NASA/TM–20210019495

Assessing Next-Gen Spacesuit Reliability: A Probabilistic Analysis Case Study

James E. Warner
Langley Research Center, Hampton, Virginia

Patrick E. Leser
Langley Research Center, Hampton, Virginia

William P. Leser
Langley Research Center, Hampton, Virginia

D. Austin Cole
Virginia Tech, Blacksburg, Virginia

Geoffrey F. Bomarito
Langley Research Center, Hampton, Virginia

August 2021

NASA STI Program Report Series

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.
- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.
- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov
- Help desk contact information:

https://www.sti.nasa.gov/sti-contact-form/
and select the "General" help request type.

NASA/TM–20210019495

Assessing Next-Gen Spacesuit Reliability: A Probabilistic Analysis Case Study

James E. Warner
Langley Research Center, Hampton, Virginia

Patrick E. Leser
Langley Research Center, Hampton, Virginia

William P. Leser
Langley Research Center, Hampton, Virginia

D. Austin Cole
Virginia Tech, Blacksburg, Virginia

Geoffrey F. Bomarito
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

August 2021

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

Abstract

Motivated by a current effort to certify the Exploration Extravehicular Mobility Unit (xEMU) spacesuit under NASA's Artemis Program, this report presents a probabilistic analysis case study focused on estimating the reliability of a spacesuit design for a potential fall scenario. Considering uncertainty in impact velocity and material parameters that govern damage in a composite spacesuit, the probability of no impact failure (PnIF) was estimated through the application of several uncertainty quantification (UQ) concepts. In particular, the focus of the case study was on providing practical solutions to common challenges of applying UQ in practice: estimating uncertainty in model parameters and alleviating the prohibitive computational burden associated with UQ when expensive, high-fidelity models are needed. Material parameter uncertainty was estimated through a combination of sensitivity analysis and Bayesian model calibration. Despite the computational model for suit impact requiring nearly one day of serial execution time, the PnIF was estimated in a tractable amount of time through a multifidelity importance sampling (MFIS) approach that leveraged surrogate modeling and high performance computing. It is shown that quantified material parameter uncertainty results in nearly 290lbf of variation in predicted contact force in the suit ($> 10\%$ variability with respect to the uncertainty-free nominal prediction). Furthermore, the MFIS prediction of PnIF agrees well with a brute force Monte Carlo simulation estimate but is significantly more efficient.

1 Introduction

Under NASA's Artemis program, the Exploration Extravehicular Mobility Unit (xEMU) spacesuit is being designed to ensure the safety of astronauts while they explore the lunar surface. The xEMU design and certification effort must strike a delicate balance between producing a spacesuit that is robust enough to safely withstand potential fall events while still satisfying stringent mass and mobility requirements. A traditional design approach that considers worst case loading and applies conservative factors of safety (FoS) to account for uncertainties in the analysis was unlikely to meet the narrow design margins. Thus, the xEMU design requirement was modified to include a probability of no impact failure (PnIF) threshold that must be verified through probabilistic analysis.

As part of a broader one year effort to help integrate modern uncertainty quantification (UQ) methodology into engineering practice at NASA, the certification of the xEMU spacesuit was selected as the primary case study. The project, titled *Uncertainty Quantification for Structures and Dynamics* under NASA's Engineering Research & Analysis (R&A) Program and henceforth referred to as the *UQ R&A Project*, aimed to provide a foundation for a long-term transition away from FoS and towards reliability-based design. The xEMU reliability case study was intended to be performed in parallel alongside the ongoing certification effort to demonstrate the advantages of UQ methodology for design. By taking a reliability-based approach to design in general, NASA can enable substantially reduced mass/costs, higher performance through narrower margins, rigorous assessments of mission risk, and ultimately, safer and more reliable flight systems. This report provides the details of a probabilistic analysis demonstration conducted under the UQ R&A Project to estimate spacesuit reliability (PnIF).

While the use of FoS-based design is still prevalent at NASA, there have been a variety of studies both highlighting the limitations of FoS and motivating the use of UQ and reliability-based design at the Agency. Nearly two decades prior to the study herein, a broad survey of uncertainty-based design methods was conducted in order to provide recommendations for research activities at NASA Langley Research Center (LaRC) [1]. While focusing on the specific disciplines of aerodynamics, controls, structures, and systems analysis, many general, application-independent benefits of uncertainty-based design were highlighted, including increased confidence in analysis tools and reduced design cycle time, cost, and risk. A later report aimed at reviewing current policies on FoS across the Agency also recommended further study on probabilistic approaches for design, noting that while traditional FoS has been useful for decades, establishing a rigorous connection between design options and reliability could be a major improvement [2]. Most recently, an approach for integrating uncertainty management into the design cycle was presented to address the more general question of how to develop confidence in design given constrained resources [3]. Here, the benefits of characterizing and quantifying uncertainties earlier in the design process were highlighted and a discussion of how UQ methodology should scale in complexity with design maturity was provided.

Several studies have investigated the relationship between the prescribed FoS and resulting reliability under various assumptions on uncertainties in load and material strength [4–6]. In particular, [6] illustrated how structural reliability varies significantly for different materials for a fixed FoS, which could lead to unacceptable rates of failure for some materials but significant overdesign for others. The potential for over conservatism when using traditional FoS was further demonstrated through a probabilistic design case study of the composite crew module (CCM) sponsored by the NASA Engineering and Safety Center (NESC) [7,8]. The study showed that a FoS of 1.4 resulted in an extremely small failure probability (1 in ∼10M) and that significant mass savings (∼12.5%) were possible using probabilistic methods. However, several practical challenges with applying UQ were noted, including the estimation of input parameter uncertainties and the associated computational expense. A more recent NESC peer review demonstrated a successful application of UQ in the context of loads and dynamics for the Space Launch System (SLS), focussing on characterizing model-form uncertainty [9]. Finally, there also exists several high-level NASA guidance documents related to UQ [10–13], including NASA-STD-7009A, Standard for Models and Simulations [10], and accompanying handbook [11] that includes sections on UQ and verification and validation. Despite this, the existence of documented cases of the successful application of UQ to engineering and design problems at NASA in reports and other publications appears relatively limited [14–16].

This report presents a probabilistic analysis case study aimed at estimating the PnIF of a next-gen spacesuit design, providing further documentation of, and motivation for, reliability-based design at NASA. In lieu of mature computational models and test data for the xEMU at the time of the UQ R&A Project, the analysis was demonstrated using existing finite element method (FEM) models and impact test data from the previous generation of spacesuits (the *Z-2*). Several common UQ concepts are described and demonstrated, including the application of sensitivity analysis and Bayesian model calibration to estimate material parameter uncertainty and reliability analysis to estimate the PnIF given estimated/assumed input uncertainties. Due to the computational expense of the full Z-2 assembly model, advanced UQ methods and high performance computing (HPC) utilization were leveraged to make the probabilistic analysis tractable. As such, potential solutions for addressing previously noted challenges of applying UQ (i.e., parameter uncertainty estimation, computational expense) [8] are highlighted.

Note that several simplifications and assumptions were introduced to the case study in order to facilitate the PnIF analysis under the time constraints of the UQ R&A Project. Most importantly, uncertainty was only considered in the impact velocity and the material parameters governing composite damage due to impact, ignoring the myriad of additional sources of variability/ignorance that would affect spacesuit impact verification in practice (manufacturing defects, impactor shape/size, lunar regolith variability, etc.). The focus, instead, was on overcoming the common, aforementioned challenges of applying UQ with expensive, high-fidelity models. The results provided are therefore strictly for demonstration purposes and do not directly reflect the true reliability of the Z-2 spacesuit.

The report is laid out as follows. Section 2 details the relevant background for the study, including a brief overview of the UQ R&A Project and a description of a general UQ workflow for estimating reliability. Section 3 describes the components of the UQ workflow in more detail and introduces the specific methods used to implement them in this study. Section 4 provides information on the case study, including a description of the computational models and test data for the Z-2 spacesuit and the simplifications and assumptions made to facilitate the analysis. Finally, Section 5 presents the results of the reliability analysis and Section 6 provides concluding remarks.

2 Background

This section introduces several important components to a practical UQ workflow, providing some background before discussing the specific methods that were leveraged in the following section. For more context about the UQ R&A Project itself that the case study was performed under, see Section 7.1 in the Appendix.

2.1 A General UQ Workflow

UQ is the science of identifying, quantifying, and reducing uncertainties associated with models, experiments, and predicted outcomes or quantities of interest (QoIs) [17]. These uncertainties may arise from a variety of sources, including experimental uncertainties (due to limited or incomplete data or limited accuracy of sensors), model discrepancy/inadequacy (due to approximate, imprecise, or simplified physics), input parameter uncertainty (the focus of model calibration), and numerical errors (due to roundoff or discretization, the focus of numerical analysis). Broadly speaking, these

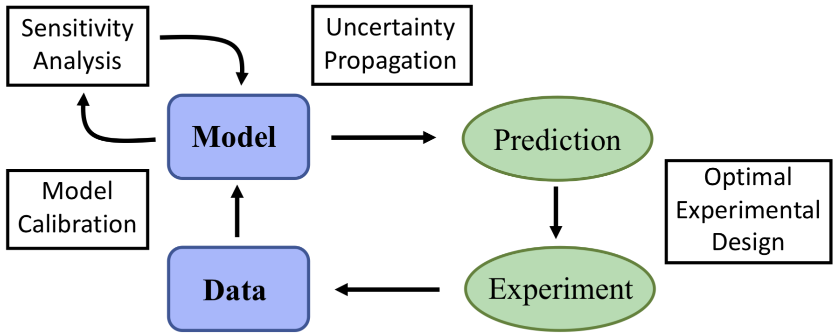


Figure 1. General UQ workflow

uncertainties may be classified as either aleatory or epistemic. Aleatory uncertainty is randomness that is inherent to the problem or system and is irreducible. Epistemic uncertainty reflects a lack of knowledge and can be reduced with additional experiment, modeling effort, etc. While distinguishing between the types of uncertainty can be useful for understanding a problem, this work assumes that both can be adequately described in a similar fashion using probability theory from a Bayesian perspective [13, 18]. Here, probability is interpreted as a degree of belief in an event or value and the ability to update initial uncertainty estimates based on observed data is fundamental. Note this is in contrast with alternative approaches that argue for separate treatment of epistemic and aleatory uncertainty [19].

The UQ R&A Project identified a general workflow with four primary components for implementing UQ for reliability analyses. Figure 1 shows a schematic of these four UQ components and where they fall within the relationship between a computational model that produces predictions and a physical experiment that generates measurement data. The purpose of each of these four concepts are provided below:

- *Uncertainty propagation* for producing probabilistic predictions of QoIs and calculating system reliability when quantified model input uncertainties are available
- *Model calibration* as a mechanism to explicitly characterize model uncertainties using experimental data in order to improve upon any initial assumptions and prior knowledge
- *Sensitivity analysis* to identify the most influential system parameters to properly focus effort/resources in a probabilistic analysis
- *Optimal experimental design* to determine the most effective experiments to perform in order to reduce uncertainty in model calibration and/or to produce the most precise predictions with uncertainty propagation

Since the ultimate goal is to make a probabilistic prediction with a model (e.g., QoI with error bars or an associated reliability value) using an uncertainty propagation method, this task tends to get the most focus and is often thought of synonymously with UQ itself. While uncertainty propagation can be performed after assigning any arbitrary uncertainties to model inputs, model calibration is a necessary initial step to quantify and reduce model uncertainties in a principled manner based on measured data while taking into account experimental uncertainty. Sensitivity analysis is typically performed in the model development stage to identify the parameters whose uncertainty has the greatest impact on the QoI so that those parameters can be the focus of model calibration and optimal experimental design efforts. Experimental design is the process of using a predictive model to inform experiments and allows testing resources to be focused on the measurements that will benefit the UQ analysis the most.

3 Approach

The UQ workflow introduced in the previous section merely outlines the general components of a probabilistic analysis, allowing for a plethora of methods and tools to be leveraged to actually implement the process. This section briefly describes the specific approach that was applied to the xEMU certification case study while introducing some relevant mathematical notation. References are supplied throughout in lieu of a detailed discussion of all concepts.

3.1 Problem Setup

In the spirit of Figure 1, consider a computational model, $\mathcal{M}$, that produces QoI predictions and an experiment, $\mathcal{E}$, that produces measurement data related to an engineering system of interest. Let $Y \in \mathbb{R}^1$ represent a scalar QoI that is dependent on both deterministic inputs, $\mathbf{x} \in \mathbb{R}^{d_x}$, and random parameters, $\boldsymbol{\Theta} \in \mathbb{R}^{d_\Theta}$ through the model, i.e.,

$$Y = \mathcal{M}(\mathbf{x}; \boldsymbol{\Theta}). \tag{1}$$

The QoI, Y, is thus a random variable because of its dependence on $\boldsymbol{\Theta}$. Let $p(y)$ and $p(\boldsymbol{\theta})$ denote the respective probability density functions (PDFs).

Additionally, let $\mathcal{D} \in \mathbb{R}^m$ represent a set of m measurements obtained through the experiment whose values are assumed to depend on $\mathbf{x}$ and $\boldsymbol{\Theta}$, i.e.,

$$\mathcal{D} = \mathcal{E}(\hat{\mathbf{x}}; \boldsymbol{\Theta}, \boldsymbol{\epsilon}). \tag{2}$$

Here, $\hat{\mathbf{x}}$ are prescribed input values defining the experimental setup and $\boldsymbol{\epsilon}$ denotes random variables associated with the measurement process itself (e.g., measurement noise/error). Thus, the measurement data, $\mathcal{D}$, are also random variables, indicating that repeated measurements for the same experimental setup will generally result in different values.

In practice, the experimental configuration in Equation (2) commonly represents a simplified system versus that of the computational model in Equation (1). For example, a component-scale experiment may be used to characterize material properties used in a simulation of a full scale assembly model. In this case, a simpler computational model may be introduced to simulate the experiment directly:

$$Y_{\mathcal{E}} = \mathcal{M}_{\mathcal{E}}(\mathbf{x}; \mathbf{\Theta}). \tag{3}$$

Since the auxiliary measured quantity, $Y_{\mathcal{E}}$, is assumed to have dependence on $\mathbf{x}$ and $\mathbf{\Theta}$, the dataset $\mathcal{D} \equiv \{\hat{Y}_{\mathcal{E}}(\hat{\mathbf{x}}^{(k)})\}_{k=1}^{m}$ can used for the purposes of calibration or validation. In the context of this case study, $\mathcal{D}$ will be used to calibrate the parameters with quantified uncertainty which will then be transferred to the full scale model $\mathcal{M}$.

3.2 Uncertainty Propagation

Uncertainty propagation methods seek to estimate $p(y)$ or other QoI statistics using the model $\mathcal{M}$ and estimated/assumed input parameter uncertainty, $p(\boldsymbol{\theta})$. As a baseline approach, this case study considers one of the most well known UQ methods, Monte Carlo simulation (MCS) [20], which can be represented succinctly as follows

$$\boldsymbol{\theta}^{(i)} \sim p(\boldsymbol{\theta}) \tag{4}$$

$$y^{(i)} = \mathcal{M}(\mathbf{x}; \boldsymbol{\theta}^{(i)}) \tag{5}$$

$$i = 1, ..., N_{MC}, \tag{6}$$

where the first equation signifies that $\boldsymbol{\theta}^{(i)}$ is a deterministic realization of the random variable, $\mathbf{\Theta}$, randomly sampled from the PDF, $p(\boldsymbol{\theta})$. As shown, MCS involves evaluating the computational model in Equation (1) N_{MC} times to produce a collection of samples, $\{y^{(i)}\}_{i=1}^{N_{MC}}$, that can then be used to estimate statistics of Y.

Since the case study described in this report is a reliability analysis, the focus is on using uncertainty propagation to estimate the probability of failure, P_f. If failure occurs when Y exceeds a critical value, y^c, then

$$P_f = P(Y \geq y^c). \tag{7}$$

Using MCS, the probability of failure can be estimated as follows

$$P_f \approx P_f^{MC} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \mathbf{1}(y^{(i)} \geq y^c) \tag{8}$$

where $\mathbf{1}$(condition) is the indicator function, equal to 1 if the condition is true and 0 otherwise. Simply put, Equation (8) estimates P_f as the percentage of the N_{MC} model evaluations performed with MCS that resulted in Y exceeding the critical value. The application of this approach is simple but it is generally computationally demanding. Since failure is a rare event in most practical problems, it can take a huge number of samples, N_{MC}, for an accurate resolution of a small P_f value. Thus, if $\mathcal{M}$ is expensive to evaluate, then MCS can quickly become infeasible. Section 3.6 describes the approaches used in this case study to make MCS-based estimates of P_f more computationally efficient while retaining accuracy.

3.3 Model Calibration

One common challenge associated with applying UQ in practice is determining an appropriate PDF, $p(\boldsymbol{\theta})$, describing uncertainty in the model parameters, $\boldsymbol{\Theta}$ [8]. While deterministic model calibration (or model updating) approaches use optimization to select single values of $\boldsymbol{\theta}$ that result in good agreement between model and experiment, probabilistic model calibration methods explicitly characterize parameter uncertainties using measurement data.

This case study considers a well-established approach of Bayesian model calibration [21] in conjunction with Markov Chain Monte Carlo (MCMC) [22] numerical sampling to quantify model parameter uncertainty. The conditional PDF of model parameters given the observed measurement data, $p(\boldsymbol{\theta}|\mathcal{D})$, is formulated using Bayes' Theorem. This PDF combines any available prior knowledge about $\boldsymbol{\theta}$ with inferred likelihood of the parameters based on predictions from the model. Intuitively, the PDF $p(\boldsymbol{\theta}|\mathcal{D})$ takes larger values for parameters that result in a close agreement between model and experiment, and smaller values for parameters that result in a larger mismatch. Since this probability distribution is often analytically intractable (not available in closed form), sampling algorithms, such as MCMC, can be leveraged to draw samples $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^{N}$ where $\boldsymbol{\theta}^{(i)} \sim p(\boldsymbol{\theta}|\mathcal{D})$ to represent model parameter uncertainty. Note that each sample drawn with MCMC, $\boldsymbol{\theta}^{(i)}$, requires an execution of the model, $\mathcal{M}_{\mathcal{E}}$, for a total of N model evaluations.

3.4 Sensitivity Analysis

When the number of uncertain parameters, d_{Θ}, is large, it can be challenging to apply many existing UQ methods. Sensitivity analysis can be used to identify a small subset of parameters whose uncertainty has the greatest influence on a quantity of interest. Then, effort and resources can be strategically used in a probabilistic analysis to analyze that important subset while the remaining parameters can be treated as deterministic with relatively little impact.

In this study, a density-based global sensitivity measure introduced by [23,24] and implemented in the SALib Python package [25] was used to quantify the sensitivity of the QoI to the model parameters. In contrast to moment-based measures [26], the density-based measure considers the influence of each parameter on the full QoI probability distribution rather than a specific statistical moment (e.g., variance). Another advantage of this measure is that its derivation does not depend on any assumptions of parameter independence, making it particularly useful for complex computational models with many correlated parameters.

3.5 Optimal Experimental Design

The experimental configuration used to acquire the measurement data, $\mathcal{D}$, can have a significant impact on the resulting quantified model parameter uncertainty, $p(\boldsymbol{\theta})$, obtained through model calibration. Optimal experimental design methods can determine the most effective experimental setup, $\hat{\mathbf{x}}$, from Equation (2), when testing is expensive and limited. For example, the number of sensors and sensor location can be selected such that the uncertainty reduction in the model parameters is maximized.

Given the short duration of the UQ R&A Project (see Section 7.1 in the Appendix), experimental testing was not feasible and thus optimal experimental design methods were not demonstrated in this case study. Instead, existing measurement data from tests conducted prior to the start of the project were used.

3.6 Computationally Efficient Approaches

A well-known challenge with applying UQ methods in practice is a potentially intractable computational cost. Intuitively, when uncertainty is considered in model parameters, the computational model must be evaluated a large number of times to assess the range of possible outcomes due to that uncertainty. This is in contrast with deterministic methods where the model is evaluated just once (or significantly fewer times) for fixed parameter settings. For the four UQ concepts described above, the number of samples and model evaluations required to implement them, N, can be considerably large ($O(10^4)$ or more), and even more so when the estimation of small failure probabilities (Equation (8)) is required. Thus, if the computational model is expensive to evaluate (e.g., a FEM simulation), the standard application of UQ can be intractable. The following sub-sections introduce approaches used in this case study to make UQ more computationally efficient.

3.6.1 Surrogate Modeling

Surrogate modeling[1] is a common approach for speeding up UQ that involves replacing the original, *high-fidelity*, model, $\mathcal{M}$, with an approximate model, $\tilde{\mathcal{M}}$, that is faster to evaluate. To apply surrogate modeling, a grid of values for the model parameters, $\{\bar{\boldsymbol{\theta}}^{(i)}\}_{j=1}^{S}$, is first generated either at random (e.g., $\bar{\boldsymbol{\theta}}^{(i)} \sim p(\theta)$), using a space-filling approach [27], or on a uniform grid within the feasible bounds of the parameters. Then, the original computational model in Equation (1) is evaluated for each parameter setting to produce the dataset of input-output pairs

$$\mathcal{S} = \{(\bar{\boldsymbol{\theta}}^{(i)}, \bar{y}^{(i)}\}_{j=1}^{S}, \tag{9}$$

where $\bar{y}^{(i)} = \mathcal{M}(\mathbf{x}; \bar{\boldsymbol{\theta}}^{(i)})$.[2] Using $\mathcal{S}$ as *training data*, one of a variety of regression/supervised machine learning and interpolation techniques can be used to construct the surrogate model such that $\tilde{\mathcal{M}}(x; \boldsymbol{\theta}^*) \approx \mathcal{M}(x; \boldsymbol{\theta}^*)$ for new parameter values, $\boldsymbol{\theta}^*$. It is considered good practice to generate an additional *test dataset*, $\mathcal{S}^*$, with input-output pairs that are unique from those in $\mathcal{S}$ in order to validate the effectiveness of the surrogate model.

In this work, Gaussian Process (GP) regression [28] is used to construct surrogate models from training data to speed up the demonstrated UQ analysis. GPs have successfully improved the efficiency of UQ for a wide variety of applications [29–31].

[1] Surrogate models are also referred to as meta models, reduced-order models, kriging, or response surfaces depending on the field and approach used to construct them.

[2] Typically, a surrogate model is constructed for a specific setting of the deterministic inputs, $\mathbf{x}$, and so multiple surrogate models must be trained if additional settings are to be analyzed.

An advantage of the method is its ability to produce probabilistic predictions which can facilitate adaptively selected training data for improved accuracy with fewer evaluations of the original computational model [32–34].

3.6.2 Multi-Fidelity Importance Sampling

While the surrogate modeling approach described in the previous section can alleviate the computational burden for UQ, it often comes at the expense of accuracy. Since $\tilde{\mathcal{M}}$ is an approximation of the original computational model $\mathcal{M}$, the resulting estimates for QoI statistics will generally have an associated error or *bias*, the magnitude of which is difficult to estimate a priori. Furthermore, even relatively accurate surrogate models with small global error can result in significant errors in probability of failure estimates if they are not particularly accurate near the failure threshold. Rather than substitute the high-fidelity model for a surrogate model, emerging multi-fidelity UQ methods [35–38] aim to combine predictions for both models (or collections of models of varying fidelity) for statistical estimators that are both efficient *and* unbiased. This case study, in particular, leveraged multi-fidelity importance sampling (MFIS) [39] to generate efficient and unbiased estimates of probability of failure.

In contrast with the MCS approach shown in Equation (8), an importance sampling estimator for probability of failure, P_f^{IS}, has the form

$$P_f \approx P_f^{IS} = \frac{1}{N_{IS}} \sum_{i=1}^{N_{IS}} \mathbf{1}(y^{(i)} \geq y^c) w(\boldsymbol{\theta}^{(i)}) \tag{10}$$

where $\boldsymbol{\theta}^{(i)}$ in this case is sampled from a new PDF, $q(\boldsymbol{\theta})$, called the *bias* distribution, and $w(\boldsymbol{\theta}) = \frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}$. The purpose of the biasing distribution is to give higher probability density to the values $\boldsymbol{\theta}$ that result in failure ($y(\boldsymbol{\theta}) \geq y^c$) so that fewer overall samples ($N_{IS} << N_{MC}$) are required to generate an accurate estimate of P_f. Since $q \neq p$ is being used as the parameter PDF, the weighting factor w acts as a correction to ensure that P_f^{IS} is an unbiased estimator.

The primary challenge with applying importance sampling is constructing an effective biasing distribution, especially in high dimensions. Doing so requires knowledge of the areas of the model parameter space that result in failure (the *failure region*) according to the computational model prediction and failure threshold. Multi-fidelity importance sampling (MFIS) overcomes this challenge by leveraging the surrogate model, $\tilde{\mathcal{M}}$, to efficiently probe the failure region and develop a PDF, $q(\boldsymbol{\theta})$, that assigns higher probability to this region.

First, an MCS is performed analogous to Equation (6) but using the surrogate model $\tilde{\mathcal{M}}$ and for a large number of samples $M >> N$. Then, a set of parameter samples, $\{\boldsymbol{\theta}_f^{(i)}\}$, that resulted in failure ($\tilde{y}^{(i)} = \tilde{\mathcal{M}}(\mathbf{x}; \boldsymbol{\theta}_f^{(i)}) \geq y^c$) are identified. A PDF can then be fit to this sample data, e.g., using a Gaussian mixture model. Samples are then drawn from the resulting bias distribution, $q(\boldsymbol{\theta})$, and evaluated using the high-fidelity model to build the probability of failure estimator according to Equation (10). In this way, MFIS provides an efficient mechanism for constructing an effective bias distribution using the surrogate model and also yields unbiased failure probability estimators by leveraging the original model to calculate P_f^{IS}. For more details of the MFIS approach, see [39].

3.6.3 High Performance Computing

As mentioned previously, UQ analyses require the repeated evaluation of potentially expensive, high-fidelity computational models. Even when more efficient surrogate modeling approaches described in Section 3.6.1 are used, the original computational model must be executed a large number of times to produce the required training dataset (Equation (9)). Here, the accuracy of the surrogate model and subsequent UQ analysis depends critically on the size, S, of this dataset, or number of model evaluations that can be completed up front.

The ability to access and properly leverage high performance computing (HPC) resources can substantially reduce the turnaround time for UQ analyses by scheduling and executing the required model evaluations in a parallel fashion. This case study relied on midrange HPC resources at NASA LaRC and custom software for automating batches of simulations and managing the resulting input/output data. The FEA model of the spacesuit considered in the case study took over 40 hours to execute in serial (on 1 processor) and 450 evaluations of the model were performed for the UQ analysis. HPC resources and parallel computing were used to perform the required 2+ years of computation time in just 2 weeks by running 15 simulations at a time in parallel, with each simulation running on 10 processors.

3.7 Discussion

It is important to note that UQ is a rapidly evolving research area and the multifidelity UQ approach leveraged in this study is just one of a large variety of potential options that could have comparable or potentially better performance. Even limiting consideration to within NASA, there are several active areas of research for UQ and reliability analysis, including polynomial chaos expansion (PCE)-based approaches [40], methods based on scenario optimization theory [41], and non-parametric variation and hybrid parametric variation (NPV/HPV) methods [9].

Out of the many advanced, modern UQ approaches, multi-fidelity methods were chosen in this project for several reasons. For starters, surrogate modeling is already a technique that has been leveraged at NASA for probabilistic analyses in several engineering projects. Therefore, adopting a multi-fidelity UQ method, which keeps the high-fidelity model rather than replacing it, has a smaller learning curve relative to many other modern UQ methods for analysts familiar with surrogate modeling. Furthermore, another advantage of most multi-fidelity UQ methods is that they perform effectively regardless of the number of random input parameters considered in a problem. This is in contrast with many modern UQ methods that seek to directly model the random input parameters, becoming exponentially more difficult as the dimension increases (i.e. the curse of dimensionality).

4 Case Study

4.1 Overview

Estimating the reliability of a spacesuit design under potential impact load scenarios represents both a challenging and relevant case study to demonstrate the utility of probabilistic methods. The design and certification of the xEMU spacesuit is important for NASA's Artemis program and an explicit probabilistic requirement was postulated in order to develop a suit that was adequately robust, lightweight, and mobile. At the time of the UQ R&A Project, this requirement read as follows: *Space suit design shall provide a Probability of no Impact Failure (PnIF) of 0.997 (TBR) or better while supporting 40 hours of two-person EVAs in the lunar surface environment.* Since mature computational models and test data were not available for xEMU, a previous generation of spacesuit, the Z-2, was substituted for demonstration. The Z-2 spacesuit was made of a S-2 glass / PW IM10 hybrid composite laminate and the primary structural components in the assembly were the hard upper torso (HUT), brief, glass visor, and primary life support system (PLSS).

While the PnIF requirement must be satisfied for a range of potential astronaut fall events, this case study focused specifically on a scenario resulting in impact to the HUT portion of the suit. Uncertainties from two primary sources were considered: 1) the material properties governing impact damage initation and progression and 2) the impact load. Failure was assumed to occur when the maximum contact force in the HUT due to an impactor exceeded a predefined threshold. The general UQ workflow described previously was applied in order to:

1. Identify the subset of material parameters that had the most influence on the contact force using sensitivity analysis (Section 3.4)

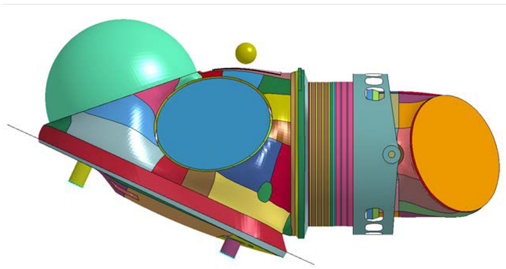
2. Quantify the uncertainty in the most influential material parameters using model calibration (Section 3.3)

3. Estimate the PnIF in the HUT by propagating uncertainty (3.2) in material properties and impact loading through the assembly impact model

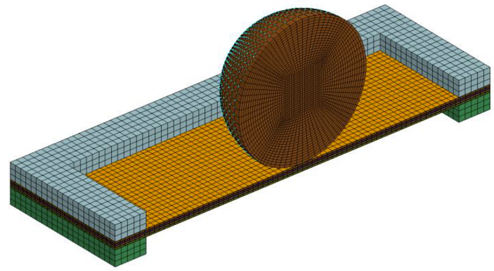
The remainder of this section describes the available computational models and experimental data used for the analysis and introduces the simplifications and assumptions made.

4.2 Models and Test Data

A model of the Z-2 spacesuit assembly was developed by the University of Delaware Center for Composite Materials using the LS-Dyna FEM software [42], shown in Figure 2 (a). The model comprised of $52,000$ shell elements and $29,700$ 3D solid elements, where solid elements were placed in the impact zones to more effectively predict damage. The fall scenario was simulated using an explicit reverse impact analysis - the suit assembly was translated in space at a prescribed velocity such that the HUT contacted a fixed 2 inch sphere - to mimic a realistic impact event. The suit was internally pressurized at 8.8 psi. The contact force during the simulation



(a)



(b)

Figure 2. LS-Dyna FEM models for simulating (a) impact on the HUT of the Z-2 assembly and (b) half symmetry impact on a test coupon of a low velocity impact experiment.

Parameter	Description
$AM1$	Coefficient of strain softening for fiber damage in direction 1
$AM2$	Coefficient of strain softening for fiber damage in direction 2
$AM3$	Coefficient of strain softening for fiber shear and crush
$AM4$	Coefficient of strain softening for matrix crack and delamination
$OMGMX$	Maximum admissible modulus reduction due to damage
$EEXPN$	Relative element expansion volume threshold for element erosion
$ELIMIT$	Axial tensile strain threshold for element erosion

Table 1. Descriptions of the key MAT162 material model parameters [44].

was recorded and the maximum force was extracted as the QoI. The assembly model required approximately 16 hours to execute using 10 CPU cores in LS-Dyna.

Progressive damage in the composite materials was modeled using the MAT162 material model [43, 44]. MAT162 was developed specifically to model 3D progressive damage in unidirectional and plain weave composites using solid elements and explicit analysis in LS-Dyna. In addition to more standard elastic and strength parameters, the material model contains several key variables that define post-damage softening behavior using continuum damage mechanics. These variables are listed in Table 1 along with short descriptions. For more information about the MAT162 material model and the definitions of the governing parameters, see [43, 44].

The damage parameters in Table 1 represent phenomenological variables that must be determined empirically for a given material by calibrating the model behavior to low velocity impact (LVI) experiments. Test data was provided for both the S-2 glass and PW IM10 composite materials to facilitate model calibration. Here, the LVI experiment considered 4-inch $\times$ 6-inch test coupons with clamped boundary conditions. An impactor with a mass of 29.883 lbs moving at velocity of 10.5 ft/sec was used and the resulting impact force versus time was measured. An LS-Dyna model of the LVI test setup was also provided, see Figure 2 (b). Due to symmetry, only half the coupon and impactor had to be analyzed.This model required approximately 2 hours to execute using 4 CPU cores in LS-Dyna.

The models and test data introduced in this section can be directly related to the notation provided in the problem setup in Section 3.1. Equation 1 represents the reverse impact analysis of the Z-2 spacesuit, where $\mathcal{M}$ denotes the LS-Dyna assembly model in Figure 2 (a), Y denotes the maximum contact force, and $\mathbf{\Theta}$ denotes MAT162 material parameters and impact velocity. The LVI test is represented by $\mathcal{E}$ in Equation (2) and $\mathcal{D}$ denotes the resulting contact force measurement data. Finally, the LS-Dyna model that simulates the LVI test in Figure 2 (b) is denoted by $\mathcal{M}_{\mathcal{E}}$ in Equation (3), where $Y_{\mathcal{E}}$ is the predicted contact force in the test coupon.

4.3 Simplifications and Assumptions

Due to both time constraints and a lack of explicit problem specifications over the course of the UQ R&A Project, several simplifications and assumptions were introduced to the case study to facilitate the PnIF analysis. Thus, the results shown in the following section are strictly for demonstration purposes and do not directly reflect the true reliability of the Z-2 spacesuit. In particular, the following simplifications were made:

1. Only impact to the HUT portion of the spacesuit is considered
   - In the xEMU certification program, the PnIF requirement must be verified for impacts to all components of the spacesuit (HUT, brief, PLSS, etc.). Here, only impact on the HUT is considered.
2. Assumed failure criteria
   - Impact failure is assumed to simply occur when the maximum contact force in the HUT exceeds 2800 lbf. This assumption was made in lieu of a more sophisticated leak rate failure criteria that was under development at the time of the UQ R&A Project.
3. Assumed impact velocity uncertainty
   - The uncertainty in the impact velocity is assumed to be described by a Gaussian distribution with mean equal to the nominal velocity used in the previous Z-2 design effort (5.90 ft/sec) and coefficient of variation of 1.25%.
4. Additional uncertainties affecting the impact scenario are ignored
   - Several additional factors that contribute to uncertainty in the impact loading were ignored, including the impactor (rock) shape and properties, manufacturing defects/variability, and the potential for energy absorption in regolith beneath the rock.
5. The likelihood of the HUT impact event is assume to be 1.0
   - In reality, the PnIF would depend not only on the probability of failure under the fall scenario considered, but also the likelihood of that fall event occurring. This likelihood was assumed to be 1.0 for the PnIF calculation.

6. Uncertainty was considered/quantified in only 1 of 2 materials in the composite laminate
    - Uncertainty was quantified using model calibration for only the S-2 glass material, while the properties of PW IM10 were fixed at nominal values determined in the previous Z-2 design effort. Note this likely results in an underestimation of uncertainty in the predicted maximum contact force and affects the PnIF value reported.

5 Results

This section presents results from applying the UQ approach described in Section 3 to the Z-2 spacesuit reliability case study introduced in Section 4. The end-to-end process of estimating the PnIF is demonstrated, including the quantification of material parameter uncertainty using LVI test data and the application of the MFIS method described in Section 3.6.2 to efficiently and accurately predict PnIF.

5.1 Quantifying Material Parameter Uncertainty

The first component of the Z-2 PnIF analysis was the estimation of the MAT162 parameter uncertainty using LVI test data. Recall from the previous section that only uncertainty in the S-2 glass material was considered. Based on subject matter expert recommendation for this material, the $OMGMX$ and $EEXPN$ parameters were fixed at values of 0.97 and 1.65, respectively, for the analysis. From here, a sensitivity analysis (Section 3.3) was conducted to identify the most/least important remaining MAT162 parameters from Table 1. The uncertainty in the subset of most influential parameters was then estimated using model calibration (Section 3.4). Since both of these methods require a large number of evaluations of the LVI experiment model (Figure 2 (a)), a surrogate model was first constructed to alleviate the computational burden associated with running the full fidelity LS-Dyna model. This section presents the results from each of these steps taken to arrive at material parameter uncertainty estimates.

5.1.1 Coupon Surrogate Model Construction

Following the discussion in Section 3.6.1, a surrogate model was developed to approximate the LS-Dyna LVI test model shown in Figure 2 (b) to allow for efficient sensitivity analysis and model calibration. First, a training dataset was generated. The following bounds were chosen for the MAT162 parameters based on their expected and feasible ranges [43, 44]:

$$
\begin{aligned}
AM1 &\in [0.001, 4.0] \\
AM2 &\in [0.001, 4.0] \\
AM3 &\in [0.001, 4.0] \\
AM4 &\in [-1.5, 1.5] \\
ELIMIT &\in [1.0, 7.0].
\end{aligned} \tag{11}
$$

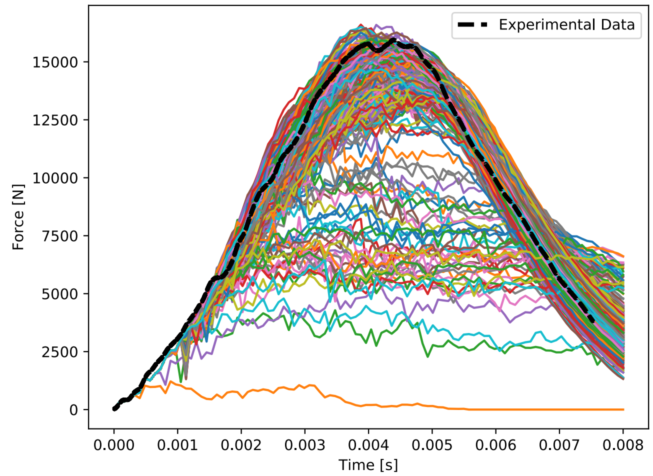


Figure 3. Force versus time outputs from 1200 evaluations of the LS-Dyna LVI test model for randomly sampled MAT162 parameter values.

The predicted force versus time response was then computed for 1200 different parameter values that were drawn uniformly at random from within the above bounds using the LVI test model. To alleviate the computational burden, 30 FEM simulations were run at a time using the automated batching software (Section 3.6.3), each of which was executed on four CPU cores such that all simulations were complete in around one day. The resulting force versus time curves are shown along with the LVI test data in Figure 3. The training dataset for the surrogate model was then produced by filtering out those curves (and corresponding MAT162 parameters) that produced a maximum force less than $14000N$, resulting in 799 input/output pairs.

Using the training dataset, a GP regression surrogate model was constructed[3] to predict the force at 50 evenly spaced points in time for a given set of MAT162 input parameter values. A separate test dataset containing 192 input/output pairs was generated using the same process outlined above to assess the effectiveness of the surrogate model. Figure 4 shows a scatter plot of predicted maximum contact force for the surrogate model (y axis) versus the FEM model, showing a reasonable degree of accuracy ($RMSE = 378.64$) for a fraction of the computational cost.

5.1.2 Sensitivity Analysis

After constructing a surrogate model that could effectively relate MAT162 material parameters to the resulting contact force in a simulated LVI test, a global sensitivity analysis was conducted to quantify the influence each parameter had on the predicted response. Specifically, the sensitivity of the peak force from LVI to the variation of the MAT162 parameters within the bounds from Equation (11) was computed following the approach in Section 3.4. The resulting parameter sensitivity ranked in

[3]The Scikit-Learn Python module was used for GP regression [45]

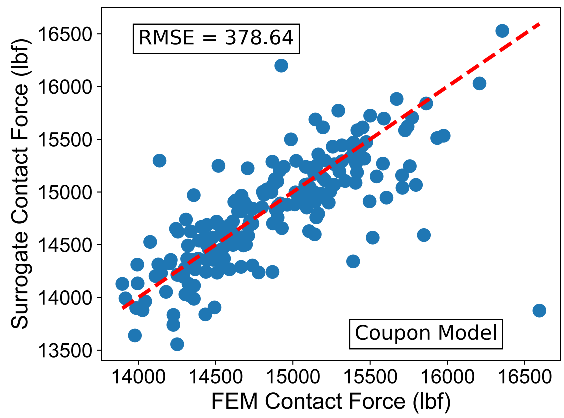


Figure 4. Surrogate model accuracy for the coupon FEM model.

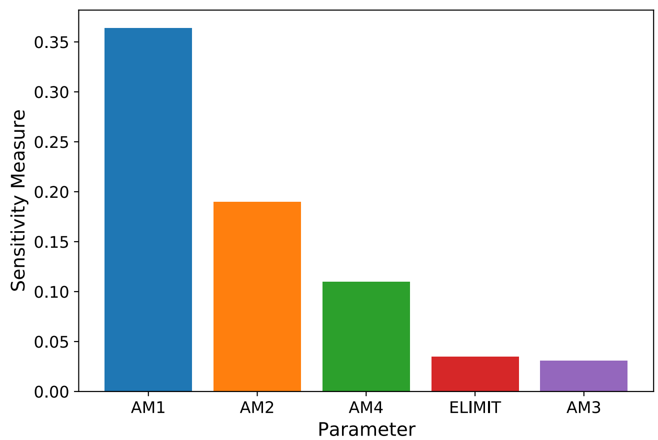
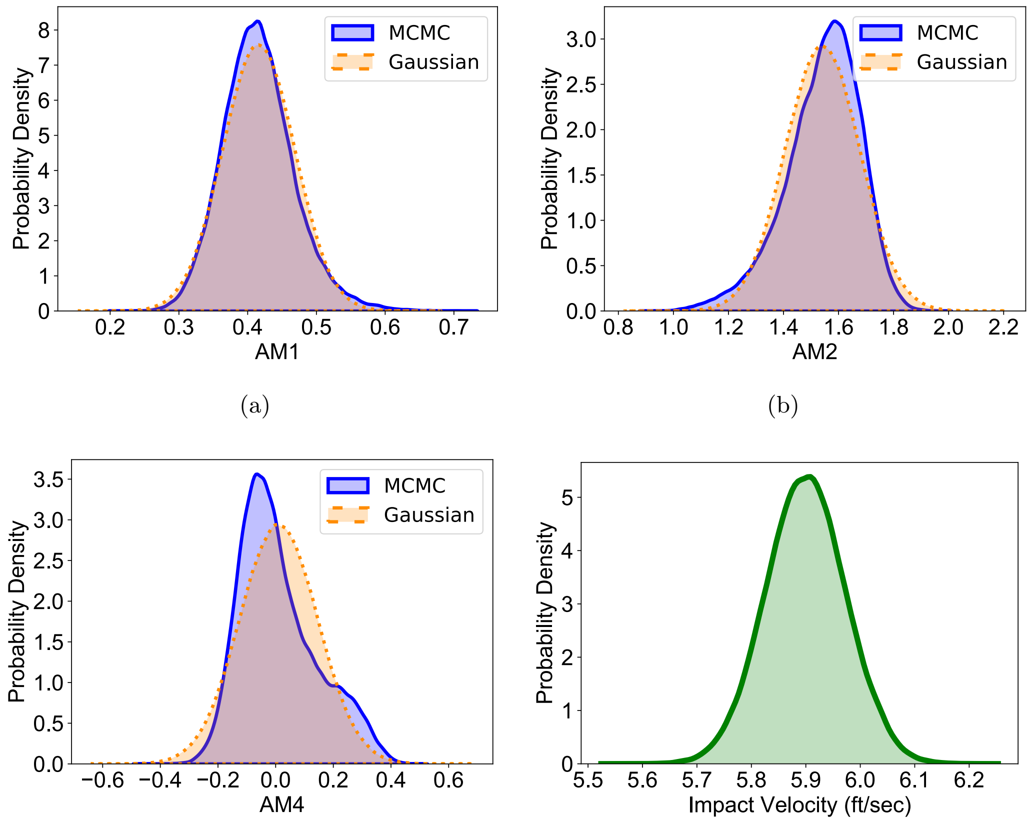


Figure 5. Sensitivity of the peak force in the coupon model to each MAT162 parameter.

order of importance is shown in Figure 5. Based on this analysis, uncertainty was quantified in the three most influential parameters, $AM1$, $AM2$, and $AM4$, while the values of the least influential parameters, $ELIMIT$ and $AM3$, were fixed at their deterministic, "best-fit" values.

5.1.3 Model Calibration

A probabilistic model calibration was performed to quantify the uncertainty in the three most influential MAT162 parameters selected via sensitivity analysis. The inverse problem was formulated using Bayesian inference and MCMC simulation was used to approximate the probability distributions of the input parameters given the available LVI test data, as described in Section 3.3. Again, the GP surrogate model was used to significantly speedup the model calibration process. A total of $500,000$ samples were generated using MCMC to estimate the uncertainty in the



(a) (b)

(c) (d)

Figure 6. Input parameter PDFs.

MAT162 parameters. The resulting probability distributions are shown by the blue curves in Figure 6 (a), (b), and (c). It can be seen that there is a significant amount of variability in the parameters, meaning there is a relatively large region of the parameter space that yield predicted responses close enough to the LVI data to be considered probable.

For illustration, the uncertainties in Figure 6 were propagated through the surrogate model to produce the full force versus time curve, shown in Figure 7. Here, each LVI test data point used in the calibration is represented by a red “x”. The probabilistically calibrated model response comprises three components: a black line representing the median prediction, a dark grey region representing the 95% credible intervals, and a light grey region representing the 95% prediction intervals. The credible intervals contain 95% of the simulated forces at a given time based on model input uncertainty alone. The prediction intervals consider both model input uncertainty and measurement noise and thus can be interpreted as containing a potential future measurement with 95% probability. The primary take away is that the probabilistic model calibration fits the data rather well and also provides a measure of uncertainty that can facilitate the reliability analysis.

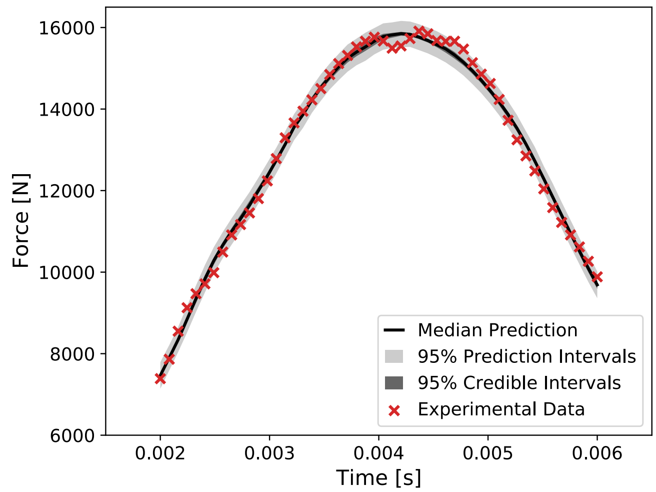


Figure 7. Comparison of the calibrated coupon (surrogate) model predictions versus the LVI test data.

5.2 Estimating Reliability (PnIF)

The section provides results from applying uncertainty propagation to estimate Z-2 spacesuit reliability (i.e., PnIF) considering uncertainty in material properties and impact velocity. The FEM model of the Z-2 spacesuit (Figure 2 (a)) was used to simulate impact to the HUT section of the suit assembly. The PnIF was calculated as $1 - P_f$, where probability of failure, P_f, is defined in Equation (7). Recall from Section 4 that Y represents the maximum contact force in the HUT portion of the suit and the failure threshold was assumed to be $y^c = 2800$ lbf. Furthermore, the impact velocity uncertainty was assumed to follow a Gaussian distribution with a mean of 5.90 ft/sec and coefficient of variation of 1.25%, shown in Figure 6 (d).

Uncertainty in material properties is based off the probabilistic model calibration presented in the previous section. In order to demonstrate the MFIS method introduced in Section 3.6.2 for predicting P_f, however, a Gaussian distribution was fit to the calibrated material parameter uncertainty using MCMC. This was done to simplify the evaluation of the input PDF, $p(\boldsymbol{\theta})$, to calculate the weights, $w(\boldsymbol{\theta})$, in Equation (10).[4] The Gaussian approximations to the input PDFs are shown by the orange curves in Figure 6 (a), (b), and (c), where good agreement with the original MCMC-based probability distributions can be seen.

In summary, the probability distribution, $p(\boldsymbol{\theta})$, for the four uncertain model parameters, $\boldsymbol{\theta} = [AM1, AM2, AM4, V_i]$, is explicitly specified as follows

$$p(\boldsymbol{\theta}) = N(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta), \tag{12}$$

[4] Note: a closed form density function was used out of convenience but is not believed to be strictly necessary. A more advanced sample-based estimate, e.g., leveraging [46], is a potential alternative.

where N is a Gaussian distribution with mean

$$\boldsymbol{\mu}_\theta = \left[4.160\text{x}10^{-1} \;\; 1.542 \;\; 1.031\text{x}10^{-2} \;\; 5.900\right]^T \tag{13}$$

and covariance matrix

$$\boldsymbol{\Sigma}_\theta = \begin{bmatrix} 2.750\text{x}10^{-3} & -4.940\text{x}10^{-3} & -3.732\text{x}10^{-3} & 0.0 \\ -4.940\text{x}10^{-3} & 1.856\text{x}10^{-2} & 3.201\text{x}10^{-3} & 0.0 \\ -3.732\text{x}10^{-3} & 3.201\text{x}10^{-3} & 5.505\text{x}10^{-2} & 0.0 \\ 0.0 & 0.0 & 0.0 & 5.439\text{x}10^{-3} \end{bmatrix} \tag{14}$$

Here, V_i denotes the impact velocity. Equation (14) shows that the probabilistic model calibration approach not only estimates joint PDFs of the material parameters but also the covariance structure describing the correlations between them. Note that the impact velocity is assumed to be independent of the material parameters here.

The PnIF was first predicted using standard MC simulation with a surrogate model of the Z-2 spacesuit FEM model. With the efficiency of the surrogate model, the effects of material uncertainty only versus combined material and impact velocity uncertainty could be studied separately. Then, an MFIS estimate was generated that combined predictions from both the surrogate model and the high-fidelity FEM model, as described in Section 3.6.2. A high-fidelity MC estimate was also generated using 2500 evaluations of the Z-2 spacesuit FEM model for comparison. The following sections provide these estimates as well as details of the Z-2 spacesuit surrogate model development.

5.2.1 Assembly Surrogate Model Construction

Following the discussion in Section 3.6.1, a surrogate model was developed to approximate the LS-Dyna Z-2 spacesuit model shown in Figure 2 (a). First, a training dataset was generated. The maximum contact force was computed and stored for 200 input samples randomly drawn from the probability distribution in Equation (12). To alleviate the computational burden, 20 FEM simulations were run at a time using the automated batching software (Section 3.6.3), each of which was executed on ten CPU cores such that all simulations needed for training data were completed in around one week.

Using the training dataset, a GP surrogate model was constructed to predict the maximum contact force in the spacesuit as a function of the four random input parameters. A separate test dataset containing 200 input/output pairs was generated to assess the effectiveness of the surrogate model. Figure 8 shows a scatter plot of predicted maximum contact force for the surrogate model versus the FEM model, showing a reasonable degree of accuracy ($RMSE = 48.67$) for a fraction of the computational cost.

5.2.2 Surrogate-Based Monte Carlo Reliability Estimate

Using the GP surrogate model as an approximation to the Z-2 FEM model, the uncertainty in the input parameters described by Equation (12) was propagated

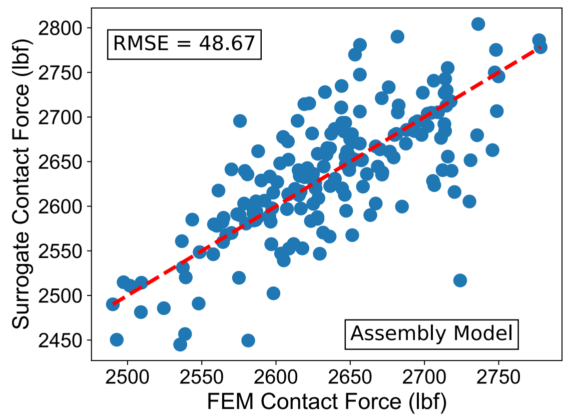


Figure 8. Surrogate model accuracy for the assembly FEM model.

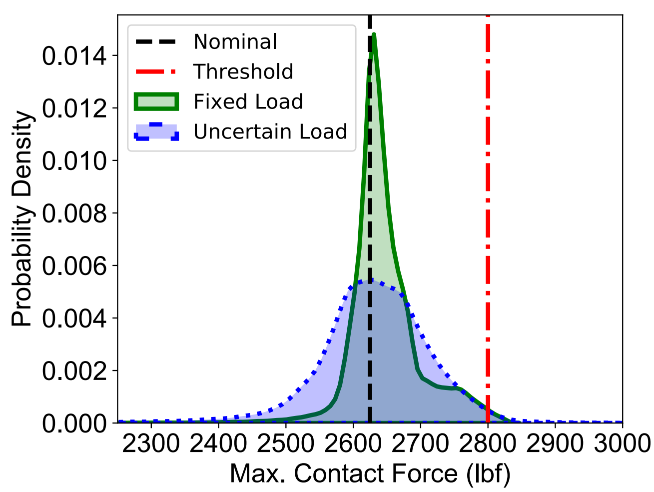


Figure 9. Distribution of maximum contact force in the suit assembly model.

	Deterministic	Uncertain Material	Uncertain Material & Load
Reliability (PnIF)	1.000	0.9918	0.9852

Table 2. PnIF calculated with a deterministic (using nominal material parameters and impact velocity) and probabilistic approach using surrogate model-based MC simulation.

to probabilistic predictions of maximum contact force using MC simulation with 10^5 samples. The resulting PDFs of maximum contact force are shown in Figure 9 for both the case of a nominal load (green) and an uncertain load (blue) due to the random impact velocity. The deterministic prediction of maximum contact force is also shown in comparison to the failure threshold of 2800 lbf. The PnIF is represented by the area under the PDFs to the left of the failure threshold in Figure 9, calculated using Equation (8) based on the samples of maximum contact force from the MC simulation. These estimated PnIF values are shown in Table 2.

For the assumptions made and the problem setup described in this report, it can be seen that the deterministic prediction of maximum contact force using nominal input parameter values is significantly lower than the failure threshold (i.e., PnIF = 1.000). However, considering the calibrated uncertainty in the MAT162 material model only with a fixed impact velocity resulted in 287.6lbf of variation[5] in the predicted contact force in the suit, representing 10.9% variability with respect to the uncertainty-free nominal prediction. In this case, the calculated PnIF was 0.9918. This result underscores the importance of probabilistic model calibration since a deterministic approach that chose a single parameter setting would have ignored the non-negligible probability of failure due to material uncertainty. Recall also that uncertainty in only one of two materials was accounted for, so the actual variability caused by material uncertainty is most likely higher. Finally, when uncertainty in the impact velocity is also considered, there is significantly more variability (469lbf, or 17.8% with respect to the nominal prediction) in the predicted maximum contact force in Figure 9, as expected. In this case, the PnIF is reduced to 0.9852.

5.2.3 Multifidelity Importance Sampling Reliability Estimate

While estimating reliability with a surrogate model-based approach is commonly used due to its computational efficiency, the resulting estimate will generally be biased due to discrepancies in the surrogate model approximation (see Figure 8). As discussed in Section 3.6.2, the MFIS method offers an unbiased reliability estimate at the expense of a moderate number of additional high-fidelity model evaluations, representing an effective balance between low- and high-fidelity MC simulation. This section details an MFIS estimate of PnIF that was produced in this case study.

To estimate PnIF with MFIS, the surrogate model described in Section 5.2.1 was used to *efficiently* generate a bias distribution, $q(\boldsymbol{\theta})$, and then the Z-2 FEM model was used to evaluate Equation (10) to yield an *unbiased* estimator. Here, MC simulation was performed with the surrogate model with 10^5 random samples from

[5] Defined as the difference between the 99^{th} and 1^{st} percentiles of the PDF

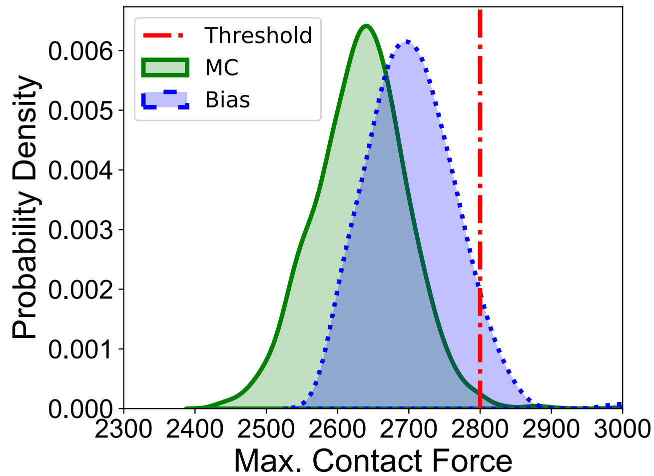


Figure 10. Distribution of maximum contact force when sampling original input distribution versus the bias distribution for MFIS.

the input distribution (Equation (12)) to identify the subset of samples, $\{\boldsymbol{\theta}_f^{(i)}\}$, that resulted in failure, i.e., maximum contact force predictions greater than 2800lbf. A Gaussian mixture model was fit to these samples to produce the bias distribution, $q(\boldsymbol{\theta})$. Then the Z-2 FEM model was evaluated for 250 random input samples drawn from $q(\boldsymbol{\theta})$ to calculate P_f^{IS} from Equation (10). Finally, the PnIF using MFIS was obtained as $1 - P_f^{IS}$.

The PDF of maximum contact force values computed using the Z-2 FEM model and the MFIS bias distribution, $q(\boldsymbol{\theta})$, is shown in Figure 10. For comparison, the PDF of 2500 predicted force samples using the original input distribution, $p(\boldsymbol{\theta})$, is shown as well. Importantly, it can be seen that the PDF from the bias distribution is shifted towards higher maximum contact force values, resulting in more samples above the failure threshold. Specifically, the bias distribution leads to a 6.4% failure rate (16 out of 250 samples) versus the 0.68% failure rate observed from the original input distribution (17 out of 2500 samples). This is the mechanism by which MFIS (and more generally, importance sampling) obtains accuracy gains for estimating P_f.

Table 3 compares the estimated PnIF using the low-fidelity (surrogate model-based) MC approach described in the previous section, using high-fidelity (Z2 FEM model-based) MC simulation, and MFIS. The table also includes the number of Z2 FEM model evaluations required to obtain the estimates. It can be seen that the MFIS PnIF estimate shows significantly closer agreement with the high-fidelity MC estimate versus the low-fidelity MC estimate (0.02% difference versus 0.81% difference). The reliability estimates for each method are further compared in Figure 11, which includes error bars representing the 95% confidence intervals. It can be seen that the confidence interval widths for the MFIS and high-fidelity MC estimates are comparable, despite MFIS requiring significantly fewer FEM model evaluations. This is a consequence of the higher failure rate achieved with the MFIS approach. The low-fidelity MC estimate and confident interval lie well outside the confidence

	Low-Fidelity MC	High-Fidelity MC	MFIS
Reliability (PnIF)	0.9852	0.9932	0.9930
# FEM Evaluations	200	2500	450

Table 3. Comparisons of reliability estimates for high- and low-fidelity MC versus MFIS.

interval for high-fidelity MC, reflecting a prediction bias due to approximation errors with respect to the high-fidelity FEM model. This highlights the potential pitfalls of relying solely on a low-fidelity surrogate model for UQ.

From the number of Z-2 FEM evaluations reported in Table 3, it can be seen that the MFIS solution represents a tradeoff between computational expense and accuracy. Both the low-fidelity MC and MFIS estimates require 200 FEM evaluations to generate training data for the surrogate model, but then MFIS required an additional 250 FEM evaluations to compute Equation (10). Both approaches also used 100000 evaluations of the surrogate model, but this expense is assumed to be negligible since the total computation time for this was ~ 10 seconds. The tradeoff between computational efficiency and accuracy observed here suggests the following "tiered" approach to reliability analysis: 1) a surrogate-model based prediction as a fast but potentially inaccurate preliminary estimate, 2) an MFIS prediction as an unbiased estimator that requires a moderate level of computational expense, 3) a brute force, high-fidelity MC estimate that is most rigorous but potentially intractable for some applications and computational models. Finally, the effectiveness of the MFIS approach is known to depend particularly on the accuracy of the surrogate model predictions near the failure threshold [39]. Accurate failure region resolution leads to more effective bias distributions which in turn increases the observed failure rate for MFIS and precision in reliability estimates. Thus, developing advanced training strategies to produce surrogate models with this characteristic for use with MFIS is the subject of ongoing and future research.

6 Conclusions

This report presented a probabilistic analysis case study focused on estimating the reliability of a next-gen spacesuit design under impact loading. While the work was motivated by the certification of the new Exploration Extravehicular Mobility Unit (xEMU), computational models and test data were used from the previous generation of spacesuit, the Z-2, for demonstration purposes. Considering uncertainty in impact velocity and the material parameters that govern damage in the composite suit, the probability of no impact failure (PnIF) for the Z-2 was estimated using an uncertainty quantification (UQ) workflow. The study focused in particular on approaches for overcoming two of the typical challenges with employing UQ methods in practice: estimating uncertainty in model parameters and alleviating the prohibitive computational burden associated with UQ when expensive, high-fidelity models are needed. For the former, sensitivity analysis was used to identify the most influential material parameters in causing the assumed failure condition and

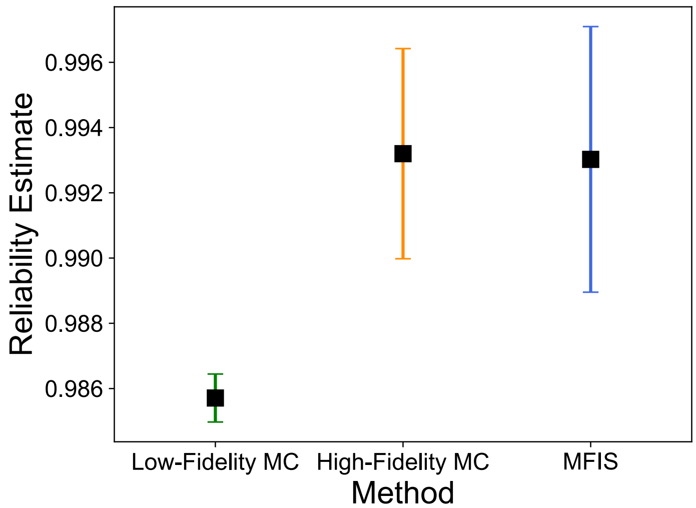


Figure 11. Reliability (PnIF) estimates with 95% confidence intervals.

then Bayesian inference was used to estimate the probability distributions of those parameters from LVI test data. For the latter, it was shown that surrogate modeling, high performance computing, and an advanced reliability analysis method, MFIS, can be leveraged to perform UQ in a computationally tractable way despite relying on expensive FEM models for predictions. Several simplifying assumptions were introduced to conduct the PnIF demonstration under time constraints, including ignoring a range of additional sources of uncertainty in astronaut fall probabilities, spacesuit manufacturing defects, impactor shapes/sizes, etc., and were listed in detail in the report. The results showed that the quantified material parameter uncertainty results in at least 10% variation in the predicted contact force with respect to the uncertainty-free nominal prediction, underscoring the importance of the model calibration process. Finally, the MFIS method was shown to produce PnIF estimates that were more accurate than relying on a surrogate model alone, but more efficient than a brute force MC simulation approach that used only the Z-2 FEM model.

References

1. Zang, T. A.; Hemsch, M. J.; Hilburger, M. W.; Kenny, S. P.; Luckring, J. M.; Maghami, P.; Padula, S. L.; and Stroud, W. J.: Needs and Opportunities for Uncertainty-Based Multidisciplinary Design Methods for Aerospace Vehicles. *NASA/TM-2002-211462*, 2002.

2. Raju, I.; Stadler, J.; Kramer-White, J.; and Piascik, R.: White Paper on Factors of Safety. *NASA/TM-2009-215723*, 2009.

3. Walker, E.; Hemsch, M. J.; and West, T.: *Integrated Uncertainty Quantification for Risk and Resource Management: A NASA Langley Perspective (Invited).* 2015. URL `https://arc.aiaa.org/doi/abs/10.2514/6.2015-0501`.

4. Zipay, J. J.; Modlin, T. C.; and Larsen, C. E.: The Ultimate Factor of Safety for Aircraft and Spacecraft - Its History, Applications, and Misconceptions. *JSCCN-33063*, 2016.

5. Verderaime, V.: Aerostructural Safety Factor Criteria Using Deterministic Reliability. *Journal of Spacecraft and Rockets*, vol. 3, no. 2, 1993.

6. Smith, J. P.: The Case for Lower Factors of Safety With Modern Aerospace Metals. *REPORTNUMBER*, 2020.

7. Mason, B.; and Krishnamurthy, T.: *Probabilistic Analysis of a Composite Crew Module*. URL `https://arc.aiaa.org/doi/abs/10.2514/6.2011-2035`.

8. Scotti, S. et al.: Probabilistic Design Case Study - Composite Crew Module. *NESC-RP-10-00611*, 2010.

9. Sills, J.; Blelloch, P.; and Kammer, D.: NESC Peer Review of Exploration Systems Development (ESD) Integrated Vehicle Modal Test, Model Correlation, Development Flight Instrumentation (DFI) and Flight Loads Readiness. *NASA/TM-2021-0009733, NESC-RP-16-01110*, 2021.

10. Standards for Models and Simulations. *NASA-STD-7009*, 2008.

11. NASA Handbook for Models and Simulations: An Implementation Guide for NASA-STD-7009. *NASA-HDBK-7009*, 2013.

12. Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners. *NASA/SP-2011-3421*, 2011.

13. Dezfuli, H.; Kelly, D.; Smith, C.; Vedros, K.; and Galyean, W.: Bayesian Inference for NASA Probabilistic Risk and Reliability Analysis. *NASA/SP-2009-569*, 2009.

14. Vander Kam, J. C.; and Gage, P.: Estimating Orion heat shield failure due to ablator cracking during the EFT-1 mission. *AIAA Thermophysics Conference*, 2016.

15. Tobin, S. A.; Brune, A. J.; and Bowes, A. L.: Probabilistic Risk Analysis and Margin Process for a Flexible Thermal Protection System. *International Conference on Flight Vehicles, Aerothermodynamics and Re-entry Missions and Engineering*, 2019.

16. Mason, B. H.; Chen, T.; Krishnamurthy, T.; Spain, C. V.; and Naser, A. S.: Probabilistic analysis and design of a raked wing tip for a commercial transport. *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2008.

17. Smith, R. C.: *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, USA, 2013.

18. O'Hagan, A.; and Oakley, J. E.: Probability is perfect, but we can't elicit it perfectly. *Reliability Engineering, and System Safety*, vol. 85, 2004, pp. 239–248.

19. Ferson, S.; and Ginzburg, L. R.: Different methods are needed to propagate ignorance and variability. *Reliability Engineering, and System Safety*, vol. 54, 1996, pp. 133–144.

20. Rubinstein, R. Y.; and Kroese, D. P.: *Simulation and the Monte Carlo Method*. Wiley Publishing, 3rd ed., 2016.

21. Kennedy, M. C.; and O'Hagan, A.: Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, 2001, pp. 425–464. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00294`.

22. Gilks, W.; Richardson, S.; and Spiegelhalter, D.: *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics, Taylor & Francis, 1995. URL `http://books.google.com/books?id=TRXrMWY_i2IC`.

23. Borgonovo, E.: A new uncertainty importance measure. *Reliability Engineering & System Safety*, vol. 92, no. 6, 2007, pp. 771–784.

24. Plischke, E.; Borgonovo, E.; and Smith, C. L.: Global sensitivity measures from given data. *European Journal of Operational Research*, vol. 226, no. 3, 2013, pp. 536–550.

25. Herman, J.; and Usher, W.: SALib: an open-source Python library for sensitivity analysis. *Journal of Open Source Software*, vol. 2, no. 9, 2017, p. 97.

26. Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; and Tarantola, S.: *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

27. McKay, M. D.: Latin Hypercube Sampling as a Tool in Uncertainty Analysis of Computer Models. *Proceedings of the 24th Conference on Winter Simulation*, WSC '92, Association for Computing Machinery, New York, NY, USA, 1992, pp. 557–564. URL `https://doi.org/10.1145/167293.167637`.

28. Rasmussen, C. E.; and Williams, C. K. I.: *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

29. Gramacy, R. B.: *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences*. Chapman Hall/CRC, Boca Raton, Florida, 2020. `http://bobby.gramacy.com/surrogates/`.

30. Su, G.; Peng, L.; and Hu, L.: A Gaussian process-based dynamic surrogate model for complex engineering structural reliability analysis. *Structural Safety*,

vol. 68, 2017, pp. 97–109. URL
https://www.sciencedirect.com/science/article/pii/S016747301730214X.

31. Rajabi, M. M.; and Ketabchi, H.: Uncertainty-based simulation-optimization using Gaussian process emulation: Application to coastal groundwater management. *Journal of Hydrology*, vol. 555, 2017, pp. 518–534. URL https://www.sciencedirect.com/science/article/pii/S0022169417307163.

32. Gramacy, R. B.; and Lee, H. K. H.: Adaptive Design and Analysis of Supercomputer Experiments. *Technometrics*, vol. 51, no. 2, 2009, pp. 130–145. URL http://www.jstor.org/stable/40586591.

33. Brochu, E.; Cora, V. M.; and De Freitas, N.: A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*, 2010.

34. Bect, J.; Ginsbourger, D.; Li, L.; Picheny, V.; and Vazquez, E.: Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, vol. 22, no. 3, 2012, pp. 773–793.

35. Peherstorfer, B.; Willcox, K.; and Gunzburger, M.: Survey of multifidelity methods in uncertainty propagation, inference, and optimization. 2018.

36. Gorodetsky, A.; Geraci, G.; Eldred, M.; and Jakeman, J. D.: A generalized approximate control variate framework for multifidelity uncertainty quantification. *Journal of Computational Physics*, 2020, p. 109257. URL http://www.sciencedirect.com/science/article/pii/S0021999120300310.

37. Bomarito, G. F.; Leser, P. E.; Warner, J. E.; and Leser, W. P.: On the Optimization of Approximate Control Variates with Parametrically Defined Estimators. 2020.

38. Giles, M. B.: Multi-level Monte Carlo path simulation. *Operations Research*, vol. 56, no. 3, 2008, pp. 607–617.

39. Peherstorfer, B.; Cui, T.; Marzouk, Y.; and Willcox, K.: Multifidelity importance sampling. *Computer Methods in Applied Mechanics and Engineering*, vol. 300, 12 2015.

40. West, T. K.; Reuter, B. W.; Walker, E. L.; Kleb, B.; and Park, M. A.: Uncertainty quantification and certification prediction of low-boom supersonic aircraft configurations. *Journal of Aircraft*, vol. 54, no. 1, January 2017, pp. 40–53.

41. Crespo, L. G.; Kenny, S. P.; and Giesy, D. P.: Interval predictory models with linear parameter dependency. *Journal of Verification, Validation, and Uncertainty Quantification*, vol. 1, no. 2, June 2016.

42. LS-DYNA Keyword User's Manual (Version 971). *Livermore Software Technology Corporation*, 2007.

43. Gama, B. A.; Bogetti, T.; and Gillespie, J. W.: Progressive Damage Modeling of Plain-Weave Composites using LS-Dyna Composite Damage Model MAT162. *Proceedings and CD Rom of 7th European LS-DYNA Conference*, May 2009.

44. Gama, B. A.: A progressive composite damage model for unidirectional and woven fabric composites: MAT162 User Manual Version 17A-2017. 2017.

45. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E.: Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, vol. 12, 2011, pp. 2825–2830.

46. Nguyen, T. L.; Septier, F.; Peters, G. W.; and Delignon, Y.: Efficient sequential Monte-Carlo Samplers for Bayesian Inference. *IEEE Transactions on Signal Processing*, vol. 64, no. 5, 2016, pp. 1305–1319.

7 Appendix

7.1 UQ R&A Project Overview

The Engineering R&A Program was launched in FY2020, aiming to restore NASA's investment in fundamental engineering capabilities by sponsoring a cohesive portfolio of engineering research across the Agency. In doing so, new tools, techniques, and methods could be elevated from research-level usage into standard NASA practice. Emphasis was also placed on leveraging relevant activities and expertise at other national labs, industry, and academia. The UQ R&A Project was selected as one of five Engineering R&A topics for funding.

The UQ R&A Project was motivated by limitations of factors of safety (FoS) and uncertainty factors that are traditionally used for design at NASA. These factors may be applied sequentially by multiple teams throughout design, stacking conservatism, inefficiencies, and cost growth. Furthermore, the commonly used 1.4 FoS in NASA standards resulted from decades of experience and so the proper specification for new vehicle types, materials, and environments poses a great challenge. Future space missions will require a rational basis for meeting stringent mass requirements and narrower margins where the application of typical FoS is impractical. Many probabilistic tools exist that can address these challenges, but best practices for their use and implementation in design and operations at NASA is lacking.

The goal of the UQ R&A Project was to help integrate modern UQ methodology and tools into engineering practice at NASA. In doing so, a foundation would be provided for a long-term transition away from FoS and towards reliability-based design. By taking a reliability-based approach to design, NASA can potentially enable substantially reduced mass/costs, higher performance through narrower margins, rigorous assessments of mission risk and the tradeoff between risk and costs, and ultimately, safer and more reliable flight systems. Under the UQ R&A Project, an emphasis was placed on broadly developing a sustained workforce in the area of UQ over the course of the originally slated four year project duration. This was to be done by establishing and documenting best practices for UQ-based analysis, identifying viable existing UQ tools/software as well as developing new, in-house capabilities, and aiming to overcome common barriers to the adoption of UQ practices, including computational intractability. The certification of the xEMU spacesuit was selected as the primary case study with which to highlight the advantages of a probabilistic analysis.

In early 2020, it was announced that the entire R&A Program was being defunded as part of broader cuts to discretionary funding across NASA. The UQ R&A Project was thus a one year effort only with a substantially reduced budget instead of the originally planned four year effort. Given the reduced scope, the main objective of the project shifted towards laying a foundation for future work to sustain progress in the areas of UQ and reliability-based design at NASA. This was done both through strategic procurement purchases and persistent advocacy for follow-on support using preliminary project results. Deliverables related to the xEMU spacesuit certification case study were given higher priority given the current importance of xEMU to NASA and the infusion opportunity it presented. The team worked quickly to put a basic UQ workflow in place and demonstrate its effectiveness with estimating the reliability for spacesuit design. In this way, initial versions of in-house UQ capabilities were developed for future use at NASA while preliminary results were used to secure follow-on support under the xEMU project.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704–0188*

1. REPORT DATE *(DD-MM-YYYY)*
01-08-2021

2. REPORT TYPE
Technical Memorandum

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Assessing Next-Gen Spacesuit Reliability: A Probabilistic Analysis Case Study

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
James E. Warner and Patrick E. Leser and William P. Leser and D. Austin Cole and Geoffrey F. Bomarito

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
347224.01.04

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, Virginia 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NASA/TM–20210019495

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified-Unlimited
Subject Category 61
Availability: NASA STI Program (757) 864-9658

13. SUPPLEMENTARY NOTES
An electronic version can be found at http://ntrs.nasa.gov.

14. ABSTRACT
Motivated by a current effort to certify the Exploration Extravehicular Mobility Unit (xEMU) spacesuit under NASA's Artemis Program, this report presents a probabilistic analysis case study focused on estimating the reliability of a spacesuit design for a potential fall scenario. Considering uncertainty in impact velocity and material parameters that govern damage in a composite spacesuit, the probability of no impact failure (PnIF) was estimated through the application of several uncertainty quantification (UQ) concepts. In particular, the focus of the case study was on providing practical solutions to common challenges of applying UQ in practice: estimating uncertainty in model parameters and alleviating the prohibitive computational burden associated with UQ when expensive, high-fidelity models are needed. Material parameter uncertainty was estimated through a combination of sensitivity analysis and Bayesian model calibration. Despite the computational model for suit impact requiring nearly one day of serial execution time, the PnIF was estimated in a tractable amount of time through a multifidelity importance sampling (MFIS) approach that leveraged surrogate modeling and high performance computing. It is shown that quantified material parameter uncertainty results in nearly 290lbf of variation in predicted contact force in the suit ($> 10\%$ variability with respect to the uncertainty-free nominal prediction). Furthermore, the MFIS prediction of PnIF agrees well with a brute force Monte Carlo simulation estimate but is significantly more efficient.

15. SUBJECT TERMS
Uncertainty Quantification, Reliability, Monte Carlo

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Information Desk (help@sti.nasa.gov)
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	34	(757) 864-9658