

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Rigorous, Defensible, Efficient Testing* (Keynote)

Catherine Warner  
Institute for Defense Analyses

DOT&E employs rigorous statistical methodologies in its assessments of test adequacy and in its operational evaluations to provide sound analysis of test results to acquisition decision makers. DOT&E has taken the initiative to promote and require these scientific best practices in operational testing and in the analysis of the data gathered from these well-designed tests. The methodologies have not only provide a rigorous and defensible coverage of the operational space; they also allow us to quantify the trade-space between testing more and answering complex questions about system performance. The results have been better test plans, more confidence in the information provided to acquisition and Service officials, and improved test efficiency. These methods encapsulate the need to "do more without more," especially in light of a highly constrained fiscal environment. Indeed, the final result of this policy is to ensure that we optimize scarce resources by neither over-testing nor under-testing weapon systems; provide sound rationale for the level of testing prescribed; and gather the data needed to provide warfighters with a basis for trust in evaluations of the performance of those weapon systems.

While there has been a lot of progress, much work remains. The implementation of these techniques is still far from widespread across all DoD T&E communities. Furthermore, we are currently not leveraging these methods in a sequential fashion to improve knowledge as we move from developmental testing to operational testing. In this talk I will discuss DOT&E's efforts to institutionalize the use of statistical methods in DoD T&E. I will share case studies highlighting successful examples and areas for improvement.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Data, Models, and Analytics*

Sallie Keller

Virginia Bioinformatics Institute at Virginia Tech

Data, models, and analytics have become ubiquitous across all areas of science and engineering. They are at the heart of scientific discovery and responsible policy and strategy development. The language used to describe data, models, and analytics is also pervasive opening opportunities for miscommunication of major results and lack of rigor in drawing conclusions. Today's focus on "big data" and "data science" is exacerbating this situation. This talk will discuss data, models, and analytics with a focus on placing these concepts in a crisp problem-solving framework. Special emphasis will be given to the collection of skills needed for the analysis of complex scenarios faced by defense and security analysts.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Malware Static Analysis Techniques Using a Multidisciplinary Approach*

Edward Wegman  
George Mason University

Most research work discussing malware detection completely dismiss signatures as being a thing of the past accusing it of suffering from a weak ability to detect new zero-day malware. This indeed could be the case if we are still referring to the classic definition of signatures that renders them specific to only a single malicious executable binary. But what if these signatures grouped more malicious executables under a single signature? It would make a valuable defense towards the fight against malware. To create such signatures we need to develop new methods and techniques to constantly advance the state of the art as malware get more and more elusive under old methods and approaches. The methods I will discuss not only give a good chance of creating effective signatures for malware, but also provide something just as important and that is giving the malware analyst an automated approach to understanding key characteristics of the analyzed malware. This work has resulted in Patent US 9,092,229.

This is joint work with Muhammed Aljammaz.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Design Oriented Modeling: A Powerful Approach for Model Selection for Definitive Screening Designs*

Bradley Jones  
SAS/JMP

Optimal designs implement the idea of model oriented design. That is, an optimal design maximizes the information about a specified model. Designed experiments often have strong symmetry (such as orthogonal columns). This suggests that analytical methods for designed experiments could profitably take advantage of what is already known about their structure. I call this idea design oriented modeling.

Definitive Screening Designs (DSDs) have a special structure with many desirable properties. They have orthogonal main effects, and main effects are also orthogonal to all second order effects. DSDs with more than five factors project onto any three factors to enable efficient fitting of a full quadratic model. However, analytical methods for DSDs employ generic tools invented for regression analysis of observational data. These approaches do not take advantage of all the useful structure that DSDs provide.

This talk introduces an analytical approach for DSDs that does take explicit advantage of the special structure of DSDs. To make the methodology clear, a step-by-step procedure for analysis will be presented using specific examples.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by  
Interface Foundation of North America

*Mining Text Network*  
(Special Session on Text Mining/Analysis)

David Banks  
Duke University

The last decade has seen substantial progress in topic modeling, and considerable progress in the study of dynamic networks. This research combines these threads, so that the network structure informs topic discovery and the identified topics predict network behavior. The data consist of text and links from all US political blogs curated by Technorati during the calendar year 2012. We find cascading patterns of discussion driven by events, and interpretable topic distributions. A particular advantage of the model used in this research is that it naturally enforces cluster structure in the topics, through a block model for the bloggers.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Modeling the Evolution and Influence of Blogger's Topic Sentiment and Opinion Profile*  
(Special Session on Text Mining/Analysis)

Michael James Garrity

Online blogging has become a popular way for people to share their opinions to a general audience. It also provides a forum in which others can leave feedback and recommend blogs to other users. Over time, certain users become popular with high numbers of recommendations and comments, and this may lead to increased influence on other bloggers' sentiment within the blogging community. Given the large number of users and topics being discussed in any blogosphere, it can be difficult to easily gauge patterns and trends in public sentiment, as well as any level of influence certain bloggers may have. In the following work, we introduce new techniques to model the evolving sentiment and blogger influence within a blogosphere.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Risk Preserving Semantic Content in Text Mining Using Multigrams*  
(Special Session on Text Mining/Analysis)

Yasmin H. Said

Text mining can be thought of as a synthesis of information retrieval, natural language processing, and statistical data mining. The set of documents being considered can scale to hundreds of thousands and the associated lexicon can be a million or more words. Information retrieval often focuses on the so-called vector space model. Clearly, the vector space model can involve very high-dimensional vector spaces. Analysis using the vector space model is done by consideration of a term-document matrix. However, the term-document matrix basically is a bag-of-words approach capturing little semantic content. We have been exploring bigrams and trigrams as a method for capturing some semantic content and generalizing the term-document matrix to a bigram-document matrix. The cardinality of the set of bigrams is in general not as big as  $n^2$ ; it is nonetheless usually considerably larger than  $n$ , where  $n$  is the number of words in the lexicon. show the results over time. These maps can be displayed on such devices as iPads or smartphones.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Design and Analysis for Constraining Dynamic Network Simulation Models  
with Observational Data*

(Special Session on Design of Experiments)

David Higden and Jiangzhuo Chen  
Virginia Bioinformatics Institute, Virginia Tech

Dynamic network simulation models can evolve a very large system of agents over a network that evolves with time. Such models are often used simulate epidemics or transportation, typically producing random trajectories, even when model parameters and initial conditions are identical. This introduces a number of challenges in designing ensembles of model runs for sensitivity analysis and computer model calibration. This talk will go through a case study of a recent epidemic, seeking to forecast the epidemic's behavior given initial administrative information. We make use of space-filling latin hypercube designs, with additional replication to deal with the random simulator, and to facilitate the analysis of this rather large amount of output produced by these model runs.



2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by  
Interface Foundation of North America

*Incorporating Confidence Intervals on the Decision Threshold in Logistic Regression*  
(Special Session on Design of Experiments)

Michael J. Kist and Rachel T. Silvestrini  
Rochester Institute of Technology

This work discusses an application of confidence intervals to the threshold decision value used in logistic regression and discusses its effect on changing the quantification of false positive and false negative errors. In doing this a grey area, in which observations are not classified as success (1) or failure (0), but rather 'uncertain' is developed. The size of this grey area is related to the level of confidence chosen to create the interval around the threshold as well as the quality of logistic regression model fit. This method shows that potential errors may be mitigated. Monte Carlo simulation and an experimental design approach is used to study the relationship between a number of responses relating to classification of observations and the following factors: threshold level, confidence level, noise in the data, and number of observations collected. An application of body armor penetration tests is used to show how the methodology would work in practice.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by  
Interface Foundation of North America

*Covering Arrays: A Tool to Construct Efficient Test Suites for Validating Software Systems*  
(Special Session on Design of Experiments)

Joseph Morgan and Ryan Lekivetz  
SAS/JMP

A homogenous covering array  $CA(N; t, k, v)$ , is an  $N \times k$  array on  $v$  symbols such that any  $t$  column projection contains all  $v^t$  level combinations at least once. In this talk we will describe key generalizations to this basic homogenous covering array model and explain why these constructs are viewed by the software engineering community as an important tool for validating software systems. Two case studies will be presented to illustrate how these constructs may be used in a real world setting to construct efficient test suites.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*CSP #579: The Health of Vietnam Era Women's Study (HealthViEWS)*  
(Special Session on Cooperative Studies at the VA)

Tracey Serpi  
Perry Point Cooperative Studies Program Coordinating Center, VA

Little is known about the long-term health and mental health status of women Vietnam veterans. For many of these women, the effects of this war are still present in their daily lives. As these women approach their mid-sixties, it is important to understand the impact of wartime deployment on health and mental health outcomes nearly 40 years later. HealthViEWS was designed to assess the prevalence of PTSD and other mental and physical health conditions for women Vietnam veterans, and to explore the relationship between posttraumatic stress disorder (PTSD) and other conditions and the Vietnam deployment experience.

This study attempted to contact approximately 8,000 women for participation in a mailed survey, telephone interview and a review of their medical records. The three primary study aims were: 1) to determine the prevalence of lifetime and current psychiatric conditions including PTSD among women who served during the Vietnam Era, 2) to characterize the physical health of women who served during the Vietnam Era, and 3) to characterize the level of current disability in women who served during the Vietnam Era. The four primary outcomes of this study were: (1) lifetime and current PTSD as measured by the Composite International Diagnostic Interview (CIDI), (2) lifetime and current Major Depressive Disorder as measured by the CIDI, and (3) cardiovascular disease and diabetes as measured by self-report and (4) the current levels of disability as measured by self-report. Women identified as serving in Vietnam, near Vietnam (in Asia during the Vietnam War) and in the US during the Vietnam War were identified from an established cohort and supplemented with self-registration and the DMDC Vietnam Roster. Military service history and locations of service were verified through military personnel records. Qualifying Veterans were sent a survey on demographics, behaviors, disability, health-related quality of life, and medical conditions. Women agreeing to be contacted were also interviewed using the modified CIDI to ascertain current and lifetime mental health conditions and exposure to traumatic events. A more extensive chart review was conducted by a clinician to validate self-report of key medical conditions.

This study represents to date the most comprehensive cohort to be developed on a group of women Vietnam veterans, and will inform future research on women veterans in future war cohorts. Such an understanding will lay the groundwork for planning and providing appropriate services for these often overlooked women veterans, as well as for the aging veteran population today.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*CSP #504: Risperidone Treatment for Military Service Related  
Post-Traumatic Stress Disorder (PTSD)*  
(Special Session on Cooperative Studies at the VA)

Karen Jones

Perry Point Cooperative Studies Program Coordinating Center, VA

CSP #504, Risperidone Treatment for Military Service Related Post-Traumatic Stress Disorder (PTSD), was a study to assess the efficacy and safety of the second-generation antipsychotic medication, risperidone, in treating Veterans with military service related chronic PTSD who had been partial- or non-responders to antidepressant medications. Secondary objectives included assessment of the effect of risperidone treatment on depression, anxiety and quality of life.

The study was a 6-month, randomized, double-blind, placebo-controlled multicenter clinical trial conducted between February, 2007, and February, 2010, at 23 Department of Veterans Affairs medical centers across the country. Of the 367 Veterans assessed for eligibility, 296 were randomized and 247 were included in the primary outcome analysis.

The primary outcome measure was the Clinician-Administered PTSD Scale (CAPS), a semi-structured interview to assess PTSD symptoms and often considered the gold standard in PTSD assessment. Other outcome measures included the Montgomery-Asberg Depression Rating Scale (MADRS), the Hamilton Anxiety Scale (HAM-A), the Clinical Global Impression scale (CGI) and the Veterans RAND 36-item Health Survey (SF-36V).

Analysis of the change in CAPS total score (change score) from baseline to 24 weeks (primary outcome analysis) revealed no significant difference between the placebo and risperidone groups. A secondary mixed models analysis of CAPS total scores at all data collection points also did not show a significant difference between the groups. Risperidone did not reduce symptoms of depression or anxiety nor did it increase quality of life. Post-hoc mixed models analyses of the CAPS subscales showed a statistically significant effect of risperidone on reducing symptoms in the re-experiencing and hyperarousal domains; however, the findings did not meet the threshold for a minimal clinically important difference, suggesting the changes on these subscales would not be considered meaningful by clinicians.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by  
Interface Foundation of North America

*Tractable Analysis of Experiments on High-dimensional Simulation Codes*

Matthew Plumlee  
University of Michigan

Simulation models have become a popular method to investigate complex systems. Analysis of large scale experiments on these models has remained a difficult problem. This paper suggests an approach for building predictors from experiments using specialized grid-based layouts. The computational savings for analysis of high-dimensional experiments can be several orders of magnitude. It is also shown that the prediction accuracy is comparable with the common space-filling designs.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Computational Geometry for Multivariate Statistics*

John Nolan  
American University

A range of multivariate statistical distributions require the ability to work with shapes and regions. We describe how some basic computational geometry allow us to go beyond 2 dimensions and explore the open source R package `mvmesh` that produces and plots multivariate shapes and regions. One application is multivariate histograms, where the bins are multivariate and possibly non-rectangular. Another example is multivariate extreme value distributions (MVEVDs), which are characterized by an angular measure on the unit simplex. A rich class of models based on partitioning the simplex that is tractable and dense.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Statistical Design and Validation of Modeling and Simulation*

Kelly McGinnity and Laura Freeman  
Institute for Defense Analyses

In many situations, collecting sufficient data to evaluate system performance against operationally realistic threats is not possible due to cost and resource restrictions, safety concerns, or lack of adequate or representative threats. Modeling and simulation tools that have been verified, validated, and accredited can be used to supplement live testing in order to facilitate a more complete evaluation of performance. Two key questions that have not been fully answered in the OT context are 1) what is the best design strategy for the simulation experiment and 2) what is the best analysis method for comparing live trials to the simulated trials for the purpose of validating the M&S? This talk will discuss various strategies for addressing these two questions. The best methodologies for designing and analyzing may vary depending on the nature of the model and how much live data is available.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Learning Bayesian Network Structure using Data from Designed Simulation Experiments*

Jarom Ballantyne, Terril Hurst, and Allan Mense  
Raytheon Missile Systems

Bayesian networks have proven to be a useful tool for evaluating the results from designed simulation experiments. A Bayesian Network model consists of (1) a set of clearly defined network variables and their values, (2) a network structure connecting variables (edges) and (3) a conditional probability table (CPT) for each of the network variables. The network structure is constructed using either information from domain experts or from experimental data. In complex simulations, a structure based on experts’ causal perceptions of the network variables can be inaccurate, and essential variable interactions are often overlooked. This leads to sub-optimal model predictive accuracy.

Learning the network structure from a given data set can be accomplished with a scoring function. Learning the network structure from simulation data has the advantage of including variable interactions as estimated by the data. However, large simulation data sets used in the network learning can result in overly complex network structures.

This paper describes a method for using the scoring function to identify dominant relationships between variables. The method allows the pruning of edges between variables that have a low impact on the overall score. Examples are given to illustrate the benefits of this learning approach. This method, along with prior knowledge and design information, is essential in developing accurate Bayesian network models in simulation environments.



2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Modeling Adversarial Sensor Networks Using Generalized Stochastic Petri Nets*

Terril Hurst, Kris Kukla, Daniel Rosser, and James Kimmet  
Raytheon Missile Systems

Bayesian networks have proven useful for designing simulation experiments to allocate and verify compliance with multi-level, probabilistic performance requirements. However, modeling a stochastic, distributed system sometimes requires a method that can more explicitly represent time-dependent attributes such as concurrency, conflict, and deadlock. Generalized stochastic Petri nets (GSPNs) offer a way to represent these attributes and to complement the analysis rigor of Bayesian networks.

This paper reports lessons learned in constructing and analyzing a GSPN for a specific type of distributed system. An adversarial sensor network (ASN) consists of “Red” and “Blue” agents, each employing multiple sensors and/or sensor modes in an engagement. As the number of sensors and/or modes increases, the GSPN can become too complex for computing the associated steady-state Markov-process solution. With skill-building practice, the analyst can apply symmetry arguments to lump states and thereby reduce the GSPN size to reach a plausible solution.

An example describes a throughput (update rate) comparison between two blue-agent polling disciplines, cyclic and random. Results are compared to a discrete-event simulation of the ASN. A discussion is included of GSPN utility in related areas, e.g., establishing CONOPS, algorithm/software requirements, rapid prototyping prior to detailed simulations, and designing software test suites.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by  
Interface Foundation of North America

*Derivation of the Distribution Function for the Tampered Brownian Motion Process Model*

Arthur Fries  
Institute for Defense Analysis

The tampered Brownian motion process (BMP) arises in the context of partial step-stress accelerated life testing when the underlying system fatigue accumulated over time is modeled by two constituent BMPs, one governing up to the predetermined time point at which the stress level is elevated and the other afterwards. A conditioning argument obtains the probability distribution function (pdf) of the corresponding time-to-failure random variable. This result has been reported and studied in the literature, but its derivation has not been published.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Summary of Findings and Recommendations from the National Research Council Report on “Reliability Growth: Enhancing Defense System Reliability”*<sup>1</sup>

Arthur Fries (Chair) Institute for Defense Analyses, W. Peter Cherry, Science Applications International Corporation, Robert Easterling, Statistical Consultant, Elsayed A. Elsayed, Rutgers University, Aparnav Huzurbazar, Los Alamos National Laboratory, Patricia A. Jacobs, Naval Postgraduate School, William Q. Meeker, Jr., Iowa State University, Nachi Nagappan, Microsoft Research, Michael Pecht, University of Maryland, Ananda Sen, University of Michigan, Scott Vander Wiel, Los Alamos National Laboratory, Constance F. Citro, Director, Committee on National Statistics, Michael L. Cohen, Study Director, Michael J. Siri, Program Associate

This paper summarizes the findings and recommendations from a recent report from the Panel on Reliability Growth Methods for Defense Systems, operating under the auspices of the Committee on National Statistics (CNSTAT) within the National Research Council (NRC). The report offers recommendations to improve defense system reliability throughout the sequence of stages that comprise US Department of Defense (DoD) acquisition processes—beginning with the articulation of requirements for new systems and ending with feedback mechanisms that document the reliability experience of deployed systems. A number of these recommendations are partially or fully embraced by current DoD directives and practice, particularly with the advent of recent DoD initiatives that elevate the importance of design for reliability techniques, reliability growth testing, and formal reliability growth modeling. The report supports the many recent steps taken by DoD, building on these while addressing associated engineering and statistical issues. The report provides a self-contained rendition of reliability enhancement proposals, recognizing that current DoD guides and directives have not been fully absorbed or consistently applied and are subject to change.

<sup>1</sup> As noted below, this presentation is comprised of extracts from the report brief and final report authored by the NRC Panel on Reliability Growth Methods. Permission from the NRC to reproduce their products here is gratefully acknowledged. Any opinions, findings, conclusions, or recommendations expressed in this paper (as well as in the original NRC publications) are those of the authors and do not necessarily reflect those of the panel’s Office of the Secretary of Defense co-sponsors (the Office of the Director of Operational Test and Evaluation (DOT&E) and the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD AT&L)).

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Modeling an Augmented Lagrangian for Blackbox Constrained Optimization*

Robert Gramacy  
University of Chicago

Constrained blackbox optimization is a difficult problem, with most approaches coming from the mathematical programming literature. The statistical literature is sparse, especially in addressing problems with nontrivial constraints. This situation is unfortunate because statistical methods have many attractive properties: global scope, handling noisy objectives, sensitivity analysis, and so forth. To narrow that gap, we propose a combination of response surface modeling, expected improvement, and the augmented Lagrangian numerical optimization framework. This hybrid approach allows the statistical model to think globally and the augmented Lagrangian to act locally. We focus on problems where the constraints are the primary bottleneck, requiring expensive simulation to evaluate and substantial modeling effort to map out. In that context, our hybridization presents a simple yet effective solution that allows existing objective-oriented statistical approaches, like those based on Gaussian process surrogates and expected improvement heuristics, to be applied to the constrained setting with minor modification. This work is motivated by a challenging, real-data benchmark problem from hydrology where, even with a simple linear objective function, learning a nontrivial valid region complicates the search for a global minimum.

A preprint can be found here: <http://arxiv.org/abs/1403.4890>

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Maximum Likelihood Estimation of a Non-homogeneous Poisson Process Software Reliability Model with the Expectation Conditional Maximization Algorithm*

Vidhyashree Nagaraju and Lance Fiondella, University of Massachusetts  
Thierry Wandji, Naval Air Systems Command

The recent National Academies report "Reliability Growth: Enhancing Defense System Reliability" [1] offers several recommendations, including use of reliability growth models to empower hardware and software reliability management teams that direct contractor design and test activities. While there are many software reliability models [2], there are relatively few tools to automatically apply these models [3]. Moreover, these tools are over two decades old and are difficult or impossible to configure on modern operating systems without a virtual machine. To overcome this technology gap, we are developing an open source software reliability tool for the Naval Air Systems Command (NAVAIR), the Department of Defense (DoD), and broader software engineering community. A key challenge posed by such a project is the stability of the underlying model fitting algorithms, which must ensure that the parameter estimates of a model are indeed those that best fit the data. If such model fitting is not achieved users who lack knowledge of the underlying mathematics may inadvertently use inaccurate predictions. This is potentially dangerous if the model underestimates important measures such as the number of faults remaining or mean time to failure (MTTF). To improve the robustness of the model fitting process, we are developing expectation maximization (EM) [4] and expectation conditional maximization (ECM) [5] algorithms to compute the maximum likelihood estimates of nonhomogeneous Poisson process (NHPP) software reliability growth models (SRGM). This talk presents an implicit ECM algorithm for the Weibull NHPP SRGM. The implicit approach eliminates computationally intensive integration from the update rules of the ECM, achieving a speedup of between 200 and 400 times that of explicit ECM methods. The enhanced performance and stability of these algorithms will ultimately benefit the DoD and contractor communities that use the open source software reliability tool.

- 1 Panel on Reliability Growth Methods for Defense Systems, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, “Reliability Growth: Enhancing Defense System Reliability,” National Research Council of the National Academies, Washington D.C, 2015.
- 2 M. Lyu, Handbook of Software Reliability Engineering, 2nd ed. New York, NY, USA: McGraw-Hill, 1996.
- 3 M. Lyu and A. Nikora, "CASRE: a computer-aided software reliability estimation tool," Computer-Aided Software Engineering, pp. 264-275, Jul 1992.
- 4 A. Dempster, L. Nan, and D. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” Journal of the Royal Statistical Society, vol. 39, no. 1, pp. 1-38, 1977.
- 5 X. Meng and D. Rubin, “Maximum likelihood estimation via the ECM algorithm: A general framework,” Biometrika, vol. 80, no. 2, pp. 267-278, June 1993.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Bayesian Hierarchical Models for Common Components Across Multiple System Configurations*

Kassie Fronczyk, Rebecca Dickinson, and Laura Freeman, Institute for Defense Analyses  
Alyson Wilson and Caleb Browning, North Carolina State University

Reliability is a high priority in the testing of military systems. A common interest is reliability growth, which tracks the change in the reliability of a system as it moves through test phases and undergoes periodic corrective actions. A Bayesian hierarchical model is used to assess the reliability of a future combat family of vehicles. The family of vehicles has been through a series of three developmental test events, where fixes occurred only during a set corrective action period. The proposed model effectively combines information across test phase and common vehicle components. By leveraging the information learned in developmental testing, we discuss a framework to determine the test length needed to demonstrate a given reliability threshold in subsequent testing.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Estimating System Reliability from Heterogeneous Data*

Caleb Browning and Alyson Wilson, North Carolina State University  
Laura Freeman, Kassie Fronczyk, and Rebecca Dickinson, Institute for Defense Analyses

Many complex systems are designed to operate in multiple environments under a variety of operating conditions, which makes it challenging to define a single system reliability. In this paper, we consider the problem of defining system reliability for a self-propelled howitzer, which is an artillery piece. A howitzer's reliability may be measured in miles driven for failure modes that are induced by driving or rounds fired for failure modes that are induced by the firing of the howitzer. The full system reliability of the howitzer, however, is defined in hours for a given use profile, which also accounts for time spent idling. In this talk, we propose a methodology for combining information from several test events under different operating conditions to estimate system reliability. The methodology accounts for correlation between the multiple failure modes. Additionally, we propose an experimental design approach to addressing the impact of the individual tasks on system reliability when multiple operational profiles will be employed the user.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*A Design of Experiments approach to evaluating Parameterization Schemes for Numerical Weather Prediction*

Jeffrey A. Smith, US Army Research Laboratory

Richard Penc, US Army Research Laboratory through the Oak Ridge Associated Universities

Numerical Weather Prediction (NWP) is the science of forecasting weather conditions at some time in the future based on present and past observations using complex mathematical/physical model(s) and advanced computational techniques. These predictions can range from hours to months and in some cases years into the future. Spatially, these codes can simulate conditions from small regional microclimates to global climates. Of particular interest to the US Army are codes that can make short term predictions that are valid for a few hours over a roughly brigade sized area of responsibility (AOR).

While both long- and short-term forecasts are of interest to a brigade commander for planning purposes, it is at the meso- and microscales where forecasts can enhance situational awareness for the echelons below brigade. Mesoscale models resolve meteorological features such as thunderstorms or microbursts that typically are on the order of few than 10 km in extent. Microscale models resolve features are often less than a kilometer in extent such as turbulent flows in an urban canyon. Temporally, these meso- and microscale models fill a gap between coarse synoptic forecasts of large scale features such as weather fronts that update typically every six-to- twelve hours. For US Army purposes, the term “Nowcast” refers to forecasts of mesoscale conditions up to six hours into the future which can be updated as frequently as every 30-60 minutes.

Nowcast models are typically high-resolution applications of research and operational mesoscale NWP models such as the Advanced Research Weather Research and Forecast [1] model (WRF-ARW) with extensions that might include finer terrain resolutions and more detailed representations of the underlying atmospheric physics. Functionally, these models take as initial conditions both very coarse-grained synoptic forecasts and observational data from various meteorological data sources to produce their higher resolution Nowcasts.

The foundation of NWP models is the conservation of mass, heat, motion, and water vapor as well as other gaseous and aerosol materials [2, 3] over the region of interest. NWP codes model these properties through a set of coupled partial differential equations with the first-guess initial and time-dependent later boundary tendencies provided by a forecast from a coarse-grained synoptic model. Because these models seek to represent climate physics at a high resolution, the transition from coarse- to fine-grain is handled via a nesting strategy so that boundary conditions enter into the model more smoothly. Typically, this nesting occurs through telescopic “multi” nesting approach tapering from an outer domain on the



2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

order of 1500km by 1500 km to an inner domain that may be a 100km by 100km in extent. In this approach, the outer nest moves the lateral boundaries as far away from the desired model center of interest as possible so that tendencies from the external coarse-grain model pass gradually into the model domain. The middle nest (or several) acts as an intermediate resolution nest(s) for ideal downscaling. Finally, the interior domain captures the domain of interest at the highest resolution desired. Such an approach is also called a “limited-area” mesoscale NWP configuration.

While the conservation balance equations govern the transport of mass, energy, etc., throughout the model, other physics (often non-resolved processes such as turbulence whose effects must be estimated) are incorporated via ‘parameterizations’ which capture effects of cloud cover, turbulence, solar radiation, etc. [4]. The collection of these conservation balance equations (partial differential equations which are linearized and solved numerically) along with the various parameterizations constitutes the typical NWP model.

A single parameterization scheme will fall into one of five groups: Microphysics, Cumulus, Radiation, Planetary Boundary Layer, or Surface effects which capture respectively physics such as moisture, clouds, solar radiation, turbulence and land cover. The sub-grid effects are a consequence of the interaction between representative schemes from each group during the simulation execution for the conservation principles to hold. Within each group, there are a number of different approaches to handling that particular effect, each approach suitable for a range of conditions.

If one considers only the surface layer and boundary layer parameterizations, there are 15 combinations of these two physical schemes possible. If one takes all the available combinations of physics options in WRF-ARW, there are over 2 million possible combinations; thus, on a battlefield or in a remote deployment it is not possible to perform forecasts with all these combinations to find the ‘single’ set that best describes the local weather. Furthermore, each scheme or combination of schemes quite often only works best in certain environments. Given that we cannot know in advance where a brigade will be deployed; we seek a combination parameterizations that works ideally well in all situations, but practically well within a defined region.

If we consider our modeling system a “grey box”, we can model the outputs (for example forecast skill) as a statistical function of the initial conditions along with the internal physics that constitutes the “grey box”. Consequently, our task becomes to appropriately sample that function in a manner that minimizes both the computational burden we face and allows us to maximize the amount of information we can extract from our set of model runs. Ideally, one should approach the selection a candidate set of parameterizations via a method that leads to robust performance under a variety of weather conditions for a given domain, and one such approach is through the use of Design of Experiments techniques.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

Here we intend a design point be a single sample of the “grey box simulation” whose value will be an output figure of merit such as forecast skill. Consequently, the design as a whole can be used to support identification of a suite of parameterization schemes that collectively produce a “skillful” forecast over a variety of conditions.

#### References

1. Skamarock, W.C., et al., A Description of the Advanced Research Wrf Version 3. 2008. NCAR Technical Note NCAR/TN-475+STR.
2. Pielke, R.A., Sr., Mesoscale Meteorological Modeling. 2nd ed. International Geophysics Series. 2002, San Diego: Academic Press.
3. Warner, T.T., Numerical Weather and Climate Prediction. 2011, Cambridge: Cambridge University Press.
4. Stensrud, D.J., Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models. 2007, Cambridge: Cambridge University Press.

2015 “Second Conference on Applied Statistics in Defense (CASD)”, sponsored by Interface Foundation of North America

*Ammunition Projectile Chemical Coating Experiment with Design Oriented Modeling*

Doug Ray, US Army Armament Research, Development and Engineering Center  
Tom Donnelly, SAS Institute Inc., JMP Division

A 22-trial Definitive Screening Design (DSD) was run to characterize the effects of 7-continuous and 2-categorical factors on the Salt-Spray Time-to-Failure (SS TTF) of chemical coatings on munitions. With only a subset of factors important, a reduced model was used to optimize the process. Conservative and aggressive regression modeling approaches are compared, including the very recently developed Design Oriented Modeling approach.