12-2019

Uncertainty and Error in Combat Modeling, Simulation, and Analysis

Jason A. Blake

UNCERTAINTY AND ERROR IN COMBAT MODELING, SIMULATION, AND ANALYSIS

DISSERTATION

Mr. Jason A. Blake, USAF
AFIT-ENS-DS-19-D-019

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-DS-19-D-019

UNCERTAINTY AND ERROR IN COMBAT MODELING, SIMULATION, AND ANALYSIS

DISSERTATION

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Mr. Jason A. Blake, USAF

AFIT-ENS-DS-19-D-019

19 December 2019

DISTRIBUTION STATEMENT A
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENS-DS-19-D-019

UNCERTAINTY AND ERROR IN COMBAT MODELING,

SIMULATION, AND ANALYSIS

DISSERTATION

Mr. Jason A. Blake, USAF

Committee Membership:

John O. Miller, PhD
Chairman

Jeffery D. Weir, PhD
Member

Douglas D. Hodson, PhD
Member

ADEDJI B. BADIRU, PhD

Dean, Graduate School of Engineering and Management

AFIT-ENS-DS-19-D-019

Abstract

Due to the infrequent and competitive nature of combat, several challenges present themselves when developing a predictive simulation. First, there is limited data with which to validate such analysis tools. Secondly, there are many aspects of combat modeling that are highly uncertain and not knowable. This research develops a comprehensive set of techniques for the treatment of uncertainty and error in combat modeling and simulation analysis.

First, Evidence Theory is demonstrated as a framework for representing epistemic uncertainty in combat modeling output. Next, a novel method for sensitivity analysis of uncertainty in Evidence Theory is developed. This sensitivity analysis method generates marginal cumulative plausibility functions (CPFs) and cumulative belief functions (CBFs) and prioritizes the contribution of each factor by the Wasserstein distance (also known as the Kantorovich or Earth Mover's distance) between the CBF and CPF. Using this method, a rank ordering of the simulation input factors can be produced with respect to uncertainty. Lastly, a procedure for prioritizing the impact of modeling choices on simulation output uncertainty in settings where multiple models are employed is developed. This analysis provides insight into the overall sensitivities of the system with respect to multiple modeling choices. The new method does not make weakly predictive models strongly predictive models, but ensures a plurality of perspectives can be reconciled during a modeling and simulation activity.

For my wife, without your love and encouragement this would not have been possible.

For my children, always pursue your dreams.

Acknowledgements

I'd like to express my sincerest thanks to my research advisor, Dr. J. O. Miller, for his support and encouragement throughout this process. His admonishment to keep things simple and his graciousness when things didn't go according to plan were key to maintaining my sanity, hopefully it did not cost him his. I'd also like to thank my committee members, Dr. Jeff Weir and Dr. Doug Hodson for their thoughtful feedback, especially during the initial phases of this work.

I also owe a debt of gratitude to the leadership of the Simulation and Analysis Facility and the Engineering Analysis Squadron for their support in this endeavor. Thank you for extending uncommon flexibility, I appreciate your trust.

Jason A. Blake

Contents

List of Figures

List of Tables

Page

UNCERTAINTY AND ERROR IN COMBAT MODELING, SIMULATION, AND ANALYSIS

I. Introduction

The purpose of this research is to explore the mechanisms by which uncertainties in combat should be incorporated in a comprehensive analysis via modeling and simulation. Existing approaches within the defense community for incorporating uncertain elements into simulation studies are ad hoc with a significant number of tools that do not facilitate straight forward exploration of system uncertainty. This problem is further compounded by the fact that there are multiple overlapping simulation toolsets which, having been individually developed by domain experts (aeronautics, signatures/sensing, communications, etc.), each have slightly different representations of entities and environmental factors. This research addresses the treatment of uncertainty and modeling error by leveraging Evidence Theory as a framework to combine multiple, potentially conflicting sources for simulation factor settings and represent the uncertainty these inputs induce on simulation output.

Modeling and simulation has been applied to a wide variety of challenging problems in research, commercial industry and government. While drawbacks include lengthy development time, software licensing fees and limited pools of qualified practitioners, it is particularly well suited to problems where 1) experimentation with the real system is prohibitively expensive (presumably more so than the modeling and simulation effort itself) and 2) there is no other way in which to reasonably conduct the desired experiments. In the Department of Defense (DoD), models of combat have been employed by systems and operations research analysts since the 1960's for exploring possible outcomes of hypothetical military conflict (Davis, 1995). The fidelity of these models has ranged from simple mathematical relationships describing attrition between two opposing forces (Lanchester, 1914) to high fidelity operator or hardware in the loop simulations for exploration of detailed system configuration changes (Haase, 2014). Over time the employment of combat simulations has expanded within the DoD including uses in operations planning, requirements analysis, operational test and training.

Due to the infrequent and competitive nature of combat, several challenges present themselves when developing a predictive simulation. First, there is limited data with which to validate such analysis tools. While it is possible to validate individual pieces of a combat simulation, such as the radar performance of a platform, to assess the integration of all mission aspects against specific threats is a much more significant effort. Attempts have been made to validate combat models in aggregate with historical data (Schramm, 2012), but this is of little value as the models themselves require heavy modification to incorporate modern or future forces, requiring further validation. Secondly, there are many aspects of combat modeling that are highly uncertain and not knowable (an unresolvable uncertainty (Bankes, 1993)), such as the exact tactics, techniques and procedures of an adversary force in response to a blue force strike.

Recognizing these issues, Dewar (1996) developed a topology of uses of distributed, real time simulations, shown in Figure 1, which delineated between strongly predictive and weakly predictive uses of these simulations. Strongly predictive models are described as having a demonstrated capacity to forecast outcomes with a high degree of accuracy. Examples of these types of models include engineering or physics based models to predict part life, strength, fatigue characteristics, etc. Alternatively, weakly predictive models suffer from moderate to high levels of parametric, structural or other uncertainties. Yet, the model still captures enough of the critical elements of the system under study to be useful in exploratory analysis. Millar (2016) extended this idea, arguing that the taxonomy applied to combat models in general. Due to the uncertainties described above, models of combat, to include Live-Virtual-Constructive simulations (LVCs), are generally considered weakly predictive simulations and thought to be most appropriately used for exploratory purposes.

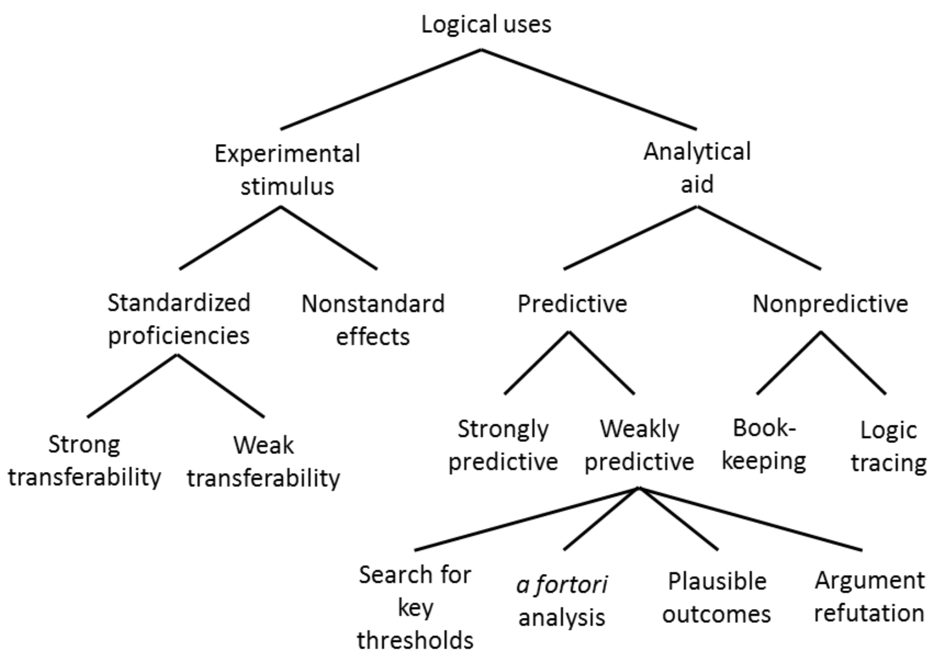

Figure 1: Logical Uses of the Distributed Interactive Simulation (DIS) System (Dewar, 1996)

An alternate philosophy with respect to use of weakly predictive models can be found in the weather and climate forecasting community. There are many models that predict the weather or climate activity in existence. The issue is that individually, these models are not accurate enough to be useful (Tebaldi, 2007). Each model includes different representations of the underlying physics, different assumptions on initial conditions, and thus different strengths. Fortunately, when these individual models are used in aggregate (referred to as a multi-model ensemble) they become sufficiently accurate for use in predicting the weather on the local news broadcast. Ensemble aggregation techniques range from simple averages across the simulation responses, to more sophisticated techniques, like Bayesian model averaging (Ajami, 2007).

There are significant differences between models of climate phenomena and models of combat. Chief among these is a source of validation data with which to compare the ensemble and individual model performances. Lacking this data for combat, it still may be useful to further consider the idea of employing multiple models of a similar combat situation and leverage the various sources of responses for either evaluating internal consistency of the ensemble or obtaining explicit bounds that incorporate various modeling perspectives. In the same way that the inadequacy of individual weather models can be overcome through an ensemble approach, the uncertainties and errors within combat simulations could be mitigated through the use of a plurality of models.

Since there is limited data for combat model validation, the techniques for aggregation in multi-model ensembles for weather forecasting are not appropriate. There would be no way to confirm that the aggregated ensemble provides any predictive improvement over any individual simulation response. But the point here is not to make weakly predictive models strongly predictive models, but to improve the insight gained through a modeling and simulation activity. The most appropriate use of data from an ensemble of similar mission level effectiveness assessments may be to check for the internal consistency among them. This would yield insight into the overall sensitivities of the system with respect to multiple modeling choices. In a setting where multiple, similar mission level modeling and simulation studies are being executed, this approach would also provide some quantitative backing to the aggregation methods, which are typically the synthesis of reports by a trusted agent of the decision maker.

One drawback of this approach might be the perceived additional cost of having to pay for multiple analyses of the same combat scenario. However, it is important to keep in mind the financial cost of choosing the wrong alternative due to unexplored uncertainties associated with modeling choices. Also, even in the current budget conscious environment, the DoD often pays for the same analysis multiple times without any thought for how the results might be cohesively integrated. At a recent review the schedule for a development planning effort was presented. Included in the schedule were three different mission level assessments of similar concepts against the same threat, with three different performing organizations, using three different modeling and simulation software packages. While it appears this was planned haphazardly, with program managers and engineers making use of available resources, it could point to the programmatic feasibility of implementing an approach where multiple similar assessments are commissioned and then quantitatively explored for discrepancies or identification of bounds in the face of uncertainty.

There is a growing body of work in applying Evidence Theory as a framework for systematic exploration and quantification of uncertainty in modeling and simulation. Applications can be found in fields where there is near zero fault tolerance and uncertainty exists within the system, such as space launch and nuclear power plant design (Oberkampf, 2002). Evidence Theory differs from probability theory in that likelihood is assigned to sets (i.e.

a range of parameter values) instead of being assigned to a probability density function. By explicitly defining ranges of uncertain input parameters and propagating them through a model, Evidence Theory bounds the true cumulative density function for a response by empirically developed cumulative plausibility functions (CPF, upper bound on probability) and cumulative belief functions (CBF, lower bound on probability) (Figure 2). Other methods within Evidence Theory allow for a quantification of the “conflict” among varying sources of information.

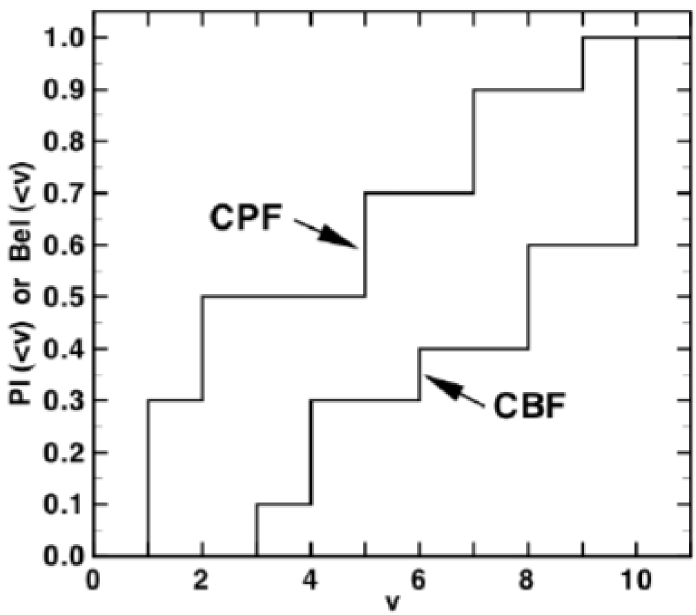

Figure 2: Depiction of Evidence Theory Bounds on CDF (Oberkampf, 2002)

1.1 Research Contributions

This research develops a comprehensive set of techniques for the treatment of uncertainty and error in combat modeling and simulation analysis. This approach enables the analyst to reconcile multiple and potentially conflicting perspectives, quantitatively. The following are a summary of the primary research contributions described within:

1) Demonstration of Evidence Theory as a framework for representation of epistemic uncertainty in combat modeling;

2) A novel procedure for sensitivity analysis of the uncertainty in modeling and simulation output with respect to several epistemically uncertain factors; and

3) Development of a method to prioritize the impact of modeling choices (or error) on simulation output uncertainty in settings where multiple models are employed in a similar context.

1.2 Outline of the Dissertation

This dissertation was prepared in k-paper format. It begins with an overarching literature review of combat modeling and methods of uncertainty quantification. Examples from the weather modeling and nuclear system safety are explored and compared with proposals for analysis of uncertainty in combat modeling and simulation.

The next three chapters were designed to be completely severable, documenting three distinct contributions to the field of combat modeling and analysis of uncertainty. Each of these articles includes an introduction, complete literature review, methodology, results and discussion, and conclusion section.

The first article demonstrates Evidence Theory as a framework for representing epistemic uncertainty in combat modeling output. This approach unifies the analysis of uncertainty in combat modeling with modern approaches to uncertainty analysis and provides a direct mapping of input uncertainty to uncertainty in the distribution of simulation output. To provide context for the Evidence Theory analysis, a traditional approach was employed in assessment of uncertainty of Blue minus Red residual forces in a Lanchester model of conflict. The results of both analyses were compared and contrasted.

In the second article, a new method for sensitivity analysis of uncertainty in Evidence Theory was developed. This sensitivity analysis method generates marginal CPFs and CBFs and prioritizes the contribution of each factor by the Wasserstein distance (also known as the Kantorovich or Earth Mover's distance) between the CBF and CPF. Using this method, a rank ordering of the simulation input factors can be produced. This method is less susceptible to ties when there are large uncertainties in outcomes with respect to the variables compared to existing approaches in the literature.

The third article builds on the first two and uses Evidence Theory to prioritize the impact of error or modeling choices on simulation output uncertainty in settings where multiple models are employed. This analysis provides insight into the overall sensitivities of the system with respect to multiple modeling choices. The new method does not make weakly predictive models strongly predictive, but provides a mechanism for reconciling a plurality of modeling perspectives during a simulation activity.

II. Literature Review

2.1 Development and Use of Combat Models

Models of combat have been employed by systems and operations research analysts since the 1960's for exploring possible outcomes of hypothetical military conflict (Davis, 1995). The fidelity of these models has ranged from simple mathematical relationships describing attrition between two opposing forces (Lanchester, 1914) to high fidelity operator or hardware in the loop simulations for exploration of detailed system configuration changes (Haase, 2014). Over time the employment of combat simulations has expanded within the DoD including, operations planning, requirements analysis, operational test and training. While each of these applications are important, of particular interest for this document, are the class of models used to support requirements analysis.

2.1.1 The Hierarchy of Combat Models

In the U. S. Department of Defense (DoD), combat models are frequently categorized into campaign, mission, engagement and engineering levels of fidelity (see Figure 3). Of primary concern in this paradigm, is the tradeoff between scope of the modeling effort and the fidelity with which entities and their interactions are represented. In general, as models move up the pyramid from engineering to campaign level analysis, the level of aggregation increases and the of combat processes are represented with less resolution (or fidelity). The key for the analyst is in appropriately choosing a model that best addresses the decision maker's question.

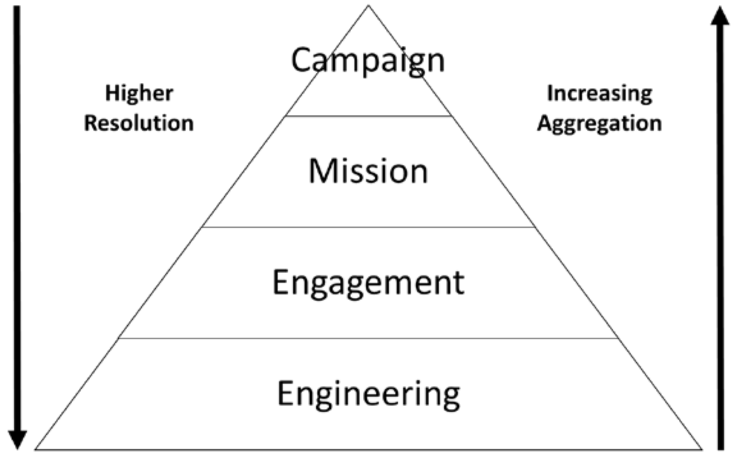

Figure 3: Combat Modeling Hierarchy (Miller, 2016)

Outside the military domain, model aggregation techniques have been widely applied for commercial purposes. Examples were found in transportation (Lee, 2004), ecology (Wu, 2002), materials science (NRC, 2008) and the electronics industry (Zhao, 2002). The aggregation methods in these fields were often defined by simple relationships, such as aggregating roadways that run in similar directions into a single “road”, or defined by physics based relationships, such as in materials processing.

Rodriguez (2008) provides an overview of statistical techniques for aggregation and disaggregation in combat modeling and simulation. Performance of each statistical technique was compared against the full model (no aggregation). Use of fitted distributions or artificial neural networks were not recommended and other statistical modeling methods produced statistically similar results as the full model. The authors also acknowledged that non statistical factors may be as important as statistical factors in choosing an appropriate aggregation method, such as comprehensibility and skill and comfort of the analyst with the technique.

2.1.2 Live-Virtual-Constructive Modeling and Simulation

An alternate taxonomy of DoD combat simulations is the Live-Virtual-Constructive (LVC) paradigm. Live simulations are the class of simulations where real people operate real equipment. An example of this would be in test where a real pilot in a real aircraft flying against synthetic, computer generated targets. In a virtual simulation, real people operate simulated systems or simulated people operate real systems. This is the class of simulation most frequently associated with the LVC paradigm and crew familiarization and training are classic examples of these activities. In the case where simulated people operate simulated systems, these exercises are referred to as constructive simulations. The operations research community within the DoD considers these analytic simulations and many variants have been developed across the engagement, mission and campaign levels of fidelity. Some prominent examples include THUNDER, STORM, Brawler, and Suppressor.

LVC exercises typically involve multiple geographically separated simulators connected via a network. Flexible communication protocols have been developed over time to ensure the vast array of assets intended for use in LVC simulations can adequately communicate. Popular communications protocols in the LVC community include; the Test and Training Enabling Architecture (TENA) (TENA, 2017), Distributed Interactive Simulation (DIS) (IEEE, 2012) and High Level Architecture (HLA) (IEEE, 2000). Each of these standards has a core set of definitions which specify the data to be transmitted across the network, the frequency with which to transmit, and standards for handling entity to entity interactions (e.g. kill removal). Choice of protocol is typically dependent on level of consistency required across the simulation network nodes, with HLA typically employed in situations requiring high levels of shared state consistency and TENA applied in Live situations where no guarantees can be made about state consistency. Details on each of these protocols can be found in their respective specifications (TENA, 2017; IEEE, 2012; IEEE, 2000).

As each of the independent nodes in an LVC exercise must process data across the network, discrepancies may occur from site to site regarding the true state of the simulation (Millar, 2016). To mitigate this outcome, software engineers can adjust the frequency with which updates are sent to other players. These actions result in increasing overall network traffic and processing burden on the simulators, potentially degrading the experience of the human operator as the simulation software begins to process data slower than real time. This phenomena is called the consistency-throughput tradeoff and has been studied extensively in the context of DoD LVC events (Sandeep, 1999; Hodson, 2009).

2.2 Uncertainty in Combat Modeling and Simulation

2.2.1 Categories of Uncertainty

The distinction between uncertainty, variability and error in a modeling and simulation study has not been consistently employed within the vocabulary of the analytical community. A useful framework for discussing variability and error in a modeling and simulation study was proposed by Oberkampf (2002). This framework proposes two kinds of uncertainty and the notion of error within a modeling and simulation context that are akin to the colloquial use of the terms uncertainty, variability, and error. Using their definition, aleatory uncertainty "describes the inherent variation associated with a physical system or environment under consideration" (Oberkampf, 2002:334). This could be thought of as the defect rate in a manufacturing process, where the same physical processes occur repeatedly, yet each part does not come off the production line exactly to specification. Similar terms for aleatory uncertainty include variability, stochastic variability, or irreducible uncertainty. A second category of uncertainty, epistemic uncertainty, was defined as “the potential deficiency in any phase or activity of the modeling process that is due to lack of knowledge” (Oberkampf, 2002:334). This category is difficult to conceptualize in most process flow modeling and simulation contexts where any potential epistemic uncertainty could be resolved by simply inspecting the process as it occurs. Epistemic uncertainty does manifest in the materials science and engineering realm, where certain model parameters would require materials testing at high temperatures and current methods preclude collecting this data.

These two types of uncertainty are related in that their impact manifests as either simulation output variation or unquantified decision risk. It’s easy to see how a simulation process that includes some representation of the aleatory uncertainties results in variation in simulation output through replication, either within run or run-to-run. Epistemic uncertainties can also induce variation in simulation output if assumptions regarding unknown aspects are enumerated and relevant inputs varied within the study. Of more concern is when epistemic uncertainties are not explicitly varied within a simulation, providing no insight into the sensitivity of simulation responses to assumptions for the analyst, resulting in unquantified decision risk. It is likely that the systematic varying of assumptions associated with epistemic uncertainties will not cover all unique possibilities, but at least relative impacts can be identified and presented to the study stakeholders to qualitatively include in their deliberations.

The popular terminology typically stops at this point with variability commonly referring specifically to aleatory uncertainty and uncertainty to epistemic uncertainty. There is a third useful distinction to make alongside the two types of uncertainty that classifies cases where choices in model abstraction and software implementation have an appreciable impact on the form of the simulation. In the manufacturing example, the modeler could choose to implement their simulation with either discrete event or agent based perspective. This decision, a choice made in the process of abstracting the physical system for simulation, can change decisions made based on modeling and simulation analysis in either a satisfactory or unsatisfactory fashion. In this case, the effect on simulation output due to the selected modeling paradigm is related to neither the natural variability of the process or elements that are unknowable regarding the system under study. To account for this scenario, Oberkampf (2002) proposes the concept of simulation error. They define simulation error as "a recognizable inaccuracy in any phase or activity of modeling and simulation that is not due to lack of knowledge" (Oberkampf, 2002:334). Unfortunately this has the connotation that someone has done something "wrong", which may not be the case. While unacknowledged errors are the term that describes errors made by the modeler, of more interest for this discussion are the acknowledged errors or errors resulting from an intentional effort in the system abstraction or simulation implementation process. The impact of error in a simulation study results in biased simulation output or explainable deviation from "truth". In a similar fashion to both epistemic and aleatory uncertainty, certain forms of error could be systematically explored to identify sensitivities to choices by the analyst.

2.2.2 Traditional Methods for Analysis with Uncertainties

Techniques have been developed to identify drivers of response variation and reduce the width of confidence intervals to improve mean estimates for comparing two systems and improve discriminatory capability such as; paired t-test, common random numbers, antithetic variates, control variates, etc. For detailed procedures on executing these techniques, see (Law, 1999) or (Banks, 2010).

More recent work has focused on assessment of the overall effect of input uncertainty on simulation output. Typically input distributions for simulation execution are "fit" based on empirical data, and reduced to a closed form distribution to make simulation implementation simple. Since these inputs are based on processes which are not guaranteed to specify the true distribution, there is uncertainty in selection of distribution family and input parameters. Since there are many such inputs in a typical simulation study, methods to identify the inputs with largest impact are desirable. In Ankenman (2012), a random-effects model was employed to estimate the ratio of input uncertainty relative to the standard error of the model. This idea is extended in Song (2013), by developing a direct model of the variance contributors based on variation in simulation response and an optimal run allocation for estimates of the marginal variances.

2.2.3 Uncertainty in an LVC Context

In the context of LVC, of particular interest to the DoD analytical community are the uncertainties and errors associated with distributed, real time simulation exercises. The primary components of an LVC can be broken down into three categories; 1) the computerized simulation players, 2) the human and 3) the distributed simulation network (Figure 4).

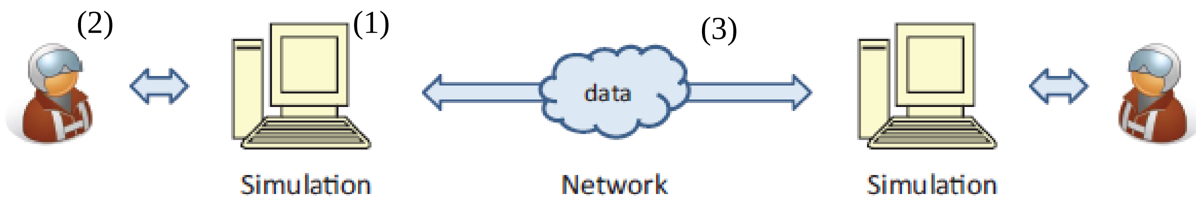

Figure 4: Architecture of an LVC Exercise (Adapted from (Hodson, 2014))

If a particular distributed, real time simulation exercise contained a live system or actual operational flight programs (OFP), they would be assumed to have the effect of reducing the uncertainty and error associated with the system representation. This assumes that the live and OFP systems are deterministic, the same input yields the same output. Humans are not deterministic, and this is the reason why they are called out as a separate domain of uncertainty and error.

The computerized simulation players include any aircraft simulator, but not the human, and any constructive player, to include player behaviors and logic, in the simulation environment. The uncertainties and errors associated within this domain would be similar to those associated with a purely constructive simulation of combat. These could include uncertainties in opposing force capabilities, size, and deployed location. Any one of these could be categorized as parametric, structural, resolvable or unresolvable.

If there was an explicit treatment (i.e. systematic exploration and changing of assumptions) of the identified uncertainties within this portion of the overall simulation, summary statistics (mean and variance) or confidence intervals of simulation responses could be captured and exploited for analysis. Techniques developed for constructive simulations to identify drivers of response variation (Ankenman, 2012; Song, 2013) or to improve mean estimates for comparing two systems (Banks, 2010; Law, 1999), are not practical for distributed simulation. These issues arise in an LVC as the software is either not centrally controlled or readily modifiable to accommodate these techniques, and there are small numbers of runs and highly correlated output (see Table 1).

Table 1: Summary of Sources of Uncertainty and Error in Modeling and Simulation Exercises

Summary of Sources of Uncertainty and Error in Modeling and Simulation			
Source	**Resulting Manifestation**	**Analysis Artifacts**	**Compensation Methods**
Computerized Simulation Elements			
- Lack of knowledge/data of underlying physical system	- Decision risk	- Stochastic inputs	- Assessment of input uncertainty (i.e. Song and Nelson)
- Stochastic variability	- Variation in simulation responses	- Confidence intervals, mean, variance	- Replications
- Configuration management/matched fidelity		- Choices in abstraction	- Design of experiments
			- Paired t test, antithetic variates, control variates, hypothesis testing
Human-in-the-Loop			
- Learning effects	- Non independent and identically distributed observations	- Correlated response data	- Experimental planning
- Operator availability	- Limited replications	- Inputs map to multiple responses	- Design of experiments
- Lack of repeatability		- Small sample sizes	- Bootstrapping
Network Effects			
- Uncontrolled processing of data	- Unverified events (non plausible outcomes)	- Inconsistent entity state data	- Manual data review
- Finite capacity			- "White cell" adjudication
			- Qualitative apriori verification

Additionally, in combat modeling there are significant uncertainties associated with lack of knowledge of the system, resulting in decision risk. There are no simple statistical techniques to reduce the risk associated with “unknown unknowns” (or “known unknowns”), outside of conducting experiments across a wide range of threat scenarios or excursions.

Two unique sources of uncertainty and error in distributed, real time simulations include the human operator in a virtual setting and the computer network that connects the players. The employment of a human operator as part of the simulation system injects aleatory uncertainty into the exercise due to the poor repeatability reported in LVC exercises. This requires the analyst to develop strategies to combat learning curve effects, low numbers of replications, and difficult to randomize experiments. The resulting data from these exercises cannot be assumed to be independent and identically distributed (iid), and is likely highly correlated. This leaves employment of common techniques that assume iid output suspect. The limited availability of representative operators in conjunction with learning curve effects limits the number of runs that can be practically or usefully run. There are many studies published in the literature that highlight these issues (Haase, 2014; Gray, 2007; Hodson, 2014). Most point to rigorous experimental planning as the best way to combat the effects of having a human operator as part of the simulation system, but there is limited guidance on the topic.

2.2.4 Methodologies for Addressing Uncertainty in Combat Modeling

One approach to addressing the weakly predictive nature of combat models is employment of multiresolution, multiperspective modeling. This, as described by Davis (2000), is a family of models that describe similar phenomena at varying levels of resolution. A key feature of the proposed modeling system would be that the family is consistent or mutually calibrated within itself. This is in contrast with the current family of Air Force modeling and simulation tools which were not designed to be integrated from the outset and require significant effort to do so.

An example of the implementation of multiresolution, multiperspective modeling was discussed in (Davis, 2000). The analysis task was assessing the effectiveness of long range fires in interdicting an invading land force. A multiresolution, multiperspective model called PGM Effectiveness Model (PEM) was developed to assess the operational scenario, but was calibrated based on results from legacy models JANUS and MADAM. Due to the differences in fidelity and scope across JANUS, MADAM and PEM, a large effort to reconcile the output across the three models was required. Because of this additional (and non-simulation) effort, significant insight was gained in the post simulation reconciliation which clarified insight for the decision maker.

In a similar vein, a National Academy of Science study advocated for a “multi-resolution analysis” approach to support the Air Force intelligence, surveillance and reconnaissance (ISR) Capability Planning and Analysis process (NAS, 2012). This proposed framework is intended to integrate elements of network analysis, sensor physics, cost analysis, operational analysis tools, and mission effectiveness analysis (Figure 5).

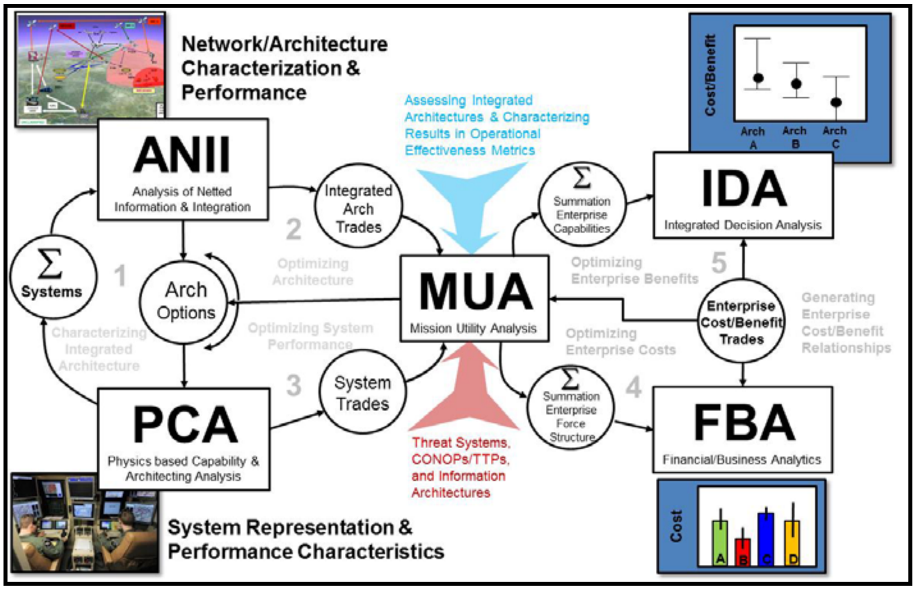

Figure 5: Air Force ISR Multi Resolution Analysis Process (NAS, 2012)

The intent of the multi-resolution analysis framework is to provide a rigorous foundation to early acquisition decisions for ISR. Current incarnations of the process include some procedures for identifying and exploring uncertainties and advanced experimental design, but commonly operate with a single constructive mission level modeling and simulation tool.

Wright (2004) presented a method for utilizing two radically different modeling paradigms for analysis of strategic airlift. The first model was a simulation of cargo flow from on load to offload point with explicit representation of the number of aircraft, routes, air based infrastructure, and other resources. The second model was a large-scale linear program with side constraints, which was initially developed to assess fleet adequacy and for use in identifying system bottlenecks. While these models share common elements, the representation of the real system is different for each model. For example, the simulation models event durations as random variables while the optimization employs mean values for these quantities. The proposed framework assessed the convergence of these model outputs based on iteratively modifying the individual models with the goal of reducing discrepancies between them. When the models were sufficiently close in output, they were considered covalid.

2.3 Design, Development, and Analysis of Model Ensembles

In the weather domain, seasonal forecasts have been shown to have better predictive capability when several independent models are combined, commonly referred to as a multi-model ensemble (Tebaldi, 2007), than when the individual predictions of those models are taken alone. There are several domain specific phenomena that contribute to the difficulties with single model analyses and the success of multi-model approaches, some of which are (Palmer, 2004):

- The physics of weather is well understood, but the phenomena are chaotic and models are sensitive to initial conditions
- Solutions to the partial differential equations that describe weather phenomena must be reduced to analytically tractable forms, which introduces computation error in final solutions
- There are numerous ways to implement approximations of the underlying physics and numerical approximations
- There is no underlying framework from which a pdf of model uncertainty can be estimated.

Developers of multi-model ensembles must consider both the number and composition of models within the ensemble as well as the method of data aggregation and assessment of ensemble predictive skill.

Multi-model superiority is not only due to error compensation, but primarily by its improved constancy and reliability across the entire predictive region (Hagedorn, 2005). A particular ensemble may not be the best model at each point within the predictive region, but generally outperforms any given individual model over the full range of cases. The success of multi-model methods in the weather community has led to the development of analogous approaches in prediction of disease outbreak (Morse, 2005) and rainfall runoff (Ajami, 2007).

2.3.1 Ensemble Construction

The analysis of multi-model ensembles rely on the principle that the component models are structurally independent from each other (Palmer, 2004). Otherwise, specific bias may be overrepresented (or underrepresented) in the ensemble and have an outsized impact on the aggregated results. This ensemble property is not necessarily knowable and ignored in practice (Hagedorn, 2005) with suitability of results a product of the validation process. Independence of a multi-model ensemble is often taken to be implied by the fact that various groups have independently developed their own models, and it follows that their construction methods were not influenced by others development efforts. These variations in model physics and numerical computation methods play a substantial role in generating the full spectrum of possible solutions.

Along with independence of the component models, the aggregator must also manage the number of models within the ensemble. Larger ensemble sizes are generally considered better, with widely used climate models reporting between 7 (Palmer, 2004) and 11 members

(Kirtman, 2014). Research has shown that predictive skill scores grow faster with ensemble size when membership is less than 30, and saturates with large numbers of members.

2.3.2 Statistics of Aggregation

Data aggregation techniques for analysis of multi-model ensemble output range from simple averages to more advanced Bayesian techniques. Model output can vary from model to model in terms of variation (or range) and mean. The impact of modeling decisions can change either of these features, changing the variation of the output or inducing bias in the response. Using historical climate data, it is possible to correct for systematic spatial shifts of each model within the ensemble (Doblas-Reyes, 2005). Canonical correlation analysis and variance inflation techniques have both been demonstrated to enhance the reliability of multi-model forecasts.

The simplest implemented multi-model forecast is developed by combining individual contributors with equal weight (Hagedorn, 2005; Kirtman, 2014). Extensions of this procedure with optimal weights developed for each contributor based on historical prediction capability have been developed, but it has proved difficult to calculate robust weights with the available training datasets. Methods that only output the grand mean lose the information associated with the differences in variation across models, motivating probabilistic techniques for data aggregation (e.g. Bayesian model averaging, etc.).

2.3.3 Assessment of Skill

Verification and validation of multi-model ensembles face similar issues as any other modeling and simulation activity. An identified key to success lies in combining independent and skillful models each with their own strengths and weaknesses (Hagedorn, 2005). Tebaldi (2007) outlines several key processes for verification and validation of model ensembles, to include a comparison of ensemble prediction to historical data, evaluation of theoretical correctness of individual model behavior, and perceived trust of the model. Generally, most climate models agree reasonably well with present day mean climate, but many diverge significantly in predictions of future climate.

To aid in the validation process, Palmer (2004) developed a modular verification system, comprised of various indices of predictive skill with metrics for both deterministic and probabilistic (output is distribution of outcomes) simulations. Deterministic validation metrics included anomaly correlation coefficient, root mean square skill score, and mean square skill score (Hagedorn, 2005). Probabilistic simulation skill is assessed with reliability diagrams, ROC skill score, Brier score, and ranked probability skill score, among several others. Issues with validation of probabilistic forecasts beyond those associated with a deterministic simulation include; improper estimates of probabilities from small-sized ensembles, insufficient number of forecast cases, and imperfect reference values due to observation errors (Doblas-Reyes, 2005).

The choice of best ensemble is sensitive to choice of model output (metric) and skill assessment method (Hagedorn, 2005), with no single measure being sufficient for comparing forecast quality across all possible ensembles (Doblas-Reyes, 2005). Skill scores have been demonstrated to grow faster with ensemble size when membership is less than 30, saturates with large numbers of members (Palmer, 2004), and multi-model ensembles have been shown to be systematically more skillful especially, when scores are averaged over large regions or long periods of time.

2.3.4 Criticisms

Tebaldi (2007) offers several criticisms of employment of multi-model ensembles. One was the notion that the performance of a forecast improves by averaging multiple models is based on the assumption the models are independent and that the errors cancel as the averages are taken. Since many models in the weather community are based on a similar understanding of the physics of weather, they are related at some level, thus not truly independent and biased.

Second is that model ensembles are often assembled out of convenience (Tebaldi, 2007), with no systemic approach to sampling models for the ensemble. This can lead to unexplainable changes in predicted performance by swapping out models in an ensemble. Stated another way, there is no practical guarantee that all model uncertainties are accounted for (and suitably independent) within a given multi-model ensemble (Kirtman, 2014) and additional models may change the aggregated results.

2.3.5 Applications

Both American (Kirtman, 2014) and European (Palmer, 2004) multi-model ensembles have been successfully developed to aid in developing accurate climate forecasts. Typical output of these systems includes sea surface temperature, two mile temperature and precipitation rates, with aggregation occurring via simple averaging techniques. The component models' configuration, resolution, etc. are left to the forecast providers and not centrally managed by the aggregator. Each of these ensembles have demonstrated smaller prediction errors than any of their individual contributors over the full range of experimental conditions.

Bayesian methods have been demonstrated as a technique for analysis of multi-model ensembles of climate characteristics. Smith (2009) developed a Bayesian framework as an objective method of quantifying output uncertainty in a multi-model ensemble using uninformative priors. In a similar vein, Tebaldi (2005) developed a Bayesian approach with hyperprior distribution parameters to assess the inter-model agreement of regional temperature predictions. This approach was used to generate univariate regression models of temperature change for each individual region and a multivariate regression model where the response was the vector of all regions' temperature change.

Another use of multi-model ensembles was found in Ajami (2007), where Bayesian model averaging was used to create a multi-model ensemble for prediction of rainfall run-off. Bayesian model averaging is a technique that weights the individual models by the likelihood that the model matches a comparison metric. The comparison metric could be historical data or a baseline model that is used as a cross validation source. Ajami (2007) built a model ensemble comprised of three hydrologic models and performed a validation based on historical data. It was found that the model ensemble improved the number of empirical observations that were within a 95% confidence interval of the model estimate by over 300% (from a maximum 22% capture to 76.3% capture).

2.4 Evidence Theory

There is a growing body of work in applying Evidence Theory, also known as Dempster-Shafer Theory, as a framework for systematic exploration and quantification of uncertainty in modeling and simulation. The theory was first introduced by Dempster (1967) and later codified by Shafer (1976). Applications can be found in fields where there is near zero fault tolerance and uncertainty exists within the system, such as space launch and nuclear power plant design (Sentz, 2002). Evidence Theory differs from probability theory in that likelihood is assigned to sets (i.e. a range of parameter values) instead of being assigned to a probability density function. By explicitly defining ranges of uncertain input parameters and propagating them through a model, Evidence Theory bounds the true cumulative density function for a response. There are three key functions in Evidence Theory; the basic probability assignment function (*bpa* or *m*), the Belief function (*Bel*) and the Plausibility function (*Pl*).

In general the basic probability assignment is not equivalent to probability as discussed in classical probability theory (although connections exist (Sentz, 2002)). Similarly to classical probability theory, the basic probability assignment is a mapping of all sets (*X*, the power set) to the interval [0, 1] and the sum of all assignments across subsets is 1. Formally, this is represented as:

$$m: P(X) \rightarrow [0, 1] \tag{1}$$

$$m(\emptyset) = 0 \tag{2}$$

$$\sum_{A \in P(X)} m(A) = 1 \tag{3}$$

Using this basic probability assignment, upper and lower bounds for an interval can be calculated. The lower bound (or Belief), for a set *A* (subset of *X*), is the sum of all basic probability assignments of the proper subsets (*B*) of the set of interest (*A*).

$$Bel(A) = \sum_{B|B \text{ is a subset of } A} m(B) \tag{4}$$

The upper bound (or Plausibility) is the sum of all the basic probability assignments of the sets (*B*) that intersect the set of interest (*A*).

$$Pl(A) = \sum_{B|B \cap A \neq \emptyset} m(B) \tag{5}$$

2.4.1 Types of Evidence

There are many ways to combine evidence from multiple different sources in Evidence Theory, but the key to appropriately applying the right combination rules is determining the type of evidence being combined. Sentz (2002) identifies four types of evidence, with varying levels of conflict: consonant, consistent, arbitrary, and disjoint.

Consonant evidence is a collection of nested sets of data where the elements of the smallest set are included in the next larger set, which is included in the next largest set, etc. (Sentz, 2002) (see Figure 6).

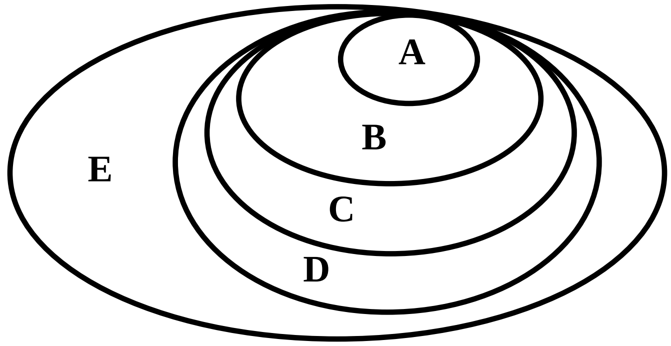

Figure 6: Consonant Evidence

Consistent evidence is the situation where at least one set is common to all other sets (Sentz, 2002). This is illustrated in Figure 7, where set *A* is a subset of each other set *B*, *C*, *D* and *E*.

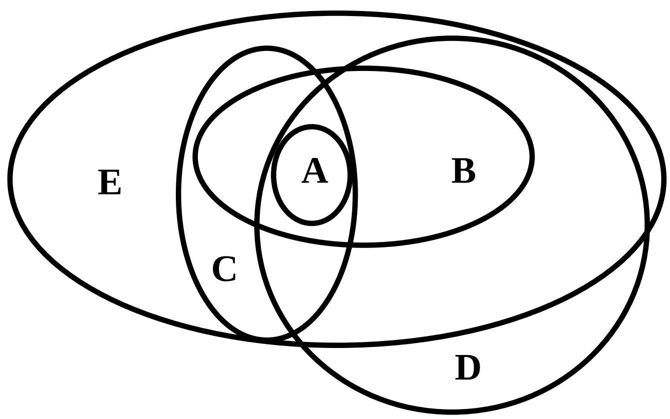

Figure 7: Consistent Evidence

Arbitrary evidence is the situation where there is no set common to all subsets (Sentz, 2002). This is illustrated in Figure 8, where there are clearly subsets that overlap but no set is common to all other sets.

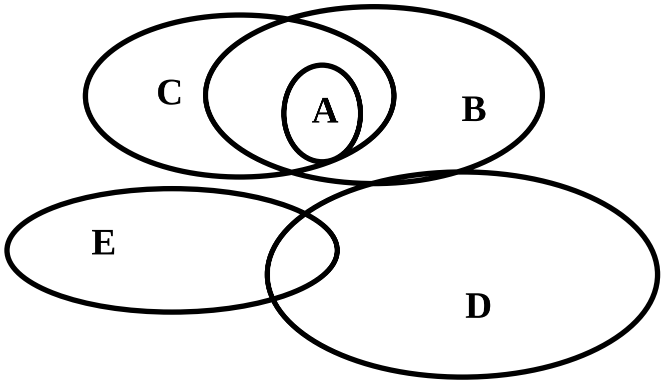

Figure 8: Arbitrary Evidence

Disjoint evidence corresponds to the situation where no two sets overlap (Sentz, 2002) (see Figure 9).

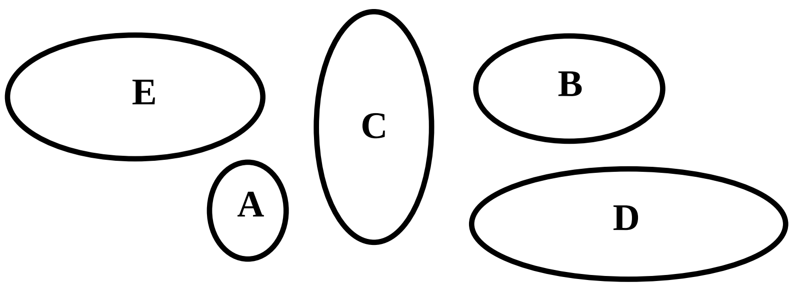

Figure 9: Disjoint Evidence

As is apparent from inspection of Figures 6 through 9, each configuration of evidence results in a different level of conflict across the available evidence. In the condition where the evidence is disjoint, the level of conflict is relatively high. Where in the case of consonant evidence, all sets share a common set, indicating relatively lower conflict. Consistent and arbitrary set configurations would exhibit a varying level of conflict, depending on the situation. These would conceivably exhibit levels of conflict between disjoint and consonant evidence.

2.4.2 Rules of Combination

The rules of combination in Evidence Theory allow data to be aggregated across multiple, potentially conflicting sources within a common frame of discernment. For the simulation context discussed in this research, these sources could be input for a common parameter from a variety of subject matter experts (SMEs), a set of epistemically uncertain inputs for an individual model, or the individual models in an ensemble. This process assumes that the sources are independent (Shafer, 1976), however this requirement is not rigorously established in practice (Tebaldi, 2007). There are many rules for combining parameter estimates in Evidence Theory, the key to providing credible insight in a given analysis is to understand or choose how conflict between sources should be considered. A survey of relevant combination rules is provided below.

Dempster's combination rule was the original combination operator that drove the conception of Evidence Theory (Dempster, 1967). Using Dempster's combination rule, the basic probability assignments from two (or more) sources is combined with a purely conjunctive operation. The formal definition of this operation (m_{12}) is below:

$$m_{12} = \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-K} \quad when\ A = \emptyset \tag{6}$$

$$m_{12}(\emptyset) = 0 \tag{7}$$

$$\text{where } K = \sum_{B\cap C=\emptyset} m_1(B)m_2(C) \tag{8}$$

It is important to note that this operation results in an aggregated mass only for intervals which overlap, giving zero mass to regions where evidence existed but did not overlap with another method. The measure of non-overlapping probability mass (or conflict) is represented by the computation of *K*.

The aggregated masses are normalized based on *K* to achieve a basic probability assignment function that resembles a probability density function from classical probability theory. Unfortunately, this choice can lead to counterintuitive results in situations involving high levels of conflict. These shortcomings are detailed in (Zadeh, 1984). Recognizing the potential pitfall, numerous other combination rules have been developed which account for level of conflict differently.

Yager's combination rule does not normalize the probability mass assignments by the degree of conflict (Yager, 1987) and distinguishes between the basic probability assignment and a new construct, the ground probability assignment (*q*). The ground probability assignment represents the evidence as provided, without being inflated based on conflict. As such, the sum of the ground probability assignments over all sets will not necessarily equal 1. The formal definition of the ground probability assignment is provided below:

$$q(A) = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{9}$$

Due to the associativity of Yager's operator, multiple pieces of evidence can be combined simultaneously. This result is shown in equation (10).

$$q(A) = \sum_{\cap_{i=1}^{n} A_i = A} m_1(A_1) m_2(A_2) \ldots m_n(A_n) \tag{10}$$

In Yager's construct for combining evidence, conflict is assigned to the universal set, with $q(\emptyset)$ (see equation (11)) having the interpretation of the degree of ignorance in the data.

$$q(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{11}$$

Yager's combination rule is equivalent to Dempster's combination rule when the degree of conflict is zero.

The Mixing Rule of combination is a popular mechanism for aggregation of disjunctive evidence (see equation (12)). This rule averages the masses (m_i) associated with a particular interval across all *i* estimates (*i* from 1 to *n*). The individual estimates can be weighted based on reliability by the multiplier, ***w***, where each w_i is the reliability associated with the *i*th source.

$$m_{1\ldots n}(A) = \frac{1}{n} \sum_{i=1}^{n} w_i m_i(A) \tag{12}$$

In contrast with Dempster's Rule of combination, evidence in conflict is preserved in the resulting BPA. Said another way, the full range of possibilities expressed in the sources are represented in the final BPA. This feature is particularly beneficial where the application of evidence theory is not to identify the most likely distribution of a particular metric, but to express the full range the distribution could be.

There are many additional rules for combining evidence, such as; Discount and Combine (Sentz, 2002), Convolutive x-Averaging (Sentz, 2002) (a generalization of the average for scalar numbers), and the qualitative combination rule (Yao, 1994) (rank ordered process for aggregating data).

2.5 Bayesian Analysis

Bayesian analysis is a method of statistical inference that allows hypotheses regarding uncertain quantities to be updated based on observed or new information. In the frequentist paradigm, the uncertainty in many real word quantities are reduced to singular values, where they are represented as distributions with unknown parameters in a Bayesian paradigm. Gelman (2014) defines a three step process for any Bayesian data analysis effort:

1. *Setting up a full probability model.* In this phase, a joint probability distribution is established for all quantities of interest in the analysis.

2. *Conditioning on observed data.* This involves calculating the posterior distribution, which is the distribution of the unobserved quantity of interest conditioned on the observed data.

3. *Evaluating the fit of the model and implications of the resulting posterior distribution.* This phase is where insight is extracted from the constructed model, reasonableness of conclusions are assessed, and sensitivity analysis is executed relative to assumptions made throughout previous steps.

2.5.1 Bayes' Rule (Gelman, 2014)

The central theme of Bayesian analysis is to make probability statements about some quantity ϑ given some data y. This quantity can be expressed as the product of two different probability densities, known as the prior distribution ($p(\theta)$) and the data distribution ($p(y|\theta)$), as shown below:

$$p(\theta, y) = p(y)p(y|\theta) \tag{13}$$

By conditioning on the obtained data *y* and applying Bayes' rule, the expression for the posterior density is obtained:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} \tag{14}$$

The term p(y) is a constant, as it does not depend on θ and is frequently written as:

$$p(\theta|y) \propto p(\theta)p(y|\theta) \tag{15}$$

Before the data *y* are collected, the probability density function of *y* is

$$\begin{aligned} p(y) &= \int p(y, \theta) d\theta \\ &= \int p(\theta)p(y|\theta) d\theta \end{aligned} \tag{16}$$

This is also referred to as the prior predictive distribution as it is not conditional on the collected data. If inference about $\check{y}$ after *y* data have been collected is desired, the posterior predictive distribution is derived (see equation (17)).

$$p(\breve{y}|y) = \int p(\breve{y}, \theta|y)d\theta \tag{17}$$

$$= \int p(\breve{y}|\theta, y)p(\theta|y)d\theta$$

$$= \int p(\breve{y}|\theta)p(\theta|y)d\theta$$

Using this distribution, inference about the unknown but observable quantities in an analysis can be made with updated quantity estimates based on observations of the process.

2.5.2 Dirichlet Processes and Bayesian Histograms

Practical application of Bayesian analysis requires the ability to compute the integrals in equation (16), which is not always easy or straight forward. Significant research efforts have been undertaken to approximate these integrals, employing Monte Carlo and/or markov chain methods (Gelman, 2014). Another line of research has identified convenient probability density functions for which simple conjugate prior distributions exist. Of specific interest to this paper is the analysis of simulation output, which has been turned into interval data for histogram comparison. This results in the simulation output being modeled as a multinomial distribution, where the possible outcomes are equivalent to the histogram bins. While other Bayesian approaches exist that may not require strict bins widths to be defined (Gelman, 2014), these are reserved for future research.

The multinomial distribution, a generalization of the binomial distribution is used to describe data for which each outcome is one of *k* discrete possibilities. Assigning *y* as the count of observations of each outcome *k*, then the posterior density is:

$$p(\theta|y) \propto \prod_{j=1}^{k} \theta_j^{y_j}, \tag{18}$$

where the sum of probabilities, $\sum_{j=1}^{k} \theta_j$, is 1 and is typically considered to implicitly condition on the number of observations, $\sum_{j=1}^{k} y_j = n$. The conjugate prior of the multinomial distribution is a generalization of the beta distribution, known as the Dirichlet. For which the probability density function is:

$$p(\theta|\alpha) \propto \prod_{j=1}^{k} \theta_j^{\alpha_j - 1}, \tag{19}$$

where θ_j's are greater than 0 and sum to 1. The resulting prior distribution for the θ_j's is Dirichlet with parameters $\alpha_j + y_j$.

Bayesian analysis of interval data produced via simulation can be accomplished through the modeling as a Bayesian histogram. Suppose a set of points $\xi = (\xi_0, \xi_1, \dots, \xi_k)$ have been defined which identify the intervals for the histogram estimate, with $\xi_0 < \xi_1 < \dots < \xi_k$ and $y_i \in [\xi_0, \xi_k]$. A probability model representation of the histogram is as follows:

$$f(y) = \sum_{h=1}^{k} 1_{\varepsilon_{h-1} < y \leq \varepsilon_h} \frac{\pi_h}{(\varepsilon_h - \varepsilon_{h-1})} \tag{20}$$

with $\pi = (\pi_1, \dots, \pi_k)$ is an unknown probability vector. To complete the Bayes specification, a Dirichlet($a_1, \dots, a_k$) prior distribution for π is assumed,

$$p(\pi|a) = \frac{\prod_{h=1}^{k} \Gamma(a_h)}{\Gamma(\sum_{h=1}^{k} a_h)} \prod_{h=1}^{k} \pi_h^{a_h - 1} \tag{21}$$

The hyperparameters, π, can be expressed as $a = \alpha\pi_0$, where

$$E(\pi|a) = \pi_0 = \left(\frac{a_1}{\sum_h a_h}, \dots, \frac{a_k}{\sum_h a_h}\right) \tag{22}$$

a is the prior average and α is a scale parameter and interpreted as a prior sample size.

The posterior distribution is then calculated as:

$$p(\pi|y) \tag{23}$$

$$\propto \prod_{h=1}^{k} \pi_h^{a_h - 1} \prod_{i: y_i \in (\varepsilon_{h-1}, \varepsilon_h]} \frac{\pi_h}{(\varepsilon_h - \varepsilon_{h-1})}$$

$$\propto \prod_{h=1}^{k} \pi_h^{a_h + n_h - 1}$$

Which by definition is Dirichlet$(\mathrm{a}_1 + n_1, \ldots, \mathrm{a}_k + n_k)$, where n_k is the number of observations within the h^{th} histogram bin. This approach is an incremental step toward development of fully non parametric Bayesian density estimation, which is reserved as a topic for future research.

Starting with equation (17), the posterior predictive distribution can be derived for this process.

$$p(\breve{y} = i|y) = \int p(\breve{y}|\theta)p(\theta|y)d\theta \tag{24}$$

$$= \int \theta_i Dir(\theta|N + a)d\theta$$

$$= \frac{N_i + a_i}{\sum_{j=1}^{k} N_j + a_j}$$

2.5.3 Applications

A survey of Bayesian methods in the context of discrete event modeling and simulation is provided in Chick (2006). These topics include uncertainty analysis, ranking and selection, input modeling and metamodeling. In the combat simulation domain, dynamic Bayesian networks were employed to analyze data created via an air combat simulation called X-Brawler (Poropudas, 2007). The data were input into a dynamic Bayesian network and used to study various courses of action within a combat situation. Kelleher (2014) developed a method using bootstrapping techniques to reduce the number of simulation runs required to train a dynamic Bayesian network as a simulation meta-model. This approach was applied to an analysis of a cruise missile defense scenario using the System Effectiveness Analysis Simulation (SEAS) simulation framework.

2.6 Summary

Significant uncertainties prevent the development of predictive combat models. Several approaches have been published in the literature for proper use of combat models given these uncertainties. These approaches deemphasize quantitative use of simulation output and advocate for them as a tool for searching for regions of relative benefit. Other domains have employed more quantitative approaches to accounting for uncertainty. Specifically, the space and nuclear power communities have employed Evidence Theory as a tool for quantifying the impact of uncertain inputs on uncertainty in simulation output. These tools may provide a means for further understanding the relationships between the uncertainties in combat modeling inputs and the uncertainty in combat modeling outputs.

III. Analysis of Epistemic Uncertainty in Combat Modeling and Simulation

3.1 Introduction

Modeling and simulation has been applied to a wide variety of challenging problems in research, commercial industry and government. While drawbacks include lengthy development time, software licensing fees and limited pools of qualified practitioners, it is particularly well suited to problems where 1) experimentation with the real system is prohibitively expensive (presumably more so than the modeling and simulation effort itself) and 2) there is no other way in which to reasonably conduct the desired experiments. In the Department of Defense (DoD), models of combat have been employed by systems and operations research analysts since the 1960's for exploring possible outcomes of hypothetical military conflict (Davis, 1995). The fidelity of these models has ranged from simple mathematical relationships describing attrition between two opposing forces (Lanchester, 1914) to high fidelity operator or hardware in the loop simulations for exploration of detailed system configuration changes (Haase, 2014). Over time the employment of combat simulations has expanded within the DoD including, operations planning, requirements analysis, operational test and training.

Due to the infrequent and competitive nature of combat, several challenges present themselves when using simulation as a tool for analyzing combat. First, there is limited data with which to validate such analysis tools. While it may be possible to validate individual pieces of a combat simulation, such as the performance of a specific radar on a specific platform, to assess the integration of all mission aspects against current and future threats is a much more significant effort. Attempts have been made to validate combat models in aggregate with historical data (Schramm, 2012), but this is of little value as the models themselves require heavy modification to incorporate modern or future forces, requiring their own distinct validation. Secondly, there are many aspects of combat modeling that are highly uncertain and not knowable (an unresolvable uncertainty (Bankes, 1993)), such as the exact tactics, techniques and procedures of an adversary force in response to a blue force strike.

Within the combat modeling community there have been several suggestions for exploring uncertainties associated with the domain (Bankes, 1993; Davis, 2000; Dewar, 1996). However, the output of these approaches does not clearly communicate the uncertainties that are buried within the inputs. The purpose of this research is to improve representation of uncertainty in combat modeling and simulation through application of Evidence Theory. It is anticipated that employment of such techniques will enable rapid visualization of system uncertainties and aid in interpretation of simulation output.

Applications of Evidence Theory can be found in fields where there is near zero fault tolerance and uncertainty exists within the system, such as space launch and nuclear power plant design (Oberkampf, 2002). In contrast with classical probability theory, likelihood is assigned to sets (i.e. a range of parameter values) instead of being assigned to a probability density function. By explicitly defining ranges of uncertain input parameters and propagating them through a model, Evidence Theory bounds the true cumulative density function for a response by empirically developed cumulative plausibility functions (CPF, upper bound on probability) and cumulative belief functions (CBF, lower bound on probability).

This chapter is organized into five major sections: Introduction, Background, Probabilistic Analysis of Epistemic Uncertainty, Analysis of Epistemic Uncertainty via Evidence Theory, and Conclusion. The Introduction provides a general overview of the context for the research in exploration and quantification of epistemic uncertainty in combat modeling. This section includes a brief introduction and analysis of issues in the literature and a synopsis of research contributions. The Background section provides a brief overview of Lanchester's Equations and their history in Combat Modeling along with an introduction to Evidence Theory. Traditional and novel approaches to analyzing uncertainty in combat modeling are presented in the Probabilistic Analysis of Uncertainty and Analysis of Epistemic Uncertainty via Evidence Theory sections, respectively. For each, the general method, results, and discussion of efficacy are provided. In the Conclusions section, a summary of the research context, contributions, and future work are identified.

3.2 Background

3.2.1 Lanchester's Equations

Lanchester's Square Law (equations (25) and (26)) were developed to model combat between two homogeneous forces where both forces use aimed fire, target acquisition time does not depend on the number of targets, target acquisition time is factored into the firepower coefficients, and the firepower coefficients are constant over time (MORS, 1994).

$$\frac{dx}{dt} = -\alpha * y(t) \text{ where } x(0) = X_0 \tag{25}$$

$$\frac{dy}{dt} = -\beta * x(t) \text{ where } y(0) = Y_0 \tag{26}$$

The equations model the change of a given force level (say *x(t)*) as a function of the opposing force level (*y(t)*) over time, given initial force sizes and estimates of the firepower coefficients. Using these equations, various quantities of interest can be explored, such as: who wins the conflict, residual forces, and duration of conflict (Figure 10).

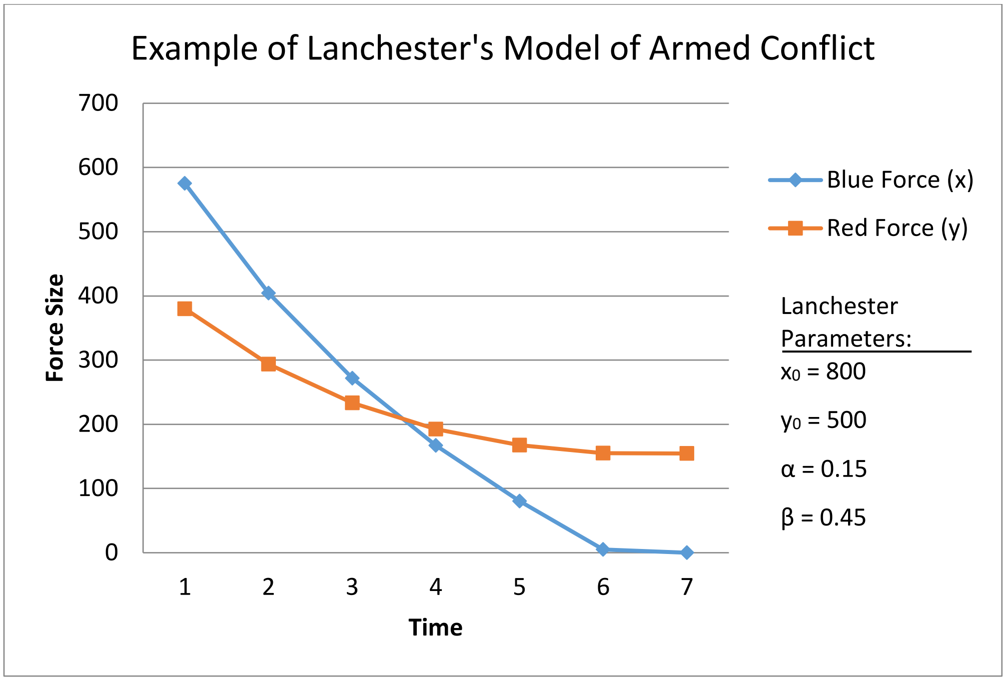

Figure 10: Example of Lanchester's Model of Armed Conflict

Lanchester's equations are appealing in part, due to their simplicity, transparency and ease of implementation. Combat analysts have employed them in assessments of several conflicts including; the Ardennes Campaign, the Battle of Kursk, and Iwo Jima (Bracken, 1995; Lucas, 2004; Schramm, 2012). However, there are numerous sources in the literature that identify deficiencies of a Lanchester model of armed conflict (Tolk, 2012). Taylor (1983) consolidates these into a single list, several of which are provided below as reference:

- No force movement
- Not verified by history
- No way to predict attrition rate coefficients
- Tactical decision processes not considered
- Battlefield intelligence not considered
- Command, control, and communications not considered
- Effects of terrain not considered
- Target priority/fire allocation not explicitly considered
- Noncombat losses are not considered

Despite these criticisms, many variants of Lanchester's equations have been developed. Extensions include incorporation of heterogeneous forces, stochastic attrition processes, reinforcements, logistics and maintenance and breakpoints (Tolk, 2012). In a more modern setting, the effects of network disruptions were represented as piecewise firepower coefficients (Schramm, 2012), enabling assessment of the impact of cyber effects on combat outcomes. Kelton (2010) describes a Lanchester model with stopping levels and stochastic reinforcements for both red and blue forces. This model was designed for implementation and analysis in the Arena modeling and simulation package.

3.2.2 Evidence Theory

There is a growing body of work in applying Evidence Theory, and specifically Dempster-Shafer (D-S) theory, as a framework for systematic exploration and quantification of uncertainty in modeling and simulation. The theory was first introduced by Dempster (1967) and later codified by Shafer (1976). Applications can be found in several scientific fields, such as space launch and nuclear power plant design (Sentz, 2002). Evidence theory differs from probability theory in that likelihood is assigned to sets (i.e. a range of parameter values) instead of being assigned to a probability density function. By explicitly defining ranges of uncertain input parameters and propagating them through a model, evidence theory bounds the true cumulative density function for a response. There are three key functions in Evidence Theory; the basic probability assignment function (*BPA* or *m*), the Belief function (*Bel*) and the Plausibility function (*Pl*).

In general the basic probability assignment is not equivalent to probability as discussed in classical probability theory (although connections exist (Sentz, 2002)). Similarly to classical probability theory, the basic probability assignment is a mapping of all sets (*X*, the power set) to the interval [0, 1] and the sum of all assignments across subsets is 1. Formally, this is represented as:

$$m: P(X) \rightarrow [0, 1] \tag{27}$$

$$m(\emptyset) = 0 \tag{28}$$

$$\sum_{A \in P(X)} m(A) = 1 \tag{29}$$

Using these basic probability assignments, upper and lower bounds for an interval can be calculated. The lower bound (or Belief), for a set *A* (subset of *X*), is the sum of all basic probability assignments of the proper subsets (*B*) of the set of interest (*A*).

$$Bel(A) = \sum_{B|B \text{ is a subset of } A} m(B) \tag{30}$$

The upper bound (or Plausibility) is the sum of all the basic probability assignments of the sets (*B*) that intersect the set of interest (*A*).

$$Pl(A) = \sum_{B|B \cap A \neq \emptyset} m(B) \tag{31}$$

The rules of combination in Evidence Theory allow data to be aggregated across multiple, potentially conflicting sources within a common frame of discernment. This process assumes that the sources are independent (Shafer, 1976), however this requirement is not rigorously established in practice (Tebaldi, 2007). There are many rules for combining evidence; the key to identifying the most appropriate method is to determine how conflict between sources should be considered. A survey of relevant combination rules is provided below.

Dempster's (1967) combination rule was the original combination operator that drove the conception of Evidence Theory. Using Dempster's combination rule, the basic probability assignments from two (or more) sources is combined with a purely conjunctive operation. The formal definition of this operation (m_{12}) is:

$$m_{12} = \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-K} \quad when\ A = \emptyset \tag{32}$$

$$m_{12}(\emptyset) = 0 \tag{33}$$

$$\text{where } K = \sum_{B\cap C=\emptyset} m_1(B)m_2(C) \tag{34}$$

It is important to note that this operation results in an aggregated mass only for intervals which overlap, giving zero mass to regions where evidence existed but did not overlap with another method. The measure of non-overlapping probability mass (or conflict) is represented by the computation of *K*.

The aggregated masses are normalized based on *K* to achieve a basic probability assignment function that resembles a probability density function from classical probability theory. Unfortunately, this choice can lead to counterintuitive results in situations involving high levels of conflict. These shortcomings are detailed in (Zadeh, 1984). Recognizing the potential pitfall, numerous other combination rules have been developed which account for level of conflict differently (see (Sentz, 2002) and (Yao, 1994) for many examples).

The Mixing Rule of combination is a popular mechanism for aggregation of disjunctive evidence (see equation (35)). This rule averages the masses (m_i) associated with a particular interval across all *i* estimates (*i* from 1 to *n*). The individual estimates can be weighted based on reliability by the multiplier, ***w***, where each w_i is the reliability associated with the *i*th source.

$$m_{1\ldots n}(A) = \frac{1}{n}\sum_{i=1}^{n} w_i m_i(A) \tag{35}$$

In contrast with Dempster's rule of combination, evidence in conflict is preserved in the resulting BPA. Said another way, the full range of possibilities expressed in the sources are represented in the final BPA. This feature is particularly beneficial where the application of evidence theory is not to identify the most likely distribution of a particular metric, but to express the full range the distribution could be.

3.3 Probabilistic Analysis of Epistemic Uncertainty

3.3.1 Methodology

In this section, we analyze a hypothetical military conflict modeled by Lanchester's Square Law with a stopping level (a force level for both red and blue where the conflict ends). For this model there are 6 key inputs; initial force levels, attrition coefficients, and stopping level for both red and blue sides. These inputs are presumed to be epistemically uncertain and that the analyst has been given ranges for each input (see Table 2). In reality it is unlikely that all input quantities will be epistemically uncertain, but were made so in this instance to introduce sufficient complexity to establish viability of uncertainty quantification approaches. In this situation, it is assumed that blue minus red residual forces is the primary quantity of interest.

Table 2: Summary of Uncertain Inputs for Lanchester's Model of Armed Conflict

Inputs		Uncertainty
X1	Blue Initial Force Size (X_0)	[900, 1000]
X2	Red Initial Force Size (Y_0)	[400, 500]
X3	Blue Attrition Coefficient (α)	[0.15, 0.25]
X4	Red Attrition Coefficient (β)	[0.3, 0.45]
X5	Blue Stopping Level	[0, 50]
X6	Red Stopping Level	[0, 25]

Bankes (1993) proposes two different modeling paradigms: exploratory vs. consolidative modeling. Consolidative modeling is the process of building a model by consolidating known facts into a single package and then using it as a surrogate for the actual system. In contrast with consolidative modeling, the exploratory modeling approach is the use of a series of experiments to explore the implications of assumptions when unresolvable uncertainties preclude building a surrogate for the system. The power of the consolidative approach lies in the assumption that the performance of the model has been compared to reality and the accuracy of the model is known to some precision. In situations where this is not feasible or where important facts about the system under study are uncertain, Bankes (1993) suggests that the exploratory modeling approach is preferable, and that treating such an endeavor as if it were a consolidative modeling effort is perilous.

To explore the impact of the proposed uncertainties on blue minus red residual forces, we employ the exploratory modeling approach as suggested in the literature. A sampling of the uncertain input distributions are propagated through the Lanchester model using a Monte Carlo approach with 1,000 runs. However, no distribution for the uncertain inputs was provided, so the problem becomes identifying the appropriate distribution to use. Significant literature references the principle of maximum entropy and that in the absence of another option, a uniform distribution should be chosen (Bankes, 1993). That convention is adopted here. Each uncertain input is assumed to be uniformly distributed with endpoints as specified in the uncertainty interval in Table 2.

3.3.2 Results

A histogram and cumulative distribution function of Blue minus Red residual forces is provided in Figure 11.

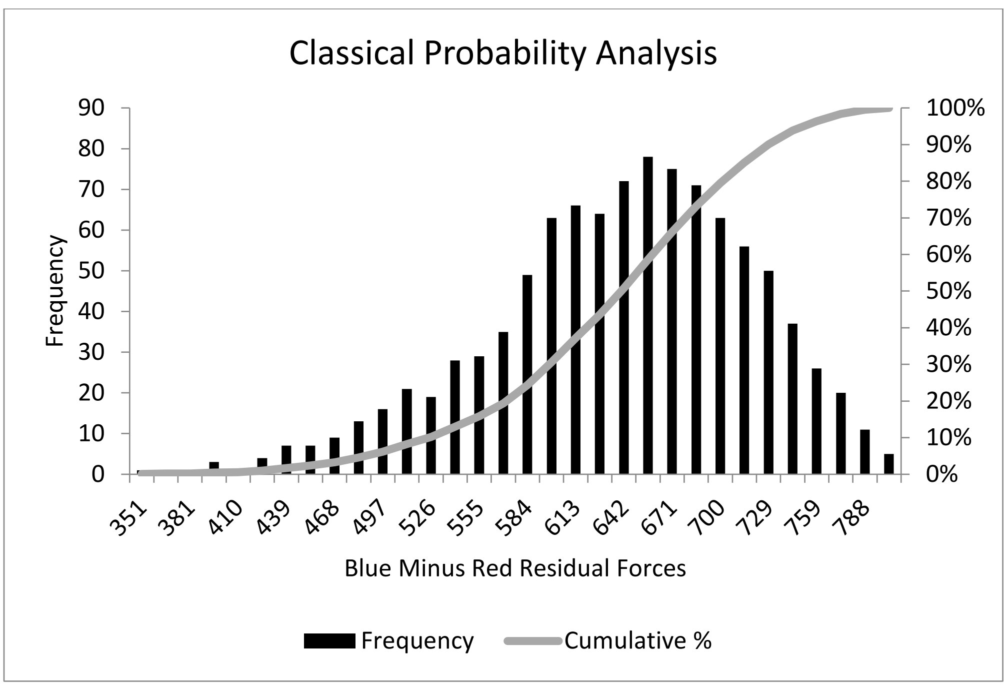

Figure 11: Propagation of Uncertainty via Classical Probability Theory

The Blue minus Red residual forces ranged from 351 to 802 units, with an average value of 633 units. Thus indicating that Blue overwhelmed Red and won the conflict across this sampling of the uncertain inputs.

3.3.3 Discussion

As an analysis of epistemic uncertainty, there are several ways the preceding analysis could have been embellished. While with the current uncertainty specification Blue always wins this conflict, the commander may have an objective value for the Blue minus Red residual forces that is tactically significant. If a commander was willing and able to specify a threshold for Blue minus Red for which the strategic objectives for the wider conflict are achieved, then those values could be reported. A sampling of some notional thresholds and resulting probability of successfully achieving that objective is provided below (Table 3).

Table 3: Probability of Achieving Blue minus Red Objective

Blue – Red Threshold	**Probability of Meeting Objective**
750	0.053
650	0.450
550	0.853
450	0.978

Additionally, there are various statistical intervals that may be of interest. Confidence intervals on the mean response are a popular way to describe simulation output. Confidence intervals for the mean Blue minus Red residual forces were computed for confidence levels of 0.99, 0.95 and 0.90 as included in Table 4. While these measures are useful for describing the central tendency of the simulation output, they do a poor job of characterizing the dispersion of the data (although the confidence intervals do capture some measure of variance).

Table 4: Confidence Intervals on the Mean of Blue minus Red Residual Forces

Confidence Level	**Confidence Interval**
0.99	[626.53, 639.44]
0.95	[628.07, 637.89]
0.90	[628.86, 637.10]

Additionally, the Blue minus Red residual forces could have been analyzed using ANOVA or regression techniques. This would have resulted in estimates of the uncertain factors impact on Blue minus Red residual forces. As this was a random sample, there would likely be correlation among input factors that may confound the analysis. This could be circumvented by employment of design of experiments techniques. There are many classes of experimental designs that would be suitable in aiding analysis of uncertainty, to include factorial and space filling designs. Employment of these designs would ensure that estimates of factor effects would not suffer from correlation among inputs.

Any analysis that is executed for this hypothetical combat situation should be completed in a manner that considers the difficulties in validating the model and the subsequent reliability in any predictive quantities. A simple presentation of the histogram as the representation of possible outcomes or confidence intervals on the mean blue minus red residual forces, in this case, is less than satisfactory. These methods emphasize the absolute quantitative information instead of the relative comparison, which is contrary to the weakly predictive nature of the model. In this way, an ANOVA (or similar analysis) procedure is preferable as relative factor effects can be identified and prioritized. Such an approach emphasizes the absolute outcomes and concentrates on general themes that can be extracted from the data.

Each of the factors in this analysis represents an epistemically uncertain quantity, meaning that there exists a true value for a given factor but that it is unknown. Since there exists a true value for this unknown quantity, it is not epistemically correct to model the factors as random variables or stochastic processes that can be simply described with means and variances. A more appropriate model for epistemic uncertainty is the range in simulation output that its possible values can take on (Ferson, 2006).

There are numerous difficulties and inconveniences that are encountered when analyzing epistemically uncertain quantities impact on simulation output considering both range and variance. It is this shortcoming that is remedied in the following section.

3.4 Analysis of Epistemic Uncertainty via Evidence Theory

3.4.1 Methodology

Evidence Theory provides a statistical framework for aggregating multiple, potentially conflicting estimates for a simulation input factor. The uncertainty in the simulation inputs is then mapped to the simulation output by feeding the input intervals through the model (Figure 12). This process generally requires the formation of basic probability assignments for uncertain factors, the development of a design matrix to collect data from the simulation, execution of the simulation runs, estimation of simulation output uncertainty intervals, and, finally, the computation of cumulative plausibility and belief functions. In this section, the Lanchester formulation and uncertainty intervals from Section 3.3 are used as a mechanism to explore Evidence Theory representation of epistemic uncertainty in blue minus red residual forces.

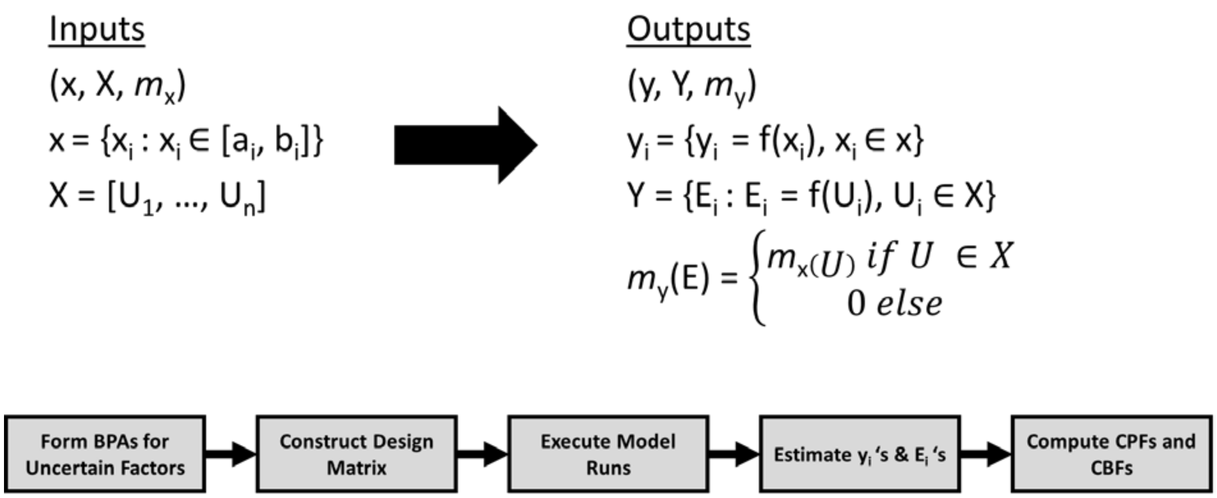

Figure 12: Method for Propagating Uncertain Inputs through Simulation

Formally, the evidence space for an input is *(x, X, m_x)*. Where *x* is the set of all possible values for that input, *X* is the set of subsets of *x* (U_i's) that represent the interval estimates for the input, and m_x is the vector of masses associated with each element of *X* (U_i). The evidence space for the corresponding simulation output is *(y, Y, m_y)*. Where y is the set of all possible values for that output, *Y* is the set of subsets of *y* (E_i's) that represent the interval estimates for the output, and m_y is the vector of masses associated with each element of *Y* (E_i). In the case of uncertain model inputs, all that is known is *X* (U_i's) and *Y* (E_i's) is not. The E_i's are thus properly considered estimates based on propagation of a corresponding U_i, and identifying the minimums and maximums produced by that input interval. The mass for a given E_i *(m_y)* is assigned based on the corresponding U_i. If a given U_i produced an E_i, its mass (m_x) becomes the new mass (m_y) associated with the new E_i.

In Evidence theory, uncertainty in the inputs is represented with basic probability assignment functions resulting from the aggregation of multiple inputs. For this demonstration we assume that three experts were asked “What is the firepower coefficient of the opposing force in 2055?”. The three plausible responses could be (Figure 13):

- Expert 1: “It is certainly between 0 and 0.75.”
- Expert 2: “Hard to say, but most likely between 0.25 and 0.85.”
- Expert 3: “Based on detailed analysis, I believe it follows this distribution: {[0.0.25]: 0.2, (0.25, 0.5]: 0.4, (0.5, 0.75]: 0.3, (0.75, 1]: 0.1}”

The word expert implies that these are subjective inputs, but it is worth mentioning that the “experts” could be outputs of other simulations, estimates derived from distinct observations, etc. The only requirement is that the individual estimates be independent. In the author’s experience, this setting is not unrealistic, especially with respect to future operational scenarios with adversaries whose systems are not completely understood.

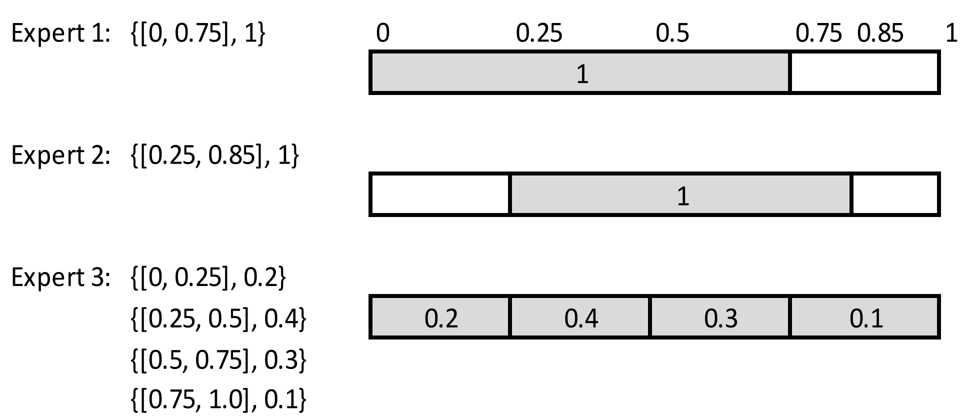

Figure 13: Summary of Expert Input

For this comparison of uncertainty representation with basic exploratory analysis procedures and Evidence Theory, all six uncertain factors in the Lanchester model are represented with the same scaled BPA represented in Figure 12. BPAs were constructed with both the Dempster and the Mixing rules of combination to illustrate differences between conjunctive and disjunctive combination rules. The final design matrix and estimation of E_i's were generated using inputs from the BPA formed with the Mixing rule of combination.

There is a direct link between constructing the design matrix for the simulation effort and how the analyst intends to estimate y_i's and E_i's. To estimate the E_i's, the maximum and minimum simulation outputs must be estimated for each intersection of the Ui's across each factor. For this six factor experiment with six U_i's for each factor, the resulting experimental design contains 6^6 (or 46,656) interval intersections for which the maximum and minimum simulation response must be estimated. Two possible approaches to estimating the y_i's and E_i's would be to use a meta-model or enumerating a large number of possibilities explicitly with the simulation. Neither approach guarantees that the global maximum or minimums have been found, which is why these quantities are frequently referred to as estimates for E.

In either case, the analyst must choose a sampling procedure. The literature contains many examples of the sampling based approached, including random sampling, factorial experiments, and space filling designed experiments (Ankenman, 2012; Kleijnen, 2006). However, if a meta-model is to be used, care should be taken to ensure that the resulting model is statistically valid. To avoid this difficulty and to demonstrate the concept of Evidence Theory as a viable method for uncertainty representation in combat modeling, an enumeration approach was selected. The E_i's were estimated by taking the maximum and minimum from running the factorial combination of the endpoints of the U_i's resulting in 6^6*2^6 (2,985,984) simulation runs. Finally, the cumulative plausibility and belief functions (CPF and CBF respectively) were computed using equations (30) and (31).

3.4.2 Results

The expert input provided in Figure 3 was aggregated per the Dempster's Rule of Combination and the Mixing Rule of Combination as in equations (32) and (35). The resulting BPA, Cumulative Plausibility Function, and the Cumulative Belief Functions are shown in Figure 14 and Figure 15. Figure 16 is the plot of the Cumulative Belief and Plausibility functions resulting from propagating the uncertain inputs through the Lanchester model of conflict supposed in section 3.4.1.

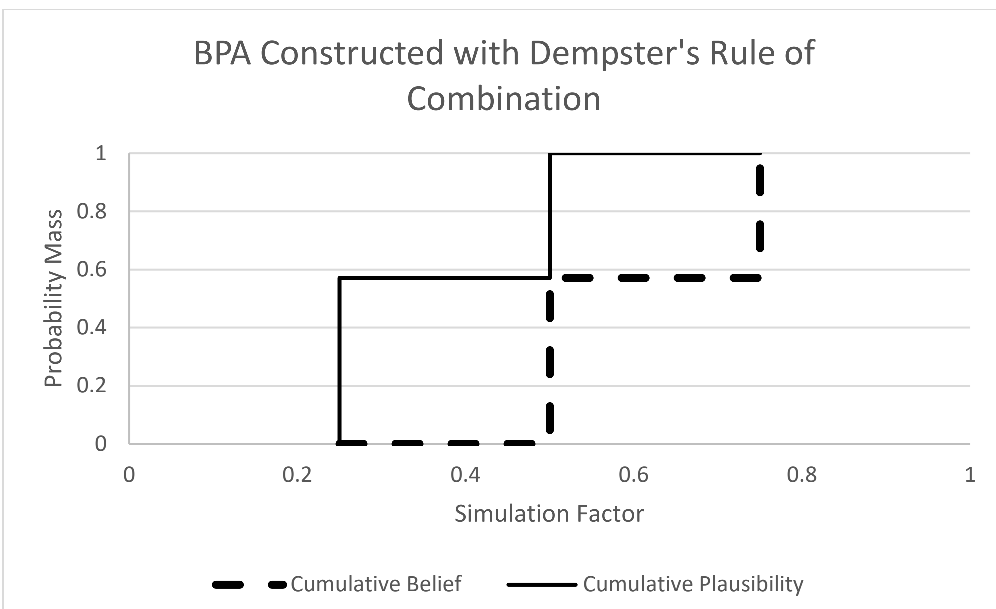

Figure 14: BPA Constructed with Dempster's Rule of Combination

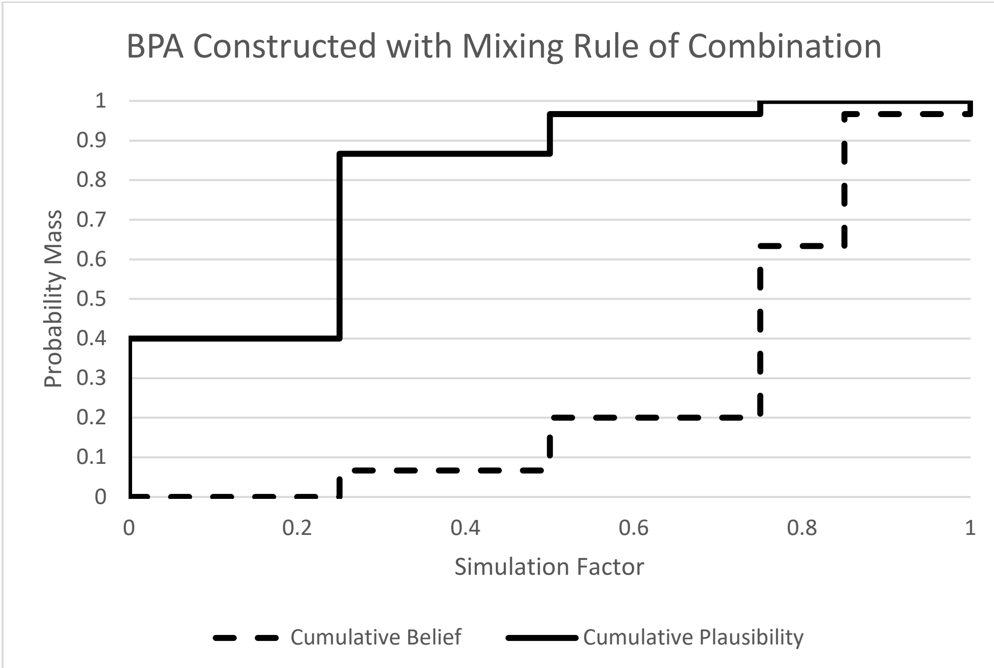

Figure 15: BPA Constructed with Mixing Rule of Combination

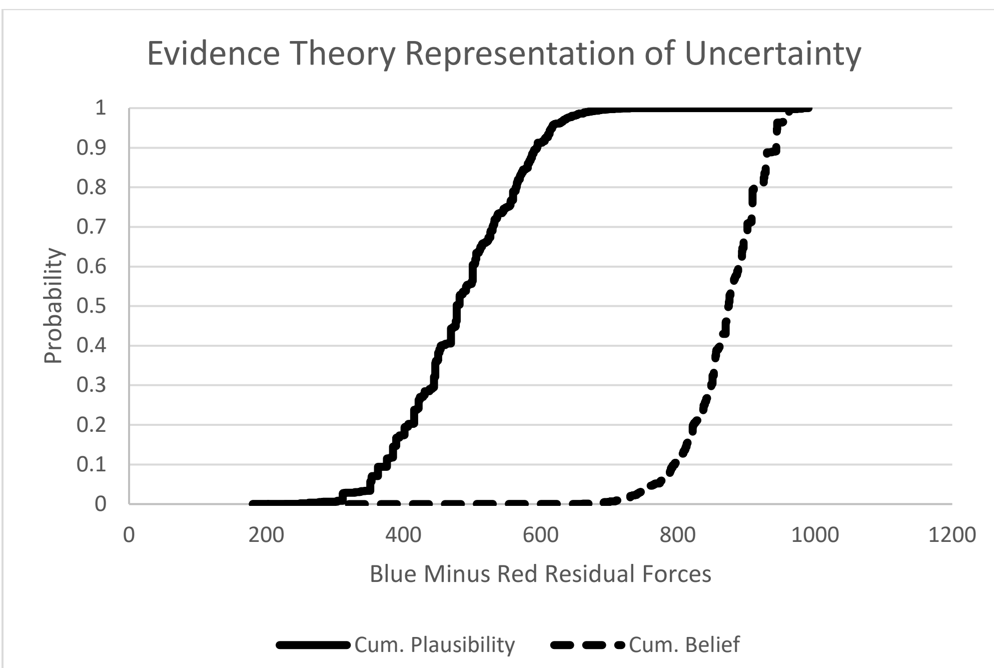

Figure 16: Evidence Theory Representation of Combat Model Uncertainty

3.4.3 Discussion

Upon comparison of the outputs from the traditional probabilistic uncertainly analysis outlined in Section 3.3 and the analysis of epistemic uncertainty with Evidence Theory, there are several noteworthy differences. The first significant difference is the manner in which the inputs for the simulation were collected. In the probabilistic analysis, the inputs were aggregated in a way that resulted in a single range for each uncertain factor, which encompassed the widest possible set of values from subject matter experts. This approach discards information about where the expert's opinions overlapped. This overlap in inputs is leveraged in Evidence Theory, by using a combination operator to create a probability assignment distribution for the factor. Using this process there is no need for the analyst to make judgments about the quality or validity of an individual estimate, and all relevant inputs can be aggregated in a transparent manner. For this demonstration, the hypothetical subject matter expert input was aggregated using both Dempster's and the mixing rules of combination (see Figure 17).

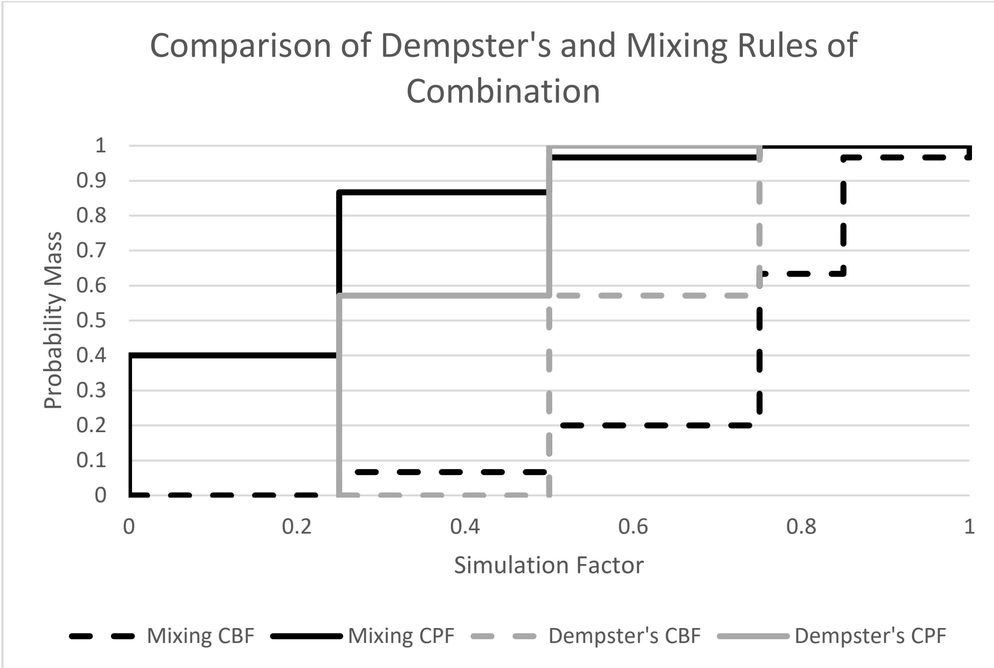

Figure 17: Comparison of Dempster's and the Mixing Rules of Combination

Dempster's rule of combination is a conjunctive operator and, as expected, truncated the range of the expert input to the space over which the factor estimates overlapped. In contrast, the mixing rule of combination is a disjunctive operator and preserved the full range of expert input. While one of the benefits of an Evidence Theory representation of uncertainty in combat modeling is the repeatable and transparent method for consolidating multiple, conflicting sources, the relative weight of each source could be adjusted. These adjustments result in different BPA's for the factors. Figure 18 plots the CBF and CPF for three different combinations of weights (***w*** from 3.2.2) for the three expert inputs used, where each w_i is the reliability associated with the *i*th source (and w_i's sum to one).

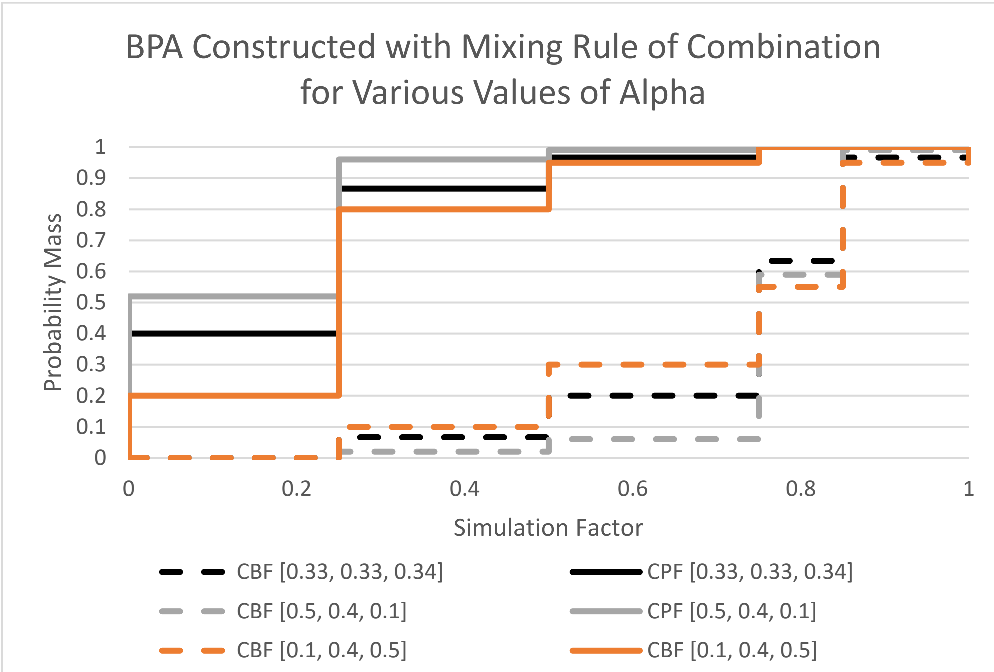

Figure 18: BPAs for Various Relative Weights across Expert Inputs

Another key difference between the probabilistic and Evidence Theory representations of uncertainty is the presentation and interpretation of summary statistics. The output of the Evidence Theory analysis is a set of two functions; the CBF and CPF, compared to the single output for the probabilistic analysis (see Figure 19).

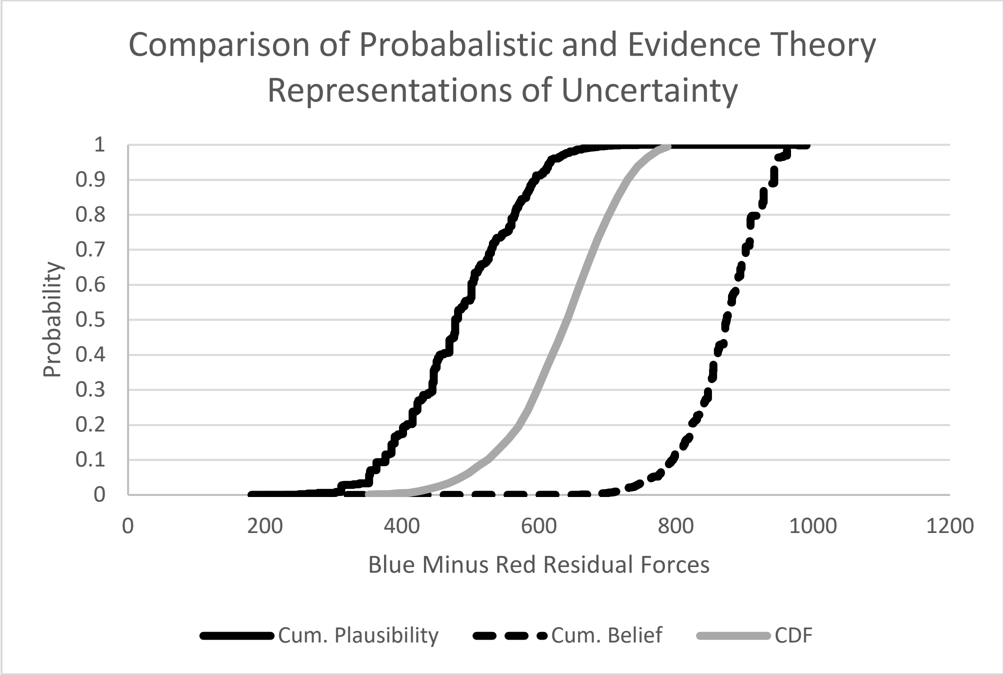

Figure 19: Comparison of Probabilistic and Evidence Theory Representations of Uncertainty

As described in the literature, the CPF and CBF are bounds on the CDF from the probabilistic analysis. In a similar manner to the analysis in Section 3.3, key summary statistics can be extracted from the probability plots except that the outputs are ranges instead of single values. For example, the mean (50^{th} percentile) in the Evidence Theory representation is reported as the range [477, 874]. This corresponds to the 50^{th} percentile value from the CPF and CBF respectively. Table 5 contains a report of various other percentiles of interest. Reporting of summary statistics as ranges provides a direct representation of uncertainties in model inputs to consumers of this analysis of the impact of uncertainties on simulation output, mitigating the propensity to interpret simulation output as absolute.

Table 5: Plausibility and Belief Percentiles

Percentile Level	Range
95th	[617, 944]
75th	[551, 908]
50th	[477, 874]
25th	[422, 838]
5th	[351, 765]

Additionally, there is an Evidence Theory analog to the analysis in section 3.3.3. If a commander was willing and able to specify a threshold for Blue minus Red for which the strategic objectives for the wider conflict are achieved, then the range of probabilities associated with that threshold could be identified. A sampling of some notional thresholds and resulting probability of successfully achieving that objective is provided below for both the baseline and Evidence Theory based analysis of uncertainty (Table 6). The baseline and Evidence Theory representations of the probability of meeting the commander's thresholds for success are very different. First, the resulting probabilities from the Evidence Theory analysis are ranges. In some cases these ranges are quite large, 0.95 and 0.97 for the 750 and 650 Blue minus Red thresholds respectively. In all cases the Evidence Theory ranges encompass the probabilities provided by the baseline analysis.

Table 6: Probability of Achieving Blue minus Red Objective

Blue – Red Threshold	Probability of Meeting Objective - Baseline	Probability of Meeting Objective – Evidence Theory
750	0.053	[0.01, 0.96]
650	0.450	[0.02, 0.99]
550	0.853	[0.25, 1]
450	0.978	[0.64, 1]

It is likely the commander's reaction to each analysis would be entirely different, depending on which analysis they were presented. Suppose the commander believes that having a Blue minus Red residual force of 650 units is strategically significant and is presented with the baseline analysis which indicates that the probability of achieving success is 0.45. Seeing that this is roughly a coin flip (and that his troops are confident they can win), he may proceed with the planned battle.

Now suppose the same commander is presented with the Evidence Theory analysis which indicates the probability of achieving success is between 0.02 and 0.99. One response could be that they take the average of the range and proceed as in the baseline analysis. Alternatively, the commander may pause and arrive at the conclusion that his analysts don't have enough information to distinguish between near certainty (0.99) that they will achieve their objectives or near certainty that they will fail (0.02). At this point, the commander may either develop another plan that is not sensitive to a Blue minus Red threshold of 650 in this battle or gather more information that can be used to reduce the uncertainty in the ability of their forces to achieve the objective.

Using this framework does not address the complicating factors which prevent model validation in a combat analysis and the subsequent limits in quantified predictive capability. Representation of uncertainties inherent to combat modeling using Evidence Theory improves representation of uncertainty by providing ranges instead of point values for metric output. Using this framework should reduce the propensity to overlook these shortcomings when reviewing analysis output and inspire a more thoughtful dialog on the causes of the range in possible outcomes observed during the study.

A direct analog for the ANOVA analysis discussed in Section 3.3 has been developed in Helton (2006) and Guo (2007). This work and an extension are detailed in Chapter IV.

3.5 Conclusion

In this chapter, Evidence Theory was demonstrated as a framework for representing epistemic uncertainty in combat modeling output. The steps for aggregating multiple, conflicting sources for simulation input data were demonstrated. A basic probability assignment was assumed for six uncertain factors; initial force size, attrition coefficient and stopping level for both Blue and Red forces in a Lanchester model of conflict. The analysis found that the proposed uncertainty configuration induced a large gap between the cumulative plausibility and belief functions for Blue minus Red residual forces, indicating large uncertainty in combat outcomes (although Blue always won!). To provide context for the Evidence Theory analysis, a traditional approach was employed in assessment of uncertainty of Blue minus Red residual forces in a Lanchester model of conflict. The results of both analyses were compared and contrasted.

The demonstration of Evidence Theory as a framework for representing outcomes in combat modeling and simulation addresses several key gaps in the literature and common practice. First, is the propensity to treat combat simulation output as predictive when, upon examining what is known and unknown regarding the model inputs, it clearly is not. This is addressed by supplementing the single output probability density functions with cumulative belief and plausibility functions from evidence theory. These functions represent bounds on probability densities given an input uncertainty specification (or basic probability assignment). Common summary statistics (i.e. mean, probability intervals, etc.) are in the form of ranges, which discourage the propensity to treat point estimates from a simulation as predictive.

Second, is that the employment of the Evidence Theory rules of combination eliminates the need to make choices about how to use multiple, potentially conflicting sources for modeling and simulation inputs. In the presence of multiple inputs, traditional approaches typically follow one of two lines of thought: 1) condense the sources into a single point estimate or 2) employ the Laplace principle of maximum entropy and assume a uniform distribution over the range of possible values. The validity of these approaches is heavily influenced by the process by which the sources are condensed to either a point estimate or range. In practice, these methods are unstructured and not well documented. In contrast, Evidence Theory provides a structure for aggregating multiple, conflicting sources in a repeatable manner without making assumptions regarding the distribution of the true value of the input.

Implementing these methods does not come without cost. Managing multiple sources and choosing the appropriate Evidence Theory rule of combination add additional complexity to the overall analysis. The need to estimate the maximum and minimum response in each input's basic probability assignment interval intersections can drive significant increases in the number of required simulation runs. Large run matrices can be mitigated through the employment of design of experiments and meta-modeling. These costs are offset by the clarity provided to the decision maker regarding how uncertainties in modeling and simulation inputs affect the analysis outcomes.

IV. Sensitivity Analysis of Uncertainty in Combat Modeling

4.1 Introduction

Due to the infrequent and competitive nature of combat, several challenges present themselves when using simulation as a tool for analyzing combat. First, there is limited data with which to validate such analysis tools. While it may be possible to validate individual pieces of a combat simulation, such as the performance of a specific radar on a specific platform, to assess the integration of all mission aspects against current and future threats is a much more significant effort. Attempts have been made to validate combat models, in aggregate, with historical data (Schramm, 2012), but this is of little value as the models themselves require heavy modification to incorporate modern or future forces, requiring their own distinct validation. Secondly, there are many aspects of combat modeling that are highly uncertain and not knowable (an unresolvable uncertainty (Bankes, 1993)), such as the exact tactics, techniques and procedures of an adversary force in response to a blue force strike.

Within the combat modeling community there have been several suggestions for exploring uncertainties associated with the domain (Bankes, 1993; Davis, 2000; Dewar, 1996). However, the output of these approaches does not clearly communicate the uncertainties that are buried within the inputs. In Chapter III, Evidence Theory was demonstrated as a framework for representing epistemic uncertainty in combat modeling output. The steps for aggregating multiple, conflicting sources for simulation input data were demonstrated with a simple Lanchester model incorporating six uncertain factors. The analysis found that the proposed uncertainty configuration induced a large gap between the cumulative plausibility and belief functions for Blue minus Red residual forces, indicating large uncertainty in combat outcomes.

This framework addresses several key gaps in the combat modeling and simulation literature and common practice. First, is the propensity to treat combat simulation output as predictive when it is not. This is addressed by representing simulation output with cumulative belief and plausibility functions from evidence theory, instead of the classical single probability density function. These functions represent bounds on probability densities given an input uncertainty specification. Common summary statistics (i.e. mean, probability intervals, etc.) are in the form of ranges, which discourages the tendency to treat point estimates from a simulation as predictive.

Second, is that the employment of Evidence Theory rules of combination eliminates the need to make choices about how to reconcile multiple, potentially conflicting sources for modeling and simulation inputs. In the presence of multiple inputs, traditional approaches typically follow one of two lines of thought: 1) condense the sources into a single point estimate or 2) employ the Laplace principle of maximum entropy and assume a uniform distribution over the range of possible values expressed by the experts. The validity of these approaches is heavily influenced by the process by which the sources are condensed to either a point estimate or range. In practice, these methods are unstructured and not well documented. In contrast, Evidence Theory provides a structure for aggregating multiple, conflicting sources in a repeatable manner without making assumptions regarding the distribution of the true value of the input.

Use of Evidence Theory as an analytical framework does not directly address the complicating factors which prevent model validation in a combat analysis and the subsequent limits in quantified predictive capability. Literature addressing this topic has suggested the adoption of exploratory approaches, where inputs are structured as a designed experiment or monte carlo sample. This sample is then internally analyzed to identify significant factors through statistical techniques like regression, ANOVA, etc. Exploratory activities produce prioritized contributors to simulation output, and deemphasize the magnitude of the simulation output variables.

The objective of this research is to extend the recent work in representation of uncertainty in combat modeling and simulation, by developing a sensitivity analysis method for identifying the factors which contribute to the overall uncertainty in simulation output. Existing approaches to this problem implement a variance based metric for factor prioritization (Helton, 2006), which requires a second analysis, or is susceptible to ties when large uncertainties are present (Guo, 2007). In a resource constrained environment, such methods would facilitate prioritization of resources with the goal of reducing uncertainty in system performance.

This chapter is organized into five major sections: Introduction, Background, Methodology, Results and Discussion, and Conclusion. The Introduction provides a general overview of the context for the research in sensitivity analysis of belief functions in Evidence Theory. This section includes a brief introduction and analysis of issues in the literature and a synopsis of research contributions. The Background section provides a summary of sensitivity analysis in modeling and simulation, sensitivity analysis of belief functions in Evidence Theory, and measures of distance in Evidence Theory. A novel approach to identifying and quantifying the impact of input uncertainty on total system output uncertainty is presented in the Methodology section. This method is demonstrated using a Lanchester model of conflict. The results of this analysis and a discussion of interesting features from this application is provided in the Results and Discussion section. In the Conclusions section, a summary of the research context, contributions, and future work are identified.

4.2 Background

4.2.1 Sensitivity Analysis of Belief and Plausibility

There has been some suggestion that analysis using Evidence Theory is sensitivity analysis (Ferson, 2006). In the sense that analysis with Evidence Theory identifies bounds on true outcome probabilities, this is reasonable. However, often what is meant in the modeling and simulation community by sensitivity analysis is a process that produces the relative, absolute or rank ordered effects for a set of variables on a measurable parameter. This type of product is not the direct result of propagating basic probability assignments (BPAs) of uncertain factors through a simulation and producing the cumulative plausibility function (CPF) and cumulative belief function (CBF) of an important measure of system performance. A specific set of techniques are required to translate the ideas of sensitivity analysis in classical probability theory to the framework of Evidence Theory.

In classical probability theory, important variables explain the largest portion of the variation in system output. In contrast, variables that are important in explaining uncertainty in Evidence Theory have the largest effect in reducing the area between the CBF and CPF. Several attempts to produce rank ordering of variables explaining uncertainty in Evidence Theory have been documented in the literature with two specific approaches summarized in the subsequent paragraphs.

Helton (2006) presented a first of its kind paper outlining three methodologies for sensitivity analysis in conjunction with Evidence Theory representations of uncertainty. The first method generically entailed using a stratified sampling procedure to generate data which could then be fit with a statistical modeling technique to identify the most critical factors *x* on outcome *y*. Latin Hypercube or random sampling were suggested as viable sampling procedures, while rank regression and squared rank differences were provided as suitable statistical modeling techniques. Where this method differs from most classic design of experiments and simulation efforts is that the sample is weighted by the respective factors' BPAs. They called this procedure exploratory sensitivity analysis.

Additionally, a stepwise procedure for construction of CBFs and CPFs was presented. The intent of this method was to enable epistemically uncertain variables to be added one-at-a-time until the CPFs' and CBFs' rate of change slowed to within an acceptable range, saving computation time. The methodology is as follows (Helton, 2006):

- Step 0: Perform Exploratory sensitivity analysis to determine the most important factors, $\hat{x}_1, \ldots, \hat{x}_n$, on the uncertainty in output, *y*. Where $\hat{x}_1$ is the most important variable, $\hat{x}_2$ is the next most important variable, etc. and *n* is the number of variables or factors.

- Step 1: Estimate a CPF and a CBF for *y* on the basis of the evidence space obtained from the original evidence space for the $\hat{x}_1$ and degenerate evidence spaces for all other variables (in which the sample spaces are assigned BPAs of 1).

- Step 2: Estimate a CPF and a CBF for *y* on the basis of the evidence space obtained from the original evidence space for the $\hat{x}_1$ and the $\hat{x}_2$ and degenerate evidence spaces for all other variables (in which the sample spaces are assigned BPAs of 1).

- Step s: Estimate a CPF and a CBF for *y* on the basis of the evidence space obtained from the original evidence space for $\hat{x}_1, \ldots, \hat{x}_s$ and degenerate evidence spaces for all other variables (in which the sample spaces are assigned BPAs of 1).

- Termination: End when no significant difference between the CBF_{s-1} and CPF_{s-1} obtained at Step *s-1* and CBF_s and CPF_s obtained at Step *s*.

This method relies on the fact that as epistemically uncertain variables are added to the computation of the CBF and CPF, the bounds they represent can only stay the same or decrease.

The final sensitivity analysis method proposed by Helton (2006) was called the summary sensitivity analysis procedure. This method decomposes the variance of outcome *y* into contributions by the individual x_i's variances. As there are many possible distributions for each x_i that are consistent with its evidence space, a sampling of assumed distributions for each x_i must be explored as part of this method. This results in not one prioritized list of sensitivities, but a set of prioritized lists. This method was demonstrated on failure probabilities of a set of actuators. For this analysis a number of different distribution assumptions were made, and the resulting decomposition was found to be invariant to choice of distributions. As this is an empirical result, there are no guarantees that this holds in general.

Guo and Du (2007) developed a one-at-a-time approach to quantify the effect of each individual variable on the uncertainty in their system response. Their method produces a series of CBFs and CPFs, where each one is generated by fixing all but one variable to an average, weighted by the factors' BPA and leaving the remaining as an epistemically uncertain variable with corresponding BPA. The distance between the CBF and CPF for a given variable, as calculated by the Kolmogorov–Smirnov (K-S) statistic, is then used to prioritize the relative importance of individual variables. This method was demonstrated on two example problems from engineering mechanics; the crank-slider mechanism and a crowned cam roller/follower's contact.

The K-S distance is the maximum difference between two CDFs (Banks, 2010), or in the context of the sensitivity analysis procedure described by Guo (2007), the difference between the CPF and CBF. This is an admirable choice in that the prioritization metric does not rely on a variance based metric (where Evidence Theory is primarily concerned with ranges) as in Helton (2006). However, in situations where there is high uncertainty and subsequently large differences between CPFs and CBFs, discriminating among the important factors may prove difficult.

Figure 20 represents a set of plausible CBFs and CBFs for six epistemically uncertain factors *x* on outcome *y*. In this configuration, significant uncertainty exists where there is a large range where the value of each CPF is 1 and the value of each CBF is 0, resulting in a K-S distance of 1 for all factors (highlighted by the red arrow in Figure 20). Referring back to the original notion of sensitivity analysis in Evidence Theory being interested with factors that minimize the area between the CPF and CBF being the most important, it is clear that variables *X5* and *X6* are not in that category. Their CBFs and CPFs are outside the bounds of all other variables, yet the K-S distance does not identify that feature.

It is not entirely clear how many runs are required to run the analysis proposed by Guo (2007). The authors reference *2n+1* as the number of iterations of uncertainty analysis required, but do not specify how many runs are required to execute each iteration. Per their own admission, this approach is expected to be computationally intensive and potentially inefficient as each uncertain variable is varied one-at-a-time. In contrast, the stepwise construction of CBFs and CPFs proposed by Helton (2006) does not require any additional simulation runs when epistemically uncertain factors are varied together.

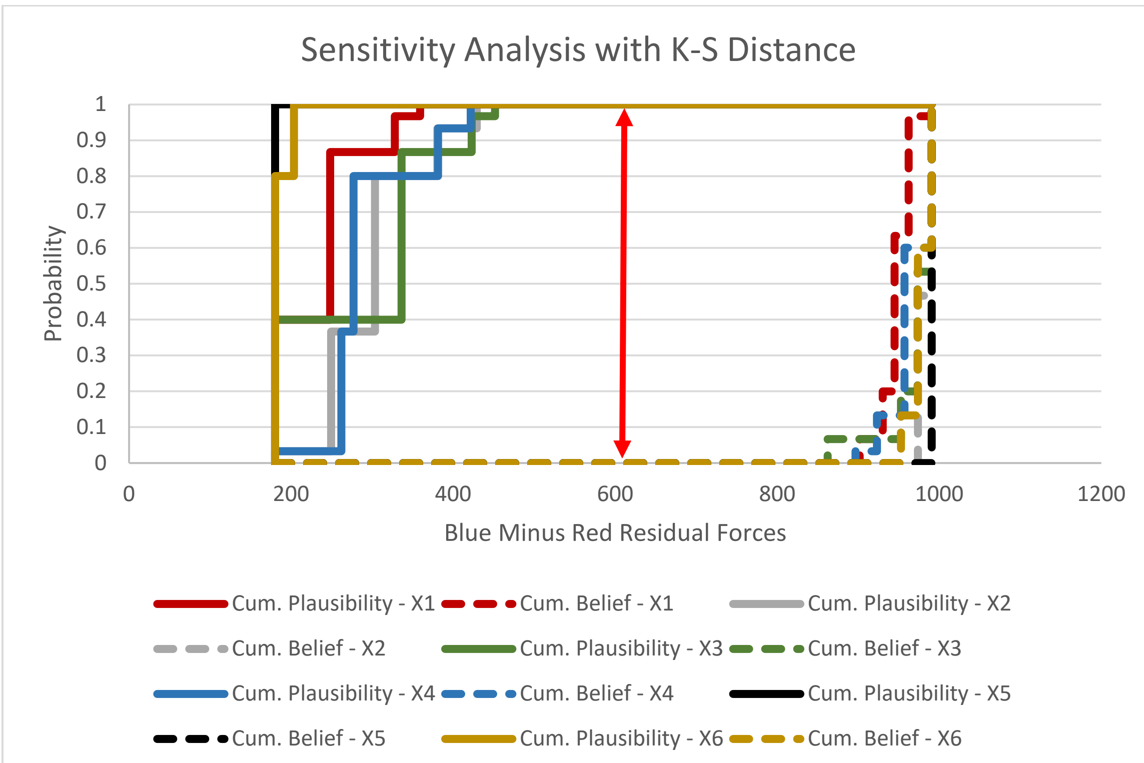

Figure 20: Ties using K-S Distance when Large Uncertainties Exist

4.2.2 Measures of Distance in Evidence Theory

Measures of distance or similarity have been developed and employed in many areas of statistical analysis, including object classification, statistical comparison of distributions, etc. Their primary goal is to provide a measure for how different two points, vectors or bodies of evidence are. They are usually constructed such that small distances indicate the two points or objects are “close” or “similar”, while large distances mean the two are “far apart” or “dissimilar”.

The formal definition of a distance metric is a measure that satisfies the four properties listed in Table 7; Non-negativity, Symmetry, Definiteness (grouped with Reflexivity and Separability) and the Triangle Inequality. The non-negativity property states that the distance between two points, in the space where the measure is defined, is always greater than or equal to zero. A distance metric must also produce the same value, regardless of the direction the distance is calculated. This is the symmetry property. The definiteness property states that the distance between two identical points is zero (reflexivity) and that when the distance is zero, the two points must have been identical (separability). Finally, the triangle inequality is the property that the distance between two points is less than or equal to the sum of the distances between each of these points and a third point.

Not all measures used in Evidence Theory satisfy all four properties. As previously stated, a true distance metric satisfies all four properties listed in Table 7. Semi-metrics satisfy all properties except the triangle inequality. Quasi-metrics are not symmetric. Pseudo-metrics are not separable, and thus not definite. Quasi-pseudo metrics only satisfy the symmetric and reflexive properties, while pre-metrics only satisfy the reflexive property. A non-metric (not show in Table 7), is a measure which does not satisfy any of the four properties.

Table 7: Axioms for metrics (adapted from Jousselme, (2012))

		Metric	Semi-Metric	Quasi-Metric	Pseudo -Metric	Quasi-Pseudo Metric	Pre-Metric
(i) Non-negativity	d(y, z) ≥ 0	X	X	X	X	X	X
(ii) Symmetry	d(y, z) = d(z, y)	X	X		X	X	
(iii) Definiteness	d(y, z) = 0 ↔ y = z	X	X	X			
(iii)’ Reflexivity	d(y, y) = 0	X	X	X	X	X	X
(iii)’’ Separability	d(y, z) = 0 → y = z	X	X	X			
(iv) Triangle Inequal.	d(y, z) ≤ d(y, t) + d(t, z)	X		X	X		

Applications of distance metrics in Evidence Theory are largely centered on finding approximations to belief functions, quickly and automatically (Han, 2018). Large numbers of estimates for a parameter value can produce BPAs which contain many foci. Propagating these foci through a model can be time consuming, even with a small (6 or less) number of estimates. The goal of this body of work is to reduce the number of elements in a belief function while maintaining the general integrity of the true function.

There are a large number of distance metrics available for use in analysis with Evidence Theory. Jousselme (2012) provides a survey of many distance measures in Evidence Theory, their properties, and relative family from which they are derived. The Minkowski family is a series of distances which includes the evidence theory analogs to the Manhattan, Euclidean and Chebyshev distances. Another way to measure the difference between two belief functions is to estimate the difference in their information content. This family contains the Evidence Theory equivalent to the Kullback-Liebler divergence from classical probability theory.

Additional families of distance measures in Evidence Theory include inner products, cosine measures, and the fidelity family of metrics.

4.3 Methodology

4.3.1 Wasserstein Distance as a Metric for Sensitivity Analysis of Uncertainty

There are two key features in a procedure for sensitivity analysis in Evidence Theory; the stepwise construction of CPFs and CBFs and the construction of a metric with which to compare the contributions of the variables in explaining the uncertainty in outcomes. The following section develops the second feature, with a comprehensive methodology produced in section 4.3.2.

The intuitive notion behind sensitivity analysis in Evidence Theory is to identify which variables reduce the area between the CPF and CBF and by how much (Ferson, 2006). This concept is more intuitive when compared against the concept of least information, or where all the BPAs are degenerate in Evidence Theory. When all BPAs are degenerate, the CPF and CBF will appear as the plot in Figure 21, where *a* is the minimum of the data and *b* is the maximum.

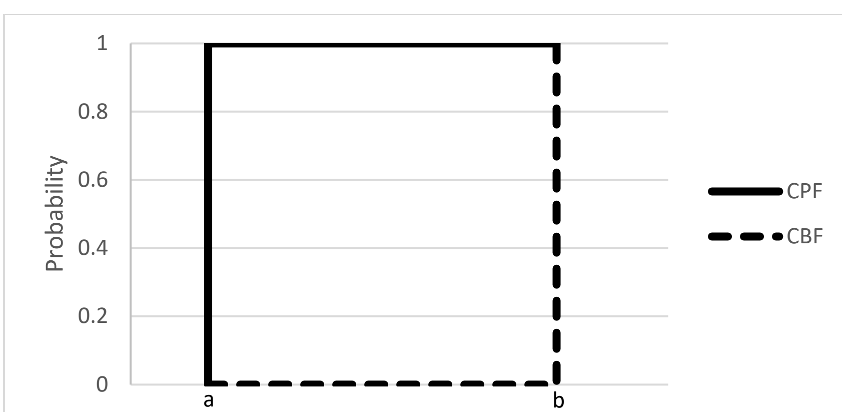

Figure 21: Concept of “Maximum Uncertainty”

From the proof in Helton (2006), as more variables with non-degenerate BPAs are added to the estimate of the CPF and CBF, the area between the two functions will always decrease.

In classical probability theory, the area between two cumulative density functions is known as the Wasserstein metric (Rüschendorf, 2001),

$$W_p(\mu, v) \coloneqq \inf_{\gamma \epsilon \Gamma(\mu, v)} \left(\int_{M x M} d(x, y)^p d_\gamma(x, y) \right)^{1/P} \tag{36}$$

where *(M, d)* is a metric space and for $p \geq 1$, $P_p(M)$ is the collection of all measures μ, v on *M* with finite p^{th} moment. This measure is also known as the Kantorovich or Earth Movers distance in the computer science community.

In our setting, we can greatly simplify this expression for the 1 dimensional case, resulting in the following expression (Rüschendorf, 2001):

$$W_1(\mu, v) = \int_{-\infty}^{\infty} \left| F_\mu(x) - G_v(x) \right| dx. \tag{37}$$

This measure satisfies the requirements for non-negativity, symmetry, definiteness, and triangle inequality, qualifying it as a true metric in the framework presented in 4.2.2. Also, assuming the body of evidence results in BPAs that meet the criteria for probability density functions, the Evidence Theory and classical probability theory representation of this metric are equivalent. In Evidence Theory F_{μ} and G_{v} represent the CBF and CPF, respectively, for a given variable.

4.3.2 *Method for Prioritizing Variables Impact on Uncertainty*

This method for sensitivity analysis in Evidence Theory is motivated by prior work developing a stepwise procedure to generate marginal CBF and CPFs and the use of statistical differences to develop a ranking of the variables impact on uncertainty. It is intended that this methodology follows an analysis executed according to the procedures discussed in Chapter III. These procedures detail the process to aggregating multiple inputs, propagating them through a model or simulation, and generating the CPF and CBF of the resulting measure of interest. Once the data is generated, the methodology is as follows:

1. Let $\Phi = \{1, \ldots, n\}$ be the set of all variable indices under consideration for this analysis (where *n* is the number of variables) and $\Omega = \{\quad\}$.
2. Iteration *k*: for each variable $x_i, i \in \Phi$;
   a. Estimate a CPF and a CBF for *y* on the basis of the evidence space obtained from the original evidence space for the x_i, the original evidence space(s) for any $x_j, \forall j \in \Omega$, and degenerate evidence spaces for all other variables (in which the sample spaces are assigned BPAs of 1).
   b. Calculate W_{1i} between the marginal CPF and CBF for variable *i*.
3. Select variable $x_s, s \in \Phi$, that minimizes $W_{1.}$. Remove *s* from set Φ and add to set Ω.

   Let $\hat{x}_k = x_s$, where $\hat{x}$ is an ordering of *x* and $\widehat{W}_{1k} = W_{1s}$.
4. Increment k and repeat steps (2) and (3) until $\Phi = \{\quad\}$.

This method is notable in several respects. First, it employs the stepwise construction of CBFs and CBFs as descried in Helton (2006) and a statistical measure of distance as in Guo (2007). Guo (2007) employs the K-S distance which has been demonstrated to be susceptible to ties when there are large uncertainties in outcomes with respect to the variables. The Wasserstein distance, however, is not likely to cause ties in the procedure except when the variables under analysis have very little relationship to the uncertainty in model outcomes, where they exhibit marginal CPFs and CBFs similar to Figure 21.

A significant difference between the method proposed by Helton (2006) and the one presented here, is that no preliminary exploratory sensitivity analysis procedure is required. Such a procedure is good practice, but not integrally linked in the new method. A modification of this procedure could be produced where it is used (as in Helton (2006)) to incrementally construct estimates of CBFs and CPFs by adding variables until the functions stop changing enough to warrant further computation.

4.3.3 Methodology Demonstration

To demonstrate the methods from sections 4.3.1 and 4.3.2, a hypothetical military conflict modeled by Lanchester's Square Law with stopping level (a force level for both red and blue where the conflict will end) for both red and blue forces is analyzed. For this model there are 6 key inputs; initial force levels (blue – *X1*, red – *X2*), attrition coefficients (blue – *X3*, red – *X4*), and stopping level for both blue (*X5*) and red (*X6*) sides. These inputs are presumed to be epistemically uncertain and that the analyst has been given ranges for each input. In this situation, it is assumed that blue minus red residual forces is the primary quantity of interest.

All six uncertain factors in the Lanchester model are represented with the same scaled BPA represented in Figure 22. The factor BPAs were formed using the Mixing Rule of combination. The final simulation run matrix was constructed to support an enumeration approach for calculating the interval estimates of the output. This overall experimental setup is exactly the same as described in Chapter III. As such; a more detailed explanation of the Lanchester model used can be found in sections 3.2.1 and 3.4.1 and more details regarding the construction of the BPAs and design matrix for this demonstration can be found in section 3.4.1. Once the simulation runs were complete, the marginal sensitivity of the model factors and overall contribution to the belief and plausibility functions were computed as described in sections 4.3.1 and 4.3.2

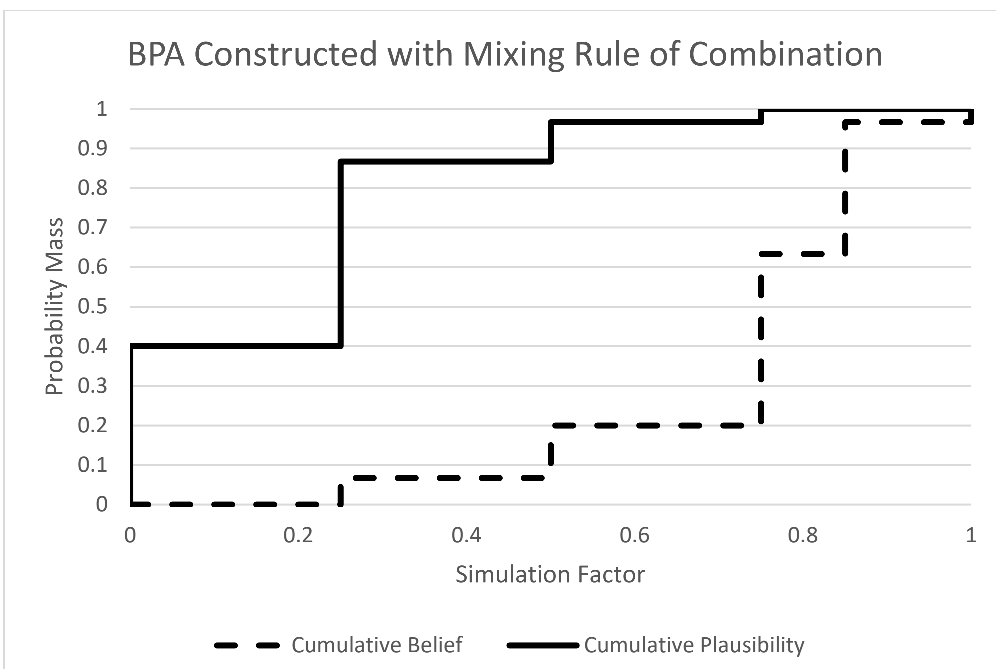

Figure 22: Notional BPA for Representing Uncertain Factors

4.4 Results and Discussion

The methodology and application were performed as described in section 4.3. The first iteration of the method resulted in six distinct sets of cumulative plausibility and belief functions (CBFs and CPFs), one for each epistemically uncertain variable in the model (Figure 23). Upon inspection, it appears that variables five and six explain the least amount of uncertainty. It is difficult to explain (by inspection) which variables explain the most uncertainty, as there is no pair of CBFs and CPFs that dominate all other sets. To clarify which variable explains the most uncertainty, the Wasserstein distance between the marginal CBFs and CPFs were computed for each variable in Table 8. These scores ranged from 810.13 to 672.53, with *X5* (blue stopping level) explaining the least uncertainty and *X4* (red attrition coefficient) explaining the most. To conclude step one, *X4* was selected as the first variable to enter the basis for the estimates of the CBF and CPF.

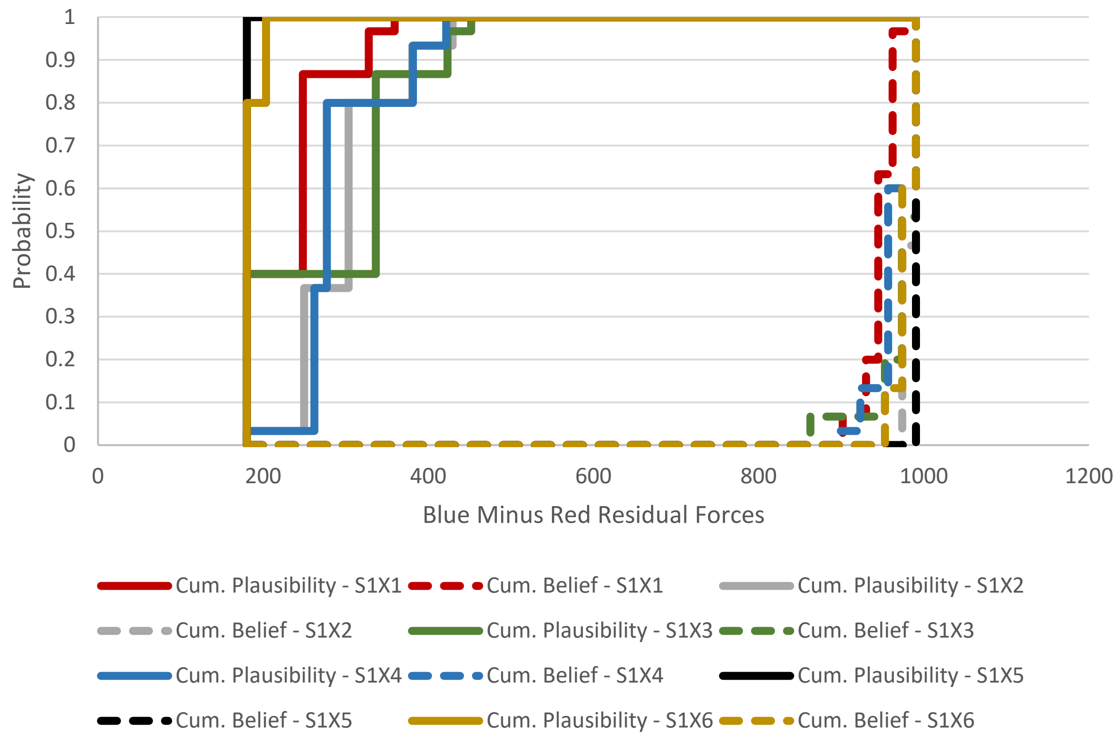

Figure 23: Sensitivity Analysis of Blue minus Red Residual Forces – Step 1

Table 8: Sensitivity Analysis – Step 1

Summary of Wasserstein Distance	
	Step 1
Initial Blue Forces (X1)	714
Initial Red Forces (X2)	683
Blue Firepower Coefficient (X3)	685
Red Firepower Coefficient (X4)	**673**
Blue Stopping Level (X5)	810
Red Stopping Level (X6)	793

In step 2, five sets of CPFs and CBFs were produced. Each was a combination of *X4* and the remaining variables non degenerate BPAs (Figure 24). Again, visual inspection reveals no obvious choice for explaining the most uncertainty, while combinations containing *X5* and *X6* appear to explain the least, as in step 1. The Wasserstein distances between the marginal CBFs and CPFs were computed for each variable (Table 9). These scores ranged from 672.53 to 570.37, with *X5* (blue stopping level) explaining the least uncertainty and *X3* (blue attrition coefficient) explaining the most. To conclude step two, *X3* was selected as the next variable to enter the basis for the estimates of the CBF and CPF.

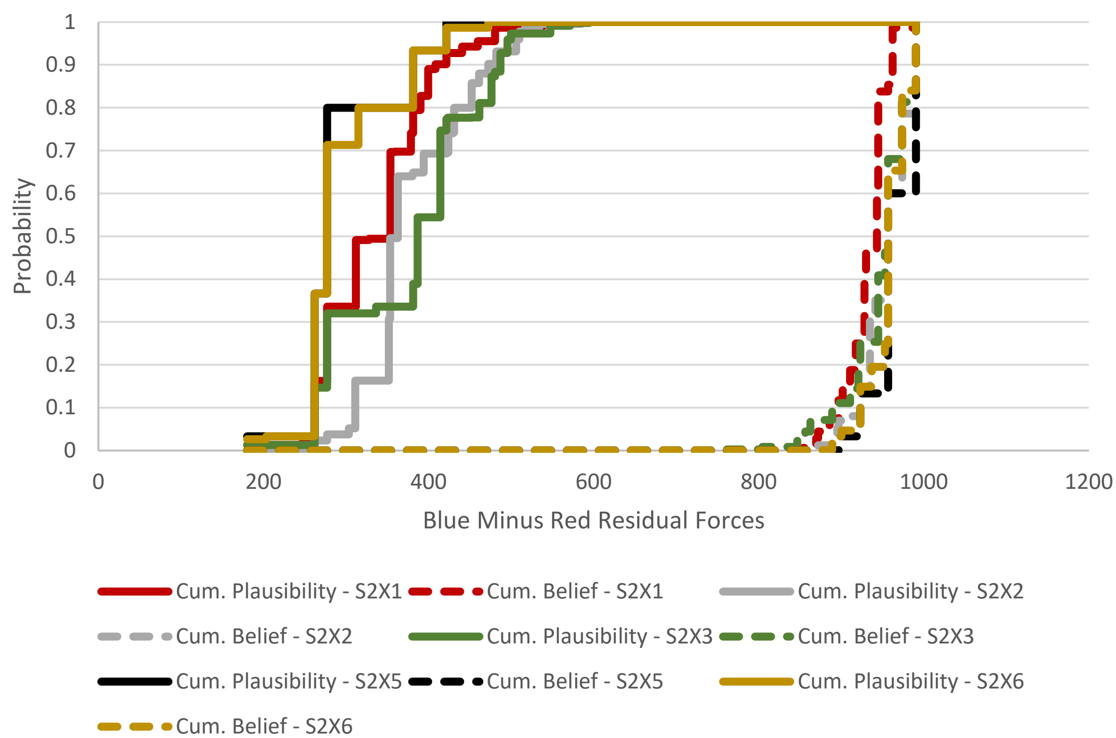

Figure 24: Sensitivity Analysis of Blue minus Red Residual Forces – Step 2

Table 9: Sensitivity Analysis – Step 2

Summary of Wasserstein Distance	
	Step 2
Initial Blue Forces (X1)	597
Initial Red Forces (X2)	574
Blue Firepower Coefficient (X3)	**570**
Red Firepower Coefficient (X4)	-
Blue Stopping Level (X5)	673
Red Stopping Level (X6)	661

The same procedure produced equivalent plots and Wasserstein distances for steps 3 through 6. At the conclusion of each step, the variable which had the smallest Wasserstein distance between their marginal CPFs and CBFs was added to the basis for the overall CPF and CBF. See Figure 25, Figure 26, Figure 27, and Figure 28 for plots of the marginal CPF and CBFs. Table 10 provides a summary of the Wasserstein distances at each step of the sensitivity analysis. The ordered importance of the variables on total uncertainty was red attrition coefficient (*X4*), blue attrition coefficient (*X3*), red initial force level (*X2*), blue initial force level (*X1*), red stopping level (*X6*) and blue stopping level (*X5*).

Table 10: Sensitivity Analysis Summary

Summary of Wasserstein Distance						
	Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**	**Step 6**
Initial Blue Forces (X1)	714	597	499	**401**	-	-
Initial Red Forces (X2)	683	574	**476**	-	-	-
Blue Firepower Coefficient (X3)	685	**570**	-	-	-	-
Red Firepower Coefficient (X4)	**673**	-	-	-	-	-
Blue Stopping Level (X5)	810	673	570	476	401	**386**
Red Stopping Level (X6)	793	661	560	458	**386**	-

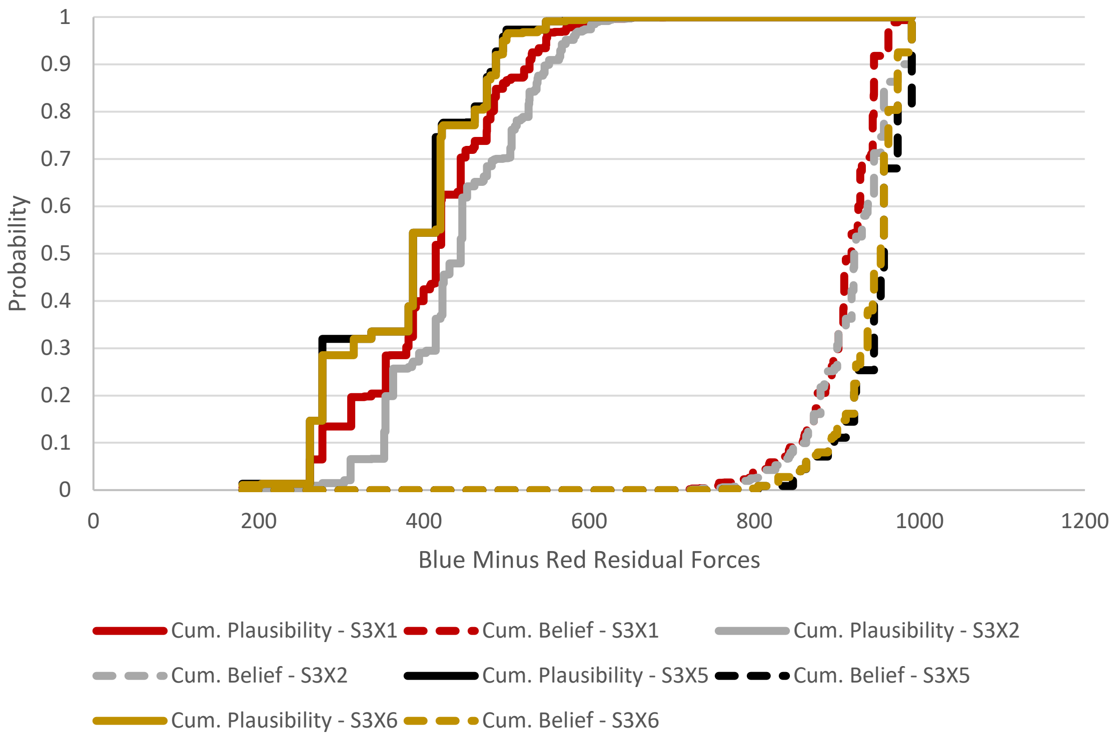

Figure 25: Sensitivity Analysis of Blue minus Red Residual Forces – Step 3

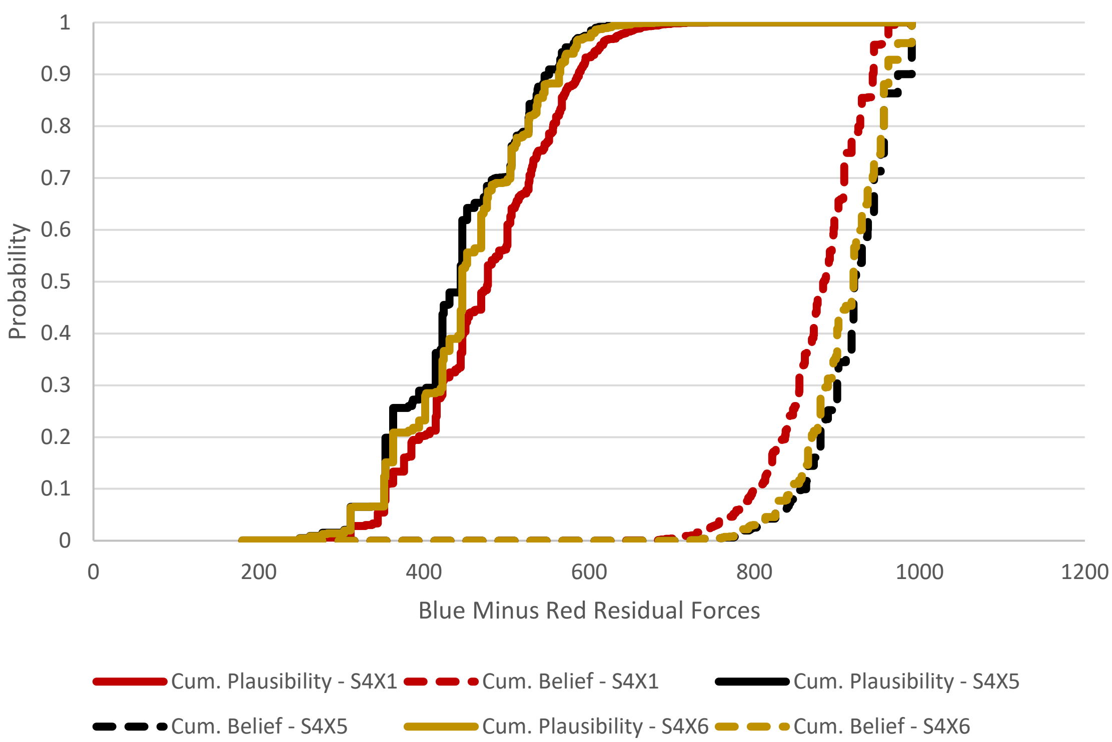

Figure 26: Sensitivity Analysis of Blue minus Red Residual Forces – Step 4

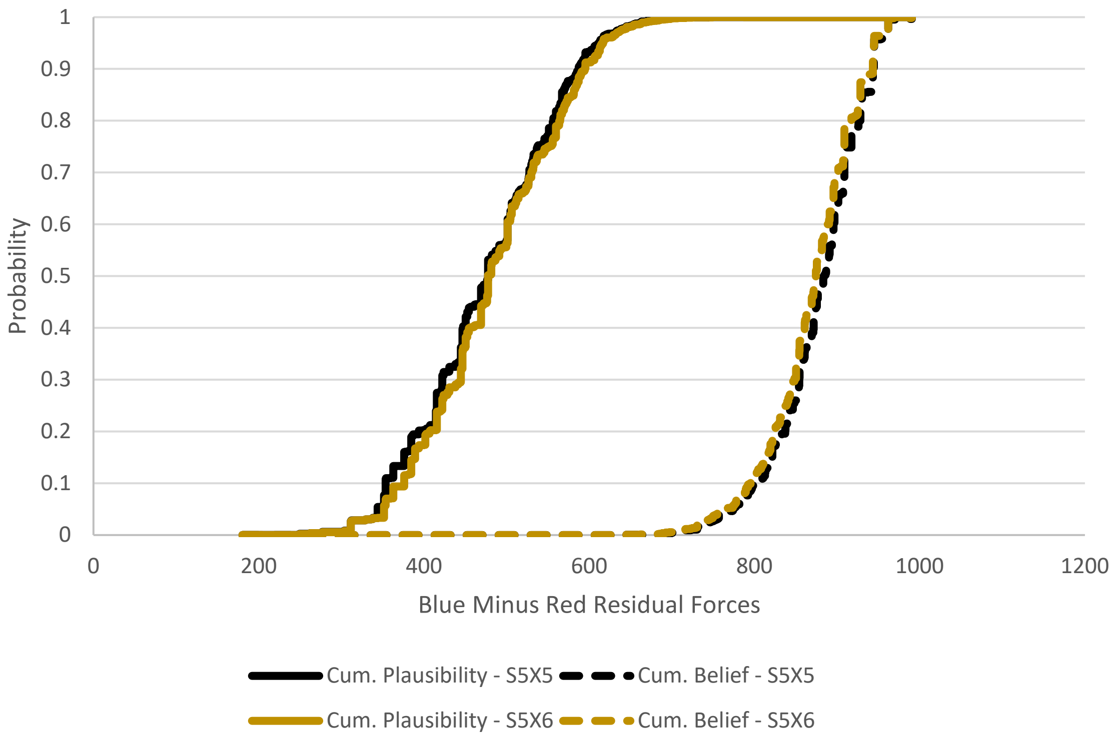

Figure 27: Sensitivity Analysis of Blue minus Red Residual Forces – Step 5

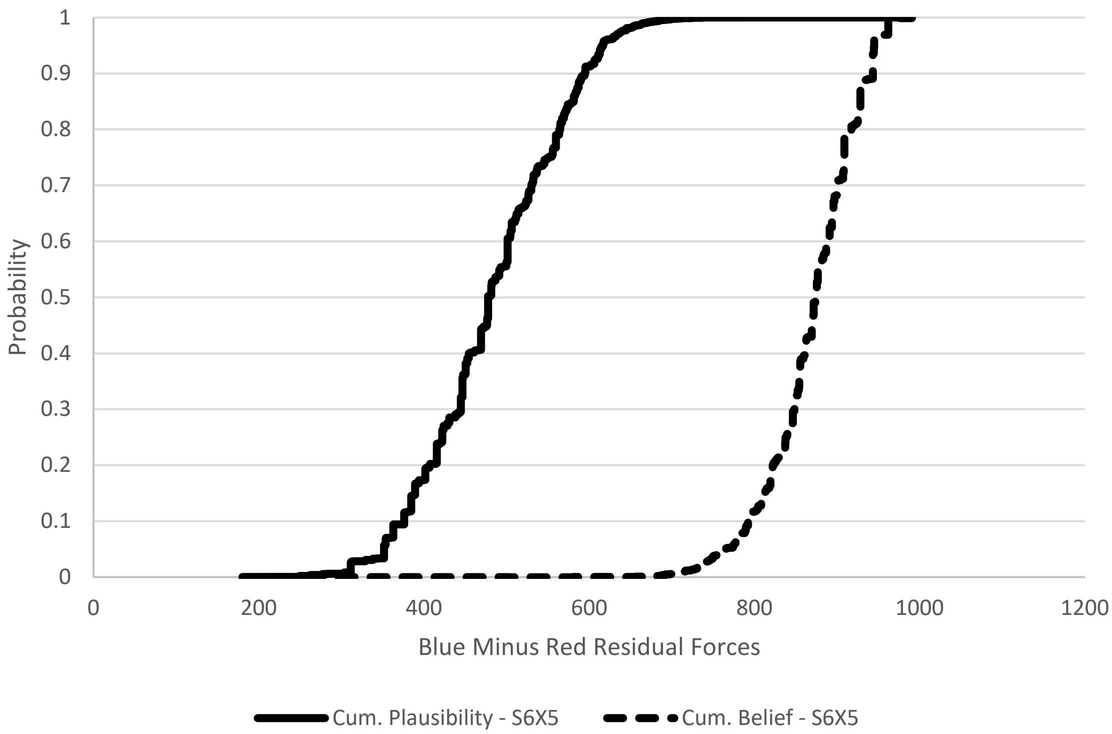

Figure 28: Sensitivity Analysis of Blue minus Red Residual Forces – Step 6

A number of model specific and method specific findings were identified as a result of this analysis. First, the Lanchester Model parameters were ordered in their explanation of total uncertainty. The Lanchester Model utilized in this demonstration had three key parameters to describe both the red and blue forces; initial force level, attrition coefficient and stopping level. In this analysis both attrition coefficients (red then blue) were the first variables added to the model, then initial force size, then stopping level. It is also interesting to note that the red variables were always first to enter the basis for the CPFs and CBFs. There is no particular explanation for this, except that it may be due to the specific uncertainty ranges used for this study. Other ranges may produce different results.

Second, the variable *X5* does not explain any uncertainty associated with the output of our model with respect to the range of inputs analyzed. The resulting marginal CPF and CBF selected in step 5 is the same as the final, full CPF and CBF. This could have been evident at the first step when it was observed that the marginal CPF and CBF resulted in no change in explained uncertainty, i.e. maximum uncertainty (see Figure 21).

The CPFs and CBFs at the end of each step are either equal to or contained within the bounds of prior steps' CPFs and CBFs (see Figure 29). This is consistent with the result in Helton (2006), where the Wasserstein distance between two CPFs and CBFs and can only decrease or remain the same when adding variables. Also, the ordering of the Wasserstein distances among the variables at a given step is not necessarily preserved across steps. For example, in step 1 the marginal CPF and CBF with the second smallest Wasserstein distance was *X2* (red initial force level). But the variable with the smallest Wasserstein distance in step 2 was variable *X3* (blue attrition coefficient), not *X2*. The number of foci increased with each step, resulting in smoother marginal CPFs and CBFs. Assigning degenerate BPAs for all factors but one in step 1, resulted in CPFs and CBFs with at most 6 foci (or steps), step 2 produced CPFs and CBFs with at most 6^2, step 3; 6^3, etc.

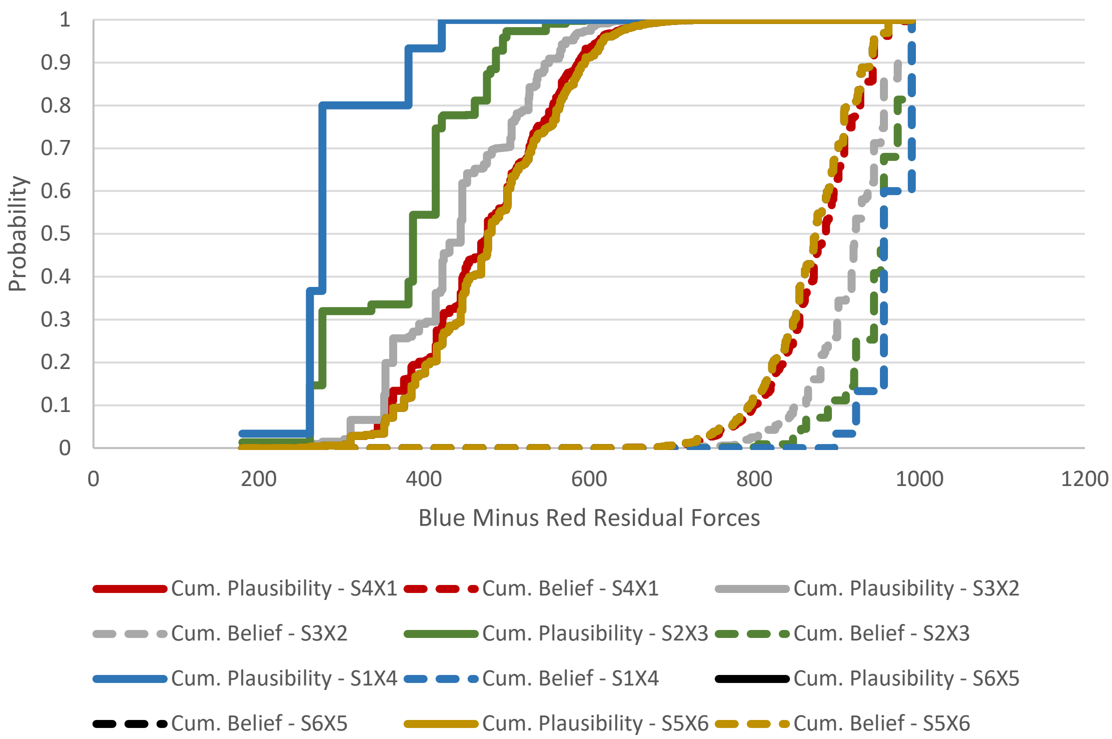

Figure 29: Sensitivity Analysis of Blue minus Red Residual Forces – Summary

4.5 Conclusion

In this chapter, a new method for sensitivity analysis of uncertainty in Evidence Theory was developed. This sensitivity analysis method generates marginal CPFs and CBFs and prioritizes the contribution of each factor by the Wasserstein distance (also known as the

Kantorovich or Earth Mover's distance) of the CBF and CPF. Using this method, a rank ordering of the simulation input factors can be produced.

This method is notable in several respects. First, it combines positive elements from the Evidence Theory literature; the stepwise construction of CBFs and CBFs and a statistical measure of distance. The new method improves on existing work that employs the K-S distance, which has been demonstrated to be susceptible to ties when there are large uncertainties in outcomes with respect to the variables. The Wasserstein distance, however is not likely to cause ties in the procedure except when the variables under analysis have very little relationship to the uncertainty in model outcomes, where they exhibit a large difference between marginal CPFs and CBFs.

A significant difference between the method proposed in Helton (2006) and the one presented here, is that no preliminary exploratory sensitivity analysis procedure is required. Such a procedure is good practice, but not integrally linked in the new method. A modification of this procedure could be produced where it is used to incrementally construct estimates of CBFs and CPFs by adding variables until the functions stop changing enough to warrant further computation.

The method was demonstrated on a notional Lanchester model of conflict with six epistemically uncertain parameters. A relative prioritization of the factors was produced, where five of six factors had distinct contributions in explaining total uncertainty while a sixth did not. The ordered importance of the variables on total uncertainty was red attrition coefficient, blue attrition coefficient, red initial force level, blue initial force level, red stopping level and blue stopping level. While the specific results of this analysis are not extensible, they are a useful example for how to apply sensitivity analysis with Evidence Theory in modeling, simulation and analysis activities.

The costs of employing this method are similar to those of generally employing Evidence Theory in analysis. Additional complexity to the overall analysis is induced via management of multiple input sources and choosing the appropriate Evidence Theory rule of combination to summarize the body of evidence. The need to recalculate CBFs and CBFs adds computation time, which is not insignificant for large numbers of uncertain variables with complex BPAs. These costs are offset by the clarity provided to the decision maker regarding the sensitivity of analysis outcome uncertainties to uncertain inputs.

V. Error Estimation via Multi-Model Methods for Analysis of Combat

5.1 Introduction

Models of combat have been employed by systems and operations research analysts since the 1960's for exploring possible outcomes of hypothetical military conflict (Davis, 1995). The fidelity of these models has ranged from simple mathematical relationships describing attrition between two opposing forces (Lanchester, 1914) to high fidelity operator or hardware in the loop simulations for exploration of detailed system configuration changes (Haase, 2014). Over time the employment of combat simulations has expanded within the DoD to support processes including, operations planning, requirements analysis, operational test, and training.

Due to the infrequent and competitive nature of combat, several challenges present themselves when using simulation as a tool for analyzing combat. First, there are many aspects of combat modeling that are highly uncertain and not knowable (an unresolvable uncertainty (Bankes, 1993)). This drives the analyst to fill knowledge gaps with best guesses for critical parameters, processes, tactics, and future force mix, when representing red and blue forces. This in of itself does not doom the utility of analysis of combat with modeling and simulation, but coupled with the fact that there is limited data with which to validate such tools, there is no way to determine an absolute bound on the impact assumptions have on simulation output in an uncertain environment. It may be possible to validate individual pieces of a combat simulation, such as the performance of a specific radar on a specific platform. However, assessing the predictive capability of a simulation across even a modest subset of mission aspects against current and future threats is impossible. Attempts have been made to validate combat models in aggregate with historical data (Schramm, 2012), but this is of little value as the models themselves require heavy modification to incorporate modern or future forces, requiring their own distinct validation, which is also impossible.

The distinction between uncertainty, variability and error in a modeling and simulation study has not been consistently employed within the vocabulary of the analytical community (Oberkampf, 2002). The risk analysis literature delineates between uncertainty and error, with uncertainties being further refined into epistemic and aleatory categories. This framework is akin to the colloquial use of the terms uncertainty, variability, and error within a modeling and simulation context. Uncertainties can manifest as either simulation output variation (aleatory uncertainties) or unquantified decision risk (if epistemic uncertainties are not enumerated and relevant inputs varied within the study).

Additionally, error is defined as "a recognizable inaccuracy in any phase or activity of modeling and simulation that is not due to lack of knowledge" (Oberkampf, 2002:334). Unfortunately this has the connotation that the modeler or analyst has done something "wrong", which may not be the case. While unacknowledged errors are the term that describes inadvertent errors made by the modeler, of interest for this paper are the acknowledged errors or errors resulting from a choice in system abstraction or simulation implementation process. The impact of such errors in a simulation study results in biased simulation output or explainable deviation from "truth". In a similar fashion to both epistemic and aleatory uncertainty, certain forms of error could be systematically explored to identify sensitivities to choices made by the analyst.

Within the combat modeling community there have been several suggestions for exploring uncertainties associated with the domain (Bankes, 1993; Davis, 2000; Dewar, 1996). However, the output of these approaches does not clearly communicate the uncertainties that are buried within the inputs. In Chapter III, Evidence Theory was demonstrated as a framework for representing epistemic uncertainty in combat modeling output. The steps for aggregating multiple, conflicting sources for simulation input data were demonstrated with a simple Lanchester model incorporating six uncertain factors. The analysis found that the proposed uncertainty configuration induced a large gap between the cumulative plausibility and belief functions for blue minus red residual forces, indicating large uncertainty in combat outcomes.

This research was extended in Chapter IV by developing a sensitivity analysis method for identifying the factors which contribute to the overall uncertainty in simulation output. Prior approaches to this problem implemented a variance based metric for factor prioritization (Helton, 2006), which requires a second analysis, or is susceptible to ties when large uncertainties are present (Guo, 2007). In a resource constrained environment, sensitivity analysis methods would facilitate prioritization of resources with the goal of reducing uncertainty in system performance.

In the weather domain, there are several errors and uncertainties that contribute to the difficulties with single model analyses, some of which are (Palmer, 2004):

- The physics of weather is well understood, but the phenomena are chaotic and models are sensitive to initial conditions
- Solutions to the partial differential equations that describe weather phenomena must be reduced to analytically tractable forms, which introduces computation error in final solutions

- o There are numerous ways to implement approximations of the underlying physics and numerical approximations
- o There is no underlying framework from which a pdf of model uncertainty can be estimated.

Despite these uncertainties and errors, seasonal forecasts have been shown to have better predictive capability when several independent models are analyzed together, commonly referred to as a multi-model ensemble (Tebaldi, 2007), than when the individual predictions of those models are taken alone. Developers of multi-model ensembles must consider both the number and composition of models within the ensemble as well as the method of data aggregation and assessment of ensemble predictive skill.

Multi-model predictive superiority is not only due to error compensation, but primarily by its improved consistency and reliability across the entire predictive region (Hagedorn, 2005). A particular ensemble may not be the best forecast at each point within the predictive region, but generally outperforms any given individual model over the full range of cases. The success of multi-model methods in the weather community has led to the development of analogous approaches in prediction of disease outbreak (Morse, 2005) and rainfall runoff (Ajami, 2007).

Since there is limited data for combat model validation, the techniques for aggregation in multi-model ensembles for weather forecasting are not appropriate. There would be no way to confirm that the aggregated ensemble provides any predictive improvement over any individual simulation response. This paper develops a method to quantify the impact of error or modeling choices on simulation output uncertainty in settings where multiple models are employed. This would yield insight into the overall sensitivities of the system with respect to multiple modeling choices. The objective is not to make weakly predictive models strongly predictive models, but to improve the insight gained through a modeling and simulation activity.

This chapter is organized into five major sections: Introduction, Background, Methodology, Results and Discussion, and Conclusion. The Introduction provides a general overview of the context for the research in quantifying the impact of error on simulation output uncertainty. The Background section provides a review of literature discussing sources of uncertainties and error in the modeling and simulation process and Evidence Theory as a representation of uncertainty in combat modeling and simulation. A novel approach to quantifying the impact of error on simulation output uncertainty is presented in the Methodology section. This method is demonstrated using three distinct Lanchester models of conflict. The results of this analysis and a discussion of interesting features from this application is provided in the Results and Discussion section. In the Conclusions section, a summary of the research context, contributions, and future work are identified.

5.2 Background

5.2.1 Sources of Uncertainties and Error in the Modeling and Simulation Process

Oberkampf (2002) provides a framework for discussing variability and error in a modeling and simulation study. This framework proposes two kinds of uncertainty and the notion of error within a modeling and simulation context that are akin to the colloquial use of the terms uncertainty, variability, and error used in the operations research community. Using their definition, aleatory uncertainty "describes the inherent variation associated with a physical system or environment under consideration" (Oberkampf, 2002:334). This could be thought of as the defect rate in a manufacturing process, where the same physical processes occur repeatedly, yet each part does not come off the production line exactly to specification. Similar terms for aleatory uncertainty include variability, stochastic variability, or irreducible uncertainty. A second category of uncertainty, epistemic uncertainty was defined as “the potential deficiency in any phase or activity of the modeling process that is due to lack of knowledge” (Oberkampf, 2002:334). This category is difficult to conceptualize in most process flow modeling and simulation contexts where any potential epistemic uncertainty could be resolved by simply inspecting the process as it occurs. Epistemic uncertainty does manifest in the materials science and engineering realm, where certain model parameters would require materials testing at high temperatures and current methods preclude collecting this data.

These two types of uncertainty are related in that their impact manifests as either simulation output variation or unquantified decision risk. It’s easy to see how a simulation process that includes some representation of the aleatory uncertainties results in variation in simulation output through replication, either within run or run-to-run. Epistemic uncertainties can also induce variation in simulation output if assumptions regarding unknown aspects are enumerated and relevant inputs varied within the study. Of more concern is when epistemic uncertainties are not explicitly varied within a simulation, providing no insight into the sensitivity of simulation responses to assumptions for the analyst, resulting in unquantified decision risk. It is likely that the systematic varying of assumptions associated with epistemic uncertainties will not cover all unique possibilities, but at least relative impacts can be identified and presented to the study stakeholders to qualitatively include in their deliberations.

The popular terminology typically stops at this point with variability commonly referring specifically to aleatory uncertainty and uncertainty to epistemic uncertainty. There is a third useful distinction to make alongside the two types of uncertainty that classifies cases where choices in model abstraction and software implementation have an appreciable impact on the form of the simulation. In the manufacturing example, the modeler could choose to implement their simulation with either discrete event or agent based perspective. This choice impacts decisions that are made in the process of abstracting the physical system for simulation, changing system representation in either a satisfactory or unsatisfactory fashion. In this case, the effect on simulation output due to the selected modeling paradigm is related to neither the natural variability of the process or elements that are unknowable regarding the system under study. To account for this scenario, Oberkampf (2002) proposes the concept of simulation error. They define simulation error as “a recognizable inaccuracy in any phase or activity of modeling and simulation that is not due to lack of knowledge” (Oberkampf, 2002:334). Unfortunately this has the connotation that someone has done something “wrong”, which may not be the case. While unacknowledged errors are the term that describes errors made by the modeler, of more interest for this discussion are the acknowledged errors or errors resulting from a choice in abstraction or simulation implementation process. The impact of error in a simulation study results in biased simulation output or explainable deviation from “truth”. In a similar fashion to both epistemic and aleatory uncertainty, certain forms of error could be systematically explored to identify sensitivities to choices by the analyst.

The Society for Modeling and Simulation (1979) provides a simple framework to understand how error and uncertainty arise throughout the modeling and simulation process

(Figure 30). This framework consists of three core elements; Reality, Conceptual Model, and Computerized Model. These elements are connected via the processes of Analysis, Programming, and Computer Simulation. Model Qualification, Verification, and Validation provide feedback from “upstream” elements to ensure the simulation processes generate a suitable product based on its preceding elements.

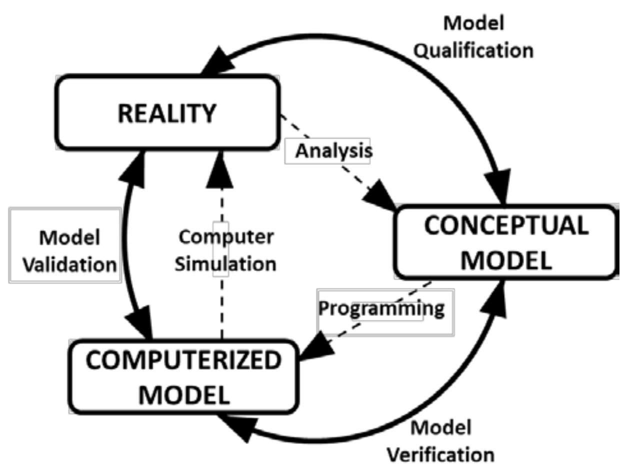

Figure 30: Society for Modeling and Simulation Framework for Simulation (SCS, 1979)

The types of uncertainty and error manifest themselves at different elements and steps throughout the modeling process. Aleatory uncertainties are manifest in reality. For all practical uses their induced stochastic behavior can be measured to suitable precision and have little impact on the remainder of the modeling process (although downstream processes may adjust how they are represented in computer code). Epistemic uncertainties manifest while developing a mental model of reality. Here the modeler aggregates what is known and what can be observed and, especially in combat modeling, makes educated assumptions when significant items are unknown or are known to exist, but are unobservable. Error manifests in both the construction of a mental model and in implementing the model in code. Classic manifestations as part of the programming process are code errors or "bugs", which are classified as unacknowledged errors. Examples of acknowledged errors would be recognizable simplifications/assumptions that are made throughout the abstraction process. The key here is that these features are recognizable when comparing the conceptual model with reality. In operations research parlance, these are referred to as uncertainties or structural uncertainties. However, these are really identifiable discrepancies between reality and the way the modeler has chosen to represent the system and thus, error. Sometimes software, processing hardware, etc. limit the implementation of reality in code. These are also errors by our definition. In this way, the modeling process can be thought of as representing the system while meeting a suitable error threshold, as measured during the model validation process.

To address the treatment of uncertainty and error, Bankes (1993) proposes two modeling paradigms: exploratory vs. consolidative modeling. Consolidative modeling is the process of building a model by consolidating known facts into a single package and then using it as a surrogate for the actual system. In contrast with consolidative modeling, the exploratory modeling approach is the use of a series of experiments to explore the implications of assumptions when unresolvable uncertainties preclude building a surrogate for the system. The power of the consolidative approach lies in the assumption that the performance of the model has been compared to reality and the accuracy of the model is known to some precision (i.e. the model can be validated (Figure 31)). In situations where this is not feasible or where important facts about the system under study are uncertain, Bankes (1993) suggests that the exploratory modeling approach is preferable, and that treating such an endeavor as if it were a consolidative modeling effort is perilous.

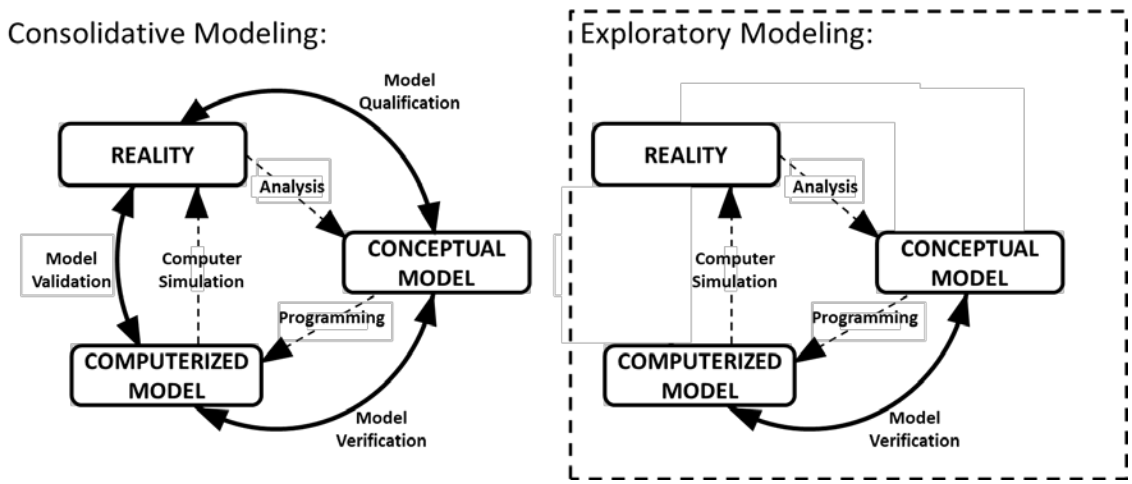

Figure 31: Consolidative vs Exploratory Modeling

Considering these issues, Dewar (1996) developed a topology of uses of distributed, real time simulations that, among other distinctions, delineated between strongly predictive and weakly predictive uses of these simulations. One factor this topology used to distinguish among possible uses of combat models was their ability to be validated. Strongly predictive models are described as having a demonstrated capacity to forecast outcomes with a high degree of accuracy (i.e. can be validated). Examples of these types of models include engineering or physics based models to predict part life, strength, fatigue characteristics, etc. Alternatively, weakly predictive models suffer from moderate to high levels of parametric, structural, or other uncertainties. Yet, the model still captures enough of the critical elements of the system under study to be useful in exploratory analysis. While these models may not be validated in the quantitative sense, they often rely on softer forms of validation (i.e. “face validation”). Millar

(2016) extended this idea, arguing the taxonomy applied to combat models in general. Due to the uncertainties inherent to combat, models of combat, to include Live-Virtual-Constructive simulations (LVCs), are generally considered weakly predictive simulations and thought to be most appropriately used for exploratory purposes.

5.2.2 Analysis of Uncertainty with Evidence Theory

There is a growing body of work in applying Evidence Theory, as a framework for systematic exploration and quantification of uncertainty in modeling and simulation. The theory was first introduced by Dempster (1967) and later codified by Shafer (1976). Applications can be found in several scientific fields, such as space launch and nuclear power plant design (Sentz, 2002). Evidence theory differs from probability theory in that likelihood is assigned to sets (i.e. a range of parameter values) instead of being assigned to a probability density function. By explicitly defining ranges of uncertain input parameters and propagating them through a model, evidence theory bounds the true cumulative density function for a response. There are three key functions in Evidence Theory; the basic probability assignment function (*BPA* or *m*), the Belief function (*Bel*) and the Plausibility function (*Pl*).

In general the basic probability assignment is not equivalent to probability as discussed in classical probability theory (although connections exist (Sentz, 2002)). Similarly to classical probability theory, the basic probability assignment is a mapping of all sets (*X*, the power set) to the interval [0, 1] and the sum of all assignments across subsets is 1. Formally, this is represented as:

$$m: P(X) \rightarrow [0,1] \tag{38}$$

$$m(\emptyset) = 0 \tag{39}$$

$$\sum_{A \in P(X)} m(A) = 1 \tag{40}$$

Using these basic probability assignments, upper and lower bounds for an interval can be calculated. The lower bound (or Belief), for a set *A* (subset of *X*), is the sum of all basic probability assignments of the proper subsets (*B*) of the set of interest (*A*).

$$Bel(A) = \sum_{B|B \text{ is a subset of } A} m(B) \tag{41}$$

The upper bound (or Plausibility) is the sum of all the basic probability assignments of the sets (*B*) that intersect the set of interest (*A*).

$$Pl(A) = \sum_{B|B \cap A \neq \emptyset} m(B) \tag{42}$$

The rules of combination in Evidence Theory allow data to be aggregated across multiple, potentially conflicting sources within a common frame of discernment. For the simulation context discussed in this research, these sources could be a set of epistemically uncertain inputs for a single simulation model or the individual combat models in a multi-model ensemble. This process assumes that the sources are independent (Shafer, 1976), however this requirement is not rigorously established in practice (Tebaldi, 2007). There are many rules for combining evidence; the key to identifying the most appropriate method is to determine how conflict between sources should be considered. A survey of relevant combination rules is provided below.

Dempster's combination rule was the original combination operator that drove the conception of D-S theory. Using Dempster's combination rule, the basic probability assignments from two (or more) sources is combined with a purely conjunctive operation. The formal definition of this operation (m_{12}) is below:

$$m_{12} = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K} \quad when\ A = \emptyset \tag{43}$$

$$m_{12}(\emptyset) = 0 \tag{44}$$

$$\text{where } K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{45}$$

It is important to note that this operation results in an aggregated mass only for intervals which overlap, giving zero mass to regions where evidence existed but did not overlap with another method. The measure of non-overlapping probability mass (or conflict) is represented by the computation of *K*.

The aggregated masses are normalized based on *K* to achieve a basic probability assignment function that resembles a probability density function from classical probability theory. Unfortunately, this choice can lead to counterintuitive results in situations involving high levels of conflict. These shortcomings are detailed in (Zadeh, 1984). Recognizing the potential pitfall, numerous other combination rules have been developed which account for level of conflict differently (see (Sentz, 2002) and (Yao, 1994) for many examples).

The Mixing Rule of combination is a popular mechanism for aggregation of disjunctive evidence (see equation (46)). This rule averages the masses (m_i) associated with a particular interval across all *i* estimates (*i* from 1 to *n*). The individual estimates can be weighted based on reliability by the multiplier, ***w***, where each w_i is the reliability associated with the *i*th source.

$$m_{1\ldots n}(A) = \frac{1}{n}\sum_{i=1}^{n} w_i m_i(A) \tag{46}$$

In contrast with Dempster's Rule of combination, evidence in conflict is preserved in the resulting BPA. Said another way, the full range of possibilities expressed in the sources are represented in the final BPA. This feature is particularly beneficial where the application of evidence theory is not to identify the most likely distribution of a particular metric, but to express the full range the distribution could be.

5.2.3 Lanchester's Model of Armed Conflict

Lanchester's Square Law (equations (47) and (48)) were developed to model combat between two homogeneous forces where both forces use aimed fire, target acquisition time does not depend on the number of targets, target acquisition time is factored into the firepower coefficients, and the firepower coefficients are constant over time (MORS, 1994).

$$\frac{dx}{dt} = -\alpha * y(t) \text{ where } x(0) = X_0 \tag{47}$$

$$\frac{dy}{dt} = -\beta * x(t) \text{ where } y(0) = Y_0 \tag{48}$$

The equations model the change of a given force level (*x(t)*) as a function of the opposing force level (*y(t)*) over time, given initial force sizes and estimates of the firepower coefficients. Using these equations, various quantities of interest can be explored, such as: who wins the conflict, residual forces, and duration of conflict (Figure 32).

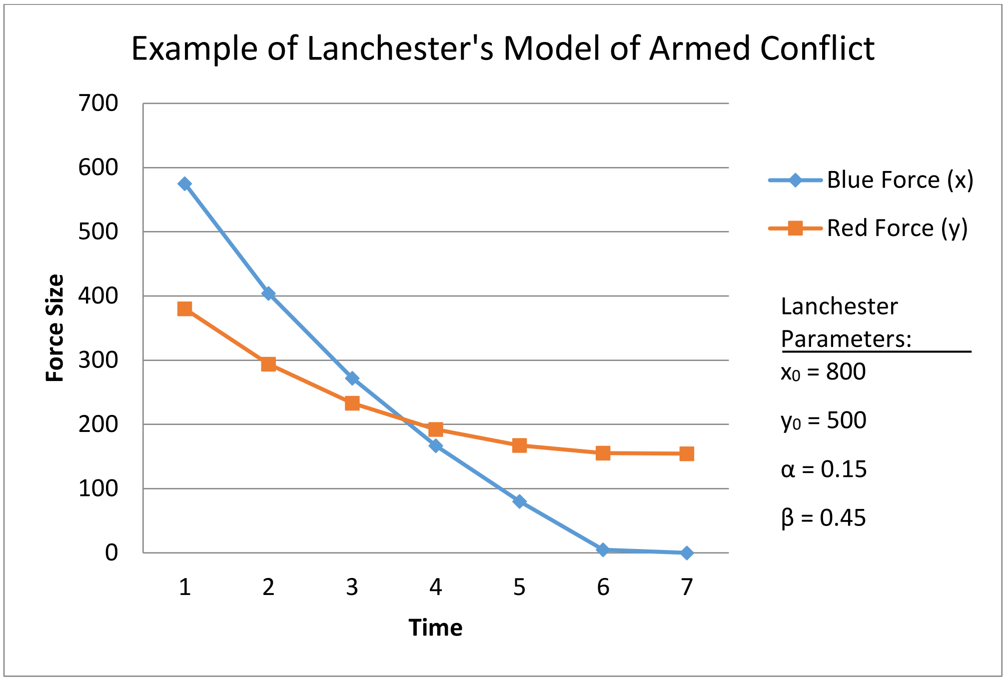

Figure 32: Example of Lanchester's Model of Armed Conflict

Lanchester's equations are appealing in part, due to their simplicity, transparency and ease of implementation. Combat analysts have employed them in assessments of several conflicts including; the Ardennes Campaign, the Battle of Kursk, and Iwo Jima (Bracken, 1995; Lucas, 2004; Schramm, 2012). However, there are numerous sources in the literature that identify deficiencies of a Lanchester model of armed conflict (Tolk, 2012). Taylor (1983) consolidates these into a single list, several of which are provided below as reference:

- No force movement
- Not verified by history

- No way to predict attrition rate coefficients
- Tactical decision processes not considered
- Battlefield intelligence not considered
- Command, control, and communications not considered
- Effects of terrain not considered
- Target priority/fire allocation not explicitly considered
- Noncombat losses are not considered

Despite these criticisms, many variants of Lanchester's equations have been developed. Extensions include incorporation of heterogeneous forces, stochastic attrition processes, reinforcements, logistics and maintenance and breakpoints (Tolk, 2012). In a more modern setting, the effects of network disruptions were represented as piecewise firepower coefficients (Schramm, 2012), enabling assessment of the impact of cyber effects on combat outcomes. Kelton (2010) describes a Lanchester model with stopping levels and stochastic reinforcements for both red and blue forces. This model was designed for implementation and analysis in the Arena discrete event simulation tool.

5.2.4 Evidence Theory Representations of Uncertainty in Combat Modeling and Simulation

In Chapter III, Evidence Theory was demonstrated as a framework for representing epistemic uncertainty in combat modeling output. There were 6 epistemically uncertain inputs; initial force levels, attrition coefficients, and stopping level for both red and blue sides. For demonstration purposes, all six uncertain factors in the Lanchester model were represented with the same scaled BPA which was constructed using the Mixing Rule of combination. The analysis found that the proposed uncertainty configuration induced a large gap between the cumulative plausibility (CPF) and belief (CBF) functions for blue minus red residual forces, indicating large uncertainty in combat outcomes.

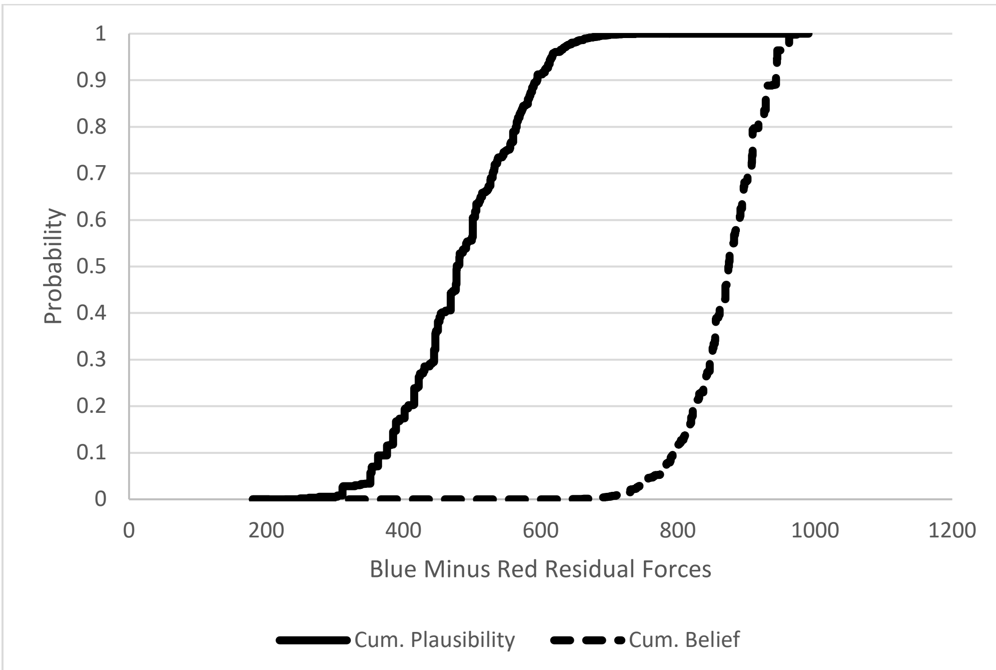

Figure 33: Evidence Theory Representation of Uncertainty

The demonstration of Evidence Theory as a framework for representing outcomes in combat modeling and simulation addresses several key gaps in the literature and common practice. First, is the propensity to treat combat simulation output as predictive when, upon examining what is known and unknown regarding the model inputs, it clearly is not. This is addressed by supplementing the single output probability density functions with cumulative belief and plausibility functions from evidence theory. These functions represent bounds on probability densities given an input uncertainty specification (or basic probability assignment).

Common summary statistics (e.g. mean, probability intervals, etc.) are in the form of ranges, which discourage the propensity to treat point estimates from a simulation as predictive.

Second, is that the employment of the Evidence Theory rules of combination eliminates the need to make choices about how to use multiple, potentially conflicting sources for modeling and simulation inputs. In the presence of multiple inputs, traditional approaches typically follow one of two lines of thought: 1) condense the sources into a single point estimate or 2) employ the Laplace principle of maximum entropy and assume a uniform distribution over the range of possible values. The validity of these approaches is heavily influenced by the process by which the sources are condensed to either a point estimate or range. In practice, these methods are unstructured and not well documented. In contrast, Evidence Theory provides a structure for aggregating multiple, conflicting sources in a repeatable manner without making assumptions regarding the distribution of the true value of the input.

In Chapter IV, the work of Chapter III was extended by developing a new method for sensitivity analysis of uncertainty in Evidence Theory. This sensitivity analysis method generates marginal CPFs and CBFs and prioritizes the contribution of each factor in reducing the Wasserstein distance (also known as the Kantorovich or Earth mover's distance) between the CBF and CPF. Using this method, a rank ordering of the model or simulation input factors was produced.

This method is notable in several respects. First, it employs the stepwise construction of CBFs and CBFs as described in Helton (2006) and a statistical measure of distance as in Guo (2007). Guo (2007) employs the K-S distance which has been demonstrated to be susceptible to ties when there are large uncertainties in outcomes with respect to the variables. The Wasserstein distance, however is not likely to cause ties in the procedure except when the variables under analysis have very little relationship to the uncertainty in model outcomes, where they exhibit marginal CPFs and CBFs similar to Figure 21.

A significant difference between the method proposed by Helton (2006) and the one presented here, is that no preliminary exploratory sensitivity analysis procedure is required. Such a procedure is good practice, but not integrally linked in the new method. A modification of this procedure could be produced where it is used (as in Helton (2006)) to incrementally construct estimates of CBFs and CPFs by adding variables until the functions stop changing enough to warrant further computation.

To demonstrate the stepwise sensitivity analysis procedure, a hypothetical military conflict modeled by Lanchester's Square Law with stopping level (a force level for both red and blue where the conflict ends) for both red and blue forces was analyzed. This overall experimental setup is exactly the same as described in Chapter III, were each E_i was estimated with an enumeration approach. As such; a more detailed explanation of the Lanchester model used can be found in sections 3.2.1 and 3.4.1 and more details regarding the construction of the BPAs and design matrix for this demonstration can be found in section 3.4.1. Once the simulation runs were complete, the marginal sensitivity of the model factors and overall contribution to the belief and plausibility functions were computed for blue minus red residual forces.

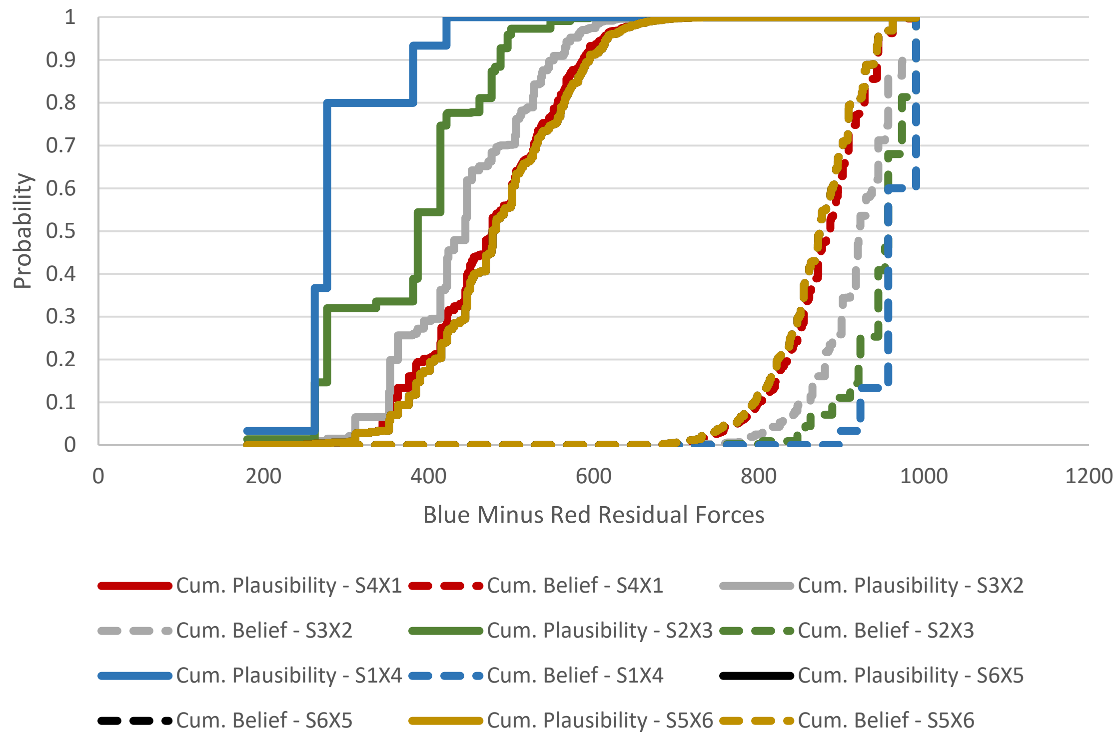

Figure 34: Sensitivity Analysis Summary of Blue minus Red Residual Forces

A relative prioritization of the factors was produced, where five of six factors had distinct contributions in explaining total uncertainty while a sixth did not. The ordered importance of the variables on total uncertainty was red attrition coefficient, blue attrition coefficient, red initial force level, blue initial force level, red stopping level and blue stopping level. While the specific results of this analysis are not extensible, they are a useful example for how to apply sensitivity analysis with Evidence Theory in modeling, simulation and analysis activities.

Implementing these methods does not come without cost. Managing multiple sources and choosing the appropriate Evidence Theory rule of combination add additional complexity to the overall analysis. The need to estimate the maximum and minimum response in each input's basic probability assignment interval intersections can drive significant increases in the number of required simulation runs. Large run matrices can be mitigated through the employment of design of experiments and meta-modeling. These costs are offset by the clarity provided to the decision maker regarding how uncertainties in modeling and simulation inputs affect the analysis outcomes.

5.3 Methodology

This research demonstrates the quantification of model error, when multiple models are employed in the analysis of a similar context. This method treats the multiple models as an epistemically uncertain quantity with an assumed BPA. Evidence Theory can then be used as in Chapter III to compute a CBF and CPF which treats the error in model form as an uncertain factor as well as the epistemically uncertain factors. The model error and uncertain factors can then be prioritized using the stepwise procedure developed in Chapter IV. The relative importance of the model error relative to the real, uncertain physical factors may drive a different interpretation of the output of the individual models. If the model error explains little of the uncertainty in the data, then the decision maker might be more qualitatively confident in the broad conclusions taken from the ensemble of models. Conversely, if model error explains much of the uncertainty in the data, then the decision maker should consider an alternate way to rationalize their decision making. The subsequent sections detail the methods for generating the CBF and CPF for an ensemble of models, a procedure for prioritization of the uncertainties and modeling error, and describe a combat scenario used to demonstrate the method.

5.3.1 *Propagating Uncertainty through Models and Computing CPFs and CBFs*

Evidence Theory provides a statistical framework for aggregating multiple, potentially conflicting estimates for a simulation factor. The uncertainty in the simulation inputs is then mapped to the simulation output by feeding the input intervals through the model. This process generally requires the formation of basic probability assignments for uncertain factors, the development of a design matrix to collect data from the simulation, execution of the simulation runs, estimation of simulation output uncertainty intervals, and, finally, the computation of cumulative plausibility and belief functions.

Formally, the evidence space for a given simulation input is (x, X, m_x). Where x is the set of all possible values for that input, X is the set of subsets of x (U_i's) that represent the interval estimates for the input, and m_x is the vector of masses associated with each element of X (U_i). The evidence space for the corresponding simulation output is (y, Y, m_y). Where y is the set of all possible values for that output, Y is the set of subsets of y (E_i's) that represent the interval estimates for the output, and m_y is the vector of masses associated with each element of Y (E_i). In the case of uncertain model inputs, all that is known is X (U_i's) and Y (E_i's) is not. The E_i's are thus properly considered estimates based on propagation of a corresponding $U_{i,}$ and identifying the minimums and maximums produced by that input interval. The mass for a given E_i (m_y) is assigned based on the corresponding U_i. If a given U_i produced an E_i, its mass (m_x) becomes the new mass (m_y) associated with the new E_i, Finally, the cumulative plausibility and belief functions were computed using equations (41) and (42).

5.3.2 *Stepwise Procedure for Prioritization of Epistemic Uncertainties and Modeling Error*

The intuitive notion behind sensitivity analysis in Evidence Theory is to identify which variables reduce the area between the CPF and CBF and by how much (Ferson, 2006). In classical probability theory, the area between two cumulative density functions is known as the Wasserstein metric (Rüschendorf, 2001),

$$W_p(\mu, v) \coloneqq \inf_{\gamma \epsilon \Gamma(\mu, v)} \left(\int_{MxM} d(x,y)^p d_\gamma(x,y) \right)^{1/P} \tag{49}$$

where *(M, d)* is a metric space and for $p \geq 1$, $P_p(M)$ is the collection of all measures μ, v on M with finite p^{th} moment. This measure is also known as the Kantorovich or Earth movers distance in the computer science community.

In our setting, we can greatly simplify this expression for the 1 dimensional case, resulting in the following expression (Rüschendorf, 2001):

$$W_1(\mu, v) = \int_{-\infty}^{\infty} \left| F_\mu(x) - G_v(x) \right| dx. \tag{50}$$

This measure satisfies the requirements for non-negativity, symmetry, definiteness, and triangle inequality, qualifying it as a true metric in the framework presented in 4.2.2. Also, assuming the body of evidence results in BPAs that meet the criteria for probability density functions, the Evidence Theory and classical probability theory representation of this metric are equivalent. F_μ and G_v represent the CBF and CPF, respectively, for a given variable.

The Wasserstein metric, in combination with a stepwise procedure to generate marginal CBF and CPFs can be used to develop a ranking of the variables impact on uncertainty. Once the procedures for analyzing uncertainties using modeling and simulation are complete (aggregating multiple inputs to form BPAs for each factor, propagating them through a model or simulation, and generating the CPF and CBF of the resulting measure of interest), the methodology is as follows:

5. Let $\Phi = \{1, \dots, n\}$ be the set of all uncertain variable indices (including model error) under consideration for this analysis (where *n* is the number of variables) and $\Omega = \{\quad\}$.

6. Iteration *k*: for each variable $x_i, i \in \Phi$;

   a. Estimate a CPF and a CBF for *y* on the basis of the evidence space obtained from the original evidence space for the x_i, the original evidence space(s) for any $x_j, \forall j \in \Omega$, and degenerate evidence spaces for all other variables (in which the sample spaces are assigned BPAs of 1).

   b. Calculate W_{1i} between marginal CPF and CBF for variable *i*.

7. Select variable $x_s, s \in \Phi$, that minimizes $W_{1\cdot}$. Remove *s* from set Φ and add to set Ω.

   Let $\hat{x}_k = x_s$, where $\hat{x}$ is and ordering of *x* and $\widehat{W}_{1k} = W_{1s}$.

8. Increment k and repeat steps (2) and (3) until $\Phi = \{\quad\}$.

At the conclusion of this procedure, $\hat{x}$ will contain the prioritized list of factors where $\hat{x}_1$ is the most important for explaining the uncertainty in simulation output and $\hat{x}_n$ is the least.

5.3.3 Error Prioritization Methodology Demonstration

To demonstrate the methodology from sections 5.3.1 and 5.3.2, a hypothetical military conflict modeled by Lanchester's Square Law with a stopping level (a force level for both red and blue where the conflict ends) and reinforcements is analyzed for three distinct models (Table 11). In Model 1, reinforcements are modeled as occurring at a constant time interval with a fixed number of reinforcements for each side. Model 2 considers stochastic reinforcements with poison inter-arrival times and exponential numbers of reinforcements. Model 3 does not implement reinforcement behavior, although it is explicitly stated as part of the problem statement. This represents the situation where an existing modeling framework does not account for all aspects of the mission under study and there are neither sufficient resources nor time to develop the appropriate capability.

Table 11: Models for Reinforcements

	Blue		Red	
	Arrivals	Number of Reinforcements	Arrivals	Number of Reinforcements
Model 1	Every 6 time units	20 units/time unit	Every 3 time units	25 units/time unit
Model 2	Poisson($\lambda = 6$ time units)	Exp($\mu = 20$ units/time unit)	Poisson($\lambda = 3$ time units)	Exp($\mu = 25$ units/time unit)
Model 3	-	-	-	-

As in Chapter III, there are 6 key uncertain inputs; initial force levels, attrition coefficients, and stopping level for both red and blue sides. These inputs are presumed to be epistemically uncertain and that the analyst has been given ranges for each input (see Table 12). In reality it is unlikely that all input quantities are epistemically uncertain, but were made so in this instance to introduce sufficient complexity to establish viability of uncertainty quantification approaches. In this situation, it is assumed that blue minus red residual forces is the primary metric of interest.

Table 12: Summary of Uncertain Inputs for Error Prioritization Demonstration

Inputs		**Uncertainty**
X1	Blue Initial Force Size (X_0)	[900, 1000]
X2	Red Initial Force Size (Y_0)	[400, 500]
X3	Blue Attrition Coefficient (α)	[0.15, 0.25]
X4	Red Attrition Coefficient (β)	[0.3, 0.45]
X5	Blue Stopping Level	[0, 50]
X6	Red Stopping Level	[0, 25]
X7	Model	[Model 1, Model 2, Model 3]

The BPA for *X1* through X6, was constructed based on estimates from three experts for the uncertain parameter values. The BPA for Model, *X7*, in this analysis is constructed as if the organization was asked: "Which method is the most appropriate for estimating blue minus red residual forces in this setting?" It is presumed that each organization (or expert) would emphatically vote that their model would be the best. Using the mixing rule of combination (with equal weights across the expert estimates), an overall BPA that spreads the mass evenly across the three models was produced. The evidence configuration for *X1* through *X6* is assumed to be the same as in Chapter III (Figure 13) with the same resulting scaled BPA (Figure 15) constructed using the mixing rule of combination.

Two possible approaches to estimating the y_i's and E_i's would be to use a meta-model or enumerating a large number of possibilities explicitly with the simulation. Neither approach guarantees that the global maximum or minimums have been found, which is why these quantities are frequently referred to as estimates for E. In either case, the analyst must choose a sampling procedure. The literature contains many examples of the sampling based approached, including random sampling, factorial experiments, and space filling designed experiments (Ankenman, 2012; Kleijnen, 2006). However, if a meta-model is to be used, care should be taken to ensure that the resulting model is statistically valid. To avoid this difficulty and to demonstrate the concept of Evidence Theory as a viable method for assessing modeling error in combat modeling, an enumeration approach was selected. The E_i's were estimated by taking the maximum and minimum from running the factorial combination of the endpoints of the U_i's resulting in 6^6*2^6 (2,985,984) runs for each simulation.

5.4 Results and Discussion

5.4.1 Uncertainty Analysis of Each Simulation Model

The resulting CBF and CPF and a prioritization of the uncertain factors was constructed for each simulation model to provide context for the multi-model analysis and resulting error prioritization (Figures 35, 36, & 37; Tables 13, 14, 15, & 16). Each simulation model produced a slightly different priority order of the uncertain factor effects on uncertainty (Table 13). Red Initial Force Size (*X2)*, Blue Firepower Coefficient (*X3*), and Red Firepower Coefficient (*X4*) were always the first three most important variables, but varied in specific order depending on the method for modeling reinforcements. Blue Firepower Coefficient (*X1*), Red Stopping Level (*X6*), and Blue Stopping Level (*X5*) were always 4th, 5th, and 6th most important respectively.

Table 13: Prioritization of Factor Effects on Uncertainty

Prioritization of Factor Effects on Uncertainty			
	Model 1	**Model 2**	**Model 3**
1	X4	X3	X4
2	X2	X2	X3
3	X3	X4	X2
4	X1	X1	X1
5	X6	X6	X6
6	X5	X5	X5

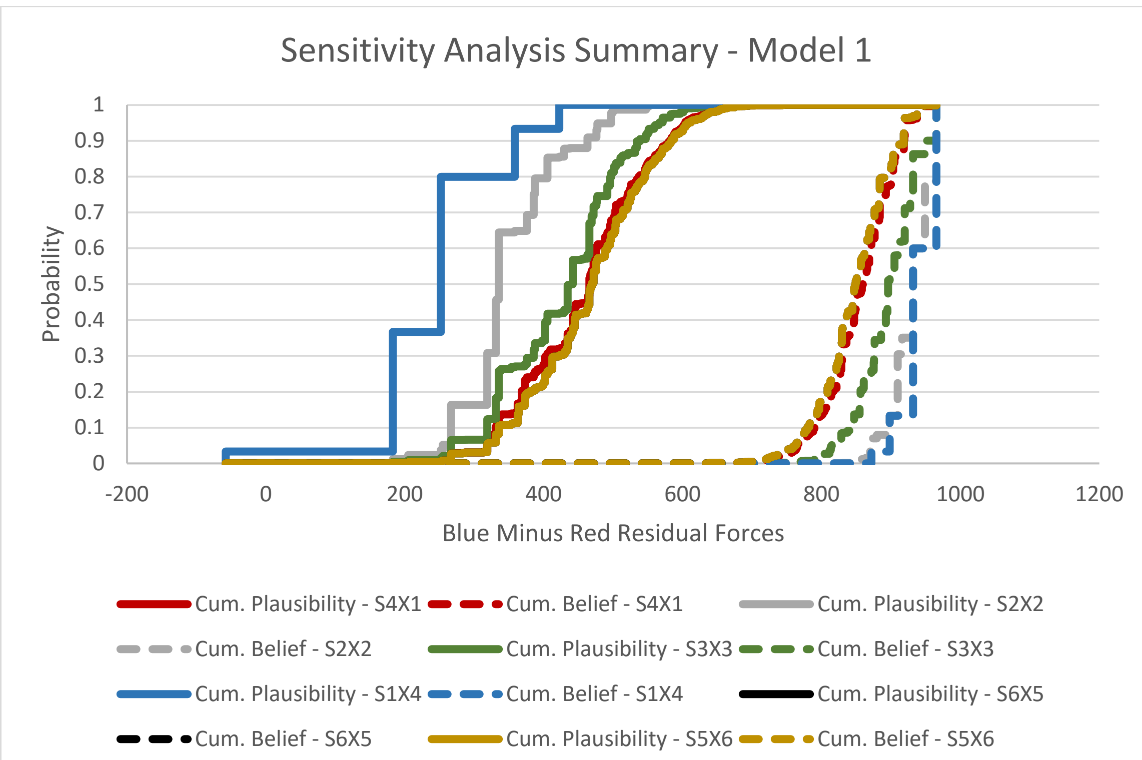

Figure 35: Summary of Uncertain Factor Prioritization for Model 1

Table 14: Summary of Wasserstein Distance for Model 1

Summary of Wasserstein Distance – Model 1						
	Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**	**Step 6**
Initial Blue Forces (X1)	812	611	708	**397**	-	-
Initial Red Forces (X2)	707	**580**	-	-	-	-
Blue Firepower Coefficient (X3)	782	586	**470**	-	-	-
Red Firepower Coefficient (X4)	**696**	-	-	-	-	-
Blue Stopping Level (X5)	1009	695	580	470	397	**381**
Red Stopping Level (X6)	1011	688	568	457	**381**	-

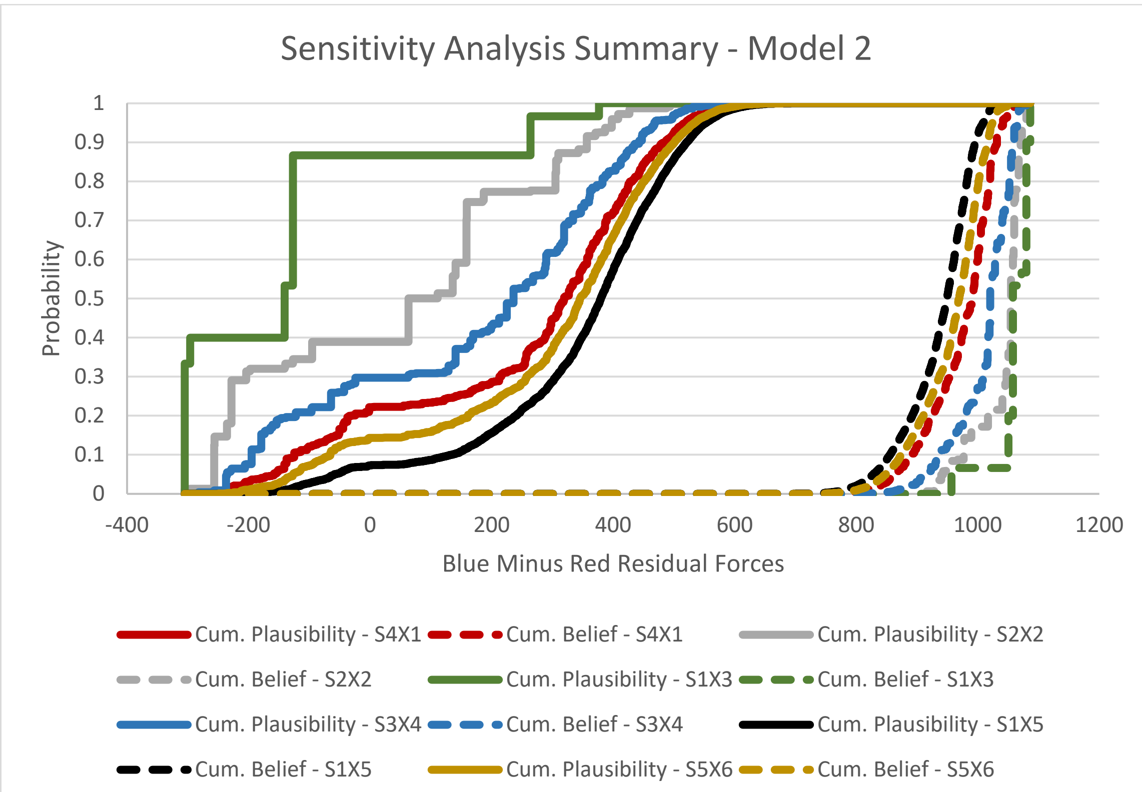

Figure 36: Summary of Uncertain Factor Prioritization for Model 2

Table 15: Summary of Wasserstein Distance for Model 2

Summary of Wasserstein Distance – Model 2						
	Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**	**Step 6**
Initial Blue Forces (X1)	1324	1076	882	**720**	-	-
Initial Red Forces (X2)	1270	**996**	-	-	-	-
Blue Firepower Coefficient (X3)	**1204**	-	-	-	-	-
Red Firepower Coefficient (X4)	1283	1004	**840**	-	-	-
Blue Stopping Level (X5)	1355	1143	938	784	663	**591**
Red Stopping Level (X6)	1357	1154	939	780	**660**	-

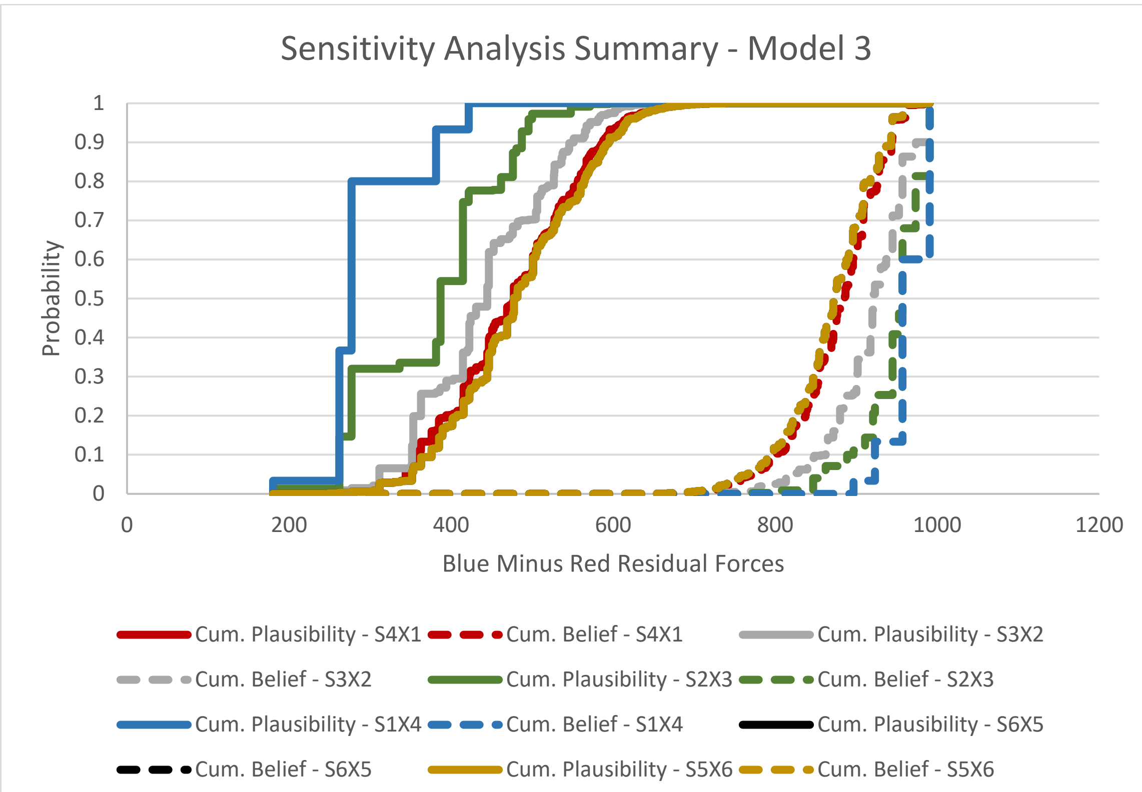

Figure 37: Summary of Uncertain Factor Prioritization for Model 3

Table 16: Summary of Wasserstein Distance for Model 3

Summary of Wasserstein Distance – Model 3						
	Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**	**Step 6**
Initial Blue Forces (X1)	714	597	499	**401**	-	-
Initial Red Forces (X2)	683	**574**	-	-	-	-
Blue Firepower Coefficient (X3)	685	570	**476**	-	-	-
Red Firepower Coefficient (X4)	**673**	-	-	-	-	-
Blue Stopping Level (X5)	810	673	570	476	401	**386**
Red Stopping Level (X6)	793	661	560	458	**386**	-

5.4.2 Error Quantification Using Evidence Theory

The multi-model analysis (or ensemble) produced a grand CPF and CBF and a prioritization of error and epistemic uncertainties for blue minus red residual forces. In this analysis the most important factor in total uncertainty was the method by which reinforcements were modeled (*X7*). This factor was followed by Red Firepower Coefficient (*X4*), Blue Firepower Coefficient (*X3*), Red Initial Force Size (*X2)*, Blue Firepower Coefficient (*X1*), Red Stopping Level (*X6*), and Blue Stopping Level (*X5*) in order from most important to least.

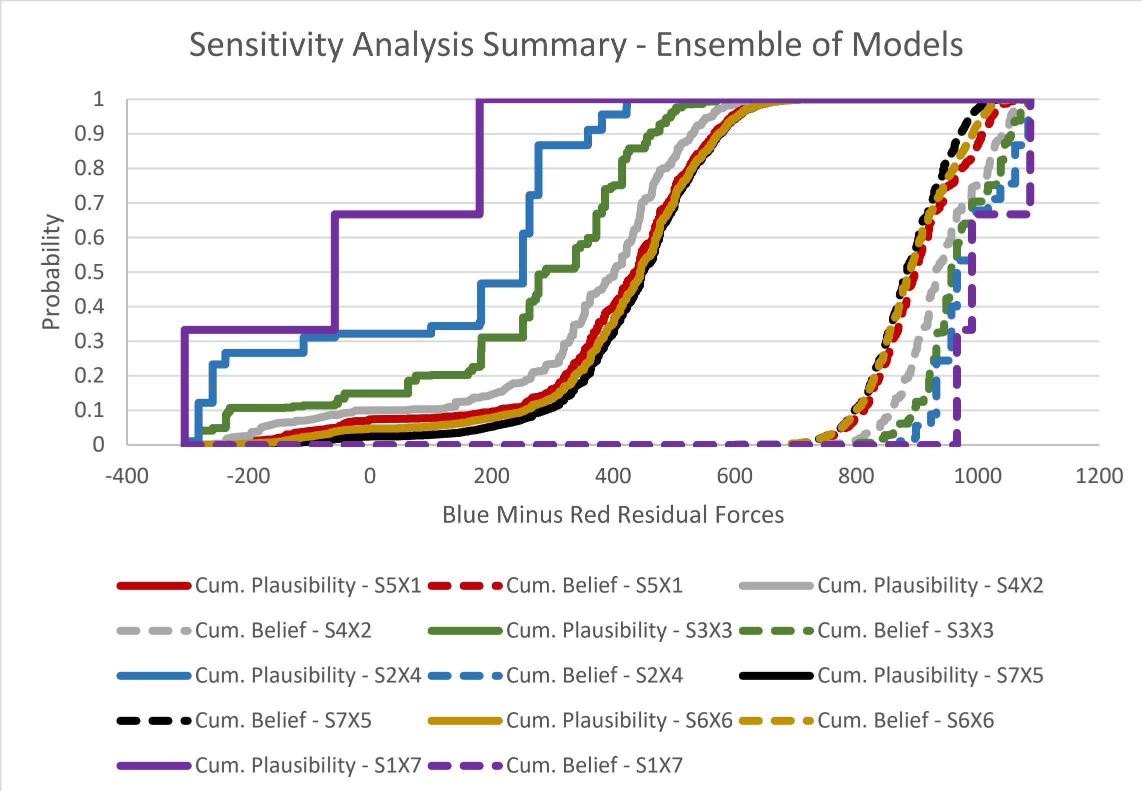

Figure 38: Summary of Uncertain Factor Prioritization for the Ensemble of Models

Table 17: Summary of Wasserstein Distance for the Ensemble of Models

Summary of Wasserstein Distance – Ensemble of Models							
	Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**	**Step 6**	**Step 7**
Initial Blue Forces (X1)	1324	950	791	628	**505**	-	-
Initial Red Forces (X2)	1269	886	758	**595**	-	-	-
Blue Firepower Coefficient (X3)	1204	890	**720**	-	-	-	-
Red Firepower Coefficient (X4)	1282	**883**	-	-	-	-	-
Blue Stopping Level (X5)	1354	1058	868	701	576	486	**452**
Red Stopping Level (X6)	1356	1053	859	693	565	**475**	-
Model (X7)	**1075**	-	-	-	-	-	-

This analysis highlights the commonly believed idea that choices in abstraction and software implementation can induce uncertainties that overwhelm natural factors of the system under study (Song, 2013). Despite this belief, systematic exploration of these choices is not revisited beyond the decision point. Using Evidence Theory and a multi-model framework for analysis of combat can highlight the impact of these choices on simulation output and begin a thoughtful dialog on how to best use simulation analysis as a decision making aid.

Seeing that the choice of how reinforcements are modeled is a significant driver of uncertainty, a commander may pause and consider how best to use these results. Further, if a commander was willing and able to specify a threshold for blue minus red for which the strategic objectives for the wider conflict are achieved, then the range of probabilities associated with that threshold could be identified. A sampling of some notional thresholds and resulting probability of successfully achieving that objective is provided below for both the baseline and Evidence Theory based analysis of uncertainty (Table 18). The individual method and multi-model Evidence Theory representations of the probability of meeting the commander's thresholds for success are different, especially for lower threshold values.

Table 18: Probability of Achieving Blue minus Red Objective with Evidence Theory

Probability of Meeting Blue Minus Red Objective				
Blue – Red Threshold	**Model 1**	**Model 2**	**Model 3**	**Ensemble of Models**
750	[0.01, 0.97]	[0.01, 0.99]	[0.01, 0.96]	[0.01, 0.98]
650	[0.02, 1.0]	[0.01, 1.0]	[0.02, 1.0]	[0.02, 1.0]
550	[0.17, 1.0]	[0.05, 1.0]	[0.25, 1.0]	[0.16, 1.0]
450	[0.58, 1.0]	[0.27, 1.0]	[0.64, 1.0]	[0.50, 1.0]

Using this framework does not address the complicating factors which prevent model validation in a combat analysis and the subsequent limits in quantified predictive capability – the ensemble does not have any more provable predictive capability than any of its constituents. Representation of errors and uncertainties inherent to combat modeling by employing multiple models and Evidence Theory improves conceptualization of uncertainty by providing ranges instead of point values for metric output. Using this framework should reduce the propensity to overlook these shortcomings when reviewing analysis output and inspire a more thoughtful dialog on the causes of the range in possible outcomes observed during the study.

5.5 Conclusion

This paper developed a method to prioritize the impact of error or modeling choices on simulation output uncertainty in settings where multiple models are employed. Prior work on representation of uncertainty in combat modeling and subsequent sensitivity analysis with respect to uncertainty with Evidence Theory was used as the basis for providing a quantitative understanding by treating model selection as an epistemically uncertain factor. This analysis provides insight into the overall sensitivities of the system with respect to multiple modeling choices. The ensemble is never the best or always the worst in terms of range in uncertainty, which is consistent with weather ensemble performance. However, the new method does not make weakly predictive models strongly predictive models, but it does ensure a plurality of perspectives are considered during a modeling and simulation activity.

This method is demonstrated using three distinct Lanchester models of conflict. Each model represented the arrival and number of reinforcements slightly differently but based on the same scenario description. Upon analysis with Evidence Theory, each of these models produced a slightly different rank ordering of the top three significant factors, while the bottom three factors were identically ranked. An analysis of the ensemble produced an aggregate ranking of the factors, which identified the method by which reinforcements were modeled as having the largest impact on uncertainty in blue minus red residual forces.

Within the combat modeling community there have been several suggestions for exploring uncertainties (Bankes, 1993; Davis, 2000; Dewar, 1996). However, the output of these approaches does not clearly communicate the uncertainties that are buried within the inputs. The methods described in this chapter explicitly generate bounds, giving the decision maker the opportunity to consider their use of the simulation output as part of their decision making process.

VI. Conclusion

This research develops a comprehensive set of techniques for the treatment of uncertainty and error in combat modeling and simulation analysis. Existing approaches within the defense community for incorporating uncertain elements into simulation studies are ad hoc with a significant number of tools that do not facilitate straight forward exploration of system uncertainty. This problem is further compounded by the fact that there are multiple overlapping simulation toolsets which, having been individually developed by domain experts (aeronautics, signatures/sensing, communications, etc.), each have slightly different representations of entities and environmental factors.

In Chapter III, Evidence Theory was demonstrated as a framework for representing epistemic uncertainty in combat modeling output. The steps for aggregating multiple, conflicting sources for simulation input data were demonstrated. A basic probability assignment was assumed for six uncertain factors; initial force size, attrition coefficient and stopping level for both blue and red forces in a Lanchester model of conflict. The analysis found that the proposed uncertainty configuration induced a large gap between the cumulative plausibility and belief functions for blue minus red residual forces, indicating large uncertainty in combat outcomes. To provide context for the Evidence Theory analysis, a traditional approach was employed in assessment of uncertainty of blue minus red residual forces in a Lanchester model of conflict. The results of both analyses were compared and contrasted.

In Chapter IV, a new method for sensitivity analysis of uncertainty in Evidence Theory was developed. This sensitivity analysis method generates marginal CPFs and CBFs and prioritizes the contribution of each factor by the Wasserstein distance (also known as the

Kantorovich or Earth Mover's distance) of the CBF and CPF. Using this method, a rank ordering of the simulation input factors can be produced. This method combines positive elements from the Evidence Theory literature; the stepwise construction of CBFs and CBFs and a statistical measure of distance, the Wasserstein Distance. Published literature employs the K-S distance, which has been demonstrated to be susceptible to ties when there are large uncertainties in outcomes with respect to the variables. The Wasserstein distance, however is not likely to cause ties in the procedure except when the variables under analysis have very little relationship to the uncertainty in model outcomes, where they exhibit a large difference between marginal CPFs and CBFs.

In Chapter V, a method to prioritize the impact of error or modeling choices on simulation output uncertainty in settings where multiple models are employed. Prior work on representation of uncertainty in combat modeling and subsequent sensitivity analysis with respect to uncertainty with Evidence Theory was used as the basis for providing a quantitative understanding by treating model selection as an epistemically uncertain factor. This analysis provides insight into the overall sensitivities of the system with respect to multiple modeling choices. The new method does not make weakly predictive models strongly predictive models, but ensures a plurality of perspectives can be reconciled during a modeling and simulation activity. This method is demonstrated using three distinct Lanchester models of conflict, each with distinct representation of the arrival and volume of reinforcements based on the same scenario description. Upon analysis, each of these models produced a slightly different rank ordering of the top three significant factors, while the bottom three factors were identically ranked. An analysis of the ensemble produced an aggregate ranking of the factors, which identified the method by which reinforcements were modeled as having the largest impact on uncertainty in blue minus red residual forces.

The demonstration of Evidence Theory as a framework for representing outcomes in combat modeling and simulation addresses several key gaps in the literature and common practice. First, is the propensity to treat combat simulation output as predictive when, upon examining what is known and unknown regarding the model inputs, it clearly is not. This is addressed by supplementing the single output probability density functions with cumulative belief and plausibility functions from evidence theory. These functions represent bounds on probability densities given an input uncertainty specification (or basic probability assignment). Common summary statistics (i.e. mean, probability intervals, etc.) are in the form of ranges, which discourage the propensity to treat point estimates from a simulation as predictive.

Second, is that the employment of the Evidence Theory rules of combination eliminates the need to make choices about how to use multiple, potentially conflicting sources for modeling and simulation analysis and their required inputs. In the presence of multiple choices, traditional approaches typically follow one of two lines of thought: 1) condense the sources into a single point estimate or 2) employ the Laplace principle of maximum entropy and assume a uniform distribution over the range of possible values. The validity of these approaches is heavily influenced by the process by which the sources are condensed to either a point estimate or range. In practice, these methods are unstructured and not well documented. In contrast, Evidence Theory provides a structure for aggregating multiple, conflicting sources in a repeatable manner without making assumptions regarding the distribution of the true value of the input. While this discussion focuses on modeling and simulation inputs, the same logic applies when choosing the appropriate modeling and simulation framework, as demonstrated in Chapter V.

Implementing these methods does not come without cost. Managing multiple sources and choosing the appropriate Evidence Theory rule of combination add additional complexity to the overall analysis. The need to estimate the maximum and minimum response in each input's basic probability assignment interval intersections can drive significant increases in the number of required simulation runs. Large run matrices can be mitigated through the employment of design of experiments and meta-modeling. These costs are offset by the clarity provided to the decision maker regarding how uncertainties in modeling and simulation inputs affect the analysis outcomes.

6.1 Future Research

There are several practical considerations for implementing these methods in a systematic way. The most significant challenge may be the generation and maintenance of the BPAs. Thinking in terms of uncertainty will be a challenge for some specialists, and it may require significant encouragement to move them from the single parameter estimate mindset to one that embraces multiple, potentially conflicting parameter estimates. A related challenge will be deciding which factors for which BPAs should be constructed and used in analysis. It will likely not be worth the effort to maintain BPAs for every factor, or even for every uncertain factor.

This research developed a procedure to prioritize the sensitivities of a simulation outcome to its input factors, assuming the BPAs were constructed from multiple independent estimators for each factor. An approach where the sensitivity of outcome uncertainty with respect to individual estimators may be desirable. Estimators could be omitted from a factor's BPAs in a structured manner, propagated through the simulation, and the effect on outcome uncertainty analyzed with a modified version of the methods from Chapter IV (Figure 39). This level of insight could provide feedback to subject matter experts regarding which estimation methods drive uncertainty and help them make decisions about how to appropriately calibrate estimate weights.

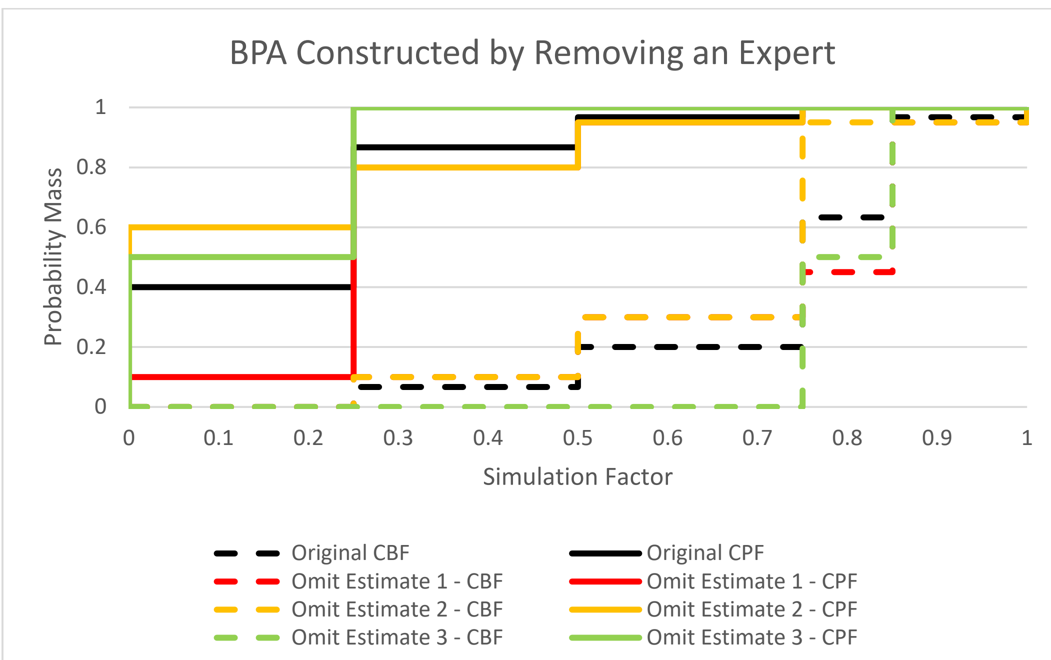

Figure 39: Sensitivity of Factor BPA to Removing an Expert Estimate

VII. References

Ajami, N. K., Duan, Q., & Sorooshian, S. (2007). An Integrated Hydrologic Bayesian Multimodel Combination Framework: Confronting Input, Parameter, and Model Structural Uncertainty in Hydrologic Prediction. *Water Resources Research*. 43(1).

Ankenman, B. E. & Nelson, B. L. (2012). A Quick Assessment of Input Uncertainty. In *Proceedings of the 2012 Winter Simulation Conference*. 1-10. IEEE.

Banks, J., Carson, J. S., Nelson, B. L., & Nicol, D. M. (2010). *Discrete-Event System Simulation* (5^{th} Edition). Upper Saddle River, NJ: Prentice Hall.

Bankes, S. (1993). Exploratory Modeling for Policy Analysis. *Operations Research*. 41(3):435-449.

Bracken, J. (1995), Lanchester Models of the Ardennes Campaign. *Naval Research Logistics*. 42:559-577.

Chick, S. E. (2006). Bayesian Ideas and Discrete Event Simulation: Why, What and How. In *Proceedings of the 2006 Winter Simulation Conference*. 96-106. IEEE.

Davis, P. K. (1995). Distributed Interactive Simulation in the Evolution of DoD Warfare Modeling and Simulation. *Proceedings of the IEEE*. 83(8):1138-1155.

Davis, P. K. (2000). Dealing with Complexity: Exploratory Analysis Enabled by Multiresolultion, Multiperspective Modeling. *In Proceedings of the 32nd Conference on Winter Simulation*. 293-302. Society for Computer Simulation International.

Davis, P. K., Bigelow, J. H., & McEver, J. (2000). Informing and Calibrating a Multiresolution Exploratory Analysis Model with High Resolution Simulation: The Interdiction Problem as a Case History. In *Proceedings of the 2000 Winter Simulation Conference*. 316-325. IEEE.

Dempster, A. P. (1967). Upper and Lower Probabilities Induced by a Multivalued Mapping. The Annals of Mathematical Statistics. 325-339.

Dewar, J. A., Bankes, S. C., Hodges, J. S., Lucas, T., & Saunders-Newton D. K. (1996). *Credible Uses of the Distributed Interactive Simulation (DIS) System*. (No. RAND/MR-607-A). Santa Monica, CA: Rand.

Doblas-Reyes, F. J., Hagedorn, R., & Palmer, T. N. (2005). The Rationale Behind the Success of Multi-Model Ensembles in Seasonal Forecasting–II. Calibration and Combination. *Tellus* A. 57(3):234-252.

Ferson, S., & Tucker, W. T. (2006). *Sensitivity in Risk Analyses with Uncertain Numbers*. Albuquerque, New Mexico: Sandia National Laboratories.

Gray, F. (2007). Experimental Designs for Development Tests in Distributed Joint Environments. *Technical Report*, Destin, Florida.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis* (3rd Edition). Boca Raton, FL: CRC press.

Goldstein, M., & Rougier, J. (2004). Probabilistic Formulations for Transferring Inferences from Mathematical Models to Physical Systems. *SIAM Journal on Scientific Computing.* 26(2):467-487.

Guo, J. & Du, X. (2007) Sensitivity Analysis with Mixture of Epistemic and Aleatory Uncertainties. AIAA Journal. 45(9):2337-2349.

Haase, C. L., Hill, R. R., & Hodson, D. D. (2014). Planning for LVC Simulation Experiments. *Applied Mathematics*. 5(14):2153-2167.

Hagedorn, R., Doblas-Reyes, F. J., & Palmer, T. N. (2005). The Rationale Behind the Success of Multi-Model Ensembles in Seasonal Forecasting –I. Basic Concept. *Tellus* A. 57(3):219-233.

Han, D., Dezert, J., & Yang, Y. (2018). Belief Interval-based Distance Measures in the Theory of Belief Functions. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 48(6):833-850.

Helton, J. C., Johnson, J. D., Oberkampf, W. L., & Sallaberry, C. J. (2006). Sensitivity Analysis in Conjunction with Evidence Theory Representations of Epistemic Uncertainty. Reliability Engineering and System Safety. 91:1414-1434.

Hodson, D., Gehl, D., & Baldwin, R. (2006). Building Distributed Simulations Utilizing the EAAGLES Framework. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).* 5(2).

Hodson, D. D., & Baldwin, R. O. (2009). Characterizing, Measuring, and Validating the Temporal Consistency of Live-Virtual-Constructive Environments. *Simulation*. 85(10):671-682.

Hodson, D. D., & Hill, R. R. (2014). The Art and Science of Live, Virtual, and Constructive Simulation for Test and Analysis. *The Journal of Defense Modeling and Simulation*. 11(2):77-89.

IEEE. Standard for Distributed Interactive Simulation, Standard 1278, 2012.

IEEE. Standard for Modeling and Simulation High Level Architecture, Standard 1516, 2000.

Jousselme, A.-L. & Maupin, P. (2012). Distances in Evidence Theory: Comprehensive Survey and Generalizations. *International Journal of Approximate Reasoning*. 53:118-145.

Kelleher, C. T. (2014). *Dynamic Bayesian Networks as a Probabilistic Metamodel for Combat Simulation* (Doctoral Dissertation, Air Force Institute of Technology).

Kelton, W.D., Sadowski, R.P. and Swets, N.B. (2010) *Simulation with Arena* (5th Edition). Boston, MA: McGraw-Hill.

Kirtman, B. P., Min, D., Infanti, J. M., Kinter III, J. L., Paolino, D. A., Zhang, Q., ... & Peng, P. (2014). The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society*, 95(4):585-601.

Lanchester, F.W. 1914. Aircraft in Warfare: The Dawn of the Fourth Arm, Parts V and VI—The Principle of Concentration. *Engineering.* 98:422-423, 452-454.

Law, A. M. & Kelton, D. M. (1999). *Simulation Modeling and Analysis* (3rd Edition). New York, NY: McGraw-Hill Higher Education.

Lee, J., Lim, Y., Chi, S. (2004). Hierarchical Modeling and Simulation Environment for Intelligent Transportation Systems. *Simulation*. Transactions of Society for Modeling and Simulation International.

Lucas, T. W. and Turkes, T. (2004), Fitting Lanchester equations to the battles of Kursk and Ardennes. *Naval Research Logistics*. 51:95-116.

Military Operations Research Society., Olson, W. K., Parry, S. H., & Youngren, M. A. (1994). Military Operations Research Analyst's Handbook. Alexandria, VA: Military Operations Research Society.

Millar, J. R., Blake, J. A., Hodson, D. D., Hill, R. R., & Miller, J. O. (2016). Sources of Unresolvable Uncertainties in Weakly Predictive Distributed Virtual Environments. In *Proceedings of the 2016 Winter Simulation Conference*. 3134-3142. IEEE.

Millar, J. R., Hodson, D. D., Peterson, G. L., & Ahner, D. K. (2016). Data Quality Challenges in Distributed Live-Virtual-Constructive Test Environments. *Journal of Data and Information Quality*. 7(1–2):2.

Miller, J O. “OPER 671 - Joint Combat Modeling I". Course Notes, 2016.

Morse, A. P., Doblas-Reyes, F. J., Hoshen, M. B., Hagedorn, R., & Palmer, T. N. (2005). A Forecast Quality Assessment of an End-to-End Probabilistic Multi-Model Seasonal Forecast System Using a Malaria Model. *Tellus* A. 57(3):464-475.

National Academy of Sciences. (2012). Capability Planning and Analysis to Optimize Air Force Intelligence, Surveillance and Reconnaissance Investments. The National Academies Press. Washington DC.

National Research Council. 2008. Integrated Computational Materials Engineering: A Transformational Discipline for Improved Competitiveness and National Security. Washington, DC: The National Academies Press. https://doi.org/10.17226/12199.

Oberkampf, W. L, DeLand, S. M., Rutherford, B. M., Diegert, K. V., & Alvin, K. F. (2002). Error and Uncertainty in Modeling and Simulation. *Reliability Engineering and System Safety*. 75(3):333-357.

Palmer, T. N., Doblas-Reyes, F. J., Hagedorn, R., Alessandri, A., Gualdi, S., Andersen, U., & Graham, R. (2004). Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER). *Bulletin of the American Meteorological Society*. 85(6):853-872.

Poropudas, J., & Virtanen, K. (2007). Analyzing air combat simulation results with dynamic Bayesian networks. In *Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come*. 1370-1377. IEEE Press.

Rodriguez, J. D., Miller, J. O., Bauer, K. W., & Neher, R. E. (2008). Methodologies for aggregating large hierarchical simulation models. In *Modeling and Simulation for Military Operations III.* 6965(69650C). International Society for Optics and Photonics.

Rougier, J. C (2007): Probabilistic Inference for Future Climate Using an Ensemble of Simulator Evaluations, *Climatic Change*. 81(3):247-264.

Roy, C. J., & Oberkampf, W. L. (2011). A Comprehensive Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing. *Computer Methods in Applied Mechanics and Engineering*. 200(25):2131-2144.

Rüschendorf, L. (2001). Wasserstein Metric. In Michiel Hazewinkel (ed.), *Encyclopedia of Mathematics*. Springer Science+Business Media B.V. / Kluwer Academic Publishers, ISBN 978-1-55608-010-4

Sandeep, S., & Zyda, M. (1999). *Networked Virtual Environments – Design and Implementation*. New York, NY: Addison Wesley.

Sentz, K., & Ferson, S. (2002). *Combination of Evidence in Dempster-Shafer Theory*. 4015. Albuquerque, NM: Sandia National Laboratories.

Shafer, G. (1976). A Mathematical Theory of Evidence. Princeton, NJ: Princeton University Press.

Smith, R. L., Tebaldi, C., Nychka, D. & Mearns, L. O. (2009). Bayesian Modeling of Uncertainty in Ensembles of Climate Models. *Journal of the American Statistical Association*. 104(485):97-116.

Schramm, H. C. (2012). Lanchester models with discontinuities: An application to networked forces. *Military Operations Research.* 17(10).

Society for Modeling and Simulation International. (1979). Terminology for model credibility. *SIMULATION.* 32(3):103–104. https://doi.org/10.1177/003754977903200304

Song, E., & Nelson, B. L. (2013). A Quicker Assessment of Input Uncertainty. In *Proceedings of the 2013 Winter Simulation Conference: Making Decisions in a Complex World*. 474-485. IEEE.

Tebaldi, C., & Knutti, R. (2007). The Use of the Multi-Model Ensemble in Probabilistic Climate Projections. *Philosophical Transactions of the Royal Society*. 365(1857).

Tebaldi, C., Smith, R. L., Nychka, D., & Mearns, L. O. (2005). Quantifying Uncertainty in Projections of Regional Climate Change: A Bayesian Approach to the Analysis of Multimodel Ensembles. *Journal of Climate*. 18(10):1524-1540.

TENA. Test and Training Enabling Architecture, October 2017. https://www.tena-sda.org/display/TENAintro/Home

Tolk, A. (2012). *Engineering Principles of Combat Modeling and Distributed Simulation*. Hoboken, NY: Wiley.

Wright, S. A., K. W. Bauer, Jr., and M. E. Oxley. 2004. The Covalidation of Dissimilarly Structured Strategic Airlift Models. *Mathematical and Computer Modelling*. 39(6-8):677-696.

Wu, J., & David, J. L. (2002). A Spatially Explicit Hierarchical Approach to Modeling Complex Ecological Systems: Theory and Applications. *Ecological Modelling*. 153(1):7-26.

Yager, R. R. (1987). On the Dempster-Shafer Framework and New Combination Rules. *Information Sciences*. 41(2):93-137.

Yao, Y. Y., & Wong, S. K. M. (1994). Representation, Propagation and Combination of Uncertain Information. *International Journal of General Systems*. 23(1):59-83.

Zadeh, L. A. (1984). Review of a Mathematical Theory of Evidence. *AI magazine*. 5(3):81.

Zhao, M., Panda, R. V., Sapatnekar, S. S., & Blaauw, D. (2002). Hierarchical Analysis of Power Distribution Networks. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. 21(2):159-168.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
19-12-2019

2. REPORT TYPE
Doctoral Dissertation

3. DATES COVERED *(From - To)*
Apr 2015 – Dec 2019

4. TITLE AND SUBTITLE
Uncertainty and Error in Combat Modeling,
Simulation, and Analysis

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Blake, Jason A.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management
(AFIT/EN)
2950 Hobson Way
WPAFB, OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENS-DS-19-D-019

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Ms. Olivia Oxford
NASIC\ACNQ
4180 Watson Way
WPAFB, Ohio 45433
olivia.oxford@us.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
NASIC\ACNQ

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution Statement A:
Approved for Public Release; Distribution Unlimited.

13. SUPPLEMENTARY NOTES
This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT
Due to the infrequent and competitive nature of combat, several challenges present themselves when developing a predictive simulation. First, there is limited data with which to validate such analysis tools. Secondly, there are many aspects of combat modeling that are highly uncertain and not knowable. This research develops a comprehensive set of techniques for the treatment of uncertainty and error in combat modeling and simulation analysis.

First, Evidence Theory is demonstrated as a framework for representing epistemic uncertainty in combat modeling output. Next, a novel method for sensitivity analysis of uncertainty in Evidence Theory is developed. This sensitivity analysis method generates marginal cumulative plausibility functions (CPFs) and cumulative belief functions (CBFs) and prioritizes the contribution of each factor by the Wasserstein distance (also known as the Kantorovich or Earth Mover's distance) between the CBF and CPF. Using this method, a rank ordering of the simulation input factors can be produced with respect to uncertainty. Lastly, a procedure for prioritizing the impact of modeling choices on simulation output uncertainty in settings where multiple models are employed is developed. This analysis provides insight into the overall sensitivities of the system with respect to multiple modeling choices.

15. SUBJECT TERMS
Combat Modeling, Uncertainty, Epistemic Uncertainty, Error, Evidence Theory

16. SECURITY CLASSIFICATION OF:
UNCLASSIFIED

a. REPORT	b. ABSTRACT	c. THIS PAGE
U	U	U

17. LIMITATION OF ABSTRACT
U

18. NUMBER OF PAGES
152

19a. NAME OF RESPONSIBLE PERSON
Dr. J. O. Miller, AFIT/ENS

19b. TELEPHONE NUMBER *(include area code)*
(937) 255-3636, x4326;
john.miller@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18