

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188				
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>							
1. REPORT DATE (DD-MM-YYYY) 28-08-2016		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 15-Oct-2015 - 14-Oct-2016			
4. TITLE AND SUBTITLE Final Report: Conference on Applied Statistics in Defense (CASD 2015)			5a. CONTRACT NUMBER W911NF-15-1-0652				
			5b. GRANT NUMBER				
			5c. PROGRAM ELEMENT NUMBER 611102				
6. AUTHORS Don R Faxon (PI), Edward J Wegman (Treasurer)			5d. PROJECT NUMBER				
			5e. TASK NUMBER				
			5f. WORK UNIT NUMBER				
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Interface Foundation of North America, Inc. 9748 Thorn Bush Drive Fairfax Station, VA 22039 -2537			8. PERFORMING ORGANIZATION REPORT NUMBER				
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO				
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 68258-MA-CF.1				
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited							
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.							
14. ABSTRACT CASD is a conference of the Interface Foundation of North America and a forum for the presentation and discussion of theoretical and applied papers relating to the use of probability and statistics for solving problems of national importance, with an emphasis on defense and national security. CASD provides valuable opportunities for constructive interaction among academic, industry, and government agency scientists. CASD also serves a nurturing role in the elevation of statistical proficiency among researchers in other disciplines who find themselves statistical practitioners because of the compelling benefits statistical science brings to research, development, and							
15. SUBJECT TERMS CASD 2015, Data Sciences							
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT		15. NUMBER OF PAGES		19a. NAME OF RESPONSIBLE PERSON	
a. REPORT	b. ABSTRACT					c. THIS PAGE	Don Faxon
UU	UU	UU	UU			19b. TELEPHONE NUMBER	
						571-437-7696	

Report Title

Final Report: Conference on Applied Statistics in Defense (CASD 2015)

ABSTRACT

CASD is a conference of the Interface Foundation of North America and a forum for the presentation and discussion of theoretical and applied papers relating to the use of probability and statistics for solving problems of national importance, with an emphasis on defense and national security. CASD provides valuable opportunities for constructive interaction among academic, industry, and government agency scientists. CASD also serves a nurturing role in the elevation of statistical proficiency among researchers in other disciplines who find themselves statistical practitioners because of the compelling benefits statistical science brings to research, development, and testing.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

<u>Received</u>	<u>Paper</u>
-----------------	--------------

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

<u>Received</u>	<u>Paper</u>
-----------------	--------------

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 27.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Book

TOTAL:

Received

Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Post Doctorates

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Under Graduate students supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:..... 0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME
Total Number:

Names of other research staff

NAME PERCENT SUPPORTED
FTE Equivalent:
Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Technology Transfer

CASD 2015 19-22 October, 2015
Interface Foundation of North America, Inc., EIN 54-1522706

Executive Summary

About CASD

The Conference on Applied Statistics in Defense (CASD) is a forum for the presentation and discussion of theoretical and applied papers related to the use of probability and statistics in solving problems of National importance. The conference provides an invaluable opportunity for interaction among academic, industry, and government scientists. It also serves a nurturing role in the elevation of statistical competence among researchers in other disciplines who find themselves statistical practitioners because of the compelling benefits statistical science brings to research, development, and testing.

The conference is structured as follows. General sessions offer invited presentations on emerging topics in statistics. Special sessions focus attention on National problem areas in which the role of statistical science can be more fully explored. Contributed sessions include papers ranging from new methodology to interesting statistical application. In the clinical sessions, a distinct feature of the conference, presenters can seek guidance from a panel of experts for problems that have not been completely or satisfactorily solved. *National problem context is strongly encouraged as the motivation or as a potential application for all work presented, but the focus is to be statistics.*

Two other important events take place at the conference. A tutorial on a statistical topic believed to be important in National application precedes each conference. The Army Wilks Award is given periodically at this conference to an individual who has made substantive contributions in the area of national defense to commemorate the service of Sam Wilks.

CASD, originally ACAS (or ICAS as it was briefly known as in 2013), was established in 1995 by six sponsoring activities within the national defense community. In the ensuing years, the organization has broadened its mission to include context areas of National importance and has attracted more industry, academia, and other government agencies. In 2013, CASD was established to reflect the broadening role. The present direction represents a natural progression from being a conference under Army direction (1995-2004) to one under the direction of the Interface (2005-2012) to one that continues under the Interface but with an expanded role in the statistical community. In name this is suggested by the replacement of Army with Interface. CASD is the successor to the Conference on the Design of Experiments in Army Research, Development, and Testing, a historic series of meetings that formally concluded in 1994 after 40 years of service to the Army.

Program in Brief

- **Welcome:**
- **Short Course:**

Bayesian Inference

Prof. Robert Gramacy (University of Chicago) will present a short course "Bayesian
Enclosure 2

Inference " on October 19 & 20 preceding the conference beginning on October 21.

- **Invited Speakers:**

- Catherine Warner, Institute for Defense Analyses
Rigorous, Defensible, Efficient Testing
- Sallie Keller, Virginia Tech
Data, Models, and Analytics
- Edward Wegman, George Mason University
Malware Static Analysis Techniques Using a Multidisciplinary Approach
- Bradley Jones, SAS/JMP
Design Oriented Modeling: A Powerful Approach for Model Selection for Definitive Screening Designs

- **Special Sessions:**

- **Design of Experiments**
Organizer: Tom Donnelly, SAS/JMP

Design and Analysis for Constraining Dynamic Network Simulation Models with Observational Data
David Higdon, Jiangzhuo Chen, Virginia BioInformatics Institute, Virginia Tech

Incorporating Confidence Intervals On the Decision Threshold in Logistic Regression
Michael J. Kist and Rachel T. Silvestrini, Industrial and Systems Engineering Department, Rochester Institute of Technology, Rochester, NY

Covering Arrays: A Tool to Construct Efficient Test Suites for Validating Software Systems
Joseph Morgan and Ryan Lekivetz, SAS/JMP
- **Text Mining**
Organizer: Ed Wegman, George Mason University

Mining Text Networks
David Banks, Duke University

Modeling the Evolution and Influence of Blogger's Topic Sentiment and Opinion Profile
Michael James Garrity

Preserving Semantic Content in Text Mining Using Multigrams
Yasmin H. Said

o **Cooperative Studies at the VA**

Organizers: Barry Bodt, Army Research Laboratory, and David Webb, Army Test and Evaluation Command

CSP #579: The Health of Vietnam Era Women's Study (HealthViEWS)
Tracey Serpi, Perry Point Cooperative Studies Program Coordinating Center, VA

CSP #504: Risperidone Treatment for Military Service Related Post-Traumatic Stress Disorder (PTSD)
Karen Jones, Perry Point Cooperative Studies Program Coordinating Center, VA

• **Contributed Papers:**

Tractable Analysis of Experiments on High-dimensional Simulation Codes
Matthew Plumlee, University of Michigan

Computational Geometry for Multivariate Statistics
John Nolan, American University

Statistical Design and Validation of Modeling and Simulation (M&S) Tools Used in Operational Testing
Kelly McGinnity and Laura Freeman, Institute for Defense Analyses

Learning Bayesian Network Structure using Data from Designed Simulation Experiments
Jarom Ballantyne, Terril Hurst, Allan Mense, Raytheon Missile Systems

Modeling Adversarial Sensor Networks Using Generalized Stochastic Petri Nets
Terril Hurst, Kris Kukla, Daniel Rosser, James Kimmet, Raytheon Missile Systems

Derivation of the Distribution Function for the Tampered Brownian Motion Process Model
Arthur Fries, Institute for Defense Analysis

Summary of Findings and Recommendations from the National Research Council Report on "Reliability Growth: Enhancing Defense System Reliability"
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

Modeling an Augmented Lagrangian for Blackbox Constrained Optimization
Robert Gramacy, University of Chicago

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

Vidhyashree Nagaraju¹, Thierry Wandji², and Lance Fiondella^{1*}, University of
Massachusetts¹, Naval Air Systems Command²

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

Kassie Fronczyk¹, Rebecca Dickinson¹, Alyson Wilson², Caleb Browning², and
Laura Freeman¹, Operational Evaluation Division, Institute for Defense
Analyses¹ and Department of Statistics, North Carolina State University²

Estimating System Reliability from Heterogeneous Data

Caleb Browning¹, Laura Freeman², Alyson Wilson¹, Kassie Fronczyk², and
Rebecca Dickinson², Department of Statistics, North Carolina State University¹,
Operational Evaluation Division, Institute for Defense Analyses²

*Ammunition Projectile Chemical Coating Experiment with Design Oriented
Modeling*

Doug Ray, U.S. Army Armament Research, Development and Engineering
Center (ARDEC), Tom Donnelly, SAS Institute Inc., JMP Division

*A Design of Experiments approach to evaluating Parameterization Schemes for
Numerical Weather Prediction (Clinical Paper)*

Jeffrey A. Smith, US Army Research Laboratory and Richard Penc, US Army
Research Laboratory through the Oak Ridge Associated Universities

- **Events:**

Mixer, Wednesday, 21 October

Wilks Luncheon, Thursday, 22 October

Short Course:

Bayesian Inference
October 19 & 20

- Prof. Robert Gramacy, University of Chicago

Course Preparation:

REPORT DOCUMENTATION PAGE (SF 298 Continuation Sheet)

Please download all files and print out the lecture notes and exercises. It will be important for you to have access to the electronic versions of the R files (in the demos directory) so you can follow along on your own computer. You can also adapt the code(s) to work on the in-class exercises, as necessary. New link to course materials: [Download Materials](#).

We were asked about R packages needed to run the code in the demos directory. Please have the following installed on your machines, along with R: coda, ellipse, mvtnorm, akima.

R should install/load any subsequent packages that are needed with the above.

Course Logistics:

Dates: October 19 & 20

Times: 9:00 AM to 5:00 PM

Location: The Hub

Rooms: 3 & 4

Course Description:

Synopsis:

This course will cover the basics of the Bayesian approach to practical and coherent statistical inference. Particular attention will be paid to computational aspects, including MCMC. Examples/practical hands-on exercises will be run gamut from toy illustration to real-world data analysis from all areas of science, with R implementations/coaching provided.

Texts:

There is no required text, however the course closely follows P.D. Hoff's *A First Course in Bayesian Statistical Methods* - Springer 2009. Some examples are borrowed from two other texts which are nice references to have. J. Albert's' *Bayesian Computation with R* - Springer 2nd ed. 2009; A. Gelman, J.B. Carlin, H.S. Stern and D.B. Rubin's *Bayesian Data Analysis* - 2nd ed. 2004

Prerequisites:

Undergraduate level Statistics, Probability, Calculus, Linear Algebra, and a familiarity with a programming language (e.g., R, Matlab, Python, C or Fortran)

Preparation:

Students are expected to install and familiarize themselves with R in advance of the course.

A short tutorial with accompanying R code is provided here:

slides <http://bobby.gramacy.com/teaching/bayes/rtutor.pdf>

code <http://bobby.gramacy.com/teaching/bayes/rtutor.R>

Additionally, some other useful resources prepared by others include

video tutorials <http://dist.stat.tamu.edu/pub/rvideos/>

step-by-step guides <http://tryr.codeschool.com>

REPORT DOCUMENTATION PAGE (SF 298 Continuation Sheet)

Schedule at a Glance:

<u>Start</u>	<u>End</u>	<u>Monday</u>	<u>Start</u>	<u>End</u>	<u>Tuesday</u>
9:00	12:00	Pre-conference Tutorial	9:00	12:00	Pre-conference Tutorial
12:00	1:00	Lunch	12:00	1:00	Lunch
1:00	4:30	Pre-conference Tutorial	1:00	4:30	Pre-conference Tutorial
<u>Start</u>	<u>End</u>	<u>Wednesday</u>	<u>Start</u>	<u>End</u>	<u>Thursday</u>
8:45	9:00	Welcome Remarks	9:00	10:00	General IV - Edward Wegman
9:00	10:00	Keynote - Catherine Warner			
10:00	10:15	Welcome - Peggy Agouris			
10:15	10:30	BREAK	10:00	10:30	BREAK
10:30	12:00	Special 1 - Text Analysis Contributed 1 - Reliability Growth	10:30	12:00	Contributed 3 - Modeling and Simulation Contributed 4: Testing
12:00	1:00	LUNCH (on own)	12:00	2:00	LUNCH (included) Award Presentation
1:00	2:30	Special 2 - Design of Experiments Contributed 2: Advances in Statistics	2:00	3:00	Special 3 - Cooperative Studies at the VA
2:30	3:00	BREAK	3:00	3:30	Closing Remarks
3:00	4:00	General Session II - Sallie Keller			
4:00	5:00	General Session III - Bradley Jones			
5:00	5:30	Clinical Session			
6:00	7:00	Social - TBD			

ABSTRACTS

Tutorial - Bayesian Inference Using R

Bayesian Inference Using R

Prof. Robert B. Gramacy, Booth School of Business, The University of Chicago

Abstract: This course will cover the basics of the Bayesian approach to practical and coherent statistical inference. Particular attention will be paid to computational aspects, including MCMC. Examples/practical hands-on exercises will the run gamut from toy illustration to real-world data analysis from all areas of science, with R implementations/coaching provided. There is no required text, however the course closely follows P.D. Hoff's *A First Course in Bayesian Statistical Methods*, Springer 2009. Some examples are borrowed from two other texts which

Enclosure 2

are nice references to have. J. Albert's' *Bayesian Computation with R*, Springer 2nd ed. 2009; A. Gelman, J.B. Carlin, H.S. Stern and D.B. Rubin' s *Bayesian Data Analysis*, 2nd ed. 2004.

General Session I - Keynote Address

Keynote Address: Rigorous, Defensible, Efficient Testing

Catherine Warner, Institute for Defense Analysis

Abstract: 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.

Special Session I, Text Analysis (parallel)

Mining Text Networks

David Banks, Duke University

Abstract: 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 U.S. 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.

Modeling the Evolution and Influence of Blogger's Topic Sentiment and Opinion Profile

Michael James Garrity

Abstract: 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.

Preserving Semantic Content in Text Mining Using Multigrams

Yasmin H. Said

Abstract: 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 nC_2 ; it is nonetheless usually considerably larger than n , where n is the number of words in the lexicon.

Contributed Session I - Reliability Growth (parallel)

Bayesian Hierarchical Models for Common Components across Multiple System Configurations

Kassie Fronczyk¹, Rebecca Dickinson¹, Alyson Wilson², Caleb Browning², and Laura Freeman¹
Operational Evaluation Division, Institute for Defense Analyses¹ and Department of Statistics,
North Carolina State University²

Abstract: 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.

***Summary of Findings and Recommendations from the National Research Council Report on "Reliability Growth: Enhancing Defense System Reliability"*¹**

Arthur Fries, Institute for Defense Analysis

Abstract: 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 U.S. 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.

[Note] ¹ 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)).

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

Vidhyashree Nagaraju¹, Thierry Wandji², and Lance Fiondella^{1*}, ¹University of Massachusetts, North Dartmouth, MA 02747, ²Naval Air Systems Command, Patuxent River, MD 20670

Abstract: The recent National Academies report “Reliability Growth: Enhancing Defense System Reliability”² 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 models³, there are relatively few tools to automatically apply these models.⁴ 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)⁵ and expectation conditional maximization (ECM)⁶ 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.

[Notes] ² 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.

³ M. Lyu, *Handbook of Software Reliability Engineering*, 2nd ed. New York, NY, USA: McGraw-Hill, 1996.

⁴ M. Lyu and A. Nikora, "CASRE: a computer-aided software reliability estimation tool," *Computer-Aided Software Engineering*, pp. 264-275, Jul 1992.

⁵ 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.

⁶ 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.

Special Session II - Design of Experiments (parallel)

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

Dave Higdon and Jiangzhuo Chen, Social Decisions Analytics Laboratory, Virginia Bioinformatics Institute, Virginia Tech

Abstract. 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.

Incorporating Confidence Intervals on the Decision Threshold in Logistic Regression

Michael J. Kist and Rachel T. Silvestrini, Industrial and Systems Engineering Department, Rochester Institute of Technology, Rochester, NY

Abstract. 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.

Covering Arrays: A Tool to Construct Efficient Test Suites for Validating Software Systems

Joseph Morgan and Ryan Lekivetz, SAS Institute Inc., JMP Division

Abstract: 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.

Contributed Session II - Advances in Statistics (parallel)

Modeling an Augmented Lagrangian for Blackbox Constrained Optimization

Robert B. Gramacy, Booth School of Business, The University of Chicago

Abstract: 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>

Computational Geometry for Multivariate Statistics

John Nolan, Math/Stat Department, American University

Abstract: 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.

Estimating System Reliability from Heterogeneous Data

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

Enclosure 2

Abstract: 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.

General Session II

Data, Models, and Analytics

Dr. Sallie Keller, Professor of Statistics and Director, Social and Decision Analytics Laboratory, Virginia Bioinformatics Institute at Virginia Tech

Abstract: 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.

General Session III

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

Bradley Jones, SAS Institute

Abstract: 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.

Clinical Session

A Design of Experiments Approach to Evaluating Parameterization Schemes for Numerical Weather Prediction

Jeffrey A Smith and Richard Penc, U.S. Army Research Laboratory

Abstract: 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 U.S. 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).

Special Session III - Studies from the Veteran's Administration

CSP 579: The Health of Vietnam Era Women's Study (HealthViEWS)

Tracey Serpi, Perry Point Cooperative Studies Program Coordinating Center, VA

Abstract: 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 U.S. 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

Enclosure 2

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.

CSP #504: Risperidone Treatment for Military Service Related Post-Traumatic Stress Disorder (PTSD)

Karen Jones, Perry Point Cooperative Studies Program Coordinating Center, VA

Abstract: 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 nonresponders 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.

Contributed Session III - Modeling and Simulation (parallel session)

Tractable Analysis of Experiments on High-dimensional Simulation Codes

Matthew Plumlee, University of Michigan

Abstract: 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.

Modeling Adversarial Sensor Networks Using Generalized Stochastic Petri Nets

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

Abstract: 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.

Learning Bayesian Network Structure using Data from Designed Simulation Experiments

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

Abstract: 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.

Contributed Session IV - Testing (parallel session)

Bayesian Hierarchical Models for Common Components across Multiple System Configurations

Kassie Fronczyk¹, Rebecca Dickinson¹, Alyson Wilson², Caleb Browning², and Laura Freeman¹.
Operational Evaluation Division, Institute for Defense Analyses¹ and Department of Statistics,
North Carolina State University²

Abstract: 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.

Derivation of the Distribution function for the Tampered Brownian Motion Process Model⁷

Arthur Fries, Institute for Defense Analysis

Abstract: 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.

[Note] ⁷ This paper is scheduled to appear in the Gnedenko e-Forum International Group on Reliability electronic journal: *Reliability: Theory & Applications*, Vol. 10, No. 3, September 2015.

Ammunition Projectile Chemical Coating Experiment with Design Oriented Modeling

Doug Ray, U.S. Army Armament Research, Development and Engineering Center (ARDEC), Tom Donnelly, SAS Institute Inc., JMP Division

Abstract: 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.

General Session IV

Malware Static Analysis Techniques Using a Multidisciplinary Approach

Edward J. Wegman⁸

Abstract: 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.

[Note] ⁸ This is joint work with Muhammed Aljammaz.

REPORT DOCUMENTATION PAGE (SF 298 Continuation Sheet)

Demographics:

Total Participants: 51 Total Presentations: 27 (includes short course)
 (Total on-site attendees: 37; Total contributing authors: 36)

Category	Count
com	20
edu	15
gov	3
mil (AF)	1
mil (Army)	4
mil (Navy)	1
mil (USMC)	3
other	2
student	2

51

Breakout of Demographics by Categories						
com	edu	gov	mil (AF)	mil (Army)	mil (Navy)	mil (USMC)
ARS	American	BLS	AFIT	AMC	C N A	MCOTEA
Boeing	Brigham Young	DOT&E	AFLCMC	AMC-AFRIMS	Navy	USMC
Census	CMU	EPA		APG	NASC	
CEW	Columbia	JPL		ARDEC	NRL	
Digital System	Cornell	L. Livermore		ARO	NSWC	
Elder Research	Duke	Los Alamos		ARL		
Engility Gp	GMU	NGA		Army		
Google	GWU	NIST		ATEC		
IDA	IASTATE	OASD		ISMED		
JRAD	Indiana	Sandia		USMA		
Mathematica	NCSU	VAMHCS		WRAIR		
Qualcomm	PSU			ARL-WSMR		
Raytheon	Rice					
SAS	RIT					
TQI	SUNY Buffalo					
	Texas Tech					
	U Miami					
	U Michigan					
	UC Davis					
	UChicago					
	UMASS					
	UMB					
	UT Austin					
	UW					
	VT					
	W&M					