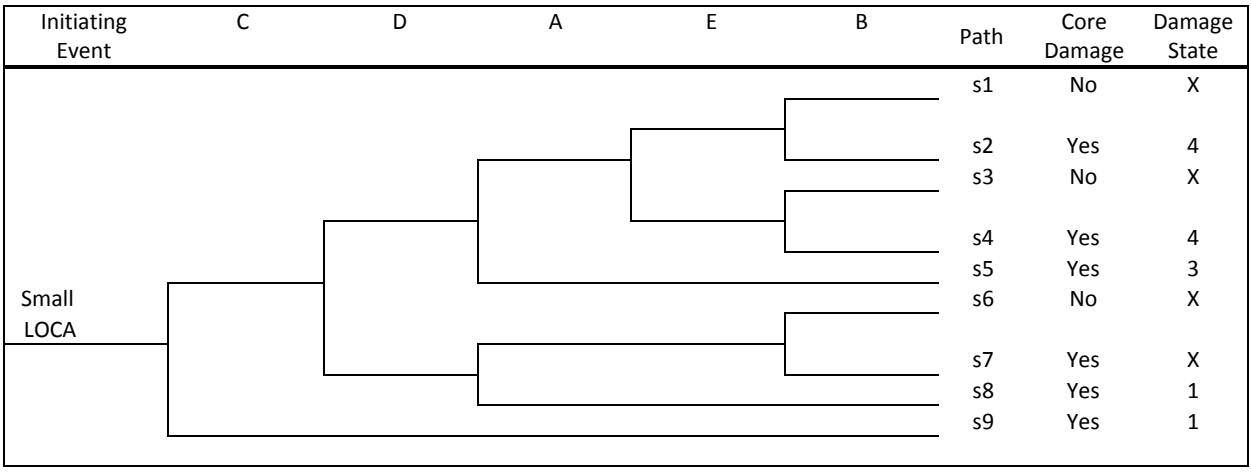

Figure 3. 13: Small Break LOCA Core Damage Event Tree

In a specific assigned core damage-state, the progression of core damage and the threats posed to the integrity of the containment are similar. For this example, there are two linked containment event trees that characterize the modes and timing of failure that are possible. The first containment event tree, shown in Figure 3. 14, characterizes early threats to containment failure. Early failure is particularly important because, for the associated magnitude of release and limited time for evacuation of the neighboring population, there is some potential for large doses to the population to result in radiation sickness sufficiently severe to result in fatality. If there is no early failure of containment (and no potential for offsite early fatalities), a late containment failure tree is analyzed, shown in Figure 3. 15. These containment event trees are based off examples in NUREG/CR-6595 [66]. In total, each run of the event tree series results in 841 unique end-state scenarios.

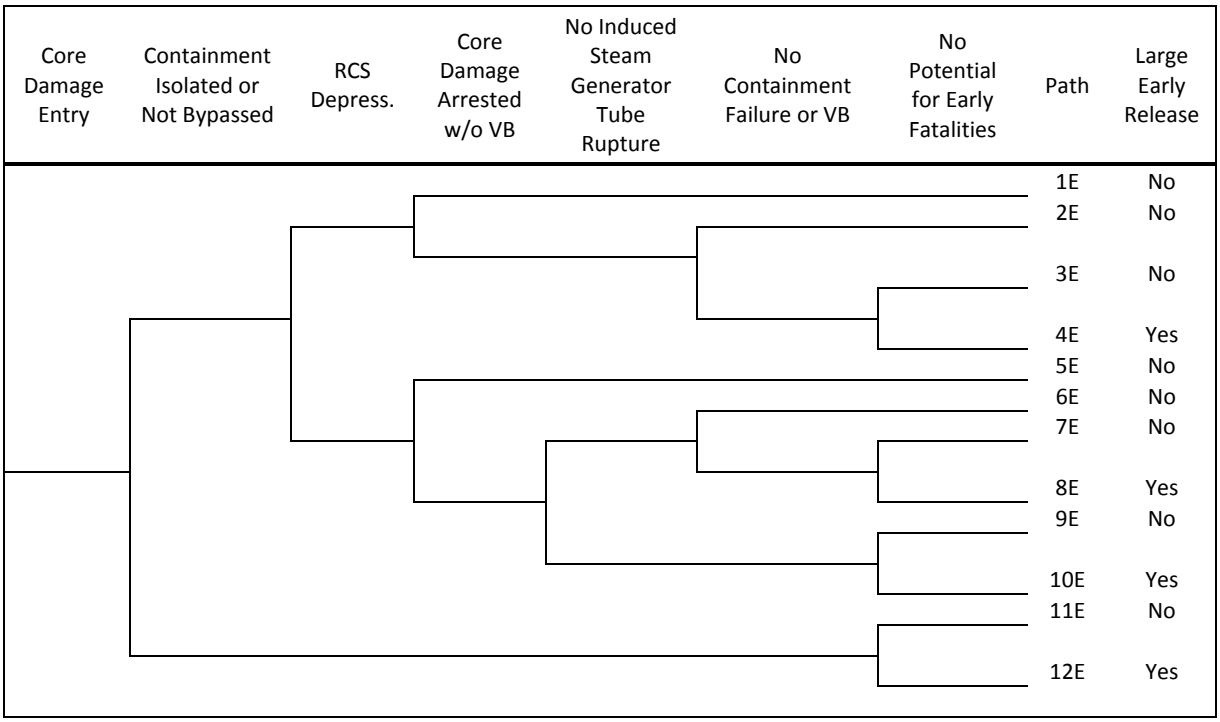

Figure 3. 14: Early Containment Failure Event Tree

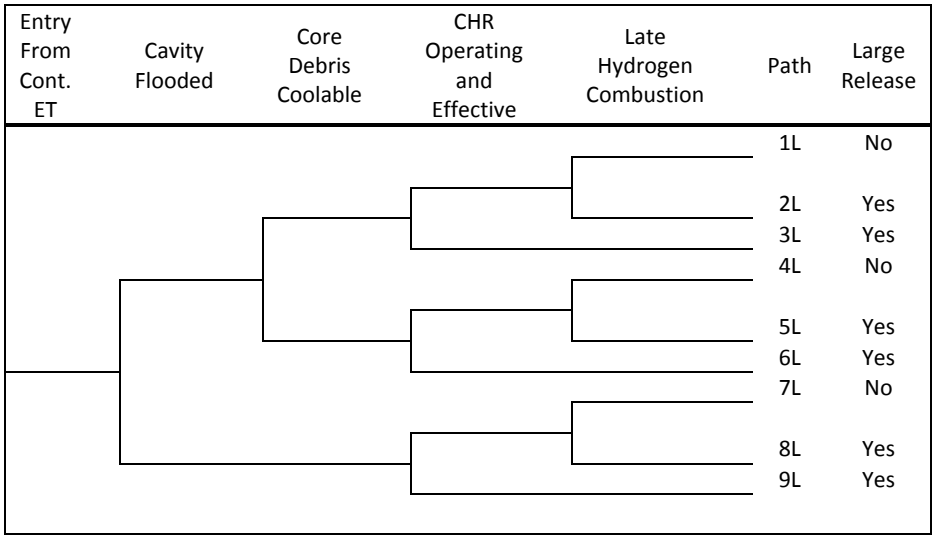

Figure 3. 15: Late Containment Failure Event Tree

The offsite dose is calculated using the Gaussian dispersion model described in NUREG-1465 [67], starting with the calculation of χ/Q in Eq. 38,

$$\frac{\chi}{Q} = \frac{e^{-\left(\frac{y^2}{2\sigma_y^2}\right)\left(\frac{h^2}{2\sigma_z^2}\right)}}{\pi\sigma_y\sigma_z u} \quad \text{Eq. 38}$$

where ,

$u = Average\ Wind\ Speed\ (m/s)$
$y = Distance\ from\ the\ plume\ axis\ in\ the\ transverse\ direction\ (m)$
$h = Release\ height\ (m)$
$\sigma_y = Pasquill - Gifford\ coefficient\ of\ Horizontal\ Diffusion\ (m)$
$\sigma_z = Pasquill - Gifford\ coefficient\ of\ Vertical\ Diffusion\ (m)$

From there, the inhalation and submersion doses can be found according to Eq. 39 and Eq. 40. The offsite dispersion and dose calculations are made at a distance of 1 km from the release point, and along the centerline of the plume (as is required for regulatory analyses [68], [69]). The dose conversion factors are found from [70]. The core radionuclide inventory was calculated using [71].

$$D_{Inh} = R \cdot Q \cdot \frac{\chi}{Q} \cdot DCF \quad \text{Eq. 39}$$

$$D_{Sub} = Q \cdot \frac{\chi}{Q} \cdot DCF \quad \text{Eq. 40}$$

where,

$D_{Inh} = Inhalation\ Dose\ (rem)$
$D_{Sub} = Submersion\ Dose\ (rem)$
$R = Inhalation\ Rate\ (m^3/s)$
$Q = Activity\ (Ci)$
$DCF = Dose\ Conversion\ Factor$

In total, there are 27 uncertain parameters. These are listed in Table 3. 7, along with their distribution shape.

Table 3. 7: PRA Event Tree Uncertainties

	Uncertainty	Distribution*
1	Small LOCA Initiating Event Frequency	Lognormal(0,1)
2	Medium LOCA Initiating Event Frequency	Lognormal(0,1)
3	Large LOCA Initiating Event Frequency	Lognormal(0,1)
4	C	Uniform(0,1)
5	D	Normal(0.05,0.01)
6	A	Normal(0.005,0.001)
7	E	Uniform(0,1)
8	B	Exponential(0.8)
9	Containment Isolated or Not Bypassed	Exponential(0.8)
10	RCS Depress.	Exponential(0.8)
11	Core Damage Arrested w/o VB	Exponential(0.8)
12	No Induced Steam Generator Tube Rupture	Exponential(0.8)
13	No Containment Failure or VB	Exponential(0.8)
14	No Potential for Early Fatalities	Exponential(0.8)
15	Cavity Flooded	Exponential(0.8)
16	Core Debris Coolable	Exponential(0.8)
17	CHR Operating and Effective	Exponential(0.8)
18	Late Hydrogen Combustion	Exponential(0.8)
19	Early Release Fraction Noble Gases	Beta(2,2)
20	Early Release Fraction Iodine	Beta(2,2)
21	Late Release Fraction Noble Gases	Uniform(0,1)
22	Late Release Fraction Iodine	Beta(2,60)
23	Wind Speed	Beta(2,2)
24	Release Height	Uniform(0,1)
25	Pasquill-Gifford Coefficient Horizontal Diffusion	Beta(2,2)
26	Pasquill-Gifford Coefficient Vertical Diffusion	Beta(0.8,5)
27	Containment Leak Rate	Beta(2,2)

*Many of the uncertainties are not the distribution of the actual parameter, but of a scaling factor or part of a larger formula

For this analysis, the figure of merit is the *mean risk*, which is defined in Eq. 41,

$$\bar{R} = \sum_{i=1}^{w} F_i \cdot D_i \qquad \text{Eq. 41}$$

where $\bar{R}$ is the mean risk, $w = 841$ or the total number of scenarios per run of the PRA, F_i is the frequency of the i-th scenario, and D_i is the offsite dose of the i-th scenario. So each run of the PRA will result in a single value for the mean risk. The 0.95-quantile mean risk of a 10^8-run CMC trial was 0.00300 rem/yr. Figure 3. 16 shows an empirical

CDF of the output for a 10^5-run CMC trial. As the figure shows, the output ranges over many orders of magnitude.

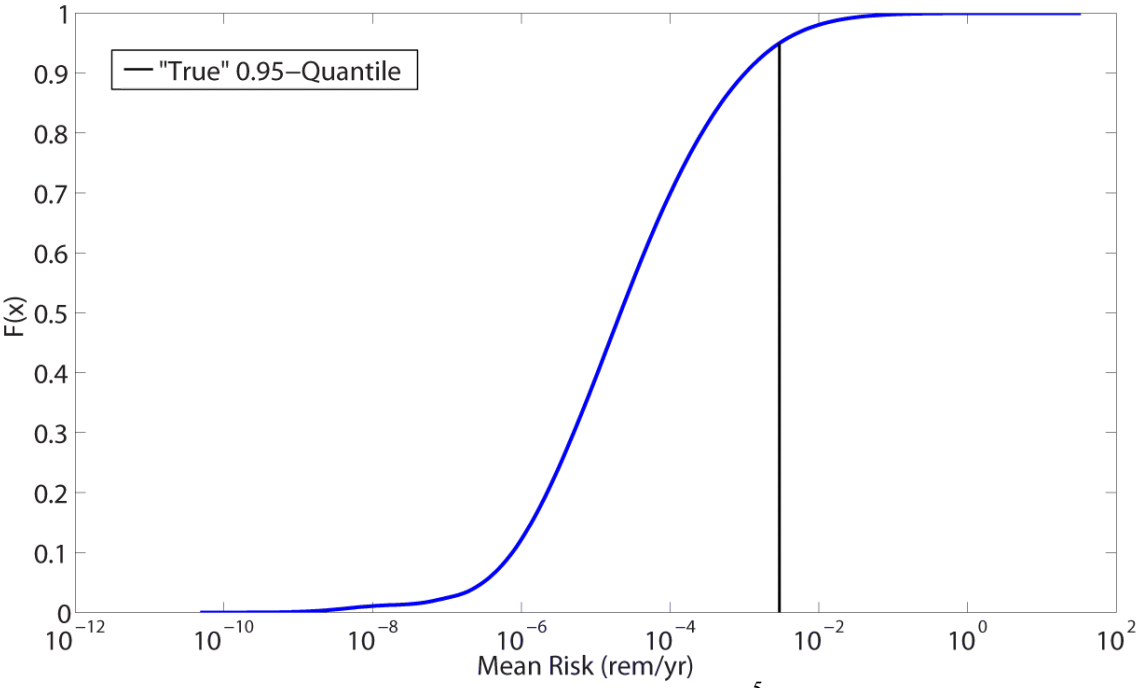

Figure 3. 16: Empirical CDF of Mean Risk 10^5-Run CMC Trial

Since this system contained 27 inputs, Resolution III OAs and Resolution II OLHCs were not available for the 16 run level. Table 3. 8 lists the arrays used for this experiment. Table 3. 9 contains the results of this experiment. Here, it is important to note that the percent difference may appear large, especially when compared to the previous examples. However, the output distribution of this analysis had a range of several orders of magnitude, so the percent difference will appear greater than the examples which had a smaller range of possible outcomes.

Table 3. 8: List of OAs Used for Each Run Level[a]

Number of Runs	OA Resolution III	OLHC Resolution II
16	X	X
32	L_{32}	OLHC.32
64	OA.64.32[b]	OLHC.64

[a]The OAs can be found in Appendix A, [b]OA.64.32 is a Resolution IV OA since no Resolution III OAs are available for that run level and number of inputs

Table 3. 9: Results of 10^5 Trials for PRA LOCA Analysis

Number of Runs[a]	Metric	CMC	LHS	OA - Res III		OLHC - Res II	
				Midpoint[b]	Sampling[b]	Midpoint[b]	Sampling[b]
16	Mean of $\tilde{\xi}$	0.00890	0.00987				
	S.D. of $\tilde{\xi}$	0.01778	0.01754				
	% Difference[c]	**217.94**	**235.30**				
32	Mean of $\tilde{\xi}$	0.00405	0.00403	0.00061	0.00409	0.00374	0.00401
	S.D. of $\tilde{\xi}$	0.00309	0.00216	0.00006	0.00300	0.00168	0.00193
	% Difference[c]	**63.83**	**49.86**	**79.66**	**60.26**	**39.70**	**46.16**
64	Mean of $\tilde{\xi}$	0.00290	0.00289	0.00060	0.00292	0.00282	0.00291
	S.D. of $\tilde{\xi}$	0.00129	0.00090	0.00004	0.00118	0.00068	0.00072
	% Difference[c]	**32.53**	**23.42**	**80.11**	**29.70**	**19.07**	**19.27**

[a]Number of runs in a single trial, [b]See Section 3.2.1, [c]See Eq. 35

As with the previous experiments, LHS and OLHC designs outperform the other methods in the percent difference metric, with LHS at 49.85, OLHC using sampling at 46.15, and CMC at 63.83, for the percent difference at the 32-run level. Once again, OLHCs using sampling and static midpoints provides a more accurate result on average than LHS. While Resolution III OAs outperform CMC sampling when using a sampling approach, the use of static midpoints results in the worst performance of all the methods,

grossly underestimating the mean risk, with a mean of 0.00061 at the 32-run level compared to a true value of 0.003 .

3.3 Discussion

The results in Section 3.2 indicate the OLHC designs, whether using static midpoints or interval sampling, are the most accurate and precise of the analyzed methods when determining the 0.95-quantile of the output distribution. The results also show that if it is necessary for an analysis to use set static values, such as midpoints, OLHCs are the preferred method over the use of a Resolution III OAs, if the quantile estimation is the only goal of the analysis. If the research is also focused on capturing input interactions, it may be necessary to go to a higher order OA. These results are not completely surprisingly since the creation of Resolution III OAs was not focused on quantile estimation, but other factors such as input screening. The poor performance of Resolution III OAs when using static midpoint is most likely a result of the low number of intervals. This means the midpoints of intervals will be far from the tails of the distribution and may not characterize certain areas of the output distribution.

The indicated ability for OLHCs to outperform LHS, even when using static midpoint values is an interesting outcome and more work should be done into possible uses of OLHCs. Certain setbacks do still exist with OLHCs though, as they can be difficult to create and may not exist for a system with a large numbers of inputs in combination with low run levels.

Chapter 4: Confidence Intervals for Quantiles

This section expands on the concept of quantile estimation, explained in Section 3, by introducing confidence intervals for the point estimate of the quantile. Confidence intervals for quantiles are currently used in nuclear reactor safety analyses as a method to demonstrate adherence to NRC safety limits. This section details these NRC criteria, and the evolution of methods used to show satisfaction of these safety requirements (Section 4.1). From there, a more detailed analysis of the current NRC-accepted sampling method is presented, along with a detailed derivation of a new VRT confidence interval method, that has recently been proven [72] (Section 4.2). Lastly, the new VRT method and the current NRC-accepted method are compared using systems designed to represent those encountered during nuclear reactor safety analyses (Section 4.3), and conclusions relating to the probability of achieving the correct conclusion during an analysis are presented (Section 4.4).

4.1. Background

4.1.1. Regulatory History

As mentioned in Section 2, the initial approach to the treatment of modeling uncertainties in regulatory analysis was to use non-mechanistic, conservative models. In the implementation of the Part 50 Appendix K of the Code of Federal Regulations [4], which describes a prescription for the conservative treatment of uncertainties in the

analysis of LOCAs, it became apparent that what was thought to be conservative might not be conservative in all cases, and that conservative regulatory models could be misleading with regard to the improvement of reactor safety. The transition to best-estimate plus uncertainty regulatory requirements began with an amendment to 10 CFR 50.46 [73] in 1988, which allowed for realistic modeling of LOCAs. While this rule-change signaled an advancement in regulatory safety analysis, the statistical requirements of the output result were vague, stating only that there should be a "high level of probability that the criteria would not be exceeded."

In 1989, the NRC issued RG 1.157 [74], which helped clarify the procedure for performing a best-estimate calculation relating to the design bases for essential safety systems. It set the standard for the handling of computational uncertainty for nuclear safety applications by stating that a 95% probability level is considered acceptable to the NRC staff for comparison of best-estimate predictions to safety limits. However, the ambiguity of the term "95% probability level" remained an issue for the analyst.

The most obvious solution to the "95% probability" requirement was to estimate the 0.95-quantile of the output distribution. One method to do this was to perform a large number of CMC random sampling runs and simply order and count the results until 95% of the runs fell below that threshold. The large number of runs required by CMC to obtain sufficient accuracy represented a major problem for safety analysts, due to minimal computing power and extended code run times. There was also the question of just how many runs would be necessary for an analyst to be able to claim that the estimate of the 0.95-quantile was sufficiently accurate.

Response-surface methods [55] were initially proposed as a way of reducing runs and increasing knowledge of the overall behavior of the parameters of interest. An advantage of this method is that it employs a fixed matrix of runs to be conducted to obtain the desired surface. This property not only gives the analyst a plan to provide to the regulator, but also produces a level of understanding about the impact of different input parameters. However, like the large-sample CMC case, run designs often needed to be very large to capture input interactions and nonlinearities, and the only way around this was to group input parameters based on the analyst's judgment [75]. In response to these considerations, methods were developed that required a smaller number of runs, but which could satisfy the regulatory guidelines.

Both Areva [76] and Westinghouse [75] developed approaches to the use of CMC using order statistics (CMC-OS) for their regulatory LOCA analyses. While the method of CMC-OS was first considered for use in the nuclear industry in the 1970's [77], it wasn't until the NRC published NUREG-1475 [78], a guide to applying statistics, in 1994 that the NRC provided a more comprehensive picture of its use for regulatory requirements [79]. Gesellschaft für Anlagen-und Reaktorsicherheit (GRS) helped bring CMC-OS to the thermal hydraulic and safety fields soon after that [80]. Major steps forward occurred in 2003 and 2004 with publications by Guba, Pál, and Makai [81], and Nutt and Wallis [82]. These works not only expanded on how CMC-OS could be used in safety analyses, but also proposed the use of CMC-OS in regards to the 95% probability reporting requirement. The solution provided by Nutt and Wallis [82] to this question was to report a 95% one sided confidence interval for the 0.95-quantile of the output

distribution. Based on the works of Wilks [83] and Wald [84], this method simulates the model using CMC to establish a confidence interval for a tolerance interval. This method was considered acceptable by the NRC in regards to the 95% probability requirement [75], and is discussed in detail in Section 4.2.1.

While the acceptance of the 95% confidence interval for the 0.95-quantile has been adopted by the NRC for satisfying design basis accident requirements, there are other safety applications for which less stringent requirements may be appropriate, such as for the analysis of beyond-design-basis events. For the analysis of these events, similar, but less stringent limits could be established, such as the use of a high value confidence interval for a lower quantile.

4.1.2. Confidence Intervals and Hypothesis Testing

This section explains, in detail, the meaning of a confidence interval, clarifies its use within a hypothesis test, and presents a comparison of confidence intervals and credible intervals. From there, a framework is developed to more rigorously present the NRC's probability requirement in terms of hypothesis testing, and the possible scenarios where errors in conclusion could occur are detailed.

4.1.2.1. Confidence Intervals

A confidence interval (CI) gives an estimated range of values which is likely to include an unknown population parameter, with the estimated range being calculated from a given set of sample data [85]. The confidence level determines how frequently the calculated interval will contain the parameter. Unlike a point estimate, which only gives a

single estimated value for a parameter, a CI gives a range in which that parameter is estimated to lie.

While the concept of a CI may seem straightforward, its meaning is constantly misinterpreted. An example will help explain this common mistake. Imagine the goal of an analysis is to estimate the location of the p-quantile ξ_p of a distribution. After n number of samples have been taken from the distribution (either through physical sampling or computer code simulations), an estimate of the quantile value is made $\tilde{\xi}_p$. This is the point estimate. In order to give more information about the possible location of the true quantile, a CI is calculated (the process to obtain a CI will be described in Section 4.2). The estimated quantile, along with the bounds of the CI are shown graphically in Figure 4. 1.

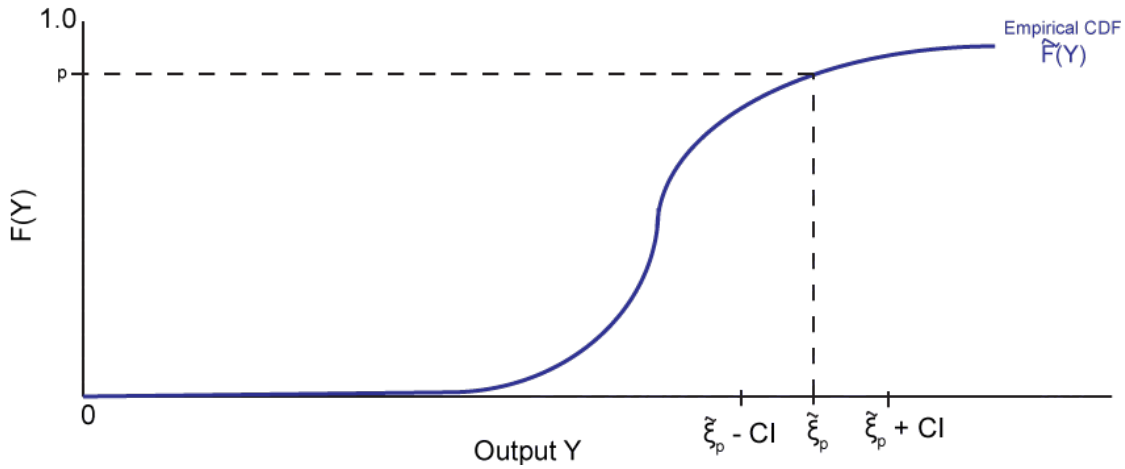

Figure 4. 1: Estimated Quantile with Confidence Interval

For this example, say a two-sided CI is constructed using a 90% confidence level. The mistake usually comes with the interpretation of this interval. Many times, CI results will

be reported with the comment that there is a 90% probability of the true parameter lying within the CI. This is a mistake. The CI gives no probabilistic information. The true parameter is fixed, not a random variable. Therefore, the probability of the true parameter lying within the CI is 0.0 or 1.0. It either does or it does not. The CI does not report uncertainty about the parameter, but uncertainty about the sampling method. If the analysis was repeated with new samples, and a new CI was constructed, 90% of the time this interval will include the true parameter. A CI only makes sense if more samples can be taken. If every possibility has been sampled, the CI becomes meaningless.

The meaning of a CI is often times confused with that of a *credible interval*. A credible interval (sometimes called a Bayesian CI), is a probabilistic statement about the location of the parameter. The reason that this statement can be made during a Bayesian analysis is because the distribution's parameter was assumed to be a random variable from the start. This is not the case in frequentist statistics, where it is assumed that there is only one true value of the parameter. With a credible interval, it is possible to make a statement such as "there is a 90% probability of the parameter lying within the credible interval."

The question could be asked here, then why use CIs instead of Bayesian credible intervals? As mentioned in Section 2.1, there was great debate over the acceptance of PRA in nuclear safety analysis during the 1980's and 1990's. A subset of this debate centered on the use of frequentist (classical) statistics versus the use of Bayesian (subjective) statistics. An issue of *Reliability Engineering & System Safety* was dedicated to the practicality of each method within a PRA [86], and the NRC weighed in with

NUREG-1489, *A Review of NRC Staff Uses of Probabilistic Risk Assessment* [87], in 1994. Here, the NRC offered guidelines on the use of PRA, including the statement that,

> "There is general agreement that both frequentist and subjectivist interpretations of probability are appropriate for use in PRA. However, one view or the other may be preferable for particular analyses."

The NRC provided guidance on the use of frequentist and Bayesian statistics in NUREG-1489 [87]. In this document, advantages and disadvantages of each method were noted. A selection of these pros and cons can be found in Table 4. 1.

Table 4. 1: Frequentist and Bayesian Pros and Cons [87]

Method	Frequentist (Classical) Statistics	Bayesian (Subjective) Statistics
Advantages	- Results depend only on data - Good estimates with large quantity of data - Historical precedence, well known and widely used	- Provides logical and unified approach to the use of prior information - Has probabilistic interpretation which can be easily propagated through a PRA - Easily updates - More applicable when generic data exists
Disadvantages	- Confidence interval has no probabilistic interpretation - Cannot use prior relevant information - Very difficult to propagate confidence intervals through fault or event tree models - Sensitive to the way data is collected	- Suitable prior must be identified and justified - Sensitive to prior distribution - Less well know and accepted, may require more effort to implement and interpret

As the table shows, the main drawbacks of frequentist statistics lie in their inability to use relevant prior information, and the difficulties when propagating uncertainty. Using this information, NUREG/CR-6823, *Handbook of Parameter Estimation for Probabilistic Risk Assessment* [88], recommends using Bayesian statistics for parameter estimation.

This would seem like a definitive answer to which technique should be used for the goal of parameter estimation and comparison to a limit, but that is not the case. When NUREG/CR-6823 refers to "parameter estimation" it means the estimation of system parameter distributions for use in a PRA. This includes component failure rates, initiating event frequencies, and equipment non-recovery probabilities. This definition is not necessarily the same as used in this document, where parameter estimation is the estimation of a true property of a distribution, such as a quantile. In the case of NUREG/CR-6823, estimating system and component parameters involves collecting experimental component data and previous plant history. These data can come from many sources, and could have been collected in different ways. This is one of the reasons why the ability to specify priors is of great advantage. Also, once the parameters are estimated, they will be inputs into a PRA, so the ease at which Bayesian uncertainties can be propagated through an analysis is another big advantage. Neither of these reasons is applicable to the analysis being conducted here.

In the regulatory analysis of comparing a system parameter to a limit value, the use of prior information presents potential difficulties. It may be problematic to justify any prior for use in a regulatory analysis, if it could be shown that this prior would modify the results in the licensee's favor. This is one of the reasons why the VRTs of control variates and importance sampling are not investigated in this work. Also, the data are not collected from many different sources, but from a single analysis that was conducted according to regulatory guidelines. So the data are regular and frequentist statistics can be applied. Lastly, the results of the analysis are the ultimate motivation for

its collection. They are not inputs into a larger system, but the final goal. This means the ease by which the confidence interval can be propagated is not a concern.

Even without those advantages, Bayesian statistics would still seem like a more natural fit to the NRC's "95% probability" requirement, since credible intervals return probabilistic information. However, this is not necessarily the case. The probability requirement is fulfilled by the use of the 0.95-quantile, not the confidence or credible interval. The use of the 0.95-quantile implies that there is a 95% probability of the output of the system being below that value. The use of the confidence interval simply provides an estimation of the quantile location. So the probability characteristics of the quantile are retained, and there is not necessarily an advantage to an additional probabilistic statement about the location of the quantile provided by the credible interval. It may be that there are other benefits to the additional probabilistic statement, but for this work that is not the case. Also, as stated in this section, the CI gives an indication of the uncertainty in the sampling method, not the system uncertainty. This is a positive in this application, since the quantile satisfies the probability requirement related to system uncertainty, and CI provides the regulator with confidence regarding how the experiment was conducted (i.e. the sampling scheme). For these reasons, the use of CIs rather than credible intervals would appear acceptable in this application.

As mentioned before, the confidence level gives a percentage of how often the true parameter will lie within the interval. The confidence limits are the upper and lower bounds of the CI. While the example in Figure 4. 1 shows a two-sided CI, it is also possible to construct a one-sided confidence interval (OSCI). This tends to be the more

useful CI for the problems described here, since the approximate location of a parameter is not the main interest, rather its location compared to a limit or goal value.

4.1.2.2. Relation to Hypothesis Testing

Usually, simply reporting CIs for an estimation of a population parameter is not considered a hypothesis test. This is because hypothesis tests relate to a single conclusion, such as statistical significance versus no statistical significance, where a CI is only the reporting of a range of plausible values for that system parameter. Many times a CI could be reworded to become a hypothesis test [28], and as Section 2.2 mentioned, there is a confidence interval approach to hypothesis testing. In this case, instead of simply reporting a CI around a sample statistic, some hypothesized value for that parameter is compared to see if it falls in or out of that interval. Increasingly, this approach to hypothesis testing is gaining favor over the use of *p*-values, and recent medical journal publications now prefer CIs to the use of *p*-values [17]. CIs are gaining preference over *p*-value testing because they are more informative than the *p*-value approach [89]. A CI provides a measure of accuracy of the parameter estimation that a point estimate and significance value do not.

The terminology laid out by the NRC to meet the 95% probability reporting requirement can be reworded in order to create a hypothesis test that would fall into the confidence interval approach category. As stated in Section 2.2, a hypothesis test begins with an assumption about a parameter, but uses a statistic for the decisionmaking process. In the case of an output distribution satisfying a regulatory limit with at least 95% probability, the null hypothesis H_0 is that the true 0.95-quantile of the output distribution

is above the regulatory limit (assigned value b). The alternative hypothesis H_1 is that the true 0.95-quantile of the system is below the regulatory limit b, as shown in Table 4. 2.

Table 4. 2: Hypothesis Test Alternatives

H_0	True 0.95-quantile $\xi_{0.95} > b$ *System fails test*
H_1	True 0.95-quantile $\xi_{0.95} < b$ *System passes test*

By making H_0 the case where the true quantile is above the limit, the default position is that the system should fail the regulatory test because there is a greater than 95% probability of the output of the system being greater than b. Since this is the null hypothesis, the analyst must prove that this is not the case, or to state it another way, the analyst must provide a statistically significant amount of evidence that the true 0.95-quantile of the output is below limit b. Obviously, the parameter $\xi_{0.95}$ is unknown. Therefore, the quantile estimator $\tilde{\xi}_{0.95}$ will be used, with a 95% OSCI, as the test statistic.

To make this more clear, Figure 4. 2 shows quantile estimator $\tilde{\xi}_p$, and a one-sided confidence interval (OSCI) for the p-quantile ξ_p (note: the OSCI actually extends to $-\infty$, but obviously negative values are not realistic).

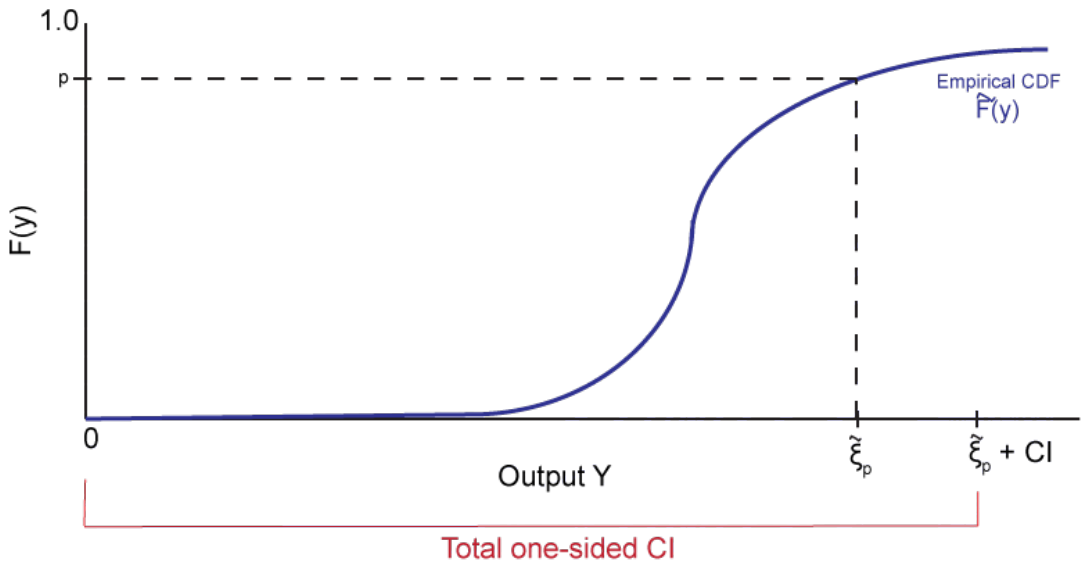

Figure 4. 2: One-sided CI for Quantile Estimator $\tilde{\xi}_p$

As the figure shows, the interval extends from 0 to value $\tilde{\xi}_p + CI$. If the limit value b falls anywhere within this interval, the system should not pass the test. Therefore, the null hypothesis should be accepted if the limit value b is within the interval 0 to value $\tilde{\xi}_p + CI$. The alternative hypothesis should be accepted if the limit value b is *not* in this interval. Since the CI is one-sided, there is only one way the limit value b could not fall in the interval, which would be by exceeding the highest bound. Using a 95% OSCI is equivalent to testing the null hypothesis at the $P < 0.05$ level. Another way of saying this is that if b falls in this interval, there is a not insignificant possibility (where the line between significant and insignificant is made by the 95% confidence) that it could be at a value equal to, or below, the true population parameter ξ_p.

Since the problem is being phrased in terms of hypothesis testing, more explanation is needed on how errors will arise in this framework. As described in Section

4.1.2.1, a 95% OSCI implies that the true parameter value will be captured by the interval 95 times out of 100. This means that 5% of the time the top bound of the OSCI created from the samples will be at a value lower than the true quantile. This scenario is shown in Figure 4. 3.

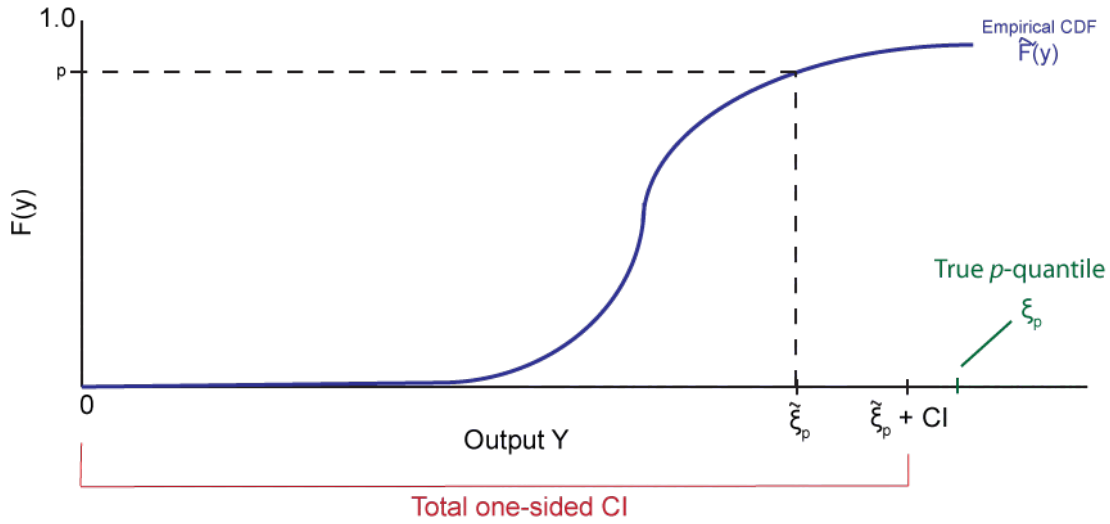

Figure 4. 3: True Quantile above OSCI

Even though the true quantile lies outside the interval created by the OSCI, this does not directly imply that a Type-I error (false positive) will be committed. A Type-I error will only occur if the limit value *b* happens to have been set at a value between the upper bound of the OSCI and the true quantile, as shown in Figure 4. 4 (it is important to note here, that the limit value *b* is a fixed parameter that was set before the analysis; the error interval in Figure 4. 4 simply shows a range created by the upper bound of the OSCI where, if the limit *b* had been placed, a Type-I error would occur).

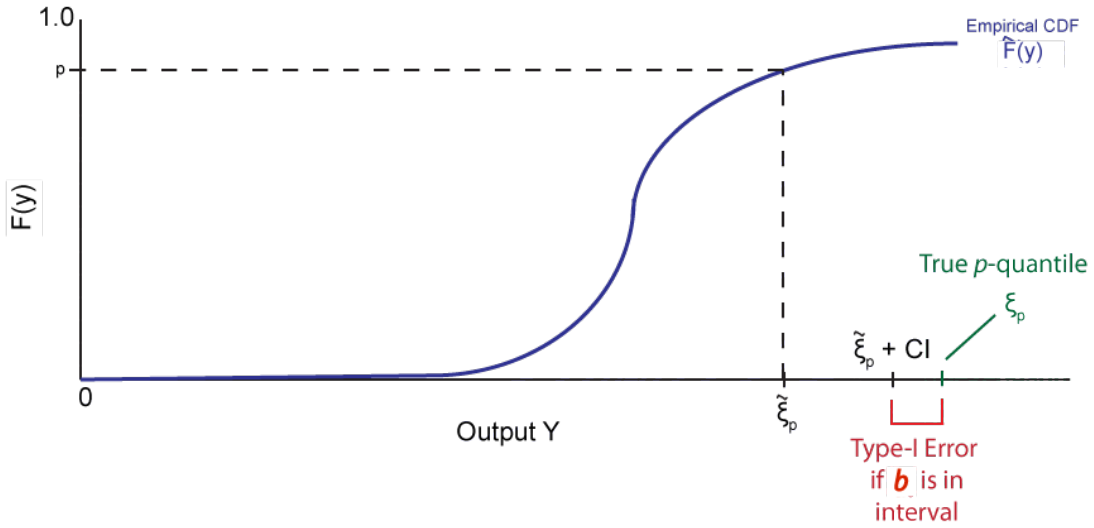

Figure 4. 4: OSCI with Type-I Error

Therefore, the size of the test α is at most 0.05 (assuming a correct 95% OSCI). The actual value for α will depend on the location of the upper bound of the OSCI in relation to limit b. The larger the distance between the upper bound of the OSCI and the true quantile, the closer α could be to 0.05, since the error interval in Figure 4. 4 will grow wider. Since the limit b is set beforehand, independently of the analysis, as the error interval grows wider, the probability of the top bound of the OSCI falling below b will increase, meaning α will get closer to 0.05. It should be noted that even if a Type-I error does not occur, there can still be mistakes caused by these ~5% of results. Even if the hypothesis test were to reach the right conclusion, the underestimation of the true quantile can lead to incorrect decisions in relation to the ranking of the severity of accidents (or whatever situation the analyst may be investigating), and this fault should not be underestimated.

A Type-II error would occur in the following scenario. In this case, the OSCI does incorporate the true quantile. However, the OSCI is so large that it also includes the limit value *b*, even though it is set above the true quantile, as show in Figure 4. 5. As the figure shows, the error can be induced by the size of the CI, but it is also possible that the quantile estimator $\tilde{\xi}_p$ is above the true quantile, which means no matter the width of the CI, there is still a chance of a Type-II error, as show in Figure 4. 6.

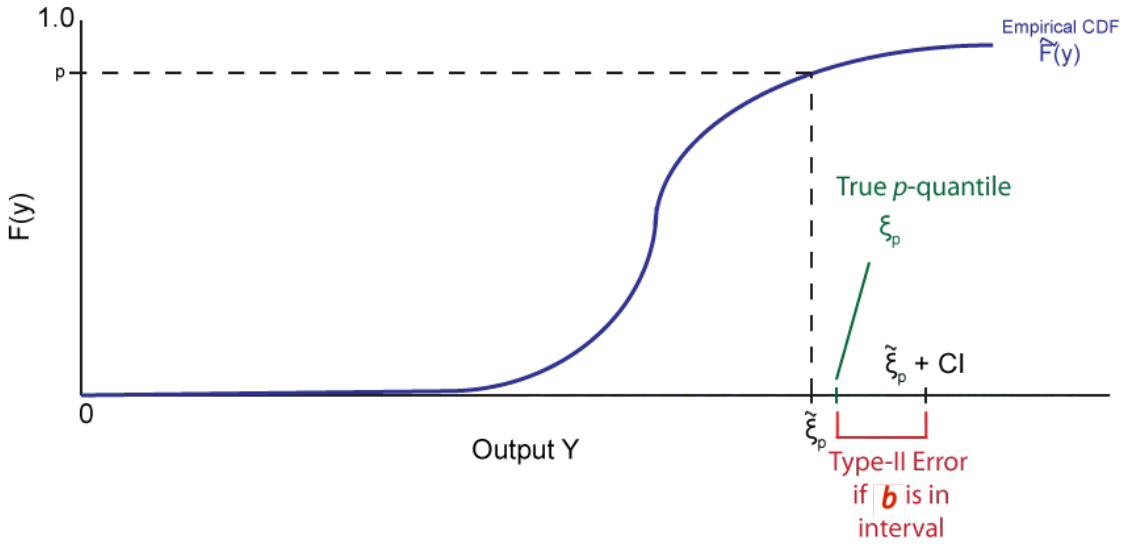

Figure 4. 5: OSCI with Type-II Error[1]

[1] In Figure 4. 5 and Figure 4. 6, it appears that the true quantile has changed value in comparison to Figure 4. 4; using frequentist statistics, the true quantile is fixed, it would actually be the empirical CDF and OSCI that have changed from the previous figure

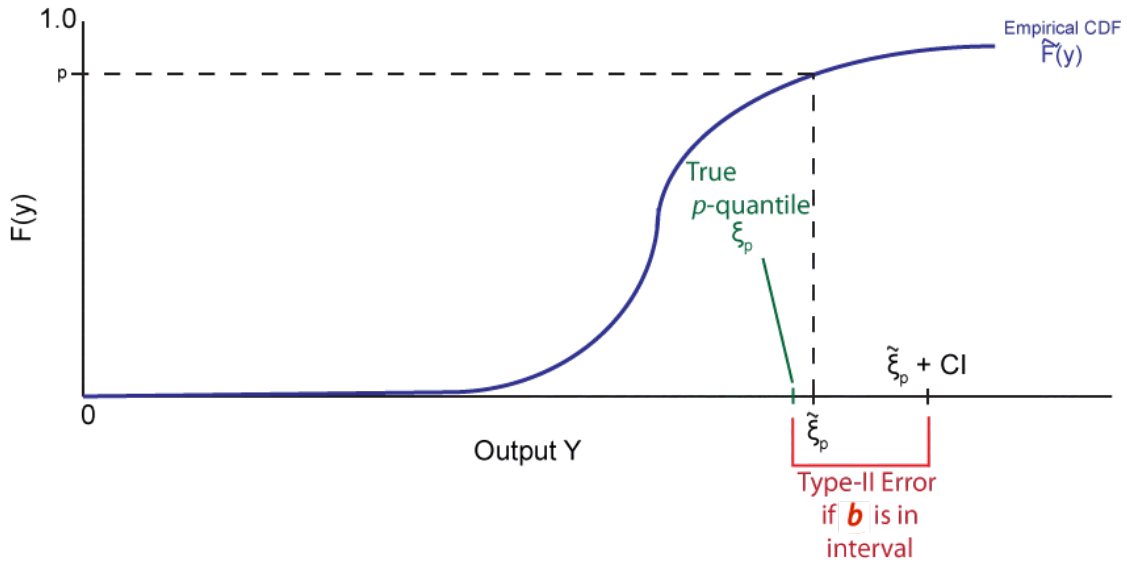

Figure 4. 6: OSCI with Type-II Error with Over Estimation of Quantile

Like α, the value for β is bound by the value of the confidence level. With a 95% confidence interval, only 5% of trials will result in the true quantile exceeding the top bound of the OSCI. This means the true quantile will be within the bounds of the OSCI 95% of the time. However, the OSCI does not give any information about the location of the true parameter within the CI, just a value to the possibility of its location within the bounds. This means β has a theoretical top bound of 0.95, if the limit value b happened to be set only slightly above the true quantile. Obviously, a test that resulted in a β value close to 0.95 would not be of much use. The actual value for β is completely dependent on the accuracy of the quantile estimation, and the precision of the CI. The probability of the upper bound of the OSCI lying above the true quantile and above the limit value b will increase as the overestimation of the 0.95-quantile by the OSCI increases. The closer the point estimate is to the true quantile, and the skinnier the CI, the smaller β will be.

In this framework it is possible to reduce the value for both α and β. To reduce α, the upper bound of the OSCI must be as close to the true quantile as possible during those ~5% of times when the OSCI does not incorporate it. To reduce β, the accuracy of the quantile estimation and the precision of the CI should both be improved. The only way to accomplish these tasks without increasing the number of samples (assuming the analyst has no control over the limit value) is to reduce the variance of the test statistic.

Beyond committing errors during regulatory analysis, there are other reasons utilities and regulators would like to increase the accuracy of resulting confidence for a quantile, or similar, value. The margin from the resulting value to the safety limit is also of use. Significant margin may allow utilities to increase reactor temperature or power, increasing profit. As Westinghouse has stated, "The quantification and tracking of the margin (to the safety limit) is most often requested by both the plant operator and the regulator,…" [75].

4.2. Methods

This section documents the procedure to establish confidence intervals for quantiles using various methods. For the techniques using VRTs, more detail is given about the derivation of the method since they have only recently been proven. There is also a small aside about the asymptotic methods and hypothesis testing. This is included here in order to offer a point of comparison for the methods detailed in Section 5, which use a different test statistic.

4.2.1. Crude Monte Carlo using Order Statistics

There are several key properties that make CMC-OS appealing to nuclear safety analysts. The biggest benefits of the CMC-OS method are that it is nonparametric and non-asymptotic. Nonparametric means that the method is independent of the outputs' probability distribution, as long as it is continuous. Since it is non-asymptotic, the validity of the confidence statement holds exactly for certain finite sample sizes *n* and does not depend on *n* growing toward infinity. Called *bracketing* by Nutt and Wallis [82], the CMC-OS method first fixes an integer $r \geq 1$ (variable *m* in Nutt and Wallis [82]) and then determines the number *n* of runs necessary so that the *r*-th largest output of the *n* runs is a valid 95/95 value, which is the upper endpoint of an 95% upper one-sided confidence interval for $\xi_{0.95}$. Then the NRC criterion is verified by checking if the 95/95 value lies below the safety limit. With this method, it is also possible to find the number *n* of runs necessary to construct a valid confidence interval for any quantile.

The required value for *n*, when $r = 1$, can be determined as follows. Suppose that *n* i.i.d. runs are performed, giving *n* i.i.d. outputs, and consider the true *p*-quantile ξ_p. Each of the *n* outputs has probability *p* of lying below ξ_p, so the probability that all *n* outputs are less than ξ_p is p^n. Thus, the probability that at least one output is larger than ξ_p is 1-p^n, so the probability that the largest of the *n* outputs is greater than ξ_p is

$$\beta = 1 - p^n. \qquad \text{Eq. 42}$$

Setting $\beta = p = 0.95$ and solving for *n* in Eq. 42 results in $n = 59$. Thus, if 59 CMC runs are conducted, then the largest (i.e., $r = 1$) of the 59 outputs is a 95/95 value. If a 95% confidence interval is desired for the 0.75-quantile, then $\beta = 0.95,\ p = 0.75$, and $n = 11$.

This means if 11 runs are conducted and ordered, the largest output can be taken as a 95/75 value.

A drawback of taking the largest of 59 runs as the 95/95 value or the largest of 11 runs as a 95/75 is that it will typically have large variance since the number of CMC runs is so small. This usually leads to a large range of possible 95/95 values, which in most cases will be conservative in the sense that they are considerably larger than the true quantile of the probability distribution of the model's output. To obtain a more *accurate* 95/95 or 95/75 value, the value of *r* can be increased, which will lead to larger run size *n*. Here, *accuracy* is defined as the distance from the 95/95 or 95/75 value to the true quantile ξ_p, and *precision* is the spread or range of possible 95/95 or 95/75 values. For $r \geq 1$, the argument used to obtain Eq. 42 can be generalized to show that the probability that the *r*-th largest of the *n* outputs is larger than ξ_p is

$$\beta = 1 - \sum_{i=n-r+1}^{n} \frac{n!}{i!\,(N-i)!} p^i (1-p)^{n-1}. \qquad \text{Eq. 43}$$

Now set $\beta = p = 0.95$ and fix $r \geq 1$ in Eq. 43. Then solving for *n* gives the number of runs needed to ensure that the *r*-th largest output of the *n* runs is a valid 95/95 value. For example, if $r = 1$, then Eq. 43 reduces to Eq. 42, resulting in $n = 59$, as before. If $r = 3$, then $n = 124$, so the third largest output from the 124 runs is a valid 95/95 value.

As stated before, the potential downsides from CMC-OS arise not from the resulting values being invalid, but from the variance and conservatism of the results. First, at lower run levels, the 95/95 value will, on average, be overly conservative. This can be seen in Figure 4. 7, based on a similar plot by Nutt and Wallis [82]. This figure

shows for the different values of r, the probability density of the 95/95 value of a CMC-OS analysis. The probability density is computed by taking the derivative of Eq. 43 with respect to p and is shown as a function of p. Thus, the function gives the likelihood of the 95/95 value lying in a small interval around the p-quantile for different values of p. For $r = 1$, which corresponds to $n = 59$, the 95/95 value is more likely to fall in an interval near the 1.0-quantile than an interval near the 0.95-quantile. Even at $r = 40$ ($n = 1008$ runs), the 95/95 value is more likely to be near the 0.96-quantile than the 0.95-quantile.

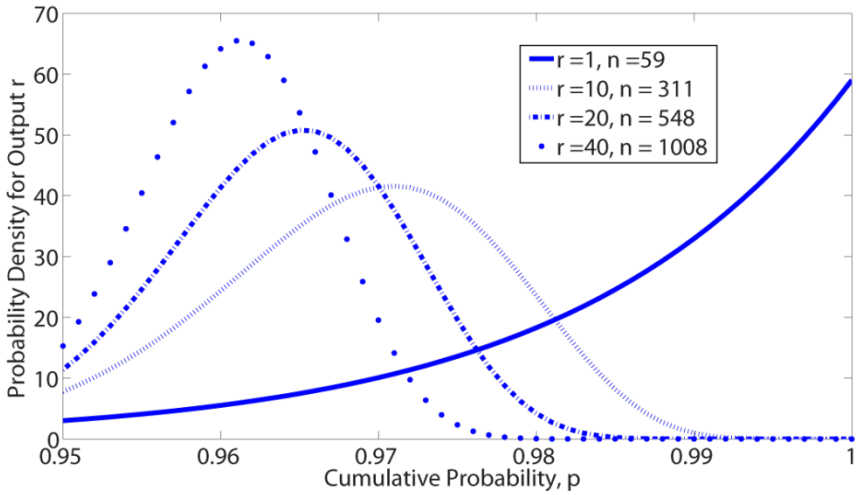

Figure 4. 7: Dependence on the Order Selected to Represent 0.95-Quantile

The same is true when trying to find a 95/75 value, as Figure 4. 8 shows. Even at $n = 886$ runs, the resulting value is more likely to be near the 0.77-quantile than the 0.75-quantile.

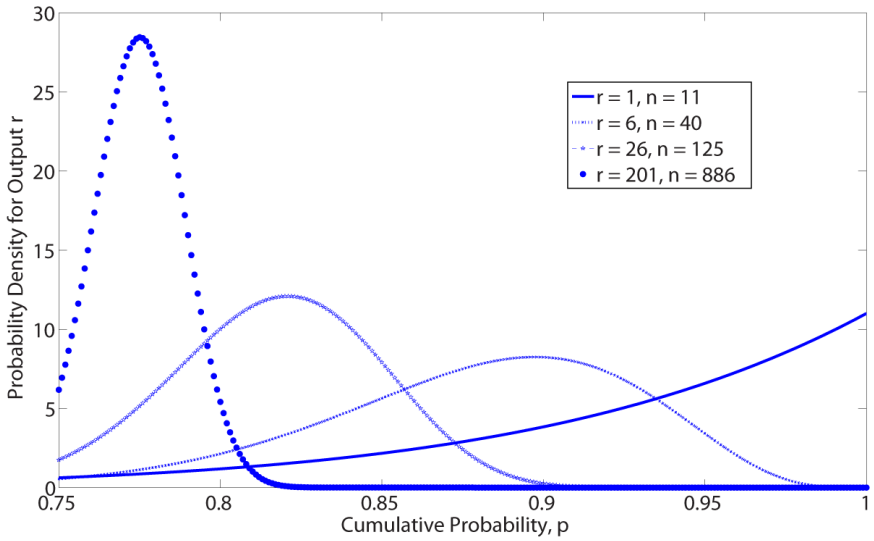

Figure 4. 8: Dependence on the Order Selected to Represent 0.75-Quantile

Secondly, as stated above, since CMC sampling is used, the variance of the 95/95 value can be very high when *r* (and subsequently *n*) is small, so the likely range of 95/95 values is large. This is even more so when estimating a 95/75, since as Figure 4. 8 shows, as few as 11 runs can be conducted to find a 95/75 value. In certain cases, this can mean that even though only ~5% of trials will fall below the actual quantile (due to the 95% confidence), there is a not-insignificant chance that they could fall well below. This could potentially cause a Type-I error during the analysis, and result in a value that is closer to the true capacity limit of the system, as shown in the safety margin characterization in Figure 2. 2.

4.2.2. Asymptotic Methods

In contrast to the CMC-OS method, which states *a priori* a set number of runs which must be conducted to establish a confidence interval for a quantile, it is also

possible to establish other confidence intervals by proving a central limit theorem (CLT) as the number of runs grows large. This method has long been known when using CMC sampling [90], but until recently, has not been proven when using variance reduction techniques (VRTs), described in Section 2.

The following sections review asymptotic confidence intervals for CMC sampling, and discuss the recent work to expand their applicability to VRTs. For this work, only LHS and AV were investigated. This is because these methods can generally be applied without using features of the system. Using a VRT that relies on detailed knowledge about the system to adjust sampling methods or outputs, such as importance sampling and some types of control variates, may cause reluctance among regulators since they cannot be applied generally. It is important to note that assumptions are still needed about the system when using LHS and AV to guarantee they reduce variance. Both methods are essentially guaranteed to reduce the variance of the output if the system is a monotone function of the inputs, meaning increasing an input value will lead to the output either always decreasing or always increasing. It is still possible to get variance reduction if this is not true, but it is not ensured [91].

For this explanation, suppose output Y from the simulation model can be represented as

$$Y = g(U_1, U_2, \dots, U_d), \qquad \text{Eq. 44}$$

where g is a given (deterministic) function having a fixed number d of arguments and $U_1, U_2, \dots, U_d$ are i.i.d. uniform[0,1] random variables. The function g, which takes the d the i.i.d. uniforms and transforms them into a single output Y, can be quite complicated,

and it may not be possible to express g in closed-form. For example, a LOCA simulation might have s input random variables that are fed into a detailed computer code, which then computes an output Y. In this case, the function g transforms the d uniforms into samples of the s input variables, runs the computer code with these inputs, and produces an output Y (in many settings, $s = d$, and each input variable X_j is sampled from its distribution G_j via inversion, i.e., $X_j = G_j^{-1}(U_j)$). Let F be the CDF of Y, so for $0 < p < 1$, the p-quantile is $\xi_p = F^{-1}(p) \equiv \inf\{x : F(x) \geq p\}$.

4.2.2.1. Review of CMC

This section reviews how to use CMC to estimate and construct an asymptotically valid confidence interval for ξ_p based on a CLT when Y has the form in Eq. 44. It is possible to generate n i.i.d. copies of Y by first generating nd i.i.d. uniform[0,1] random variables $U_{i,j},\ i = 1,2,\dots,n,\ j = 1,2,\dots,d$, where d is as defined in Eq. 44. These uniforms can be arranged in an $n \times d$ array

$$\begin{matrix} U_{1,1} & U_{1,2} & \cdots & U_{1,d} \\ U_{2,1} & U_{2,2} & \cdots & U_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ U_{n,1} & U_{n,2} & \cdots & U_{n,d} \end{matrix}, \qquad \text{Eq. 45}$$

where the ith row is used to generate the ith output Y_i, i.e.,

$$Y_1 = g(U_{1,1}, U_{1,2}, \dots, U_{1,d})$$

$$Y_2 = g(U_{2,1}, U_{2,2}, \dots, U_{2,d})$$

$$\vdots$$

$$Y_n = g(U_{n,1}, U_{n,2}, \dots, U_{n,d})$$

Each Y_i has the distribution F because the d entries in the ith row of Eq. 45 are i.i.d. uniforms, as required by Eq. 44. Moreover, $Y_1, Y_2, \ldots, Y_n$ are independent by the independence of the n rows in Eq. 45. Then the CMC p-quantile estimator is computed as $\hat{\xi}_{p,n} = \hat{F}_n^{-1}(p)$, where $\hat{F}_n$ is defined in Eq. 33 (for the following derivations, a hat $\wedge$ will be used to denote an estimated parameter using CMC sampling, and the tilde ~ will be used to denote an estimated parameter using a VRT).

To establish a CI for ξ_p based on the CMC point estimator $\hat{\xi}_{p,n}$, it must be shown that $\hat{\xi}_{p,n}$ satisfies a CLT as the number n of samples grows large. One way of establishing this is by first proving that $\hat{\xi}_{p,n}$ satisfies a so-called *Bahadur representation*; see [92]. Let f denote the derivative, when it exists, of F, and assume that $f(\xi_p) > 0$. Now consider the following heuristic argument. When n is large, $\hat{F}_n \approx F^{-1}(p) = \xi_p$. Because $F(\xi_p) = p$ by definition, it can be seen that $F(\hat{\xi}_{p,n}) \approx p$, so a Taylor approximation yields

$$\begin{aligned} p &\approx F(\hat{\xi}_{p,n}) \\ &\approx F(\xi_p) + f(\xi_p)(\hat{\xi}_{p,n} - \xi_p) \\ &\approx \hat{F}_n(\xi_p) + f(\xi_p)(\hat{\xi}_{p,n} - \xi_p), \end{aligned}$$

where the last approximation holds because $\hat{F}_n \approx F$. Rearranging terms leads to $\hat{\xi}_{p,n} = \xi_p + [p - \hat{F}_n(\xi_p)]/f(\xi_p)$, which approximates a quantile estimator by a linear transformation of a CDF estimator.

Bahadur [92] makes this argument mathematically rigorous. In particular, suppose that the second derivative F'' of F exists and is bounded in a neighborhood of ξ_p, and that $f(\xi_p) > 0$. Then Bahadur proves that

$\hat{\xi}_{p,n} = \xi_p + \frac{p - \hat{F}_n(\xi_p)}{f(\xi_p)} + R'_n$, where $R'_n = O(n^{-3/4} \log n)$ as $n \to \infty$ with probability 1. Eq. 46

This is known as a Bahadur representation.

Under weaker conditions, Ghosh [93] establishes a variant of a weaker version of Eq. 46, which will be useful and sufficient for the following proof. Specifically, let p_n be a perturbed value of p converging to p as $n \to \infty$, and let $\hat{\xi}_{p_n,n} = \hat{F}_n^{-1}(p_n)$ (working with a perturbed p_n rather than a fixed p will allows an asympotitc CI to be constructed for $\xi_p$when applying VRTs). Also, let $\Rightarrow$ denote convergence in distribution (Section 1.2.4 of [94]). Then [93] shows that if $f(\xi_p) > 0$, then

$$\hat{\xi}_{p_n,n} = \dot{\xi}_{p_n} + \frac{p - \hat{F}_n(\xi_p)}{f(\xi_p)} + R_n \qquad \text{Eq. 47}$$

with

$$\sqrt{n}R_n \Rightarrow 0 \text{ as } n \to \infty \qquad \text{Eq. 48}$$

where

$$\dot{\xi}_{p_n} = \xi_p + \frac{p_n - p}{f(\xi_p)} \qquad \text{Eq. 49}$$

when $p_n = p + O(1/\sqrt{n})$. If f is also continuous in a neighborhood of ξ_p, then Eq. 47 and Eq. 48 hold for all $p_n \to p$ with

$$\dot{\xi}_{p_n} = F^{-1}(p_n). \qquad \text{Eq. 50}$$

The results in Eq. 47 and Eq. 48 ensure that the CMC quantile estimator $\hat{\xi}_{p,n}$ satisfies a CLT. To show this, fix $p_n = p$ in Eq. 47 so $\dot{\xi}_{p_n} = \xi_p$, rearrange terms and scale by $\sqrt{n}$ to obtain

$$\sqrt{n}(\hat{\xi}_{p,n} - \xi_p) = \sqrt{n}\left(\frac{p - \hat{F}_n(\xi_p)}{f(\xi_p)}\right) + \sqrt{n}R_n. \qquad \text{Eq. 51}$$

Eq. 33 shows that $\hat{F}_n(\xi_p)$ is the sample average of i.i.d. indicator functions $I(Y_i \leq \xi_p)$, $i = 1,2,\ldots,n$, each of which has mean p and variance $0 < p(1-p) < \infty$. Hence, the first term on the right side of Eq. 51 satisfies a CLT (see p. 28 of [94]), with limit $N(0, p(1-p)/f^2(\xi_p))$ as $n \to \infty$, where $N(a, b^2)$ denotes a normal random variable with mean a and variance b^2. The second term on the right side of Eq. 51 vanishes (in distribution) as $n \to \infty$ by Eq. 48, so Slutsky's theorem (p. 19 of [94]) ensures that $\sqrt{n}(\hat{\xi}_{p,n} - \xi_p) \Rightarrow N(0, p(1-p)/f^2(\xi_p))$ as $n \to \infty$, or equivalently,

$$\frac{\sqrt{n}}{\sqrt{p(1-p)}\lambda_p}(\hat{\xi}_{p,n} - \xi_p) \Rightarrow N(0,1) \quad \text{as } n \to \infty, \qquad \text{Eq. 52}$$

where

$$\lambda_p = \frac{1}{f(\xi_p)}, \qquad \text{Eq. 53}$$

which is known as the *sparsity function* [95] or the *quantile density function* [96]. One interpretation of the CLT is that the left of Eq. 52 will have approximately a standard (i.e., mean 0, variance 1) normal distribution for large n.

The CLT in Eq. 52 illustrates one reason why a Bahadur representation is useful. The latter shows that a quantile estimator can be approximated as a linear transformation of a CDF estimator, which typically is a sample average so it satisfies a CLT. Thus, a Bahadur representation provides insight into why a quantile estimator, which is not a sample average, satisfies a CLT.

Once the CLT in Eq. 52 has been established, it can then be unfolded to obtain a confidence interval for ξ_p. Let $z_\beta = \Phi^{-1}(1-\beta)$ for any $0 < \beta < 1$, where Φ is the CDF of $N(0,1)$. Then

$$1-\alpha = P\{-z_{\alpha/2} \le N(0,1) \le z_{\alpha/2}\}$$

$$\approx P\left\{-z_{\alpha/2} \le \frac{\sqrt{n}}{\sqrt{p(1-p)}\lambda_p}(\hat{\xi}_{p,n} - \xi_p) \le z_{\alpha/2}\right\}$$

$$= P\left\{\hat{\xi}_{p,n} - z_{\alpha/2}\frac{\sqrt{p(1-p)}\lambda_p}{\sqrt{n}} \le \xi_p \le \hat{\xi}_{p,n} + z_{\alpha/2}\frac{\sqrt{p(1-p)}\lambda_p}{\sqrt{n}}\right\}$$

where the approximation holds for large *n* by the CLT. Hence,

$$\left[\hat{\xi}_{p,n} - z_{\alpha/2}\frac{\sqrt{p(1-p)}\lambda_p}{\sqrt{n}}, \hat{\xi}_{p,n} + z_{\alpha/2}\frac{\sqrt{p(1-p)}\lambda_p}{\sqrt{n}}\right] \equiv \left[\hat{\xi}_{p,n} \pm z_{\alpha/2}\frac{\sqrt{p(1-p)}\lambda_p}{\sqrt{n}}\right] \quad \text{Eq. 54}$$

is an asymptotically valid (two-sided) $100(1-\alpha)\%$ confidence interval for ξ_p. Since λ_p is unknown, for the CI in Eq. 54 to be implementable in practice, it must be replaced with a *consistent* estimator $\hat{\lambda}_{p,n}$; i.e., $\hat{\lambda}_{p,n} \Rightarrow \lambda_p$ as $n \to \infty$. If such an estimator exists, then

$$J_n = \left[\hat{\xi}_{p,n} \pm z_{\alpha/2}\frac{\sqrt{p(1-p)}\hat{\lambda}_{p,n}}{\sqrt{n}}\right]$$

is another asymptotic two-sided $100(1-\alpha)\%$ CI for ξ_p, which is *asymptotically valid* in the sense that

$$P\{\xi_p \in J_n\} \to 1-\alpha \quad as \quad n \to \infty.$$

Now the key issue is constructing a consistent estimator $\hat{\lambda}_{p,n}$ of λ_p from Eq. 53. Since $\lambda_p = 1/f(\xi_p) = \frac{d}{dp}F^{-1}(p) = \lim_{h\to 0}[F^{-1}(p+h) - F^{-1}(p-h)]/(2h)$ by the chain rule of differentiation, a natural estimator for λ_p is the (central) finite difference

$$\hat{\lambda}_{p,n} = \frac{\hat{F}_n^{-1}(p+h_n) - \hat{F}_n^{-1}(p-h_n)}{2h_n}, \quad \text{Eq. 55}$$

where $h_n > 0$ is a user-specified (small) parameter known as the *bandwidth* or *smoothing parameter* (see Section VII.1 of [97] or Section 7.1 of [39] for overviews of finite-difference estimators). If $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$, then [98] and [99] prove the consistency of $\hat{\lambda}_{p,n}$ as $n \to \infty$. More detail on this estimator is provided in Section 4.2.2.5.

Rather than a two-sided CI for ξ_p, an asymptotic upper one-sided $100(1-\alpha)\%$ CI for ξ_p can be developed

$$\left(-\infty, \hat{\xi}_{p,n} + z_\alpha \frac{\sqrt{p(1-p)}\hat{\lambda}_{p,n}}{\sqrt{n}}\right]. \qquad \text{Eq. 56}$$

Setting $\alpha = 0.05$ (so $z_\alpha = 1.645$), then the upper endpoint of Eq. 56 is an asymptotically valid 95/95 (resp., 95/75) value for CMC as $n \to \infty$ when $p = 0.95$ (resp., $p = 0.75$).

To implement this procedure for a OSCI, the following code in Figure 4. 9 can be used, where *p* is the quantile, *n* is the number of runs, *ceil* is the round-up function, and *NN* is the standard normal critical point for the desired confidence level. This code first estimates the quantile $\hat{\xi}_{p,n}$, called *Xi*, using the ordered results *Y_ordered*, and the round –up function. Then the CFD in Eq. 55 is calculated. Next, the quantity to the right-hand side of the plus sign in Eq. 56 is calculated using these results. Lastly, the OSCI in Eq. 56 is calculated.

```
%%% CMC Asymptotic
    Xi=Y_ordered(ceil(p*n));       % Quantile Estimation using ordered results
    CFDhigh=Y(ceil((p+hn)*n));     % CFD High Point
    CFDlow=Y(ceil((p-hn)*n));      % CFD Low Point
    CFD=(CFDhigh-CFDlow)/((ceil((p+hn)*n)-ceil((p-hn)*n))/(n));  %CFD
    add_on=NN*((sqrt(p*(1-p))*CFD)/sqrt(n));  % Confidence Term
    Xi_w_conf=Xi+add_on;           % Quantile Estimation plus Confidence
```

Figure 4. 9: MATLAB Code Implementation of CMC Quantile Asymptotic Method

The hypothesis test at the beginning of this section can now be written more rigorously using the above asymptotic CMC formulation. The quantile-estimator (here the quantile will be 0.95 for ease of reference) is $\hat{\xi}_{0.95}$ and satisfies the following CLT:

$$\frac{\sqrt{n}}{\tau}(\hat{\xi}_{0.95} - \xi_{0.95}) \approx N(0,1) \qquad \text{Eq. 57}$$

for large n, where

$$\tau = \frac{\sqrt{0.95(1-0.95)}}{f(\xi_{0.95})} \qquad \text{Eq. 58}$$

as can be seen in Eq. 56. Here, $\hat{\tau}$ will be an estimator for τ (using the CFD in Eq. 55). Then the upper endpoint of the OSCI U is

$$U = \hat{\xi}_{0.95} + z\frac{\hat{\tau}}{\sqrt{n}}, \qquad \text{Eq. 59}$$

where z is the standard normal critical point for a 95% confidence interval, and a 95/95 criterion is satisfied when $U \leq$ limit value,

For a limit value b, the hypothesis test alternatives for a comparison to the true 0.95-quantile become:

H_0: $\xi_{0.95} > b$

H_1: $\xi_{0.95} \leq b$

Then Eq. 59 can be rearranged to show when to reject H_0,

$$reject\ H_0\ if\ and\ only\ if\ \frac{\hat{\xi}_{0.95} - b}{\hat{t}/\sqrt{n}} \leq -z \qquad \text{Eq. 60}$$

Eq. 60 is equivalent to saying that the 95/95 value is less than the limit b. Similarly,

$$accept\ H_0\ if\ and\ only\ if\ \frac{\hat{\xi}_{0.95} - b}{\hat{t}/\sqrt{n}} > -z \qquad \text{Eq. 61}$$

which is the same as saying the 95/95 value is greater than the limit b.

Comment on CMC and Quantile Test

When using samples that are i.i.d., like those with CMC, it is possible to conduct a hypothesis test known as the *quantile test*. The quantile test is a type of binomial test that investigates the hypothesized location of a distribution quantile. It will appear very similar to the hypothesis test framework laid out in Section 4.1.2, but it will help demonstrate the relation between CMC-OS and the asymptotic CMC method [100].

The easiest way to explain the quantile test is through example. Imagine a random variable U. An analyst wants to take n samples from U to see if the 0.75-quantile of U is greater than 20 (for example). If this is true, then $< 75\%$ of the n samples should be less than 20, and $> 25\%$ of the samples should be more than 20. If this is not the balance seen in the samples, then it will give an indication of the direction of the true 0.75-quantile. The hypothesis choices are similar to before:

H_0: The 0.75-quantile is less than or equal to 20
H_1: The 0.75-quantile is greater than 20

This can be rephrased in terms of probability, where

H_0: $P(U \leq 20) \geq 0.75$

H_1: $P(U \leq 20) < 0.75$

which is essentially a binomial test because the samples are either less than or equal to 20, or not. This can be thought of the number of successes and failures in a binomial test (also similar to the derivation of CMC-OS in Section 4.2.1.). In this example, the size of the test α will be assigned 0.05. So using the binomial distribution in Eq. 62, it's possible to find how many samples must fall above 20 in order for the significance of the results to exceed $1 - \alpha$.

$$P(U < u) = \left(\frac{n!}{k!\,(n-k)!}\right) p^k (1-p)^{n-k} \qquad \text{Eq. 62}$$

If $n = 15$ samples, $p = 0.75$, and the number of samples < 20 is k, the binomial formula gives the following probabilities:

$$P(k \leq 13) = 0.9198$$
$$P(k \leq 14) = 0.9866$$

This means that in order to satisfy the significance level $1 - \alpha = 0.95$, at least two samples out of 15 should exceed 20 in order for H_1 to accepted.

As can be seen, the process of using CMC-OS in comparison to a limit value is essentially a quantile test. In the example just given, if $n = 59$ samples and the quantile $p = 0.95$, then the result would be $P(k \geq 59) = 0.0485$, which means there is a less than 0.05 probability of the hypothesized value being exceeded by the true 0.95-quanitle, which is the same result as using the CMC-OS method. This relates to the asymptotic CMC method because of the normal approximation of the binomial distribution. The normal approximation to a binomial distribution $B(n, p)$ is $N(np, np(1-p))$. This

means instead of using the binomial formula in Eq. 62, the number of successes needed to achieve the significance level can be found using Eq. 63,

$$t_{(1-\alpha)} = np + z_{(1-\alpha)}\sqrt{np(1-p)} \qquad \text{Eq. 63}$$

where z is the standard normal critical value, and $n - t_{(1-\alpha)}$ is the number of success necessary out of n samples. Eq. 63 is simply the mean of the normal distribution plus the standard deviation times a scaling factor, but closely resembles the asymptotic CMC method in Eq. 56.

The same formula can be used to establish CIs for quantiles too. The difference between this result and the one shown in the asymptotic CMC results in Eq. 56 is the desired information. The result here would give the rank of the ordered sample that would be closest to the desired confidence level. For example, if 100 samples were taken and the 95/95 value was desired, Eq. 63 would return 98.585, which would mean the 99 ordered result would satisfy the 95/95. However, this means the confidence will actually exceed 95% since the result did not fall directly on an ordered result, so the solution will be conservative. The asymptotic CMC method outlined in Section 4.2.2.1 does not return an ordered result, but the actual value of the bounds of the CI. Both methods can be viewed as a result of the CLT, however the form in Eq. 63 is non-asymptotic. Due to this similarity between the methods, it would be assumed that asymptotic CMC should be less conservative than CMC-OS at low run levels, but as the number of runs increases, the two methods will converge to the same solution.

4.2.2.2. VRTS

For the case when applying VRTs, Chu and Nakayama [72] have developed methods for constructing asymptotically valid CIs for ξ_p, which is presented here. Let $\tilde{F}_n$ be an estimate of the CDF F, where $\tilde{F}_n$ is obtained by simulating using a VRT with sampling budget n (Sections 4.2.2.3 and 4.2.2.4 give examples for some specific VRTs). Then a VRT p-quantile estimator is

$$\tilde{\xi}_{p,n} = \tilde{F}_n^{-1}(p).$$

The asymptotic validity of the method in [72] for constructing a CI for ξ_p based on $\tilde{\xi}_{p,n}$ relies on showing that the VRT quantile estimator satisfies a Bahadur representation analogous to the CMC version in Eq. 47 and Eq. 48. Specifically, let $\tilde{\xi}_{p_n,n} = \tilde{F}_n^{-1}(p_n)$ be the VRT p_n-quantile estimator, with p_n a perturbed value of p, and assume that $f(\xi_p) > 0$. Then Chu and Nakayama develop a set of general conditions (denoted as Assumptions A1, A2, and A3 in [72], these assumptions are given in Appendix B) on the VRT CDF estimator $\tilde{F}_n$ to ensure that

$$\tilde{\xi}_{p_n,n} = \dot{\xi}_{p,n} + \frac{p - \tilde{F}_n(\xi_p)}{f(\xi_p)} + R_n \qquad \text{Eq. 64}$$

with

$$\sqrt{n}R_n \Rightarrow 0 \quad \text{as} \quad n \to \infty, \qquad \text{Eq. 65}$$

where $\dot{\xi}_{p,n}$ is as in Eq. 49 when $p_n = p + O(1/\sqrt{n})$. If f is further assumed to be continuous in a neighborhood of ξ_p and Assumption A2 in [72] is slightly strengthened, then Eq. 64 and Eq. 65 hold with $\dot{\xi}_{p,n}$ defined in Eq. 50 for all $p_n \to p$ as $n \to \infty$. Chu and Nakayama [72] show that their Assumptions A1, A3 and the stronger version of A2 hold

(under various moment conditions) for importance sampling, combined importance sampling and stratification, antithetic variates and control variates; Nakayama [101] establishes the same for a type of Latin hypercube sampling.

As in the case of CMC in Section 4.2.2.1, the Bahadur representation in Eq. 64 and Eq. 65 implies that the VRT p-quantile estimator satisfies a CLT

$$\frac{\sqrt{n}}{\kappa_p}(\tilde{\xi}_{p,n} - \xi_p) \Rightarrow N(0,1) \quad \text{as} \quad n \to \infty, \qquad \text{Eq. 66}$$

where

$$\kappa_p = \psi_p \lambda_p, \qquad \text{Eq. 67}$$

ψ_p^2 is the asymptotic variance in the CLT

$$\sqrt{n}(p - \tilde{F}_n(\xi_p)) \Rightarrow N(0, \psi_p^2) \qquad \text{Eq. 68}$$

for the VRT CDF estimator $\tilde{F}_n$ at ξ_p, and λ_p is defined in Eq. 55. The value of ψ_p depends on the particular VRT used and equals $\sqrt{p(1-p)}$ for CMC (compare Eq. 52 and Eq. 66). It turns out that developing a consistent estimator $\tilde{\psi}_{p,n}$ of ψ_p is straightforward; the following sections present such estimators for specific VRTs. The value of λ_p is independent of the VRT applied, and can be estimated using a (central) finite difference

$$\tilde{\lambda}_{p,n} = \frac{\tilde{F}_n^{-1}(p + h_n) - \tilde{F}_n^{-1}(p - h_n)}{2h_n}, \qquad \text{Eq. 69}$$

where $h_n > 0$ is the bandwidth. Chu and Nakayama [72] prove that if their Assumptions A1-A3 hold and $f(\xi_p) > 0$, then

$$\tilde{\lambda}_{p,n} \Rightarrow \lambda_p \quad \text{as} \quad n \to \infty \qquad \text{Eq. 70}$$

for $h_n = c/\sqrt{n}$ for any constant $c > 0$. If it is assumed that f is continuous in a neighborhood of ξ_p and a slightly stronger version of Assumption A2 from [72] holds, then Eq. 70 is true for bandwidths satisfying

$$h_n \to 0 \quad \text{and} \quad \sqrt{n}h_n \to b \quad \text{for some} \quad b \in (0, \infty] \quad \text{as} \quad n \to \infty. \qquad \text{Eq. 71}$$

For example, $h_n = cn^{-v}$ satisfies Eq. 71 for constants $c > 0$ and $0 < v \leq 1/2$. Thus, an asymptotically valid two-sided $100(1-\alpha)\%$ CI for ξ_p when applying a VRT is

$$\left[\tilde{\xi}_{p,n} \pm z_{\alpha/2} \frac{\tilde{\psi}_{p,n}\tilde{\lambda}_{p,n}}{\sqrt{n}}\right]. \qquad \text{Eq. 72}$$

Also, an asymptotically valid upper one-sided $100(1-\alpha)\%$ CI for ξ_p is

$$\left(-\infty, \tilde{\xi}_{p,n} + z_{\alpha} \frac{\tilde{\psi}_{p,n}\tilde{\lambda}_{p,n}}{\sqrt{n}}\right], \qquad \text{Eq. 73}$$

whose upper endpoint is an asymptotically valid 95/95 (resp., 95/75) value when applying the VRT with $\alpha = 0.05$ and $p = 0.95$ (resp., $p = 0.75$).

Eq. 70 holds when $h_n = c/\sqrt{n}$ for some constant $c > 0$ because of the following (the other cases of h_n satisfying Eq. 70 can be handled in a similar manner; see [72] for details). In Eq. 69, note that $\tilde{F}_n^{-1}(p + h_n) = \tilde{\xi}_{p+h_n,n}$ and $\tilde{F}_n^{-1}(p - h_n) = \tilde{\xi}_{p-h_n,n}$, so the Bahadur representation in Eq. 64 and Eq. 65 can be used to analyze the finite difference. Since $p \pm h_n = p \pm c/\sqrt{n} = p + O(1/\sqrt{n})$, $\dot{\xi}_{p,n}$ is set as in Eq. 49; thus, Eq. 64 and Eq. 65 imply there exist $R_{n,1}$ and $R_{n,2}$ satisfying $\sqrt{n}R_{n,i} \Rightarrow 0$, $i = 1,2$, such that

$$\tilde{\lambda}_{p,n} = \frac{1}{2h_n}\left[\left(\xi_p + \frac{p + h_n - p}{f(\xi_p)} + \frac{p - \tilde{F}_n(\xi_p)}{f(\xi_p)} + R_{n,1}\right)\right.$$

$$\left. - \left(\xi_p + \frac{p - h_n - p}{f(\xi_p)} + \frac{p - \tilde{F}_n(\xi_p)}{f(\xi_p)} + R_{n,2}\right)\right]$$

$$= \frac{1}{2h_n}\left[\frac{2h_n}{f(\xi_p)} + R_{n,1} - R_{n,2}\right]$$

$$\Rightarrow \frac{1}{f(\xi_p)} = \lambda_p$$

as $n \to \infty$ since $R_{n,i}/(2h_n) = \sqrt{n}R_{n,i}/(2c) \Rightarrow 0$ for $i = 1,2$, Eq. 65.

4.2.2.3. Antithetic Variates

Instead of generating independent outputs as in CMC, the method of antithetic variates (AV) generates outputs in negatively correlated pairs, which can reduce variance; see Section V.3 of [97] for an overview of AV. If both Y and Y' each have marginal distribution F and are negatively correlated, then (Y, Y') is called an *AV pair*. One way to simulate such a pair is to generate d i.i.d. $uniform[0,1]$ random variables $U_1, U_2, \ldots, U_d$, and then set $Y = g(U_1, U_2, \ldots, U_d)$ and $Y' = g(1 - U_1, 1 - U_2, \ldots, 1 - U_d)$, where g is from Eq. 44. Clearly, Y has CDF F by Eq. 44, but Y' also does since each $1 - U_j$ is also $uniform[0,1]$. If g is monotonic in each argument U_j, then Y and Y' are guaranteed to be negatively correlated (p. 181 of [102]), which will ensure a variance reduction, as will be shown shortly. AV can still result in a variance reduction when g is not monotonic in each argument, but it may be difficult to prove.

To estimate ξ_p using AV, generate $n/2$ AV pairs $(Y_i, Y_{i'})$, $i = 1,2, \ldots, n/2$, where n is even. This can be accomplished by generating i.i.d. uniforms as in Eq. 45, but with $n/2$

rows rather than n. Then set $Y_i = g(U_{i,1}, U_{i,2}, \ldots, U_{i,d})$ and $Y_{i'} = g(1 - U_{i,1}, 1 - U_{i,2}, \ldots, 1 - U_{i,d})$. The then AV estimator $\tilde{F}_n$ of the CDF F can be computed as

$$\tilde{F}_n(y) = \frac{1}{n/2} \sum_{i=1}^{n/2} \frac{1}{2} [I(Y_i \leq y) + I(Y_{i'} \leq y)], \quad \text{Eq. 74}$$

and the resulting AV estimator of ξ_p is $\tilde{\xi}_{p,n} = \tilde{F}_n^{-1}(p)$.

It can be shown that for each y, $\tilde{F}_n(y)$ has no greater variance than the CMC estimator $\hat{F}_n(y)$ in Eq. 33. Note that

$$\begin{aligned} Var[\tilde{F}_n(y)] &= \left(\frac{1}{n/2}\right)^2 \sum_{i=1}^{n/2} \frac{1}{4} Var[I(Y_i \leq y) + I(Y_{i'} \leq y)] \\ &= \left(\frac{1}{n/2}\right) \frac{1}{4} (Var[I(Y \leq y)] + Var[I(Y' \leq y)] + 2Cov[I(Y \leq y), I(Y' \leq y)]) \\ &= \left(\frac{1}{n/2}\right) \frac{1}{2} (F(y)(1 - F(y)) + Cov[I(Y \leq y), I(Y' \leq y)]) \quad \text{Eq. 75} \\ &\leq \frac{F(y)(1 - F(y))}{n} = Var[\hat{F}_n(y)], \quad \text{Eq. 76} \end{aligned}$$

since the negative correlation of Y_i and $Y_{i'}$ implies the same for $I(Y \leq y)$ and $I(Y' \leq y)$ because $g(x) = I(x \leq y)$ is monotonic in x (p. 181 of [102]). Thus, AV can reduce the variance of the estimator of $F(y)$ compared to CMC, which leads to the AV p-quantile estimator $\tilde{\xi}_{p,n}$ having smaller variance.

Avramidis and Wilson [103] develop this AV estimator of ξ_p, which they prove satisfies the CLT in Eq. 66, but they do not consider the estimation of the asymptotic variance κ_p^2 in the CLT to construct a CI for ξ_p. To address this issue, Chu and Nakayama [72] prove that the AV CDF estimator $\tilde{F}_n$ satisfies their Assumptions A1, A3 and the stronger version of A2 without any extra conditions required. Thus, if $f(\xi_p) > 0$, then the

AV quantile estimator satisfies the Bahadur representation in Eq. 64 and Eq. 65 with $\dot{\xi}_{p_n}$ in Eq. 49 for $p_n = p + O(1/\sqrt{n})$. Moreover, Eq. 64 and Eq. 65 hold with $\dot{\xi}_{p_n}$ in Eq. 50 for any $p_n \to p$ when f is also continuous in a neighborhood of ξ_p. In either case, this implies the AV p-quantile estimator $\tilde{\xi}_{p,n}$ satisfies the CLT in Eq. 66. To derive an expression for ψ_p^2 in Eq. 68 when applying AV, which is needed to determine κ_p in Eq. 67, note that by Eq. 74, $\tilde{F}_n(\xi_p)$ is the sample average of $n/2$ quantities $Z_i \equiv [I(Y_i \le \xi_p) + I(Y_{i'} \le \xi_p)]/2$. Since the Z_i, $i = 1,2,\dots,n/2$, are i.i.d. with finite variance, the CLT in Eq. 68 holds with $\psi_p^2 = 2Var[Z_i]$. Hence, it follows from Eq. 75 that

$$\begin{aligned}\psi_p^2 &= Var[I(Y \le \xi_p)] + Cov[I(Y \le \xi_p), I(Y' \le \xi_p)] \\ &= p(1-p) + E[I(Y \le \xi_p)\, I(Y' \le \xi_p)] - E[I(Y \le \xi_p)]\, E[I(Y' \le \xi_p)] \\ &= p(1-2p) + P\{Y \le \xi_p, Y' \le \xi_p\}\end{aligned}$$

since $E[I(Y \le \xi_p)] = E[I(Y' \le \xi_p)] = p$. Chu and Nakayama[72] show that

$$\tilde{\psi}_{p,n}^2 = p(1-2p) + \frac{1}{n/2}\sum_{i=1}^{n/2} I(Y_i \le \tilde{\xi}_{p,n}, Y_{i'} \le \tilde{\xi}_{p,n}) \qquad \text{Eq. 77}$$

consistently estimates ψ_p^2, so $\tilde{\psi}_{p,n} \Rightarrow \psi_p$ as $n \to \infty$. Substituting Eq. 77 into Eq. 72 and Eq. 73 then results in asymptotically valid two-sided and one-sided $100(1-\alpha)\%$ CIs for ξ_p when applying AV.

Computing the AV quantile estimator and constructing the corresponding CI require inverting the AV CDF estimator in Eq. 74, which can be accomplished as follows. Define $A_{2i-1} = Y_i$ and $A_{2i} = Y_{i'}$ for $i = 1,2,\dots,n/2$. Let $A_{(1)} \le A_{(2)} \le \cdots \le A_{(n)}$ be the order

statistics of the $A_j, j = 1,2,\ldots,n$. Then for any $0 < q < 1$, it is possible to compute $\tilde{F}_n^{-1}(q) = A_{(\lceil nq \rceil)}$, where $\lceil \cdot \rceil$ denotes the round-up function.

To implement AV within a computer code, the procedure in Figure 4. 10 can be used. As with the asymptotic CMC code in Figure 4. 9, first the quantile estimation *Xi* is made using the ordered results of all *n* samples, called *Y_tot*. Then the CFD is calculated using all *n* samples. Next, the summation in Eq. 77 is found using MATLAB's built-in indicator function. This is used to find $\tilde{\psi}_{p,n}$, called *psi*, then the additional confidence term is added to the quantile estimation.

```
%%% Antithetic Variates
    Xi=Y_tot(n*p);                  % Quantile Estimation using all n samples
    CFDhigh=Y_tot(ceil((p+hn)*n));  % CFD High Point
    CFDlow=Y_tot(ceil((p-hn)*n));   % CFD Low Point
    CFD=(CFDhigh-CFDlow)/((ceil((p+hn)*n)-ceil((p-hn)*n))/n); %CFD
    prob=mean(Y <= Xi & YY <= Xi);  % Indicator Function sum in Eq. 77
    psi=sqrt(p*(1-2*p)+prob);       % Psi calculation
    add_on=NN*((psi*CFD)/sqrt(n));  % Confidence Term
    Xi_w_conf=Xi+add_on;            % Quantile Estimation plus Confidence
```

Figure 4. 10: MATLAB Code Implementation of AV Quantile Confidence Method

4.2.2.4. Latin Hypercube Sampling

As detailed in Section 3.1.2, Latin hypercube sampling (LHS) is an extension of stratified sampling (Chapter 5 of [104]) in multiple dimensions, and it induces correlations among the outputs, which can reduce variance. It has been used frequenctly in nuclear engineering [40], although not presently for the calculation of 95/95 values in uncertainty analyses of LOCAs. Avramidis and Wilson [103] develop LHS quantile

estimators, but they do not develop CIs based on the estimators. Nakayama [101] shows how the general framework of [72] applies to a type of replicated LHS (rLHS), thus allowing the construction of an asymptotically valid CI for a quantile when using rLHS.

Rather than generating a single LHS sample of size n, the basic idea of rLHS in [101] is to generate the $n = mt$ samples as m independent LHS samples, each of size t. For each independent LHS sample $k = 1,2,\dots,m$, let $U_{i,j}^{(k)}$, for $1 \le i \le t$ and $1 \le j \le d$, be td i.i.d. $uniform[0,1]$ random variables, which can be arranged as a $t \times d$ array

$$\begin{matrix} U_{1,1}^{(k)} & U_{1,2}^{(k)} & \cdots & U_{1,d}^{(k)} \\ U_{2,1}^{(k)} & U_{2,2}^{(k)} & \cdots & U_{2,d}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ U_{t,1}^{(k)} & U_{t,2}^{(k)} & \cdots & U_{t,d}^{(k)} \end{matrix}.$$

Then let $\pi_j^{(k)} = (\pi_j^{(k)}(i): i = 1,2,\dots,t)$ for $1 \le j \le d$ and $1 \le k \le m$ be dm independent permutations of $(1,2,\dots,t)$, which are also independent of the $U_{i,j}^{(k)}$. Thus, $\pi_j^{(k)}(i)$ is the value to which i is mapped in the permutation $\pi_j^{(k)}$. Then define

$$V_{i,j}^{(k)} = \frac{\pi_j^{(k)}(i) - 1 + U_{i,j}^{(k)}}{t} \quad \text{for} \quad 1 \le i \le t \quad \text{and} \quad 1 \le j \le d.$$

For each $1 \le k \le m$, arrange the $V_{i,j}^{(k)}$ into a $t \times d$ array

$$\begin{matrix} V_{1,1}^{(k)} & V_{1,2}^{(k)} & \cdots & V_{1,d}^{(k)} \\ V_{2,1}^{(k)} & V_{2,2}^{(k)} & \cdots & V_{2,d}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ V_{t,1}^{(k)} & V_{t,2}^{(k)} & \cdots & V_{t,d}^{(k)} \end{matrix}. \qquad \text{Eq. 78}$$

It is straightforward to show that each $V_{i,j}^{(k)}$ has a $uniform[0,1]$ distribution. Moreover, by the independence of the permutations $\pi_1^{(k)}, \pi_2^{(k)}, \dots, \pi_d^{(k)}$, the columns of Eq. 78 are independent. Thus, defining

$$\begin{aligned} Y_1^{(k)} &= g(V_{1,1}^{(k)}, V_{1,2}^{(k)}, \ldots, V_{1,d}^{(k)}), \\ Y_2^{(k)} &= g(V_{2,1}^{(k)}, V_{2,2}^{(k)}, \ldots, V_{2,d}^{(k)}), \\ &\vdots \\ Y_t^{(k)} &= g(V_{t,1}^{(k)}, V_{t,2}^{(k)}, \ldots, V_{t,d}^{(k)}), \end{aligned} \quad \text{Eq. 79}$$

and each $Y_i^{(k)}$ has CDF F by Eq. 44. But the rows in Eq. 78 are dependent because all of the entries in column j depend on the same permutation $\pi_j^{(k)}$, so $Y_1^{(k)}, Y_2^{(k)}, \ldots, Y_t^{(k)}$ are dependent. Let $Y_1^{(k)}, Y_2^{(k)}, \ldots, Y_t^{(k)}$ be an *LHS case* of run size t. Now replicating this procedure m independent times leads to

$$\begin{matrix} Y_1^{(1)} & Y_1^{(2)} & \cdots & Y_1^{(m)} \\ Y_2^{(1)} & Y_2^{(2)} & \cdots & Y_2^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ Y_t^{(1)} & Y_t^{(2)} & \cdots & Y_t^{(m)} \end{matrix}, \quad \text{Eq. 80}$$

where each column in Eq. 80 corresponds to one LHS case of run size t, as in Eq. 79. The columns in Eq. 80 are independent since the m LHS cases are generated independently, but the entries within a column are dependent since they are generated using LHS. The $n = mt$ values in Eq. 80 are an rLHS sample with m cases, each with run size t.

The rLHS estimator of the CDF F is then

$$\tilde{F}_{m,t}(y) = \frac{1}{mt} \sum_{k=1}^{m} \sum_{i=1}^{t} I(Y_i^{(k)} \le y), \quad \text{Eq. 81}$$

and the rLHS p-quantile estimator is $\tilde{\xi}_{p,m,t} = \tilde{F}_{m,t}^{-1}(p)$. Nakayama [101] proves that if $f(\xi_p) > 0$, then the following Bahadur representation holds:

$$\tilde{\xi}_{p_m,m,t} = \dot{\xi}_{p_m} + \frac{p - \tilde{F}_{m,t}(\xi_p)}{f(\xi_p)} + R_{m,t} \quad with \quad \sqrt{m} R_{m,t} \Rightarrow 0 \quad as \quad m \to \infty \quad \text{with } t \text{ fixed,} \quad \text{Eq. 82}$$

where $\dot{\xi}_{p_m} = \xi_p + (p_m - p)/f(\xi_p)$ when $p_m = p + O(1/\sqrt{m})$. If in addition f is continuous is a neighborhood of ξ_p, then Eq. 82 holds with $\dot{\xi}_{p_m} = F^{-1}(p_m)$ for all $p_m \to p$ as $m \to \infty$

(these results are established in [101] by proving that Assumptions A1, A3 and the stronger version of A2 from [72] hold for LHS). It then follows that $\tilde{\xi}_{p,m,t}$ satisfies the CLT

$$\frac{\sqrt{m}}{\kappa_p}\left(\tilde{\xi}_{p,m,t} - \xi_p\right) \Rightarrow N(0,1) \quad as \quad m \to \infty \quad \text{with } t \text{ fixed,} \qquad \text{Eq. 83}$$

where κ_p has the form in Eq. 67. To construct a CI for ξ_p based on Eq. 83, an estimator for $\kappa_p = \psi_p \lambda_p$ must be developed.

As before, the Bahadur representation in Eq. 82 allows a consistent estimator for λ_p to be developed, which is needed to construct a CI for ξ_p based on the CLT in Eq. 83. If $f(\xi_p) > 0$, then

$$\tilde{\lambda}_{p,m,t} = \frac{\tilde{F}_{m,t}^{-1}(p + h_m) - \tilde{F}_{m,t}^{-1}(p - h_m)}{2h_m} \qquad \text{Eq. 84}$$

satisfies $\tilde{\lambda}_{p,m,t} \Rightarrow \lambda_p$ as $m \to \infty$ with t fixed when $h_m = c/\sqrt{m}$ for any constant $c > 0$. If f is also continuous in a neighborhood of ξ_p, then $\tilde{\lambda}_{p,m,t} \Rightarrow \lambda_p$ as $m \to \infty$ for fixed t for any $h_m \neq 0$ satisfying $h_m \to 0$ and $\sqrt{m}h_m \to b$ for some $b \in (0, \infty]$ as $m \to \infty$.

To derive an expression for ψ_p^2 in Eq. 68, note that $\tilde{F}_{m,t}(y) = \frac{1}{m}\sum_{k=1}^{m} W^{(k)}(y)$, where

$$W^{(k)}(y) = \frac{1}{t}\sum_{i=1}^{t} I(Y_i^{(k)} \leq y).$$

Now $W^{(1)}(\xi_p), W^{(2)}(\xi_p), \ldots, W^{(m)}(\xi_p)$ are i.i.d. with finite variance since $0 \leq W^{(k)}(\xi_p) \leq 1$. Thus, $\tilde{F}_{m,t}(\xi_p)$ satisfies the CLT in Eq. 68 with

$$\psi_p^2 = Var[W^{(k)}(\xi_p)].$$

Nakayama [101] develops

$$\tilde{\psi}^2_{p,m,t} = \frac{1}{m-1}\sum_{k=1}^{m}[W^{(k)}(\tilde{\xi}_{p,m,t}) - \bar{W}_m]^2$$

as a consistent estimator (as $m \to \infty$ with t fixed) of ψ_p^2, where

$$\bar{W}_m = \frac{1}{m}\sum_{k=1}^{m} W^{(k)}(\tilde{\xi}_{p,m,t}).$$

Substituting $\tilde{\xi}_{p,m,t}$, $\tilde{\lambda}_{p,m,t}$, and $\tilde{\psi}_{p,m,t}$ for $\tilde{\xi}_{p,n}$, $\tilde{\lambda}_{p,n}$, and $\tilde{\psi}_{p,n}$, respectively, in Eq. 72 and Eq. 73 then result in asymptotically valid two-sided and one-sided $100(1-\alpha)\%$ CIs for ξ_p when applying rLHS.

Constructing the CIs requires inverting the rLHS CDF estimator in Eq. 81, which can be done as follows. Define $B_{(k-1)m+i} = Y_i^{(k)}$ for $i = 1,2,\dots,t$, and $k = 1,2,\dots,m$. Let $B_{(1)} \le B_{(2)} \le \cdots \le B_{(mt)}$ be the order statistics of the B_j, $j = 1,2,\dots,mt$. Then for any $0 < q < 1$, then the q-quantile value is $\tilde{F}^{-1}_{m,t}(q) = B_{(mtq)}$.

This technique can be implemented in a computer code using Figure 4. 11. Once again, *Y_tot* is the ordered results from all *m* cases, which is used to find the quantile estimation *Xi*. Then the CFD is calculated using all $n = mt$ samples. Next, the individual values of *W* for each case are found using MATLAB's built-in indicator function, and the ordered results from each LHS case, which are called *y*. The mean of these values for *W* is used to find $\bar{W}_m$, called *W_bar*. This is used in the calculation of $\tilde{\psi}_{p,m,t}$, called *psi*. This is than used to calculate the addition confidence term which is added to the quantile estimation to result in a one-side CI.

```
%%% rLHS
    Xi=Y_tot(ceil(p*m*t));            % Quantile Estimation
    CFDhigh=Y_tot(ceil((p+hm)*m*t)); % CFD High Point
    CFDlow=Y_tot(ceil((p-hm)*m*t));  % CFD Low Point
    CFD=(CFDhigh-CFDlow)/((ceil((p+hm)*m*t)-ceil((p-hm)*m*t))/(m*t));  %CFD
    for num=1:m                       % Loop through Cases
        Wmk(num)=mean(y(num,:)<=Xi); % Indicator Function Summation for W
    end
    W_bar=mean(Wmk);                  % W_bar Calculation
    psi=sqrt((1/(m-1))*sum((W_bar-Wm).^2));  % Psi calculation
    add_on=NN*((psi*CFD)/sqrt(m));    % Confidence Term
    Xi_w_conf=Xi+add_on;              % Quantile Estimation plus Confidence
```

Figure 4. 11: MATLAB Code Implementation of LHS Quantile Confidence Method

There is a tradeoff between the amount of variance reduction from rLHS and the rate of the convergence of the confidence interval's coverage. If an analyst takes many cases of a small run size, meaning large m but small t, then the asymptotics will converge more quickly, since the large m will help satisfy the CLT in Eq. 83. But the small run sizes will not reduce the quantile estimator's variance by as much as large run sizes would. As the run size t increases, the quantile estimator will have lower variance, but to remain at the same number n of total runs, the number m of cases must be reduced, and the coverage can suffer.

4.2.2.5. Derivative Estimation and Bandwidth

Section 4.2.2.1-4.2.2.4 considered central finite-difference (CFD) estimators Eq. 55, Eq. 69, and Eq. 84 to estimate the derivative λ_p in Eq. 53. This method of derivative estimation is very similar to the brute force method of sensitivity analysis described in Section 2.3.2.1. Implementing these estimators in practice requires the user to specify the bandwidth h_n (or h_m), and the particular choice for h_n can have a large impact on the

quality of the estimators. Previous work with asymptotic CMC provides some guidance on the choice for h_n. For example, [98] and [99] show that under certain conditions, taking $h_n = c_1 n^{-1/5}$ for some constant c_1 asymptotically minimizes the mean-square error of the CFD estimator of λ_p. Also, the coverage error of CIs can be asymptotically minimized by taking $h_n = c_2 n^{-1/3}$ for some constant c_2; see [105]. The values of c_1 and c_2 depend on the CDF F and p, and these papers provide data-based methods for estimating c_1 and c_2.

The CFD estimators in Eq. 55, Eq. 69, and Eq. 84 are each symmetric in the sense that the inverse of the estimated CDF is evaluated at perturbed values that are symmetric about p. However, the symmetric CFD estimator often overestimates λ_p when $p \approx 1$, as in the case of the 0.95-quantile. To see why, suppose that the CDF $F(y)$ has a density $f(y)$ that is differentiable and strictly decreasing for all y sufficiently large (this is true for many common distributions, including the normal, lognormal, gamma and Weibull). Thus, the density's derivative $f'(y) < 0$ for all sufficiently large y. Then defining $Q(p) = F^{-1}(p)$ as the quantile function, its first derivative $Q'(p) = \lambda_p = 1/f(\xi_p) > 0$ and its second derivative $Q''(p) = -f'(\xi_p)/f^3(\xi_p) > 0$ for $p \approx 1$. Hence, as shown in Figure 4. 12, $Q(p)$ is increasing and convex for $p \approx 1$. This leads to the symmetric CFD typically overestimating λ_p because $Q(p)$, whose derivative $Q'(p) = \lambda_p$ is being estimated with the finite difference, has slope that is increasing as p increases. Similarly, a backward finite-difference (BFD) estimator such as $[\tilde{F}_n^{-1}(p) - \tilde{F}_n^{-1}(p - h_n)]/h_n$ will often underestimate λ_p, as shown in Figure 4. 13.

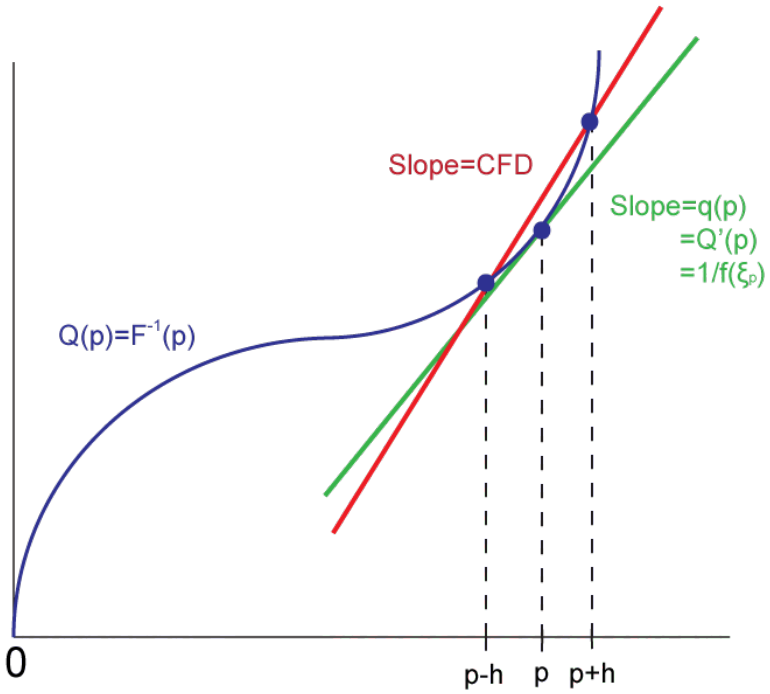

Figure 4. 12: Overprediction of Derivative using CFD

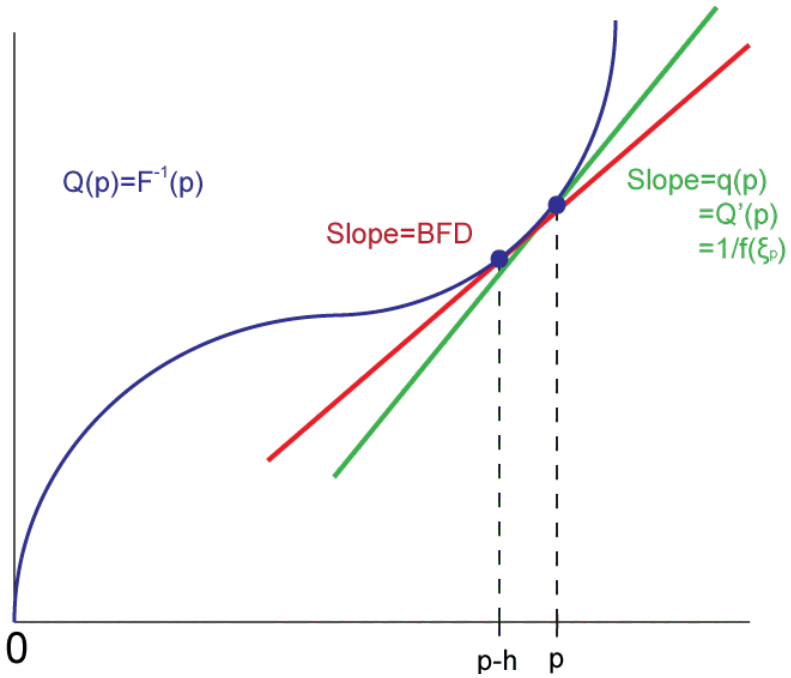

Figure 4. 13: Underprediction of Derivative using BFD

This suggests that it may be more accurate to estimate λ_p by using an *asymmetric* CFD estimator $[\tilde{F}_n^{-1}(p+h_n) - \tilde{F}_n^{-1}(p-h_n')]/(h_n+h_n')$, where $h_n' \neq h_n$. Figure 4. 12 shows that choosing $h_n' > h_n > 0$ may be beneficial because the slope of $Q(p)$ increases as p grows. Experiments were carried out using these asymmetric CFD esimators, and more detail is available in Appendix C (other estimators of λ_p are also possible; for example, Falk [106] develops a kernel estimator of λ_p for CMC, and Nakayama [107] considers another type of kernel estimator when using importance sampling).

4.3 Experiments

Nakayama demonstrated the method for estimating the confidence interval in Eq. 72 and Eq. 73 on a small stochastic activity network (SAN) [72], [101]. However, since the present work focuses on the use of these methods in nuclear safety analysis, the systems detailed in Section 3.2 were used to compare the techniques detailed in Section 4.2 since they would more closely mimic common safety analysis situations. This included starting with the simple nonlinear equation (Section 4.3.1), moving to a response-surface surrogate for the thermal-hydraulic computer code RELAP5 (Section 4.3.2), conducting a PRA involving the comparison of beyond-design-basis accidents to a risk limit curve (Section 4.3.3), and finally, using a large severe-accident analysis computer code (Section 4.3.4). The results presented here will focus on the comparison between CMC-OS and the asymptotic methods using a symmetric CFD for the derivative estimator. More information on the experiments with an asymmetric CFD can be found in Appendix C.

4.3.1. Nonlinear Equation

First, the nonlinear equation detailed in Section 3.2.2 was used to compare the methods of establishing confidence intervals for quantiles. This included using only normally distributed inputs, then non-normal inputs.

4.3.1.1. Normal Inputs

As in Section 3.2.2, a large CMC experiment with 10^8 runs was conducted in order to estimate the true 0.75- and 0.95-quantile of the system. The 0.95-quantile was chosen in order to see how these methods would perform when trying to satisfy the 95/95 criterion. The 0.75-quantile was chosen in order to test the applicability to possible future, less stringent, requirements, such as a 95/75. The result was a 0.95-quantile of 40.6457, which would be considered the "true" 0.95-quantile, and a "true" 0.75-quantile of 29.3887.

These quantiles were found in order for the distance between the calculated 95/95 and 95/75 values and the "true" quantiles to be found. This distance would be considered a measurement of the accuracy of the 95/95 and 95/75 values. It is important to point out that poor accuracy, as defined here, does not mean that the 95/95 or 95/75 values are not valid, but that significant overestimation of the quantiles is not desired and could potentially lead to Type-II errors, or incorrect safety analysis decisions. Figure 4. 14 shows the output distribution of the system for a 10^5-run CMC trial, which is shown simply to give the reader an idea of the range of possible outputs. The output is fairly compact at lower levels, but does have a long tail at higher values, meaning the higher quantiles are separated by a larger margin than the low quantiles.

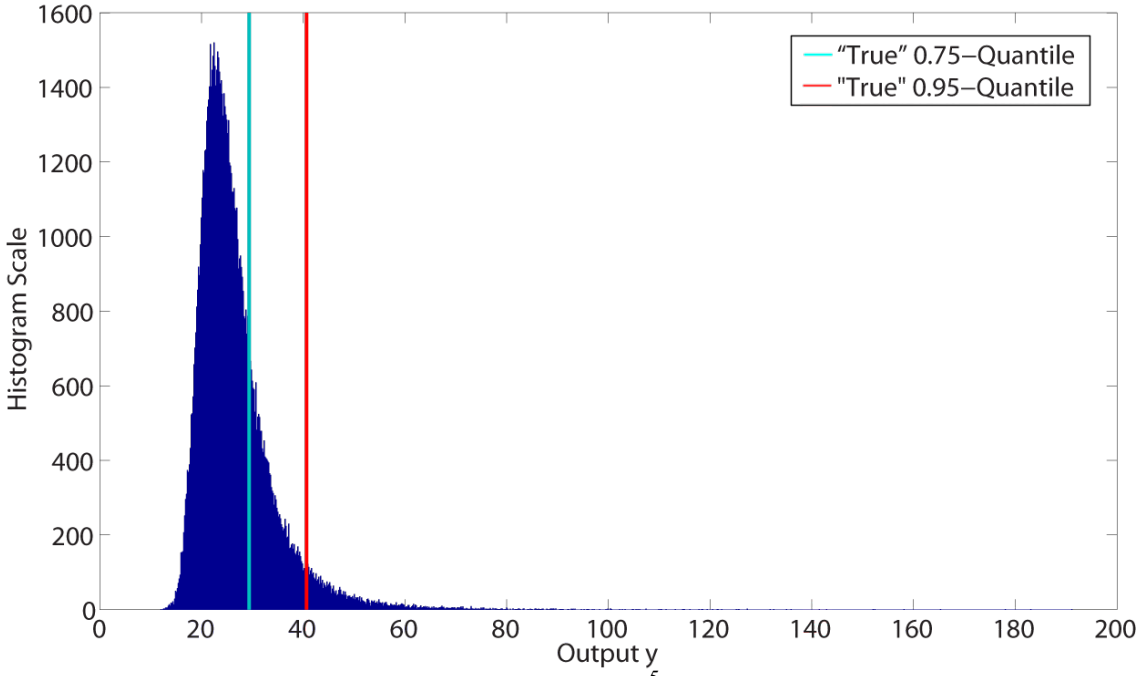

Figure 4. 14: Histogram of 10^5 Run CMC Trial

Each method to calculate a 95/95 and 95/75 value was repeated for 10^4 trials. Here, a trial is a complete experiment that would be undertaken during a safety analysis. For example, one CMC-OS trial may consist of 59 computer code runs to find a single 95/95 value. For each method, 10^4 trials were conducted so that the spread of possible 95/95 and 95/75 values could be found. This gives information about the precision of each method. Once again, poor precision does not mean that the 95/95 or 95/75 values are invalid, but a technique that provides these values over a large range is undesirable, especially if decisions are to be made about the system based on the results.

Each method was tested at several different total run levels. These run levels are based off the results for n in Eq. 43 for CMC-OS. They start with the lowest possible run

level using CMC-OS ($r = 1$), and increase from there. For rLHS, several representative values were chosen for the run size *t* for which $mt \approx n$. This gives information about the tradeoff between the size of *m* and *t* as described at the end of Section 4.2.2.4.

As described in Section 4.2.2.5, the derivative λ_p in Eq. 53 is estimated using a CFD estimator, and determining the proper bandwidth h_m is not trivial. Small changes in the bandwidth parameters *c* and *v* in $h_m = cn^{-v}$ can greatly impact the calculation of the CI. Also, while there has some been some guidance provided on the selection of these parameters when using the CFD to establish asymptotic CIs during a CMC simulation, there is less direction when using VRTs since asymptotic CIs have only recently been proven. Nakayama [101] provides some insight into the selection of *v* from the SAN example, which appeared to show $v = 1/2$ was more efficient for estimating quantiles close to 1. So for the experiments presented here, $v = 1/2$ was used for all run levels. Small experiments were conducted with *v* set to other values, but the results are not presented here.

Table 4. 3 notes the values chosen for the constant *c*. For each run level, asymptotic CMC, AV, and rLHS all used the same values for *c* and *v*. It is unlikely that there is one set of optimal values for all three simulation methods, but the values were kept constant so that a consistent comparison could be made between the methods. The value for *c* is smaller at lower run levels to ensure that in the CFD in Eq. 55, the inverse of the CDF estimator is evaluated at a point strictly less than 1. As the number of runs increases, the value for *c* can grow without resulting in a value of the inverse CDF estimator exceeding 1. Different values for *c* were used depending whether the 0.75- or

0.95-quantile was being estimated. Since the 0.75-quantile is further from the top value of the inverse CDF at 1, a wider bandwidth could be used. At low run levels, the selected value for *c* can cause completely different qualitative results. This is due to the coarseness of the estimated CDF, since it is constructed with relatively few samples. So a small change in the value of *c* can mean the values selected from the inverse CDF for the CFD estimator could differ by a wide margin. As the number of runs grows large (> 500), the selection of *c* becomes less impactful (although not trivial) since the estimated CDF is more developed. While both *c* and *v* could have been picked individually for each system tested, the goal is to determine values which are applicable to many systems, and that are not problem-specific.

Table 4. 3: Values for *c* in Bandwidth h_m (or h_n) $= cn^{-v}$ for Varying Run Sizes *n*

n	***c***	***p***
59	0.3	
93	0.3	
124	0.4	
311	0.5	0.95
548	0.5	
1008	0.5	
2004	0.5	
11	0.8	
29	0.8	
40	1.0	
135	1.25	0.75
246	1.25	
459	1.75	
886	1.75	

For each experiment, a detailed look at particular scenarios is presented first, followed by the complete numerical results.

As stated, the first run level conducted was based on the lowest value for which a 95/95 value could be found using CMC-OS. Figure 4. 15 shows a comparison of the histograms of the 10^4 95/95 values for CMC-OS at 59 runs and rLHS at 60 runs (m=6, t=10). The numerical results are in Table 4. 4. It is important to note that these are not histograms of the output distribution of the system, but rather for the 95/95 values for 10^4 complete trials. This means that 59 CMC-OS runs were conducted for each trial, and each trial resulted in one 95/95 value. As these results show, the rLHS method was not only more accurate, with a mean of 50.23 compared to a mean of 57.79 using CMC-OS, but also more precise, with about half the standard deviation. Both methods had ~5% of trials fall below the "true" quantile (called % Below "true" in Table 4. 4), which is to be expected with a 95% confidence interval. These results mean that the rLHS method is less likely to cause both Type-I and Type-II testing errors.

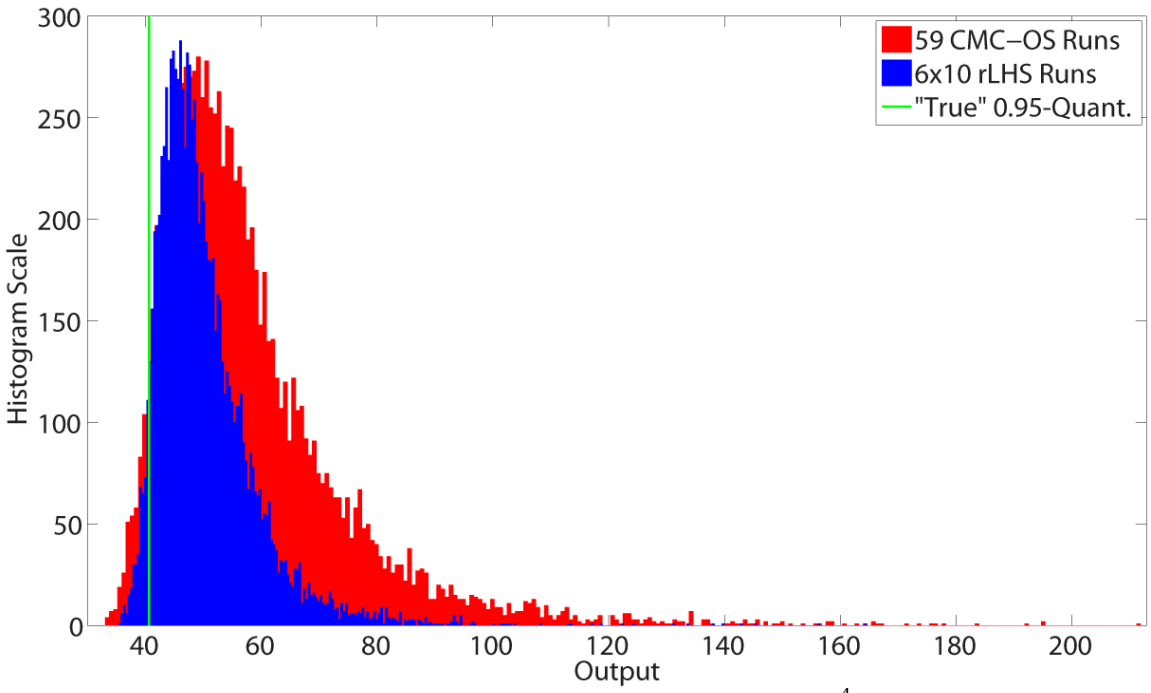

Figure 4. 15: Comparison of 95/95 Value Histograms for 10^4 Trials – 59 Runs

Table 4. 4: Comparison of 95/95 Values for 10^4 Trials – 59 Runs

	6x10 rLHS	59 CMC-OS
Mean of 10^4 95/95 Values	50.23	57.79
S.D. of 10^4 95/95 Values	8.56	16.52
% Below "true"	5.40%	5.10%

Figure 4. 16 and Table 4. 5 show the same results, but now for 124-run CMC-OS trials and 120-run rLHS trials (m=12, t=10). The scale on Figure 4. 16 is kept the same as in Figure 4. 15 to show the reduction in variance that naturally occurs with increased run size. The trend of rLHS being both more accurate and more precise continues at this higher run level.

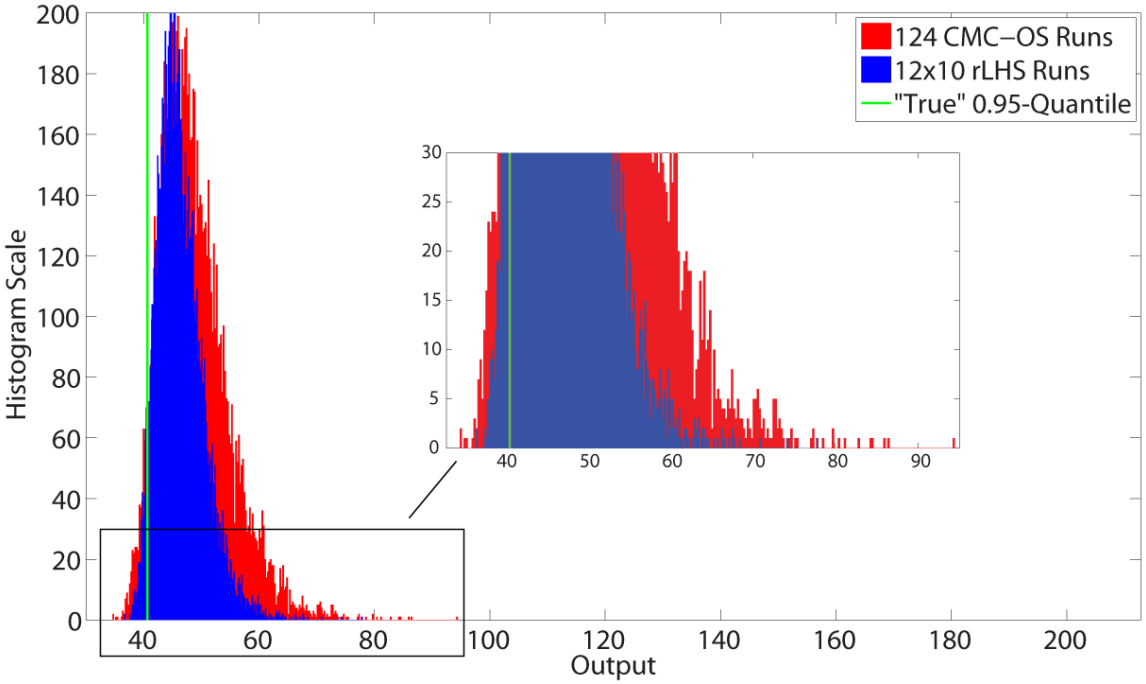

Figure 4. 16: Comparison of 95/95 Value Histograms for 10^4 Trials – 124 Runs

Table 4. 5: Comparison of 95/95 Values for 10^4 Trials – 124 Runs

	12x10 rLHS	124 CMC-OS
Mean of 10^4 95/95 Values	46.38	48.80
S.D. of 10^4 95/95 Values	4.06	6.14
% Below "true"	4.15%	4.90%

Before viewing the complete numerical results, there are several important points to note; the first being the number of runs conducted. Since the number *n* of runs conducted was based on levels for CMC-OS, it may not have been possible for AV (which needs an even number of runs) and rLHS (which used several different values for *t* in this work) to achieve that exact number. Therefore, run values for AV and rLHS were chosen as close as possible. For rLHS, this was done by dividing the number of runs

necessary for CMC-OS by the number chosen for *t*, then taking the closest whole number for the value for *m*. For example, if 59 CMC-OS runs were conducted and for rLHS $t = 10$, $59/10 = 5.9$, so $m = 6$ resulting in 60 total runs. Once again, representative values were chosen for *t*, but these are not the only options. Certain run levels were not conducted for some sizes of *t*, since the value for *m* would have been equal to two, and too low to properly satisfy the CLT. That is why there are several blank areas on the tables at low run levels. An exception was made at the lowest run level when finding a 95/75 ($n = 11$) to provide a comparison to CMC-OS, as will be seen.

Next, five metrics are provided for the numerical results of the asymptotic CMC, AV, and rLHS methods. These include:

1. The mean of the 95/95 or 95/75 values over all trials
2. The standard deviation of the 95/95 or 95/75 values over all trials
3. The percent of trials that fell below the "true" quantile (this is expected to be ~5%)
4. The coverage, or the percent of trials where the "true" quantile falls within the constructed 90% two-sided confidence interval for ξ_p (this confidence interval is defined in Eq. 54 for CMC, and Eq. 72 for AV and rLHS). Its expected value is ~90%.
5. The average value for the derivative estimator $\hat{\lambda}_p$ or $\tilde{\lambda}_p$ over all trials

Only the first three values are given for CMC-OS since no derivative estimation is necessary, and only a one-sided CI was found (although two-sided CIs are also possible).

Lastly, all tables include a comparison between asymptotic CMC, AV, and rLHS using a central finite-difference estimator for the derivative λ_p and an exact value for λ_p, which was calculated numerically using a CFD and a large-run CMC trial. This is presented in order to measure the effect of having to estimate the derivative λ_p.

The complete results for the experiment with the nonlinear equation using normal inputs, when calculating 95/95 values, can be found in Table 4. 6.

Table 4. 6: 95/95 Results 10^4 Trials – Nonlinear Eq. Normal Inputs

		CFD for λ_p					**Exact λ_p**				
n	**CMC-OS**	**CMC**	**AV**	**rLHS**			**CMC**	**AV**	**rLHS**		
				***t*=10**	***t*=20**	***t*=30**			***t*=10**	***t*=20**	***t*=30**
	57.79	54.28	52.23	50.23	48.56		49.60	47.22	45.99	44.80	**Mean***
	16.52	11.88	10.82	8.56	7.43		5.33	4.01	3.16	3.05	**S.D. ***
59	5.10	5.36	7.75	5.40	5.40		0.95	1.67	2.25	6.41	**% Below**
		90.31	91.18	93.31	93.53		85.69	95.00	93.34	90.68	**Covg.**
		256.76	260.75	269.15	277.06		156.60	156.60	156.60	156.60	**Avg. $\tilde{\lambda}_p$**
	50.96	49.03	49.83	47.51	45.56	45.92	46.86	47.42	45.83	44.18	44.53
	8.45	6.69	6.84	5.10	4.32	3.92	3.74	3.68	2.89	2.65	2.63
93	5.48	6.59	4.63	4.58	8.51	4.63	2.07	1.06	1.25	8.22	5.65
		90.27	92.57	92.16	88.95	89.40	89.69	91.39	89.89	87.44	84.72
		215.05	218.02	213.45	213.03	218.92	156.60	156.60	156.60	156.60	156.60
	48.80	48.03	48.06	46.38	45.26	45.09	45.47	45.48	44.39	43.61	43.44
	6.14	5.39	5.45	4.06	3.38	3.33	3.03	2.99	2.24	2.11	2.15
124	4.90	5.07	4.86	4.15	5.17	5.35	3.81	2.97	2.90	7.79	9.55
		93.52	94.03	94.64	92.35	92.18	90.71	92.04	92.67	86.37	85.30
		236.28	236.08	233.71	236.20	238.26	156.60	156.60	156.60	156.60	156.60
	44.77	44.57	44.58	43.76	42.91	42.92	43.91	43.89	43.23	42.55	42.51
	2.78	2.59	2.56	1.95	1.47	1.44	1.97	1.94	1.50	1.24	1.26
311	4.88	4.76	4.45	3.63	5.37	4.81	3.40	3.08	2.58	5.80	6.72
		92.56	93.28	93.48	91.69	91.97	89.98	90.45	89.99	88.76	87.96
		189.40	190.79	190.81	186.29	189.13	156.60	156.60	156.60	156.60	156.60
	43.45	43.28	43.26	42.77	42.31	42.23	43.04	43.02	42.56	42.14	42.07
	1.89	1.82	1.76	1.34	1.08	1.02	1.49	1.43	1.10	0.95	0.93
548	5.62	6.02	5.73	4.30	5.34	5.40	4.36	3.98	3.17	5.26	6.03
		90.84	91.58	92.20	91.45	91.07	89.47	90.18	89.82	89.33	88.87
		172.86	173.02	174.75	175.02	173.91	156.60	156.60	156.60	156.60	156.60
	42.62	42.49	42.49	42.11	41.83	41.76	42.39	42.39	42.03	41.77	41.71
	1.27	1.26	1.22	0.95	0.77	0.73	1.08	1.04	0.82	0.70	0.67
1008	5.54	6.33	5.64	5.39	5.89	5.58	4.76	4.13	3.66	5.07	5.44
		90.19	91.12	90.67	90.18	90.35	89.75	90.75	89.89	89.35	89.29
		165.09	165.03	165.82	165.59	165.29	156.60	156.60	156.60	156.60	156.60
	42.03	41.93	41.91	41.62	41.48	41.49	41.89	41.87	41.59	41.45	41.46
	0.87	0.87	0.84	0.66	0.54	0.52	0.77	0.75	0.58	0.50	0.48
2004	4.95	6.36	6.05	6.15	5.75	4.60	5.08	4.44	4.69	4.98	4.31
		89.65	90.09	89.75	89.76	90.16	89.37	90.47	89.65	89.27	89.17
		161.98	162.28	161.92	162.65	161.87	156.60	156.60	156.60	156.60	156.60

* Mean and S.D. of the 10^4 95/95 Values

As the table shows, CMC-OS and asymptotic CMC converge to approximately the same solution as the number of runs grows large. This is to be expected, as the comment in Section 4.2.2.1 explains the relation between the two methods. Using an

exact value for λ_p at the lowest run level, asymptotic CMC has fairly poor coverage. This may seem odd since the only part of the formula left to estimate is the quantile. However, at this low run level, the quantile estimator may not have converged. This is compounded with the use of the round-up function, which has a larger impact at low run levels. So the coverage appears poor. When using asymptotic CMC with an estimated λ_p at this run level, the derivative is overestimated; this causes the CI to increase in size, so the effects of the poor quantile estimation are not as obvious.

The complete table of results also shows the tradeoff between run size and case number when using rLHS. The more cases, the quicker the convergence, since it is the number of cases which satisfies the CLT. However, the larger the run size, the more variance reduction will be seen when compared to CMC. As the table shows, the results when $t = 10$ tend to converge the fastest, but the accuracy and precision of the result improves when using $t = 20$ and $t = 30$. Lastly, AV does show variance reduction when compared to CMC and CMC-OS, but not to the extent of rLHS.

The following figures and tables show the comparisons when finding a 95/75 value. The lowest run level possible for CMC-OS is 11 runs. Figure 4. 17 and Table 4. 7 show the results for 11-run CMC-OS trials and 10-run rLHS trials (m=2, t=5).

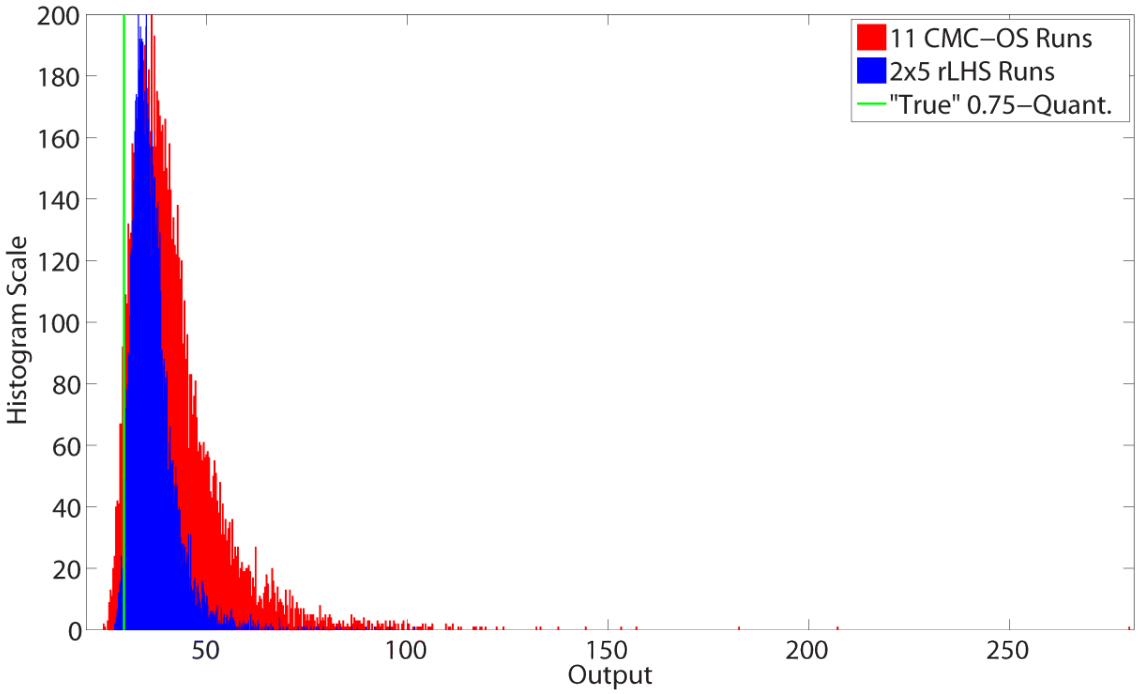

Figure 4. 17: Comparison of 95/75 Value Histograms for 10^4 Trials – 11 Runs

Table 4. 7: Comparison of 95/75 Values for 10^4 Trials – 11 Runs

	2x5 rLHS	11 CMC-OS
Mean of 10^4 95/75 Values	35.88	42.08
S.D. of 10^4 95/75 Values	4.63	12.02
% Below "true"	0.93%	4.22%

As to be expected, the variance of the resulting CMC-OS 95/75 is very large due to how few CMC runs are being conducted. While the rLHS method is much more accurate and precise, less than 1% of the runs fell below the "true" 0.75-quantile. This might seem like a positive at first, but it is a sign that the method has not converged to the proper coverage level. This is not surprising since the number of cases $m = 2$, and large m is

needed for the CLT to hold. More cases are necessary for the coverage of the rLHS method to converge.

Due to the high variance at 11 runs, it is likely the analyst would perform more runs to find a 95/75 value. Figure 4. 18 and Table 4. 8 show the results for 40-run CMC-OS trials and 40-run rLHS trials (m=8, t=5). Once again, the rLHS method outperforms the CMC-OS method. However, with m now equal to 8, the rLHS has ~4% of trials falling below the "true" quantile and is closer to convergence. Varying the values for c and v in the derivative estimator for rLHS may also improve convergence, but this would not be known in a real analysis since only one trial is conducted.

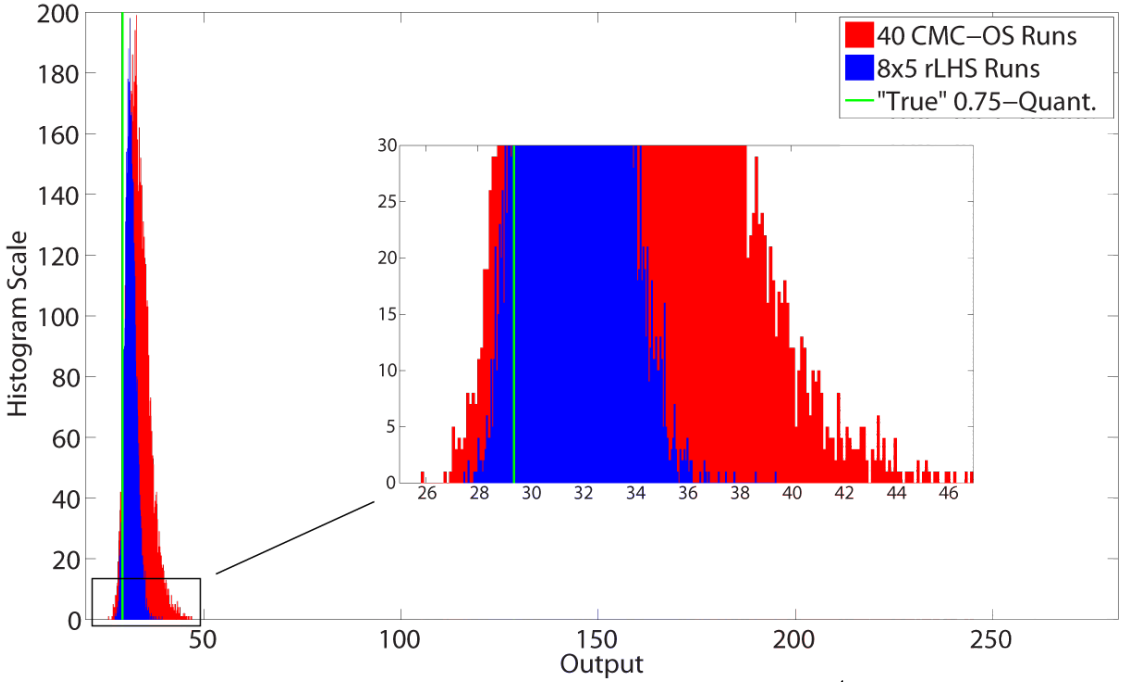

Figure 4. 18: Comparison of 95/75 Value Histograms for 10^4 Trials – 40 Runs

Table 4. 8: Comparison of 95/75 Values for 10^4 Trials – 40 Runs

	8x5 rLHS	40 CMC-OS
Mean of 10^4 95/75 Values	31.48	33.47
S.D. of 10^4 95/75 Values	1.33	2.74
% Below "true"	3.91%	4.23%

The complete numerical results for the 95/75 values can be found in Table 4. 9. As the table shows, the rLHS method has not converged at the $n = 11$ run level, since only two LHS cases are conducted. However, by the $n = 29$ run level, the rLHS method at $t = 5$ has ~5% of trials below the true quantile, and the coverage is almost 90%. So convergence appears to occur quickly. The coverage appears worse at the $n = 40$ level for rLHS, but this is probably a result of the change in the bandwidth values at that level. A look at the average value for $\tilde{\lambda}_p$ shows that the derivative estimation actually got worse at this level (from 30.48 to 32.93, with an actual value of 25.43). As the rest of the results show, this was the only run level, other than $n = 11$, in which the coverage was not approximately equal to 90%. Other results to note from the table are that AV once again reduces variance when compared to CMC-OS and asymptotic CMC, but not to the extent of rLHS. Also, asymptotic CMC shows less conservatism than CMC-OS at low run levels, but they converge to approximately the same values at high run levels, which is to be expected. Lastly, the tradeoff between run size and the number of cases when using rLHS continues to be present, as $t = 5$ converges faster, but $t = 10$ and $t = 15$ provide a greater variance reduction.

Table 4. 9: 95/75 Results 10^4 Trials – Nonlinear Eq. Normal Inputs

		CFD for λ_p					Exact λ_p				
n	CMC-OS	CMC	AV	rLHS			CMC	AV	rLHS		
				t=5	t=10	t=15			t=5	t=10	t=15
11	42.08	38.22	37.76	35.88			35.79	34.46	33.05		**Mean***
	12.02	7.44	7.10	4.63			3.71	2.42	2.04		**S.D.***
	4.22	5.38	3.68	0.93			0.95	0.01	1.64		**% Below**
		90.05	95.71	94.96			87.77	97.34	86.09		**Covg.**
		36.97	40.69	49.54			25.43	25.43	25.43		**Avg. $\tilde{\lambda}_p$**
29	34.44	33.65	33.48	31.94	31.65		32.62	32.94	31.57	31.38	
	3.54	3.07	2.77	1.67	1.59		2.01	1.71	1.34	1.40	
	4.37	5.45	3.63	4.51	5.77		3.56	0.19	4.54	6.76	
		92.47	94.76	90.44	85.66		91.20	94.76	87.44	82.29	
		33.22	29.51	30.48	29.59		25.43	25.43	25.43	25.43	
40	33.47	32.75	32.78	31.48	31.23	30.95	32.02	32.06	30.97	30.79	30.66
	2.74	2.37	2.16	1.33	1.33	1.45	1.70	1.40	1.09	1.15	1.27
	4.23	5.62	3.05	3.91	7.10	14.96	4.41	0.74	6.96	11.16	16.19
		92.79	96.24	94.09	89.18	72.13	91.28	96.12	88.92	83.35	69.98
		31.92	31.90	32.93	32.35	31.20	25.43	25.43	25.43	25.43	25.43
135	31.21	31.13	30.98	30.51	30.29	30.40	31.01	30.85	30.44	30.22	30.34
	1.17	1.13	0.94	0.65	0.59	0.61	0.96	0.76	0.61	0.56	0.58
	4.97	5.10	3.07	3.59	5.90	4.25	3.57	1.53	3.55	6.65	4.71
		90.71	95.57	90.57	90.04	88.47	89.80	95.50	88.79	88.20	86.43
		27.37	27.52	27.60	27.53	27.39	25.43	25.43	25.43	25.43	25.43
246	30.66	30.61	30.60	30.14	30.10	30.05	30.56	30.54	30.11	30.08	30.02
	0.81	0.79	0.68	0.47	0.44	0.45	0.70	0.57	0.44	0.42	0.43
	5.16	5.17	2.65	4.98	4.98	6.39	3.88	1.40	4.88	5.02	6.83
		90.89	95.53	90.50	89.63	89.25	89.98	95.25	89.48	88.61	88.13
		26.63	26.65	26.73	26.57	26.72	25.43	25.43	25.43	25.43	25.43
459	30.24	30.29	30.24	29.97	29.89	29.89	30.25	30.20	29.94	29.87	29.87
	0.57	0.57	0.47	0.34	0.31	0.31	0.52	0.42	0.32	0.30	0.30
	6.18	5.08	2.83	3.99	5.11	4.87	4.15	1.95	3.91	5.65	5.35
		91.05	95.72	91.34	91.15	90.73	89.99	95.37	90.08	89.66	89.11
		26.62	26.66	26.72	26.70	26.68	25.43	25.43	25.43	25.43	25.43
886	30.03	30.02	30.01	29.78	29.76	29.75	30.00	29.99	29.77	29.75	29.74
	0.39	0.39	0.33	0.24	0.22	0.22	0.37	0.30	0.23	0.22	0.22
	4.83	5.19	2.51	4.71	4.78	4.85	4.37	1.78	4.57	4.62	4.74
		90.50	95.61	91.00	90.42	90.24	90.11	95.41	90.58	89.98	89.57
		26.05	26.02	25.99	26.01	26.07	25.43	25.43	25.43	25.43	25.43

* Mean and S.D. of the 10^4 95/75 Values

4.3.1.2. Non-Normal Inputs

Next, the nonlinear equation experiment was repeated, but with the non-normal inputs described in Section 3.2.2. As before, a 10^8-run CMC experiment was conducted in order to determine the "true" quantiles. Here, the 0.95-quantile was found to be 106.727, and the 0.75-quantile was 34.216. Figure 4. 19 shows the output distribution for

a 10^5-run CMC trial. Once again, this is done to give the reader an idea of the output distribution shape. Compared to the previous example in Section 4.3.1.1, this output has a fatter right tail, which results in the 0.75- and 0.95-quantiles being much further apart.

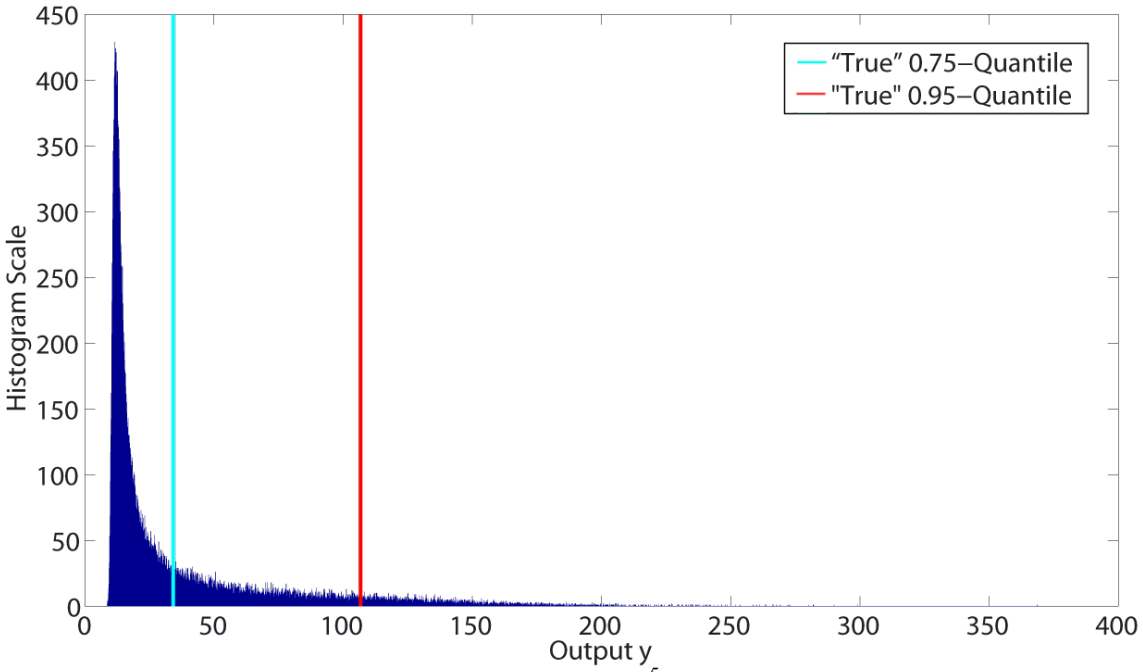

Figure 4. 19: Histogram of 10^5 Run CMC Trial

The same procedure as the previous example was followed, starting with a comparison between 10^4 trials of 59 CMC-OS runs and 60 rLHS runs (m=6, t=10). Figure 4. 20 and Table 4. 10 have these results. The trend continues with rLHS being more precise and accurate. However, at this level over 6% of rLHS trials fell below the "true" quantile. This could again be a sign that the asymptotics have not converged yet for proper coverage, or that the values for c and v are not appropriate. Nevertheless, even though the rLHS has a greater percentage of 95/95 values fall below the "true" quantile,

these values still do not fall as low as values when using CMC-OS. As mentioned in Section 4.1.2.2, the probability of committing a Type-I error is not only dependent on the percentage of trials that fall below the true quantile, but on the distance from those trials to the true quantile. So even though the rLHS method at this run level may not have converged to the proper coverage level, it is not possible to say whether the rLHS method would be more likely to experience a Type-I error than CMC-OS without knowing the placement of the limit value.

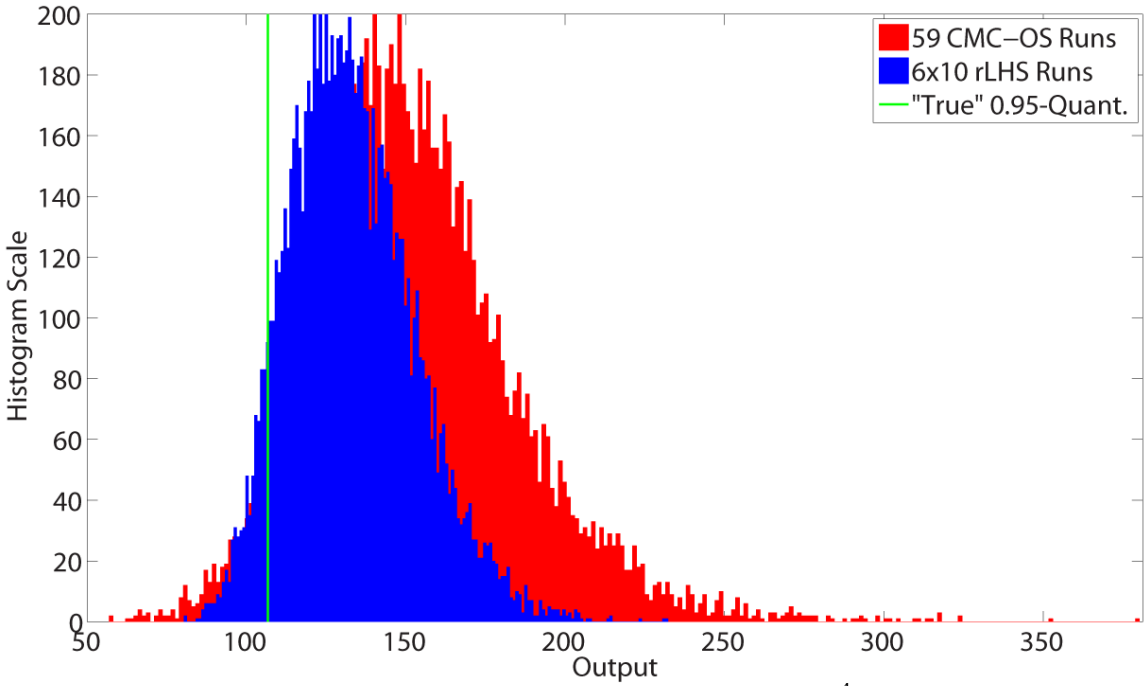

Figure 4. 20: Comparison of 95/95 Value Histograms for 10^4 Trials – 59 Runs

Table 4. 10: Comparison of 95/95 Values for 10^4 Trials – 59 Runs

	6x10 rLHS	59 CMC-OS
Mean of 10^4 95/95 Values	133.68	153.12
S.D. of 10^4 95/95 Values	19.90	31.63
% Below "true"	6.64%	4.82%

Figure 4. 21 and Table 4. 11 have the results for 124-run CMC-OS trials and 120-run rLHS trials (m=12, t=10). For rLHS, the percent below the "true" quantile is now closer to 5% and may be a sign that the coverage is converging properly.

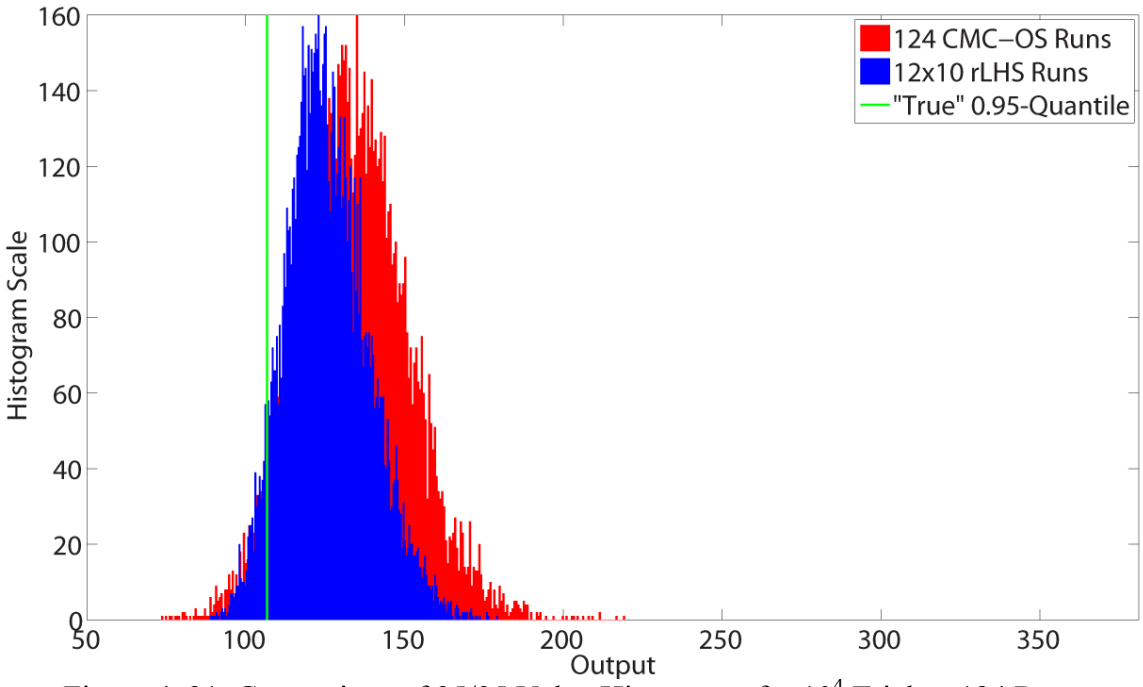

Figure 4. 21: Comparison of 95/95 Value Histograms for 10^4 Trials – 124 Runs

Table 4. 11: Comparison of 95/95 Values for 10^4 Trials – 124 Runs

	12x10 rLHS	124 CMC-OS
Mean of 10^4 95/95 Values	125.79	133.98
S.D. of 10^4 95/95 Values	13.01	17.25
% Below "true"	5.72%	5.01%

The complete numerical results are in Table 4. 12. As the table shows, the rLHS method using $t = 10$, while not converged at $n = 59$, appears to converge at the next

highest run level of $n = 93$. However, the coverage is worse at the $n = 124$ level. A closer inspection shows that, once again, the derivative estimation got worse with increasing run level (from 814.74 to 853.58, with an actual value of 745.73). This again is a result of changing the bandwidth parameters as the number of runs increases. This shows the impact that small changes in these parameters can have at low run levels.

Table 4. 12: 95/95 Results 10^4 Trials – Nonlinear Eq. Non-normal Inputs

		CFD for λ_p					**Exact λ_p**				
n	**CMC-OS**	**CMC**	**AV**	**rLHS**			**CMC**	**AV**	**rLHS**		
				***t*=10**	***t*=20**	***t*=30**			***t*=10**	***t*=20**	***t*=30**
	153.12	151.04	138.67	133.68	127.97		145.13	132.92	129.10	123.36	**Mean***
	31.63	28.53	25.00	19.90	16.59		21.11	19.31	13.94	12.93	**S.D.***
59	4.82	4.98	9.00	6.64	7.36		3.71	9.23	3.37	7.78	**% Below**
		89.13	88.54	91.57	91.22		89.34	89.12	94.02	90.11	**Covg.**
		872.37	873.46	877.69	911.32		745.73	745.73	745.73	745.73	**Avg. $\bar{\lambda}_p$**
	139.86	137.53	141.98	131.54	122.20	125.21	134.28	138.22	129.58	119.33	123.39
	21.43	20.61	20.96	16.71	13.58	13.22	16.81	16.14	12.49	12.90	11.23
93	5.01	6.01	3.73	5.02	10.68	5.82	5.25	2.52	1.80	16.12	5.24
		88.98	92.72	90.95	86.64	87.84	90.14	94.48	91.96	80.99	86.86
		833.13	836.20	814.74	823.99	839.44	745.73	745.73	745.73	696.73	745.73
	133.98	131.75	132.23	125.79	120.71	120.40	127.74	128.14	123.14	118.48	118.07
	17.25	16.46	16.49	13.01	11.25	11.05	14.22	14.17	10.41	9.83	9.62
124	5.01	6.18	5.79	5.72	9.61	8.99	7.36	6.75	4.27	11.57	11.44
		91.23	91.90	92.18	87.15	87.55	90.13	90.53	92.43	83.68	83.42
		870.07	871.86	853.58	875.24	878.60	745.73	745.73	745.73	745.73	745.73
	123.01	122.74	123.19	119.25	115.15	115.18	121.57	121.99	118.40	114.59	114.53
	9.82	9.88	9.80	7.81	6.17	6.16	9.20	9.03	6.81	5.71	5.79
311	4.85	5.26	4.28	4.57	7.69	7.85	5.29	4.89	3.62	7.78	8.52
		90.44	92.10	91.26	88.82	88.65	90.13	90.84	90.44	87.13	87.09
		803.44	803.77	802.32	800.32	805.29	745.73	745.73	745.73	745.73	745.73
	118.18	118.02	118.10	115.73	113.26	112.96	117.59	117.69	115.39	112.98	112.67
	7.27	7.37	7.24	5.72	4.61	4.46	7.02	6.79	5.13	4.30	4.20
548	5.74	6.17	5.65	5.33	7.39	7.54	6.22	5.07	4.03	7.23	7.54
		89.53	90.34	89.82	88.98	88.89	89.70	90.83	90.06	88.27	87.78
		773.57	773.07	776.79	781.79	781.79	745.73	745.73	745.73	745.73	745.73
	115.26	115.16	114.97	113.18	111.56	111.26	114.96	114.79	113.02	111.44	111.16
	5.34	5.40	5.27	4.10	3.30	3.22	5.19	5.04	3.75	3.09	3.04
1008	5.69	5.73	5.59	5.16	6.63	7.61	5.60	5.25	4.51	6.38	6.87
		89.40	89.70	90.19	89.78	88.14	89.75	90.18	89.98	89.21	88.51
		762.97	761.96	764.98	765.89	764.29	745.73	745.73	745.73	745.73	745.73
	112.77	112.56	112.40	111.03	110.19	110.26	112.45	112.31	110.96	110.13	110.20
	3.67	3.74	3.70	2.89	2.33	2.25	3.66	3.58	2.70	2.21	2.14
2004	4.99	5.93	6.19	6.58	6.70	5.59	5.88	5.98	5.66	5.92	5.11
		89.64	89.25	89.14	89.27	89.59	89.93	89.49	89.48	89.22	89.28
		759.30	757.59	756.36	758.79	760.86	745.73	745.73	745.73	745.73	745.73

* Mean and S.D. of the 10^4 95/95 Values

Also, the table shows that while the rLHS method with larger run sizes provides even more variance reduction, it takes much longer to converge in this case. The coverage level converges to ~90% fairly quickly, but the percent below the true quantile stays above 5% even at higher run levels.

Figure 4. 22, Figure 4. 23, Table 4. 13 and Table 4. 14 have the results for estimating 95/75 values using CMC-OS and rLHS at 11 runs and 886 runs. As with the previous example, the rLHS method is much more precise and accurate than the CMC-OS method. However, once again the rLHS method has less than 5% of trials below the "true" quantile at the lowest run level, which is a sign that the CLT asymptotics have not yet converged. By the large run level in Table 4. 14, the proper coverage level has been established, and even at this high run level, rLHS is still more accurate and precise.

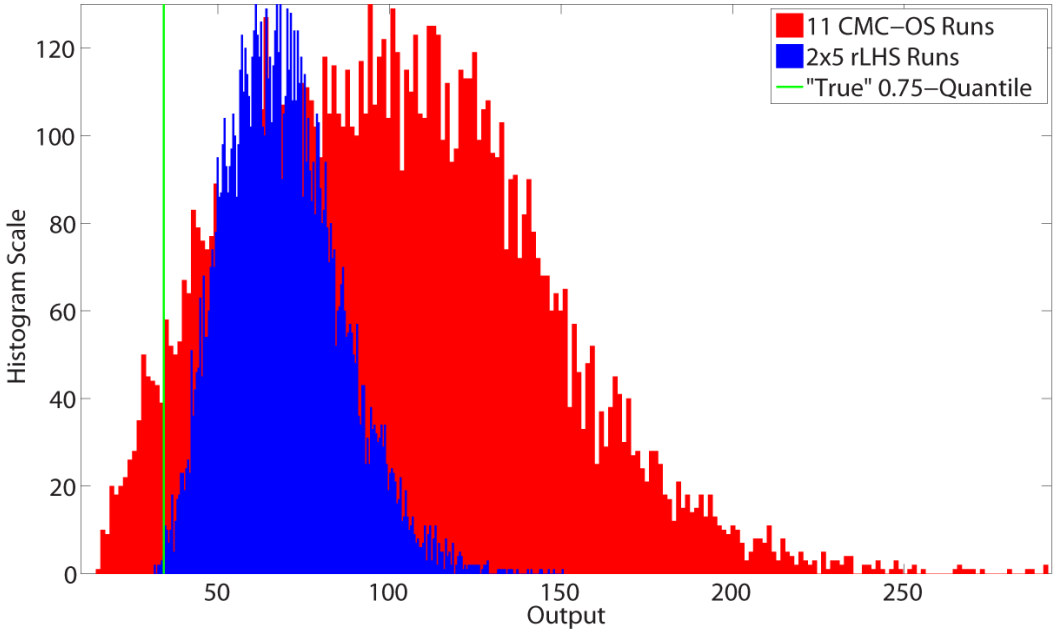

Figure 4. 22: Comparison of 95/75 Value Histograms for 10^4 Trials – 11 Runs

Table 4. 13: Comparison of 95/75 Values for 10^4 Trials – 11 Runs

	2x5 rLHS	11 CMC-OS
Mean of 10^4 95/75 Values	68.90	100.45
S.D. of 10^4 95/75 Values	16.50	42.36
% Below "true"	0.14%	4.18%

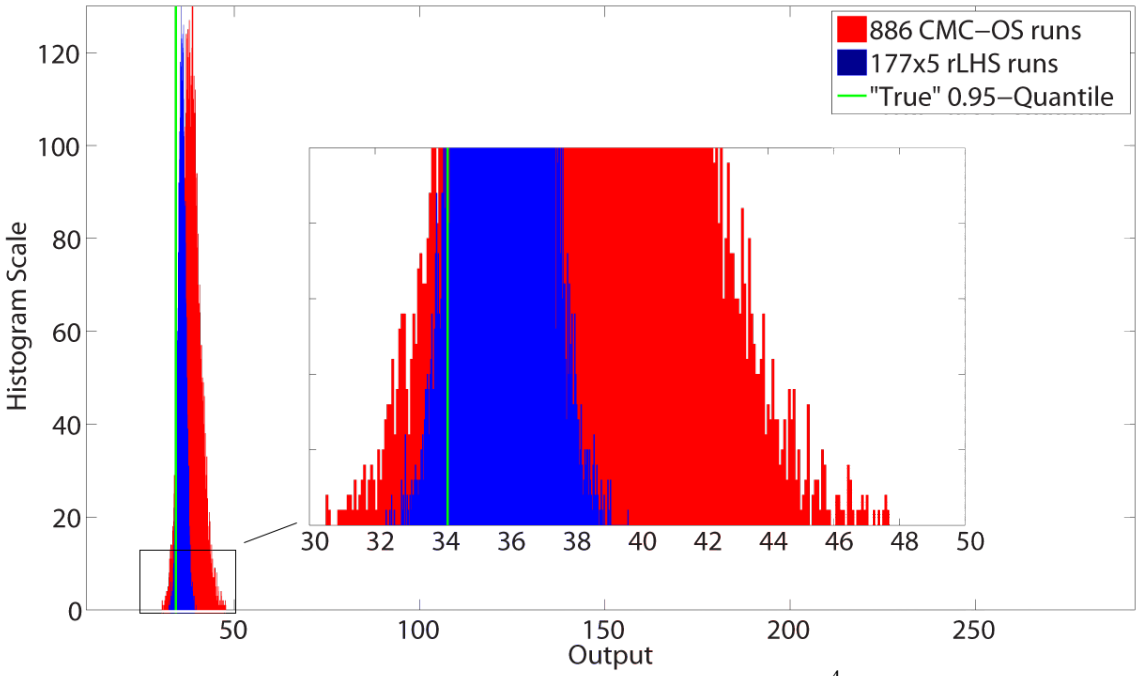

Figure 4. 23: Comparison of 95/75 Value Histograms for 10^4 Trials – 886 Runs

Table 4. 14: Comparison of 95/75 Values for 10^4 Trials – 886 Runs

	177x5 rLHS	886 CMC-OS
Mean of 10^4 95/75 Values	35.88	38.19
S.D. of 10^4 95/75 Values	1.00	2.51
% Below "true"	4.82%	4.98%

The complete results are in Table 4. 15. Some of the same issues from the normal input example continue here, with non-convergence at the lowest run level, and the

derivative estimation getting worse at $n = 40$, due to the bandwidth parameter changes. However, it also takes rLHS, with $t = 5$, longer to converge in this example than with normal inputs. A look at the results shows that even when using the exact value for the derivative, convergence does not occur until the run level is greater than 200. This means the issue is most likely due to the difficulty in estimating the quantile and not with the bandwidth parameters. However, even when the coverage is poor, rLHS always errors on the conservative side, meaning $< 5\%$ of trials fall below the true quantile, rather than $> 5\%$, which could increase the probability of a Type-I error.

Also, the table shows that the complexity of this equation, while increasing the difficulty of the quantile estimation, also benefits rLHS, since it provides a far more accurate and precise solution than CMC-OS, even before complete convergence. This means that even though rLHS is not providing the exact coverage level of 90%, it is still far less likely to result in a solution that may cause a Type-I or Type-II error than CMC-OS since it returns 95/75 values which more properly characterize the 0.75-quantile.

Table 4. 15: 95/75 Results 10^4 Trials – Nonlinear Eq. Non-normal Results

		CFD for λ_p					Exact λ_p				
n	**CMC-OS**	**CMC**	**AV**	**rLHS**			**CMC**	**AV**	**rLHS**		
				***t*=5**	***t*=10**	***t*=15**			***t*=5**	***t*=10**	***t*=15**
	100.45	81.39	64.12	68.90			75.40	58.59	58.71		**Mean***
	42.36	35.62	22.55	16.50			22.76	13.78	7.54		**S.D.***
11	4.18	6.71	6.76	0.14			0.01	0.01	0.01		**% Below**
		85.38	90.38	99.60			85.40	94.47	99.72		**Covg.**
		180.09	185.13	214.13			152.20	152.20	152.20		**Avg. $\tilde{\lambda}_p$**
	67.62	60.64	55.72	47.29	44.55		55.12	54.57	44.75	43.16	
	22.44	18.47	15.99	6.68	5.04		11.92	10.38	5.12	4.44	
29	4.73	5.63	5.23	0.85	0.64		0.02	0.01	0.69	0.56	
		91.75	90.58	97.20	96.07		92.74	91.89	97.01	95.34	
		193.96	161.83	188.67	176.72		152.20	152.20	152.20	152.20	
	61.48	55.08	52.32	44.49	42.01	38.34	51.07	48.16	41.67	39.76	37.52
	18.23	14.50	12.10	5.34	4.38	8.83	9.98	8.41	4.30	3.77	7.11
40	4.43	5.16	3.88	1.79	1.72	39.19	0.30	0.26	2.66	5.41	39.19
		92.62	94.07	97.68	97.11	16.95	93.62	93.23	96.37	92.10	16.95
		187.79	197.47	198.94	202.00	196.95	152.20	152.20	152.20	152.20	152.20
	46.34	45.21	42.52	39.29	37.78	38.18	44.38	41.77	38.80	37.43	37.88
	8.08	7.49	5.88	2.65	2.11	2.14	5.79	4.57	2.56	2.05	2.08
135	4.93	5.40	6.25	2.48	3.62	2.42	1.72	2.35	3.73	4.96	3.30
		90.45	90.97	93.57	93.95	89.78	90.29	91.40	91.32	92.13	87.28
		165.73	167.17	169.48	166.92	166.17	152.20	152.20	152.20	152.20	152.20
	42.48	41.81	40.66	37.50	37.09	36.54	41.43	40.20	37.27	36.94	36.41
	5.46	5.23	4.23	1.89	1.60	1.52	4.21	3.41	1.88	1.57	1.49
246	4.72	5.85	4.77	3.97	3.18	5.69	2.64	2.34	5.12	3.68	6.67
		90.70	91.99	93.10	92.37	90.35	90.49	90.99	91.78	91.38	88.91
		160.55	164.22	162.88	160.61	160.42	152.20	152.20	152.20	152.20	152.20
	39.57	39.72	38.56	36.51	36.17	36.04	39.48	38.56	36.37	36.07	35.94
	3.59	3.53	2.91	1.38	1.18	1.08	3.02	2.91	1.38	1.17	1.07
459	5.92	4.99	5.26	4.49	4.41	4.02	2.94	5.26	5.57	5.10	4.80
		91.50	91.43	92.77	92.04	91.70	90.83	91.45	91.29	90.62	90.22
		159.37	160.20	161.19	160.43	160.55	152.20	152.20	152.20	152.20	152.20
	38.19	38.02	37.37	35.88	35.67	35.50	37.95	37.31	35.84	35.64	35.47
	2.51	2.45	1.99	1.00	0.85	0.79	2.17	1.77	1.00	0.84	0.78
886	4.98	5.35	4.71	4.82	3.91	4.98	3.48	3.14	5.22	4.14	5.12
		90.82	91.52	91.10	91.18	90.46	90.77	91.02	90.71	90.50	90.00
		155.21	155.22	155.21	155.38	154.91	152.20	152.20	152.20	152.20	152.20

* Mean and S.D. of the 10^4 95/75 Values

4.3.2. LOCA Response Surface

The LOCA response surface detail in Section 3.2.3 was again used as the next step to a more realistic safety analysis scenario. Historically, the large LOCA has represented the most extreme challenge as the design basis for a plant's emergency core cooling system. For best-estimate plus uncertainty analysis, a 95/95 criterion has been

imposed on the entire spectrum of LOCA break sizes. Because the likelihood of a very large pipe break is extremely small, consideration has been given to using risk-informed requirements for emergency core cooling systems. In this approach, a transition break size would be established based on the expected frequency of breaks as a function of size. Below the transition break size, the 95/95 criterion would still be imposed. Above the transition break size, less conservatism would be applied (such as a 95/75 criterion).

The result of a 10^8-run CMC experiment yielded a "true" 0.95-quantile of 1683.65°F and 0.75-quantile of 1607.07°F. Figure 4. 24 shows the distribution of a 10^5-run CMC trial. In this case, the upper tail decays very quickly, so the 0.75- and 0.95-quantiles are fairly close together.

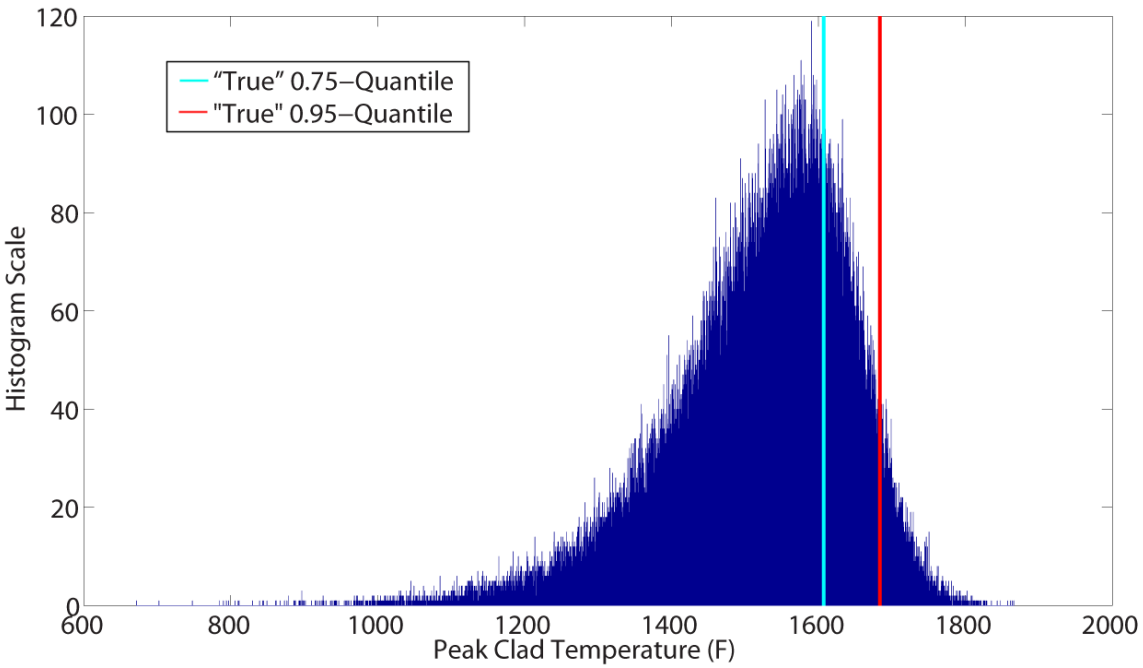

Figure 4. 24: Histogram of 10^5 Run CMC Trial

As before, the first test conducted was at the minimum number of runs necessary to find a 95/95 value using CMC-OS. Figure 4. 25 and Table 4. 16 show the results for 59-run CMC-OS trials and 60-run rLHS trials (m=6, t=10). The rLHS method continues to show better performance than CMC-OS, but less so than in the previous examples. This is most likely due to the output distribution shape. Unlike the previous non-linear equation examples, the output distribution does not have a long tail at the higher quantiles. Since CMC-OS only uses a single value to calculate a 95/95 value, it is less likely that this value is significantly larger than the true 0.95-quantile. So the variance reduction from rLHS is not as large. Also, the rLHS analysis has >6% of trials falling below the "true" quantile versus the 5% expected. Once again, this could be related to convergence, or to the selection of bandwidth parameters. It may have been possible to select bandwidth parameters that resulted in exactly 5% of trials falling below the "true" quantile, but again the challenge is finding bandwidth parameters that are applicable to a variety of systems and sample sizes.

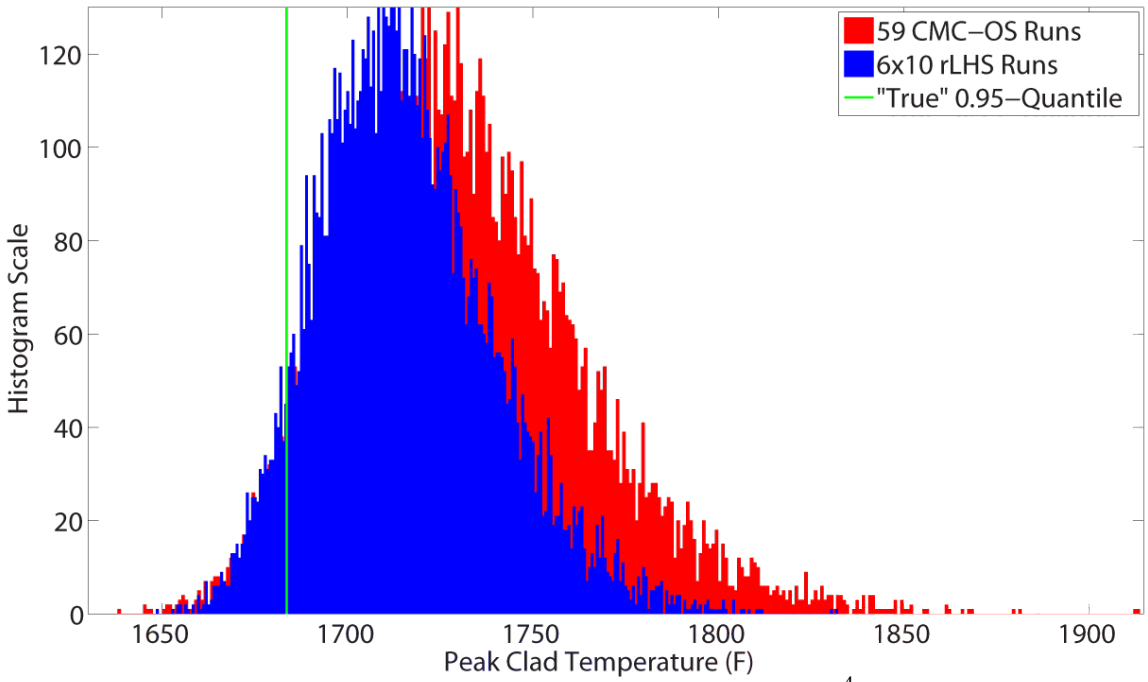

Figure 4. 25: Comparison of 95/95 Value Histograms for 10^4 Trials – 59 Runs

Table 4. 16: Comparison of 95/95 Values for 10^4 Trials – 59 Runs

	6x10 rLHS	59 CMC-OS
Mean of 10^4 95/95 Values	1,715.98	1,730.89
S.D. of 10^4 95/95 Values	23.32	31.83
% Below "true"	6.58%	4.73%

Figure 4. 26 and Table 4. 17 have the results for 124-run CMC-OS trials and 120-run rLHS trials (m=12, t=10). At this level, the rLHS is closer to the proper coverage level (as the complete results in Table 4. 18 show, the coverage level is ~90%). If this were a real safety analysis, the rLHS method would result in a 95/95 value that was, on average, 10°F lower than using CMC-OS.

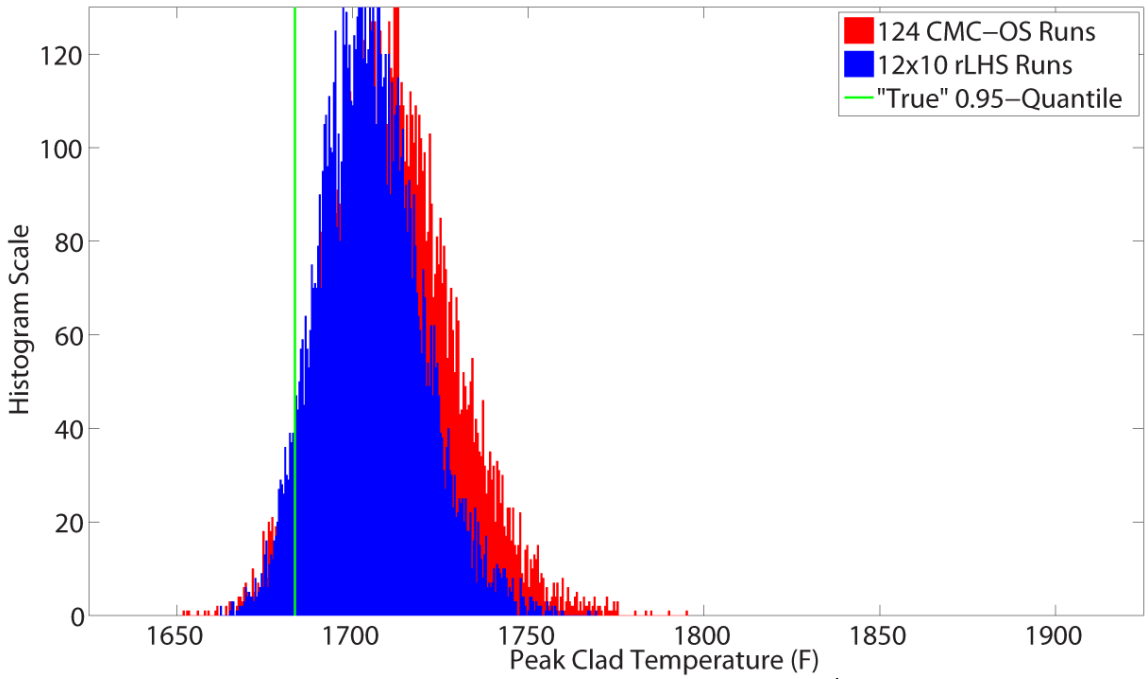

Figure 4. 26: Comparison of 95/95 Value Histograms for 10^4 Trials – 124 Runs

Table 4. 17: Comparison of 95/95 Values for 10^4 Trials – 124 Runs

	12x10 rLHS	124 CMC-OS
Mean of 10^4 95/95 Values	1,701.51	1,711.62
S.D. of 10^4 95/95 Values	11.66	17.90
% Below "true"	5.61%	4.69%

The complete results can be found in Table 4. 18. As with the previous examples, convergence appears to occur fairly quickly, although here it takes until the $n = 93$ run level to approach the appropriate values. Also like the previous examples, changing the bandwidth parameters at lower run levels causes the derivative estimation to get worse at $n = 311$, but it does not have a great effect on the coverage, which is still ~90%. More

work needs to be done on how to properly increase the bandwidth of the derivative estimation at these low run levels, and more will be said on this topic in Section 4.4.

Table 4. 18: 95/95 Results 10^4 Trials – LOCA Response Surface

		CFD for λ_p					Exact λ_p				
n	CMC-OS	CMC	AV	rLHS			CMC	AV	rLHS		
				t=10	*t*=20	*t*=30			*t*=10	*t*=20	*t*=30
	1,730.9	1,727.6	1,715.5	1,716.0	1,712.5		1,720.4	1,708.4	1,708.6	1,707.5	**Mean***
	31.83	28.30	24.70	23.32	25.62		20.19	17.83	16.59	22.91	**S.D.***
59	4.73	4.91	9.00	6.58	10.10		2.96	7.71	6.25	13.31	**% Below**
		89.12	88.94	91.20	86.26		88.70	90.04	91.24	82.74	**Covg.**
		849.91	853.12	867.07	866.65		695.70	695.70	695.70	739.92	**Avg. $\bar{\lambda}_p$**
	1,717.3	1,714.4	1,718.2	1,709.9	1,705.9	1,707.3	1,710.2	1,713.5	1,707.2	1,703.6	1,704.7
	21.67	20.50	21.28	18.23	19.90	19.73	16.02	15.48	13.95	15.75	15.92
93	4.80	5.84	4.34	5.74	12.11	9.67	4.68	2.05	3.74	9.90	8.54
		89.08	92.13	89.60	83.23	81.63	89.30	93.98	90.26	84.93	81.35
		807.86	807.37	783.16	771.50	792.16	695.70	695.70	695.70	695.70	695.70
	1,711.6	1,709.0	1,708.9	1,701.5	1,704.3	1,703.4	1,703.9	1,704.1	1,701.5	1,700.5	1,699.8
	17.90	16.41	16.24	11.66	14.93	15.63	13.42	13.00	11.66	12.09	12.99
124	4.69	5.31	4.87	5.61	7.22	9.38	6.32	5.06	5.61	8.23	10.51
		92.26	92.96	90.78	89.92	86.35	89.90	91.59	90.78	87.31	83.95
		851.29	845.28	695.70	837.51	834.45	695.70	695.70	695.70	695.70	695.70
	1,699.4	1,699.0	1,699.6	1,697.3	1,695.4	1,695.8	1,697.6	1,698.1	1,696.0	1,694.4	1,694.6
	10.08	9.77	9.56	8.28	8.11	8.43	8.71	8.44	7.32	7.19	7.50
311	5.38	5.23	4.29	4.56	6.67	6.93	5.27	3.85	3.99	6.39	6.94
		90.84	92.24	91.40	89.83	89.28	89.63	91.05	90.41	88.97	87.97
		766.44	767.96	766.79	754.68	764.04	695.70	695.70	695.70	695.70	695.70
	1,694.9	1,694.6	1,694.6	1,693.2	1,692.5	1,692.4	1,694.1	1,694.0	1,692.2	1,692.0	1,691.9
	7.10	7.09	6.94	6.15	6.01	5.99	6.49	6.29	6.11	5.46	5.49
548	5.51	5.92	5.32	5.41	6.51	6.74	5.57	4.79	7.56	6.02	6.44
		90.28	90.87	90.13	90.01	89.34	89.74	90.63	86.81	89.63	88.67
		729.53	731.8	733.23	734.49	732.66	695.70	695.70	655.47	695.70	695.70
	1,691.8	1,691.6	1,691.5	1,690.6	1,690.1	1,690.0	1,691.4	1,691.3	1,690.4	1,689.9	1,689.9
	5.10	5.15	5.07	4.42	4.31	4.30	4.80	4.68	4.06	3.98	3.99
1008	5.25	5.83	6.00	5.33	6.55	6.68	5.38	5.21	4.41	5.61	5.82
		89.73	89.69	89.76	89.33	88.72	89.52	89.85	90.13	89.56	89.12
		715.13	714.66	715.42	714.66	713.76	695.70	695.70	695.70	695.70	695.70
	1,689.4	1,689.2	1,689.1	1,688.3	1,688.2	1,688.4	1,689.1	1,689.0	1,688.2	1,688.2	1,688.3
	3.59	3.63	3.58	3.09	2.99	3.00	3.44	3.36	2.90	2.81	2.83
2004	5.34	6.12	5.81	6.48	5.95	5.76	5.35	5.25	5.67	4.96	5.10
		89.26	89.43	89.37	89.66	88.83	89.62	89.44	89.78	90.19	89.01
		708.07	709.23	706.91	706.91	707.70	695.70	695.70	695.70	695.70	695.70

* Mean and S.D. of the 10^4 95/95 Values

Figure 4. 27 and Figure 4. 28 show the results for 95/75 estimates, with numerical results in Table 4. 19 and Table 4. 20. As with the previous examples, the rLHS method is closer to the "true" quantile. However, once again it has <5% of trials below "true" at

the lowest run level, since only two LHS cases are being performed to satisfy the CLT. As the run level gets higher, and the rLHS method converges to the proper coverage level, it still provides a better characterization of the 0.75-quantile, as Figure 4. 28 and Table 4. 20 show at ~246 runs.

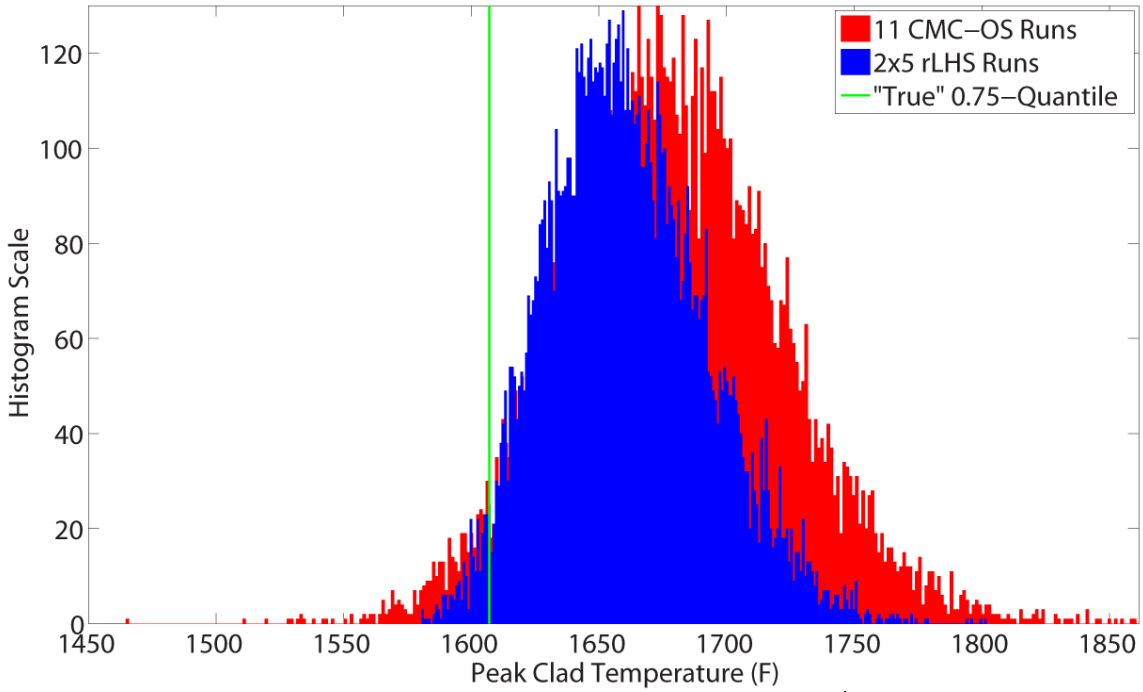

Figure 4. 27: Comparison of 95/75 Value Histograms for 10^4 Trials – 11 Runs

Table 4. 19: Comparison of 95/75 Values for 10^4 Trials – 11 Runs

	2x5 rLHS	11 CMC-OS
Mean of 10^4 95/75 Values	1,660.96	1,677.04
S.D. of 10^4 95/75 Values	30.93	42.55
% Below "true"	2.50%	4.35%

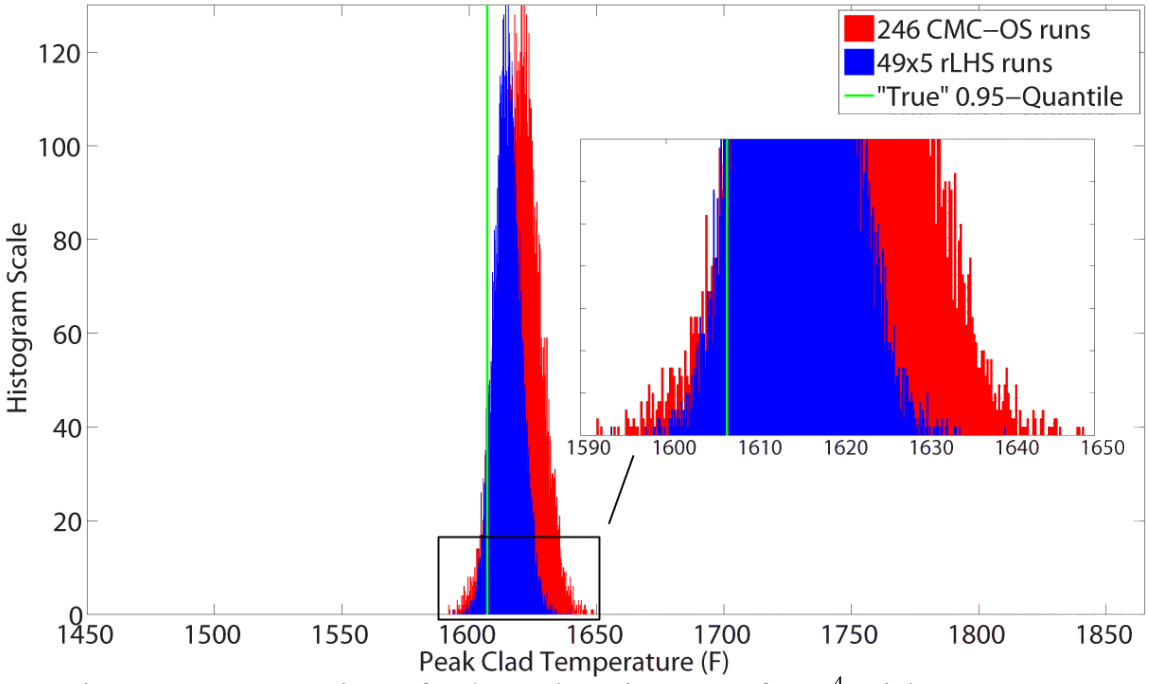

Figure 4. 28: Comparison of 95/75 Value Histograms for 10^4 Trials – 246 Runs

Table 4. 20: Comparison of 95/75 Values for 10^4 Trials – 246 Runs

	49x5 rLHS	246 CMC-OS
Mean of 10^4 95/75 Values	1,615.03	1,619.82
S.D. of 10^4 95/75 Values	5.08	7.73
% Below "true"	5.61%	5.20%

The complete results are in Table 4. 21. Once again, while not converged at $n = 11$, rLHS appears to have converged by $n = 29$. However, the number of trials falling below the true quantile increases to over 5% at $n = 40$. At first, this might seem to be caused by changing the bandwidth parameters (as in the previous examples), but there is a difference here. A closer look shows that using the exact value for the derivative, at this run level, results in over 9% of trials falling below the true quantile.

That means the error at this run level is more likely caused by difficulty estimating the quantile, rather than the derivative estimation. By the next highest run level, the quantile estimation issue is resolved. Also, rLHS provides a larger gain here than when estimating the 0.95-quantile. At $n = 29$, rLHS returns a 95/75 value that is 15°F lower on average, than the result when using CMC-OS.

Table 4. 21: 95/75 Results 10^4 Trials – LOCA Response Surface

		CFD for λ_p					**Exact λ_p**				
n	**CMC-OS**	**CMC**	**AV**	**rLHS**			**CMC**	**AV**	**rLHS**		
				***t*=5**	***t*=10**	***t*=15**			***t*=5**	***t*=10**	***t*=15**
11	1,677.0	1,675.7	1,649.2	1,661.0			1,669.7	1,645.6	1,654.8		**Mean***
	42.55	42.29	31.67	30.93			36.10	25.56	24.48		**S.D.***
	4.35	4.64	8.01	2.50			4.83	5.14	1.82		**% Below**
		89.61	89.65	95.30			90.62	92.82	95.37		**Covg.**
		311.07	303.97	321.47			282.85	282.85	282.85		**Avg. $\tilde{\lambda}_p$**
29	1,646.6	1,643.7	1,641.6	1,631.6	1,628.8		1,640.5	1,641.2	1,630.7	1,628.1	
	23.50	23.59	21.35	16.20	16.78		22.54	17.50	14.84	15.80	
	4.94	5.90	4.22	5.52	8.53		6.84	1.82	5.36	8.50	
		90.14	92.47	88.55	82.50		90.17	94.46	88.00	81.74	
		307.11	286.46	298.65	296.91		282.85	282.85	282.85	282.85	
40	1,640.8	1,636.9	1,631.4	1,625.5	1,623.6	1,623.1	1,634.6	1,630.0	1,623.9	1,622.3	1,622.0
	19.74	19.62	17.08	13.42	13.93	22.45	19.21	15.21	12.74	13.28	20.92
	4.25	6.25	7.34	7.96	11.46	27.03	7.73	6.10	9.34	12.63	27.00
		90.59	90.02	89.58	84.18	44.50	89.61	90.71	87.18	82.21	44.44
		303.08	297.92	306.96	306.65	308.19	282.85	282.85	282.85	282.85	282.85
135	1,624.6	1,624.9	1,620.7	1,618.9	1,616.5	1,617.9	1,624.6	1,620.4	1,618.7	1,616.3	1,617.7
	10.53	10.69	8.96	7.07	6.62	6.79	10.57	8.37	6.79	6.46	6.57
	4.73	4.85	6.38	4.33	7.46	5.19	5.06	5.51	4.27	7.85	5.37
		89.47	90.15	89.38	88.37	87.53	89.85	90.57	88.78	87.72	86.84
		289.10	288.07	290.34	290.02	290.02	282.85	282.85	282.85	282.85	282.85
246	1,619.9	1,619.9	1,617.9	1,615.0	1,614.8	1,614.3	1,619.7	1,617.8	1,614.9	1,614.7	1,614.2
	7.81	7.93	6.69	5.07	4.83	4.95	7.83	6.34	4.93	4.71	4.85
	5.12	5.29	5.29	5.68	5.29	6.87	5.45	4.43	5.55	5.24	6.83
		89.56	89.79	90.02	89.33	88.12	89.77	90.36	89.91	89.38	87.94
		286.98	285.72	286.99	286.37	287.41	282.85	282.85	282.85	282.85	282.85
459	1,615.9	1,616.7	1,614.6	1,612.8	1,612.5	1,612.5	1,616.6	1,614.5	1,612.8	1,612.4	1,612.4
	5.61	5.68	4.77	3.73	3.49	3.46	5.66	4.60	3.65	3.44	3.40
	5.65	4.45	5.67	5.97	5.90	5.90	4.58	5.37	6.12	5.77	5.82
		90.31	90.46	89.97	89.98	88.83	90.28	90.46	89.32	89.84	88.88
		286.08	286.03	286.76	286.97	286.43	282.85	282.85	282.85	282.85	282.85
886	1,613.8	1,613.8	1,612.7	1,611.3	1,611.1	1,611.0	1,613.8	1,612.7	1,611.3	1,611.1	1,611.0
	4.05	4.09	3.46	2.65	2.49	2.49	4.09	3.38	2.60	2.46	2.46
	5.03	4.88	5.41	5.04	5.17	5.75	5.21	4.80	5.05	5.05	5.64
		90.24	89.33	90.28	89.55	89.27	90.10	89.89	90.22	89.57	89.38
		284.01	283.80	284.34	284.07	283.99	282.85	282.85	282.85	282.85	282.85

* Mean and S.D. of the 10^4 95/95 Values

4.3.3. PRA Event Tree

The PRA event tree analysis detailed in Section 3.2.4 was repeated for these techniques. However, instead of the figure of merit being mean risk, a more detailed characterization of the output was desired. Recall, each time the PRA is carried out, 841 unique scenarios are created. Each one of these scenarios has a frequency and a consequence assigned to it. Even though there are 841 unique scenarios, there are actually only 13 different consequence levels. This is due to the fact that the offsite dose is mostly dependent on the time of release and release fraction of core inventory. As the event trees in Section 3.2.4 show, there are four possible core damage-states (1, 2, 3, 4), and three possible times of release (early, late, and leakage). This gives 12 possible levels of consequence. In addition to these 12 levels, there is also the release associated with those events that had no core damage. This consequence is related to the radioactive material released from the primary system into containment, and then leaked out of containment. These consequence levels can be seen in Table 4. 22.

Table 4. 22: Consequence Bins

Bin	Core Damage State	Time of Release
1	1	Early
2	1	Late
3	1	Leakage
4	2	Early
5	2	Late
6	2	Leakage
7	3	Early
8	3	Late
9	3	Leakage
10	4	Early
11	4	Late
12	4	Leakage
13	Undamaged	Leakage

Each of the 841 unique scenarios falls into one of these 13 consequence bins. Therefore, the frequencies of the scenarios that fall into each bin can be summed. This will give a final output of 13 bins each with a consequence level and a frequency. Figure 4. 29 shows an example output from the PRA analysis, with the 13 points on a consequence versus frequency plot.

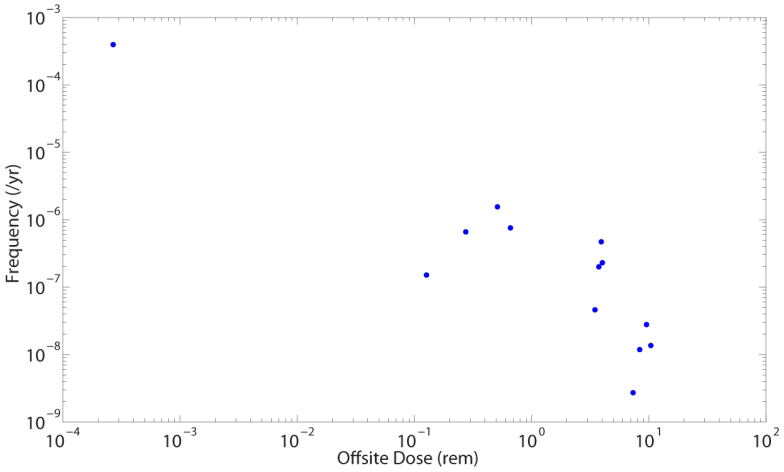

Figure 4. 29: Example of PRA Output

These points can also be used to create a complementary cumulative distribution function (CCDF). This is done by summing the frequency of the events, starting with the event with the largest consequence. Figure 4. 30 shows the CCDF version of the output for the example listed in Figure 4. 29.

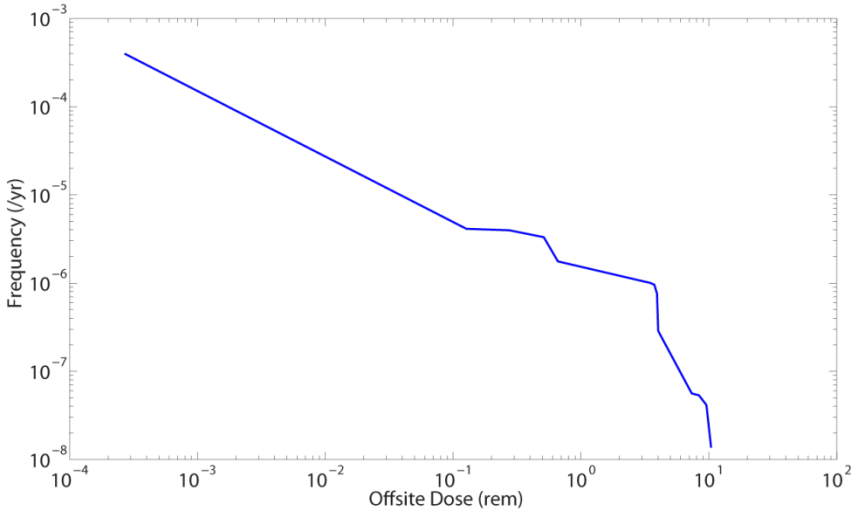

Figure 4. 30: Example CCDF Output of PRA

Historically, the results of PRAs have been used to provide risk insights but have not been required to satisfy quantitative risk limits. In NUREG-1860, the NRC developed a draft technology-neutral framework that could be applied in the future to advanced nuclear power plant designs independent of the type of design [14]. NUREG-1860 introduces the concept of a frequency-consequence limit curve in which the PRA scenarios would be aggregated into Licensing Basis Events (LBEs), each of which would be required to fall below the limit curve presented Figure 4. 31. Although the LBEs are based on the results of PRAs, the manner in which the characteristic frequency and consequence of an LBE is determined and compared with the limit curve does not actually constrain the risk. If the risk analyst refines the risk assessment, for example by dividing a small break loss of coolant accident into two break sizes, it becomes easier to satisfy the criteria. In theory, it would be possible to have an infinite number of LBEs with infinite risk and still satisfy the limit curve.

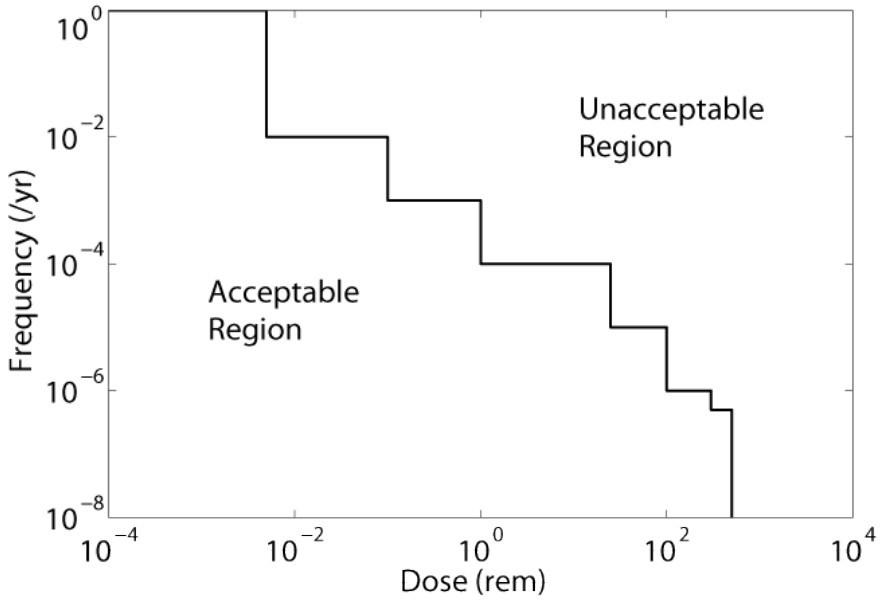

Figure 4. 31: Technology Neutral Framework F-C Limit Curve

Because of these concerns about the NUREG-1860 limit curve approach, an alternative limit curve approach has been proposed [108] in which a limit curve is used that establishes a bound on the CCDF of the LBEs (and thus a bound on risk), as seen in Figure 4. 32. In this case the limit curve has a slope of -1 (on log-log scale) in the low consequence region and a slope of -1.5 for higher consequence events. The curves are pinned at an offsite dose of 25 rem at a frequency of 1E-4 per year, since this is the site boundary dose limit for design basis accidents.

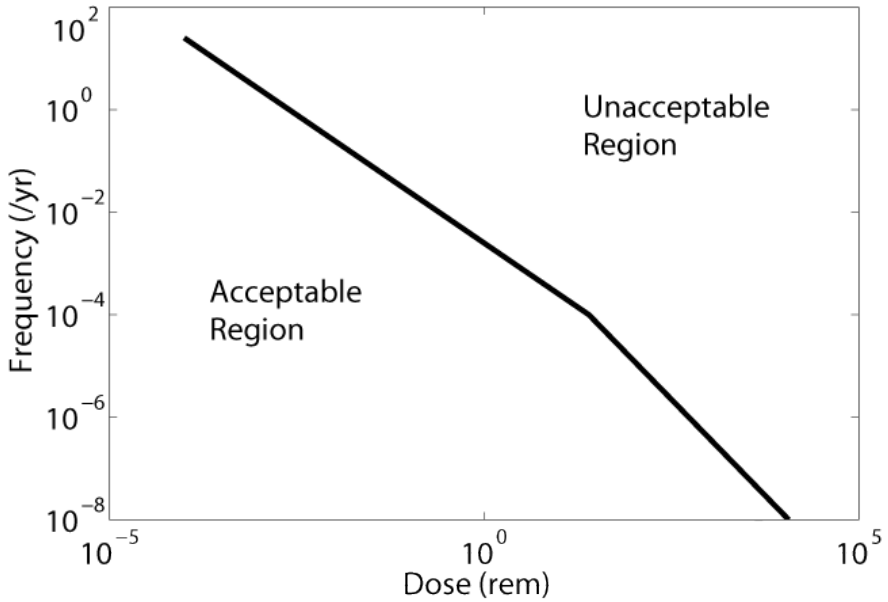

Figure 4. 32: Proposed CCDF Limit Curve

The historical display of PRA results as presented in NUREG-1150 [6] (for which a 100-run LHS design was examined) shows the family of CCDFs obtained through the performance of the uncertainty analysis. Curves showing the 0.05-quantile, median, mean and 0.95-quantile of these CCDFs are presented graphically, but no consideration is given to the confidence level of these statistics. NUREG-1855 [7] provides more guidance about the reporting of PRA results in relation to a regulatory guideline, but stops short of giving specific requirements. Instead, it states analysts should provide "A qualitative statement of confidence in the conclusion and how it has been reached" and that "to support the statement of confidence, the analyst should identify the key sources of uncertainty that were addressed." The metric usually provided by a PRA is the mean or higher quantile. In the following examples, the 0.75- and 0.95-quantiles will be used, with a high level of confidence (95%).

4.3.3.1. Comparison with a Risk Limit Curve

The first PRA analysis undertaken sought to compare the output results to the CCDF limit curve presented in Figure 4. 32. While the construction of the CCDF curve for a single PRA run is straightforward, the creation of a quantile or 95/95 CCDF curve is more complex. This is due to several causes. First, there is uncertainty not only in the frequency of the scenarios, but in the consequence. This means that scenarios are shifted along both the *x* and *y* axes. Secondly, due to the uncertainties, a single run's CCDF curve may be in the higher regions of the output distribution at one part of the plot, but be in the lower regions in another part. This means that a 95/95, or even a 0.95-quantile, curve cannot be selected directly from the resulting curves of the *n* number of runs. Instead, a curve must be created by point-by-point comparison.

In order to create a quantile CCDF curve of the resulting distribution, the resulting points from within each consequence bin were viewed directly. Figure 4. 33 shows how this was done. For each of the 13 consequence bins, the points from each run of the PRA create a spread of possible values. Figure 4. 33 shows the results for 100 CMC runs for the 13^{th} consequence bin of Table 4. 22 (the bin with the lowest consequence level, as shown by the comparison to the example in Figure 4. 29). The spread covers both the *x* and *y* axes. In order to find a 0.95-quantile value, the 0.95-quantile consequence and the 0.95-quantile frequency are determined. Here, the 0.95-quantile dose is ~0.0129 rem and the 0.95-quantile dose is ~0.00184 /yr. A new point is created using these values, and is considered the 0.95-quantile point for that consequence bin.

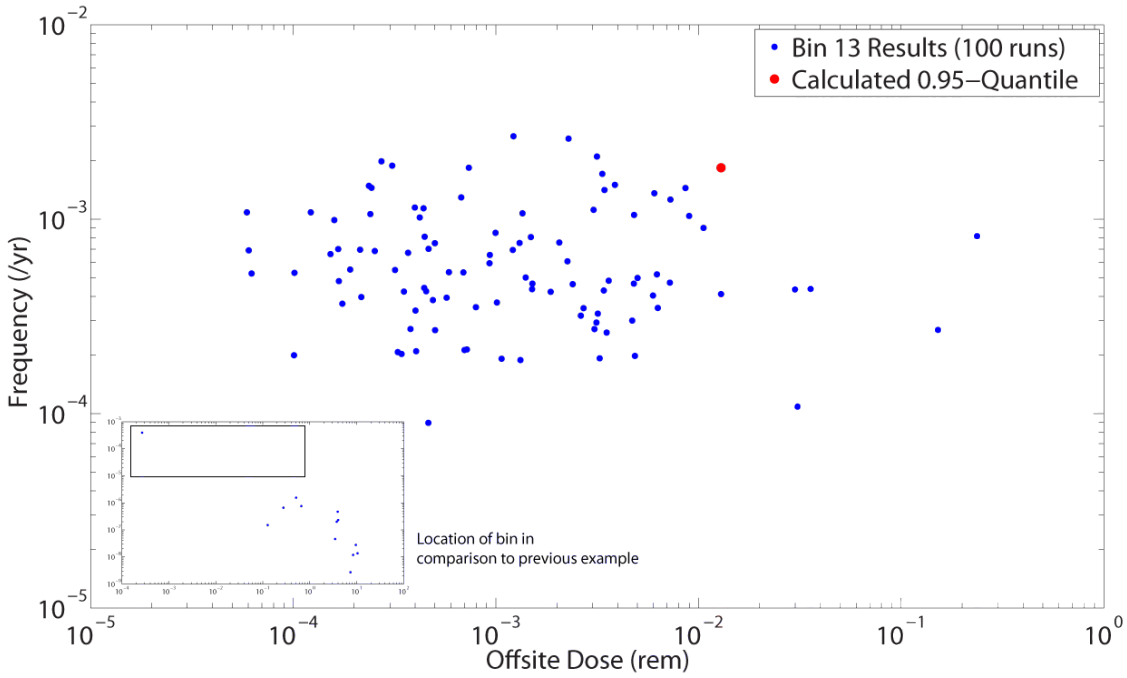

Figure 4. 33: Demonstration of Quantile Estimation

Obviously, the created quantile point will be overly conservative because it does not represent the output of an actual scenario, but a conglomeration of the worst consequences and frequencies. As the plot shows, the points that had higher consequences fell at lower frequencies, and the points with higher frequencies were located at lower consequence levels, meaning the constructed 0.95-quantile point is not necessarily realistic. A less conservative technique may have been to find a type of Euclidean distance, in log-log space, of each point to the projected limit value in Figure 4. 32. Then use this distance as the output metric, meaning the runs would be sorted based on the value of this distance. This would mean each point is now only a function of one variable, distance, rather than an x and y coordinate. The initial calculation of this distance would be more difficult, but it would simplify the analysis. However, using the

technique shown in Figure 4. 33 to calculate a quantile provided a consistent means of comparing the confidence interval methods described in Section 4.2, and the error is in the conservative direction for both the frequency and consequence.

In order to create a quantile CCDF curve, the process described above was repeated for each of the 13 consequence bins. This results in Figure 4. 34, where a 0.95-quantile point has been found for each of the 13 bins. While the results may look confusing, the actual calculation of the individual bins was not difficult due to the fact that the consequence bins always fell in the same order. This was the case since the core damage states and times of release magnitudes would stay in the same order regardless of the value of the uncertainties.

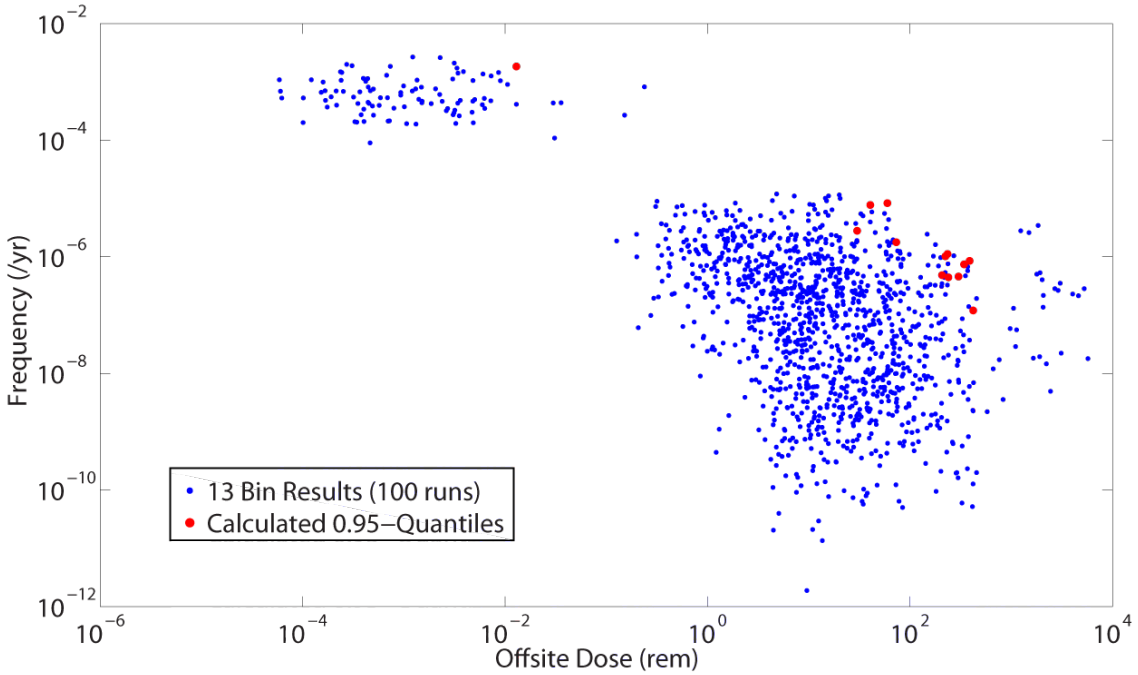

Figure 4. 34: Quantile Calculation for all 13 Bins

Using the results of Figure 4. 34, a 0.95-quantile CCDF curve could be created. This is shown in Figure 4. 35, in comparison to the 13 consequence bin points, and in Figure 4. 36, in comparison to the 100 CCDFs created by the 100 individual PRA runs.

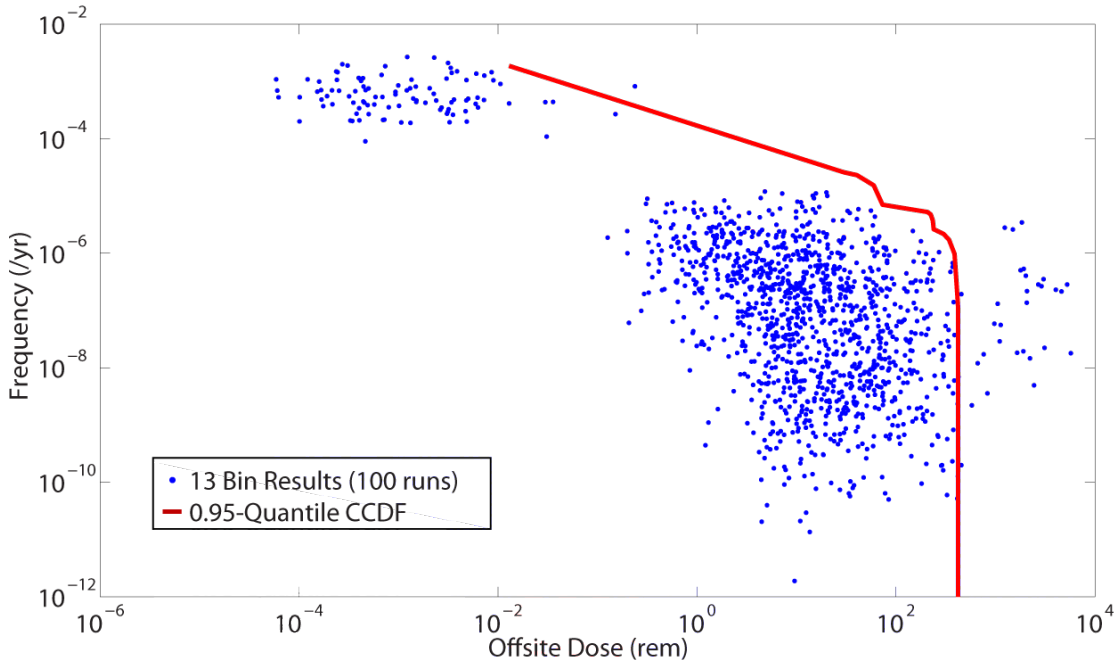

Figure 4. 35: 0.95-Quantile CCDF Curve

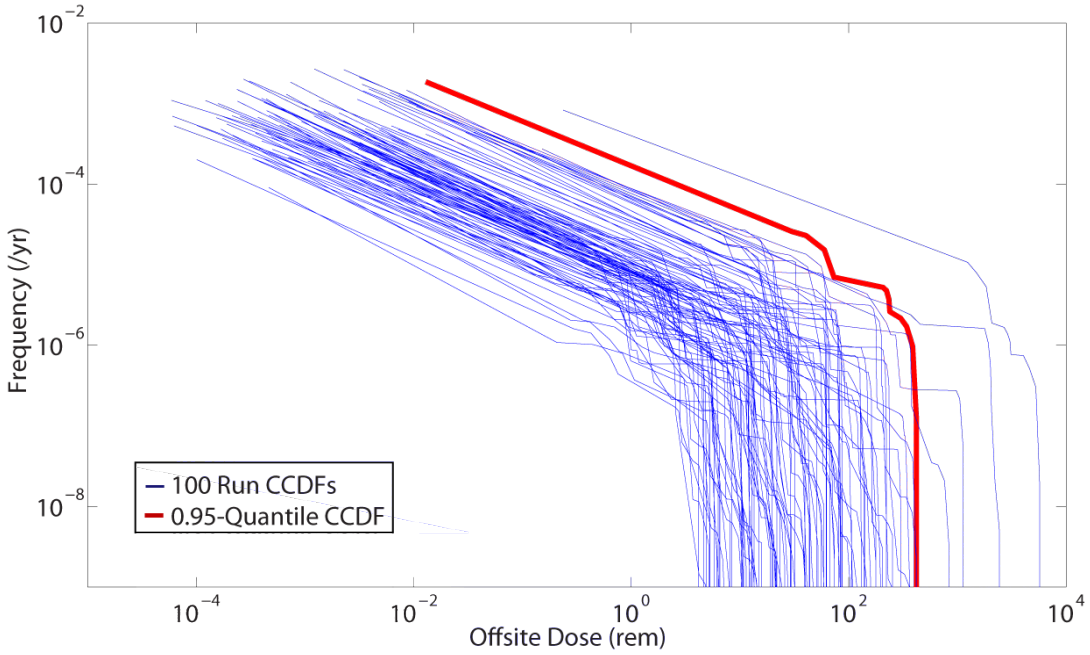

Figure 4. 36: 0.95-Quantile CCDF Curve with 100 CCDFs

A similar technique was used to form confidence intervals. For the asymptotic methods, a CI was found in respect to the consequence and in respect to a frequency. These were then used to construct a total CI point for that consequence bin, and this process was repeated for all the consequence bins in order to form a CI CCDF curve. For CMC-OS, if 59 runs were conducted, for example, the highest consequence value for that bin, and the highest frequency value for that bin were combined to form a 95/95 point. If 93 runs were conducted, the second highest values of consequence and frequency were combined to form a 95/95 point, and so on.

A 10^6-run CMC trial was conducted first to establish the "true" 0.75- and 0.95-quantile CCDF curve. Figure 4. 37 shows the results of a smaller, 25,000-run CMC trial since it was infeasible to plot the large trial. This plot is presented just to show the spread

of possible outcomes and where the “true” 0.75- and 0.95-quantile fall. As the figure shows, the 0.95-quantile does not satisfy the limit curve because it violates the line at ~200 rem. The 0.75-quantile does satisfy the limit curve.

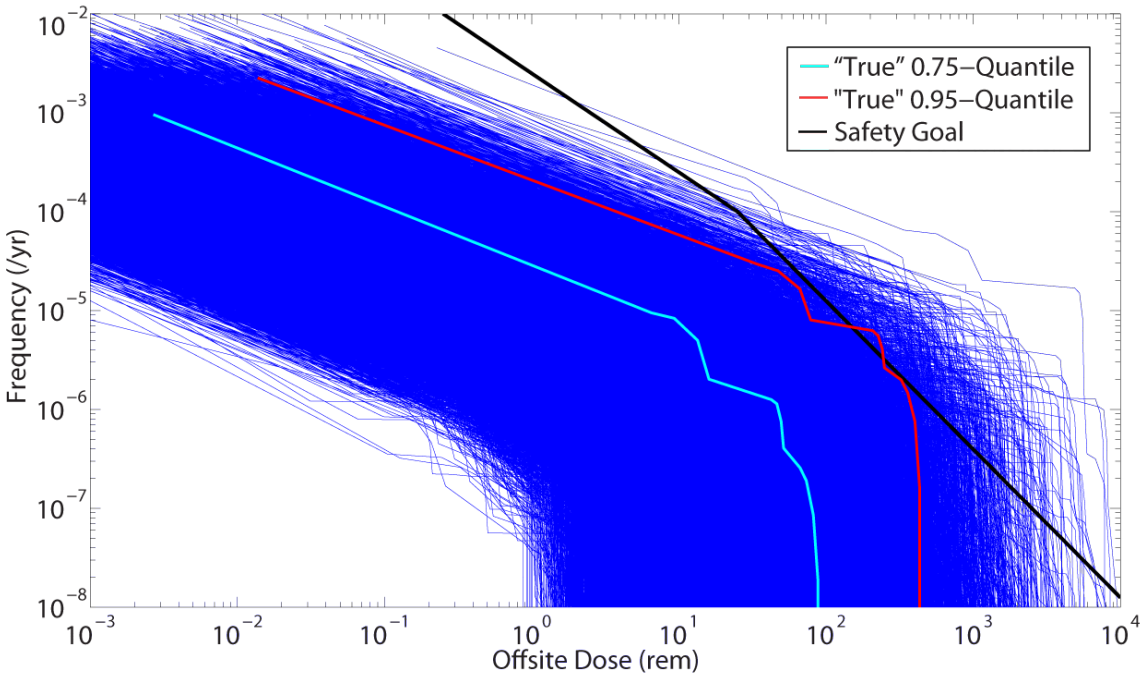

Figure 4. 37: CCDF Curves for 25,000 Run CMC Trial

Next, Figure 4. 38 show the 95/95 CCDF curves for 10^4 trials of 59 CMC-OS runs. It is important to remember that each trial consisted of 59 individual runs, where each run resulted in its own CCDF curve. The CCDF curves on Figure 4. 38 represent the 95/95 value CCDF curve of each of the 10^4 trials. Here, all 10^4 95/95 curves are compared to the “true” 0.95-quantile curve, and the candidate safety goal, presented as a black line. If the CCDF curve lies to the left of the safety curve, it satisfies the safety limit. As can be seen, the “true” 0.95-quantile violates the safety limit at one interval

around 200 rem. Therefore, the result of this analysis should conclude that the system does not pass.

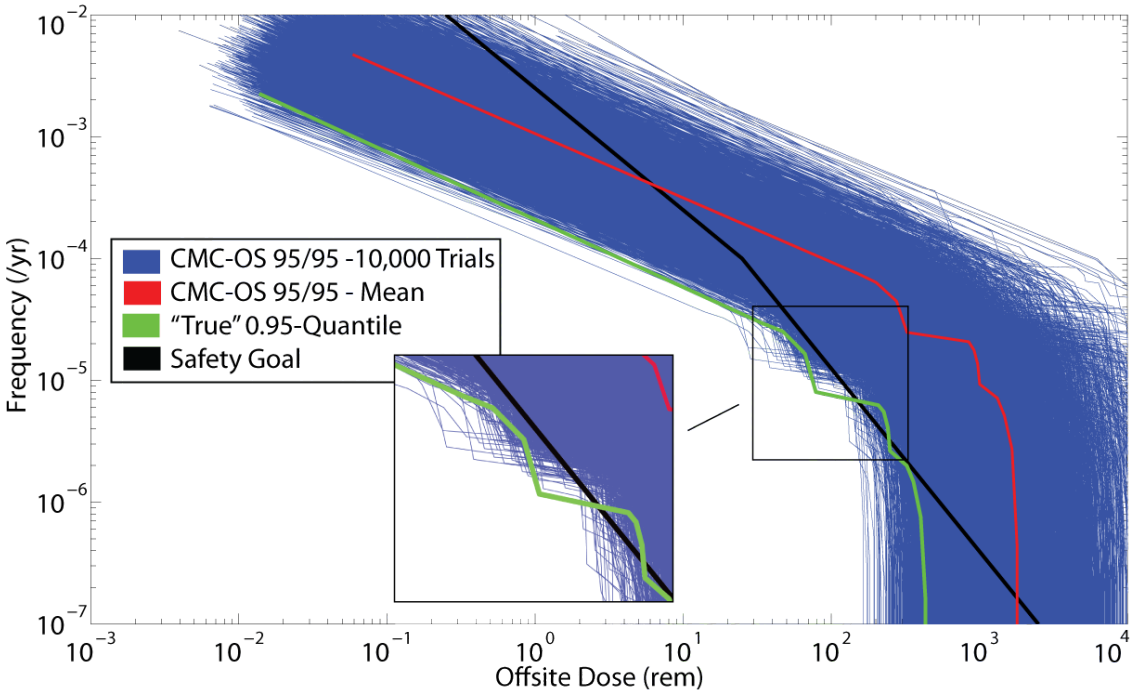

Figure 4. 38: Comparison of 95/95 Curves to Limit Curve for 10^4 CMC-OS Trials

Next, Figure 4. 39 shows the same results, but for 10^4 trials of a rLHS design with m=6 and t=10. Since the characterization of these curves results in a large amount of data, only these qualitative plots are presented as evidence of the reduction in margin by the rLHS method.

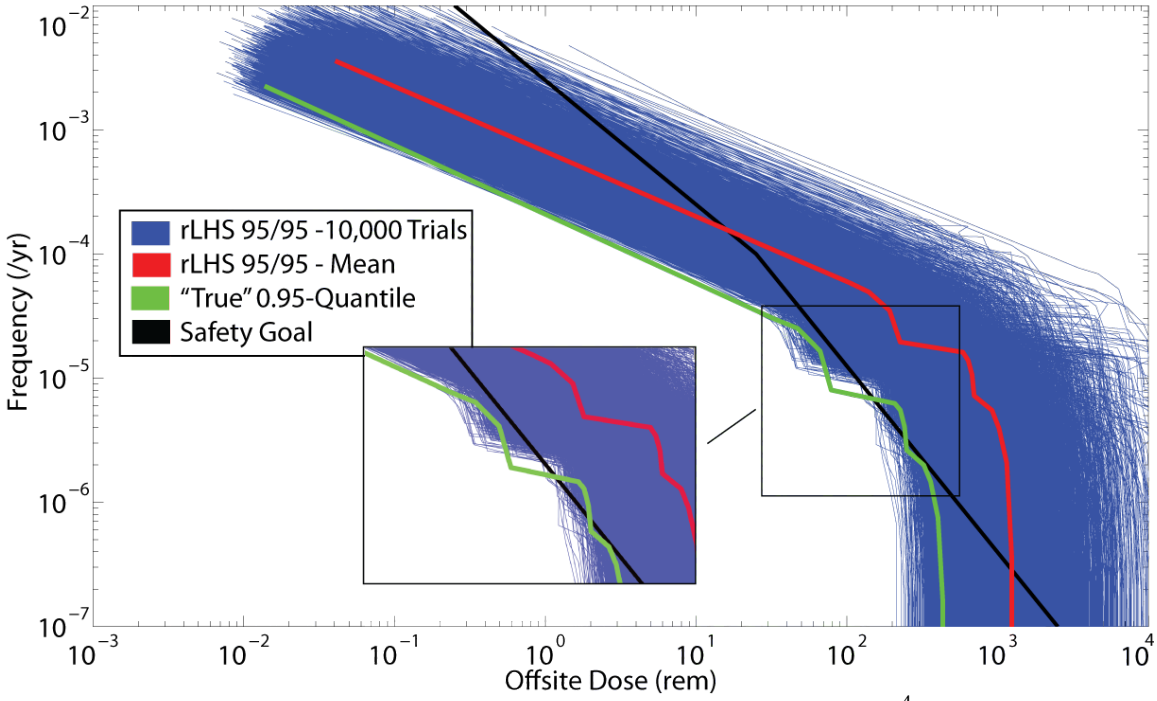

Figure 4. 39: Comparison of 95/95 Curves to Limit Curve for 10^4 rLHS Trials

There are several factors to note from these results. First, 95/95 CCDF curves from CMC-OS and rLHS both have means that fail the safety goal limit, which is the correct conclusion, but at this low run level, both methods have means *well* above the "true" 0.95-quantile, and do not characterize the curve well. If a criterion as stringent as 95/95 was imposed, it would be necessary to perform more runs if more informative results were required. Second, and more importantly, is that for the CMC-OS trials, there were a few trials where all points of the 95/95 CCDF curve satisfied the safety goal. The "true" 0.95-quantile does not satisfy the safety goal, since it violates the curve at ~200 rem and 10^{-5} /yr. This means that there is small chance that an analyst could commit a Type-I error, or believe, falsely, that the system had fulfilled the safety goal. This error does not occur with the rLHS trials.

Since the 95/95 requirement may be overly-stringent for this type of analysis, 95/75 values were also calculated. Here, Figure 4. 40 and Figure 4. 41 show the results for 10^4 trials of 11 CMC-OS runs and 10 rLHS runs (m=2, t=5).

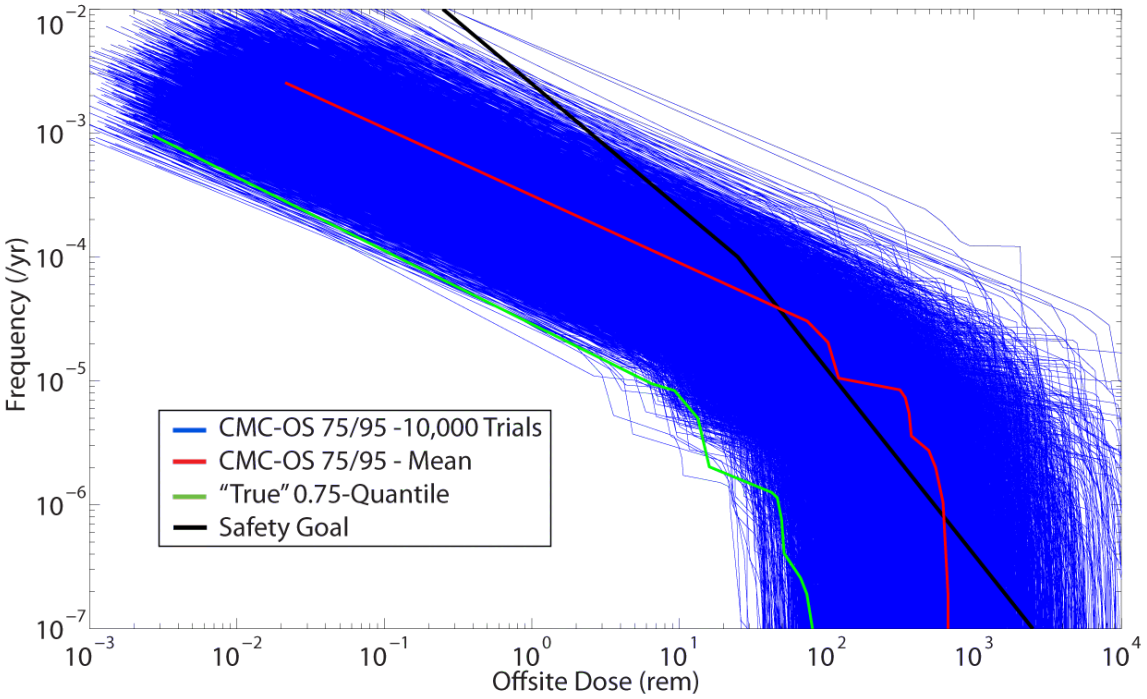

Figure 4. 40: Comparison of 95/75 Curves to Limit Curve for 10^4 CMC-OS Trials

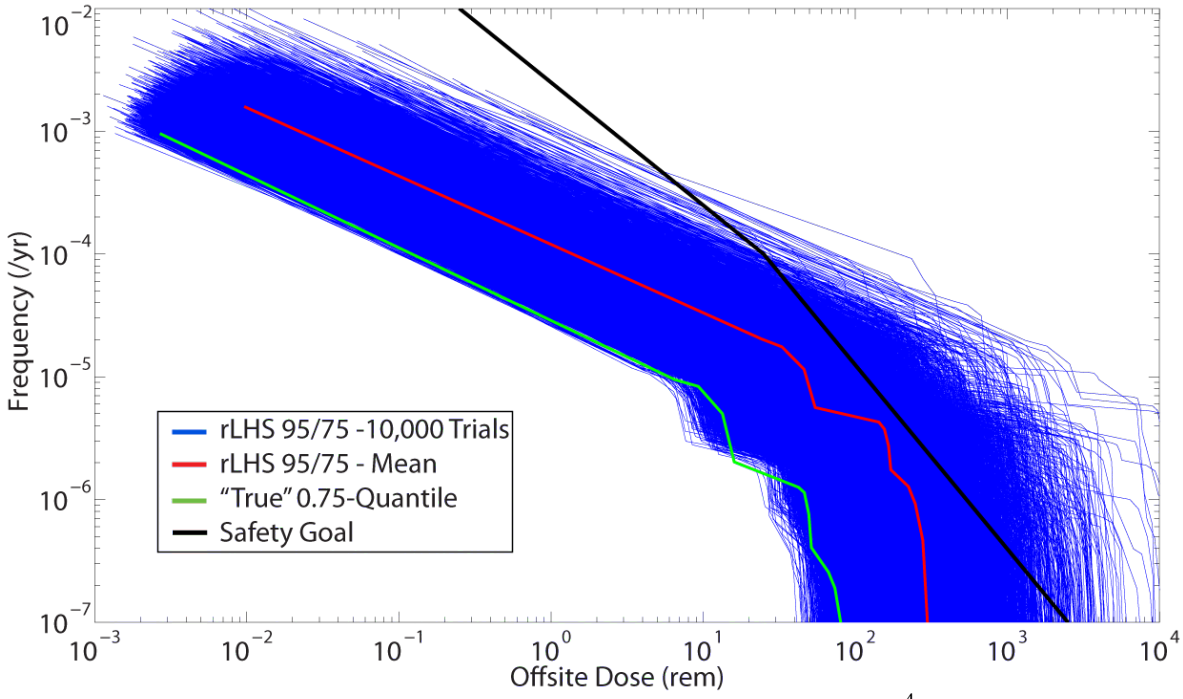

Figure 4. 41: Comparison of 95/75 Curves to Limit Curve for 10^4 rLHS Trials

At this run level, the mean of the CMC-OS trials violates the safety limit, while the mean of the rLHS trials does not. This example demonstrates how the large variance of CMC sampling can lead to incorrect conclusions being made from the analysis. Even though the rLHS method does have trials that also violate the safety goal, it is far less likely that a Type-II error would be committed with the rLHS method than with CMC-OS method.

Figure 4. 42 and Figure 4. 43 show similar results but for 10^4 trials of 40 CMC-OS runs and 40 rLHS runs (m=8, t=5). Once again, the large variance of the CMC-OS causes some trials to violate the safety limit, and could lead to a Type-II error. None of the rLHS trials violate the safety limit.

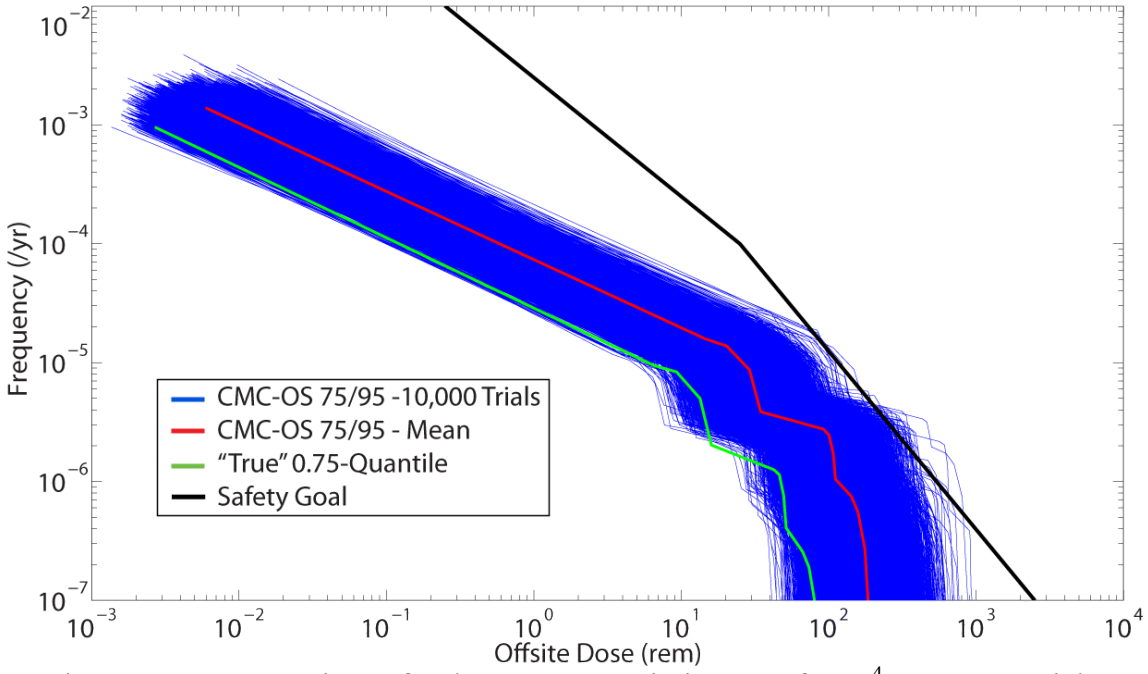

Figure 4. 42: Comparison of 95/75 Curves to Limit Curve for 10^4 CMC-OS Trials

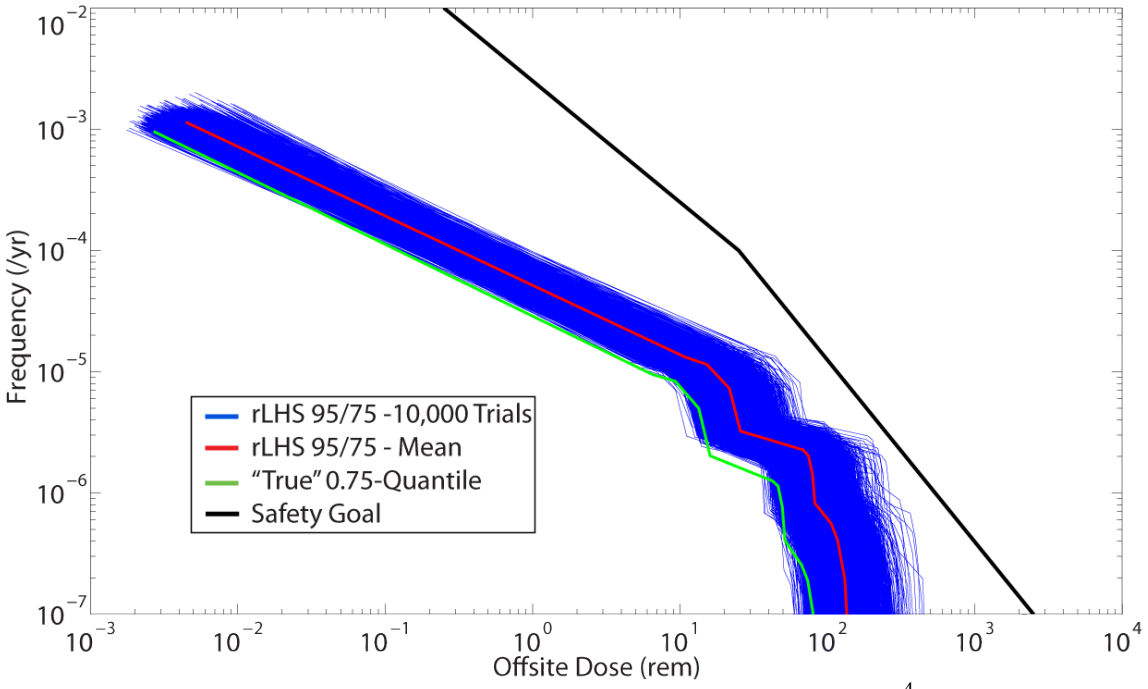

Figure 4. 43: Comparison of 95/75 Curves to Limit Curve for 10^4 LHS Trials

4.3.3.2. Comparison with an LBE Limit Curve

Next, a comparison was made to the LBE limit curve presented in Figure 4. 31. For this example, the 13 consequence bins described in Table 4. 22 were assumed to be analogous to the LBEs of NUREG-1860. Figure 4. 44 shows a comparison of offsite dose consequence bins for 10^4 trials of a 59-run CMC-OS and 60-run LHS design (m=6, t=10). Even though there are 13 consequence bins, only three are shown to keep the figure legible and to illustrate the trend. The rectangles are designed using the 0.01- and 0.99-quantile consequence and frequencies values of the 95/95 values of those bins from all 10^4 trials. This means on the x-axis, the left side of the rectangle is at the 0.01-quantile offsite dose from 10^4 95/95 values, and the right side of the rectangle is at the 0.99-quantile of the offsite dose from 10^4 95/95 values. The top and bottom are the same but for the 0.01- and 0.99-quantile frequencies of the 10^4 95/95 values. This helps show the range of possible 95/95 values, for each bin, using that method. If points lie to the left of the curve, they satisfy the limit curve. As can be seen, the three "true" values satisfy the frequency-consequence limit curve.

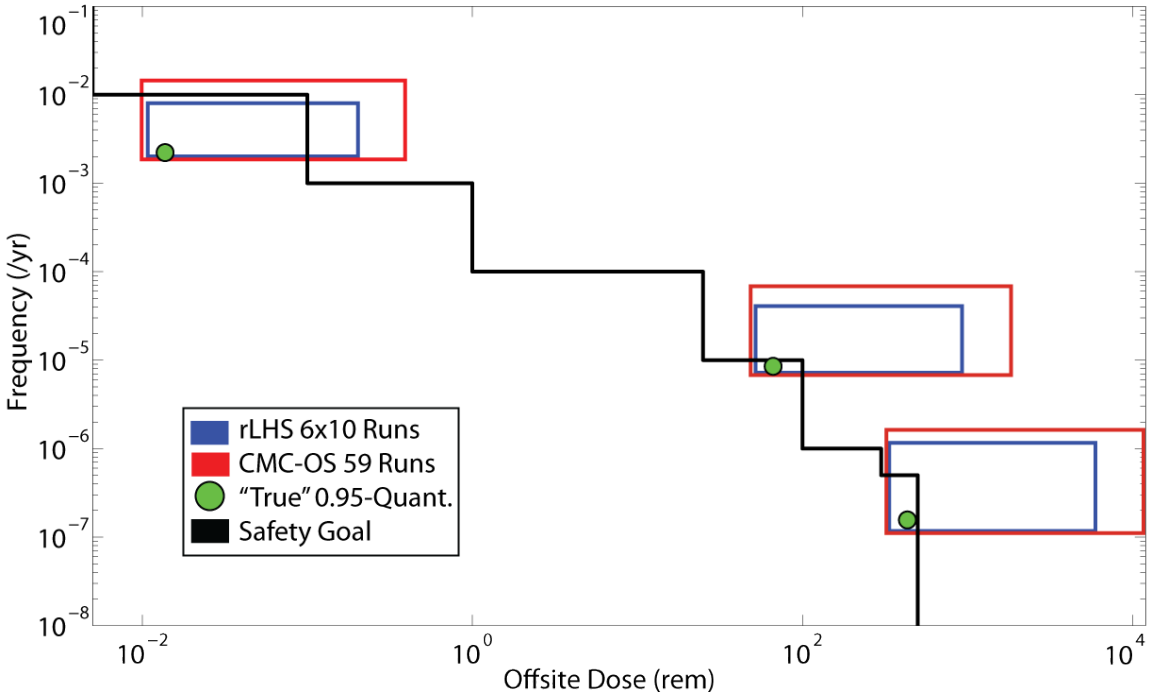

Figure 4. 44: Comparison of 95/95 Value Consequence Bins to Limit Curve

As the figure shows, even though the "true" points satisfy the curve, both methods greatly overestimate the value. However, rLHS does get slightly closer than CMC-OS. This test was repeated with 95/75 values, with Figure 4. 45 showing the results for 10^4 trials of 11-run CMC-OS and 10-run rLHS (m=2, t=5). Here, the rLHS method is noticeably better at approximating the location of the "true" quantile point. The CMC-OS range covers higher and lower values than the rLHS method. Figure 4. 46 repeats this for 40-run CMC-OS trials and 40-run rLHS trials (m=8, t=5). As the figure shows, the rLHS method once again provides a better characterization of the 0.75-quantile points.

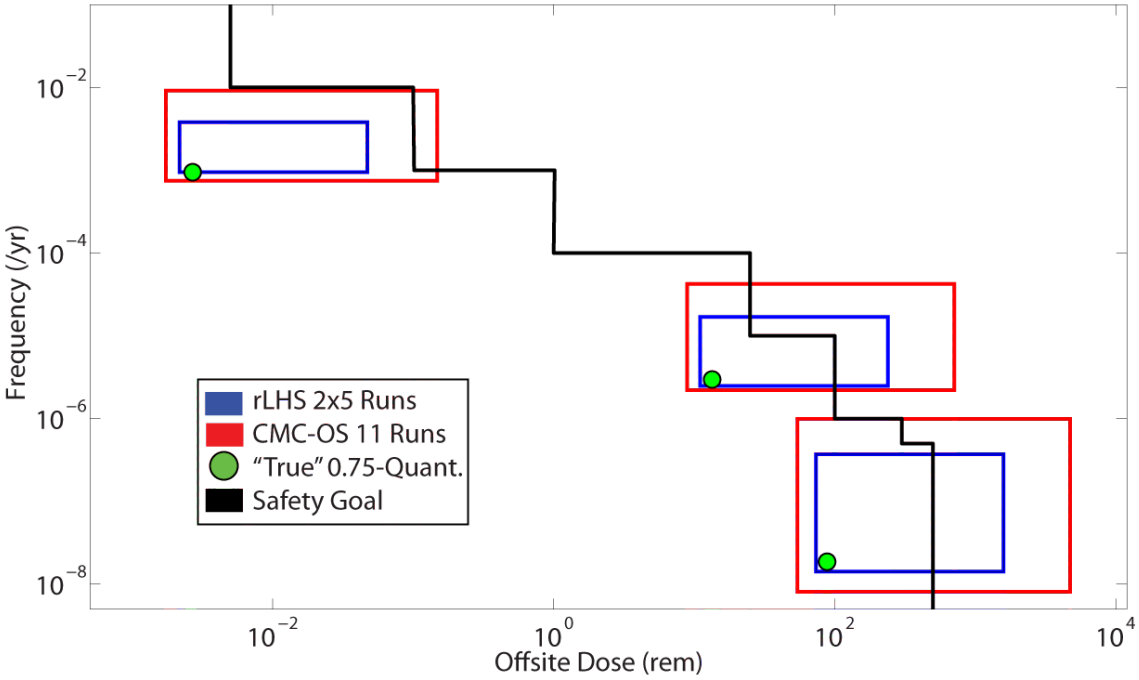

Figure 4. 45: Comparison of 95/75 Value Consequence Bins to Limit Curve

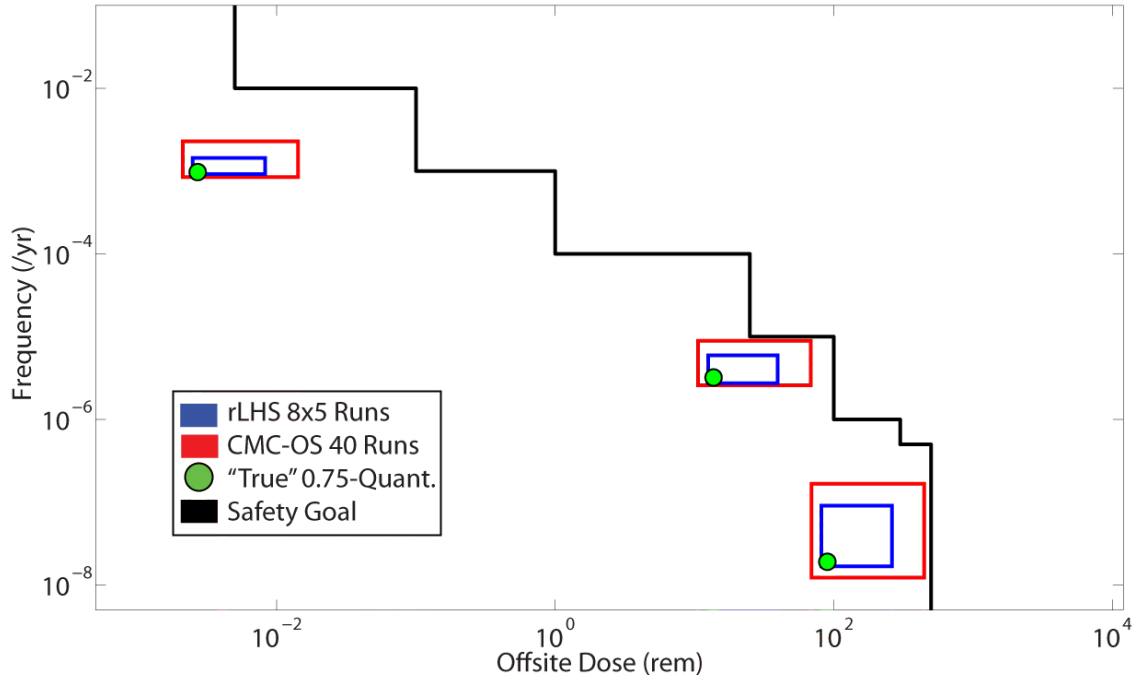

Figure 4. 46: Comparison of 95/75 Value Consequence Bins to Limit Curve

4.3.4. MELCOR LOCA Analysis

The next analysis was conducted to compare the methods using an actual nuclear power plant severe accident analysis computer code. The code used for this analysis was MELCOR 2.1, developed by Sandia National Lab [109]. This code was chosen not only because it is used in real nuclear safety analyses, but because it represents a "large and complex" model. Here, a large model is one requiring significant amounts of human, computational, or other resources in its construction and operation [18]. Complex means the system is made up of a large number of parts that interact in a nonsimple way [110]. Morgan and Henrion actually use NRC "general purpose regulatory model" computer codes as an example of a large and complex system [18]. The scenario chosen was based on a MELCOR demonstration problem presented in CR-6119 [111]. It represents a large break LOCA at the now retired Zion Nuclear Power Plants (ZNPP) near Chicago.

MELCOR is not an NRC-approved computer code for the performance of the analysis of loss of coolant accidents for regulatory submittals, like the RELAP5 computer code discussed earlier. The treatment of some two-phase flow phenomena is not of the level of fidelity required for regulatory-analyses. MELCOR is primarily used for the analysis of severe accidents in which, for example, there is not only a pipe break leading to loss of coolant, but also a failure of the emergency core cooling system, as done in this example. Nevertheless, MELCOR does a detailed nodalization of the reactor coolant system, models fuel pin heat and clad oxidation, and solves the Navier-Stokes flow equations.

Both units at ZNPP are Westinghouse four-loop pressurized water reactors (PWRs) with large, dry containments. The MELCOR nodalization of the plant can be seen in Figure 4. 47, with a diagram of the core nodalization in Figure 4. 48. The plant nodalization is split into two loops. The first loop represents the single loop in the plant with the pressurizer, and the other loop represents a combination of the other three loops of the plant.

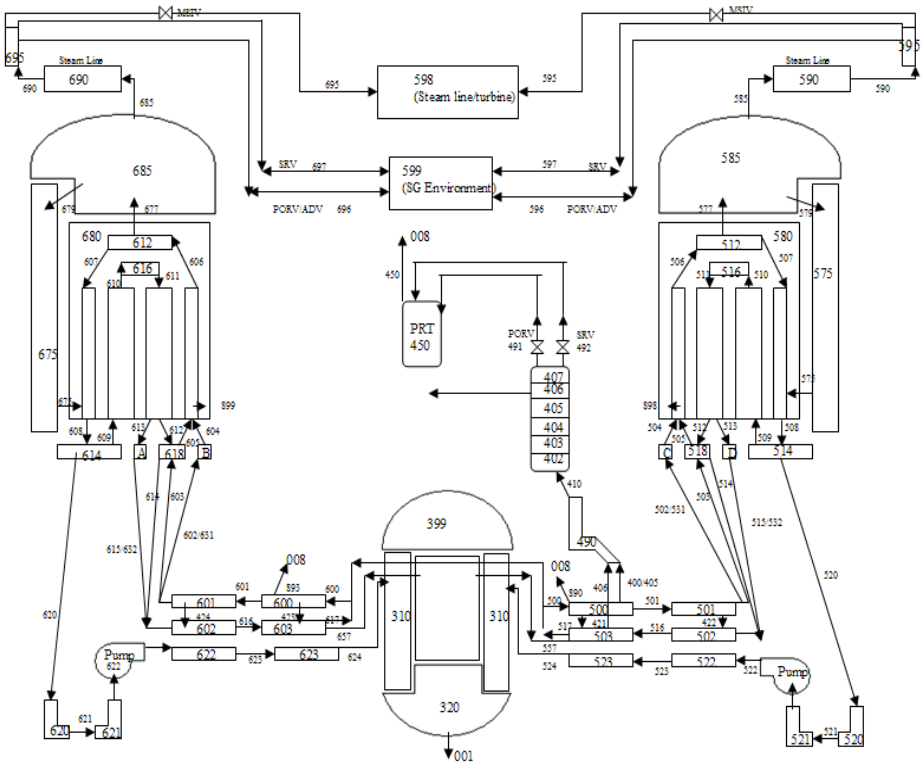

Figure 4. 47: ZNPP MELCOR Nodilization [111]

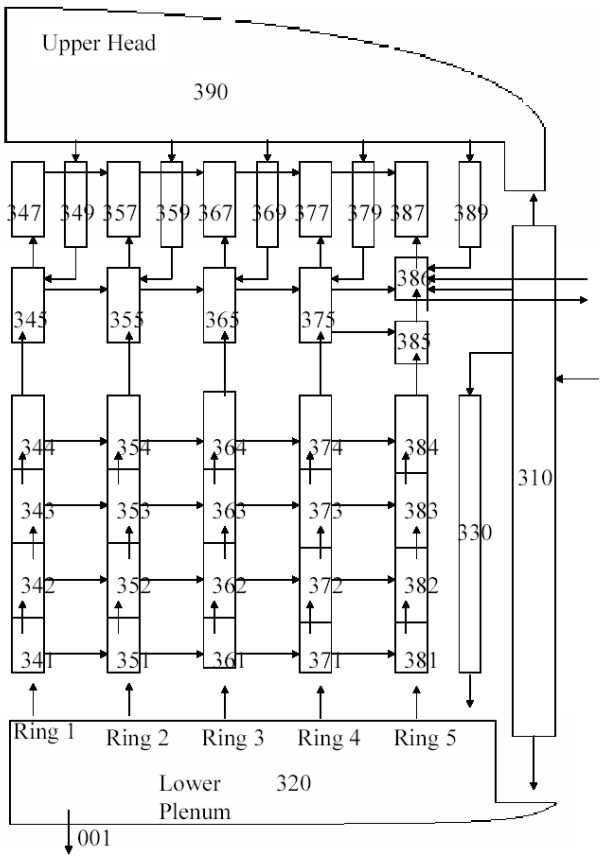

Figure 4. 48: ZNPP MELCOR Core Nodilization [111]

The scenario represents a double-ended guillotine rupture of the pressurizer loop at the reactor coolant pump inlet (node 521 in Figure 4. 47). Following the break, three Emergency Core Cooling Systems (ECCS) should activate: high pressure injection (HPI), which is provided by the charging pumps, intermediate pressure injection (IPI), which is provided by the safety injection pumps, and low pressure injection (LPI), which is provided by the residual-heat-removal pumps. However, in this scenario, their flowrates are considered uncertain, and the time of activation of LPI is delayed and uncertain. Table 4. 23 Contains a full list of uncertainties.

Table 4. 23: MELCOR LOCA Analysis Uncertainties

	Uncertainty	Distribution*
1	HPI Flowrate	Beta(2,5)
2	IPI Flowrate	Beta(2,5)
3	LPI Flowrate	Beta(2,5)
4	LPI Activation Time	Uniform(300,1100)
5	Decay Heat Multiplier	Normal(0.0,2.57)
6	Accumulator Temperature	Uniform(3250,3350)
7	Accumulator Pressure	Uniform(0.0706,0.0716)
8	Accumulator Volume	Uniform(24.07,26.07)
9	Refueling Water Storage Tank Volume	Uniform(3150,3250)
10	Reactor Power	Uniform(3.25e9,3.35e9)

*Many of the uncertainties are not the distribution of the actual parameter, but of a scaling factor or part of a larger formula

The break occurs at time 0 sec, with LPI activation occurring anywhere from 300 to 1100 seconds after. The analysis ends shortly after the activation of LPI, since even its minimum flow condition in this experiment is sufficient to temporarily cool the core.

The output of interest is again the PCT of the core, which is compared to the NRC limit of 2200°F [73]. Due to the long run-times of the MELCOR analysis, only CMC-OS, asymptotic CMC, and rLHS were evaluated. First, the "true" 0.95-quantile of the system was calculated using a 5,000-run CMC experiment. This returned a "true" 0.95-quantile of 1293.16°F. The empirical CDF in Figure 4. 49 shows the shape of the distribution. What is interesting to note, from this figure, is the slope of the distribution near the higher quantiles. While the slope is fairly constant until the 0.90-quantile, it quickly steepens, and there is almost a 1000°F range between the 0.90- and 0.99-quantile. This sensitivity is due to the heat released from zirconium oxidation, which increases exponentially with temperature.

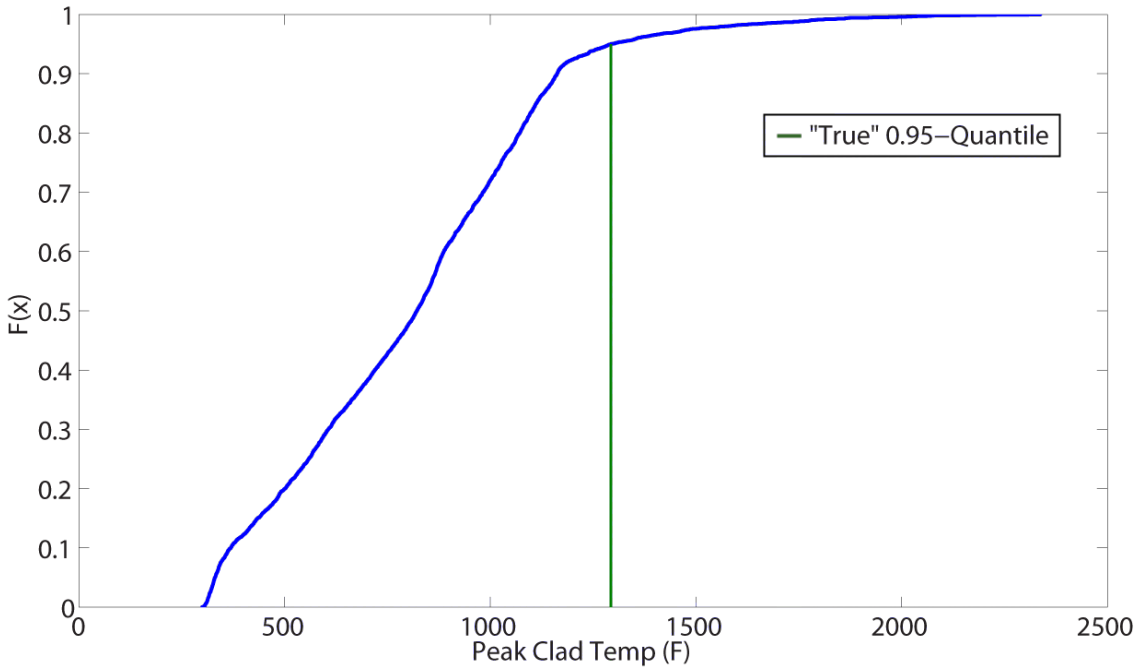

Figure 4. 49: Empirical CDF of Peak Clad Temp. – 5,000 Runs

Due to the time burden when running a large, complex code like MELCOR, unlike the previous examples, 10^4 independent trials of each method could not be performed. Instead, for CMC-OS and asymptotic CMC, a large 5000-run trial was conducted. For the analysis here, a random number of the 5000 outputs would be drawn as a trial output. For example, at $n = 59$, 59 of the 5000 outputs were chosen at random, and treated as a separate trial. This process was repeated for 10^4 trials. Obviously, this introduces some correlation in the results, since the same output value will be used more than once. However, the likelihood of pulling the exact same 59 outputs is extremely small. Using 5000 outputs, this means there are $8.88x10^{137}$ different combinations of 59 runs possible. The repeated use of trials has a bigger effect on CMC-OS than asymptotic CMC due to the way the CIs are calculated. Using asymptotic CMC, the sample variance

is used to calculate the CI, meaning many of the output results are used in the calculation. However, using CMC-OS, simply the highest value of the 59 outputs is taken as the 95/95 value. This means the highest values of the 5000 outputs will be repeatedly chosen as the 95/95 values. This could influence the results, and more will be said on this topic.

For rLHS, 500 cases of 10 LHS runs were conducted. If the desired run level was 60, then 6 of these 500 cases would be chosen at random in order to produce a trial. For rLHS, the number of combinations is smaller at $2.11x10^{13}$ since the choice is 6 out of 500, instead of 59 out of 5000. Like asymptotic CMC, since the CI for rLHS found by calculating a sample variance, the effect of repeating results is less than with CMC-OS.

The results for 10^4 trials of CMC-OS and asymptotic CMC at 59 runs, and rLHS at 60 runs ($t = 10, m = 6$) are shown in Table 4. 24. Here, the coverage and "percent below true" results appear to show that the asymptotic methods are, at the very least, close to convergence, with coverage values at ~90% and "percent below true" near 5%. Also, rLHS results in a 95/95 value that is, on average, about 150°F closer to the actual 0.95-quantile than those 95/95 values found with CMC-OS, with a mean of the rLHS 95/95 values of 1590.53°F compared to 1740.14°F when using CMC-OS. This could mean a very significant gain in margin for an operating power plant. The values for the standard deviation should be viewed with caution since, as mentioned above, by using repeated output values, certain results may occur multiple times (especially in the case of CMC-OS) and influence the spread of 95/95 values.

Table 4. 24: MELCOR - Comparison of 95/95 Values for 10^4 Trials – 59 Runs

	6x10 rLHS	59 CMC	59 CMC-OS
Mean of 10^4 95/95 Values	1590.53	1668.42	1740.14
S.D. of 10^4 95/95 Values	197.13	253.26	275.38
% Below "true"	5.67%	5.01%	5.10%
Coverage	89.67%	87.53%	

In order to confirm that the repeated use of output values did not drastically sway the results, another analysis was performed, but without using the same output result more than once. This was done by performing only 80 trials. Since 5000 CMC runs were performed, $59 * 80 = 4720$, which meant that output results could be chosen without repetition (the values were chosen without replacement). For rLHS, $6 * 80 = 480$, which is less than the 500 cases conducted. So again, this would prevent output results from being used multiple times. These results are in Table 4. 25. As the table shows, the results are nearly identical to the repeated trial results in Table 4. 24 (since only 80 trials are being performed, the statistical sample is not that large, so the methods may not have exactly 5% of trials "below true"). So it appears that the repeated trial results from above are accurate, and that the gain when using rLHS is real.

Table 4. 25: MELCOR - Comparison of 95/95 Values for *80 Trials* – 59 Runs

	6x10 rLHS	59 CMC	59 CMC-OS
Mean of 80 95/95 Values	1580.32	1665.52	1742.31
S.D. of 80 95/95 Values	190.65	262.26	270.82
% Below "true"	6.10%	3.80%	6.32%
Coverage	91.46%	91.14%	

This analysis was repeated for the $n = 93$ run level. Table 4. 26 shows the results for 10^4 trials of 93 CMC-OS and asymptotic CMC runs, and 90 rLHS runs ($t = 10$, $m = 9$), using the repeated trial method described above. While CMC-OS at this run level shows a substantial improvement in accuracy compared to the $n = 59$ run level, with over a 100°F reduction in margin, it still results in a 95/95 value that, on average, is approximately 100°F higher than the resulting value when using rLHS.

Table 4. 26: MELCOR - Comparison of 95/95 Values for 10^4 Trials – 93 Runs

	9x10 rLHS	93 CMC	93 CMC-OS
Mean of 10^4 95/95 Values	1527.21	1559.53	1622.69
S.D. of 10^4 95/95 Values	159.92	191.33	209.85
% Below "true"	5.91%	6.01%	4.92%
Coverage	89.42%	89.64%	

Once again, this run level was examined without using repeated output values, by conducting only 50 trials. The results for this experiment are shown in Table 4. 27, and like the previous example, they show very little variation from the repeated trial results (again, the important value is the mean, since the statistical sample for the "% below true" is small).

Table 4. 27: MELCOR - Comparison of 95/95 Values for *50 Trials* – 93 Runs

	9x10 rLHS	93 CMC	93 CMC-OS
Mean of 50 95/95 Values	1524.71	1552.93	1623.14
S.D. of 50 95/95 Values	163.16	194.70	218.62
% Below "true"	4.00%	8.00%	6.00%
Coverage	92.00%	88.00%	

These results show a potentially large improvement in accuracy by using rLHS instead of CMC-OS. This is caused, in part, by the shape of the output distribution. As Figure 4. 49 showed, the higher quantiles of the output distribution spanned over 1000°F. Since CMC-OS only uses the top output result, or the second highest result when $n = 93$, it could be choosing values which are substantially higher than the true 0.95-quantile. Figure 4. 7 in Section 4.2.1 showed that CMC-OS is more likely to return a 95/95 value near the 0.99- or 1.0-quantile at the $n = 59$ run level. Since those higher quantiles are far from the 0.95-quantile in this example, CMC-OS induces a large amount of excess conservatism. This situation is avoided using the asymptotic methods, since the quantile is estimated directly, and the CIs are calculated using a sample variance which takes into account more than one point of the output samples. This leaves these methods less vulnerable to one or two very high output values.

4.4. Discussion

These experiments indicate that rLHS can provide more accurate and precise confidence intervals for quantiles than CMC-OS. This would mean a reduction in the probability of both Type-I and Type-II errors. However, the rLHS method is not without its faults. As several results showed, at low run levels, the method may not have converged. This can result in too many trials resulting in a 95/95 or 95/75 value falling below the actual quantile. An interesting point though is that even when this did occur, the rLHS trial results still did not fall as far below the "true" quantile as some CMC-OS trials. So it is not possible to say whether this would result in more Type-I errors than

CMC-OS without knowing the actual location of the safety limit, which will be examined in more detail in Section 5.

There may be ways to help resolve the convergence issue. Additional experiments on different types of systems can lead to more guidance about the proper selection of the parameters of the derivative estimator. Also, it is possible to improve coverage of the constructed rLHS confidence interval by replacing the normal critical point with a critical point from a Student-t distribution with m-1 degrees of freedom, where m is the number of LHS cases. Since the Student-t distribution has somewhat heavier tails than a normal distribution, this results in slightly wider and more conservative CIs. It may help to ensure that the number of trials falling below the true quantile does not exceed 5%, but this will also reduce the accuracy. Lastly, as explained in Section 4.2.2.4, it appears that conducting more cases of a smaller size (increase m, decrease t) aids in the convergence, since the validity of the CLT requires that the number m of cases grows large.

Specifically, more work should be conducted on how to increase the bandwidth of the CFD at very low run levels. As several examples showed, the derivative estimation actually got worse as the number of runs increased because the bandwidth parameters were changing. One interesting note on this point though is that a better derivative estimation at low run levels does not necessarily mean better coverage. As the derivative estimation improves, the width of the CI decreases. This means the quantile estimation plays a bigger role. If the quantile estimation has not converged, which can be the case at low run levels, then it will dominate the error and disrupt the coverage level. Even though

this quantile estimation error is present when the derivative is overestimated, the increased width of the CI tends to negate the errors caused by the quantile estimation.

Finally, it should be noted that this work in no way challenges the validity of CMC-OS for the calculation of confidence intervals for quantiles. Conversely, all the experiments carried out here demonstrated that the results of the CMC-OS did have ~95% confidence of exceeding the desired quantile. However, when estimating a 0.95-quantile, CMC-OS is vulnerable to returning a 95/95 value which considerably overestimates the true quantile when the output distribution has a fatter tail at these higher quantiles. This is a result of CMC-OS only using a single output value to derive a 95/95 value. The further the extremes of the output distribution (i.e. the 0.99-quantile) are from the 0.95-quantile, the greater the probability that the CMC-OS 95/95 value will be a greater distance from the 0.95-quantile.

4.4.1. Applicability to OLHC

Given the performance benefits of OLHCs, shown in Section 3, it would be assumed that the use of OLHCs when establishing CIs for quantiles would also be superior to ordinary LHS. However, this was not the case. This section details the experiments conducted, and offers explanations about the cause of the error.

4.4.1.1. Experiments

As with the other methods detailed at the beginning of Section 4, the use of OLHCs, using the CI technique in Section 4.2.2.3, was tried on the several representative systems. Upon starting these tests, problems appeared with the outputs of the OLHC's

analyses. As run sizes grew, the confidence levels were not converging properly. Table 4. 28 shows a comparison of the results when using OLHCs to ordinary rLHS for the nonlinear equation with normal inputs (the rLHS results are the same as in Table 4. 6).

Table 4. 28: Comparison of rLHS and OLHC Results

n	**rLHS**			**OLHC**	
	t=10	t=20	t=30	t=16	t=32
	50.23	48.56		45.12	**Mean***
	8.56	7.43		8.32	**S.D.***
59	5.40	5.40		32.78	**% Below**
	93.31	93.53		42.52	**Covg.**
	269.15	277.06		299.21	**Avg. $\bar{\lambda}_p$**
	47.51	45.56	45.92	46.21	47.61
	5.10	4.32	3.92	3.96	3.73
93	4.58	8.51	4.63	3.81	0.44
	92.16	88.95	89.40	88.97	84.22
	213.45	213.03	218.92	237.31	248.06
	46.38	45.26	45.09	45.16	46.62
	4.06	3.38	3.33	3.23	3.21
124	4.15	5.17	5.35	4.31	0.27
	94.64	92.35	92.18	92.27	91.14
	233.71	236.20	238.26	259.08	269.03
	43.76	42.91	42.92	42.63	43.72
	1.95	1.47	1.44	1.48	123
311	3.63	5.37	4.81	7.71	0.29
	93.48	91.69	91.97	90.42	81.97
	190.81	186.29	189.13	198.12	180.88
	42.77	42.31	42.23	41.87	43.23
	1.34	1.08	1.02	1.04	0.89
548	4.30	5.34	5.40	11.36	0.06
	92.20	91.45	91.07	86.95	67.37
	174.75	175.02	173.91	179.09	165.49
	42.11	41.83	41.76	41.41	42.77
	0.95	0.77	0.73	0.74	0.63
1008	5.39	5.89	5.58	14.83	0.02
	90.67	90.18	90.35	83.79	43.32
	165.82	165.59	165.29	168.15	153.35
	41.62	41.48	41.49	41.04	42.53
	0.66	0.54	0.52	0.51	0.44
2004	6.15	5.75	4.60	22.03	0.01
	89.75	89.76	90.16	77.31	12.35
	161.92	162.65	161.87	165.30	150.86

* Mean and S.D. of the 10^4 95/95 Values

As the results show, OLHC clearly does not converge to the correct coverage levels, with a coverage level of ~77% at $n = 2004$ when $t = 16$ for OLHC, and a coverage level of about 12% when $t = 32$. Also, the trends are not even consistent between the different run sizes. When $t = 16$, the percent of trial below true errs to the low side as the number of runs grows with an increasing percentage falling below the true quantile (22% of trials when $n = 2004$), but when $t = 32$, the percent below true errs on the high side (only 0.01% of trials when $n = 2004$). Even when the OLHC are close to the correct coverage levels, the results tend to be worse than with ordinary rLHS.

The use of OLHCs was shown in Section 3 to establish more accurate and precise quantile estimations than LHS on the same system, and the CIs are based on these quantile estimations. It would seem that OLHCs should provide more accurate and precise CIs also, but the results show the opposite. There are several possible explanations for this phenomenon. The next step was to investigate why this was the case.

4.4.1.2. Analysis of Error

First, as mentioned in Section 3, the quantile estimation method used here has not been proven for OLHCs. Since this has not been proven, OLHCs no longer fall under the CI proof in Section 4.2.2.3. However, it seems unlikely that this alone is the reason for the error. As the Section 3 results showed, OLHCs were better at estimating the quantile values than ordinary LHS. So even if it has not been mathematically proven, heuristically, the quantile estimation method appears to work fine for OLHCs.

The second possible reason for the OLHCs lack of convergence could be with how the actual experiments were carried out. The OLHCs used in the experiments were based off a larger OLHC design. For example, the 16-run OLHCs were created using a 16- run OLHC that could handle up to 12 possible inputs. This means it had 12 columns. Since only four inputs were needed for the nonlinear equation, the order of the columns could be changed to create a new 4-input OLHC design. The number of possible designs can be calculated using Eq. 85,

$$Number\ of\ permutations = \frac{n!}{(n-r)!} \qquad \text{Eq. 85}$$

where n is the number to choose from, and r is the number chosen, as explained in Section 3. This means there was a total of 11880 possible permutations of the 4-input, 16-run OLHC design. This may seem like a large number, but remember, each trial consisted of multiple OLHC cases, each with its own design. Then 10^4 trials were conducted, so the number of designs adds up fast. Take for example the 16-run OLHC experiment at the 548-run level. This means each trial consisted of 34 cases ($548/16 = 34.25$). So 34 OLHC designs were used for each trial, and then this was repeated for 10^4 trials. This means, in total, there were 340,000 OLHC designs used. Obviously designs were repeated, since there were only 11880 permutations possible. Since the designs were repeated, it is possible that the resulting values could be biased towards those particular designs. An experiment was conducted to see if this was the case.

If it is true that the only reason the coverage levels are not correct is because of the repeated OLHC designs, then this problem should not exist if less trials are conducted and run designs are not repeated. To test this, the 16-run OLHC experiment from Table 4.

28 was repeated for the 548-run level. This time, a variety of trial sizes was used, starting with only 100 trials. At the 548-run level, there were 34 cases per trial, so this means a total of 3400 OLHC designs were used, which is far below the maximum of 11880. This was then repeated for 200, 500, 1000, 5000, 10000, and 50000 trials. If the repeated OLHC designs are biasing the result, the coverage level should be correct at the lower trial levels, and then get worse as the number of trials increases. This test was also conducted for a 32-run OLHC design at the 311- and 548-run level, in order to make sure the results were consistent. These results can be found in Figure 4. 50.

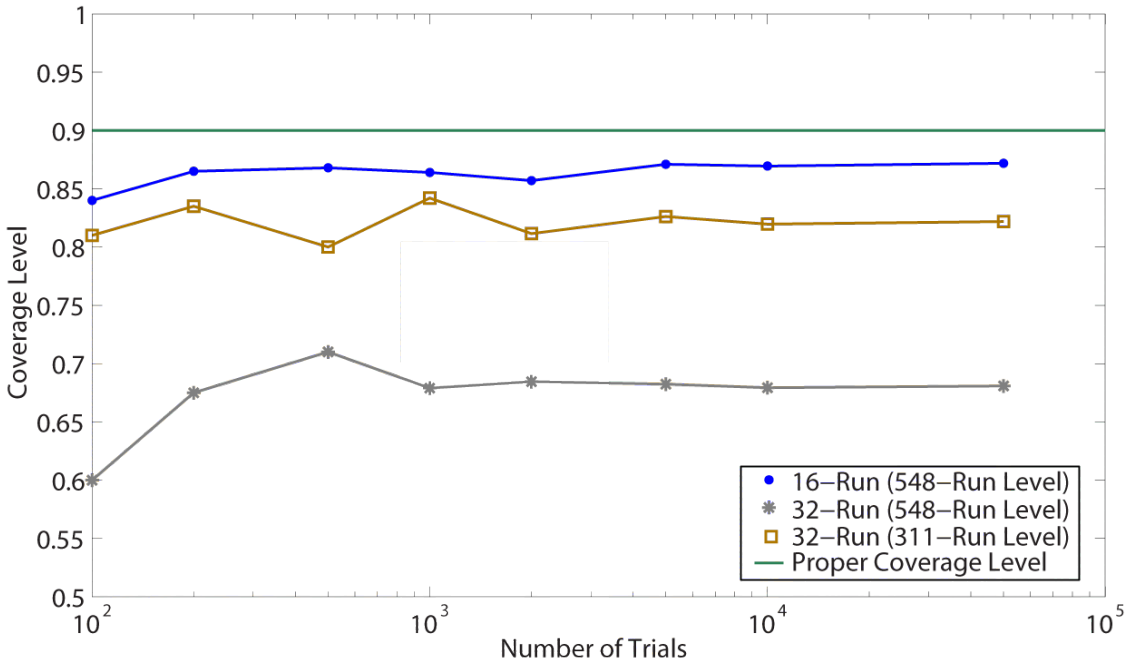

Figure 4. 50: Coverage Level with Differing Trial Numbers

As the figure shows, the coverage levels do not start out at 0.90, and then get worse. Instead, they remain fairly consistent. This would seem to be a definitive answer. The

repeated OLHC designs appear to have no effect on the incorrect coverage levels, since the coverage level is still incorrect even when the designs are not repeated many times. There must be a separate cause for the incorrect coverage levels.

The third possible explanation has to do with how CIs are established on LHS designs. Recall from Section 4.2.2.3 that since the results of a LHS design are not i.i.d., multiple cases of LHS are used since those multiple LHS designs are i.i.d. because they are created randomly. There is a difference here between ordinary rLHS and OLHCs. As explained in the previous paragraph, the OLHC designs are constructed from random permutations of a larger OLHC design. This could mean that the results of the OLHCs are not truly i.i.d. If they are not i.i.d., then the resulting values are correlated in some fashion which violates the derivation in Section 4.2.2.3. More analysis was conducted to see if this could be the cause.

As shown in Section 3, it is already known that OLHCs provide a more accurate and precise quantile estimation than ordinary rLHS when using the nonlinear equation with normal inputs. However, in Section 3, the quantile estimation was made using a single LHS or OLHC case. This is not how the quantile is estimated for the CI. In Eq. 81, the quantile is estimated using all the runs from the *m* number of cases. So even if OLHCs provided a better estimation in a single case, they may not provide a better estimation when combining many cases. A quick comparison of the empirical CDFs created from combining these cases showed a potential issue. Figure 4. 51 shows this comparison for rLHS and OLHC trial at the 1008-run level.

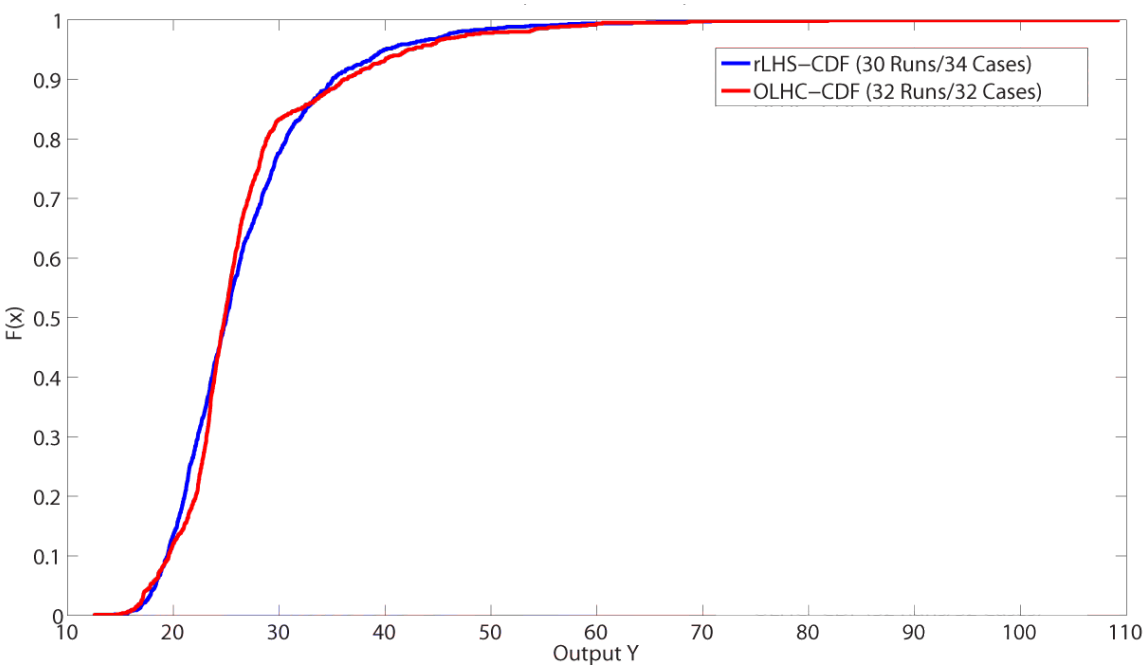

Figure 4. 51: Combined Empirical CDF Comparison

As the figure shows, ordinary rLHS creates a smooth empirical CDF curve, while the OLHC design creates a curve with several bends or knees. This results in a different 0.95-quantile estimation for the OLHC method than the rLHS method. To see if this trial was unique or represented a trend, the quantile estimations over 10^4 trials was recorded for each method at each run level. These results can be seen in Figure 4. 52, which shows the average quantile estimation, over all cases, for 10^4 trials.

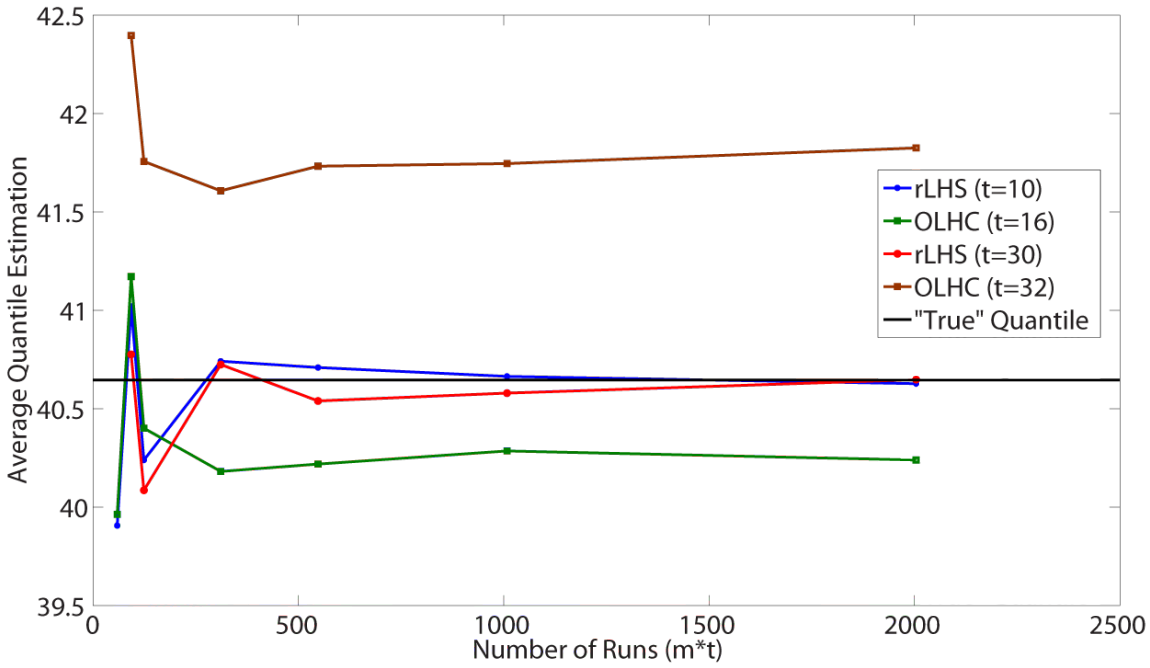

Figure 4. 52: Quantile Estimation Convergence Comparison

As the figure shows, even though the two rLHS methods fluctuate some at lower levels, they converge to the true quantile. However, the OLHC trials do not. They converge to values above and below the true quantile (which explains why, in Table 4. 28, one errored to the high side, and one erred to the low side). Clearly, even though OLHCs performed better when estimating a quantile using a single case, they will incorrectly estimate a quantile when combining cases. Why this is the case is not yet exactly known, but as explained above, it is most likely related to the OLHCs not being completely independent and identically distributed. More work is needed to see if this is in fact the only cause of error, and if so, whether there is a method that can be used to establish CIs with OLHCs.

4.4.2. Application to Risk-Informed Safety Margin Characterization

As mentioned in Section 2.1.1, there has been recent work investigating the use of a Risk-Informed Safety Margin Characterization (RISMC), depicted in Figure 2. 3. While this may seem like a radically different approach to safety margin calculation, the methods used to demonstrate adherence to the goal may not change. As shown in Figure 2. 3, the RISMC is a comparison between a capacity curve and a load curve. The most straightforward way to estimate the risk of failure would be to calculate an *overlap coefficient* (OVL) [112]. The OVL is an indicator of how much the range of the two distributions overlap. While recent work has been done in this field to investigate techniques to establish CIs [113], the fact that the distributions in this analysis are empirically derived and not standard distributions, like a normal or exponential, makes finding this coefficient more complicated. Other work has focused on nonparametrics [114][115], but the problem here remains difficult since the comparison is between the extremes of the two distributions, which are not easy to characterize without many runs being conducted. Instead, the use of quantiles may accomplish the same goal, but in a way that is easier to implement.

For example, if physical or computational experiments were conducted to determine the capacity of the system, this would result in a distribution similar to the one found in Figure 4. 53 (although it is unlikely to resemble a normal distribution).

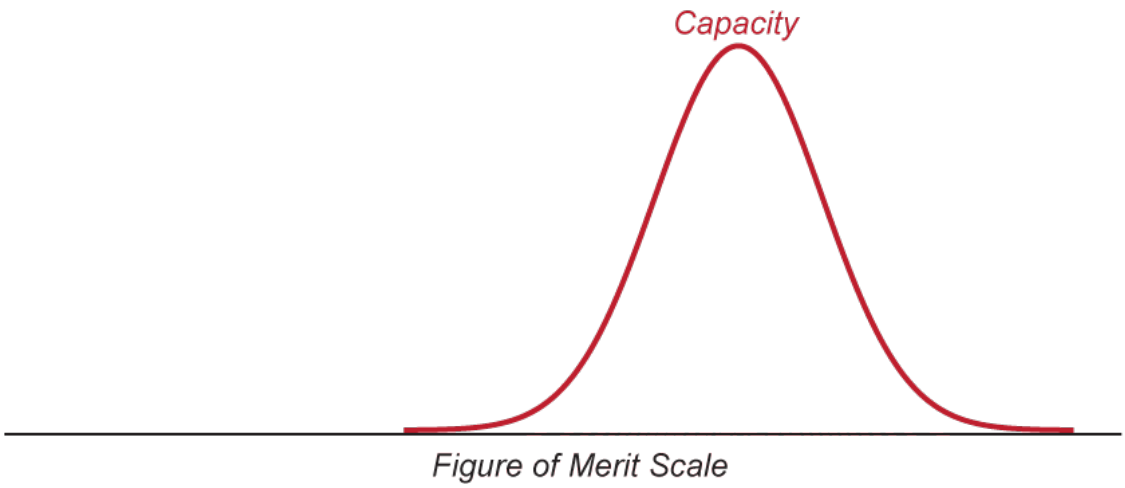

Figure 4. 53: Capacity Distribution

From there, the rule-making body, which is this case is the regulator, could specifying the amount of overlap between this curve and the load curve in different ways.

The first, and easiest, possible comparison is to choose a low quantile of the capacity curve, and a high quantile of the load curve, and check to make sure the low quantile of the capacity curve is at a higher value. Since this is done using empirical data, OSCIs can be used in place of the direct quantile estimation. Figure 4. 54 shows how a OSCI (or credible interval if the data is from many sources) could be found for the low quantile (in this case the 0.05-quantile) of the capacity curve.

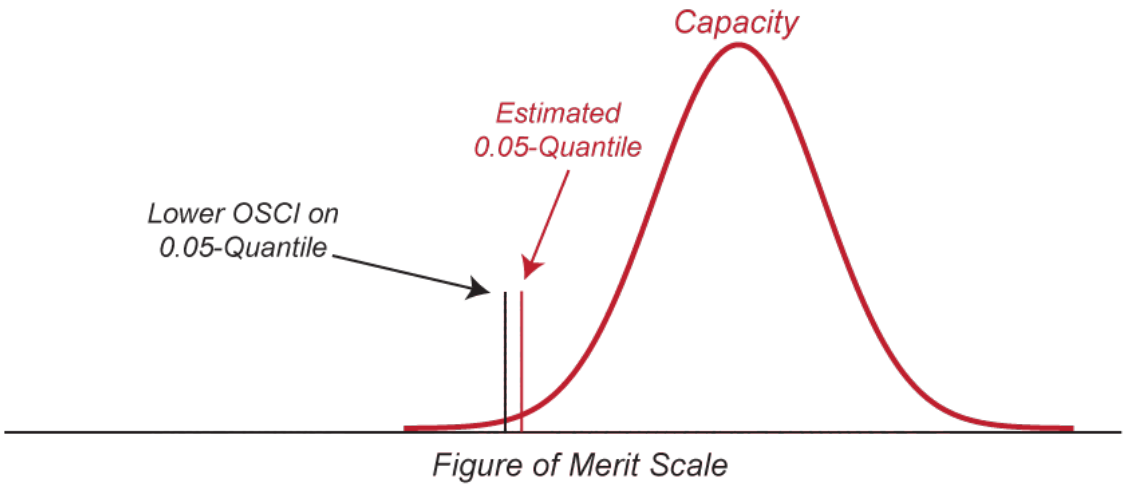

Figure 4. 54: Capacity Distribution with Example Limit

Then this value could be compared to an upper OSCI for a high quantile on the load distribution, which would be found by a utility, as seen in Figure 4. 55.

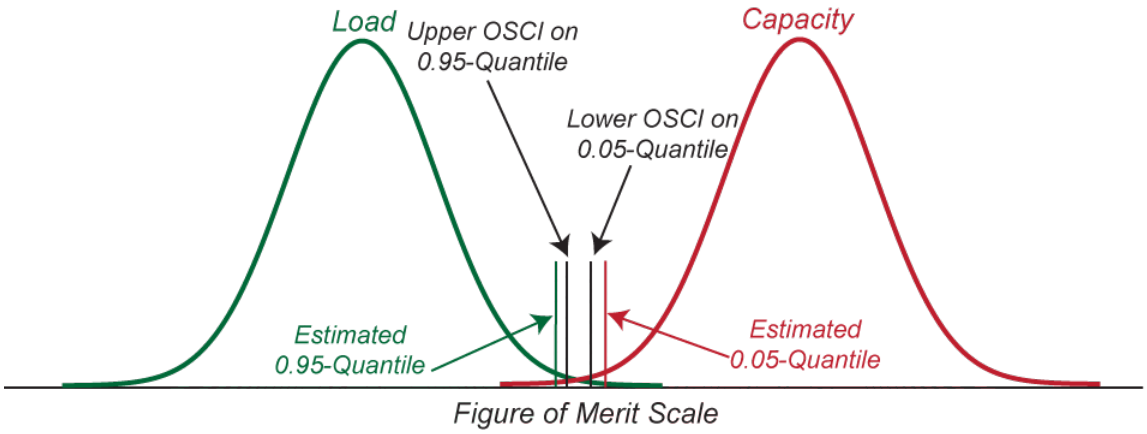

Figure 4. 55: Capacity versus Load with OSCI's

If the upper OSCI of the 0.95-quantile of the load is below the lower OSCI of the 0.05-quantile of the capacity, then the system passes the test. This comparison is essentially

the same as placing a limit on the amount of overlap between the two curves. However, unlike the OVL, estimating these quantiles does not require the extremes of the two distributions to be characterized in great detail.

The second possible technique would be to specify a low quantile on the capacity curve, as before, and then attempt to calculate how much of the load curve is above that limit. Figure 4. 56 shows what this would look like.

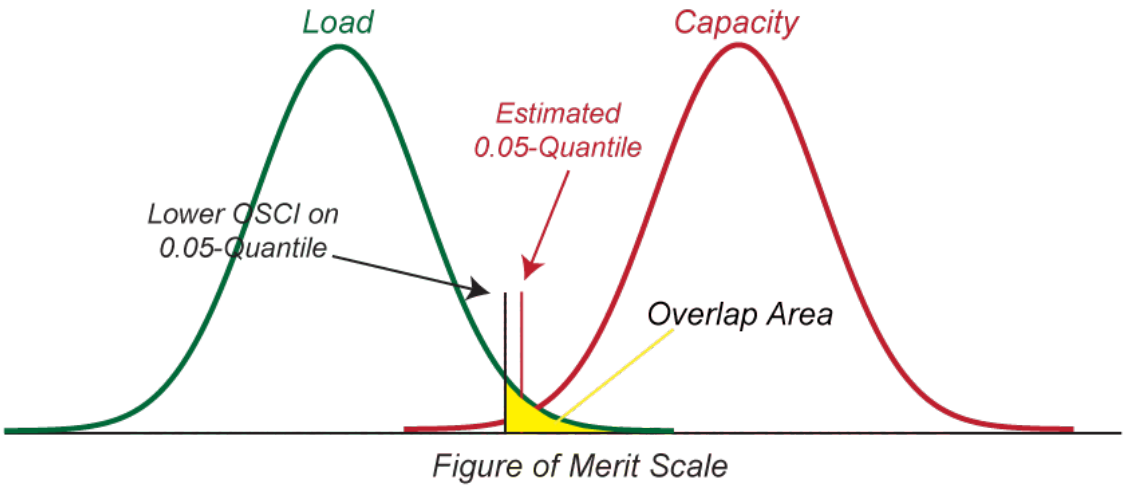

Figure 4. 56: Comparison with Low Quantile and Overlap

This would be done by interpolating the quantile of this value on the load distribution. For example, the limit value (the OSCI for the 0.05-quantile) may fall between the 0.93- and 0.94-quantiles of the load distribution. A limit could be placed on how much overlap would be permitted. This is easier than a direct calculation of the overlap of the two distributions because the comparison is made to a single value. However, this procedure becomes essentially the same as the comparison of quantiles in the previous paragraph.

The only difference is that the amount of overlap is computed then compared to the limit on that amount, and the previous method uses the maximum amount of overlap (which in that case was the 0.95-quantile) and compares it directly, but the end result would be the same.

Lastly, the previous method could be restated in terms of probability. This method is investigated in detail in Section 5. As will be shown, this method will also directly find the probability of overlap between the values, and CIs can be established on this probability.

4.4.3. Adding Cases and Possibility of Error

When using the rLHS method, multiple cases of LHS runs are performed. Unlike CMC-OS, there is no set number of total runs that needs to be conducted. This means, using rLHS, an analyst would perform some number of LHS cases, and then analyze the results to find a OSCI and compare it to a regulatory limit. However, an analyst could always add another LHS case to the results and reanalyze the results. This may be a cause of concern among some regulators, who think that analysts will try to "game" the system, or in other words, if the initial conclusion is not the desired result, they will continue to take LHS cases in the hopes that the next result will be more favorable. Obviously, this is possible, but how large is the danger associated with it?

To test this, an experiment was conducted using the nonlinear equation with normal inputs (described in Section 3.2.2), in order to discover the consequences of adding additional LHS cases. An initial trial was conducted of $m = 6$ cases of $t = 10$ runs (60 runs total). Using this data, a 95/95 value was found, and compared to a pseudo

limit value placed at 40.5, which is just below the 0.95-quantile of 40.646. The limit value being placed here meant that the system should not pass the test since the limit is below the 0.95-quantile. This is a very challenging experiment since the limit and 0.95-quantile are so close. From there, the conclusion of pass or fail would be recorded, and then an additional LHS case of 10 runs would be added to the results (70 runs total), and the conclusion would be analyzed again. From there, the amount of times the correct conclusion was reached at each step could be compared.

First, Table 4. 29 shows the conclusion probability results for this analysis for 10^4 trials. Here, the likelihood of getting the correct conclusion increases as the extra LHS case is added, which is to be expected.

Table 4. 29: Conclusion Probabilities for 10^4 Trials

Conclusion	After 60 runs (m=6, t=10)	After additional 10 runs (70 total)
Fail Test	9525	9730
Pass Test	475	270

The real question is how many of the 9525 trials which appeared to fail the test at the 60 run level, would incorrectly appear to pass the test after the extra cases was added. Table 4. 30 has these results, which show that of the 9525 trials which initially failed the test, only 28 would switch to a pass after the addition of an extra case. This means there was only a 0.3% chance of success using this method to "game the system."

Table 4. 30: Conclusion Probabilities after Initial Failure Conclusion

Conclusion	After 70 runs
Fail Test	9497
Pass Test	28
Total	9525

This probability may seem low, but a better comparison would be against CMC-OS, since it is also possible to conduct additional runs in the hopes of getting a different result. For example, if an analyst conducted 59 runs to determine a 95/95 value, and this value failed against a safety limit, an additional 34 CMC runs could be conducted to achieve the next highest run level, 93. Then the results could be tested again.

The framework from the previous example was repeated, but in this case, 60 rLHS runs were conducted, than an additional three LHS cases of 10 runs were added. This was compared against CMC-OS at 59 runs, then at 93 runs. Table 4. 31 shows the conclusions percentages after each step for 10^4 trials.

Table 4. 31: Comparison of Conclusions between CMC-OS and rLHS for 10^4 Trials

	CMC-OS		**rLHS**	
Conclusion	After 59 Runs	After additional 34 Runs (93 total)	After 60 runs (m=6, t=10)	After additional 30 runs (90 total)
Fail Test	9510	9535	9538	9600
Pass Test	490	465	462	400

As the results show, at ~60 runs, rLHS is slightly more accurate, with 9538 trials resulting in the correct conclusion, compared to 9510 with CMC-OS. However, once

again the question is how many of these trials that initially showed a failure (which is correct), became a "pass" when the additional runs were added. Table 4. 32 shows these results. For CMC-OS, of the original 9510 failure trials, 230 concluded that the system passed after the 34 runs were added. That is ~2.4%. For rLHS, 219 of the 9538 trials went from failing to passing, or ~2.3% (the reason this percentage is higher than the 0.3% when adding only one additional case, as shown in Table 4. 30, is because as more cases are added, the influence of the initial cases is reduced, so the correlation between the original conclusion and the new conclusion becomes smaller). So even though it is possible to try to beat the system using both methods, it appears to not be any more likely to occur with rLHS than with CMC-OS.

Table 4. 32: Conclusion Probabilities at ~90 Runs after Initial Failure Conclusion

Conclusion	CMC-OS	rLHS
Fail Test	9280	9319
Pass Test	230	219
Total	9510	9538

Chapter 5: Quantiles vs. Probability

Even though the current NRC-approved method of satisfying the probability requirement in [74] is to establish confidence intervals for quantiles, there may be alternatives. The most literal interpretation of the NRC requirement would be to establish confidence intervals for a probability, rather than a quantile. This may seem like restating the same thing, but there are differences between the two statements.

By establishing a confidence for a quantile, a statement is being made about the location of that particular parameter of the output distribution. For a safety analysis, this value is found, then used to compare against the safety limit. However, finding a confidence for a probability combines these two steps. Here, the probability of the output exceeding the safety limit is found directly, and then a confidence is found on that probability.

In terms of hypothesis testing, it is similar to the framework laid out in Section 4, but involves the use of a different test statistic. So even using the same data, the probability method (P-method) and quantile method (Q-method) may produce different conclusions. More detail on this hypothesis test, along with methods to find these asymptotic confidence intervals for a probability are described in the next section.

5.1. Methods

As in Section 4.2, suppose there is a system with output *Y* and CDF *F*. Unlike Section 4, the goal here is to estimate the probability of the output *Y* being less than a constant b (the limit value). To state this more rigorously, take θ to be this probability, where $\theta = P(Y \leq b) = F(b)$, and assuming $0 < \theta < 1$. If θ is greater than 0.95, that means there is a greater than 95% probability that the output *Y* is less than the limit value b. In order to satisfy a 95/95 criterion, the goal is to provide a 95% lower OSCI (LOSCI) for θ. So if $(L, +\infty)$ is a 95% LOSCI for θ, the 95/95 criterion is satisfied if $L \geq 0.95$. This means there is a 95% confidence that there is at least a 95% probability that the limit value will be greater than the output of the system. The derivation of the methods in Sections 5.1.1, 5.1.2, and 5.1.3 are summarized from [116].

5.1.1. CMC using Probability Method

This task can be accomplished using CMC since $\theta = E[I(Y \leq b)]$, so it can be estimated using a sample average, and the LOSCI is the same as when estimating the mean. If $Y_1, Y_2, \ldots, Y_n$ are i.i.d. samples from the CDF F, as in Sections 3 and Section 4, let $V_i = I(Y_i \leq b)$, so $V_1, V_2, \ldots, V_n$ are also i.i.d., with mean θ. Therefore, the sample average $\bar{V}_n$ can be found using Eq. 86.

$$\bar{V}_n = \frac{1}{n}\sum_{i=1}^{n} V_i \qquad \text{Eq. 86}$$

Then the sample variance S_n^2 can be found using Eq. 87.

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} (V_i - \bar{V}_n)^2 \qquad \text{Eq. 87}$$

The LOSCI for θ can be found in Eq. 88,

$$\left(\bar{V}_n - z\frac{S_n}{\sqrt{n}}, +\infty\right) \qquad \text{Eq. 88}$$

where z is the standard normal critical point for the confidence level desired, as before. This method is asymptotically valid since the sample average $\bar{V}_n$ satisfies the CLT in Eq. 89,

$$\sqrt{n}(\bar{V}_n - \theta) \Longrightarrow N(0, \sigma^2) \qquad \text{Eq. 89}$$

as $n \to \infty$, and the sample standard deviation $S_n \Longrightarrow \sigma$ as $n \to \infty$.

As in the previous sections of this work, an example of how this technique can be implemented in a computer code can be found in Figure 5. 1, where *limit* is the safety limit, and *NN* is the normal critical point for the desired confidence level. This code compares each output value to the limit value, in order to calculate V_i, called *V*, and $\bar{V}_n$, called *V_bar*. These values are then used to find the root of the sample variance S_n, called *Sn*. Finally, the probability plus confidence is found.

```
%%% CMC P-Method
    V=(Y <= goal);                          % V Calculation using Indicator Fun.
    V_bar=mean(V);                          % V_bar calculation
    Sn=sqrt((1/(n-1))*sum((V-V_bar).^2));% Sn Calculation
    prob_w_conf=V_bar-NN*(Sn/sqrt(n));    % Final Prob. with Confidence
```

Figure 5. 1: MATLAB Code Implementation of CMC Probability Method

5.1.2. rLHS using Probability Method

As in Section 4, a different method must be used to find a confidence interval for a probability using LHS. Once again, this is due to the fact the outputs of the LHS design

are not i.i.d. Therefore, a new approach is needed to find a sample average, which can be used to satisfy a CLT. As before, this is done by taking multiple cases *m*, each with *t* number of runs. Since the *m* cases are generated independently, their outputs can used to create a sample average.

Here, Eq. 90 shows how the average value for V can be found for one LHS case. The nomenclature is the same as before, where $Y_{i,j}$ is the *i*th output from the *j*th LHS case, and $V_{i,j} = I(Y_{i,j} \leq b)$.

$$W_j(t) = \frac{1}{t}\sum_{i=1}^{t} V_{i,j} \qquad \text{Eq. 90}$$

Then Eq. 91 shows the sample average of the case averages $W_1(t), W_2(t), \ldots, W_m(t)$, which are i.i.d.

$$\bar{W}_{m,t} = \frac{1}{m}\sum_{j=1}^{m} W_j(t) \qquad \text{Eq. 91}$$

Eq. 90 and Eq. 91 can be simplified to Eq. 92, which shows that $\bar{W}_{m,t}$ is the sample average over all $V_{i,j}$.

$$\bar{W}_{m,t} = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{t}\sum_{i=1}^{t} V_{i,j} = \frac{1}{mt}\sum_{j=1}^{m}\sum_{i=1}^{t} V_{i,j} \qquad \text{Eq. 92}$$

Since each $Y_{i,j} \sim F$, the expected value of $\bar{W}_{m,t}$ is still θ, as with the CMC example. This can be seen in Eq. 93.

$$E[\bar{W}_{m,t}] = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{t}\sum_{i=1}^{t} E[V_{i,j}] = \frac{1}{mt}\sum_{j=1}^{m}\sum_{i=1}^{t} E[I(V_{i,j} \leq b)] = \theta \qquad \text{Eq. 93}$$

The sample variance $S_{m,t}^2$ can be found using the sample average $\overline{W}_{m,t}$ and the value of $W_j(t)$ from each of the *m* cases, as seen in Eq. 94.

$$S_{m,t}^2 = \frac{1}{m-1}\sum_{j=1}^{m}(W_j(t) - \bar{W}_{m,t})^2 \quad \text{Eq. 94}$$

From there, a CLT can be satisfied with either $m \to \infty$ with *t* remaining fixed, or with $t \to \infty$ as *m* remains fixed. Practically, it is easier to implement the method with *t* remaining fixed (since the LHS design must be made with the value for *t* known beforehand), so that will be the method detailed here. Since $0 \leq V_{i,j} \leq 1$, then $0 \leq W_j(t) \leq 1$, as seen in Eq. 90. Also, since $W_j(t)$ is bounded, it must have finite variance. Since $W_1(t), W_2(t), \ldots, W_m(t)$ are i.i.d., the CLT in Eq. 95 holds

$$\sqrt{m}(\overline{W}_{m,t} - \theta) \Longrightarrow N(0, \sigma_t^2) \quad \text{Eq. 95}$$

as $m \to \infty$ with *t* fixed. Since $0 < \theta < 1$ is assumed, it ensures that σ_t^2, or the sample variance $S_{m,t}^2$ is strictly positive. Then an asymptotically valid LOSCI can be found in Eq. 96,

$$\left(\overline{W}_{m,t} - z\frac{S_{m,t}}{\sqrt{m}}, +\infty\right) \quad \text{Eq. 96}$$

as $m \to \infty$ with fixed *t*, and $S_{m,t} \Longrightarrow \sigma_t$ as $m \to \infty$ with fixed *t*.

This method can be implemented using the following computer code in Figure 5. 2, where again *limit* is the safety limit, and *NN* is the normal critical point for the desired confidence level. Here, the code cycles through each LHS case, calculating $V_{i,j}$, and $W_j(t)$ in Eq. 90, called *V_bar*, since it can also be viewed as the average value of $V_{i,j}$ for

that particular LHS case. This is then used to find $\overline{W}_{m,t}$, called *W_bar*, and the root of the variance $S_{m,t}$, called *Smt*.

```
%%% rLHS P-Method
    for j=1:m                          % For each LHS case
        V(j)=(Y(j,:) <= goal);         % V for the runs in that LHS case
        V_bar(j)=mean(V(j));           % V_bar for that LHS case
    end
    W_bar=(1/m)*sum(V_bar);                        % W_bar Calculation
    Smt=sqrt((1/(m-1))*sum((V_bar-W_bar).^2)); % Smt Calculation
    prob_w_conf=W_bar-NN*(Smt/sqrt(m));        % Final Prob. with Confidence
```

Figure 5. 2: MATLAB Code Implementation of rLHS Probability Method

5.1.3. Probability Test Statistic and Hypothesis Testing

As described at the start of Section 5, using the P-method is another way to perform the hypothesis test described in Section 4. The only difference is the test statistic used in the calculation. Rewording Eq. 89 gives

$$\frac{\sqrt{n}}{\hat{\sigma}}(\bar{V}_n - \theta_b) \approx N(0,1) \qquad \text{Eq. 97}$$

as $n \to \infty$, where θ_b is the probability of the output Y exceeding the limit value b, and $\hat{\sigma}$ is the estimator of the variance of Y, and can be found from Eq. 87. Using the CLT, the LOSCI for θ_b, which will be denoted *L,* would equal

$$L = \bar{V}_n - z\frac{\hat{\sigma}}{\sqrt{n}} \qquad \text{Eq. 98}$$

where z is the standard normal critical value for that confidence level. In order to satisfy the 95% confidence for the 95% probability criterion, $L \geq 0.95$. The hypothesis test choices are straightforward.

H_0: $\theta_b < 0.95$

H_1: $\theta_b \geq 0.95$

The scenario where H_0 is rejected can be described by rearranging Eq. 98,

$$reject\ H_0\ if\ and\ only\ if\ \frac{\bar{V}_n - 0.95}{\hat{\sigma}/\sqrt{n}} \geq z \qquad \text{Eq. 99}$$

which is equivalent to $L \geq 0.95$. Similarly,

$$accept\ H_0\ if\ and\ only\ if\ \frac{\bar{V}_n - 0.95}{\hat{\sigma}/\sqrt{n}} < z \qquad \text{Eq. 100}$$

which is the same as $L < 0.95$.

Obviously, one big drawback with the P-method compared to the Q-method is that there is no information regarding the margin to the limit value in terms of the figure of merit. While the analyst will know the probability margin, it can be hard to translate this back into the output units. However, the Q-method from Section 4 can be applied to the data also in order to get this information if it was desired, since both methods are conducted post-process.

5.2. Experiments

For this analysis, the nonlinear equation, using both normal and non-normal inputs, and the LOCA response surface were used to provide a comparison between the P- and Q-method. Only these systems were used because there were many different types

of confidence techniques to implement, so it was necessary to use system which that were not computationally intensive.

5.2.1. Nonlinear Equation

5.2.1.1. Normal Inputs

Once again, the first system to be used in the analysis was the nonlinear equation with normal inputs, detailed in Section 3.2.2. Here, the result of interest was the percentage of correct conclusion, or the percent of trials where the analyst would arrive at the correct decision regarding whether the system satisfied or violated a limit. This test was done by assigning arbitrary limits at certain quantiles of the output distribution. In a real analysis, the limit value would not be derived from the output distribution, but would be set by the regulator according to some certain safety constraints. However, this was done here so that consistent limits could be examined across multiple systems. When estimating a 95% confidence for a 0.95-quantile (Q-method) or 95% confidence for a 95% probability (P-method), the results were compared to a limit value at the 0.90-quantile of the output distribution, and at the 0.98-quantile. When estimating a 95% confidence for a 0.75-quantile or 95% confidence for a 75% probability, the results were compared to a limit value at the 0.70- and 0.80-quantile of the output distribution, since a limit at these values would be close to the true 0.75-quantile, and present a challenging test situation.

The first experiment compared a 95% confidence for the 0.95-quantile or for the 95% probability to a limit value at the true 0.90-quantile. Figure 5. 3 shows where the

true 0.95-quantile and limit value lie on the output distribution. Here the correct conclusion of the analysis is that the system fails the test, since the true 0.95-quantile is higher than the limit at the 0.90-quantile.

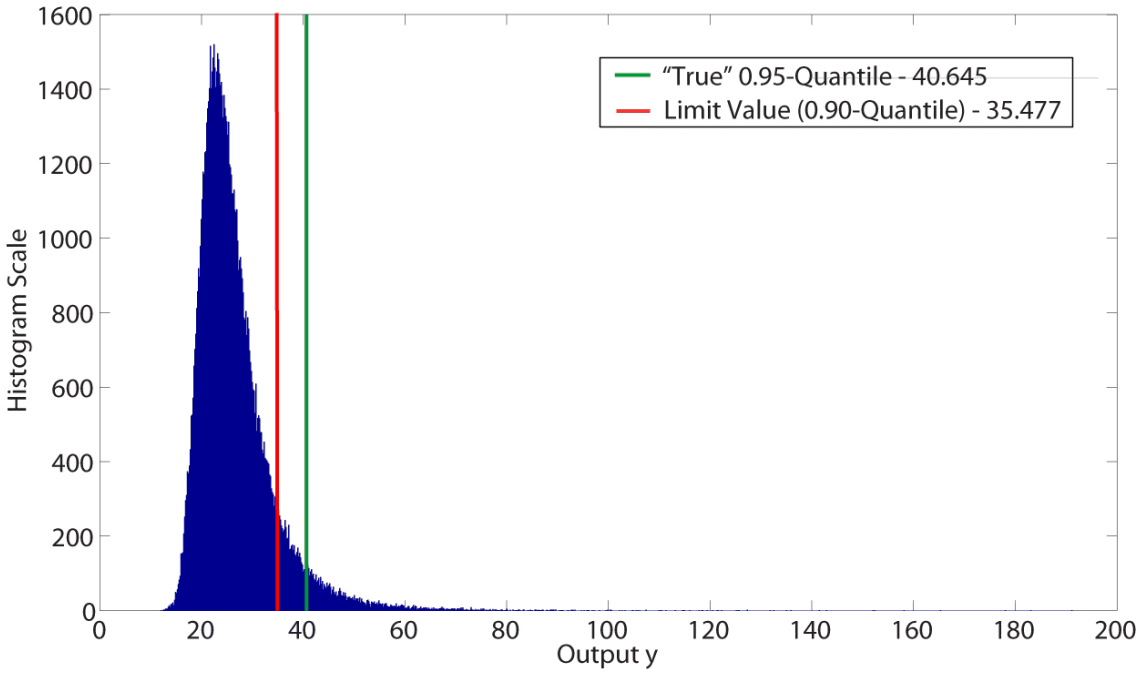

Figure 5. 3: Comparison of Limit and 0.95-Quantile

Since the limit value is below the 0.95-quantile, there are only two possible conclusions: correctly identifying that the system fails the limit, or conversely, incorrectly finding that the system is under the limit. This means the only error possible is a Type-I error, a false positive. This will occur when the result of the analysis falls below the limit value in the case of the quantile analysis, or when the probability of falling below the limit is above 0.95 in the case of the probability analysis. The possible outcomes using Q-method are shown in Figure 5. 4. As this figure shows, the 95% OSCI

could fall below the true 0.95-quantile value and still result in the correct conclusion. It is the distance which the 95/95 value falls below the true 0.95-quantile that will determine whether a Type-I error occurs.

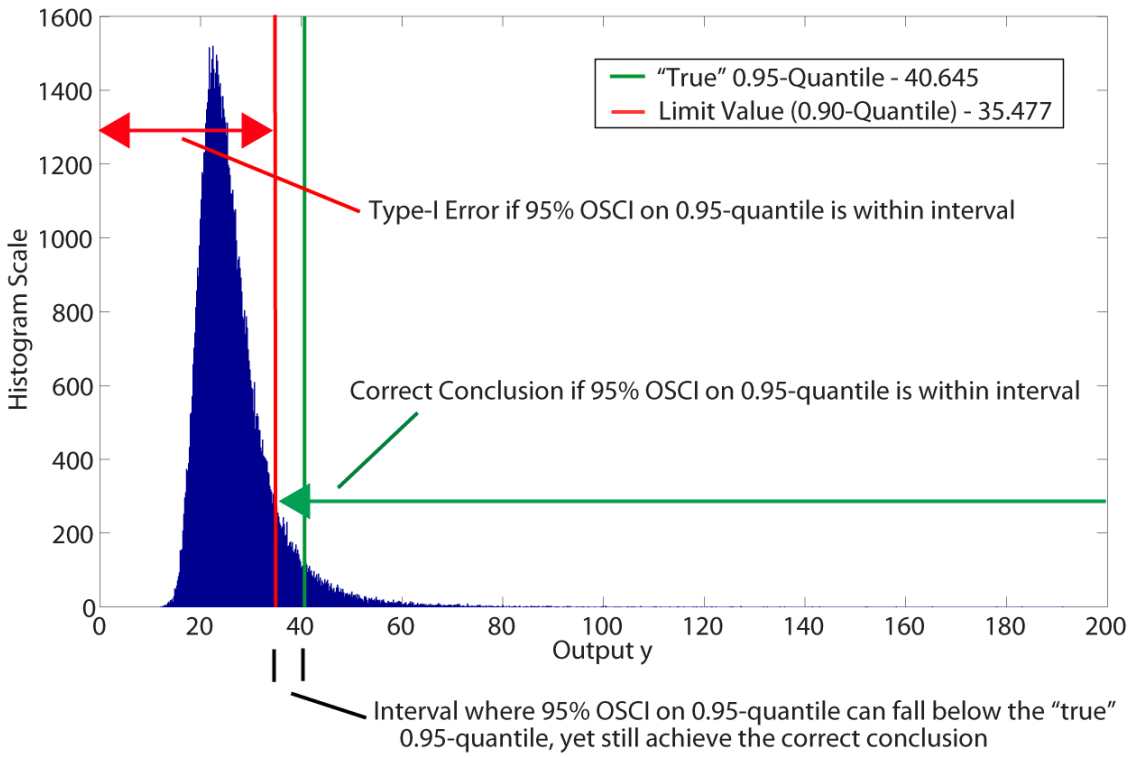

Figure 5. 4: Limit Value with Possible Conclusions

It is a similar situation for the P-method, where the correct probability outcome is 90%. Even if the probability outcome is over 90%, an error will only occur if the outcome is $\geq 95\%$.

Since there is a 95% confidence in the result of either analysis technique, the value for α, the probability of Type-I errors, is bounded at 5% (assuming the asymptotic methods have converged properly). However, how close α is to 5% depends on the

accuracy and precision of the analysis. The more accurate and precise the method, the smaller the value for α will be. For this experiment, 10^4 trials were conducted with each method. The results of this analysis can be seen in Figure 5. 5. Here, the five methods compared are:

1. rLHS: 95% OSCI for 0.95-quantile (Q-Method)
2. rLHS: 95% LOSCI for probability (P-Method)
3. CMC: 95% OSCI for 0.95-quantile (Q-Method)
4. CMC: 95% LOSCI for probability (P-Method)
5. CMC-OS: 95% OSCI for 0.95-quantile

Each method was tested at several different run levels (59, 93,124, 311, 548, and 1008), with 10^4 trials being conducted at each run level. As the figure shows, all methods had Type-I errors occur less than 5% of the time, with rLHS using the Q-method committing the least amount of errors. rLHS using the P-method was a close second, followed by CMC using the Q-method and CMC-OS. CMC using the P-method was by far the worst performer.

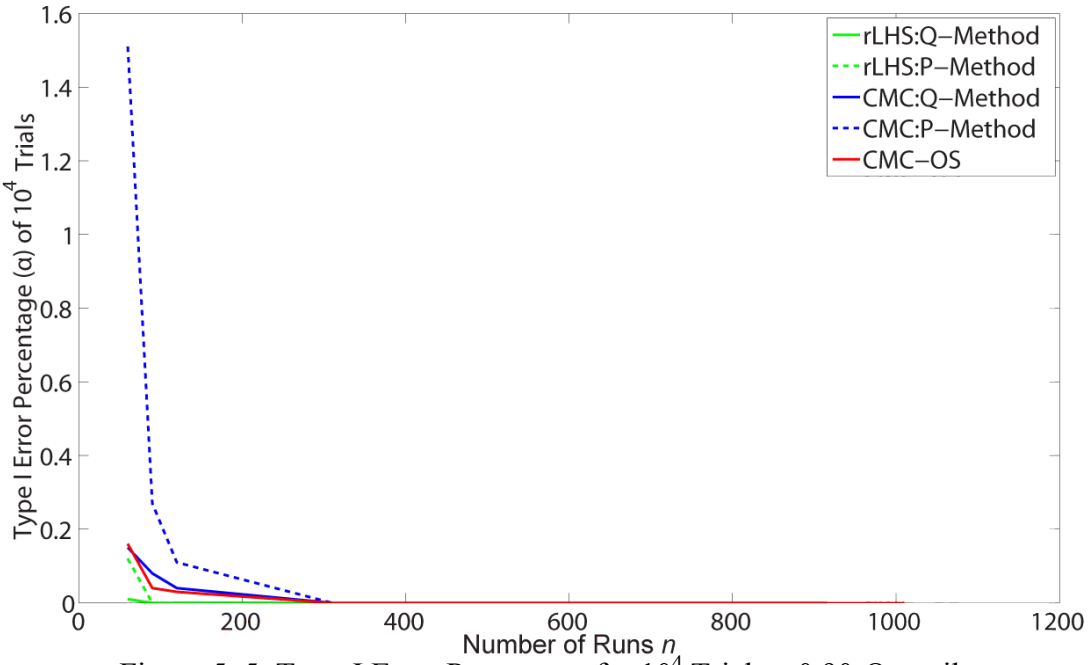

Figure 5. 5: Type-I Error Percentage for 10^4 Trials – 0.90-Quantile

The numerical results can be found in Table 5. 1, where the percentage of Type-I errors is given. Since the error percentage was so low, the test was repeated with the limit at the 0.94-quantile, which was a more challenging scenario. These results can be seen in Table 5. 2. Here, the trends from the 0.90-quantile limit continue, with the rLHS Q-method being the best performer, and the Q-method, in general, outperforming the P-method. The P-method also appears to take a longer time to converge, with several values exceeding the upper-bound of α for a properly converged OSCI of 5%, such as with CMC at $n = 59$ (12.6%) and rLHS with $t = 20$ and $n = 93$ (13.8%).

Table 5. 1: α Percentage – Nonlinear Eq. Normal Inputs – 0.90-Quantile

				rLHS					
		CMC		Q-Method			P-Method		
n*	**CMC-OS**	**Q-Method**	**P-method**	**t=10**	**t=20**	**t=30**	**t=10**	**t=20**	**t=30**
59	0.15	0.15	1.51	0.01	0.0	X	0.12	0.24	X
93	0.04	0.08	0.27	0.0	0.0	0.0	0.0	0.12	0.01
124	0.03	0.04	0.11	0.0	0.0	0.0	0.0	0.01	0.0
153	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
311	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1008	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

*10^4 Trials per run level

Table 5. 2: α Percentage – Nonlinear Eq. Normal Inputs – 0.94-Quantile

				rLHS					
		CMC		Q-Method			P-Method		
n*	**CMC-OS**	**Q-Method**	**P-method**	**t=10**	**t=20**	**t=30**	**t=10**	**t=20**	**t=30**
59	2.45	2.72	12.60	2.10	1.63	X	5.57	2.95	X
93	2.54	2.97	7.70	1.23	2.05	0.97	3.56	13.8	4.66
124	1.74	2.93	5.36	1.37	1.53	1.66	2.04	4.23	1.95
153	0.87	1.55	1.85	0.61	0.41	0.44	0.79	0.42	0.58
311	0.44	0.88	0.84	0.16	0.14	0.18	0.14	0.15	0.10
1008	0.15	0.27	0.29	0.05	0.01	0.03	0.04	0.0	0.01

*10^4 Trials per run level

In the next analysis, the limit value was placed at the true 0.98-quantile, which meant the system should pass the test since the limit is above the true 0.95-quantile. Here, the two possible conclusions are that the system correctly passes the test, or committing a Type-II error, a false negative, where the system does not appear to pass the test. Figure 5. 6 shows the possible conclusions and their related intervals. As the figure shows, if the 95% OSCI for the 0.95-quantile falls below the true quantile, there is no error. However, there is very little room to over-estimate the quantile. If the 95% OSCI for the 0.95-quantile falls above 48.453, the analyst will commit a Type-II error.

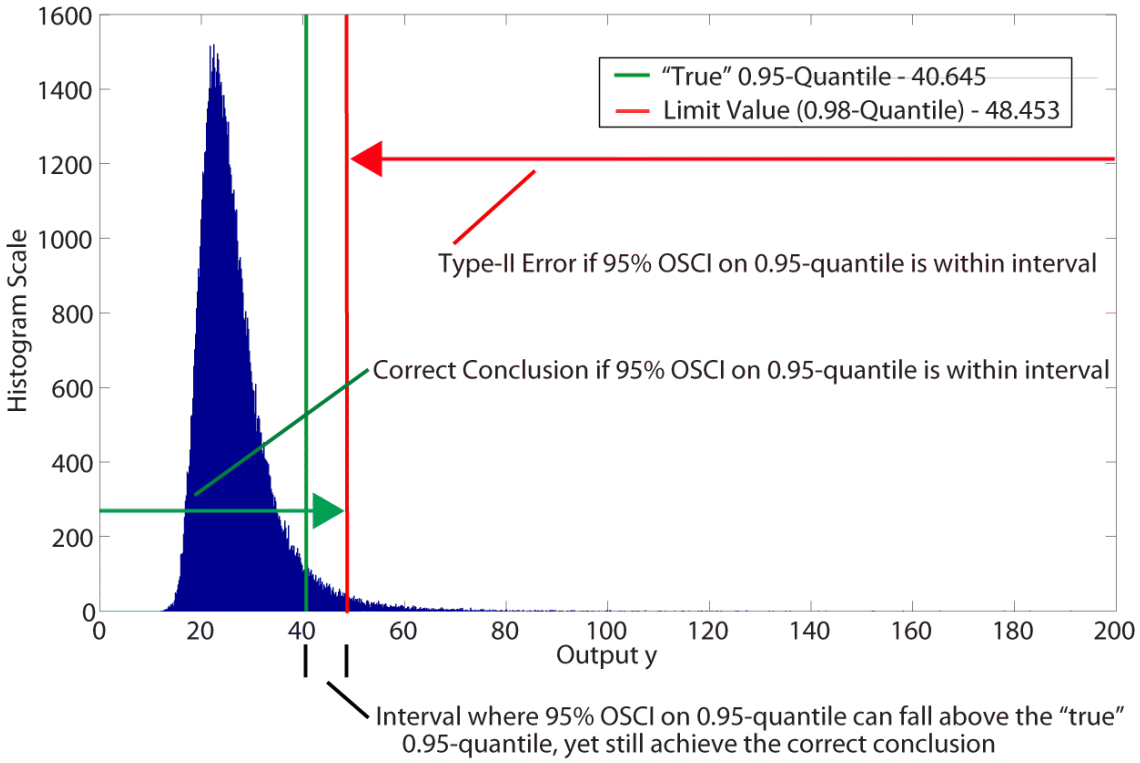

Figure 5. 6: Limit Value with Possible Conclusions

In respect to the probability method, the correct probability outcome should be 0.98. However, as long as the outcome result is greater than 0.95, the system will still pass the test.

The results of this analysis can be found in Figure 5. 7, with numerical results in Table 5. 3, which gives the percentage of Type-II errors out of 10^4 trials. In this case, the P-method using both rLHS and CMC outperforms the other techniques, with about a 50% reduction in errors at $n = 59$ compared to the Q-method results. Using rLHS with the Q-method achieves fewer errors than CMC using the Q-method at every run level, and CMC-OS, which performed the worst and resulted in $\beta \approx 70\%$ when the run level was

59. This means only ~30% of the time would the analyst have correctly concluded that the system should pass the test.

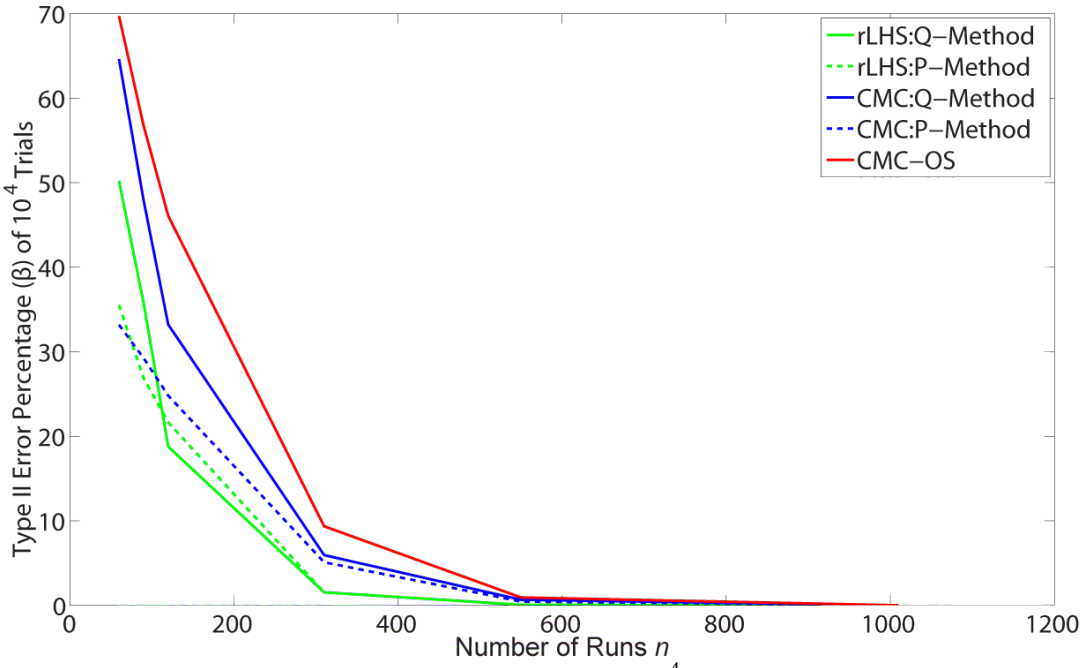

Figure 5. 7: Type-II Error Percentage for 10^4 Trials – 0.98-Quantile

Table 5. 3: β Percentage – Nonlinear Eq. Normal Inputs – 0.98-Quantile

				rLHS					
		CMC		Q-Method			P-Method		
n*	**CMC-OS**	**Q-Method**	**P-method**	**t=10**	**t=20**	**t=30**	**t=10**	**t=20**	**t=30**
59	69.68	64.60	33.17	50.19	40.71	X	35.52	36.74	X
93	56.70	47.89	29.23	35.71	19.74	22.12	26.86	16.74	10.00
124	46.05	33.2	24.79	18.76	9.33	7.58	21.62	19.11	12.23
153	9.34	5.93	5.09	1.53	0.08	0.05	1.52	0.12	0.08
311	0.92	0.65	0.44	0.05	0.0	0.0	0.06	0.0	0.0
1008	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

*10^4 Trials per run level

This analysis was repeated for the 0.75-quantile, and a limit value placed at the 0.70-quantile. The outcomes of this analysis are similar to the ones described in Figure 5.

4, where only a Type-I error is possible. For these tests, the following run levels were used: 11, 29, 40, 135, 246, 459, and 886. These results can be found in Figure 5. 8 and Table 5. 4. As the figure shows, even though the value for α is bound at 5%, the P-method results have values greater than 5%. This is because, at the lowest run level, only 11 runs were conducted. This may be too small for some of the asymptotic methods to converge, which is why the α error percentage is greater than 5%. At the next highest run level, all the methods are correctly under 5%. As the results show, rLHS using the Q-method is once again the best performer with very few Type-I errors (under 0.5% when using rLHS and $t = 5$). This is similar to the 0.90-quantile limit value results in Figure 5. 5.

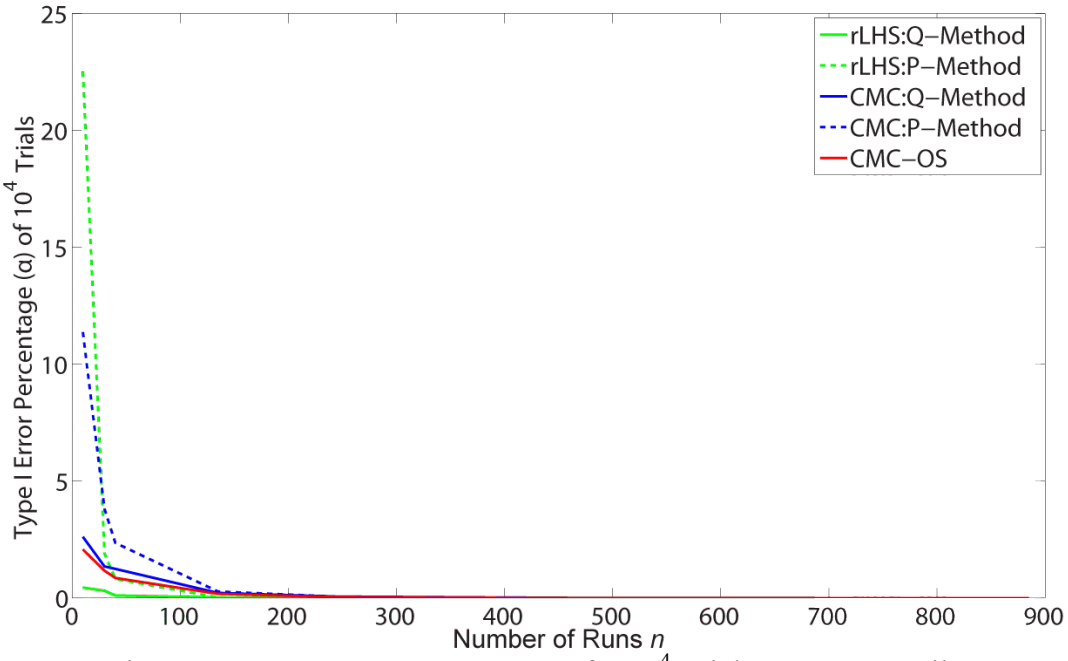

Figure 5. 8: Type-I Error Percentage for 10^4 Trials – 0.70-Quantile

Table 5. 4: α Percentage – Nonlinear Eq. Normal Inputs – 0.70-Quantile

				rLHS					
		CMC		Q-Method			P-Method		
n*	**CMC-OS**	**Q-Method**	**P-method**	**t=5**	**t=10**	**t=15**	**t=5**	**t=10**	**t=15**
11	2.07	2.61	11.37	0.44	X	X	22.51	X	X
29	1.16	1.34	3.80	0.30	0.29	7.34	1.89	0.29	2.38
40	0.85	1.24	2.35	0.10	0.28	3.44	0.81	0.28	0.30
135	0.17	0.21	0.27	0.0	0.0	0.01	0.02	0.03	0.03
246	0.03	0.05	0.05	0.0	0.0	0.0	0.0	0.0	0.0
459	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
886	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

*10^4 Trials per run level

This analysis was repeated using a limit at the 0.80-quantile, which is similar to the situation presented in Figure 5. 6, where only a Type-II error is possible. The results are presented in Figure 5. 9 and Table 5. 5. Here, rLHS using the Q-method and CMC-OS are at the same level for β when $n = 11$, but the rLHS method quickly outperforms the CMC-OS method as the number of runs increases, and convergence improves. The P-method begins as the best performer, but soon becomes essentially equivalent to the Q-method. Once again, these discrepancies at the lowest run level are a consequence of non-convergence at 11 runs. It is important to note the drastic improvement in correct conclusion when using the rLHS methods at the intermediate run levels, when compared to the CMC methods. At $n = 246$, the LHS methods are below 10% error, while the CMC methods are at 30-40%.

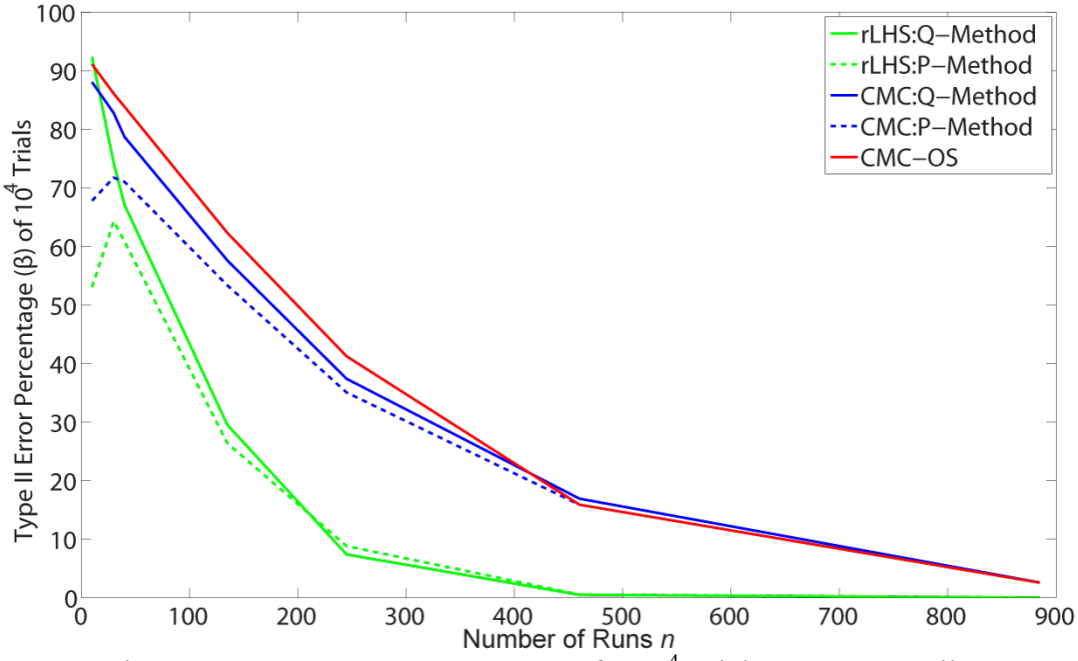

Figure 5. 9: Type-II Error Percentage for 10^4 Trials – 0.80-Quantile

Table 5. 5: β Percentage – Nonlinear Eq. Normal Inputs – 0.80-Quantile

				rLHS					
		CMC		Q-Method			P-Method		
n*	**CMC-OS**	**Q-Method**	**P-method**	**t=5**	**t=10**	**t=15**	**t=5**	**t=10**	**t=15**
11	91.11	88.06	67.79	92.35	X	X	53.05	X	X
29	86.09	82.77	71.75	74.23	66.2	47.12	64.29	55.37	48.61
40	83.84	78.69	71.02	67.07	59.39	53.91	60.82	65.51	62.18
135	62.28	57.59	53.33	29.50	17.40	23.64	26.36	21.59	23.42
246	41.24	37.39	35.05	7.39	5.54	4.75	8.82	5.91	6.25
459	15.88	16.92	15.88	0.49	0.25	0.22	0.54	0.34	0.25
886	2.57	2.59	2.57	0.0	0.0	0.0	0.01	0.0	0.0

*10^4 Trials per run level

5.2.1.2. Non-normal Inputs

These tests were repeated for the nonlinear equation with non-normal inputs detailed in Section 3.3.1.2. Here, only the numerical results are presented, with Table 5. 6 containing the results for a limit value at the 0.90-quantile, and Table 5. 7 containing the

results for a limit value at the 0.94-quantile. The trend from the example with normal inputs continues, with the Q-method outperforming the P-method when the limit value is less than the 0.95-quantile, with essential zero errors for the Q-method when the limit was at the 0.90-quantile, and less than 3% errors with a limit at the 0.94-quantile. Also, rLHS continues to incur less Type-I errors than the CMC-OS approach. As with the previous example, it appears that the P-method takes longer to converge, with several results over the bound of 5%, when the limit is placed at the 0.94-quantile.

Table 5. 6: α Percentage – Nonlinear Eq. Non-normal Inputs – 0.90-Quantile

				rLHS					
		CMC		Q-Method			P-Method		
n*	**CMC-OS**	**Q-Method**	**P-method**	**t=10**	**t=20**	**t=30**	**t=10**	**t=20**	**t=30**
59	0.16	0.20	1.29	0.0	0.0	X	0.01	0.0	X
93	0.06	0.09	0.32	0.0	0.0	0.0	0.0	0.11	0.0
124	0.01	0.0	0.06	0.0	0.0	0.0	0.0	0.01	0.0
153	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
311	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1008	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

*10^4 Trials per run level

Table 5. 7: α Percentage – Nonlinear Eq. Non-normal Inputs – 0.94-Quantile

				rLHS					
		CMC		Q-Method			P-Method		
n*	**CMC-OS**	**Q-Method**	**P-method**	**t=10**	**t=20**	**t=30**	**t=10**	**t=20**	**t=30**
59	2.82	3.04	12.91	2.21	1.64	X	4.92	1.45	X
93	2.22	2.99	7.74	1.13	2.26	0.60	2.67	18.25	3.72
124	1.86	2.98	5.25	1.87	2.06	1.78	1.77	7.15	1.22
153	1.04	1.70	2.05	0.42	0.41	0.34	0.55	0.26	0.27
311	0.47	0.90	0.86	0.17	0.14	0.09	0.17	0.11	0.05
1008	0.19	0.30	0.31	0.02	0.0	0.0	0.03	0.0	0.0

*10^4 Trials per run level

This test was repeated for a limit at the 0.98-quantile, with the results in Table 5. 8. Once again, the error percentages are very high at low run levels (greater than 30% for

all methods), but the P-method incurs about half as many errors as the Q-method, which is consistent with the previous example. Also, rLHS using the Q-method greatly outperforms CMC-OS with about a 40% reduction in errors.

Table 5. 8: β Percentage – Nonlinear Eq. Non-normal Inputs – 0.98-Quantile

				rLHS					
		CMC		Q-Method			P-Method		
n*	**CMC-OS**	**Q-Method**	**P-method**	**t=10**	**t=20**	**t=30**	**t=10**	**t=20**	**t=30**
59	70.10	70.21	33.57	41.61	28.18	X	34.08	35.10	X
93	56.82	52.80	29.53	36.36	16.12	20.79	25.68	18.01	17.57
124	46.39	40.28	24.21	21.36	8.27	7.45	20.77	20.60	19.62
153	10.47	10.17	5.39	2.25	0.19	0.15	1.39	0.20	0.29
311	1.03	1.04	0.62	0.08	0.01	0.0	0.03	0.02	0.0
1008	0.02	0.0	0.01	0.0	0.0	0.0	0.0	0.0	0.0

*10^4 Trials per run level

As before, these tests were repeated for the 0.75-quantile, with results in Table 5. 9 and Table 5. 10. The trends established before continue, with the Q-method being the best performer when the limit is below the estimated quantile, and the P-method performing better when the limit is above the estimated quantile. What is also interesting to note is that for this system, rLHS using the Q-method did not commit any errors when the limit was at the 0.90- or 0.70-quantile, and it was the only method to not contain any errors. In the case of the 0.70-quantile, this is most likely a result of convergence issues at the $n = 11$ run level. As the results in Section 4 showed, at this run level, the number of trials falling below the true quantile was much less than 5%. The results in Table 5. 10 appear to confirm this since, at $n = 11$, rLHS using the Q-method has the highest error rate, but at every other run level, rLHS is much better than the CMC methods (such as at

$n = 135$, where rLHS using the Q-method and $t = 5$ has a 5.89% error probability, compared to 61.78% using CMC-OS).

Table 5. 9: α Percentage – Nonlinear Eq. Non-normal Inputs – 0.70-Quantile

				rLHS					
		CMC		Q-Method			P-Method		
n*	**CMC-OS**	**Q-Method**	**P-method**	**t=5**	**t=10**	**t=15**	**t=5**	**t=10**	**t=15**
11	1.85	6.83	11.24	0.0	X	X	24.20	X	X
29	1.11	4.96	3.72	0.0	0.0	1.91	1.64	0.17	0.04
40	1.03	4.33	2.66	0.0	0.0	1.32	0.43	0.05	0.0
135	0.13	0.46	0.33	0.0	0.0	0.0	0.0	0.0	0.0
246	0.04	0.14	0.09	0.0	0.0	0.0	0.0	0.0	0.0
459	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
886	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

*10^4 Trials per run level

Table 5. 10: β Percentage – Nonlinear Eq. Non-normal Inputs – 0.80-Quantile

				rLHS					
		CMC		Q-Method			P-Method		
n*	**CMC-OS**	**Q-Method**	**P-method**	**t=5**	**t=10**	**t=15**	**t=5**	**t=10**	**t=15**
11	91.48	78.04	67.90	96.53	X	X	22.39	X	X
29	86.10	68.76	71.79	63.26	57.38	19.44	32.10	30.22	29.52
40	83.36	64.05	71.32	43.80	35.12	28.05	38.36	39.01	40.04
135	61.78	53.07	53.64	5.89	0.78	1.23	3.21	1.51	1.10
246	41.19	34.46	35.84	0.20	0.03	0.0	0.28	0.07	0.03
459	16.27	16.48	16.27	0.0	0.0	0.0	0.0	0.0	0.0
886	3.01	2.69	3.01	0.0	0.0	0.0	0.0	0.0	0.0

*10^4 Trials per run level

5.2.2. LOCA Response Surface

Lastly, the tests were repeated using the LOCA response surface detailed in Section 3.2.3. Here, the complete list of results presented in Table 5. 11. A closer examination of these results shows the same trend as the two previous examples. The Q-method incurs fewer errors when the limit is below the estimated quantile, and the P-method incurs fewer errors when the limit is above the estimated quantile. Also, the rLHS

approach using the Q-method consistently outperforms CMC-OS. It should be noted that for the 0.94-quantile, the CMC-OS method appears to have less errors than the rLHS Q-method, but if the results from Table 4. 18 are viewed again, it shows that the rLHS Q-method had over 6% of trials below the true quantile, which means there were convergence issues at low run levels for that design. This discrepancy disappears at the next highest run level, and the rLHS and CMC-OS methods end up essentially equivalent. Once again, when the limit value is above the true quantile, the improvement when using the rLHS methods can be very large, with an approximately 50% reduction in errors when compared to the CMC methods.

Table 5. 11: Incorrect Conclusion Percentages – LOCA Resp. Surf.

				rLHS					
		CMC		Q-Method			P-Method		
*n**	CMC-OS	Q-Method	P-method	t=10	t=20	t=30	t=10	t=20	t=30
0.90-Quantile									
59	0.25	0.24	1.45	0.06	0.08	X	0.33	0.15	X
93	0.07	0.04	0.31	0.04	0.03	0.02	0.08	0.52	0.06
124	0.02	0.01	0.14	0.01	0.02	0.01	0.02	0.01	0.02
153	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
311	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1008	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.94-Quantile									
59	2.91	2.81	11.85	3.10	4.35	X	7.23	6.81	X
93	2.15	2.45	7.36	2.01	4.54	3.88	5.11	10.41	6.82
124	1.64	2.72	5.31	2.24	3.11	4.00	3.67	3.17	2.63
153	0.76	1.43	1.67	0.94	1.02	1.37	1.19	1.03	1.31
311	0.44	0.63	0.89	0.30	0.45	0.32	0.31	0.41	0.35
1008	0.14	0.29	0.27	0.07	0.06	0.06	0.04	0.04	0.04
0.98-Quantile									
59	69.52	68.67	33.62	51.35	44.93	X	34.46	34.40	X
93	56.59	50.46	29.05	42.32	33.36	35.98	27.47	23.15	28.08
124	45.29	39.46	23.26	28.99	23.37	23.05	22.34	26.06	30.70
153	9.97	8.67	4.82	3.78	1.93	2.48	2.33	2.34	2.41
311	0.80	0.60	0.35	0.15	0.17	0.16	0.11	0.14	0.13
1008	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.70-Quantile				**t=5**	**t=10**	**t=15**	**t=5**	**t=10**	**t=15**
11	2.13	4.51	11.11	0.39	X	X	22.28	X	X
29	1.19	3.81	3.94	0.62	0.54	7.79	1.40	1.70	2.34
40	0.64	2.60	2.01	0.60	0.87	3.17	0.57	0.47	0.34
135	0.18	0.27	0.31	0.01	0.01	0.0	0.03	0.0	0.0
246	0.02	0.07	0.09	0.0	0.0	0.0	0.0	0.0	0.0
459	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
886	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.80-Quantile									
11	91.48	85.32	68.63	91.43	X	X	55.19	X	X
29	86.13	75.88	71.96	67.39	64.72	47.47	66.42	58.70	50.53
40	83.73	72.77	71.50	55.71	53.75	50.65	62.06	68.99	64.39
135	61.68	60.50	53.14	33.40	21.27	28.19	29.31	25.77	25.85
246	40.98	40.89	34.97	10.90	8.15	7.22	10.88	8.05	8.43
459	15.42	18.92	15.42	1.04	0.59	0.65	1.10	0.57	0.56
886	2.79	3.27	2.79	0.02	0.0	0.01	0.0	0.0	0.0

*10^4 Trials per run level

5.3. Analysis of Results

As the results show, the Q-method appears to outperform the P-method when the limit value is less than the estimated quantile, but the P-method performs better when the

limit value is above the estimated quantile. The following calculations in Sections 5.3.1, 5.3.2, and 5.3.3 are summarized from [116], and shed light on why this is the case.

5.3.1. P-Method Analysis

To compare the Q-method and P-method, define the quantile level q with $0 < q < 1$, and let the limit value $b \equiv b_q = F^{-1}(q)$. So when $q < 0.95$, the correct conclusion will be accepting H_0, or that the system should not pass the test, since the limit b is below the true 0.95-quantile. When $q > 0.95$, the correct conclusion is to reject H_0, meaning the system passes the test, since the limit b is above the true 0.95-quantile.

For the P-method, when $q < 0.95$, the correct conclusion occurs when $L < 0.95$, which is the same as accepting H_0, and it occurs with probability $P(L < 0.95)$. When $q > 0.95$, the correct conclusion occurs when $L \geq 0.95$, which is the same as rejecting H_0, and occurs with probability $P(L \geq 0.95)$.

The probability of achieving the correct conclusion using the P-method can be developed by approximating $P(L < 0.95)$ and $P(L \geq 0.95)$ for different values of q. Since $b \equiv b_q = F^{-1}(q)$, then $E[V] = P(Y \leq b_q) = q$, where $V = I(Y \leq b_q)$, as described in Section 5.1.1, and the variance $\sigma_{b_q}^2$ of Y is $q(1-q)$. Then specializing the CLT in Eq. 97 to the case when $b = b_q$,

$$\frac{\sqrt{n}}{\hat{\sigma}_n}(\bar{V}_n - q) \approx N(0,1) \qquad \text{Eq. 101}$$

for large n, since $\theta_b = q$, and the variance estimator

$$\hat{\sigma}_n^2 \approx \sigma_{b_q}^2 = q(1-q) \qquad \text{Eq. 102}$$

for large *n*. Therefore, using the hypothesis test statistic laid out in Eq. 100 of Section 5.1.3, the probability of $(L < 0.95)$ can be approximated by first adding the deviation of the limit value from the quantile $(0.95 - q/\hat{\sigma}/\sqrt{n})$ to both sides,

$$P(L < 0.95) = P\left(\frac{\bar{V}_n - 0.95}{\hat{\sigma}/\sqrt{n}} < z\right) = P\left(\frac{\bar{V}_n - q}{\hat{\sigma}/\sqrt{n}} < z + \frac{0.95 - q}{\hat{\sigma}/\sqrt{n}}\right). \quad \text{Eq. 103}$$

Then substituting the results from Eq. 101 into the left-hand side of the operator, and using Eq. 102 on the right-hand side gives,

$$P(L < 0.95) \approx P\left(N(0,1) < z + \frac{0.95 - q}{\sqrt{q(1-q)/n}}\right) = \Phi\left(z + \sqrt{n}\frac{0.95 - q}{\sqrt{q(1-q)}}\right) \quad \text{Eq. 104}$$

for large *n*, where Φ is the CDF of a standard normal. Also, the inverse case for $P(L \geq 0.95$ can be found,

$$\begin{aligned} P(L \geq 0.95) = 1 - P(L < 0.95) &\approx 1 - \Phi\left(z + \sqrt{n}\frac{0.95 - q}{\sqrt{q(1-q)}}\right) \\ &= \Phi\left(-z - \sqrt{n}\frac{0.95 - q}{\sqrt{q(1-q)}}\right) \end{aligned} \quad \text{Eq. 105}$$

for large *n* by the symmetry of the normal density function.

When $q < 0.95$, by Eq. 104, the probability of the correct conclusion for the P-method satisfies,

$$P(L < 0.95) \approx \Phi\left(z + \sqrt{n}\frac{0.95 - q}{\sqrt{q(1-q)}}\right) \to 1 \quad \text{Eq. 106}$$

as $n \to \infty$ since $\sqrt{n}(0.95 - q)/\sqrt{q(1-q)} \to \infty$ as $n \to \infty$.

When $q > 0.95$, by Eq. 105, the probability of the correct conclusion for the P-method satisfies,

$$P(L \geq 0.95) \approx \Phi\left(-z + \sqrt{n}\frac{0.95 - q}{\sqrt{q(1-q)}}\right) \to 1 \qquad \text{Eq. 107}$$

as $n \to \infty$ since $q > 0.95$ ensures $\sqrt{n}(0.95 - q)/\sqrt{q(1-q)} \to \infty$ as $n \to \infty$.

The following will now compare how quickly the probabilities of correct conclusion converge to 1 for the P-method as sample size *n* grows for a fixed q. When $q < 0.95$, the approximation to the probability of correct conclusion in Eq. 106 satisfies,

$$\Phi\left(z + \sqrt{n}\frac{0.95 - q}{\sqrt{q(1-q)}}\right) > \Phi(\text{z}) = 0.95 \qquad \text{Eq. 108}$$

for *all n* since $\sqrt{n}(0.95 - q)/\sqrt{q(1-q)} > 0$ for $q < 0.95$. Therefore, as long as the CLT approximation in Eq. 106 holds, meaning as long as the asymptotics have converged properly, then the probability of correct conclusion is *always* greater than 0.95 when $q > 0.95$. This is consistent with the experiments in Section 5.2, where the probability of correct conclusion is greater than 0.95 (meaning $\alpha \leq 0.05$), even when *n* is not very large (other than the trials at $n = 11$ where the asymptotics had not converged properly).

On the other hand, when $q > 0.95$, the probability of correct conclusion in Eq. 107 also converges to 1, but the approximate probability

$$\Phi\left(-z + \sqrt{n}\frac{0.95 - q}{\sqrt{q(1-q)}}\right) \qquad \text{Eq. 109}$$

is *not always* greater than 0.95 for all *n* because of the $-z$ in the argument. Rather Eq. 109 exceeds 0.95 only when *n* is large enough so that $-z + \sqrt{n}\frac{0.95-q}{\sqrt{q(1-q)}} > z$; i.e., when $n > \frac{4z^2 q(1-q)}{(0.95-q)^2}$, which can be quite large when $q \approx 0.95$. In other words, when q is only slightly larger than 0.95, the probability of the correct conclusion can be much less than 1

unless *n* is very large. This is consistent with the results in Section 5.2, when the limit value was placed at $q = 0.98$.

To put this another way, when $q < 0.95$, the correct conclusion is to accept H_0. Since H_0 is the null hypothesis and gets the "benefit of doubt", as explained in Section 2.1.2, it is not difficult to establish. This can be seen in Eq. 98, where if q is only slightly less than 0.95, the point estimate $\overline{V_n}$ is given extra help by subtracting $z\hat{\sigma}/\sqrt{n}$, thus making it more likely that $L < 0.95$.

On the other hand, when $q > 0.95$, the correct conclusion is to reject H_0, but this is harder to establish (as in the judicial example in Section 2, it would require evidence "beyond a reasonable doubt"). In Eq. 98, it can be seen that when q is only slightly greater than 0.95, it is harder for $L \geq 0.95$ because $z\hat{\sigma}/\sqrt{n}$ is subtracted from the point estimate $\overline{V_n}$, so it must also overcome the value of the $z\hat{\sigma}/\sqrt{n}$ term.

5.3.2. Q-Method Analysis

For the Q-method, assuming the same criterion laid out in the previous section, when $q < 0.95$, the correct conclusion occurs when $U > b_q$ (where *U* is the OSCI for the 0.95-quantile, as defined in Eq. 59), which is the same as accepting H_0, and occurs with probability $P(U > b_q)$. This means the system fails the test. When $q > 0.95$, the correct conclusion occurs when $U \leq b_q$, which is the same as rejecting H_0, and occurs with probability $P(U \leq b_q)$.

As with the P-method, approximations can be found for these probabilities. Since $U \leq b_q$ is equivalent to rejecting H_0 at level 0.05, starting from the hypothesis test

detailed in Section 4.2.2, and subtracting the deviation from the true quantile to the limit, $\frac{\sqrt{n}}{\hat{\tau}}(\xi_{0.95} - b_q)$, to both sides gives,

$$P(U \le b_q) = P\left(\frac{\tilde{\xi}_{0.95,n} - b_q}{\hat{\tau}/\sqrt{n}} \le -z\right) = P\left(\frac{\tilde{\xi}_{0.95,n} - \xi_{0.95}}{\hat{\tau}/\sqrt{n}} \le -z - \frac{\sqrt{n}}{\hat{\tau}}(\xi_{0.95} - b_q)\right) \quad \text{Eq. 110}$$

Then using the CLT approximation in Eq. 57 and $\hat{\tau} \approx \tau$,

$$P(U \le b_q) \approx P\left(N(0,1) \le -z - \frac{\sqrt{n}}{\tau}\right) = \Phi\left(-z - \frac{\sqrt{n}}{\tau}(\xi_{0.95} - b_q)\right) \quad \text{Eq. 111}$$

Also, as with the P-method, the conversing case can be found

$$\begin{aligned} P(U > b_q) = 1 - P(U < b_q) &\approx 1 - \Phi\left(-z - \frac{\sqrt{n}}{\tau}(\xi_{0.95} - b_q)\right) \\ &= \Phi\left(z + \frac{\sqrt{n}}{\tau}(\xi_{0.95} - b_q)\right) \end{aligned} \quad \text{Eq. 112}$$

once again by the symmetry of the normal density function.

When $q < 0.95$, by Eq. 112, the probability of the correct conclusion satisfies

$$P(U > b_q) \approx \Phi\left(z + \frac{\sqrt{n}}{\tau}(\xi_{0.95} - b_q)\right) \to 1 \quad \text{Eq. 113}$$

as $n \to \infty$ since $b_q = F^{-1}(q) < F^{-1}(0.95) = \xi_{0.95}$ for $q < 0.95$.

When $q > 0.95$, by Eq. 111, the probability of correct conclusion is

$$P(U \le b_q) \approx \Phi\left(-z + \frac{\sqrt{n}}{\tau}(b_q - \xi_{0.95})\right) \to 1 \quad \text{Eq. 114}$$

as $n \to \infty$ since $q > 0.95$ implies $b_q = F^{-1}(q) > \xi_{0.95}$, so $\frac{\sqrt{n}}{\tau}(b_q - \xi_{0.95}) \to \infty$ as $n \to \infty$.

The same comparison of the rates of convergence to 1 as *n* grows large can be made as in the previous section, with the same results. The probability of correct conclusion converges faster when $q < 0.95$ than when $q > 0.95$.

5.3.3. Comparison between Methods

The following calculations seek to prove why the Q-method outperforms the P-method when $q < 0.95$. The probabilities of correct conclusion for both methods were provided earlier in Eq. 106 and Eq. 113. The only difference between these approximations is their arguments to Φ, with $z + \sqrt{n}\frac{0.95-q}{\sqrt{q(1-q)}}$ for the P-method, and $z + \sqrt{n}\frac{\xi_{0.95}-b_q}{\tau}$ for the Q-method. If it can be shown that this argument for the Q-method is larger than that for the P-method when $q < 0.95$, then it will give some explanation of why the Q-method outperforms the P-method in the experiments in Section 5.2 when $q < 0.95$.

The goal is to show,

$$\frac{\xi_{0.95} - b_q}{\tau} > \frac{0.95 - q}{\sqrt{q(1-q)}}. \qquad \text{Eq. 115}$$

A first-order Taylor approximation for b_q gives

$$b_q = F^{-1}(q) \approx F^{-1}(0.95) + \frac{q - 0.95}{f(\xi_{0.95})} = \xi_{0.95} + \frac{q - 0.95}{f(\xi_{0.95})} \qquad \text{Eq. 116}$$

since $\frac{d}{dp}F^{-1}(p) = 1/f(F^{-1}(p)) = 1/f(\xi_p)$ by the chain rule of calculus. Using the definition of τ in Eq. 58,

$$\frac{\xi_{0.95} - b_q}{\tau} \approx \frac{(0.95 - q)/f(\xi_{0.95})}{\sqrt{0.95(1 - 0.95)/f(\xi_{0.95})}} \qquad \text{Eq. 117}$$

$$= \frac{(0.95 - q)}{\sqrt{0.95(1 - 0.95)}}$$

This result satisfies the goal of showing that Eq. 115 holds because

$$\frac{(0.95 - q)}{\sqrt{0.95(1 - 0.95)}} > \frac{0.95 - q}{\sqrt{q(1 - q)}} \qquad \text{Eq. 118}$$

for $0.05 < q < 0.95$.

A similar analysis can show why the P-method outperforms the Q-method when $q > 0.95$. The probabilities for correct conclusion using the P- and Q-methods are given by Eq. 107 and Eq. 114. Here, the differences are still in respect to their arguments to Φ; the P-method's argument is now $-z + \sqrt{n}\frac{0.95-q}{\sqrt{q(1-q)}}$, and the Q-method's argument is $-z + \sqrt{n}\frac{\xi_{0.95}-b_q}{\tau}$. Therefore, the definition of τ in Eq. 58 and of b_q in Eq. 116 implies

$$\frac{b_q - \xi_{0.95}}{\tau} \approx \frac{(q - 0.95)/f(\xi_{0.95})}{\sqrt{0.95(1 - 0.95)}/f(\xi_{0.95})} = \frac{q - 0.95}{\sqrt{0.95(1 - 0.95)}} < \frac{q - 0.95}{\sqrt{q(1 - q)}} \qquad \text{Eq. 119}$$

since $q > 0.95$. Therefore, the argument to Φ for the P-method is larger than that for the Q-method when $q > 0.95$. This means the probability of the P-method reaching the correct conclusion converges to 1 faster than the Q-method when $q > 0.95$, and explains the results seen in Section 5.2.

5.4. Combined Methods

Since the Q-method outperforms when $q < 0.95$, and the P-method outperforms when $q > 0.95$, efforts were made to combine the two methods in a way that would result in a higher percentage of correct conclusions. This technique would seek agreement between the two methods before a decision was made.

For this technique, rLHS will be used since it has proved to offer the greatest variance reduction in the experiments in Section 4.3 and Section 5.3. The procedure starts with the analyst carrying out some number of LHS cases *m*, where each case has *t* runs, as before. After a minimum number of cases have been simulated, the results will be analyzed to see if the Q-method and P-method are in agreement in regards to their conclusion. If they are, that conclusion will be taken as the final result. If they are not, an additional LHS case will be conducted, and then the results will be viewed again. This will continue until both methods are in agreement. The flowchart in Figure 5. 10 shows an example of this analysis using $t = 10$ and a minimum $m = 6$. The hope is that waiting until both methods are in agreement will lead to fewer incorrect conclusions than each method individually, even if it means conducting additional cases in order to get agreement.

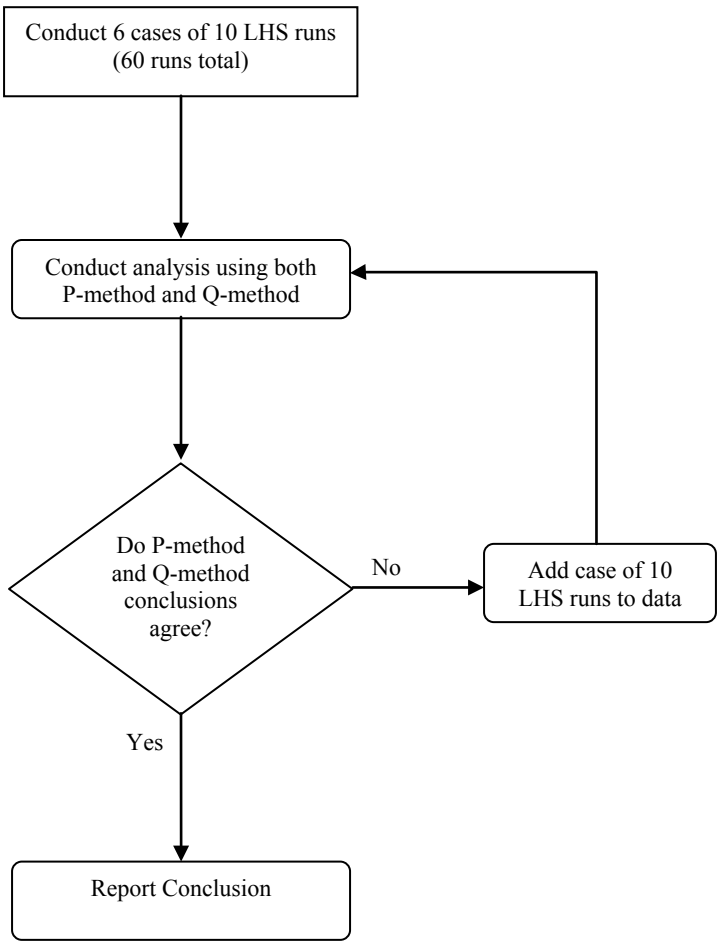

Figure 5. 10: Combined Method Flowchart

It should be noted here, that there are dangers when carrying out multiple hypothesis tests, with differing test statistics, from the same data set. The biggest danger is in regards to "cherry-picking" the desired result from multiple hypothesis tests without disclosing the full extent of the tests and their results [117]. This is not the case here, since both test statistics will be used and reported. However, since both tests use the same data, their results will be correlated to an extent. Another possible danger by combining test statistics in the manner presented here is that the attributes of the 95% confidence, in

relation to the probability of committing a Type-I error (α), are not necessarily guaranteed. As shown in Section 4 and at the beginning of Section 5, using a 95% confidence provided a top bound for α of 5%. However, by combining the two test statistics in this simple fashion, it no longer guarantees this bound for α. There are ways to design processes like these, referred to as sequential decision procedures, to ensure that the confidence property remains [118], but that analysis was not done for this technique.

Several test cases were conducted with this technique, but a direct comparison to the results in Section 4 and Section 5 is difficult. This is due to the fact that the amount of runs needed to reach a conclusion using this combined method will not be known beforehand. The runs will continue until the methods are in agreement. This means the results will not present a direct comparison to the previous results at assigned run levels.

The first comparison used the nonlinear equation with normal inputs, detailed in Section 3.2.2. For this test, when estimating a 95/95, the minimum number of runs conducted before a conclusion could be reached was 60. This level was chosen since the methods had not previously been examined at lower run levels for estimating a 95/95, since 59 was the minimum for CMC-OS. Table 5. 12 shows the results with a limit value at the 0.94-quantile, which is a very challenging situation. Here, the results presented are the number of trials, out of 10^4, that reached a conclusion, whether correct or incorrect, at that run level. These results can be compared to the results in Table 5. 2, where prescribed run levels were used. As Table 5. 12 shows, overall, the combined method had a success rate of 98.13%. This compares to a success rate of 97.90% when using the rLHS Q-method alone at 60 runs, and 94.43% when using the rLHS P-method alone at 60

runs. So there seems to be a slight improvement, even though it is not a direct comparison. The reason the total in Table 5. 12 does not equal 10,000 is because some trials did not reach a consensus between the two methods before the maximum amount of cases was added.

Table 5. 12: Correct vs. Error Nonlinear Eq. Normal Inputs – 0.94

	Runs Conducted									
	60	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**Total**
Correct Conclusion	9420	275	97	8	6	5	0	1	0	9813
Error (α)	118	2	58	4	4	1	0	0	0	108

This test was repeated with a limit at the 0.98-quantile, as seen in Table 5. 13. These results can be compared to Table 5. 3. Here, the combined method had a success rate of 62.01%. Using the rLHS Q-method alone, at 60 runs, returned a success rate of 49.81%, so there is a stark improvement by using the combined method. However, the rLHS P-method had a success rate at 60 runs of 64.48%. So using the combined method did not quite achieve the kind of reduction in errors that using the P-method alone provided.

Table 5. 13: Correct vs. Error Nonlinear Eq. Normal Inputs – 0.98

	Runs Conducted									
	60	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**Total**
Correct Conclusion	4428	197	1145	284	35	3	57	0	34	6201
Error (β)	2850	589	260	54	10	20	2	6	1	3795

Another test was conducted to see the influence of starting the analysis at a higher minimum run level, since most of the errors in the previous example occurred at the lowest run level. In this case, the minimum for *m* was 10, or 100 total runs. These results can be found in Table 5. 14. Here, the success rate is 76.43%, while using the rLHS Q-method alone at 90 runs had a success rate of 64.29%, and using the P-method alone had a rate of 73.14%. So the combined method did provide a higher probability of success than using Q-method or P-method alone.

Table 5. 14: Correct vs. Error Nonlinear Eq. Normal Inputs – 0.98 (Higher Start)

	Runs Conducted										
	100	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**Total**
Correct Conclusion	6607	31	742	9	167	2	4	44	2	19	7643
Error (β)	1702	466	42	107	3	11	4	8	1	2	2349

These tests were repeated for the nonlinear equation using non-normal inputs. Table 5. 15 shows the results for a limit at the 0.94-quantile. This resulted in a correct conclusion percentage of 98.20%, which once again was only modestly better than using the Q-method alone, which had a success rate at 60 runs of 97.79%, and the P-method alone (95.08%).

Table 5. 15: Correct vs. Error Nonlinear Eq. Non-normal Inputs – 0.94

	Runs Conducted									
	60	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**Total**
Correct Conclusion	9487	235	70	13	7	7	0	0	0	9820
Error (α)	120	2	52	4	1	0	1	0	0	180

With the limit placed at the 0.98-quantile, the success rate was 64.62%, which was better than the Q-method alone (58.38%), but slightly worse than the P-method alone (65.92%) at the 60 run level. These results can be seen in Table 5. 16. So again, there appears to be an advantage to the combined method since it's better than both the P- and Q-method with a limit below the quantile, and better than the Q-method alone with a limit above the quantile.

Table 5. 16: Correct vs. Error Nonlinear Eq. Non-normal Inputs – 0.98

	Runs Conducted									
	60	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**Total**
Correct Conclusion	4936	143	833	313	73	15	87	4	27	6462
Error (β)	2486	710	222	56	15	27	0	12	1	3533

Lastly, the combined method was tested on the LOCA response surface from Section 3.2.3. Table 5. 17 shows the results for a limit placed at the 0.94-quantile. The success rate was 98.58%, which was slightly better than the Q-method alone (96.90%), and better than the P-method alone (92.77%), but P-method had not converged properly at the 60 run level, since the correct conclusion percentage is <95%. One interesting result is that the combined method appears to have prevented that same error from occurring here. Even though the P-method may not have been properly converged at the lowest run level, the addition of the Q-method prevented a large number of incorrect conclusions being reported.

Table 5. 17: Correct vs. Error LOCA Response Surface – 0.94

	Runs Conducted									
	60	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**Total**
Correct Conclusion	9158	345	118	15	8	7	4	1	0	9658
Error (α)	210	18	93	8	2	3	3	0	3	342

Table 5. 18 shows the results with a limit at the 0.98-quantile. Here, the combined method had a success rate of 60.98%, which was far better than the Q-method alone (48.65%) and only slightly worse than the P-method alone (65.54%) at 60 runs. The apparent trend from the previous two tests continues, with a slight improvement over the other methods when the limit is below the quantile estimation, and a large improvement over the Q-method when the limit is above the quantile.

Table 5. 18: Correct vs. Error LOCA Response Surface – 0.98

	Runs Conducted									
	60	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**Total**
Correct Conclusion	4266	284	1023	272	97	24	71	8	24	6098
Error (β)	2902	606	288	37	21	29	3	7	2	3899

5.5. Discussion

The results of this section appear to show that establishing confidence intervals for a probability could be used for a direct comparison to regulatory limits. However, the experimental results, and subsequent discussion, showed that the P-method is less efficient at reaching the correct conclusion when the limit value is below the quantile of interest. On the other hand, the exact opposite outcome occurs when the limit value is

above the quantile of interest, since the results and discussion show that the P-method is more efficient than the Q-method at arriving at the correct conclusion. Both the P-method and Q-method for rLHS also appeared less likely to commit errors, regardless of the location of the limit value, when compared to CMC-OS, as long as the asymptotics had converged properly (which was not always the case at the lowest run levels). When the limit was above the quantile, the increased probability of correct conclusion when using rLHS compared to CMC-OS, could be sizable.

In an effort to take advantage of the positive aspects of both methods, a technique was devised that required agreement between the methods before a decision could be made. Through experiments, it was shown that this technique can improve the probability of correct conclusion when the limit is below the quantile, when compared to the P- and Q-method alone. When the limit is above the quantile, the results also showed that the combined method can greatly improve the probability of correct conclusion when compared to the Q-method alone, and return essentially equivalent probabilities when compared to the P-method alone. However, this combined method also has drawbacks, since the number of runs needed is not known beforehand, and it requires the data to be analyzed repeatedly (though time/effort for the data analysis should be small in comparison to the time needed for large, complex code runs). It is also possible to conduct a trial that does not result in agreement even after a large amount runs. While this possibility was small, it could result in many more code runs, meaning lost time and expensive computational costs. Also, it may be possible to combine the P- and Q-

methods within a more rigorous framework that ensures the 95% confidence level is preserved [116].

Chapter 6: Conclusions and Recommendations for Future Work

The issue of comparison between computer code outputs and regulatory limits was defined more rigorously through the use of hypothesis testing. This framework provided guidance on how to increase the probability of correct conclusion. Since certain factors are out of the analyst's control, such as the placement of limit values, the system characteristics, and the acceptable level of error, the only technique left to increase the correct conclusion percentage was to decrease the variance of the analysis result. Therefore, VRTs were analyzed in order to gauge their applicability to this goal.

The first step in this analysis involved investigating methods to increase the accuracy and precision of a point estimate of the 0.95-quantile. This included a comparison between CMC and LHS, but also a detailed study of the use of OAs and OLHCs. As was expected, LHS provided a more accurate and precise quantile estimation than CMC. However, OLHCs outperformed regular LHS, even when using static midpoints, rather than sampled values. This result is particularly important, as the use of static midpoints simplifies that process of modifying input distributions post-analysis, without the need for re-running the computer code. The potential of OLHCs appears promising, and a more thorough investigation into their use should be done. Unfortunately, the results from Section 4 showed that the method to establish CIs for the quantiles of an rLHS design may be unsuitable for use with OLHCs. A closer look indicated that this was a result of the dependence between different OLHC designs. It

may be possible to create OLHC designs that are truly independent, and that may resolve this issue. Lastly, the use of higher resolution OAs was shown to be inappropriate for the estimation of the 0.95-quantile, if static midpoints were used. This appears to be a consequence of the size of the intervals created using an OA. Since fewer intervals are used, they become wider, and the tails of the distributions are not analyzed when using midpoints. However, using higher resolution OAs with sampling did slightly outperform the CMC method, but not to the extent of LHS or OLHC.

Section 4 provided an investigation into the applicability of new methods to establish CIs for the quantiles of an output distribution created using a VRT [72] to the field of nuclear safety analysis. The results demonstrated that rLHS and AV can provide a more accurate and precise result, especially when estimating a CI for a quantile value near a long distribution tail. However, there are convergence issues at very low run levels. There may be ways to alleviate some of these convergence issues (or at least ensure that the error is on the conservative side), and as more tests are completed, increased guidance on the selection of bandwidth parameters, or other appropriate derivative estimation techniques, will arise. Interestingly, the results also showed that a more accurate derivative estimation at the lowest run levels may not always be desirable. While the increased accuracy decreases the size of the CI, it may also make the result more vulnerable to error in the quantile point estimate. Even when the rLHS method had not converged to the proper levels, it always provided resulting values over a smaller range than CMC-OS. As the results from Section 5 showed, at essentially every run level (even the lowest levels, where convergence may not have occurred), the probability of

error when using rLHS was equivalent to or smaller than using CMC-OS. Finally, it was shown that the possibility of "gaming" the system was no greater when using rLHS than with CMC-OS.

It was also shown that a more direct interpretation of the NRC probability requirement was possible by establishing CIs for the probability of exceeding a safety limit value. This technique appeared to have a higher probability of correct conclusion, when compared to the Q-method, if the limit value was above the quantile of interest, but the opposite was true when the limit was below the quantile. A detailed look into the mathematics behind both methods demonstrated why this is the case. One issue using the probability method is that it only provides information about margin to the safety limit in terms of probability, not the output units of the system or limit value. However, since both the Q- and P-method are performed post-process, it is possible to use both techniques and derive the desired data.

Since both the P- and Q-method could be applied to the same data, a technique was devised that attempted to take advantage of the positive properties of both methods. This technique waited for agreement between the two methods before establishing a conclusion. It appeared that this method improved the probability of correct conclusion, but a direct comparison to previous results was difficult, since the number of runs necessary for agreement is not known *a priori*. This is also the biggest downside to the method, since the analyst has no idea how many runs will be necessary, and there is a small, but nonzero, possibility of needing to perform a large amount of runs.

Based on this information, recommendations about the applicability of these techniques for use in nuclear safety analysis can be made. First, a look into NRC recommendations on risk assessment procedures showed that the use of CIs rather than Bayesian credible intervals appears to be appropriate for this application. This is the case since prior information is difficult to use in regulatory analyses, the results will not be propagated through more systems, the data are created in a normal way, and the probability requirement is fulfilled by the use of the quantile, while the CI only provides information about the accuracy of the sampling scheme. Also, if a RISMC is to be used, the use of quantiles may provide similar information to an overlap probability, without the need for detailed information about the extremes of the output distribution. Since this information is not needed, the amount of runs necessary for the analysis should be reduced. Also, the use of lower quantiles, such as the 0.75-quantile, with high confidence may be appropriate for regulatory limits on beyond-design-basis accidents. Techniques were shown on how this could be done in comparison to the limit curve in the TNF, and a proposed CCDF limit curve.

The recommendation of the asymptotic methods, like rLHS, for use in regulatory analyses is more nuanced. While there were convergence issues for these methods at the lowest levels possible when using CMC-OS, the deviation from the proper coverage levels was never extreme (meaning the coverage levels were off by only a couple percent). The results from Section 5 also showed that the rLHS consistently had a higher probability of correct conclusion than CMC-OS. So if the regulator's first concern is arriving at the correct conclusion, rLHS would appear to be acceptable. From the utility

point-of-view, the use of rLHS could result in a substantial increase in accuracy. This may provide more opportunity for increasing reactor properties, like temperature and power. For both the regulator and utility, the increased accuracy and precision can provide better guidance about which accidents have the greatest associated risk. Correctly prioritizing accident scenarios based on risk is one of the best ways to systematically increase safety, as the Tower Perrin firm report pointed out. So even considering the convergence issues, it is hard to justify why the rLHS method would not be acceptable for use in safety analysis.

Perhaps the greatest opportunity for future work, along this line of research, is associated with the examination of other VRTs. It is possible for control variates and importance sampling to provide much greater variance reduction than even rLHS. As mentioned in Section 4, there are additional constraints which must be considered when using these methods in relation to if/how previous knowledge is used. However, if acceptable approaches are found, these methods may be able to offer great benefits in reducing regulatory error, increasing margin to safety limits, and generally improving the knowledge and characterization of the output distribution of safety analyses.

Bibliography

[1] 107th Congress; 1st Session, "Atomic Energy Act of 1954," *NUREG-0980,* 1954 (Revised 2002).

[2] Nuclear Regulatory Commission, History Staff, "A Short History of Nuclear Regulation 1946-2009," *NUREG-0175,* 2010.

[3] W. Ergen, "Emergency Core Cooling: Report of the Advisory Task Force on Power Reactor Emergency Cooling," AEC/NRC, 1967.

[4] "10 CFR Part 50 Appendix K," *Code of Federal Regulations ,* Amended June 1, 2000.

[5] Nuclear Regulatory Commisson, "Reactor Safety Study: An Assessment of Accident Risk in U.S. Commerical Nuclear Power Plants," *WASH-1400,* 1975.

[6] Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, "Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants," *NUREG-1150,* 1990.

[7] Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, "Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making," *NUREG-1855,* 2009.

[8] Nuclear Regulatory Commission, "Commission Issuance of White Paper on Risk-Informed and Performance-Based Regulation," *Yellow Announcement # 019,* 1999.

[9] Nuclear Regulatory Commission, "An Approach for Using Probablisitic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis," *RG 1.174,* 2002.

[10] Nuclear Regulatory Commission, "An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities," *RG 1.200,* 2007.

[11] ASME/American Nuclear Society, "Standard for Level 1/Large Early Release Frequency Probabilistic Risk Assessment for Nuclear Power Plant Applications," *ASME/ANS RA-Sa-2009,* 2009.

[12] 60 FR 42622, "Use of Probabilistic Risk Assessment Methods in Nuclear Activities: Final Policy Statement," *Federal Register,* vol. 60, p. 42622, 1995.

[13] Nuclear Regulatory Commission, Risk Management Task Force, "A Proposed Risk Management Regulatory Framework," *NUREG-2150,* 2012.

[14] Nuclear Regulatory Commission: Office of Nuclear Regulatory Research, "Feasibility Study for a Risk-Informed and Performance-Based Regulatory Structure for Future Plant Licensing," *NUREG-1860,* 2007.

[15] Idaho National Lab, "Risk-Informed Safety Margin Characterization," *INL/CON-09-15549,* 2009.

[16] H. M. Park, "Hypothesis Testing and Statistical Power of a Test," The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University, 2008.

[17] J. Baptist du Prel, G. Hommel, B. Rohrig and M. Blettner, "Confidence Interval of P-Value?," *Deutsches Ärzteblatt International,* vol. 106, no. 19, pp. 335-339, 2009.

[18] M. Morgan, M. Henrion and M. Small, Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, Cambridge University Press, 1992.

[19] L. Zadeh, "Fuzzy Sets," *Information and Control,* vol. 8, pp. 338-353, 1965.

[20] L. Zadeh, "Fuzzy Sets as the Basis for a Theory of Possibility," *Fuzzy Sets,* vol. 1, pp. 3-28, 1978.

[21] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.

[22] A. Dempster, "Upper and Lower Probabilities induced by a Multivalued Mapping," *The Annals of Mathematical Statistics ,* vol. 38, no. 2, pp. 325-339, 1967.

[23] Y. Ben-Haim, Info-Gap Decision Theory, Haifa Israel: Academic Press, 2006.

[24] D. G. Cacuci, M. Ionescu-Bujor and I. M. Navon, Sesnsitivity and Uncertainty Analysis: Applications to Large-Scale Systems, Boca Raton: CRC Press, 2005.

[25] A. Saltelli, K. Chan and E. Scott, Sensitivity Analysis, John Wiley & Sons, 2008.

[26] C. Daniel, "One-at-a-Time-Plans," *Journal of the American Statistical Association,* vol. 68, p. 353, 1973.

[27] M. Morris, "Factorial Sampling Plans for Preliminary Computational Experiments," *Technometrics,* vol. 33, p. 161, 1991.

[28] NIST/SEMATECH, "e-Handbook of Statistical Methods," 2012. [Online].

Available: http://www.itl.nist.gov/div898/handbook/. [Accessed May 2012].

[29] G. Taguchi and S. Konishi, Orthogonal Arrays and Linear Graphs, Dearborn, MI: ASI Press, 1987.

[30] S. Cotter, "A Screening Design for Factorial Experiments with Interactions," *Biometrika,* vol. 66, p. 317, 1979.

[31] T. Andres and W. Hajas, "Using Iterated Fractional Factorial Design to Screen Parameters in Sensitivity Analysis of a Probabilistic Ris Assessment Model," in *Proceedings of the Joint International Conference on Mathematical Methods and Supercomputing in Nuclear Applications*, Karlsruhe, Germany April19-23, 1993.

[32] B. Bettonvil, Detection of Important Factors by Sequential Bifurcation, Tilburg, The Netherlands: Tilburg University Press, 1990.

[33] A. Dunker, "Efficient Calculations of Sensitivity Coefficients for Complex Atmospheric Models," *Atmospheric Environment,* vol. 15, p. 1155, 1981.

[34] A. Dunker, "The Decoupled Direct Method for Calculating Sensitivity Coefficients in Chemical Kinetics," *Journal of Chemical Physics,* vol. 81, p. 2385, 1984.

[35] D. Miller and M. Frenklach, "Sensitivity Analysis and Parameter Estimation in Dynamic Modeling of Chemical Kinetics," *International Journal of Chemical Kinetics,* vol. 15, pp. 677-696, 1983.

[36] D. Cacuci, "Sensitivity Theory for Nonlinear Systems. I. Nonlinear Functional Analysis Approach," *Journal of Mathematical Physics,* vol. 22, p. 2794, 1981.

[37] D. Cacuci, "Senstivity Theory for Nonlinear Systems. II. Extensions to Additional

Classes of Responses," *Journal of Mathematical Physics,* vol. 22, p. 2803, 1981.

[38] J. E. Till and H. Grogan, Radiological Risk Assessment and Enviromental Analysis, New York: Oxford Univeristy Press, 2008.

[39] P. Glasserman, Monte Carlo Methods in Financial Engineering, New York: Springer, 2004.

[40] J. Helton and F. Davis, "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems," *Reliability Engineering and System Safety,* vol. 81, pp. 23-69, 2003.

[41] W. H. Press, B. Flannery, S. Teukolsky and W. Vetterling, Numerical Recipes in C, Cambridge: Cambridge University Press, 1988.

[42] A. Bjorck, Numerical Methods for Least Squares Problems, Philadelphia: SIAM, 1996.

[43] O. Roderick, M. Anitescu and P. Fischer, "Polynomial Regression Approaches Using Derivative Information for Uncertainty Quantification," *Nuclear Science and Engineering ,* vol. 164, no. 2, pp. 122-139, 2010.

[44] N. Smirnov, "Tables for Estimating the Goodness of Fit of Emperical Distributions," *Annals of Mathematical Statistics,* vol. 19, p. 279, 1948.

[45] T. Anderson, "On the Distribution of the Two-Sample Cramer-von Mises Criterion," *The Annals of Mathematical Statistics,* vol. 33, no. 3, pp. 1148-1159, 1962.

[46] W. Kruskal, "Historical Notes on the Wilcoxon Unpaired Two-Sample Test,"

Journal of the American Statistical Association, vol. 52, no. 279, pp. 356-360, 1957.

[47] M. McKay, "Evaluating Prediction Uncertainty. Technical Report: NUREG/CR-6311," US Nuclear Regulatory Commissionn and Los Alamos National Laboratory, 1995.

[48] S. Hora and R. Iman, "A Comparison of Maximum/Bounding and Bayesian/Monte Carlo for Fault Tree Uncertainty Analysis. Technical Report: SAND85-2839," Sandia National Laboratories, Albuquerque, NM, 1986.

[49] I. Sobol', "Sensitivity Analysis for Non Linear Mathematical Models," *Matematicheskoe Modelirovanie,* vol. 2, pp. 112-118, 1990.

[50] P. Jansen, W. Rossing and R. Daamen, "Monte Carlo Estimation of Uncertainty Contributions from Several Independent Multivariate Sources," *Predictability and Nonlinear Modeling in Natural Sciences and Economics ,* pp. 334-3443, 1994.

[51] R. Cukier, C. Fortuin, K. Schuler, A. Petschek and J. Schaibly, "Study of the Sensitivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients," *Journal of Chemical Physics,* vol. 59, pp. 3873-3878, 1973.

[52] R. Cukier, H. Levine and K. Schuler, "Nonlinear Sensitivity Analysis of Multiparameter Model Systems," *Journal of Computational Physics,* vol. 26, pp. 1-42, 1978.

[53] R. Cukier, J. Schaibly and K. Schuler, "Study of the Senstivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients," *Journal of Computational Physics,*

vol. 63, pp. 1140-1149, 1975.

[54] J. Schiably and K. Schuler, "Study of the Sensitivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients," *Journal of Chemical Physics,* vol. 59, pp. 3879-3888, 1973.

[55] G. Box and K. Wilson, "On the Experimental Attainment of Optimum Conditions," *Journal of the Royal Statistical Society,* vol. 13, no. 1, 1951.

[56] G. Box and G. Tiao, Bayesian Inference in Statistical Analysis, Wiley, 1973.

[57] A. Gelman, J. Carlin, H. Stern and D. Rubin, Bayesian Data Analysis, Chapman & Hall/CRC, 2009.

[58] D. Denison, B. Mallick and F. Smith, "Bayesian MARS," *Statistics and Computing,* vol. 8, pp. 337-346, 1998.

[59] M. D. McKay, R. J. Beckham and W. J. Conover, "A Comparison of Three Methods for Selecting Input Variables in the Analysis of Output from a Computer Code," *Technometrics,* vol. 21, pp. 239-245, 1979.

[60] G. Wyss and K. Jorgensen, "A User's Guide to LHS: Sandia's Latin Hypercube Sampling Software," *Sandia National Lab,* 1998.

[61] B. Tang, "Orthogonal Array-Based Latin Hypercubes," *Journal of the American Statistical Association,* vol. 88, 1993.

[62] B. Tang, C. Lin, D. Bingham and R. Sitter, "A New and Flexiible Method for Constructing Designs for Computer Experiments," *Institute of Mathematical Studies,* vol. 38, 2010.

[63] D. Rasmuson, D. Anderson and J. Mardekian, "Use of 3^n Parallel Flats Fractional Factorial Designs in Computer Code Uncertainty Analysis," *Department of Energy,* 1979.

[64] A. French, "Response Surface Modeling of a Large Break Loss of Coolant Accident," MS Thesis, 2008.

[65] RELAP5-3D Code Development Team, "RELAP5-3D Code Manual," *Idaho National Lab,* 2005.

[66] W. T. Pratt, V. Mubayi, T. L. Chu, G. Martinez-Guridi and J. Lehner, "An Approach for Estimating the Frequencies of Various Contaiment Failure Modes and Bypass Events," *NUREG/CR-6595,* 2004.

[67] L. Soffer, S. B. Burson, C. M. Ferrell, R. Y. Lee and J. N. Ridgely, "Accident Source Terms for Light-Water Nuclear Power Plants," *NUREG-1465,* 1995.

[68] U.S. Atomic Energy Commission, "Assumptions used for Evaluating the Potential Radiological Consequences of a Loss of Coolant Accident for Boiling Water Reactors," *RG 1.3,* 1974.

[69] U.S. Atomic Energy Commission, "Assumptions used for Evaulting the Potential Radiological Consequences of a Loss of Coolant Accident for Pressurized Water Reactors," *RD 1.4,* 1974.

[70] U.S. Environmental Protection Agency, "Cancer Risk Coefficients for Environmental Exposure to Radiation," Federal Guidance Report No. 13, EPA-402-R-99-001, 1999.

[71] Oak Ridge National Laboratoy, "A User's Manual for the ORIGEN2 Computer Code," *ORNL/TM-71-75,* 1980.

[72] F. Chu and M. K. Nakayama, "Confidence Intervals for Quantiles when Applying Variance-Reduction Techniques," *ACM Transaction On Modeling and Computer Simulation,* vol. 36, pp. Article 7 (25 pages plus 12-page online-only appendix), 2012.

[73] "10 CFR 50.46," *Code of Federal Regulations,* Ammended 2007.

[74] Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, "Best-Estimate Calculations of Emergency Core Cooling System Performance," *RG 1.157,* 1989.

[75] C. Frepoli, "Overview of Westinghouse Realistic Large Break LOCA Evaluation Model," *Science and Technology of Nuclear Installations,* vol. 2008, 2008.

[76] R. Martin and L. O'Dell, "AREVA's Realistic Large Break LOCA Analysis Methodology," *Nuclear Engineering and Design,* vol. 235, pp. 1713-1725, 2005.

[77] J. Jaech, "On the Use of Tolerance Intervals in Acceptance Sampling by Attributes," *Journal of Quality Technology,* vol. 4, no. 2, 1972.

[78] Nuclear Regulatory Commission, Office of Nuclear Reactor Regulation, "Applying Statistics," *NUREG-1475,* 1994.

[79] R. Martin and W. Nutt, "Perspectives on the Application of Order-Statistics in Best-Estimate Plus Uncertainty Nuclear Safety Analysis," *Nuclear Engineering and Design,* vol. 241, no. 1, pp. 274-284, 2011.

[80] H. Glaeser and R. Pochard, "Review on Uncertainty Methods for Thermal Hydraulic Computer Codes," *International Conference on New Trends in Nuclear Systems,* vol. 1, pp. 447-455, 1994.

[81] L. Pal and M. Makai, "Remarks on Statistical Aspects of Safety Analysis of Complex Systems," *Reliability Engineering and System Safety ,* vol. 80, no. 3, pp. 217-232, 2003.

[82] W. T. Nutt and G. B. Wallis, "Evaluation of Nuclear Safety from the Outputs of Computer Codes in the Presence of Uncertainties," *Reliability Engineering and System Safety,* vol. 83, no. 1, pp. 57-77, 2004.

[83] S. Wilks, "Order Statistics," *Bullentin of the American Mathematical Society,* vol. 1, 1948.

[84] A. Wald, "An Extension of Wilks Method for Setting Tolerance Limits," *Annals of Mathematical Statistics,* vol. 14, 1943.

[85] V. Easton and J. McColl, "Statistics Glossary v1.1," 1997. [Online]. Available: http://www.stats.gla.ac.uk/steps/glossary/. [Accessed May 2012].

[86] *Reliability Engineering & System Safety,* vol. 23, no. 4, pp. 247-323, 1988.

[87] Nuclear Regulatory Commission, "A Review of NRC Staff uses of Probablistic Risk Assessment," *NUREG-1489,* 1994.

[88] Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, "Handbook of Parameter Estimation for Probablisitic Risk Assessment," *NUREG/CR-6823,* 2003.

[89] J. Sim and N. Reid, "Statistical Inference by Confidence Intervals: Issues of Interpretation and Utilization," *Journal of the American Physical Therapy Association,* vol. 79, pp. 186-195, 1999.

[90] M. M. Siddiqui, "Distribution of Quantiles in Samples from a Bivariate Population," *Journal of Research of the National Bureau of Standards B,* vol. 64, pp. 145-150, 1960.

[91] M. Stein, "Large Sample Properties of Simulations Using Latin Hypercube Sampling," *Technometrics,* vol. 29, no. 2, pp. 143-151, 1987.

[92] R. R. Bahadur, "A Note on Quantiles in Large Samples," *Annals of Mathmatical Statistics,* vol. 37, pp. 577-580, 1966.

[93] J. K. Ghosh, "A New Proof of the Bahadur Representation of Quantiles and an Application," *Annals of Mathematical Statistics,* vol. 42, pp. 1957-1961, 1971.

[94] R. J. Serfling, Approximation Theorems of Mathematical Statistics, New York: John Wiley & Sons, 1980.

[95] J. W. Tukey, "Which Part of the Sample Contains Information?," *Proc Natl Acad Sci USA,* vol. 53, pp. 127-134, 1965.

[96] E. Parzen, "Density Quantile Estimation Approach to Statistical Data Modeling," in *Smoothing Techniques for Curve Estimation*, vol. Springer, Berlin, 1979.

[97] S. Asmussen and P. Glynn, Stochastic Simulation: Algorithms and Analysis, New York: Springer, 2007.

[98] D. A. Bloch and J. L. Gastwirth, "On a Simple Esimate of the Reciprocal of the

Density Function," *Annals of Mathematical Statistics,* vol. 39, pp. 1083-1085, 1968.

[99] E. Bofinger, "Esimation of a Density Function using Order Statistics," *Australian Journal of Statistics ,* vol. 17, pp. 1-7, 197.

[100] W. J. Conover, Practical Nonparametric Statistics, New York: John Wiley & Sons, 1999.

[101] M. K. Nakayama, "Asymptotically Valid Confidence Intervals for Quantiles and Values-at-risk when Applying Latin Hypercube Sampling," *International Journal on Advances in Systems and Measurements,* vol. 4, pp. 86-94, 2011.

[102] S. Ross, Simulation, San Diego: Academic Press, 1997.

[103] A. N. Avramidis and J. R. Wilson, "Correlation-induction Techniques for Estimating Quantiles in Simulation," *Operations Research,* vol. 46, pp. 574-591, 1998.

[104] W. G. Cochran, Sampling Techniques, New York: Wiley, 1977.

[105] P. Hall and S. J. Sheather, "On the Distribution of a Studentized Quantile," *Journal of the Royal Statistical Society B,* vol. 50, pp. 381-391, 1988.

[106] M. Falk, "On the Estimation of the Quantile Density Function," *Statistics & Probability Letters,* vol. 4, pp. 69-73, 1986.

[107] M. K. Nakayama, "Asymptotic Properties of Kernel Density Estimators when Applying Importance Sampling," *Proceedings of the 2011 Winter Simulation Conference,* pp. 556-568, 2011. Institute of Electrical and Electronics Engineers.

[108] A. Brunett, R. Denning and T. Aldemir, "Application of Limit Curves to the Risk-Informed Regulations of SFRs," in *Transactions of the 2009 American Nuclear Society Annual Meeting*, Atlanta, GA, 2009.

[109] Sandia National Laboratories, "MELCOR Computer Code Manuals - Vol. 1: Primer and Users' Guide," *NUREG/CR-6119,* 2011.

[110] H. A. Simon, Sciences of the Artificial, 2d ed., Cambridge: MIT Press, 1982.

[111] Sandia National Laboratories, "MELCOR Computer Code Maunals - Vol. 3: Demonstration Problems," *NUREG/CR-6119,* 2001.

[112] M. Payton and M. N. Greenstone, "Overlappying Confidence Intervals or Standard Error Intervals: What do they mean in terms of Statistical Significance," *Journal of Insect Science,* vol. 3, no. 34, 2003.

[113] M. Mulekar and M. S.N., "Confidence Intervals Estimation of Overlap: Equal Means Case," *Computational Statistics & Data Analysis ,* vol. 34, no. 2, pp. 121-137, 2000.

[114] A. Helu, S. H. and R. Vogel, "Nonparametric Overlap Coefficient Estimation Using Ranked Set Sampling," *Journal of Nonparametric Statistics,* vol. 23, no. 2, pp. 385-397, 2011.

[115] S. Mizuno, T. Yamaguchi, A. Fukushima, Y. Matsuyama and Y. Ohashi, "Overlap Coefficient for Assessing the Similarity of Parmacokinetic Data between Ethnically Different Populations," *Clinical Trials,* vol. 2, pp. 174-181, 2005.

[116] T. Aldemir, R. Denning, D. Grabaskas and M. Nakayama, "A Comparison of

Methods for Performing Regulatory Analyses," *To be submitted,* 2012.

[117] Committee on Professional Ethics, "Ethical Guidelines for Statistical Practice," American Statistical Association, 1999.

[118] M. Sobel and A. Wald, "A Sequential Decision Procedure for Choosing One of Three Hypotheses Concerning the Unknown Mean of a Normal Distribution," *The Annals of Mathematical Statistics,* vol. 20, no. 4, pp. 502-522, 1949.

[119] A. S. Hedayat, N. Sloane and J. Stufken, Orthogonal Arrays: Theory and Applications, Chicago: Springer-Verlag, 1999.

Appendix

Appendix A: Orthogonal Arrays

The following figures list the OAs used in Section 3; they are taken from [119]. Some of the OLHC designs are not listed here due to their size, but they were created using the methods outlined in [62].

Run	Input				
	1	2	3	4	5
1	1	1	1	1	1
2	1	2	2	2	2
3	1	3	3	3	3
4	1	4	4	4	4
5	2	1	2	3	4
6	2	2	1	4	3
7	2	3	4	1	2
8	2	4	3	2	1
9	3	1	3	4	2
10	3	2	4	3	1
11	3	3	1	2	4
12	3	4	2	1	3
13	4	1	4	2	3
14	4	2	3	1	4
15	4	3	2	4	1
16	4	4	1	3	2

Figure A. 1: $L_{16'}$ – 16 Run Resolution III OA – 4 Levels

Run	Input														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
3	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
4	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
5	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1
7	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1
8	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2
9	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
10	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1
11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1
12	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2
13	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1
14	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2
15	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2
16	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1

Figure A. 2: L_{16} – 16 Run Resolution III OA – 2 Levels

Run	Input								
	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1	1	1	1	1
2	1	2	2	2	2	2	2	2	2
3	1	3	3	3	3	3	3	3	3
4	1	4	4	4	4	4	4	4	4
5	2	1	1	2	2	3	3	4	4
6	2	2	2	1	1	4	4	3	3
7	2	3	3	4	4	1	1	2	2
8	2	4	4	3	3	2	2	1	1
9	3	1	2	3	4	1	2	3	4
10	3	2	1	4	3	2	1	4	3
11	3	3	4	1	2	3	4	1	2
12	3	4	3	2	1	4	3	2	1
13	4	1	2	4	3	3	4	2	1
14	4	2	1	3	4	4	3	1	2
15	4	3	4	2	1	1	2	4	3
16	4	4	3	1	2	2	1	3	4
17	1	1	4	1	4	2	3	2	3
18	1	2	3	2	3	1	4	1	4
19	1	3	2	3	2	4	1	4	1
20	1	4	1	4	1	3	2	3	2
21	2	1	4	2	3	4	1	3	2
22	2	1	4	2	3	4	1	3	2
23	2	3	2	4	1	2	3	1	4
24	2	4	1	3	2	1	4	2	3
25	3	1	3	3	1	2	4	3	2
26	3	2	4	4	2	1	3	3	1
27	3	3	1	1	3	4	2	2	4
28	3	4	2	2	4	3	1	1	3
29	4	1	3	4	2	4	2	1	3
30	4	2	4	3	1	3	1	2	4
31	4	3	1	2	4	2	4	3	1
32	4	4	2	1	3	1	3	4	2

Figure A. 3: $L_{32'}$ – 32 Run Resolution III OA – 4 Levels

	Input																												
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2	2	2
4	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
5	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1	1	1	2	2
6	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1	2	2	2	2	1	1
7	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	1	1
8	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1	2	2	2	2	1	1	1	1	1	1	1	1	2	2
9	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1
10	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1	2	2	1	1	2	2
11	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2	1	1	2	2	2	2	1	1	2	2
12	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	1	1	2	2	1	1
13	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
14	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2	2	2	1	1	1	1
15	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1	1	1
16	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2	2	2	1	1	1	1	2	2	1	1	2	2	2	2
17	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
18	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1	2	1	2	1	2	1
19	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2	1	2	1	2	2	1	2	1	2	1
20	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	1	2	1	2	1	2
21	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	1	2	2	1
22	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2	2	1	2	1	1	2
23	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1	1	2
24	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2	2	1	2	1	1	2	1	2	1	2	1	2	2	1
25	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2
26	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2	2	1	1	2	2	1
27	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	1	2	2	1	2	1	1	2	2	1
28	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	1	2	2	1	1	2
29	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1
30	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	2	1	1	2	1	2
31	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2	1	2
32	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	2	1	1	2	1	2	2	1	1	2	2	1	2	1

Figure A. 4: L_{32} – 32 Run Resolution III OA – 2 Levels

Run	Input																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	3	1	2	0	2	0	1	2	2	3	2	2	2	1	3	0	1	1	0	0
3	2	1	2	3	0	3	0	2	3	3	1	3	3	3	2	1	0	2	2	0	0
4	3	2	3	1	0	1	0	3	1	1	2	1	1	1	3	2	0	3	3	0	0
5	0	1	3	1	2	0	2	0	1	2	2	3	2	2	2	1	3	0	1	1	0
6	1	2	2	3	2	2	2	1	3	0	1	1	0	0	3	2	3	1	0	1	0
7	2	0	1	2	2	3	2	2	2	1	3	0	1	1	0	0	3	2	3	1	0
8	3	3	0	0	2	1	2	3	0	3	0	2	3	3	1	3	3	3	2	1	0
9	0	2	1	2	3	0	3	0	2	3	3	1	3	3	3	2	1	0	2	2	0
10	1	1	0	0	3	2	3	1	0	1	0	3	1	1	2	1	1	1	3	2	0
11	2	3	3	1	3	3	3	2	1	0	2	2	0	0	1	3	1	2	0	2	0
12	3	0	2	3	3	1	3	3	3	2	1	0	2	2	0	0	1	3	1	2	0
13	0	3	2	3	1	0	1	0	3	1	1	2	1	1	1	3	2	0	3	3	0
14	1	0	3	1	1	2	1	1	1	3	2	0	3	3	0	0	2	1	2	3	0
15	2	2	0	0	1	3	1	2	0	2	0	1	2	2	3	2	2	2	1	3	0
16	3	1	1	2	1	1	1	3	2	0	3	3	0	0	2	1	2	3	0	3	0
17	0	0	1	3	1	2	0	2	0	1	2	2	3	2	2	2	1	3	0	1	1
18	1	3	0	1	1	0	0	3	2	3	1	0	1	0	3	1	1	2	1	1	1
19	2	1	3	0	1	1	0	0	3	2	3	1	0	1	0	3	1	1	2	1	1
20	3	2	2	2	1	3	0	1	1	0	0	3	2	3	1	0	1	0	3	1	1
21	0	1	2	2	3	2	2	2	1	3	0	1	1	0	0	3	2	3	1	0	1
22	1	2	3	0	3	0	2	3	3	1	3	3	3	2	1	0	2	2	0	0	1
23	2	0	0	1	3	1	2	0	2	0	1	2	2	3	2	2	2	1	3	0	1
24	3	3	1	3	3	3	2	1	0	2	2	0	0	1	3	1	2	0	2	0	1
25	0	2	0	1	2	2	3	2	2	2	1	3	0	1	1	0	0	3	2	3	1
26	1	1	1	3	2	0	3	3	0	0	2	1	2	3	0	3	0	2	3	3	1
27	2	3	2	2	2	1	3	0	1	1	0	0	3	2	3	1	0	1	0	3	1
28	3	0	3	0	2	3	3	1	3	3	3	2	1	0	2	2	0	0	1	3	1
29	0	3	3	0	0	2	1	2	3	0	3	0	2	3	3	1	3	3	3	2	1
30	1	0	2	2	0	0	1	3	1	2	0	2	0	1	2	2	3	2	2	2	1
31	2	2	1	3	0	1	1	0	0	3	2	3	1	0	1	0	3	1	1	2	1
32	3	1	0	1	0	3	1	1	2	1	1	1	3	2	0	3	3	0	0	2	1
33	0	0	2	1	2	3	0	3	0	2	3	3	1	3	3	3	2	1	0	2	2
34	1	3	3	3	2	1	0	2	2	0	0	1	3	1	2	0	2	0	1	2	2
35	2	1	0	2	2	0	0	1	3	1	2	0	2	0	1	2	2	3	2	2	2
36	3	2	1	0	2	2	0	0	1	3	1	2	0	2	0	1	2	2	3	2	2
37	0	1	1	0	0	3	2	3	1	0	1	0	3	1	1	2	1	1	1	3	2
38	1	2	0	2	0	1	2	2	3	2	2	2	1	3	0	1	1	0	0	3	2
39	2	0	3	3	0	0	2	1	2	3	0	3	0	2	3	3	1	3	3	3	2
40	3	3	2	1	0	2	2	0	0	1	3	1	2	0	2	0	1	2	2	3	2
41	0	2	3	3	1	3	3	3	2	1	0	2	2	0	0	1	3	1	2	0	2
42	1	1	2	1	1	1	3	2	0	3	3	0	0	2	1	2	3	0	3	0	2
43	2	3	1	0	1	0	3	1	1	2	1	1	1	3	2	0	3	3	0	0	2
44	3	0	0	2	1	2	3	0	3	0	2	3	3	1	3	3	3	2	1	0	2
45	0	3	0	2	3	3	1	3	3	3	2	1	0	2	2	0	0	1	3	1	2
46	1	0	1	0	3	1	1	2	1	1	1	3	2	0	3	3	0	0	2	1	2
47	2	2	2	1	3	0	1	1	0	0	3	2	3	1	0	1	0	3	1	1	2
48	3	1	3	3	3	2	1	0	2	2	0	0	1	3	1	2	0	2	0	1	2
49	0	0	3	2	3	1	0	1	0	3	1	1	2	1	1	1	3	2	0	3	3
50	1	3	2	0	3	3	0	0	2	1	2	3	0	3	0	2	3	3	1	3	3
51	2	1	1	1	3	2	0	3	3	0	0	2	1	2	3	0	3	0	2	3	3
52	3	2	0	3	3	0	0	2	1	2	3	0	3	0	2	3	3	1	3	3	3
53	0	1	0	3	1	1	2	1	1	1	3	2	0	3	3	0	0	2	1	2	3
54	1	2	1	1	1	3	2	0	3	3	0	0	2	1	2	3	0	3	0	2	3
55	2	0	2	0	1	2	2	3	2	2	2	1	3	0	1	1	0	0	3	2	3
56	3	3	3	2	1	0	2	2	0	0	1	3	1	2	0	2	0	1	2	2	3
57	0	2	2	0	0	1	3	1	2	0	2	0	1	2	2	3	2	2	2	1	3
58	1	1	3	2	0	3	3	0	0	2	1	2	3	0	3	0	2	3	3	1	3
59	2	3	0	3	0	2	3	3	1	3	3	3	2	1	0	2	2	0	0	1	3
60	3	0	1	1	0	0	3	2	3	1	0	1	0	3	1	1	2	1	1	1	3
61	0	3	1	1	2	1	1	1	3	2	0	3	3	0	0	2	1	2	3	0	3
62	1	0	0	3	2	3	1	0	1	0	3	1	1	2	1	1	1	3	2	0	3
63	2	2	3	2	2	2	1	3	0	1	1	0	0	3	2	3	1	0	1	0	3
64	3	1	2	0	2	0	1	2	2	3	2	2	2	1	3	0	1	1	0	0	3

Figure A. 5: OA.64 – 64 Run Resolution III OA – 4 Levels

Run	Input																															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
3	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
4	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0
5	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
6	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0
7	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0
8	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1
9	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
10	0	0	0	1	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1	0	1	1	1	1	1	1	0	1	0	0	0
11	0	0	1	0	1	1	1	0	1	1	0	1	0	0	0	1	0	0	1	0	1	1	1	0	1	1	0	1	0	0	0	1
12	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	0
13	0	1	0	0	1	0	1	1	1	0	1	1	0	1	0	0	0	1	0	0	1	0	1	1	1	0	1	1	0	1	0	0
14	0	1	0	1	1	1	0	0	1	0	1	0	0	0	1	1	0	1	0	1	1	0	0	1	0	1	0	0	0	0	1	1
15	0	1	1	0	0	1	0	1	1	0	0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	0	0	1	1	0	1	0
16	0	1	1	1	0	0	1	0	1	0	0	0	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	0	1	1	0	1
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
19	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
20	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
21	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
22	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1
23	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1
24	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0
25	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
26	0	0	0	1	0	1	1	1	1	1	1	0	1	0	0	0	1	1	1	0	1	0	0	0	0	0	0	1	0	1	1	1
27	0	0	1	0	1	1	1	0	1	1	0	1	0	0	0	1	1	1	0	1	0	0	0	1	0	0	1	0	1	1	1	0
28	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	0	1	1	0	0	0	1	1	0	0	0	1	1	1	0	0	1
29	0	1	0	0	1	0	1	1	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	0	1	0	0	1	0	1	1
30	0	1	0	1	1	1	0	0	1	0	1	0	0	0	1	1	1	0	1	0	0	0	1	1	0	1	0	1	1	1	0	0
31	0	1	1	0	0	1	0	1	1	0	0	1	1	0	1	0	1	0	0	1	1	0	1	0	0	1	1	0	0	1	0	1
32	0	1	1	1	0	0	1	0	1	0	0	0	1	1	0	1	1	0	0	0	1	1	0	1	0	1	1	1	0	0	1	0
33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
34	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
35	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
36	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
37	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
38	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1
39	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1
40	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0
41	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
42	1	1	1	0	1	0	0	0	0	0	0	1	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1	0	1	1	1
43	1	1	0	1	0	0	0	1	0	0	1	0	1	1	1	0	1	1	0	1	0	0	0	1	0	0	1	0	1	1	1	0
44	1	1	0	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	0	0	0	1	1	1	0	0	1
45	1	0	1	1	0	1	0	0	0	1	0	0	1	0	1	1	1	0	1	1	0	1	0	0	0	1	0	0	1	0	1	1
46	1	0	1	0	0	0	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	0	1	1	0	1	0	1	1	1	0	0
47	1	0	0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	0	0	1	1	0	1	0	0	1	1	0	0	1	0	1

Continued

Continued

48	1	0	0	0	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0	0	1	1	0	1	0	1	1	1	0	0	1	0
49	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
51	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
52	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0
53	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
54	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	1	0
55	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0
56	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1
57	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
58	1	1	1	0	1	0	0	0	0	0	0	1	0	1	1	1	0	0	0	1	0	1	1	1	1	1	0	1	0	0	0	0
59	1	1	0	1	0	0	0	1	0	0	1	0	1	1	1	0	0	0	1	0	1	1	1	0	1	1	0	1	0	0	0	1
60	1	1	0	0	0	1	1	0	0	0	1	1	1	0	0	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	0
61	1	0	1	1	0	1	0	0	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	1	0	1	1	0	1	0	0
62	1	0	1	0	0	0	1	1	0	1	0	1	1	1	0	0	0	1	0	1	1	1	0	0	1	0	1	0	0	0	1	1
63	1	0	0	1	1	0	1	0	0	1	1	0	0	1	0	1	0	1	1	0	0	1	0	1	1	0	0	1	1	0	1	0
64	1	0	0	0	1	1	0	1	0	1	1	1	0	0	1	0	0	1	1	1	0	0	1	0	1	0	0	0	1	1	0	1

Figure A. 6: OA.64.32 – 64 Run Resolution IV OA – 2 Levels

Run	Input											
	1	2	3	4	5	6	7	8	9	10	11	12
1	-15	5	9	-3	7	11	-11	7	-9	3	-15	5
2	-13	1	1	13	-7	-11	11	-7	-1	-13	-13	1
3	-11	7	-7	-11	13	-1	-1	-13	9	-3	15	-5
4	-9	3	-15	5	-13	1	1	13	1	13	13	-1
5	-7	-11	11	-7	11	-7	7	11	5	15	-3	-9
6	-5	-15	3	9	-11	7	-7	-11	13	-1	-1	-13
7	-3	-9	-5	-15	1	13	13	-1	-5	-15	3	9
8	-1	-13	-13	1	-1	-13	-13	1	-13	1	1	13
9	1	13	13	-1	-9	3	-15	5	11	-7	7	11
10	3	9	5	15	9	-3	15	-5	3	9	5	15
11	5	15	-3	-9	-3	-9	-5	-15	-11	7	-7	-11
12	7	11	-11	7	3	9	5	15	-3	-9	-5	-15
13	9	-3	15	-5	-5	-15	3	9	-7	-11	11	-7
14	11	-7	7	11	5	15	-3	-9	-15	5	9	-3
15	13	-1	-1	-13	-15	5	9	-3	7	11	-11	7
16	15	-5	-9	3	15	-5	-9	3	15	-5	-9	3

Figure A. 7: OLHC.16 – 16 Run Resolution II OLHC

Appendix B: VRT Confidence Interval Derivation Assumptions

From [72],

- Let $\hat{F}_n$ be an estimated CDF created using a VRT.

Assumptions (using notation described in Section 4.2.2):

A1. $P\{\hat{F}_n(x)$ is monotonically increasing in x$\} \to 1$ as $n \to \infty$.

A2. For every $a_n = O\left(n^{-1/2}\right)$,

$$\sqrt{n}\left[\left(F(\xi_p + a_n) - F(\xi_p)\right) - \left(\hat{F}_n(\xi_p + a_n) - \hat{F}_n(\xi_p)\right)\right] \Rightarrow 0, \text{ as } n \to \infty.$$

A3. $\sqrt{n}\left[\hat{F}_n(\xi_p) - F(\xi_p)\right] \Rightarrow N(0, \psi_p^2)$ as $n \to \infty$ for some $0 < \psi_p < \infty$.

Appendix C: Complete Confidence for Quantiles Results

The following tables provide the complete results for the experiments performed in Section 4. In additional to the results presented within Section 4, these tables also include the use of asymmetric CFDs for the derivative estimation for the asymptotic methods. As explained in Section 4.2.2.5, using a symmetric CFD may result in the overestimation of the derivative when used near the upper quantiles of the empirical CDF. The hope was that an asymmetric CFD, which was more heavily weighted toward the lower ranges of the distribution, would provide a more accurate derivative estimation. In the tables, three asymmetric CFDs are used. These are titled "Asym CFD for λ" followed by a number: 1.25, 1.50, and 2.00. These numbers indicate the weighting of the asymmetric CFD. For example, 1.25 means that the CFD is found using a high-side point of $F^{-1}(\tilde{\xi}_{p,n} + h_n)$ and a low-side point of $F^{-1}(\tilde{\xi}_{p,n} - 1.25 * h_n)$. So the higher the multiplier, the more heavily the CFD is weighted toward the lower regions of the distribution.

There are also two version of the symmetric CFD presented. One is labeled as "CFD for λ", and the other is labeled "CFD(rounding) for λ". The difference between these two methods has to do with the calculation of the denominator of the CFD in Eq. 55. The first method uses $2 * h_n$ for the denominator. While this may seem like the obvious solution, there is a potential problem. Remember, for the calculation of the inverse CDF F^{-1}, the round-up function is used. This means the distance between the

two points in the numerator of Eq. 55 may not actually be $2 * h_n$ (if a symmetric CFD is used). Instead, the round-up function will cause the distance to be slightly different. In the method with the "rounding" label, the denominator is calculated using the exact distance between the points in the numerator, and not $2 * h_n$. These results were the ones presented in Section 4, since they appeared to offer a more accurate estimation of the derivative.

Table C. 1 and Table C. 2 contain the results for the nonlinear equation with normal inputs when finding a 95/95 and 95/75 value. The main point to note from these results, which has not been previously mentioned, is the accuracy of the asymmetric CFD methods. At lower run levels, the asymmetric CFDs do provide a more accurate derivative estimation than the symmetric CFDs. However, this has some unintended effects when estimating a 95/95 value. As the table shows, when using a symmetric CFD at $n = 59$, the overestimation of the derivative caused the coverage of the CI to be too wide (i.e. >90%), but this also kept the number of 95/95 values falling below the true quantile to remain around 5%. Using the asymmetric CFD, the derivative estimation was more accurate, but the increased accuracy caused the bounds of the CI to narrow. While this meant the coverage was closer to being correct (i.e. ~90%), the amount of trials falling above or below that interval was not equal. Instead, more trials errored to the low side. A closer inspection of the results (not listed here), show that at this run level, the quantile esimation $\tilde{\xi}_{95,m,t}$ tends to error slightly below the true quantile. So with the narrower CI, more 95/95 values will error to the low side, causing the "% below" to be >5%. The exact opposite happens at the lowest run level when estimating a 95/75. Here,

since the quantile estimator $\tilde{\xi}_{75,m,t}$ tends to error above the actual quantile, not only does the coverage improve at $n = 11$ using the asymmetric CFD, but the "% below" also gets closer to 5%. Also, when finding a 95/75, the difference betweent the derivative estimations when using a symmetric or asymmetric CFD is lessened. This is due to the 0.75-quantile lying further from the extremes of the CFD, so the slope of the inverse CFD is also smaller. So the symmetric CFD does not overestimate the derivative to the same extent as when using a 0.95-quantile.

The problem with the asymmetric CFDs, as shown in the tables, is that as the number of runs grows large, they begin to underestimate the derivative. The most accurate derivative estimation may use an asymmetric CFD when the run size is small, then move to a symmetric CFD as the run size grows. It may be possible to use an asymmetric weighting coefficient, similar to the 1.25, 1.50, and 2.00 from above, that is inversely proportional to the number of runs. For example, a formula like $30/n$ could be used for the coefficient. That way, when $n = 60$, the weighting coefficient would be 0.5, but when $n = 500$, it would be only 0.06. This is just one possibility, and more work should be done investigating this method. However, it may be the case that no general formula applies to many systems, and that the best methods are problem-specific.

Table C. 1: Complete Results for Non-linear Eq. Normal Inputs - 95/95

		CFD for λ					**CFD (Rounding) for λ**					**Asym CFD for λ (1.25)**					**Asym CFD for λ(1.5)**					**Asym CFD for λ (2.0)**					**Exact λ**				
				LHS					LHS					LHS					LHS					LHS					LHS		
n	CMC-OS	CMC	AV	*t*=10	*t*=20	*t*=30	CMC	AV	*t*=10	*t*=20	*t*=30	CMC	AV	*t*=10	*t*=20	*t*=30	CMC	AV	*t*=10	*t*=20	*t*=30	CMC	AV	*t*=10	*t*=20	*t*=30	CMC	AV	*t*=10	*t*=20	*t*=30
	57.79	55.24	52.34	51.02	49.24		54.28	51.45	50.23	48.56		53.85	50.94	49.78	48.17		53.27	50.38	49.28	47.72		51.84	49.00	48.05	46.65		49.60	46.78	45.99	44.80	
	16.52	12.39	10.97	9.09	7.93		11.88	10.36	8.56	7.43		11.59	10.01	8.25	7.14		10.96	9.39	7.68	6.62		9.89	8.36	6.75	5.76		5.33	4.04	3.16	3.05	
59	5.10	4.84	6.74	4.61	4.68		5.36	7.87	5.40	5.40		5.77	8.58	5.92	5.88		5.97	9.00	6.13	6.35		7.04	11.21	7.82	8.43		0.95	2.76	2.25	6.41	
		92.04	92.24	94.49	94.49		90.31	90.84	93.31	93.53		89.43	89.90	92.59	92.87		89.22	89.75	92.60	92.51		86.37	87.07	90.27	89.93		85.69	93.13	93.34	90.68	
		277.35	280.05	289.56	298.07		256.76	260.31	269.15	277.06		247.61	248.93	257.39	264.95		235.26	236.55	244.44	250.69		204.56	205.75	212.49	217.48		156.60	156.60	156.60	156.60	
	50.96	49.33	50.67	47.86	45.17	46.20	49.03	50.38	47.51	45.56	45.92	48.89	50.16	47.50	44.95	45.90	48.10	49.70	46.85	44.73	45.39	47.53	48.60	46.37	44.13	45.00	46.86	47.71	45.83	44.18	44.53
	8.45	6.83	7.46	5.25	4.11	4.06	6.69	7.32	5.10	4.32	3.92	6.51	7.07	5.01	3.91	3.84	6.17	6.75	4.74	3.74	3.60	5.81	6.19	4.44	3.44	3.36	3.74	3.68	2.89	2.65	2.63
93	5.48	6.15	4.05	4.14	9.80	4.07	6.59	4.36	4.58	8.51	4.63	6.59	4.39	4.44	10.30	4.64	8.25	4.53	5.67	11.02	5.91	8.92	6.11	6.26	13.94	6.89	2.07	0.45	1.25	8.22	5.65
		91.09	94.03	93.07	87.20	90.59	90.27	93.55	92.16	88.95	89.40	90.77	93.89	92.86	86.98	89.89	87.72	93.69	90.46	86.34	87.39	86.62	91.11	89.26	82.40	85.49	89.69	92.61	89.89	87.44	84.72
		222.99	227.77	224.99	198.47	230.76	215.05	220.84	213.45	213.03	218.92	211.15	215.62	213.11	190.13	217.92	190.03	204.42	191.80	181.67	196.13	174.55	177.97	175.93	159.19	179.26	156.60	156.60	156.60	156.60	156.60
	48.80	49.03	48.29	46.55	45.40	45.23	48.03	48.21	46.38	45.26	45.09	47.52	47.68	46.06	44.99	44.83	47.02	47.17	45.65	44.65	44.49	46.37	46.50	45.00	44.10	43.96	45.47	45.52	44.39	43.61	43.44
	6.14	6.69	5.63	4.12	3.44	3.39	5.39	5.60	4.06	3.38	3.33	5.13	5.32	3.88	3.23	3.17	4.90	5.07	3.69	3.06	3.00	4.58	4.72	3.40	2.81	2.75	3.03	3.03	2.24	2.11	2.15
124	4.90	6.59	4.68	3.88	4.94	5.00	5.07	4.76	4.15	5.17	5.35	5.75	5.43	4.74	5.78	6.05	6.68	6.37	5.51	6.79	7.24	8.01	7.76	7.29	8.83	9.40	3.81	2.72	2.90	7.79	9.55
		90.27	94.13	94.99	92.73	92.70	93.52	94.01	94.64	92.35	92.18	92.66	93.05	94.06	91.61	91.31	91.42	91.83	92.98	90.02	89.84	89.57	90.04	90.43	87.20	86.86	90.71	91.82	92.67	86.37	85.30
		215.05	241.76	240.01	242.57	244.69	236.28	239.30	233.71	236.20	238.26	220.43	223.09	221.50	223.68	225.51	204.93	207.46	205.89	207.73	209.49	184.62	186.75	181.11	182.46	184.01	156.60	156.60	156.60	156.60	156.60
	44.77	44.65	44.77	43.83	42.79	42.87	44.57	44.70	43.76	42.91	42.92	44.40	44.51	43.62	42.71	42.72	44.18	44.28	43.46	42.58	42.59	43.88	43.97	43.22	42.37	42.41	43.91	43.97	43.23	42.55	42.51
	2.78	2.61	2.60	1.96	1.45	1.43	2.59	2.58	1.95	1.47	1.44	2.55	2.54	1.92	1.44	1.41	2.51	2.48	1.88	1.42	1.39	2.44	2.41	1.83	1.39	1.37	1.97	1.91	1.50	1.24	1.26
311	4.88	4.45	3.95	3.34	6.19	5.17	4.76	4.18	3.63	5.37	4.81	5.43	4.86	4.16	6.66	6.33	6.31	5.79	4.99	8.03	7.40	7.69	6.99	6.27	10.33	9.20	3.40	2.80	2.58	5.80	6.72
		93.17	94.24	94.04	90.37	91.48	92.56	93.84	93.48	91.69	91.97	91.65	92.83	92.59	89.76	89.88	90.13	91.45	91.07	87.94	88.24	87.78	89.31	88.41	84.44	85.58	89.98	91.27	89.99	88.76	87.96
		193.32	195.30	195.07	177.04	185.63	189.40	191.65	190.81	186.29	189.13	180.83	182.49	182.36	170.74	174.20	170.20	171.59	171.62	160.91	164.27	155.37	156.38	156.43	144.79	150.54	156.60	156.60	156.60	156.60	156.60
	43.45	43.35	43.33	42.82	42.91	42.21	43.28	43.27	42.77	42.31	42.23	43.22	43.21	42.72	42.21	42.14	43.06	43.05	42.63	42.14	42.07	42.87	42.86	42.45	42.02	41.95	43.04	43.02	42.56	42.14	42.07
	1.89	1.83	1.77	1.35	1.47	1.02	1.82	1.76	1.34	1.08	1.02	1.81	1.75	1.34	1.07	1.02	1.79	1.73	1.32	1.06	1.02	1.76	1.70	1.30	1.05	1.01	1.49	1.43	1.10	0.95	0.93
548	5.62	5.69	5.36	3.94	5.37	5.61	6.02	5.80	4.30	5.34	5.40	6.35	6.17	4.61	6.55	6.44	7.44	7.33	5.24	7.40	7.38	9.17	8.86	7.01	9.32	9.23	4.36	3.51	3.17	5.26	6.03
		91.34	91.92	92.87	91.69	90.79	90.84	91.16	92.20	91.45	91.07	90.38	90.86	91.86	89.76	89.64	88.51	89.04	90.78	88.50	88.29	85.72	86.63	87.36	85.65	85.43	89.47	90.61	89.82	89.33	88.87
		177.22	176.95	178.84	186.29	172.13	172.86	172.60	174.75	175.02	173.91	168.78	168.64	170.39	165.34	164.46	158.25	158.05	162.79	158.27	157.41	146.02	145.99	147.40	146.09	145.22	156.60	156.60	156.60	156.60	156.60
	42.62	42.51	42.52	42.12	41.80	41.73	42.49	42.51	42.11	41.83	41.76	42.51	42.46	42.06	41.76	41.69	42.37	42.39	42.02	41.72	41.66	42.26	42.28	41.93	41.66	41.59	42.39	42.41	42.03	41.77	41.71
	1.27	1.26	1.24	0.95	0.77	0.73	1.26	1.24	0.95	0.77	0.73	1.26	1.23	0.95	0.77	0.73	1.25	1.23	0.94	0.77	0.73	1.24	1.22	0.94	0.77	0.73	1.08	1.06	0.82	0.70	0.67
1008	5.54	6.17	5.98	5.28	6.19	6.22	6.33	6.12	5.39	5.89	5.58	6.17	6.55	5.95	6.96	7.15	7.62	7.20	6.64	7.66	7.94	8.71	8.47	7.96	9.06	9.40	4.76	4.08	3.66	5.07	5.44
		90.41	90.34	90.87	89.67	89.34	90.19	90.07	90.67	90.18	90.35	90.41	89.41	89.66	88.64	88.30	88.19	88.49	88.52	87.65	87.13	86.12	86.06	86.23	85.32	84.85	89.75	89.90	89.89	89.35	89.29
		166.39	166.59	166.96	162.33	160.44	165.09	165.28	165.82	165.59	165.29	166.39	160.52	160.89	156.81	155.18	154.65	154.95	155.30	151.73	150.26	144.73	145.02	145.30	142.47	141.12	156.60	156.60	156.60	156.60	156.60
	42.03	41.94	41.91	41.63	41.49	41.47	41.93	41.90	41.62	41.48	41.49	41.91	41.88	41.59	41.46	41.45	41.87	41.83	41.57	41.44	41.44	41.80	41.77	41.52	41.40	41.40	41.89	41.86	41.59	41.45	41.46
	0.87	0.87	0.85	0.66	0.54	0.52	0.87	0.84	0.66	0.54	0.52	0.87	0.84	0.65	0.54	0.52	0.86	0.84	0.65	0.54	0.52	0.86	0.84	0.65	0.54	0.52	0.77	0.75	0.58	0.50	0.48
2004	4.95	6.26	6.12	6.06	5.59	4.99	6.36	6.23	6.15	5.75	4.60	6.71	6.51	6.78	6.43	5.36	7.46	7.26	7.11	6.74	5.90	8.45	8.25	8.22	7.63	6.94	5.08	4.69	4.69	4.98	4.31
		89.82	90.07	89.94	89.95	89.39	89.65	89.90	89.75	89.76	90.16	89.25	89.27	88.80	88.77	88.90	88.04	88.12	88.38	88.34	88.15	86.30	86.50	86.55	86.71	86.19	89.37	89.74	89.65	89.27	89.17
		162.83	162.93	162.93	163.67	158.86	161.98	162.09	161.92	162.65	161.87	159.32	159.29	156.97	157.55	155.78	153.71	153.71	153.75	154.25	152.54	145.79	145.77	145.80	146.23	144.76	156.60	156.60	156.60	156.60	156.60

Table C. 2: Complete Results for Non-linear Eq. Normal Inputs - 95/75

		CFD for λ					**CFD (Rounding) for λ**					**Asym CFD for λ (1.25)**					**Asym CFD for λ(1.5)**					**Asym CFD for λ (2.0)**					**Exact λ**				
n	CMC-OS	CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS		
				t=5	t=10	t=15			t=5	t=10	t=15			t=5	t=10	t=15			t=5	t=10	t=15			t=5	t=10	t=15			t=5	t=10	t=15
	42.08	37.76	35.06	35.69			38.22	35.72	35.88			37.49	35.00	35.69			36.77	34.67	34.98			36.37	33.93	34.32			35.79	33.19	33.05		
	12.02	7.17	5.44	4.51			7.44	5.88	4.63			6.85	5.18	4.51			6.45	4.89	3.86			5.96	4.41	3.42			3.71	2.36	2.04		
11	4.22	6.04	8.23	1.01			5.38	6.70	0.93			6.13	7.74	1.01			7.46	8.28	1.36			7.30	10.86	2.15			0.95	0.08	1.64		
		88.56	90.05	94.38			90.05	91.98	94.96			88.31	91.26	94.38			85.21	90.59	93.46			85.11	87.34	90.81			87.77	95.67	86.09		
		34.83	36.55	47.97			36.97	40.52	49.54			33.33	36.23	47.97			29.99	34.24	41.98			28.13	29.86	36.35			25.43	25.43	25.43		
	34.44	33.90	32.82	31.73	31.46		33.65	33.13	31.94	31.65		33.55	32.67	31.61	31.36		33.30	32.70	31.62	31.38		32.91	32.52	31.44	31.21		32.62	32.66	31.57	31.38	
	3.54	3.16	2.50	1.61	1.53		3.07	2.60	1.67	1.59		3.01	2.45	1.58	1.50		2.92	2.44	1.58	1.51		2.77	2.37	1.53	1.46		2.01	1.69	1.34	1.40	
29	4.37	4.96	5.58	5.77	7.03		5.45	4.36	4.51	5.77		5.63	6.28	6.62	8.01		6.53	5.96	6.33	8.12		7.96	6.83	7.91	9.73		3.56	0.41	4.54	6.76	
		93.06	90.69	87.68	82.94		92.47	92.92	90.44	85.66		92.42	89.77	86.35	81.33		91.42	90.85	87.20	81.83		89.21	89.59	84.29	78.63		91.20	93.34	87.44	82.29	
		34.69	26.80	27.82	27.01		33.22	29.36	30.48	29.59		32.47	25.49	26.39	25.66		30.59	25.77	26.45	25.82		27.63	24.24	24.14	23.57		25.43	25.43	25.43	25.43	
	33.47	32.85	32.28	31.55	31.29	30.89	32.75	32.20	31.48	31.23	30.95	32.56	32.05	31.36	31.12	30.83	32.43	31.94	31.36	31.04	30.77	32.12	31.73	31.06	30.86	30.63	32.02	31.59	30.97	30.79	30.66
	2.74	2.40	1.99	1.34	1.35	1.42	2.37	1.96	1.33	1.33	1.45	2.32	1.92	1.30	1.29	1.39	2.27	1.89	1.30	1.27	1.36	2.18	1.84	1.23	1.22	1.29	1.70	1.38	1.09	1.15	1.27
40	4.23	5.17	4.60	3.55	6.61	15.25	5.62	4.97	3.91	7.10	14.96	6.67	5.75	4.85	7.98	15.72	7.34	6.34	4.85	8.92	16.07	9.17	7.88	7.72	10.99	17.16	4.41	2.70	6.96	11.16	16.19
		93.46	93.85	94.59	89.97	71.59	92.79	93.32	94.09	89.18	72.13	91.56	92.25	92.90	88.03	71.17	90.78	91.55	92.90	86.89	70.61	88.26	89.59	89.01	83.90	68.77	91.28	91.96	88.92	83.35	69.98
		32.80	32.93	33.84	33.25	30.23	31.92	32.04	32.93	32.35	31.20	30.24	30.35	31.14	30.63	28.96	29.05	29.25	31.14	29.41	27.80	26.30	26.94	26.98	26.62	25.17	25.43	25.43	25.43	25.43	25.43
	31.21	31.13	30.70	30.51	30.27	30.40	31.13	30.71	30.51	30.29	30.40	31.04	30.66	30.46	30.23	30.35	30.99	30.59	30.42	30.21	30.32	30.88	30.52	30.36	30.14	30.26	31.01	30.60	30.44	30.22	30.34
	1.17	1.13	0.91	0.65	0.59	0.61	1.13	0.91	0.65	0.59	0.61	1.12	0.90	0.65	0.59	0.61	1.11	0.90	0.65	0.59	0.61	1.10	0.89	0.65	0.59	0.60	0.96	0.76	0.61	0.56	0.58
135	4.97	5.12	6.19	3.61	6.29	4.28	5.10	6.03	3.59	5.90	4.25	6.03	6.91	4.50	7.12	5.06	6.72	7.91	5.00	7.85	5.53	8.02	9.25	6.18	9.73	6.93	3.57	4.02	3.55	6.65	4.71
		90.67	91.15	90.53	89.41	88.40	90.71	91.36	90.57	90.04	88.47	89.01	90.16	88.88	88.43	86.97	88.03	88.69	87.98	87.37	86.13	85.44	86.61	85.54	84.60	83.26	89.80	90.92	88.79	88.20	86.43
		27.33	27.43	27.55	26.99	27.34	27.37	27.57	27.60	27.53	27.39	25.90	26.49	26.09	26.09	25.90	25.11	25.17	25.27	25.25	25.10	23.33	23.74	23.48	23.42	23.34	25.43	25.43	25.43	25.43	25.43
	30.66	30.63	30.45	30.14	30.11	30.05	30.61	30.43	30.14	30.10	30.05	30.59	30.41	30.12	30.09	30.03	30.54	30.37	30.10	30.07	30.01	30.48	30.32	30.06	30.03	29.97	30.56	30.38	30.11	30.08	30.02
	0.81	0.80	0.67	0.47	0.44	0.45	0.79	0.67	0.47	0.44	0.45	0.79	0.67	0.47	0.44	0.45	0.79	0.67	0.47	0.45	0.45	0.78	0.66	0.47	0.45	0.45	0.70	0.58	0.44	0.42	0.43
246	5.16	4.86	4.40	4.99	4.81	6.32	5.17	4.66	4.98	4.98	6.39	5.38	4.96	5.50	5.39	7.01	6.27	5.71	6.20	5.92	7.64	7.20	6.71	7.38	7.17	9.35	3.88	3.30	4.88	5.02	6.83
		91.43	91.72	90.36	89.99	89.43	90.89	91.11	90.50	89.63	89.25	90.55	90.82	89.73	89.04	88.43	89.09	89.38	88.70	88.26	87.50	87.19	87.54	86.81	86.08	85.11	89.98	90.29	89.48	88.61	88.13
		27.17	27.23	26.64	26.88	26.90	26.63	26.69	26.73	26.57	26.72	26.26	26.33	25.80	26.02	26.06	25.08	25.15	25.04	25.22	25.26	23.73	23.78	23.70	23.81	23.61	25.43	25.43	25.43	25.43	25.43
	30.24	30.29	30.09	29.97	29.89	29.89	30.29	30.09	29.97	29.89	29.89	30.26	30.06	29.95	29.87	29.87	30.23	30.04	29.93	29.86	29.85	30.18	30.01	29.90	29.83	29.83	30.25	30.06	29.94	29.87	29.87
	0.57	0.57	0.46	0.34	0.31	0.31	0.57	0.46	0.34	0.31	0.31	0.57	0.46	0.34	0.31	0.31	0.56	0.46	0.34	0.31	0.31	0.56	0.46	0.34	0.31	0.31	0.52	0.41	0.32	0.30	0.30
459	6.18	5.08	5.63	4.05	5.16	4.97	5.08	5.62	3.99	5.11	4.87	5.72	6.47	4.70	6.03	5.58	6.36	7.04	5.21	6.73	6.28	7.38	8.33	6.39	8.04	7.61	4.15	4.54	3.91	5.65	5.35
		91.06	91.00	91.24	91.08	90.58	91.05	91.03	91.34	91.15	90.73	89.81	89.84	90.23	89.70	89.61	88.81	88.89	89.30	88.81	88.52	86.79	86.79	87.11	86.72	86.13	89.99	90.49	90.08	89.66	89.11
		26.62	26.68	26.63	26.68	26.51	26.62	26.70	26.72	26.70	26.68	25.65	25.70	25.68	25.70	25.77	24.77	25.02	25.01	25.02	24.88	23.42	23.62	23.50	23.61	23.50	25.43	25.43	25.43	25.43	25.43
	30.03	30.02	29.91	29.77	29.76	29.75	30.02	29.91	29.78	29.76	29.75	30.00	29.90	29.76	29.75	29.74	29.98	29.89	29.76	29.74	29.73	29.96	29.86	29.74	29.72	29.72	30.00	29.90	29.77	29.75	29.74
	0.39	0.39	0.33	0.24	0.22	0.22	0.39	0.33	0.24	0.22	0.22	0.39	0.33	0.24	0.23	0.22	0.39	0.33	0.24	0.23	0.22	0.39	0.33	0.24	0.23	0.22	0.37	0.30	0.23	0.22	0.22
886	4.83	5.21	5.35	4.73	4.91	4.85	5.19	5.29	4.71	4.78	4.85	5.75	5.93	5.28	5.46	5.41	6.16	6.40	5.71	5.99	5.96	6.97	7.14	6.72	7.02	7.04	4.37	4.54	4.57	4.62	4.74
		90.47	90.28	90.98	90.18	90.23	90.50	90.39	91.00	90.42	90.24	89.55	89.41	90.17	89.32	89.24	88.81	88.58	89.36	88.48	88.27	87.21	86.93	87.67	86.72	86.63	90.11	89.62	90.58	89.98	89.57
		26.00	26.01	25.96	25.91	26.04	26.05	26.06	25.99	26.01	26.07	25.34	25.36	25.31	25.24	25.37	24.72	24.74	24.71	24.64	24.75	23.62	23.63	23.60	23.55	23.64	25.43	25.43	25.43	25.43	25.43

Table C. 3 and Table C. 4 contain the full results for the nonlinear equation with non-normal inputs. The results of the asymmetric CFD are very similar to the previous example. When finding a 95/95 value, the asymmetric CFD provides a more accurate derivative estimation. However, this increased accuracy causes the CI to narrow, and with the quantile estimation again erroring to the low side, more than 5% of the 95/95 trials fall below the true quantile. Again, the opposite occurs when finding a 95/75, since the quantile estimation tends to error to the high side, so the narrower CI improves the coverage, and the "% below". When using the larger weighting for the asymmetric CFD when finding a 95/75, the derivative is underestimated even at the lowest run level. Once again, this is because the derivative of the inverse CDF at the 0.75-quantile is already much smaller since it is further from 1.00, and the symmetric CFD estimation is better than at the 0.95-quantile. Also, as in the previous example, the asymmetric CFDs begin to underestimate the derivative as the number of runs grows large.

Table C. 3: Complete Results for Non-linear Eq. Non-normal Inputs - 95/95

		CFD for λ					**CFD (Rounding) for λ**					**Asym CFD for λ (1.25)**					**Asym CFD for λ(1.5)**					**Asym CFD for λ (2.0)**					**Exact λ**				
n	CMC-OS	CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS		
				t=10	t=20	t=30			t=10	t=20	t=30			t=10	t=20	t=30			t=10	t=20	t=30			t=10	t=20	t=30			t=10	t=20	t=30
	153.12	154.50	141.66	136.05	129.99		151.04	138.67	133.68	127.97		149.59	136.95	132.31	126.80		149.30	136.77	132.76	126.84		145.41	133.13	129.91	124.40		145.13	132.92	129.10	123.36	
	31.63	29.52	25.81	20.61	17.25		28.53	25.00	19.90	16.59		28.12	24.55	19.49	16.21		27.69	24.46	19.53	16.10		26.62	23.63	18.52	15.25		21.11	19.31	13.94	12.93	
59	4.82	4.02	7.71	5.29	5.89		4.98	9.00	6.64	7.36		5.43	9.88	7.54	8.37		5.46	10.29	7.22	8.31		6.77	12.94	9.21	10.71		3.71	9.23	3.37	7.78	
		91.50	90.35	93.30	92.98		89.13	88.54	91.57	91.22		88.14	87.31	90.32	89.98		89.03	87.62	91.62	90.49		86.57	84.71	89.19	87.68		89.34	89.12	94.02	90.11	
		946.44	939.69	944.24	980.42		872.37	873.46	877.69	911.32		841.28	835.28	839.32	871.49		835.05	831.24	852.48	870.94		751.68	750.39	772.39	786.70		745.73	745.73	745.73	745.73	
	139.86	138.68	143.07	132.83	120.86	126.19	137.53	141.98	131.54	122.20	125.21	138.08	142.43	132.60	121.05	125.93	134.92	141.67	130.11	121.04	124.06	134.12	138.05	129.80	119.33	123.64	134.28	138.22	129.58	119.33	123.39
	21.43	20.85	21.22	17.02	13.20	13.52	20.61	20.96	16.71	13.58	13.22	20.70	20.91	16.93	13.37	13.40	20.09	20.65	16.37	13.39	12.84	19.93	19.92	16.12	12.90	12.62	16.81	16.14	12.49	12.90	11.23
93	5.01	5.55	3.49	4.31	12.59	5.01	6.01	3.73	5.02	10.68	5.82	5.90	3.64	4.41	12.82	5.29	7.66	3.95	5.94	13.14	7.01	8.28	5.28	6.26	16.12	7.25	5.25	2.52	1.80	16.12	5.24
		89.95	93.26	92.28	84.16	89.58	88.98	92.72	90.95	86.64	87.84	90.31	93.57	92.87	84.72	89.35	86.88	93.60	89.88	84.66	86.04	87.20	91.63	90.02	80.99	85.73	90.14	94.48	91.96	80.99	86.86
		863.91	862.48	858.81	767.71	884.85	833.13	836.20	814.74	823.99	839.44	847.78	847.08	850.80	772.47	870.86	763.00	828.93	765.72	770.04	783.77	741.43	741.69	755.07	696.73	762.20	745.73	745.73	745.73	696.73	745.73
	133.98	132.03	132.52	126.36	121.16	120.84	131.75	132.23	125.79	120.71	120.40	130.54	130.99	125.32	120.30	119.98	129.25	129.67	124.40	119.55	119.22	128.07	128.51	122.80	118.30	118.00	127.74	128.14	123.14	118.48	118.07
	17.25	16.51	16.54	13.11	11.38	11.19	16.46	16.49	13.01	11.25	11.05	16.32	16.35	12.97	11.17	10.95	16.19	16.25	12.83	10.97	10.73	16.17	16.25	12.54	10.63	10.37	14.22	14.17	10.41	9.83	9.62
124	5.01	6.03	5.56	5.31	9.04	8.36	6.18	5.79	5.72	9.61	8.99	7.16	6.52	6.16	10.11	9.72	7.89	7.66	7.13	11.09	11.13	9.16	8.99	8.78	13.19	12.97	7.36	6.75	4.27	11.57	11.44
		91.51	92.21	92.80	87.98	88.43	91.23	91.90	92.18	87.15	87.55	90.10	91.04	92.05	86.69	86.81	89.41	89.80	91.01	85.34	84.88	88.37	88.65	88.82	82.51	82.15	90.13	90.53	92.43	83.68	83.42
		879.02	880.82	876.61	898.85	902.31	870.07	871.86	853.58	875.24	878.60	832.65	833.75	834.77	851.71	854.74	792.51	793.16	797.81	810.92	813.31	755.76	757.15	733.68	742.92	745.61	745.73	745.73	745.73	745.73	745.73
	123.01	123.08	123.50	119.52	114.69	115.00	122.74	123.19	119.25	115.15	115.18	122.36	122.79	119.01	114.69	114.66	121.74	122.16	118.57	114.36	114.35	121.03	121.45	118.08	113.81	114.03	121.57	121.99	118.40	114.59	114.53
	9.82	9.90	9.82	7.84	6.11	6.13	9.88	9.80	7.81	6.17	6.16	9.90	9.81	7.83	6.16	6.13	9.89	9.80	7.81	6.14	6.12	9.92	9.82	7.77	6.07	6.10	9.20	9.03	6.81	5.71	5.79
311	4.85	4.91	4.04	4.16	8.74	8.30	5.26	4.28	4.57	7.69	7.85	5.82	4.92	5.00	8.80	9.23	6.53	5.65	5.72	9.88	10.24	7.68	6.72	6.54	11.71	11.32	5.29	4.89	3.62	7.78	8.52
		91.11	92.57	92.02	87.05	87.91	90.44	92.10	91.26	88.82	88.65	89.83	91.37	90.87	87.33	86.77	88.76	90.27	89.88	85.75	85.47	87.21	88.90	88.60	83.02	84.09	90.13	90.84	90.44	87.13	87.09
		820.06	819.08	820.24	760.56	790.39	803.44	803.77	802.32	800.32	805.29	784.88	784.49	786.46	759.70	760.84	754.07	754.11	757.29	731.35	734.52	719.40	719.53	725.13	682.85	706.03	745.73	745.73	745.73	745.73	745.73
	118.18	118.31	118.40	115.94	113.19	112.89	118.02	118.10	115.73	113.26	112.96	118.00	118.09	115.71	113.05	112.74	117.42	117.51	115.50	112.90	112.60	116.95	117.05	114.95	112.63	112.34	117.59	117.69	115.39	112.98	112.67
	7.27	7.39	7.26	5.74	4.61	4.45	7.37	7.24	5.72	4.61	4.46	7.41	7.27	5.76	4.63	4.47	7.40	7.26	5.77	4.63	4.48	7.44	7.27	5.74	4.62	4.47	7.02	6.79	5.13	4.30	4.20
548	5.74	5.65	5.13	4.92	7.54	7.78	6.17	5.65	5.33	7.39	7.54	6.35	5.62	5.43	8.15	8.51	7.28	6.69	5.91	8.71	9.11	8.44	7.65	7.00	9.66	10.05	6.22	5.07	4.03	7.23	7.54
		90.33	91.28	90.66	88.72	88.49	89.53	90.34	89.82	88.98	88.89	89.60	90.64	90.09	88.10	87.73	87.99	89.12	89.48	87.39	86.91	86.51	87.71	87.63	86.15	85.49	89.70	90.83	90.06	88.27	87.78
		793.09	792.57	794.94	773.79	773.78	773.57	773.07	776.79	781.79	781.79	772.72	772.25	774.76	757.30	756.81	734.48	734.06	756.51	740.83	740.46	703.68	703.60	707.92	710.17	709.65	745.73	745.73	745.73	745.73	745.73
	115.26	115.23	115.04	113.19	111.46	111.12	115.16	114.97	113.18	111.56	111.26	115.06	114.89	113.10	111.38	111.05	114.90	114.74	112.99	111.30	110.98	114.61	114.45	112.78	111.15	110.84	114.96	114.79	113.02	111.44	111.16
	5.34	5.41	5.27	4.16	3.29	3.21	5.40	5.27	4.10	3.30	3.22	5.42	5.30	4.12	3.31	3.24	5.44	5.31	4.13	3.32	3.25	5.45	5.33	4.13	3.32	3.25	5.19	5.04	3.75	3.09	3.04
1008	5.69	5.60	5.47	5.27	7.13	8.48	5.73	5.59	5.16	6.63	7.61	6.20	5.96	5.53	7.66	8.78	6.59	6.35	6.00	8.03	9.25	7.45	7.16	6.75	9.05	10.12	5.60	5.25	4.51	6.38	6.87
		89.67	89.97	90.06	88.99	87.02	89.40	89.70	90.19	89.78	88.14	88.92	89.44	89.93	88.35	86.82	88.48	88.84	89.30	87.80	86.28	87.34	87.62	88.21	86.63	85.26	89.75	90.18	89.98	89.21	88.51
		769.00	767.99	768.22	750.81	741.86	762.97	761.96	764.98	765.89	764.29	754.22	754.01	755.65	738.94	730.91	740.42	740.55	742.10	727.18	719.54	714.68	714.85	716.75	704.11	698.38	745.73	745.73	745.73	745.73	745.73
	112.77	112.59	112.43	111.05	110.21	110.19	112.56	112.40	111.03	110.19	110.26	112.54	112.39	110.94	110.13	110.18	112.40	112.25	110.92	110.11	110.16	112.23	112.09	110.92	110.01	110.06	112.45	112.31	110.96	110.13	110.20
	3.67	3.74	3.71	2.89	2.33	2.24	3.74	3.70	2.89	2.33	2.25	3.76	3.72	2.90	2.33	2.25	3.77	3.72	2.92	2.34	2.27	3.78	3.73	2.92	2.35	2.27	3.66	3.58	2.70	2.21	2.14
2004	4.99	5.91	6.13	6.46	6.57	5.87	5.93	6.19	6.58	6.70	5.59	6.07	6.31	7.06	7.16	6.09	6.62	6.77	7.23	7.27	6.19	7.27	7.42	7.23	7.92	6.82	5.88	5.98	5.66	5.92	5.11
		89.71	89.37	89.40	89.50	88.98	89.64	89.25	89.14	89.27	89.59	89.54	89.29	88.60	88.61	88.84	88.82	88.66	88.48	88.60	88.82	88.12	87.70	88.48	87.68	87.76	89.93	89.49	89.48	89.22	89.28
		763.26	761.55	761.07	763.52	746.72	759.30	757.59	756.36	758.79	760.86	757.41	756.27	742.46	745.33	743.38	739.12	738.26	738.20	741.08	738.99	718.65	716.99	738.20	720.16	718.36	745.73	745.73	745.73	745.73	745.73

Table C. 4: Complete Results for Non-linear Eq. Non-normal Inputs - 95/75

		CFD for λ					**CFD (Rounding) for λ**					**Asym CFD for λ (1.25)**					**Asym CFD for λ(1.5)**					**Asym CFD for λ (2.0)**					**Exact λ**				
				LHS					LHS					LHS					LHS					LHS					LHS		
n	CMC-OS	CMC	AV	t=5	t=10	t=15	CMC	AV	t=5	t=10	t=15	CMC	AV	t=5	t=10	t=15	CMC	AV	t=5	t=10	t=15	CMC	AV	t=5	t=10	t=15	CMC	AV	t=5	t=10	t=15
		79.15	64.12	61.88			81.39	67.49	68.90			76.17	62.10	59.46			72.82	59.54	57.37			68.70	55.33	53.66			75.40	58.59	58.71		
		34.71	22.55	13.92			35.62	23.84	16.50			33.60	21.83	12.83			32.30	20.83	11.97			30.76	19.35	10.79			22.76	13.78	7.54		
11	100.4528	7.23	6.76	0.53			6.71	5.69	0.14			8.05	7.35	0.70			9.32	8.52	0.89			10.91	11.05	1.78			0.01	0.01	0.01		
		83.53	90.38	98.58			85.38	92.27	99.60			80.93	89.74	98.14			77.30	87.88	97.67			72.38	83.48	95.61			85.40	94.47	99.72		
		169.68	185.13	171.43			180.09	205.22	214.13			155.78	173.10	156.75			140.20	157.83	144.02			120.99	132.79	121.49			152.20	152.20	152.20		
		61.78	55.72	46.14	43.68		60.64	57.57	47.29	44.55		59.55	54.32	45.27	42.98		57.62	53.52	44.64	42.53		54.50	51.29	43.14	41.41		55.12	54.57	44.75	43.16	
		18.84	15.99	6.43	4.91		18.47	16.64	6.68	5.04		18.21	15.55	6.23	4.78		17.67	15.24	6.06	4.69		16.78	14.44	5.78	4.55		11.92	10.38	5.12	4.44	
29	67.62201	5.09	5.23	1.43	1.30		5.63	4.46	0.85	0.64		6.03	6.08	2.07	1.89		7.13	6.49	2.49	2.28		9.26	8.44	4.30	4.23		0.02	0.01	0.69	0.56	
		92.65	90.58	95.83	94.43		91.75	92.64	97.20	96.07		91.26	89.09	94.88	92.76		89.49	88.43	93.94	91.67		85.95	83.95	90.30	87.09		92.74	91.89	97.01	95.34	
		202.60	161.83	172.23	161.33		193.96	177.28	188.67	176.72		185.72	150.11	159.68	149.11		171.14	143.40	150.65	141.13		147.55	124.72	129.22	121.49		152.20	152.20	152.20	152.20	
		55.66	52.32	44.83	42.26	38.02	55.08	51.83	44.49	42.01	38.34	53.63	50.62	43.65	41.36	37.77	52.21	49.44	42.80	40.71	37.50	49.65	47.32	41.30	39.58	37.10	51.07	48.16	41.67	39.76	37.52
		14.65	12.10	5.39	4.43	8.19	14.50	11.98	5.34	4.38	8.83	14.17	11.70	5.19	4.25	7.69	13.86	11.42	5.03	4.12	7.17	13.26	10.90	4.80	3.92	6.44	9.98	8.41	4.30	3.77	7.11
40	61.47786	4.87	3.88	1.54	1.47	39.19	5.16	4.25	1.79	1.72	39.19	6.28	5.24	2.55	2.47	39.19	7.54	6.38	3.64	3.69	39.19	10.48	9.05	6.53	7.25	39.19	0.30	0.26	2.66	5.41	39.19
		93.09	94.07	97.98	97.54	16.95	92.62	93.35	97.68	97.11	16.95	91.18	91.81	96.74	95.98	16.94	89.45	90.14	95.42	94.44	16.93	84.88	85.50	91.90	89.80	16.90	93.62	93.23	96.37	92.10	16.95
		193.00	197.47	204.46	207.61	179.79	187.79	192.14	198.94	202.00	196.95	174.93	179.03	185.02	187.73	166.10	162.38	166.19	170.83	173.40	151.52	139.57	143.14	146.09	148.31	129.77	152.20	152.20	152.20	152.20	152.20
		45.20	42.52	39.28	37.70	38.17	45.21	42.57	39.29	37.78	38.18	44.44	42.00	38.92	37.45	37.90	43.88	41.44	38.65	37.23	37.70	42.83	40.63	38.14	36.81	37.32	44.38	41.77	38.80	37.43	37.88
		7.49	5.88	2.65	2.10	2.14	7.49	5.89	2.65	2.11	2.14	7.37	5.80	2.64	2.09	2.11	7.28	5.70	2.64	2.08	2.09	7.10	5.56	2.61	2.07	2.06	5.79	4.57	2.56	2.05	2.08
135	46.33832	5.41	6.25	2.52	3.95	2.44	5.40	6.12	2.48	3.62	2.42	6.52	7.46	3.45	5.18	3.21	7.50	8.93	4.49	6.67	3.97	9.98	11.23	6.49	10.14	6.16	1.72	2.35	3.73	4.96	3.30
		90.41	90.97	93.48	93.50	89.73	90.45	91.15	93.57	93.95	89.78	88.36	89.23	91.42	91.68	87.35	86.48	86.87	89.37	89.61	85.31	81.72	82.83	84.77	84.51	80.27	90.29	91.40	91.32	92.13	87.28
		165.46	167.17	169.20	163.64	165.90	165.73	168.06	169.48	166.92	166.17	153.09	156.63	156.52	153.31	153.39	143.93	145.50	147.07	143.87	144.04	126.81	129.15	129.26	126.51	126.68	152.20	152.20	152.20	152.20	152.20
		41.96	40.66	37.48	37.12	36.56	41.81	40.54	37.50	37.09	36.54	41.54	40.31	37.30	36.96	36.42	41.10	39.94	37.12	36.81	36.29	40.45	39.40	36.82	36.55	36.04	41.43	40.20	37.27	36.94	36.41
		5.25	4.23	1.89	1.60	1.52	5.23	4.21	1.89	1.60	1.52	5.20	4.19	1.90	1.60	1.51	5.14	4.14	1.91	1.60	1.51	5.06	4.06	1.91	1.59	1.49	4.21	3.41	1.88	1.57	1.49
246	42.48112	5.50	4.77	4.02	3.06	5.57	5.85	5.12	3.97	3.18	5.69	6.44	5.73	5.13	3.84	6.77	7.56	6.85	6.34	4.58	7.95	9.52	8.90	8.58	6.56	10.73	2.64	2.34	5.12	3.68	6.67
		91.34	91.99	93.05	92.75	90.53	90.70	91.41	93.10	92.37	90.35	89.64	90.47	91.44	91.11	88.70	87.73	88.29	89.82	89.48	86.84	83.93	84.58	86.21	85.56	82.63	90.49	90.99	91.78	91.38	88.91
		163.78	164.22	162.33	162.53	161.54	160.55	160.98	162.88	160.61	160.42	154.52	154.91	153.43	153.32	152.55	144.79	145.12	145.21	144.98	144.34	130.50	130.84	130.80	130.38	128.97	152.20	152.20	152.20	152.20	152.20
		39.72	38.56	36.51	36.17	36.02	39.72	38.30	36.51	36.17	36.04	39.41	38.09	36.36	36.05	35.93	39.13	37.70	36.24	35.95	35.83	38.66	38.34	36.02	35.76	35.66	39.48	38.56	36.37	36.07	35.94
		3.53	2.91	1.38	1.18	1.08	3.53	2.89	1.38	1.18	1.08	3.51	2.87	1.39	1.18	1.08	3.49	2.83	1.40	1.18	1.08	3.44	2.49	1.40	1.18	1.07	3.02	2.91	1.38	1.17	1.07
459	39.5655	4.98	5.26	4.53	4.43	4.14	4.99	6.48	4.49	4.41	4.02	6.13	7.51	5.80	5.72	5.30	7.11	9.81	6.93	6.92	6.39	9.11	3.57	9.59	9.16	8.55	2.94	5.26	5.57	5.10	4.80
		91.51	91.43	92.70	92.00	91.53	91.50	89.47	92.77	92.04	91.70	89.74	87.67	90.77	89.84	89.57	87.84	83.51	89.00	87.87	87.50	83.96	90.46	84.75	83.63	83.47	90.83	91.45	91.29	90.62	90.22
		159.41	160.20	161.05	160.29	159.54	159.37	150.82	161.19	160.43	160.55	150.01	143.04	151.50	150.81	151.09	141.61	128.58	143.68	142.96	142.51	127.50	152.20	129.12	128.47	128.13	152.20	160.34	152.20	152.20	152.20
		38.02	37.37	35.87	35.67	35.49	38.02	37.38	35.88	35.67	35.50	37.85	37.24	35.80	35.60	35.44	37.70	37.11	35.73	35.55	35.38	37.43	36.89	35.60	35.44	35.29	37.95	37.31	35.84	35.64	35.47
		2.45	1.99	1.00	0.85	0.79	2.45	1.99	1.00	0.85	0.79	2.45	1.98	1.00	0.85	0.79	2.44	1.98	1.01	0.86	0.79	2.42	1.96	1.01	0.85	0.79	2.17	1.77	1.00	0.84	0.78
886	38.18945	5.38	4.71	4.85	3.97	5.00	5.35	4.68	4.82	3.91	4.98	6.18	5.44	5.84	4.78	5.89	6.89	6.16	6.70	5.50	6.68	8.78	7.74	8.55	7.37	8.55	3.48	3.14	5.22	4.14	5.12
		90.76	91.52	91.06	91.09	90.44	90.82	91.58	91.10	91.18	90.46	89.36	89.97	89.59	89.64	88.95	87.91	88.50	88.08	88.04	87.39	84.50	85.43	84.77	84.49	83.79	90.77	91.02	90.71	90.50	90.00
		154.94	155.22	155.03	154.77	154.73	155.21	155.49	155.21	155.38	154.91	148.07	148.40	148.12	147.93	147.88	141.71	142.03	141.69	141.56	141.47	130.22	130.53	130.18	130.06	129.92	152.20	152.20	152.20	152.20	152.20

Table C. 5 and Table C. 6 present the complete results for the LOCA response surface experiment. As in the previous two examples, when finding a 95/95 value at $n = 60$, the quantile estimation of the asymptotic methods tends to underestimate the true quantile. This means the asymmetric CFD methods improve the coverage, with the accurate derivative estimation, but the "% below" is >5%. Also, unlike the previous two examples, when finding a 95/75, the asymmetric CFDs actual overestimate the derivative when compared to the symmetric CFDs, at the lowest run level. This is the only experiment where that phenomena is seen. As Figure 4. 24 showed, the output distribution was very compact at the high end, and the 0.75-quantile was very close to the peak of the histogram. This means, at $n = 11$, the low-side point of the asymmetric CFD is going to fall to a much lower value, and the derivative estimation will actually be too large. As the run size grows larger, this problem disappears, and the asymmetric CFD underestimates the derivative when compared to the symmetric CFDs.

Table C. 5: Complete Results for LOCA Resp. Surf. - 95/95

		CFD for λ					**CFD (Rounding) for λ**					**Asym CFD for λ (1.25)**					**Asym CFD for λ(1.5)**					**Asym CFD for λ (2.0)**					**Exact λ**				
n	CMC-OS	CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS		
				t=10	*t*=20	*t*=30			*t*=10	*t*=20	*t*=30			*t*=10	*t*=20	*t*=30			*t*=10	*t*=20	*t*=30			*t*=10	*t*=20	*t*=30			*t*=10	*t*=20	*t*=30
	1,730.89	1,731.00	1,718.40	1,718.35	1,715.08		1,727.64	1,715.48	1,715.98	1,712.50		1,726.22	1,713.81	1,713.87	1,711.01		1,725.87	1,713.44	1,713.54	1,710.69		1,722.17	1,709.89	1,710.10	1,707.53		1,720.44	1,708.39	1,708.63	1,707.53	
	31.83	29.34	25.59	24.53	26.92		28.30	24.70	23.32	25.62		27.88	24.20	22.91	24.88		27.33	23.80	22.56	24.54		26.01	22.71	21.26	22.91		20.19	17.83	16.59	22.91	
59	4.73	4.02	7.71	5.82	8.74		4.91	9.00	6.58	10.10		5.33	9.82	7.68	11.02		5.13	9.79	7.65	11.20		6.11	11.90	9.41	13.31		2.96	7.71	6.25	13.31	
		91.20	90.67	92.49	88.11		89.12	88.94	91.20	86.26		88.14	87.73	89.93	84.96		89.00	88.36	90.67	85.53		86.73	86.10	88.63	82.74		88.70	90.04	91.24	82.74	
		922.08	917.81	925.59	932.36		849.91	853.12	867.07	866.65		819.62	815.83	822.75	828.77		812.18	807.66	815.01	820.37		732.80	728.97	736.07	739.92		695.70	695.70	695.70	739.92	
	1,717.33	1,715.46	1,719.24	1,711.35	1,704.07	1,708.59	1,714.35	1,718.19	1,709.93	1,705.86	1,707.31	1,714.74	1,718.47	1,710.72	1,704.00	1,708.01	1,711.70	1,717.59	1,708.01	1,703.78	1,705.58	1,710.76	1,714.08	1,707.23	1,701.48	1,704.91	1,710.18	1,713.55	1,707.18	1,703.63	1,704.68
	21.67	20.77	21.57	18.63	19.18	20.32	20.50	21.28	18.23	19.90	19.73	20.44	21.12	18.30	19.11	19.93	19.79	20.70	17.62	18.99	18.88	19.44	19.77	17.31	18.14	18.46	16.02	15.48	13.95	15.75	15.92
93	4.80	5.42	4.04	5.05	13.70	8.72	5.84	4.34	5.74	12.11	9.67	5.58	3.95	5.32	13.82	9.08	7.16	4.12	6.89	13.98	10.85	7.54	5.37	7.11	16.33	11.10	4.68	2.05	3.74	9.90	8.54
		89.97	92.80	90.94	81.02	83.55	89.08	92.13	89.60	83.23	81.63	90.37	93.13	91.05	81.58	83.32	87.35	93.40	88.18	81.71	80.11	87.26	91.12	88.09	78.63	79.94	89.30	93.98	90.26	84.93	81.35
		837.72	832.74	825.52	718.80	835.01	807.86	807.37	783.16	771.50	792.16	818.33	814.20	806.42	716.76	815.34	736.49	793.03	725.78	710.05	733.81	711.24	708.45	702.18	642.07	710.64	695.70	695.70	695.70	695.70	695.70
	1,711.62	1,709.24	1,709.21	1,706.07	1,704.92	1,703.96	1,708.96	1,708.93	1,701.51	1,704.29	1,703.35	1,707.67	1,707.62	1,704.65	1,703.54	1,702.66	1,706.32	1,706.30	1,703.46	1,702.38	1,701.56	1,705.12	1,705.07	1,703.46	1,700.54	1,699.78	1,703.95	1,704.08	1,701.51	1,700.47	1,699.77
	17.90	16.46	16.29	14.56	15.11	15.85	16.41	16.24	11.66	14.93	15.63	16.17	15.94	14.26	14.68	15.37	15.93	15.67	14.06	14.35	15.01	15.63	15.39	14.06	13.85	14.43	13.42	13.00	11.66	12.09	12.99
124	4.69	5.18	4.71	4.95	6.83	8.92	5.31	4.87	5.61	7.22	9.38	6.03	5.65	5.91	7.74	10.02	6.80	6.58	6.69	8.62	11.04	7.81	7.59	6.69	10.48	13.03	6.32	5.06	5.61	8.23	10.51
		92.53	93.22	92.99	90.46	87.08	92.26	92.96	90.78	89.92	86.35	91.54	92.14	91.89	89.30	85.79	90.34	90.89	90.87	88.15	84.39	89.19	89.88	90.87	85.75	81.89	89.90	91.59	90.78	87.31	83.95
		860.04	853.97	853.80	860.11	856.97	851.29	845.28	695.70	837.51	834.45	811.20	804.93	805.50	810.90	808.90	769.36	764.20	765.10	769.32	767.73	732.09	726.20	765.10	703.12	701.14	695.70	695.70	695.70	695.70	695.70
	1,699.44	1,699.34	1,699.92	1,697.56	1,694.72	1,695.52	1,699.02	1,699.62	1,697.26	1,695.35	1,695.76	1,698.58	1,699.16	1,696.89	1,694.64	1,694.96	1,697.92	1,698.50	1,696.33	1,694.14	1,694.48	1,697.18	1,697.74	1,695.68	1,693.32	1,693.94	1,697.58	1,698.13	1,696.03	1,694.40	1,694.62
	10.08	9.81	9.60	8.31	8.02	8.39	9.77	9.56	8.28	8.11	8.43	9.75	9.55	8.27	8.07	8.32	9.69	9.49	8.23	8.01	8.26	9.63	9.42	8.17	7.91	8.20	8.71	8.44	7.32	7.19	7.50
311	5.38	4.88	4.01	4.40	7.92	7.37	5.23	4.29	4.56	6.67	6.93	5.76	4.69	5.09	8.08	8.15	6.54	5.28	5.70	9.26	8.94	7.53	6.30	6.64	11.02	9.79	5.27	3.85	3.99	6.39	6.94
		91.45	92.81	91.90	87.98	88.63	90.84	92.24	91.40	89.83	89.28	90.10	91.75	90.74	88.28	87.54	88.86	90.59	89.62	86.75	86.38	87.39	89.09	88.30	84.18	85.31	89.63	91.05	90.41	88.97	87.97
		782.30	782.59	783.91	717.19	749.90	766.44	767.96	766.79	754.68	764.04	744.95	745.54	746.14	712.38	717.64	712.54	713.58	713.80	682.32	689.46	675.72	676.69	677.17	633.69	657.75	695.70	695.70	695.70	695.70	695.70
	1,694.88	1,694.89	1,694.83	1,693.47	1,692.38	1,692.29	1,694.61	1,694.55	1,693.25	1,692.48	1,692.38	1,694.54	1,694.48	1,693.17	1,692.12	1,692.29	1,693.95	1,693.90	1,692.89	1,691.87	1,691.80	1,693.45	1,693.38	1,692.23	1,691.43	1,691.37	1,694.09	1,694.00	1,692.23	1,692.00	1,691.93
	7.10	7.11	6.96	6.17	6.00	5.98	7.09	6.94	6.15	6.01	5.99	7.11	6.95	6.17	6.00	5.98	7.08	6.91	6.15	5.99	5.99	7.05	6.89	6.11	5.96	5.96	6.49	6.29	6.11	5.46	5.49
548	5.51	5.57	4.95	5.23	6.68	6.96	5.92	5.32	5.41	6.51	6.74	6.13	5.49	5.49	7.36	6.96	7.07	6.62	6.16	8.08	8.29	8.11	7.76	7.56	9.14	9.20	5.57	4.79	7.56	6.02	6.44
		90.97	91.57	90.70	89.73	89.03	90.28	90.87	90.13	90.01	89.34	90.28	90.83	90.29	88.84	89.03	88.60	89.11	89.33	87.89	87.31	86.98	87.31	86.81	86.54	85.95	89.74	90.63	86.81	89.63	88.67
		747.93	750.25	750.36	726.97	725.16	729.53	731.79	733.23	734.49	732.66	724.86	727.30	727.38	706.75	725.16	686.76	688.49	705.93	687.24	687.37	653.67	654.30	655.47	653.26	653.17	695.70	695.70	655.47	695.70	695.70
	1,691.85	1,691.68	1,691.53	1,690.65	1,689.93	1,689.84	1,691.62	1,691.47	1,690.60	1,690.06	1,690.04	1,691.49	1,691.34	1,690.48	1,689.79	1,689.71	1,691.31	1,691.16	1,690.48	1,689.66	1,689.59	1,690.98	1,690.84	1,690.48	1,689.41	1,689.36	1,691.40	1,691.26	1,690.42	1,689.89	1,689.88
	5.10	5.15	5.07	4.43	4.30	4.28	5.15	5.07	4.42	4.31	4.30	5.16	5.07	4.43	4.31	4.29	5.15	5.07	4.43	4.31	4.30	5.14	5.05	4.43	4.31	4.30	4.80	4.68	4.06	3.98	3.99
1008	5.25	5.67	5.85	5.24	6.92	7.26	5.83	6.00	5.33	6.55	6.68	6.10	6.20	5.74	7.37	7.60	6.64	6.71	5.74	7.85	8.20	7.65	7.52	5.74	8.74	9.02	5.38	5.21	4.41	5.61	5.82
		90.08	89.93	89.95	88.62	87.67	89.73	89.69	89.76	89.33	88.72	89.52	89.41	89.29	88.17	87.26	88.77	88.71	89.29	87.40	86.43	87.00	87.35	89.29	86.25	85.28	89.52	89.85	90.13	89.56	89.12
		720.78	720.31	720.36	700.58	692.81	715.13	714.66	715.42	714.66	713.76	703.77	702.93	703.05	685.57	678.68	687.63	686.63	703.05	671.52	665.40	658.60	657.99	703.05	646.04	640.67	695.70	695.70	695.70	695.70	695.70
	1,689.42	1,689.20	1,689.13	1,688.32	1,688.26	1,688.27	1,689.17	1,689.10	1,688.29	1,688.23	1,688.35	1,689.13	1,689.06	1,688.18	1,688.13	1,688.22	1,688.98	1,688.91	1,688.13	1,688.08	1,688.17	1,688.79	1,688.73	1,687.96	1,687.92	1,688.02	1,689.07	1,689.00	1,688.21	1,688.15	1,688.28
	3.59	3.63	3.58	3.09	2.99	3.00	3.63	3.58	3.09	2.99	3.00	3.64	3.58	3.09	2.99	3.01	3.64	3.58	3.10	3.00	3.02	3.64	3.58	3.10	3.00	3.02	3.44	3.36	2.90	2.81	2.83
2004	5.34	5.96	5.69	6.34	5.83	6.11	6.12	5.81	6.48	5.95	5.76	6.18	5.90	6.93	6.40	6.35	6.78	6.56	7.26	6.58	6.46	7.64	7.46	7.84	7.38	7.24	5.35	5.25	5.67	4.96	5.10
		89.52	89.63	89.58	89.87	88.18	89.26	89.43	89.37	89.66	88.83	89.19	89.38	88.79	89.05	88.02	88.37	88.33	88.46	88.90	87.86	87.19	87.07	87.47	87.72	86.72	89.62	89.44	89.78	90.19	89.01
		711.77	712.94	711.31	711.31	694.55	708.07	709.23	706.91	706.91	707.70	703.45	704.22	691.12	691.85	688.34	684.94	684.72	683.77	684.60	680.69	661.30	661.35	660.08	660.73	657.24	695.70	695.70	695.70	695.70	695.70

Table C. 6: Complete Results for LOCA Resp. Surf. - 95/75

		CFD for λ					**CFD (Rounding) for λ**					**Asym CFD for λ (1.25)**					**Asym CFD for λ (1.5)**					**Asym CFD for λ (2.0)**					**Exact λ**				
n	CMC-OS	CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS			CMC	AV	LHS		
				t=5	t=10	t=15			t=5	t=10	t=15			t=5	t=10	t=15			t=5	t=10	t=15			t=5	t=10	t=15			t=5	t=10	t=15
	1,677.04	1,671.88	1,644.18	1,650.31			1,675.74	1,649.17	1,660.96			1,675.43	1,651.51	1,653.75			1,668.78	1,653.87	1,658.62			1,679.06	1,651.73	1,657.97			1,669.68	1,645.63	1,654.84		
	42.55	41.58	30.74	28.53			42.29	31.67	30.93			42.12	35.51	28.78			41.03	37.41	29.20			42.26	37.63	28.29			36.10	25.56	24.48		
11	4.35	5.33	10.59	4.93			4.64	8.01	2.50			4.89	9.73	4.15			6.31	9.65	3.01			4.28	11.33	2.73			4.83	5.14	1.82		
		88.12	86.30	90.06			89.61	89.65	95.30			90.30	89.84	93.10			87.11	90.20	95.54			92.45	88.63	95.61			90.62	92.82	95.37		
		293.10	274.21	257.36			311.07	303.97	321.47			309.65	317.90	277.93			278.69	331.97	306.69			326.55	319.21	302.47			282.85	282.85	282.85		
	1,646.59	1,645.51	1,638.63	1,629.50	1,626.95		1,643.70	1,641.61	1,631.62	1,628.83		1,644.46	1,638.30	1,629.31	1,626.79		1,643.68	1,641.50	1,631.33	1,628.52		1,642.73	1,642.42	1,630.99	1,628.32		1,640.49	1,641.18	1,630.69	1,628.11	
	23.50	23.76	20.87	15.83	16.23		23.59	21.35	16.20	16.78		23.82	21.02	15.96	16.36		23.82	21.82	16.45	16.97		23.89	22.77	16.34	16.93		22.54	17.50	14.84	15.80	
29	4.94	5.15	5.52	7.00	10.34		5.90	4.22	5.52	8.53		5.75	6.01	7.51	10.72		6.13	5.03	6.29	9.57		6.75	5.38	6.67	9.83		6.84	1.82	5.36	8.50	
		91.31	90.03	85.45	78.99		90.14	92.47	88.55	82.50		90.82	90.04	85.72	79.19		90.49	93.24	88.79	82.37		90.11	93.59	88.57	81.91		90.17	94.46	88.00	81.74	
		320.78	261.50	272.63	271.04		307.11	286.46	298.65	296.91		312.87	258.77	269.88	268.38		306.96	285.56	293.90	291.31		299.78	293.23	288.95	287.70		282.85	282.85	282.85	282.85	
	1,640.78	1,637.81	1,632.13	1,626.09	1,624.18	1,621.68	1,636.87	1,631.37	1,625.49	1,623.64	1,623.05	1,636.05	1,630.72	1,624.96	1,623.18	1,622.38	1,636.65	1,631.55	1,625.34	1,623.57	1,622.01	1,636.21	1,631.85	1,625.07	1,623.34	1,622.52	1,634.59	1,629.98	1,623.95	1,622.27	1,622.00
	19.74	19.67	17.15	13.50	14.07	21.12	19.62	17.08	13.42	13.93	22.45	19.66	17.21	13.46	13.91	21.80	19.81	17.62	13.61	14.09	21.44	19.93	18.38	13.63	14.08	21.83	19.21	15.21	12.74	13.28	20.92
40	4.25	5.76	6.85	7.35	10.74	27.19	6.25	7.34	7.96	11.46	27.03	6.82	8.06	8.69	12.18	27.15	6.65	7.85	8.72	11.83	27.22	6.96	8.55	9.11	12.36	27.10	7.73	6.10	9.34	12.63	27.00
		91.33	90.81	90.43	85.12	44.07	90.59	90.02	89.58	84.18	44.50	89.89	89.44	88.73	83.37	44.45	90.62	90.40	89.00	84.12	44.33	90.44	90.34	88.55	83.48	44.52	89.61	90.71	87.18	82.21	44.44
		311.49	306.18	315.47	315.16	281.34	303.08	297.92	306.96	306.65	308.19	295.82	290.92	299.36	299.38	294.78	301.13	299.93	304.41	304.85	286.53	297.29	303.20	299.90	300.61	295.19	282.85	282.85	282.85	282.85	282.85
	1,624.60	1,624.92	1,620.61	1,618.89	1,616.30	1,617.89	1,624.95	1,620.69	1,618.91	1,616.50	1,617.91	1,624.46	1,620.66	1,618.59	1,616.31	1,617.63	1,624.54	1,620.32	1,618.66	1,616.34	1,617.69	1,624.35	1,620.50	1,618.54	1,616.20	1,617.58	1,624.56	1,620.43	1,618.67	1,616.29	1,617.71
	10.53	10.69	8.96	7.07	6.61	6.78	10.69	8.96	7.07	6.62	6.79	10.73	9.07	7.11	6.69	6.80	10.78	9.11	7.17	6.74	6.85	10.82	9.23	7.19	6.77	6.86	10.57	8.37	6.79	6.46	6.57
135	4.73	4.86	6.44	4.33	7.82	5.19	4.85	6.38	4.33	7.46	5.19	5.48	6.68	4.86	8.08	5.70	5.51	7.25	4.99	8.34	5.77	5.65	7.28	5.35	8.81	6.22	5.06	5.51	4.27	7.85	5.37
		89.43	90.07	89.36	87.69	87.49	89.47	90.15	89.38	88.37	87.53	88.72	90.20	88.60	87.80	86.87	88.98	89.64	88.76	87.65	87.02	88.92	90.12	88.36	87.19	86.55	89.85	90.57	88.78	87.72	86.84
		288.63	286.54	289.87	284.33	289.55	289.10	288.07	290.34	290.02	290.02	281.14	287.44	282.09	284.68	281.97	282.48	280.70	283.67	285.15	283.58	279.38	284.25	280.34	281.16	280.31	282.85	282.85	282.85	282.85	282.85
	1,619.86	1,620.15	1,618.17	1,614.96	1,614.87	1,614.35	1,619.89	1,617.95	1,614.99	1,614.78	1,614.29	1,620.06	1,618.09	1,614.92	1,614.82	1,614.31	1,619.74	1,617.81	1,614.90	1,614.78	1,614.28	1,619.67	1,617.75	1,614.86	1,614.72	1,614.13	1,619.70	1,617.84	1,614.90	1,614.71	1,614.20
	7.81	7.94	6.70	5.07	4.84	4.96	7.93	6.69	5.07	4.83	4.95	7.97	6.76	5.11	4.88	5.01	7.99	6.79	5.15	4.91	5.04	8.02	6.85	5.19	4.94	5.07	7.83	6.34	4.93	4.71	4.85
246	5.12	4.84	5.03	5.71	5.08	6.75	5.29	5.29	5.68	5.29	6.87	5.02	5.26	5.98	5.50	6.96	5.54	5.78	6.14	5.72	7.26	5.80	6.08	6.52	5.93	7.67	5.45	4.43	5.55	5.24	6.83
		90.52	90.38	89.93	89.67	88.34	89.56	89.79	90.02	89.33	88.12	90.35	90.33	89.91	89.55	88.31	89.52	89.73	89.73	89.50	88.13	89.37	89.62	89.63	89.29	87.66	89.77	90.36	89.91	89.38	87.94
		292.76	291.47	286.03	289.78	289.41	286.98	285.72	286.99	286.37	287.41	290.68	289.38	284.68	287.75	288.14	283.68	282.12	283.68	286.10	287.00	282.21	280.54	282.18	283.86	280.94	282.85	282.85	282.85	282.85	282.85
	1,615.93	1,616.69	1,614.55	1,612.82	1,612.50	1,612.46	1,616.69	1,614.55	1,612.83	1,612.50	1,612.49	1,616.58	1,614.46	1,612.76	1,612.43	1,612.46	1,616.49	1,614.46	1,612.76	1,612.44	1,612.40	1,616.46	1,614.41	1,612.73	1,612.40	1,612.37	1,616.58	1,614.47	1,612.76	1,612.43	1,612.43
	5.61	5.68	4.77	3.73	3.49	3.45	5.68	4.77	3.73	3.49	3.46	5.71	4.81	3.75	3.52	3.48	5.72	4.83	3.77	3.54	3.50	5.75	4.86	3.79	3.56	3.52	5.66	4.60	3.65	3.44	3.40
459	5.65	4.45	5.68	5.99	5.93	6.07	4.45	5.67	5.97	5.90	5.90	4.62	6.12	6.42	6.33	6.23	4.79	6.08	6.52	6.42	6.43	4.95	6.41	6.91	6.46	6.74	4.58	5.37	6.12	5.77	5.82
		90.32	90.45	89.94	89.95	88.59	90.31	90.46	89.97	89.98	88.83	90.03	90.01	89.39	89.45	88.62	89.76	90.22	89.38	89.45	88.46	89.59	90.01	89.01	89.45	88.15	90.28	90.46	89.32	89.84	88.88
		286.14	285.77	286.51	286.71	284.63	286.08	286.03	286.76	286.97	286.43	282.85	282.49	283.17	283.42	284.66	280.22	282.69	283.43	283.68	281.78	279.08	280.94	281.66	281.61	279.85	282.85	282.85	282.85	282.85	282.85
	1,613.81	1,613.83	1,612.71	1,611.34	1,611.13	1,610.99	1,613.84	1,612.72	1,611.35	1,611.15	1,611.00	1,613.80	1,612.68	1,611.32	1,611.11	1,610.97	1,613.76	1,612.65	1,611.30	1,611.09	1,610.95	1,613.71	1,612.61	1,611.26	1,611.06	1,610.92	1,613.82	1,612.70	1,611.33	1,611.13	1,610.99
	4.05	4.09	3.46	2.65	2.49	2.49	4.09	3.46	2.65	2.49	2.49	4.11	3.48	2.67	2.51	2.50	4.12	3.50	2.68	2.52	2.52	4.13	3.51	2.69	2.53	2.53	4.09	3.38	2.60	2.46	2.46
886	5.03	4.92	5.44	5.05	5.23	5.77	4.88	5.41	5.04	5.17	5.75	5.13	5.48	5.37	5.32	6.02	5.27	5.67	5.53	5.57	6.19	5.40	6.01	5.86	5.85	6.48	5.21	4.80	5.05	5.05	5.64
		90.19	89.27	90.24	89.38	89.24	90.24	89.33	90.28	89.55	89.27	89.94	89.38	90.06	89.33	88.95	89.81	89.07	89.86	89.17	88.86	89.72	88.77	89.60	88.93	88.55	90.10	89.89	90.22	89.57	89.38
		283.52	283.31	284.01	282.94	283.66	284.01	283.80	284.34	284.07	283.99	282.07	281.70	282.33	281.22	282.14	280.59	280.22	280.95	279.74	280.77	278.44	277.93	278.71	277.63	278.58	282.85	282.85	282.85	282.85	282.85