Taylor & Francis
Taylor & Francis Group

Quality Engineering

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/lqen20

Uncertainty quantification for ground vehicle vulnerability simulation

Thomas H. Johnson, Dhruv K. Patel, John T. Haman, Jeremy S. Werner & Dave Higdon

To cite this article: Thomas H. Johnson, Dhruv K. Patel, John T. Haman, Jeremy S. Werner & Dave Higdon (29 Aug 2024): Uncertainty quantification for ground vehicle vulnerability simulation, Quality Engineering, DOI: 10.1080/08982112.2024.2394437

To link to this article: https://doi.org/10.1080/08982112.2024.2394437

Published online: 29 Aug 2024.

Submit your article to this journal

Article views: 15

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=lqen20

QUALITY ENGINEERING
https://doi.org/10.1080/08982112.2024.2394437

Taylor & Francis
Taylor & Francis Group

CASE REPORT

Check for updates

Uncertainty quantification for ground vehicle vulnerability simulation

Thomas H. Johnson[a], Dhruv K. Patel[a], John T. Haman[a], Jeremy S. Werner[b], and Dave Higdon[c]

[a]Institute for Defense Analyses, Alexandria, Virginia, USA; [b]Operational Test and Evaluation, Department of Defense, Washington, DC, USA; [c]Department of Statistics, Virginia Tech, Blacksburg, Virginia, USA

ABSTRACT

A vulnerability assessment of a combat vehicle uses modeling and simulation (M&S) to predict the vehicle's vulnerability to enemy attack. The system-level output of the M&S is the probability that the vehicle's mobility is degraded as a result of the attack. The M&S models this system-level phenomenon by decoupling the attack scenario into a hierarchy of subsystems. Each subsystem addresses a specific scientific problem, such as the fracture dynamics of an exploded munition, or the ballistic resistance provided by the vehicle's armor. For each subsystem, laboratory testing is conducted to gather data to fit a subsystem-level model. The M&S hierarchically connects the subsystem-level models to enable prediction of the system-level output. As part of the DoD's ongoing effort to improve M&S using verification, validation, and uncertainty quantification, we present a case study that propagates uncertainties in the hierarchy of sub-models to the system-level output.

KEYWORDS

design of experiments; live fire test and evaluation; propagation of uncertainty; uncertainty quantification; validation; verification; vulnerability assessment

1. Introduction

As part of acquiring a new combat vehicle, the Department of Defense (DoD) conducts survivability testing and then submits a live fire test and evaluation (LFT&E) report to Congress before full-rate production of the vehicle begins. Live fire testing of a new combat vehicle includes Full-Up System-Level (FUSL) tests, as well as various combinations of component-, subsystem-, and system-level tests. FUSL testing is mandated by Congress, and involves "realistic survivability testing for vulnerability of the system in combat by firing munitions likely to be encountered in combat at the system configured for combat." (U.S. Code Title 10 Section 4172). FUSL tests are inherently costly and generate minimal data quantities; nevertheless, they offer unparalleled battlefield-realistic vulnerability insights prior to the vehicle's deployment.

The LFT&E report includes a vulnerability assessment that examines the characteristics of the vehicle that cause it to suffer degradation or loss of capability to perform the designated mission as a result of enemy attack. The assessment primarily draws upon data from live fire testing, supported also by information from design analyses, examinations of combat and safety incidents, controlled damage experiments, developmental and operational tests, as well as modeling and simulation (M&S).

Subsystem-level tests, such as armor characterization or ballistic exploitation tests, are more affordable and produce data in greater quantities than FUSL tests, are executed in a controlled laboratory environment, and address specific scientific and engineering problems. For example, the armor of the combat vehicle undergoes isolated testing on a dedicated test stand within a laboratory setting to assess its ballistic resistance capabilities, independent of its integration with the rest of the platform. Results from subsystem-level tests independently contribute to a vulnerability assessment, and are also used in the development of M&S.

The DoD has developed a specialized form of M&S designed to approximate the intricate physics involved in the destruction of combat platforms when subjected to enemy fire. DoD scientists address the problem by breaking down the physical system into a hierarchy of subsystems, each further composed of its own interconnected subsystems. The M&S is similarly built up from a hierarchy of subsystem-level models (sub-models for short) that are individually fit to subsystem-level test data. The M&S hierarchically arranges the sub-models to enable prediction of the system-level quantity of interest (QOI), such as the state of damage or residual functionality of the platform. The M&S has a fast execution time because the sub-models

CONTACT Thomas Johnson thjohnso@ida.org Institute for Defense Analyses, Alexandria, VA.
© 2024 Institute for Defense Analyses. Published with license by Taylor & Francis Group, LLC

generally do not use finite element methods, and instead rely on fast-running empirical algorithms, such as physics-based approximations of governing equations, statistical models, mathematical models, look-up tables, and surrogate models.

This form of M&S was developed decades ago and has been continuously improved since. The Army and Air Force individually develop and maintain their own M&S for assessing system-level vulnerability, called the Advanced Joint Effectiveness Model (AJEM), and Computation of Vulnerable Areas and Repair Times (COVART), respectively. Typical improvements to these M&S include new sub-models to address physical phenomena of new and emerging threat systems, as well as improvements to existing sub-models by incorporating new subsystem-level test data.

The DoD is also continuously improving the verification, validation, and uncertainty quantification (VVUQ) that is applied to this form of M&S. VVUQ are three interdependent tasks that are crucial toward ensuring the reliability and accuracy of M&S, and are defined as follows (Ghanem, Higdon, and Owhadi 2017):

- *Verification* is the process of determining that a model or simulation implementation and its associated output accurately represent the developer's conceptual description and specifications.
- *Validation* is the process of determining the degree to which a model or simulation and its associated outputs are an accurate representation of the real world from the perspective of the intended uses of the model.
- *Uncertainty Quantification* is the rigorous analysis and measurement of variations in mathematical models and simulations, aiding in understanding the consistency of the predictions with the phenomena they represent, and decision-making in complex systems.

It is well known that the output of AJEM and COVART depends on inputs that are uncertain, and is based on sub-models that are imperfect representations of reality. When the resulting uncertainty in the system-level QOI can be effectively quantified through VVUQ, M&S significantly enhances its value to decision makers involved in the acquisition process (Ghanem, Higdon, and Owhadi 2017).

Two common analyses within VVUQ are sensitivity and uncertainty analysis. Sensitivity analysis examines the impact of variations in the inputs of M&S on its corresponding output. This proves useful for identifying inputs, or interactions between inputs, that exert the most influence on the output. By doing so, researchers enhance their understanding of the M&S, detect anomalous M&S behaviors, and prioritize areas necessitating refinement.

On the other hand, an uncertainty analysis, which is the central focus of this case study, differs from sensitivity analysis in the way it handles M&S inputs. In an uncertainty analysis, inputs are treated as uncertain and represented as random variables, associated with one or more probability distributions that are subsequently propagated through the M&S to yield system-level output. The primary focus of an uncertainty analysis lies in the distribution of the output, shedding light on the variability and robustness of the M&S under different input conditions.

An uncertainty analysis applied to AJEM/COVART presents signficiant challenges. AJEM/COVART comprises extensive libraries with hundreds of empirical sub-models, each featuring diverse formulations and varying levels of complexity. Some sub-models draw from ample test data, while others rely on minimal or no data. Furthermore, while certain sub-models accommodate uncertainty quantification using statistical or probabilistic methods, others produce deterministic outputs through simple look-up tables, even when their outputs are recognized as uncertain.

This case study aims to demonstrate a comprehensive uncertainty analysis using a toy example. That is, we develop a simplified form of the M&S, using sub-models that we statistically formulate and fit to notional data. The simplified M&S retains the same number of hierarchical layers as the original, but the number of sub-models and the complexity of each sub-model have been substantially reduced. The purpose of the uncertainty analysis is to propagate forward the parameter and input uncertainties that accumulate in the hierarchy of sub-models to the system-level QOI. We are unaware of any past research that has had this goal.

The article is organized as follows. Section 2 details the methods, which starts with an overview of the case study, and then addresses two subsections, detailing the methods used to develop the simplified M&S and to propagate uncertainty. Section 3 presents the results of the propagation of uncertainty, and Section 4 provides a discussion.

2. Methods

The simplified M&S considers a ground combat vehicle being attacked by indirect artillery, drawing inspiration from a similar textbook example (Driels

2004). This scenario starts with the detonation of an artillery warhead at a specific position and orientation above the vehicle. As the warhead detonates, it disperses fragments in all directions, commonly known as shrapnel. The simplified M&S tracks a single fragment as it breaches the vehicle. The mass and velocity of the fragment in this example depend on the tilt angle of the warhead when it detonates. If the fragment penetrates the vehicle's turret roof armor, it has the potential to damage three components essential to the vehicle's mobility. The output of the simplified M&S is the probability of degraded mobility, while the input is the warhead tilt angle. An illustration of the attack scenario appears in Figure 1.

Deitz proposed a framework for decomposing the elements of a vulnerability problem into "levels" of testable characteristics (Deitz et al. 2009). Level one addresses the initial conditions of the engagement, which typically include the engagement geometry, location of impact, and settings relating to the weapon's potential and kinetic energy. Here, the initial conditions are the mass and velocity of the fragment that impacts the turret roof armor, which depends on the orientation of the warhead as it detonates above the vehicle. The first of three experiments that we use to develop the simplified M&S is called the Warhead Airburst Experiment. Its objective is to study how the mass and velocity of fragments dispersed by the warhead, which subsequently impact the turret roof armor, vary with changes in the warhead's tilt angle.

Level two describes the state of components after the attack, corresponding in this case study to the probability that component one, two, or three is damaged. Obtaining this result requires two separate experiments. The first, called the Armor Coupon Experiment, characterizes the probability that the fragment will penetrate the turret roof armor, in addition to the velocity and mass of the fragment as it exits through the armor and into the interior of the vehicle. The second, called the Component Damage Experiment, assesses the probability that the vehicle's critical components are damaged due to fragment impact.

Level three evaluates the platform's functionality following the engagement. Here, we focus on the vehicle's mobility, and fault trees are employed to correlate the status of individual components with the platform's overall state of mobility, which is viewed as binary: mission-capable, or non-mission-capable and in need of repair. The output of the fault tree sub-model is the system-level QOI, the probability of degraded mobility.

As a whole, the simplified M&S is composed of seven independent sub-models. Six are statistical regression models independently fit to data obtained from one of three experiments. The seventh is a deterministic fault tree model not derived from test data. The M&S hierarchically arranges the seven sub-models, such that the outputs of sub-models lower in the hierarchy serve as inputs to sub-models higher in the hierarchy, as shown in Figure 2. The primary input to the simplified M&S is the warhead tilt angle, while the system-level output, also known as the QOI, is the probability of degraded mobility. The following subsections detail the methods used to develop the sub-models and to propagate uncertainty.

2.1. Methods used to develop the simplified M&S

We formulate sub-models *I* through *VI* as individual Bayesian regression models. These sub-models are

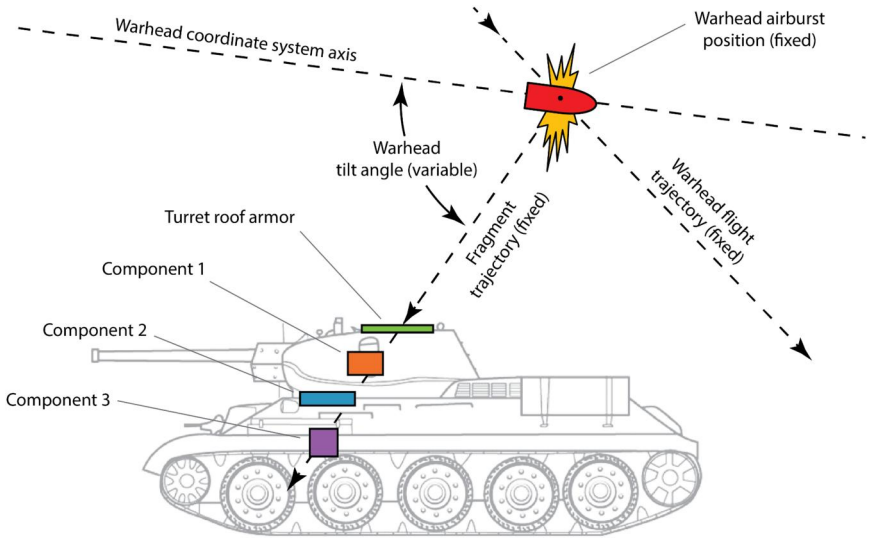

Figure 1. The engagement scenario modeled by the simplified M&S.

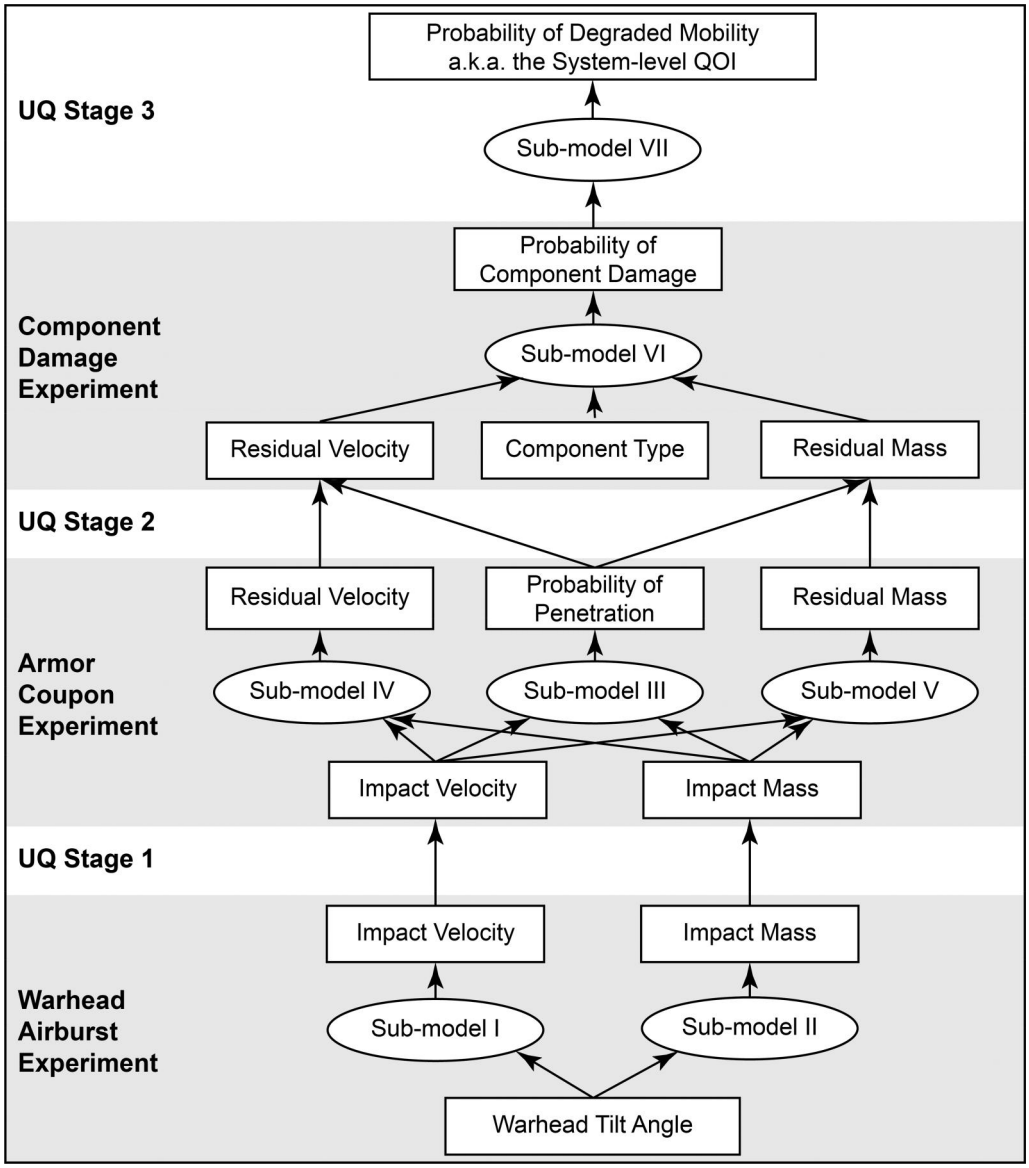

Figure 2. Hierarchical structure of simplified M&S.

fitted to data that we notionally generate, as if gathered from three real-life experiments. The subsequent subsections elaborate on these data and sub-models in more depth.

2.1.1. Formulation of sub-models I and II

Sub-models *I* and *II* are fit to notional data obtained from an envisioned Warhead Airburst Experiment, which is a standard DoD test for characterizing the fragments dispersed by an exploded warhead (JTCG/ME 2019). The experiment is conducted at an outdoor test range, where the warhead is detonated a few feet above the ground in a fixed horizontal position. Test instrumentation typically includes high speed cameras and flash screens to record fragment velocity, and an array of witness panels to retrieve and weigh fragments. The notional data encompasses a single warhead detonation, and consists of thousands of records, each of which provides the mass, velocity, and trajectory angle of an individual fragment.

The input (a.k.a. explanatory variable, independent variable, covariate, or factor) to sub-models *I* and *II* is trajectory angle. Trajectory angle is measured with respect to the warhead's two-dimensional coordinate system, with zero degrees corresponding to a fragment dispersed from the front of the warhead, 90 degrees corresponding to the side, and 180 degrees corresponding to the rear. Thirteen witness panels form a semi-circle around the warhead. Each panel corresponds to a trajectory angle, which is treated as a numeric categorical factor (panel 1 $\Rightarrow$ 0 deg, panel 2 $\Rightarrow$ 15 deg, panel 3 $\Rightarrow$ 30 deg, …., panel 13 $\Rightarrow$

180 deg). After the warhead detonates, engineers retrieve and weigh each fragment from each witness panel, analyze high speed camera footage to compute each fragment's velocity, and record a trajectory angle category for each fragment. The trajectory angle serves as the input for both sub-models *I* and *II*; however, the two sub-models produce distinct outputs that are modeled separately.

The output (a.k.a. the response variable or dependent variable) of sub-model *I* is the velocity of a single fragment that is ejected from the exploded warhead. Let y_{ij}^{I} denote the velocity of the *j*th fragment from the *i*th witness panel, where the Roman numeral superscript denotes that this response variable belongs to sub-model *I*. Let x_i denote the trajectory angle $(x_1 = 0\,\text{deg}, x_2 = 15\,\text{deg}, x_3 = 30\,\text{deg}, \ldots, x_{13} = 180\,\text{deg})$ of the fragments retrieved from the *i*th witness panel, which is treated as categorical in the data collection process, but is modeled as continuous in sub-models *I* and *II*. The number of fragments collected per bin is on the order of 100, and is denoted as n_i^I. We spatially model velocity as a function of trajectory angle, thus sub-model *I* takes the form of a Bayesian Gaussian process model, as shown in Equation 1.

$$\begin{aligned} y_{ij}^{I} &\sim \text{normal}(f_i^I, \sigma^I) \quad \forall i \in \{1,2,\ldots,13\} \text{ and } j \in \{1,2,\ldots,n_i^I\} \\ f^I &\sim \text{multivariate normal}\big(0, K(x|\alpha^I, \rho^I)\big) \\ \alpha^I &\sim \text{half normal}(0,.05) \\ \rho^I &\sim \text{inv gamma}(.05,.05) \\ \sigma^I &\sim \text{half normal}(0,1) \end{aligned} \tag{1}$$

Sub-model *I* assumes the velocity is normally distributed with an unknown standard deviation (σ^I), and an unknown mean (f^I) modeled as a Gaussian process (e.g., Gramacy 2020). The model uses the exponentiated quadratic function (K) to compute the covariance matrix of f^I. We code the response variable and tilt angle factor onto a zero-to-one scale, and choose weakly informative priors for the noise term (σ^I), the marginal standard deviation (α^I), and the length-scale parameter (ρ^I).

The output of sub-model *II* is the mass of a single fragment that is ejected from the exploded warhead. Let y_{ij}^{II} denote the mass of the *j*th fragment from the *i*th witness panel, and x_i is the trajectory angle of the fragments retreived from the *i*th witness panel. We spatially model fragment mass as a function of trajectory angle, thus sub-model *II* takes the form of a spatial Weibull model, as shown in Equation 2.

$$\begin{aligned} y_{ij}^{II} &\sim \text{Weibull}(\kappa^{II}, f_i^{II}) \quad \forall i \in \{1,\ldots,13\} \text{ and } j \in \{1,2,\ldots,n_i^{II}\} \\ \log\big(f^{II}\big) &= L^{II}\eta^{II} \\ L^{II} &= \text{cholesky decompose}\big(K(x|\alpha^{II}, \rho^{II})\big) \\ \kappa^{II} &\sim \text{half normal}(0,1) \\ \eta_i^{II} &\sim \text{normal}(0,1) \quad \forall i \in \{1,\ldots,13\} \\ \alpha^{II} &\sim \text{half normal}(0,1) \\ \rho^{II} &\sim \text{lognormal}(0,.2) \end{aligned} \tag{2}$$

Sub-model *II* assumes the fragment mass is distributed according to the Weibull distribution, with unknown scale parameter (f^{II}) modeled as a Gaussian process and shape parameter κ^{II} with prior distribution given by a standard half normal distribution.

Unlike sub-model *I*, the Gaussian process in sub-model *II* uses the latent variable formulation that multiplies the Cholesky factor of the covariance matrix by a vector of univariate normals (η^{II}), which implies that f^{II} is distributed as a multivariate normal random variable. Another difference is that a log link function is used to bound f^{II} between zero and infinity to satisfy the domain of f^{II} in accordance with the Weibull distribution. Similar to sub-model *I*, the exponentiated quadratic function (K) is used to compute the covariance matrix of f^{II}. Weakly informative priors are assigned to α^{II}, ρ^{II}, and η^{II}.

The simplified M&S assumes that the measured fragment trajectory angle is equivalent to the warhead tilt angle, as depicted in the engagement scenario. Furthermore, the simplified M&S assumes that the measured fragment's velocity and mass are equivalent to the velocity and mass of the fragment that impacts the turret roof armor, as depicted in the engagement scenario. Henceforth, we will refer to the input to sub-models *I* and *II* as warhead tilt angle, and we will refer to the outputs of sub-models *I* and *II* as impact velocity and impact mass respectively.

2.1.2. Formulation of sub-models III, IV, and V

Sub-models *III*, *IV*, and *V* are fit to notional data obtained from an envisioned armor coupon experiment, which is a standard type of DoD test for assessing the ability of an armor coupon to resist penetration from fragmentation (Military Standard 1997; Johnson et al. 2014). The Armor Coupon experiment is typically conducted at an indoor range, where a specialized gun is used to fire an individual manufactured fragment at an armor coupon that is fixed to a test stand. The tester can adjust the amount of charge that is loaded into the gun to precisely control the fragment's velocity. The fired fragments are

manufactured to be representative of real-world shrapnel, and are available in a limited number of different masses. The notional data encompasses a test series involving 120 trials, where each trial involves a single fragment being fired at an armor coupon that is representative of the turret roof armor.

The inputs to sub-models *III*, *IV*, and *V* that are controlled in the armor coupon experiment are the impact velocity and impact mass of the fragment that impacts the armor coupon. The designed experiment specifies the fragment impact mass as 1, 5, or 20 grains. Forty trials are executed for each impact mass, and the velocities of those 40 trials are determined using a sequential procedure used in the DoD, called 3Pod (Wu and Tian 2014). 3Pod is "sequential" in that it specifies the impact velocity of the next trial based on whether the previous trial resulted in a penetration or not. The inputs to sub-models *III*, *IV*, and *V* are identical, but their outputs are different and are modeled separately.

The output of sub-model *III* is a binary variable, y_k^{III}, indicating whether the fragment from the kth trial penetrated the armor, where $k = 1, 2, ..., n^{III}$; and n^{III} is the number of trials in the armor coupon experiment, which is equal to 120. Let $y_k^{III} = 1$ if the fragment penetrated, and $y_k^{III} = 0$ if it didn't. The independent variables that are controlled in the experiment are impact velocity (v_k) and impact mass (m_k). The linear predictor (μ_k^{III}) includes an intercept parameter (β_A^{III}), two main effects parameters (β_B^{III} and β_C^{III}), and an interaction parameter (β_D^{III}). The priors for the parameters are chosen for their lack of informativeness. Sub-model *III* takes the form of a logistic regression model, which appears in Equation 3.

$$\begin{aligned} y_k^{III} &\sim \text{Bernoulli}(\mu_k^{III}) \quad \forall k \in \{1, 2, ..., n^{III}\} \\ \text{logit}(\mu_k^{III}) &= \beta_A^{III} + \beta_B^{III} m_k + \beta_C^{III} v_k + \beta_D^{III} m_k v_k \\ \beta_A^{III}, \beta_B^{III}, \beta_C^{III}, \beta_D^{III} &\sim \text{normal}(0, 10) \end{aligned} \tag{3}$$

The sub-model IV output, known as residual velocity, represents the velocity of the fragment immediately after penetrating the armor, but this value is only recorded if the fragment penetrates the armor coupon. Let y_l^{IV} denote the residual velocity of the lth fragment that penetrates the armor coupon, where $l = 1, 2, ..., n^{IV}$. The linear model includes an intercept parameter (β_A^{IV}), two main effects parameters (β_B^{IV} and β_C^{IV}), and an interaction parameter (β_D^{IV}), where the priors are weakly informative. Sub-model *IV* takes the form of the linear model that appears in Equation 4.

$$\begin{aligned} y_l^{IV} &\sim \text{normal}(\mu_l^{IV}, \sigma^{IV}) \quad \forall l \in \{1, 2, ..., n^{IV}\} \\ \mu_l^{IV} &= \beta_A^{IV} + \beta_B^{IV} m_l + \beta_C^{IV} v_l + \beta_D^{IV} m_l v_l \\ \beta_A^{IV}, \beta_B^{IV}, \beta_C^{IV}, \beta_D^{IV} &\sim \text{normal}(0, 10) \\ \sigma^{IV} &\sim \text{normal}(0, 10) \end{aligned} \tag{4}$$

Similar to sub-model *IV*, the output of sub-model *V* is observed only if the fragment penetrates the armor coupon. When penetration occurs, theory holds that the fragment mass that exits may increase or decrease based on the impact mechanisms involved (Zukas et al. 1992). Here, we assume the mass increases due to plug formation. Plug formation is a mechanism that causes a chunk or "plug" of coupon material to exit with the fragment, causing an increase in its mass.

Accordingly, the output of sub-model *V* is the difference between the impact mass and residual mass, which is defined as $y_l^V = u_l - m_l$, where u_l and m_l denote the residual mass and impact mass, respectively, of the lth fragment that penetrates the armor, where $l = 1, 2, ..., n^V$; and n^V is the number of fragments that penetrate the armor. The independent variables are impact velocity (v_l) and impact mass (m_l). The linear predictor (μ_l^V) includes an intercept coefficient (β_A^V), two main effects coefficients (β_B^V and β_C^V), and an interaction coefficient (β_D^V), where the priors are weakly informative. The output is continuous, greater than zero, and skewed positive, thus sub-model *V* takes the form of the exponential regression model with log link that appears in Equation 5.

$$\begin{aligned} y_l^V &\sim \text{exponential}(\mu_l^V) \quad \forall l \in \{1, 2, ..., n^V\} \\ \log(\mu_l^V) &= \beta_A^V + \beta_B^V m_l + \beta_C^V v_l + \beta_D^V m_l v_l \\ \beta_A^V, \beta_B^V, \beta_C^V, \beta_D^V &\sim \text{normal}(0, 1) \end{aligned} \tag{5}$$

2.1.3. Formulation of sub-model VI

Sub-model *VI* is fit to notional data obtained from a component damage experiment, which is a test for assessing the fragility of individual components. The component damage experiment has a similar setup and conduct as the armor coupon experiment. The difference is that a vehicle component (such as an engine, gearbox, or radiator) is attached to the test stand instead of an armor coupon. After each trial, engineers inspect the component and determine

whether it was damaged (lost functionality) or undamaged (retained functionality).

The inputs to sub-model *VI* that are controlled in the component damage experiment are fragment impact mass, fragment impact velocity, and component type. Fragment impact mass is controlled as either 1, 5, or 20 grains, while component type specifies the component under test, which we generically refer to as component 1, 2, or 3. Fragment impact velocity is controlled using 3Pod. Forty trials are executed per unique combination of component type and fragment impact mass, resulting in 120 trials per component type, and 360 trials in total.

The output of sub-model *VI*, y_q^{VI}, indicates whether the qth trial resulted in a damaged component. Let $y_q^{VI} = 1$ if the component is damaged (loses functionality), and $y_q^{VI} = 0$ if it is not, where $q = 1, 2, ..., n^{VI}$; and n^{VI} is the number of trials in the experiment (equal to 360). The linear predictor (μ_q^{VI}) includes an intercept parameter (β_A^{VI}), two main effects paramaters (β_B^{VI} and β_C^{VI}), and an interaction parameter (β_D^{VI}). The linear predictor parameterizes the three types of components using two dummy paramaters (β_E^{VI} and β_F^{VI}) and two indicator variables (c_q^{VI} and d_q^{VI}), and the priors are weakly informative. Sub-model *VI* takes the form of the logistic regression model that appears in Equation 6.

$$\begin{aligned} y_q^{VI} &\sim \text{Bernoulli}(\mu_q^{VI}) \quad \forall q \in \{1, 2, ..., n^{VI}\} \\ \text{logit}(\mu_q^{VI}) &= \beta_A^{VI} + \beta_B^{VI} m_q^R + \beta_C^{VI} v_q^R + \beta_A^{VI} m_q^R v_q^R + \\ &\quad \beta_E^{VI} c_q^{VI} + \beta_F^{VI} d_q^{VI} \\ \beta_A^{VI}, \beta_B^{VI}, \beta_C^{VI}, \beta_D^{VI}, \beta_E^{VI}, \beta_F^{VI} &\sim \text{normal}(0, 10) \end{aligned} \tag{6}$$

From the perspective of the engagement scenario, the fragment impact velocity and fragment impact mass that are controlled in this experiment are equivalent to the residual velocity (the output of sub-model *IV*) and residual mass (the output of sub-model *V*). Furthermore, the simplified M&S assumes that the fragment's velocity and mass are unchanged as it passes through each component in sequence, without any changes or interactions affecting its velocity or mass.

2.1.4. Formulation of sub-model *VII*

Sub-model *VII* differs from all of the previous sub-models in that it is deterministic and has no unknown parameters that are estimated using experimental data. We formulate sub-model VII using a fault tree analysis that mirrors a textbook example (Driels 2004). A fault tree analysis is a deductive procedure, based on the engineering design of the vehicle, used to determine the various combinations of component failures that could cause a system-level failure, such as a lack of mobility.

The inputs to this fault tree model are the predicted outputs of sub-model VI: the probability of damage of components 1, 2, and 3, which are denoted as $p(\mathcal{C}_1)$, $p(\mathcal{C}_2)$, and $p(\mathcal{C}_3)$, respectively. We assume the three components are critical and non-redundant toward the state of the vehicle's mobility, which implies that the fault tree connects the three components in series, one after another.

The output of sub-model *VII* is the probability of degraded mobility. As in the textbook example, the output is formulated as a function of the inputs, using the survivor rule function, $f(\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3)$, which computes the probability of degraded mobility as shown in Equation 7.

$$f(\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3) = 1 - \prod_{s=1}^{3} (1 - p(\mathcal{C}_s)) \tag{7}$$

2.2. Methods used to propagate uncertainty

The case study's objective is to quantify the uncertainty in the system-level QOI by propagating forward the uncertainties accumulated throughout the hierarchy of sub-models. This accumulation includes parameter uncertainty and input uncertainty. The present subsection defines these uncertainties, and describes the methods used to propagate them forward to the system-level QOI.

Parameter uncertainty is quantified in the model fitting process. Sub-models *I* through *VI* individually take the form of a probabilistic model that relates the data collected in the experiment to the uncertain parameters in the sub-model. Each sub-model is independently fit using Bayesian modeling software, which employs Bayes' theorem to combine the prior probability distribution of the parameters with the probabilistic model of the experimental data to obtain the joint posterior probability distribution of the parameters. This posterior describes the uncertainty in the sub-model's parameters.

After the fitting process is complete, a sub-model can be used to make predictions corresponding to new, unobserved input values, while accounting for parameter uncertainty. In this case study, the predicted output of each sub-model is the posterior predictive distribution (PPD). The PPD is the distribution of possible future sub-model outputs corresponding to new, unobserved inputs values (i.e., interpolation or extrapolation). PPD computation is

standard practice in Bayesian regression modeling (e.g., Gelman et al. 2013). Standard PPD computation accounts for parameter uncertainty but not input uncertainty, because the new, unobserved inputs are typically treated as known values that are fixed as constants in the computation.

We account for input uncertainty using the forward propagation of input uncertainty. This VVUQ technique treats inputs as uncertain, assigns them corresponding distributions, and then propagates them to the output of the sub-model (e.g., Ghanem, Higdon, and Owhadi 2017). We apply this technique by modifying the standard PPD computation. Instead of treating the new, unobserved inputs as fixed, the modified PPD computation treats the inputs as uncertain. We developed the following steps to compute the modified PPD:

1. Formulate the model.
2. Collect the data.
3. Apply Bayes' theorem to obtain the parameter's posterior distribution.
4. Assign a distribution to each of the new, unobserved inputs.
5. Draw a single sample of the new, unobserved inputs from the distribution assigned in the previous step.
6. Draw a single sample of parameters from the parameter's posterior distribution.
7. Draw a predictive sample from the model likelihood by plugging in the inputs and parameters drawn in the previous two steps.
8. Repeat Steps 5–7 many times to obtain a distribution of predictive samples, which is the modified PPD.

The modified PPD computation accounts for both input and parameter uncertainty. The M&S's hierarchical structure specifies that the outputs of lower-level sub-models serve as inputs to higher-level sub-models for the purpose of prediction, as earlier summarized in Figure 2. More specifically, in Step 4 of the modified PPD computation, it is the computed PPD of the lower-level sub-model that is assigned to the new, unobserved input of the higher-level sub-model for the purpose of prediction. The computed modified PPD accounts for input and parameter uncertainty of the present sub-model, as well as the input and parameter uncertainties of the sub-models that come before it in the hierarchy.

Quantifying uncertainty in the system-level QOI begins by independently fitting each sub-model to its corresponding experimental data. Then, we propagate uncertainty from one layer of the hierarchy to the next using the following stages.

- *Stage 0* computes an independent standard PPD as the predictive output for each of sub-models *I* and *II*, for known values of warhead tilt angle ranging between 0 and 180 degrees in small increments.
- *Stage 1* computes an independent modified PPD as the predictive output for each of sub-models *III*, *IV*, and *V*. The assigned input uncertainty distributions are the computed results from Stage 0.
- *Stage 2* computes an independent modified PPD as the predictive output for sub-model *VI*. The assigned input uncertainty distributions are the computed results from Stage 1.
- *Stage 3* propagates the input uncertainties to the output of sub-model *VII*, which yields the uncertainty in the system-level QOI. This stage does not account for parameter uncertainty, because sub-model VII is deterministic and is not fit to test data. The assigned input uncertainty distribution is the computed results from Stage 2.

The stages can be expressed analytically. For example, as part of Stage 1, consider the computation of the modified PPD for sub-model *III*. Let $\tilde{y}^I$ and $\tilde{y}^{II}$ denote the predicted response at the unobserved warhead tilt angle $(\tilde{x})$. And let $\tilde{y}^{III}$ denote the predicted response at the unobserved impact velocity $(\tilde{v})$ and unobserved impact mass $(\tilde{m})$. Additionally, let θ^I, θ^{II}, and θ^{III} denote the vector of unknown parameters for sub-models *I*, *II*, and *III*, respectively. Using Gelman's notation, the standard PPD for $\tilde{y}^I$, $\tilde{y}^{II}$, and $\tilde{y}^{III}$ can be expressed as

$$p(\tilde{y}^I|y^I,\tilde{x}) = \int p(\tilde{y}^I|\theta^I,\tilde{x})p(\theta^I|y^I)d\theta^I, \tag{8}$$

$$p(\tilde{y}^{II}|y^{II},\tilde{x}) = \int p(\tilde{y}^{II}|\theta^{II},\tilde{x})p(\theta^{II}|y^{II})d\theta^{II}, \tag{9}$$

and

$$p(\tilde{y}^{III}=1|y^{III},\tilde{v},\tilde{m}) = \int p(\tilde{y}^{III}=1|\theta^{III},\tilde{v},\tilde{m})p(\theta^{III}|y^{III})d\theta^{III}. \tag{10}$$

Then, to propagate uncertainty from sub-models *I* and *II* to sub-model *III*, we independently set the unobserved inputs ($\tilde{v}$ and $\tilde{m}$) within Equation 10 equal to Equations 8 and 9, respectively. That is,

$$p(\tilde{v}|y^I,\tilde{x}) \equiv p(\tilde{y}^I|y^I,\tilde{x}), \text{and } p(\tilde{m}|y^{II},\tilde{x}) \equiv p(\tilde{y}^{II}|y^{II},\tilde{x}). \tag{11}$$

Consequently, we obtain the modified PPD for sub-model *III*, which is expressed as

$$p(\tilde{y}^{III} = 1|y^{III}, y^{II}, y^{I}, \tilde{x}) = \iiint p(\tilde{y}^{III} = 1|\theta^{III}, \tilde{m}, \tilde{v}) p(\theta^{III}|^{III}) p(\tilde{m}|y^{II}, \tilde{x}) p(\tilde{v}|y^{I}, \tilde{x}) d\tilde{v} d\tilde{m} d\theta^{III} \quad (12)$$

3. Results

The results are organized by the different stages of uncertainty propagation, starting with Stage 0. After data from the Warhead Airburst Experiment is collected, sub-models *I* and *II* are individually fit to this data using Bayesian modeling software, enabling the Stage 0 computation of the standard PPD for each of sub-models *I* and *II* for values of warhead tilt angle, ranging from 0 to 180 degrees in small increments, as shown in Figure 3. This result includes parameter uncertainty, but not input uncertainty because warhead tilt angle is treated as a known input to the simplified M&S.

After data from the Armor Coupon Experiment is collected, sub-models *III*, *IV*, and *V* are each independently fit to this data using Bayesian modeling software. Then, according to Stage 1, the modified PPD is computed for sub-models *III*, *IV*, and *V*, accounting for input uncertainty and the uncertainty in each sub-model's parameters, as shown on the left side of Figure 4. The input uncertainty is represented as a function of warhead tilt angle in the Stage 1 computation. Consequently, the output of sub-models *III*, *IV*, and *V* are expressed as a function of warhead tilt angle. Instead of propagating the entire input uncertainty distribution, the right side of Figure 4 propagates the mean of this distribution, depicting the effects of input uncertainty propagation.

Once data from the component damage experiment is collected, sub-model *VI* is fit to this data using Bayesian modeling software. Then, according to Stage 2, the modified PPD is computed for sub-model *VI*, accounting for input uncertainty and the uncertainty in the sub-model's parameters, as shown on the left side of Figure 5. Again, the input uncertainty (the Stage 1 result) is represented as a function of warhead tilt angle in the Stage 2 computation. Consequently, the output of sub-model *VI* is expressed as a function of warhead tilt angle. The result on the left side of Figure 5 includes parameter uncertainty from sub-models *I* through *VI*, and input uncertainty from sub-models *III* through *VI*. The right side of Figure 5 propagates the mean of the input uncertainty distributions from sub-models *III* through *VI*, illustrating the effects of input uncertainty propagation.

The third and final stage propagates the uncertainty accumulated in the second stage through sub-model *VII* (the fault tree model) to obtain the uncertainty in the system-level QOI, probability of degraded mobility. Sub-model *VII* has no parameter uncertainty, because it is deterministic and includes no uncertain parameters. The third stage result that appears on the left side of Figure 6 is the primary result of the case study, capturing parameter uncertainties from sub-

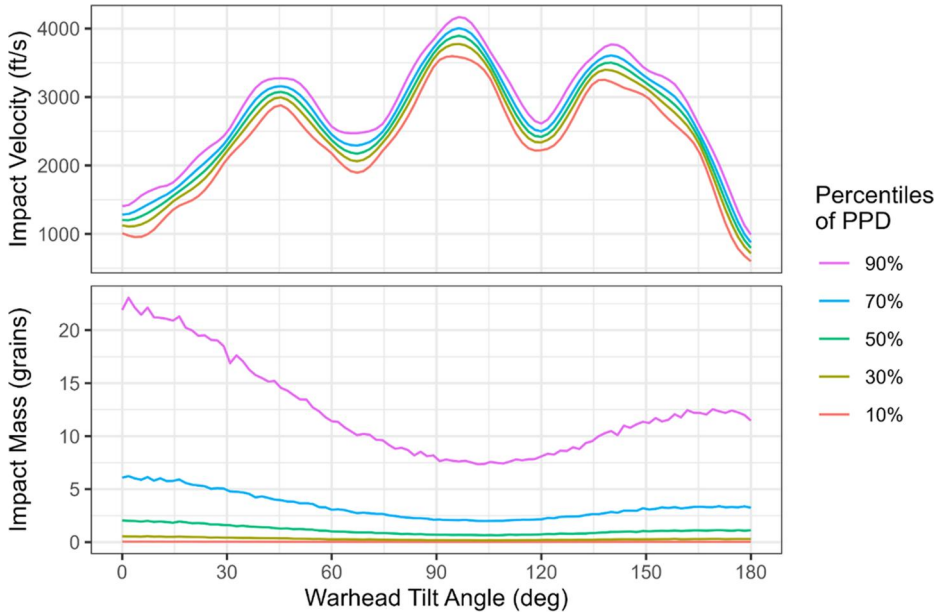

Figure 3. Stage 0 Results.

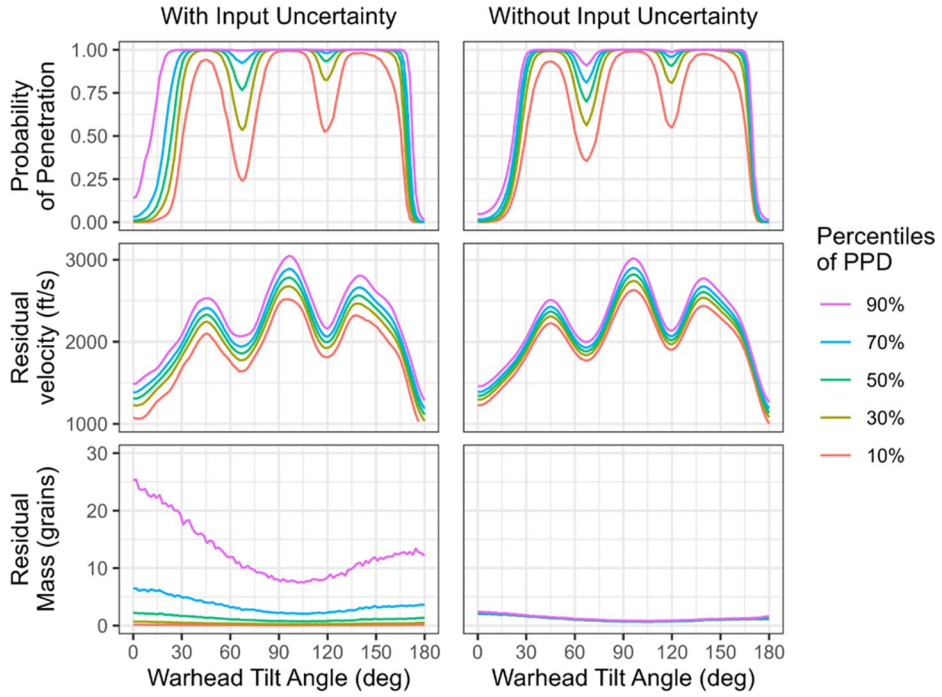

Figure 4. Stage 1 Results.

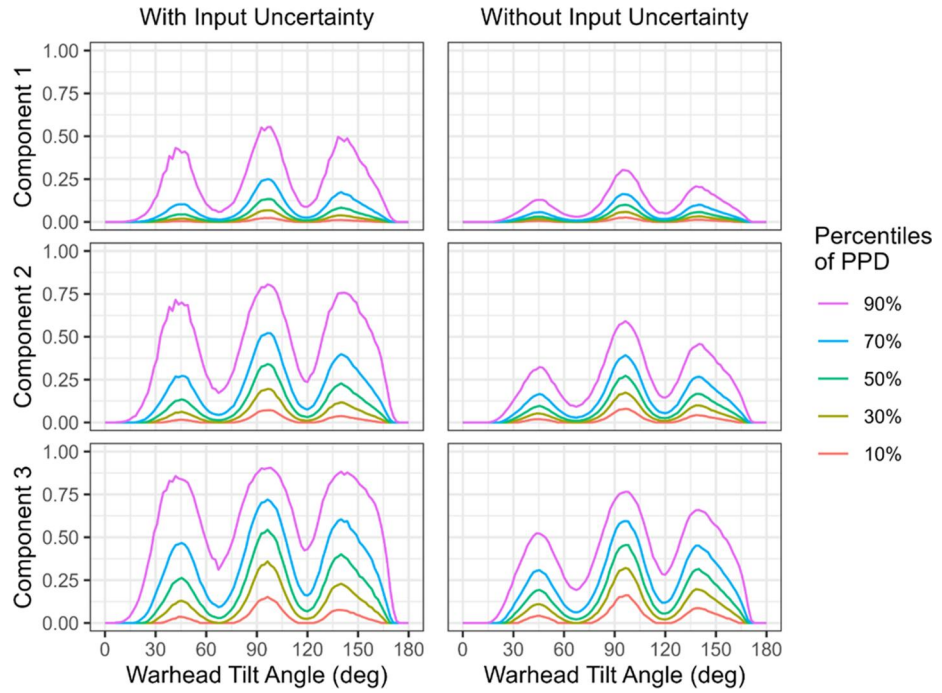

Figure 5. Stage 2 Results.

models *I* through *VI*, and input uncertainties from sub-models *III* through *VII*. The right side of Figure 6 propagates the mean of the input uncertainty distributions from sub-models *III* through *VII*, providing a clear visualization of the effects of input uncertainty propagation.

4. Discussion

The approaches described here help the Department of Defense stay at the forefront of the scientifically rigorous application of M&S to live fire test and evaluation. This topic is especially germane given the current technological and defense landscape, wherein

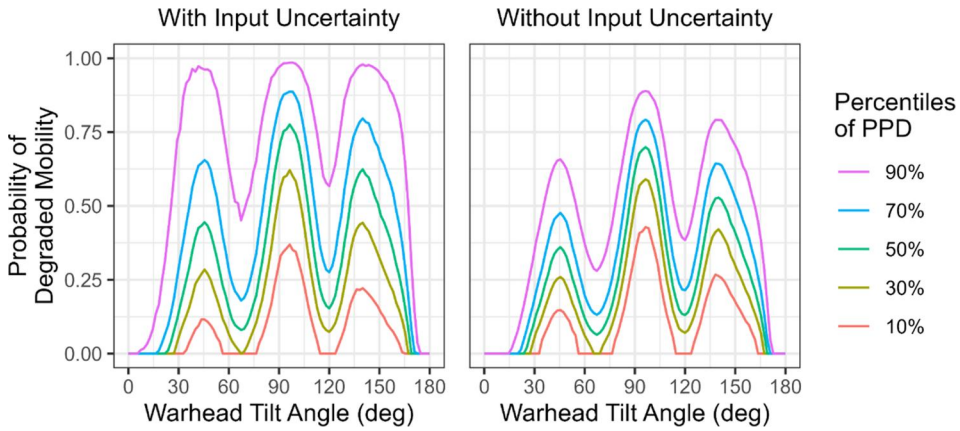

Figure 6. Stage 3 Results.

joint warfighting capabilities and their supporting materiel solutions, such as Combined Joint All-Domain Command and Control, have emerged as the principal paradigm for future deterrence and warfighting.

Emerging joint capabilities will stress the Department's current T&E capabilities. For instance, comprehensive live testing of future capabilities may be stressed due to environmental, fiscal, safety, classification, and ethical constraints. Therefore, the Department will become more dependent on M&S to evaluate the efficacy and interoperability of our systems. Additionally, the future joint warfighting paradigm requires evaluations to be focused on mission threads and kill webs, with a strong emphasis on understanding the mission effect.

To this end, the Department will need to use modern quantitative methods and computing technologies to collect and synthesize data from specific threat engagements to assess measures of performance, such as platform vulnerability. The Department will continue to rely on and improve upon empirically derived reduced order models, in addition to other types of M&S, including the integration of numerous M&S into a "digital arena" where complex joint warfighting scenarios can be interrogated in a comprehensive way.

4.1. Summary of results

The case study demonstrates an approach to quantify uncertainty in the output of a specific type of M&S, typically used in live fire vulnerability assessments. We formulated each of the six sub-models within the simplified M&S as an independent Bayesian regression model, and connected the sub-models in a hierarchy. We then quantified the uncertainties in each sub-model, and propagated them forward, through the hierarchy, to the system-level QOI.

Figure 6 presents the primary result of the case study. It shows on the vertical axis the uncertainty in the system-level QOI, probability of degraded mobility, for a range of values of the initial condition on the horizontal axis, warhead tilt angle. The uncertainty in the system-level QOI is expressed as percentiles of the PPD.

Warhead tilt angle was an initial condition of the attack scenario, as well as the input to the entire simplified M&S. It appeared at the lowest layer of the hierarchy (an input to sub-model *I* and *II*), and was by design the only input treated as known in the propagation of uncertainty. Every other sub-model input was treated as uncertain, and was assigned a distribution that originated as the PPD of the sub-model output that came before it. This hierarchical dependence enabled the result of each stage of the propagation of uncertainty to be expressed in terms of the initial condition, warhead tilt angle.

The uncertainty in the QOI that appears in Figure 6 was influenced at each stage of the propagation. The undulations in the uncertainty, cresting at tilt angles of 45, 100, and 140 degrees, are artifacts of the uncertainty derived from sub-model *I*. The system-level QOI's asymmetric credible interval is attributed to the skewed, positive PPDs that were output from sub-models *II* and *V*. The propagation of uncertainty through sub-model *VI* transformed the credible intervals onto a probability scale, ranging between zero and one. The cascading uncertainty grew and transformed as they were propagated from one layer of the hierarchy to the next.

The uncertainty in the system-level QOI stems from parameters and inputs. Parameter uncertainty was independently estimated for each sub-model (for sub-models *I* through *VI*) as a joint posterior distribution of the parameters, obtained in the model fitting process. Input uncertainty, on the other hand, was

assigned to each sub-model's input (for sub-models *III* through *VII*). Input uncertainty was defined as a distribution, equal to the PPD of the sub-model that came before it in the hierarchy. The effects of input uncertainty become clear by comparing the left and right sides of Figures 4, 5, and 6.

Included within parameter uncertainty are additional sources, including sampling uncertainty and model uncertainty. Sampling uncertainty captures the variability of the sampled data, while model uncertainty arises due to discrepancies between the formulated model and the true underlying process. Parameter uncertainty can be reduced by sampling more data, reducing variability in the data *via* an improved data collection process, or by formulating a better-fitting model, including the use of informative priors.

4.2. Value of primary result

The primary result presented in Figure 6 offers decision makers valuable insights into the uncertainties of the QOI predictions and the specific conditions influencing the predictions' accuracy and reliability. This heightened understanding empowers decision makers to make more informed and robust choices, taking potential risks and their probabilities into account.

To illustrate, consider a situation where a specified warhead tilt angle carries significant importance to an operational scenario, and has therefore been assigned a requirement: the probability of degraded mobility at the specified tilt angle must be less than a threshold value. In this context, the primary result of the case study becomes a powerful tool for conducting a probabilistic assessment. Through manipulation of the PPD shown in Figure 6, one can compute the probability of passing the requirement at the specified tilt angle.

The primary result also provides valuable insights for decision makers regarding resource allocation. For instance, when planning further testing and M&S development, one can create hypothetical data to assess its impact on uncertainty. Trade studies can then identify specific subsystem-level tests and conditions that offer the best reduction in system-level QOI uncertainty. These study results can effectively inform resource allocation decisions.

4.3. Comparison of UQ approach to past work in live fire community

Presently, we are unaware of any past work involving AJEM or COVART that attempted to quantify the uncertainty in the system-level QOI by accounting for all input and parameter uncertainties throughout the hierarchy of sub-models, as we applied to the simplified M&S in this case study. Indeed, such a comprehensive uncertainty analysis is a challenging task, given the diversity and complexity of sub-models in AJEM and COVART.

Past work has, however, conducted sensitivity analyses. COVART researchers conducted a series of separate sensitivity analyses to determine which sub-model inputs had the greatest influence on system-level QOIs (Leakeas et al. 2021; Staley 2022; Leakeas and Stewert 2023). This included a sensitivity analysis involving the geometric drawing of the target, a component probability of damage sensitivity analysis, and a threat characterization sensitivity analysis. In each study, the researchers developed and executed a designed experiment, consisting on the order of 100 COVART runs, in which the COVART inputs were systematically varied from run to run and the corresponding outputs were recorded. The primary finding from this work was that the probability of component damage due to fire was highly influential.

In a similar line of effort, AJEM researchers conducted sensitivity analyses (Andres 2020, Andres 2023) that varied the resolution of the CAD drawing, elements of the Behind Armor Debris model, and the modeled thickness of the shot-line. The main takeaway from this work was that perturbations to parameters of the Behind Armor Debris model had the largest effect on the system-level QOI, foreshadowing potential future sub-model improvements.

This case study is different than the COVART and AJEM sensitivity analyses because the goal here was to quantify uncertainty in the system-level QOI, using an objective, data-driven approach, capturing all parameter and input uncertainties throughout the simplified M&S. The intent was not to identify which inputs had the most influence on the outputs, as in the COVART and AJEM sensitivity analyses.

4.4. Comparison to other VVUQ techniques

VVUQ techniques are typically tailored to the type of M&S under consideration. AJEM/COVART are different than many M&S that are addressed in VVUQ literature, because AJEM/COVART are empirical models that are computationally inexpensive and non-black-box. For this reason, this case study employed certain VVUQ techniques and not others.

Calibration is a common VVUQ technique that we applied in this case study. The purpose of calibration

is to adjust a set of parameters associated with a model so that the agreement between model prediction and experimental observation is maximized. (Trucano et al. 2006). The type of calibration method employed often relies on the form of model being calibrated.

The simplified M&S comprised six independent statistical regression models. We used a standard method for fitting Bayesian regression models, called Markov Chain Monte Carlo (MCMC), to calibrate the set of parameters of each sub-model to the corresponding test data. Each regression model fit achieved convergence using 10 chains, with 5000 iterations per chain in the MCMC process. Regression model fitting is rarely referred to as calibration in VVUQ literature (e.g., Ghanem, Higdon, and Owhadi 2017), but it is a form of calibration nonetheless.

Another calibration method focuses on the statistical adjustment of engineering models (Joseph and Yan 2015). Engineering models include fast-running, non-numerical, closed-form expressions, which are often physics-based approximations of governing equations. These models are not black-box, and are not usually statistical regression models. This calibration method adjusts the engineering model using a fast, statistical approach, which preserves the original formulation of the engineering model. Even though this approach was not employed in this case study, it should be considered in future work because many sub-models within AJEM/COVART are engineering models.

VVUQ literature most often concentrates on the calibration of computationally expensive, black-box M&S, such as finite element models or hydro codes (e.g., Ghanem, Higdon, and Owhadi 2017). This type of M&S often has the capability to predict the QOI independent of test data. This contrasts with the simplified M&S, which was entirely dependent on test data to predict the QOI. Calibration of the computationally expensive M&S often involves development and execution of a designed experiment, which systematically varies the M&S parameters, M&S inputs, and test conditions. The resulting data set is then used to calibrate the M&S parameters to maximize the agreement between the M&S predictions and test data. This type of approach was not employed in this case study, because the M&S under consideration is not computationally expensive, not capable of predicting the QOI without test data, nor is it black-box.

Another common VVUQ technique is surrogate modeling. Again, this technique is often employed when the M&S is computationally expensive, and black-box. The surrogate model is fit to M&S QOI predictions. A surrogate model is fast-running and computationally inexpensive to run, and often takes the form of a flexible, non-linear spatial model, such as a Gaussian process model. Once a surrogate model is fit to the computationally expensive M&S predictions, it can be used to predict the QOI in a computationally inexpensive, fast-running manner. A surrogate model was not needed to serve as a surrogate of the entire simplified M&S because the simplified M&S was already fast-running and computationally inexpensive. Despite this fact, surrogate modeling did prove useful in this case study for individual sub-model formulation, as we demonstrated for sub-models *I* and *II*.

One last VVUQ technique is discrepancy modeling, which is a type of validation. A discrepancy model models the discrepancy in the QOI between the M&S prediction and experimental data. This technique is often used when the M&S is black-box, and when the M&S produces a QOI prediction that is independent of test data. After the discrepancy is modeled, it is often added onto the M&S prediction. This correction, which is sometimes referred to as combined prediction, is intended to produce a more accurate prediction of the QOI than the uncorrected M&S prediction. Combined prediction and calibration are the two primary techniques for correcting M&S using test data. We opted not to use discrepancy modeling and combined prediction in this case study, because calibration was more straightforward, given that the sub-models were formulated as statistical regression models.

4.5. Limitations

The simplified M&S had many limitations relative to the actual M&S. For starters, the simplified M&S is composed of only six statistical regression models and one deterministic fault tree model. The actual M&S includes libraries of hundreds of sub-models, such as the Fast Air Target Penetration or Projectile Penetration libraries, capable of simulating a vast array of threat-target engagement scenarios. These sub-models are empirical and are primarily formulated as engineering equations, incorporating many additional factors not considered in the simplified M&S. Other sub-models in the actual M&S include look-up tables and statistical models, both stochastic and deterministic. Despite these differences, the similarities were substantial enough to envision the possibility of applying this work to the actual M&S, in one form or another. It is our hope that this case study will inspire future AJEM/COVART uncertainty analyses.

This case study did not account for uncertainty due to a misspecified hierarchical dependence. That is, there is likely uncertainty due to representing the attack scenario as an integration of a limited number of subsystem-level phenomena. The integration of these isolated phenomena likely fails to capture the entire physics of the complete system. It may be possible to better capture this uncertainty using an approach that conducts validation and calibration across more than one level of the hierarchy. The simplified M&S only used test data from a given level to calibrate the sub-model of that same level, but other research has demonstrated techniques to calibrate and validate multiple component-level sub-models that span more than one level, using system-level test data (Mullins et al. 2013; Li and Mahadevan 2016).

The approach in this case study could not easily be applied to AJEM/COVART in the near future, because some sub-models within AJEM/COVART do not accommodate uncertainty quantification. One reason for this is that some sub-models are based on little to no test data. Another reason is that some were developed as simple look-up tables and were not formulated using a statistical framework. Future work should aim to improve these sub-models.

4.6. Future work

Future work could identify sub-models that are in need of improvement. A study could prioritize sub-models that have two qualities: the greatest influence on the system-level QOI, and the greatest influence on the uncertainty in the system-level QOI. The first quality could be quantified using the results of the recently conducted sensitivity studies. Quantifying the second quality would require additional work, and may provide additional weight to sub-models that are not formulated statistically or are based on little to no test data.

Sub-models with the highest priority could be subject to future improvements. A comprehensive, high-cost investment could involve conducting additional testing to acquire new data, which could then be used to improve the sub-model. Another high-cost option would be to use the new data to reformulate the sub-model using engineering first principles.

An alternative low-cost investment could involve recalibrating an existing sub-model to existing data. A prime candidate for this endeavor is an existing sub-model that does not accommodate uncertainty quantification, produces a deterministic output, was originally formulated as a physics-based engineering equation, and was originally calibrated to test data that is still currently available. The recalibration could follow an established procedure (Joseph and Yan 2015), employing a fast, statistical approach to recalibrate the sub-model. As a result, the sub-model could accommodate uncertainty quantification, while maintaining its original engineering formulation.

Another low-cost investment could involve recalibrating *and* reformulating an existing sub-model to existing data. Some look-up tables could be reformulated as statistical regression models to enable uncertainty quantification. In cases where the look-up table is based on little data, a simple generalized linear model could be employed, with increased emphasis on support from subject matter experts to encode prior information into the model. In cases with ample data, a more flexible and capable regression model could be employed, such as a generalized additive model or surrogate model, which may improve uncertainty quantification as well as predictive accuracy.

About the authors

Thomas H. Johnson is a data scientist and engineer who specializes in test design, statistical modeling, and modeling and simulation validation. He has worked at the Institute for Defense Analyses for 13 years, supporting live fire and operational testing of armored vehicles and rotorcraft.

Dhruv Patel is a research staff member at the Institute for Defense Analyses in Alexandria, Virginia. At IDA, he contributes to multiple research teams, including the Test Science and the Net-centric groups. His primary research interests involve modeling and simulation, uncertainty quantification, and network theory. He earned a PhD in statistics and operations research from the University of North Carolina at Chapel Hill in 2023.

John Haman is a research staff member at the Institute for Defense Analyses in Alexandria, Virginia. At IDA, Dr. Haman leads the Test Science research group, a team of statisticians, psychologists, and neuroscientists that provides statistics and human factors expertise, methodologies, trainings, and guidance to its DOD sponsors. He earned a PhD in statistics from Bowling Green State University in 2018 with a thesis on skewed data analysis.

Dave Higdon is a professor in the Statistics Department at Virginia Tech. He is an expert in Bayesian statistical modeling of environmental and physical systems, combining physical observations with computer simulation models for prediction and inference. Dr. Higdon has served on several advisory groups concerned with statistical modeling and uncertainty quantification and co-chaired the NRC Committee on Mathematical Foundations of Validation, Verification, and Uncertainty Quantification.

Jeremy Werner was appointed DOT&E's Science Advisor in December 2021 after initially starting at DOT&E as an Action Officer for Naval Warfare in August 2021. Before then, Jeremy was at Johns Hopkins University

Applied Physics Laboratory (JHU/APL), where he founded a data science-oriented military operations research team that transformed the analytics of an ongoing military mission. Jeremy received a Ph.D. in physics from Princeton University where he was an integral contributor to the Compact Muon Solenoid collaboration in the experimental discovery of the Higgs boson at the Large Hadron Collider at CERN, the European Organization for Nuclear Research in Geneva, Switzerland.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Andres, C. D. 2020. *Uncertainty quantification and sensitivity analysis methodology for the Advanced Joint Effectiveness Model (AJEM)*. Technical Report CCDC DAC-TR-2020-044. Aberdeen, MD: U.S. Army CCDC Data & Analysis Center, Aberdeen Proving Ground.

Andres, C. D. 2023. Uncertainty quantification and sensitivity analysis for the Advanced Joint Effectiveness Model (AJEM) behind-armor debris characteristics. Aberdeen, MD: Army Operations Research Symposium, Aberdeen Proving Ground.

Deitz, P. H., H. L. Reed, J. T. Klopcic, and J. N. Walbert. 2009. *Fundamentals of ground combat system ballistic vulnerability/lethality*. Reston, VA: The American Institute of Aeronautics and Astronautics.

Driels, M. R. 2004. *Weaponeering: Conventional weapon system effectiveness*. Reston, VA: American Institute of Aeronautics and Astronautics.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian data analysis*. Boca Raton, FL: CRC press.

Ghanem, R., D. Higdon, and H. Owhadi. 2017. *Handbook of uncertainty quantification*, vol. 6. Cham, Switzerland: Springer.

Gramacy, R. B. 2020. *Surrogates: Gaussian process modeling, design and optimization for the applied sciences*. Boca Raton, FL: Chapman Hall/CRC. http://bobby.gramacy.com/surrogates/.

Johnson, T. H., L. Freeman, J. Hester, and J. L. Bell. 2014. A comparison of ballistic resistance testing techniques in the Department of Defense. *IEEE Access*.2:1442–55. doi:10.1109/ACCESS.2014.2377633.

Joseph, V. R., and H. Yan. 2015. Engineering-driven statistical adjustment and calibration. *Technometrics* 57 (2): 257–67. doi:10.1080/00401706.2014.902773.

Joint Technical Coordinating Group for Munitions Effectiveness. 2019. Test and Data Reduction Procedures for Munitions. 61 JTCG/ME 1-9, Rev 6.

Leakeas, C. L., A. Piedel, M. Swart, and R. Cairo, Jr. 2021. Penetration and vulnerability modeling and simulation confidence and sensitivity. Technical Report JASPO-V-18-02-001. AFLCMC/EZJA, Wright-Patterson Air Force Base, OH.

Leakeas, C. L., and R. K. Stewert. 2023. HE threat sensitivity studies final report. Technical Report JASPO-V-21-01. AFLCMC/EZJA, Wright-Patterson Air Force Base, OH.

Li, C., and S. Mahadevan. 2016. Role of calibration, validation, and relevance in multi-level uncertainty integration. *Reliability Engineering & System Safety* 148:32–43. doi:10.1016/j.ress.2015.11.013.

Military Standard. 1997. V50 Ballistic Test for Armor, MIL-STD-662F, US Army Research Laboratory, Adelphi, MD.

Mullins, J., C. Li, S. Sankararaman, S. Mahadevan, and A. Urbina. 2013. Uncertainty quantification using multi-level calibration and validation data. In *54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 1872. Reston, VA: American Institute for Aeronautics and Astronautics (AIAA).

Staley, T. 2022. COVART V&V. Technical Report Unpublished Briefing. AFLCMC/EZJA, Wright-Patterson Air Force Base, OH.

Trucano, T. G., L. P. Swiler, T. Igusa, W. L. Oberkampf, and M. Pilch. 2006. Calibration, validation, and sensitivity analysis: What's what. *Reliability Engineering & System Safety* 91 (10–11):1331–57. doi:10.1016/j.ress.2005.11.031.

Wu, C. F. J., and Y. Tian. 2014. Three-phase optimal design of sensitivity experiments. *Journal of Statistical Planning and Inference* 149:1–15. doi:10.1016/j.jspi.2013.10.007.

Zukas, J. A., T. Nicholas, H. F. Swift, L. B. Greszczuk, D. R. Curran, and L. E. Malvern. 1992. *Impact dynamics*. Malabar, FL: Krieger Publishing Company.