Science of Test Research Consortium:

Year Two Final Report

Raymond R. Hill

Douglas C. Montgomery

G. Goeffrey Vining

Rachel T. Silvestrini

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.
AFIT/EN/TR-13-01

Abstract

The Science of Test (SOT) Research Consortium is sponsored and funded by both Office of Secretary of Defense (OSD) Director, Operational Test and Evaluation (DOT&E) and by the Test Resource Management Center (TRMC) within the OSD Director, Developmental Test and Evaluation. The consortium members include the Air Force Institute of Technology, Department of Operational Sciences, Arizona State University, School of Computing, Informatics, and Design Systems Engineering, Virginia Tech, Statistics Department and Naval Postgraduate School, Operations Research Department. The SOT research effort commenced in early CY 2011. This report summarizes the accomplishments through the second year of the research consortium. The consortium collectively contributes to the theory of the statistical rigor of test with contributions to the body of knowledge and to the practice of test with applied methods and consultations.

Science of Test Research Consortium:

Year Two Final Report

I. Introduction

The Science of Test (SOT) Research Consortium is sponsored and funded by both Office of Secretary of Defense (OSD) Director, Operational Test and Evaluation (DOT&E) and by the Test Resource Management Center (TRMC) within the OSD Director, Developmental Test and Evaluation. The consortium members include the Air Force Institute of Technology, Department of Operational Sciences, Arizona State University, School of Computing, Informatics, and Design Systems Engineering, Virginia Tech, Statistics Department and Naval Postgraduate School, Operations Research Department.

The SOT effort commenced in early CY 2011 with an official kick-off meeting held at the ASU campus in March 2011. A status update to OSD DOT&E was provided in September 2011 and again in February 2012 (see Appendix A for slides). A second consortium meeting was held in December 2011, once again at the ASU campus. The third meeting of the consortium is scheduled for Oct 2, 2013 in St. Louis, MO.

II. Goals and Objectives

The Department of Defense (DOD) Test Enterprise is responsible for test and evaluation policy, planning, execution, and analysis of test results. Additionally, this community is responsible for resource allocation, and test workforce development. Each of the Services plays a critical role both internally with respect to service unique test requirements as well as collaboratively on joint test concerns. The Service Operational Test Agency (OTA) Commanders and the DOT&E are dependent on credible test methodology and analysis to provide critical information and program assessment data during program evaluation within the test phases of acquisition.

The overall goal of this research is to support the integration of advanced statistical rigor and mathematical foundations into the test domain. The increased technical capabilities are designed to improve the 'science of test' across the DOD and leverage academic research to provide solutions to challenging problems facing the test communities. The research is ideally conducted in collaboration with service test organizations to ensure applicability of test design and methodologies to DOD systems being developed within the acquisition process. These systems could benefit from research on applying advancement test design methodology to their test plans and processes.

Specific goals and objectives as defined at the consortium meetings and as updated to OSD DOT&E are:

- Conduct basic research into the myriad challenges that remain in the statistical design of experiments;
- Make fundamental contributions to the body of knowledge associated with statistically-valid test and evaluation;
- Provide methods and processes to improve the statistical rigor associated with DoD test and evaluation;
- Transition to practice using our "interested parties"; and
- Support OSD and Service goals with respect to education and training of the future workforce.

Appendix B contains the general project summary as detailed in the proposal delivered by AFIT on behalf of the SOT Research Consortium.

III. Accomplishments

The SOT objectives are focused on basic research into statistical methods for test and evaluation. Graduate student and Research Assistant research are the primary mechanisms to achieving these objectives. Research papers published to date are listed in Appendix C. Graduate theses and doctoral dissertation efforts completed and in-progress are listed in Appendix D. As applicable these documents will be made available via the SOT research website, at http://dev.scienceoftest.com. A discussion of specific accomplishments by consortium member follows.

Arizona State University

ASU has recently completed major development and begun testing and content aggregation for the DOE website project, and is conducting research concerning the cost-optimal use of non-regular experimental designs. The website has been constructed to provide DOD personnel access to academic papers and videos on topics ranging from introductory DOE tutorials to advanced methodologies published in the most current literature. The website boasts a sophisticated and scalable indexing system that enables fast and intuitive searching of a multitude of DOE topics. New content is being continuously added, including research papers published by consortium members and video lectures by consortium professors.

Doctoral research has involved the cost-optimal selection of screening experimental designs and pathologies/strategies for Bayesian experimental design with nonlinear models. Capt Brian Stone has developed a methodology to select screening designs based on minimum expected experimental cost. He has submitted a paper on this topic to the journal of Quality Engineering and presented the research at the Military Operations Research Society Symposium (MORSS), the Quality and Productivity Research Conference (QPRC), and the Operations Analyst (OA) forum. Currently Capt Stone is researching the construction of no-confounding screening designs which have the potential to save significant experimental costs that an experimenter may incur using traditional designs.

Mr. Hassler has continued development on the DOE website project and, in concert with Dr. Silvestrini, has investigated problems with using Bayesian experimental design with nonlinear models. Bayesian nonlinear experimental design is currently implemented in commercial software (e.g. JMP) yet little is known about the performance of such methods. Mr. Hassler has identified several situations where prior probability interacts with the rate of convergence for asymptotic results that make up the canonical analysis for nonlinear models. Currently Mr. Hassler is investigating possible guidelines for prior probability specification and corrections to the design procedure to avoid said pathologies.

Virginia Tech

The primary objective of the 2012 Virginia Tech research effort was to finish Jennifer Kensler's dissertation entitled, "Analysis of Reliability Experiments with Random Blocks and Subsampling," which she successfully defended in July 2012. The first paper from her dissertation illustrated how to take a naïve approach for analyzing a blocked reliability experiment with subsampling. As such, it represented an extension of an approach proposed in Freeman and Vining (2010). This paper was submitted to the Journal of Quality Technology. The review recommended a "revise and resubmit." However, the suggested revisions were more in line with the proposed second paper, which proposes a more rigorous maximum likelihood approach and is an extension of Freeman and Vining (2012). As a result, we decided to revise the first paper and submit it to Quality Engineering and to submit the proposed second paper as the revise and submit version to the Journal of Quality Technology.

We expect to submit the paper to Quality Engineering by November 1, 2012. We hope to submit the paper to the Journal of Quality Technology by January1, 2013.

The second objective of our research efforts was to work with Laura Freeman at IDA on a series of small projects. We were able to pull together a small group of graduate students (Jennifer Kensler and Rebecca Dickinson) and one post-doc (Anne Ryan) . We have undertaken two specific projects at this point.

The first small project is entitled "A Tutorial on Planning Experiments." This paper takes a detailed look at the basic process for planning experiments. As such, it builds upon the process outlined in Coleman and Montgomery (Technometrics 1993). The primary audience for this tutorial is quality and engineering practitioners with some basic background in the statistical planning of experiments. This group includes many military personnel involved in the development and testing of weapon systems. We intend to submit this paper to Quality Engineering by November 1, 2012.

The second small project is an extension of the first. Its focus is the planning of experiments within highly constrained regions. Its basic point is that experimenters must avoid naïve approaches when confronted with highly constrained, irregular regions. Both naïve applications of classical and optimal designs have serious issues. Interestingly, this paper outlines circumstances where the classical design approaches work extremely well when a naïve application of optimal designs do not. The key to success is to apply creative approaches to the definitions of factors and their levels. Such approaches improve the effectiveness of both classical and optimal designs. We hope to submit this paper early in 2013.

A third objective of our research efforts was to develop appropriate training materials on basic statistics and experimental designs for Air Force and other military personnel. We completed work on a basic training module for the Defense Acquisitions University in early 2012. Anne Ryan and Jennifer Kensler were essential to this effort.

A fourth objective of this research project was to recruit civilian personnel to support the military efforts to apply sound experimental design strategies in weapons development and testing. Jennifer Kensler began working at the new AFIT Center of Excellence in August 2012.

The final 2012 objective was to recruit another graduate student to follow Dr. Kensler. Rebecca Dickinson has agreed to work on this project. She worked the summer of 2012 as an intern at IDA under the supervision of Laura Freeman.

Ms. Dickinson intends to extend Dr. Kensler's work to split-plot experimental situations where some of the experimental factors are much harder to change than others. We expect Ms. Dickinson to make her formal dissertation proposal in the last half of 2013.

Air Force Institute of Technology

The AFIT research focused on support to the KC-46 program, to the 46th Test Group in their vulnerability research, and to the Simulation Analysis Facility (SIMAF) Live, Virtual, Constructive (LVC) testing. Research supporting the KC-46 developed a response surface methodology for examining time series data collected via experimental design. This methodology is being examining as a way to improve high-fidelity simulator validation efforts. The methodology is also being applied as we consider ways to reduce the overall test size and improve the design efficiency of the Aerial Refueling Airplane Simulator Qualification (ARASQ) flight test protocol. The methodology was published in Quality and Reliability Engineering International.

Research supporting the vulnerability research within the 46th Test Group focused on developing an empirical model of the flash events that occur when missile fragments impact aircraft causing flashes and aircraft skin penetration. The model is currently implemented in the Joint survivability analysis model, COVART, and is designated AFITFlash 1.0. Current research is improving the predictive capabilities of the model focusing on a version for lethality as well as survivability, and adding a stochastic component.

SIMAF support consisted of validation of radar models for a large-scale study and providing statistical planning support during the early development of a new large-scale study.

Doctorial research supported is focused on two topics. Mr. Alex Gutman is examining supersaturated designs and Lt Col Shane Dougherty is examining non-linear screening designs for aeronautical engineering test applications. Both efforts plan to use flight test data from the Air Force Test Center, Edwards AFB, CA.

Naval Postgraduate School

The NPS research has focused on situations were data are scarce. This research is studying and developing alternative techniques for generating distributions based on data sets that are very small. Also, looking at the issue associated with the misspecification of Bayes' prior distributions in the definition of non-linear optimal experimental designs. This research is focused on studying the effect for several non-linear optimal design models including general linear models (GLM). The work is being written up for journal submission in November 2012. Future research will focus on power estimates for the GLM based on design choice and sample size. A new avenue of work will examine experimental design selection to reduce bias in parameter estimates.

There has also been some focus on workforce development. A five-day course, "Design of Experiments for IT Systems Test and Evaluation" was developed and presented at Ft. Huachuca in January 2012 and at Ft. Meade in February 2012. Plans are to create a shortened version and make it publically available via the SOT website.

Appendix A: February SOT Update to OSD DOT&E

Science of Test Research Consortium

VirginiaTech Invent the Future

ASU Ira A. Fulton Schools of Engineering ARIZONA STATE UNIVERSITY

NPS PRAESTANTIA PER SCIENTIAM

Science of Test Research Consortium Research Status

Science of Test Research Consortium

Overall Goals & Objectives

- Conduct basic research into the myriad challenges that remain in the statistical design of experiments
- Make fundamental contributions to the body of knowledge associated with statistically-valid test and evaluation
- Provide methods and processes to improve the statistical rigor associated with DoD test and evaluation
  - Transition to practice using our "interested parties"
- Support OSD and Service goals with respect to education and training of the future workforce

Arizona State University

ASU's Efforts

- Cost-constrained experimental designs
- In developmental testing, test planners work within a budget to design the best possible test
- There are trade-offs between experimental cost and:
  - Prediction variance optimality (G or I-optimality)
  - Coefficient estimate variance optimality (D-optimality)
  - Power
- Multi-criteria experimental designs are Pareto-optimal designs which simultaneously consider two or more of these criteria

Virginia Tech
Geoff Vining

Virginia Tech's Efforts

- Training Materials
  - Anne Ryan (post doc) and Jennifer Kensler (grad student)
  - Supervised by Laura Freeman and Geoff Vining
  - Continuous Learning Modules (CLMs) for the DAU
  - Completed Basic Statistics Sequence
  - Vetted by the DoD Community
- Ongoing Current Research
  - Jennifer Kensler's Dissertation
    - Random Block Experiments for Reliability Data
    - Estimated Completion Date: Summer 2012
  - Experimental Designs in Highly Constrained Regions

Virginia Tech's Efforts

- New Initiatives
  - Experimental Designs for Irregular Regions
    - Jennifer Kensler, Anne Ryan, Rebecca Dickinson (grad student)
    - Supervised by Laura Freeman and Geoff Vining
    - Target Journal: Quality Engineering
    - Expected Submission Date: March-April 2012
  - Other Topics (Roll-Out Over 2012)
    - "Combining" Information from Multiple Test Events
    - How to do Analysis with Continuous Data
      - Chemical Agent: Instead of the binomial response of whether a chemical agent is detected, use the time until detection.
      - Mine Detector: Range
    - Split Plots for Operational Realism

Naval Postgraduate School
Rachel Silvestrini

NPS' Efforts

- Mentored thesis student CDR Eric Lednicky
  - CDR Lednicky complete thesis (Sept 2011) entitled: Analytically Quantifying Gains in the Test and Evaluation Process Through Capabilities-Based Analysis
    - Illustrated the positive effect of incorporating DOE and M&S techniques throughout the entire T&E process
    - Quantitatively demonstrate the benefits of CBT&E over SBT&E
- Co-authored a paper with CDR Lednicky based on the thesis work
  - Paper accepted for publication: Lednicky, E. and Silvestrini, R.T. "Quantifying Gains Using the Capabilities-Based Test and Evaluation Method," 2011, Accepted for publication in *Quality and Reliability Engineering International*

Science of Test Research Consortium

NPS' Efforts

- Published a paper with Doug Montgomery, Jim Simpson, and Greg Hutto:
  - Johnson, R.T., Simpson, J.R., Montgomery, D.C., and Hutto, G.T. (2012) "Designed Experiments for the Defense Community," *Quality Engineering*, 24, pp. 60 – 79.
- Mentoring 1 PhD Students at NPS, Rebecca Black
  - Misspecification of Bayes priors in non-linear optimal design
  - This research is focused on studying the effect of misspecification of the prior estimates on parameters in optimal design generation for several different non-linear models including GLMs
- Co-mentoring 2 students with Doug Montgomery at ASU
  - Brian Stone and Edgar Hassler

NPS' Efforts

- Published a paper with Doug Montgomery, Jim Simpson, and Greg Hutto:
  - Johnson, R.T., Simpson, J.R., Montgomery, D.C., and Hutto, G.T. (2012) "Designed Experiments for the Defense Community," *Quality Engineering*, 24, pp. 60 – 79.
- Mentoring 1 PhD Students at NPS, Rebecca Black
  - Misspecification of Bayes priors in non-linear optimal design
  - This research is focused on studying the effect of misspecification of the prior estimates on parameters in optimal design generation for several different non-linear models including GLMs
- Co-mentoring 2 students with Doug Montgomery at ASU
  - Brian Stone and Edgar Hassler

AFIT Research
Raymond Hill

AFIT's Efforts

- Initial model of Ballistic Impact Flash Characterization to improve vulnerability and survivability analysis efforts
- Helped to define an statistically based experimental design process for the conduct of Live, Virtual and Constructive simulations
  - Haase, C. L., R. R. Hill, "An Algorithmic Foldover Procedure for Nearly Orthogonal Arrays with Projection" *International Journal of Experimental Design and Process Optimization*, Vol. 2, No. 3, 2011.

AFIT's Efforts

- Creating a model of Small UAV incident data to provide a predictive tool for the Air Force Research Lab, Eglin AFB
- Examining the variance influences on retinal photothermal damage threshold studies for the Human Performance Wing
  - An initial predictive capability for human risk assessment due to laser weapon testing ground effects
- Working with SIMAF on A2D2 project to define project goals, objectives and tests using statistical rigor

AFIT's Efforts

- Refined and delivered to ASC/EN through the 46th OL/ACS Ballistic Impact Flash models for entry and exit side flash
- Helped KC-46 program and 46th OL/ACS define recent live-fire test and conducting analysis of test results
- Examining Aerial Refueling Airplane Simulator Qualification (*ARASQ*) flight test program for KC-46
  - Developed new methods for examining flight test data
  - New methods to using flight test data within simulator development
  - Examining more efficient test matrices for use in program

Appendix B: Narrative Background from Initial SOT Research Consortium Proposal (AFIT Proposal 2011-55)

General Overview of Proposal

This proposal offers a multi-university, multi-disciplinary research consortium with the specific focus of improving test science and the underlying statistical theories associated with experimental data analysis to drastically improve the state-of-the-art in analytical capabilities as applied to the test and evaluation of DoD systems. This consortium of AFIT, NPS, ASU, and VaTech will conduct basic and applied research for test and evaluation, deliver multiple high-quality studies each year, advance the theories and methods needed to make test science a fundamental aspect of systems acquisition and test and evaluation, and help to improve the overall technical expertise of the DoD acquisition and test workforce.

The project involves two significant thrust areas: (1) Experimental Design and statistical theory, and (2) Live, Virtual, Constructive (LVC) Environments for test and evaluation.

Experimental Design

<u>Background</u>

Design of Experiments (DOE) is a systematic experimental planning process that ensures the collected data can be effectively analyzed using appropriate statistical models. Testing involves sequences of experiments. Effective testing requires effective experimentation. Effective experimentation ensures the experimental strategy matches the overall test objectives, maximizes the information obtained, minimizes the resources expended to achieve the test, and provides decision makers the information needed to make informed, timely decisions.

The DoD is incorporating DOE into its test enterprise; testing is a critical aspect of any Department of Defense (DoD) systems acquisition. However, there remains uncertainty about what this DOE adoption means and how the principles of DOE can be applied across the DoD test and evaluation enterprise. The goal of this research proposal is to answer these uncertainties through a program of research (theory and methods) and education (continuing and graduate). Achieving these goals will have tremendous impact on the DoD systems acquisition process in general and the test and evaluation aspect of that process in particular.

When a system is designed, developed and tested, before full-scale production and deployment, statistically designed experiments and advanced statistical analyses play a significant role in the following ways:

1. Identifying sources of variability impacting final system performance
2. Optimizing the controllable parameters of the system to obtain the desired performance;
3. Making the design robust to uncontrollable factors such as weather, climate, or

operational use;
4. Reducing total development lead time;
5. Reducing total development cost; and
6. Increasing overall knowledge of the system and its level of performance.

DoD experimentation covers all aspects of developmental and operational testing ranging from fully computer-based testing to systems-of-systems operational testing. The DoD test enterprise covers a wide range of domains:

- Air systems;
- Naval systems;
- Ground systems;
- Cyber systems; and
- Information technology systems.

These systems also range in the level of human involvement, from fully manned to unmanned, from fully human controlled to fully autonomous.

To test across such a broad range of systems, a full range of technologies and methodologies must be synergistically employed. These include:

- Principles of design of experiments;
- Use of live, virtual and constructive simulations;
- Principles of systems engineering to include building early test influence;
- Principles of reliability, maintainability and usability; and
- Integrated developmental and operational planning coupled with full requirements traceability.

Our research consortium will involve DoD-wide concerns and leading experts to undertake the complex task of devising the theories and developing the methodologies whereby the DoD test enterprise can fully embrace the systems engineering of test.

The following represent some of the broad areas of interest that will become focused areas of research, based on DoD needs, sponsor desires, and consortium capabilities.

1. **Systems engineering of test**: How do DOE principles get applied from a systems acquisition point of view? Rather than designing specific tests using DOE, how can we design a test program using the principles of DOE? Such an effort will address developmental and operational testing incorporating considerations such as system requirements, test resources and system risk.
2. **Systems-of-systems testing and the role of Live, Virtual and Constructive (LVC) simulations in this testing:** LVC has long been used in the training and mission rehearsal arenas. The question now is how this technology might assist the systems acquisition and testing community? Answering this question requires examining issues associated with LVC verification, validation and accreditation as well as how to incorporate such LVC testing within the current, live-testing constructs.
3. **Computer simulation experimentation:** To offset the growing cost of live test, use of detailed computer simulations will continue to grow in importance. This component of the research will address the use of Monte-Carlo simulations (e.g., campaign warfare models) and physics-based simulations (e.g., 6 degree-of-freedom kinetic/ballistics, computational fluid dynamics (CFD)) and how the data from such models can be used for immediate and future system decisions.

4. **Specific design issues:** The field of DOE is not complete; there are myriad issues that must be solved. For instance, how to conduct experiments in face of operational constraints, how to effectively conduct testing in a sequential fashion, how to build and examine designs containing quantitative and non-quantitative variables, or how to conduct experiments when the focus of the test is on "the edge of the operational envelope" such as is often the case with DoD systems. These issues, and others, will require solutions (theory and methods) as the DoD enterprise continues to incorporate the principles of DOE.

Live, Virtual, Constructive Test Environments (AFIT Lead)

Background

Live virtual and constructive (LVC) simulation environments provide powerful means to improve the test and evaluation (T&E) of DoD weapon systems. LVC environments have already made their mark in the mission rehearsal and training domains; these domains do not require much in the way of statistical experimental design. The capability to realize large scale operations involving a wide variety of systems (diversity) all interacting within some operational scenario of interest means the LVC environments provide arguably the only way to feasibly test modern weapon systems.

Problem Statement

LVC experiments have been used mostly for training and largely qualitative assessments of capabilities. To support acquisition system decision-making, as is the case in T&E, increased statistical rigor is required in planning and conduct of LVC experimentation. Principals of experimental design and test execution must be examined, applied and adopted for LVC experimentation to support T&E.

Focused Research Objectives

Investigate, adapt and support the application of statistical experimental design principals to LVC experimentation. This will include methodologies for:

- defining LVC test objectives
- defining LVC test matrixes
- defining LVC test schedules
- instrumentation requirements for LVC data collection
- experimental designs based on computer experimentation and small sample size requirements
- analysis of LVC test results
- statistical validity of LVC test results

Advanced Experimental Designs for Simulation and Live Experimentation (NPS Lead)

Background

It is widely recognized across science and engineering disciplines that effective testing requires effective experimentation strategies. The analytical team must ensure the experimental strategy matches the overall test objectives, maximizes the information obtained, minimizes the resources expended, and provides decision makers with the information needed to make informed, timely decisions. Traditionally, scientists equate experimental strategies with “physical testing,” but methods for experimenting on a simulation are equally important. If the simulation is being used to support or influence decisions made in the real world, then improvements made in obtaining the results directly translates to improvements for the real system in question.

Information from simulation experiments and live experiments can be joined together to improve the T&E process and greatly reduce the acquisition life cycle of a project. This requires the study of statistical techniques that are aimed at robust experimental design strategies as well as efficient and statistically sound methods for combining data across multiple sources.

Problem Statement

Experimental design methods for testing on computer simulation model are fundamentally different than methods for experimenting in a live physical situation. More research effort should be placed on sequential experimental design, experimental design for non-linear models, experimental design for use in the combined simulation/live studies, and statistical methods for estimating distribution with very few data points. Doing this will improve the current T&E efforts and the capability of the experimental design and analysis methods currently in place.

Focused Research Objectives

Develop and improve techniques for experimenting and subsequently analyzing the data across the simulation and physical domain with respect to T&E efforts. Research focused in the areas of:

- sequential experimentation
- experiments for non-linear or generalized linear models
- combining results from simulated and live experiments
- evaluation of robust parameter settings based on simulated and live experiments
- estimating statistical distributions when data is scant

Experiments with Physical and Resource Constraints (ASU Lead)

Background

Many experiments in the DoD environment involve either physical constraints on what can be tested (such as specific combinations of factors) or how the tests can be performed (such as inability to completely randomize all of the runs), or limitations on time, funding and available of test articles that effectively reduce sample sizes. It is critical to design experiments in these situations that take these constraints into account and yet produce the quality of information that good decision-making requires. Design and analysis tools for these types of problems are only in their infancy. We will develop methods to account for constraints such as randomization restrictions, sequential experimentation, incorporating prior information from previous tests or development data, and optimal allocation of resources within a complex test environment.

Problem Statement

Many experiments in the DoD environment involve hard-to-change factors that lead to restrictions on complete randomization and simultaneously involve nuisance variables (such as weather effects) and constraints on the operating space. These experiments can also involve nested factors, random effects and covariates (factors that cannot be precisely controlled for the experiment but can be measured). Many experiments involve response or outcome variables that are heavy-tailed (non-normal) or are binary (such as success probabilities). Software is a large and growing component of many systems. Software testing and testing systems with embedded software is an area of experimental design that has had little formal study and few methods of general applicability are available. Investigation of these problems and developing valid approaches and solutions will greatly improve the T&E effort.

Focused Research Objectives

Some typical research objectives to be accomplished include:

- Extensions of split plot designs and variations. Split plot designs arise when one or more factors are hard to change, restricting complete randomization and introducing a model term that acts as both a classic blocking factor, as well as a model factor with both main effects and interactions. The extension of design methods for split-plot experiments to incorporate nested factors, unbalanced test structures, covariates and noise factors will provide new methodology for T&E.
- Complex designs power analysis and sample size determination.
- Designs for problems with non-normal responses where the generalized linear model is required to obtain adequate models for the response. This will include factorials, response surface type experiments and complex factor structures such as split-plots.
- Incorporation of statistics of extremes into the analysis of experiments so that tail probability performance of weapon systems and sensors can be accurately predicted.
- Design and analysis for software and systems with embedded software. Development of guidelines, principles and applications for experimentation on these problems.

Experimental Designs to Produce Reliable and Robust Weapon Systems (Va Tech Lead)
Background

Warriors in the field must have weapon systems known to be highly reliable and robust to uncontrollable factors such as the environmental conditions at use. Reliability in this context has two important meanings. The warrior expects the weapon system to perform as expected each time it is used (short term). Reliability also means that the weapon system continues to perform consistently over long periods of time. Robustness in this context means that the proper weapon system functions well over all the expected conditions for the system's use. Well-developed experimental design strategies can greatly improve all of these crucial aspects.

Many good rigorous experimental design and analysis strategies exist for improving product and process robustness. However, such approaches need adaption for the T&E of DoD weapon systems. On the other hand, rigorous experimental design and analysis strategies for reliability exist only for overly simplistic and naïve situations. For example, there are no current strategies that rigorously handle blocking, subsampling, or split-plot situations. No current methodologies exist for system robustness involving reliability data. As a result few, if any, of the current rigorous methodologies for reliability data are truly appropriate for the T&E of a weapon system.

Problem Statement

Warriors in the field require reliable and robust weapon systems. The experimentation required to ensure such systems requires fundamental research to adapt current methodologies and to develop new, more appropriate experimental designs and analyses. The new and adapted methodologies will enable the T&E of DoD weapon systems to provide more effective tools to the field.

Focused Research Objectives

Develop and improve experimental designs and analyses suitable for the T&E of DoD weapon systems reliability and robustness in the following ways:

- review current experimental strategies for the analysis of reliability data and determine their suitability
- develop new experimental designs and analyses for reliability data
- develop appropriate sequential strategies for such experiments with reliability data
- review and adapt current robust experimentation strategies
- develop robust experimentation strategies with reliability data

Appendix C: Research Papers Developed In Support of Consortium Research

Journal Papers

Storm, S., R. R. Hill, and J. J. Pignatiello. May 2012. A Response Surface Methodology for Modeling Time Series Response Data. Accepted for publication in Quality and Reliability Engineering International.

Chambal, S P., J. Kitchen, R. R. Hill and A. J. Gutman, September 2011, Acquisition and Testing, DT/OT Testing: The Need for Two-Parameter Requirements. Accepted for publication in *Quality and Reliability Engineering International*.

Haase, C. L., R. R. Hill, and D. Hodson. September 2011. Using Statistical Experimental Design to Realize LVC Potential in T&E, *International Test and Evaluation Journal*, Vol. 32, No. 3, 288-297.

Haase, C. L., and R. R. Hill. 2011. An Algorithmic Foldover Procedure for Nearly Orthogonal Arrays with Projection, *Journal of Experimental Design and Process Optimization*, Vol. 2, No. 3, 191-201.

Hill, R. R., D. A. Leggio, S. R. Capehart, A. G. Roesener. October 2011. Examining Improved Experimental Designs for Wind Tunnel Testing using Monte Carlo Sampling Methods. *Quality and Reliability Engineering International*, Vol. 27, Issue 6, 795-803.

Haase, C. L., R. R. Hill, and D. Hodson. February 2011. *Planning for LVC Experiments*, submitted to Systems Engineering.

Wolf, S. E., R. R. Hill, and J. J. Pignatiello. June 2012. *Using Neural Networks and Logistic Regression to Model Small Unmanned Aerial System Accident Results for Risk Mitigation,* submitted to Military Operations Research.

Wolf, S. E., R. R. Hill and J. J. Pignatiello. July 2012. *Analysis of an Intervention for Small Unmanned Aerial System (SUAS) Accidents,* submitted to Quality Engineering, LQEN-2012-0056.

Stone, B., D. C. Montgomery, E. Hassler and R. T. Silvestrini. 2012. *An Expected Cost Method for Selecting Screening Experiments.* Submitted to Quality Engineering.

Silvestrini, R. T., W. Parker and G. Sammito. 2012. *Design of Experiments for Information Technology Systems: What Program Managers Should Know about the Plan and Design Phases.* To be published in Defense AT&L Magazine.

Silvestrini, R. T. and E. Lednicky. 2012. *Quantifying Gains Using the Capabilities-Based Test and Evaluation Method.* Accepted for publication in Quality and Reliability Engineering International.

Stone, B., D. C. Montgomery, E. Hassler and R. T. Silvestrini. 2012. *An Expected Cost Methodology for Screening Design Selection.* Submitted to Quality Engineering, LQEN-2012-0059

Freeman, L.A., Ryan, A.G., Kensler, J.K, Dickinson, R.M., and Vining, G.G. (2012) "A Tutorial on the Planning of Expeirments," to be submitted to Quality Engineering.

Freeman, L.A., Ryan, A.G., Kensler, J.K, Dickinson, R.M., and Vining, G.G. (2013) "Planning Experiments for Highly Constrained Regions," to be submitted to Quality Engineering.

Kensler, J.K, Freeman, L.A., and Vining, G.G. (2012) "Analysis of Reliability Experiments with Random Blocks and Subsampling," to be resubmitted to the Journal of Quality Technology.

Kensler, J.K, Freeman, L.A., and Vining, G.G. (2012) "A Practitioner's Approach to the Analysis of Reliability Experiments witth Random Blocks and Subsampling," to be submitted to Quality Engineering.

Conference Papers

Wolf, S, J. J. Pignatiello, R. R. Hill. May 2012. A Magnitude Robust Control Chart for Monitoring Process Disperson. Proceedings of the 2012 Industrial and Systems Engineering Research, Orlando, Fl.

Wooddell, D. A., C. M. Schubert-Kabban, R. R. Hill. January 2012. An Analysis of the Influences of Biological Variance, Measurement Error, and Uncertainty on Retinal Photothermal Damage Threshold Studies. Proceedings of International Society for Optics and Photonics, 2012 Photonics West Conference, San Francisco, CA.

Talafuse, Thomas, Hill, R. R. and Bestard, J. April 2011. Characterization of Ballistic Impact Flashes Empirical Model Development. Proceedings of the the 52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Denver CO, April 4-7, 2011.

Hill, R. R., T. Talafuse, and J. Bestard. 2011. Characterization of Ballistic Impact Flashes – Developing Empirical Models for Survivability Analyses, Proceedings of the 2011 Industrial Engineering Research

Haase, C. L. and R. R. Hill. 2011. Applying Experimental Design to Live, Virtual, and Constructive (LVC) Experiments. Proceedings of the 2011 Industrial Engineering Research Conference, Reno NV, May 21-25, 2011.

Appendix D: Masters Theses and Dissertations Supported

Theses
(completed)

1. Haase, C. March 2011. Tailoring the Experimental Design Process for LVC Experiments.
2. Hosket, J. March 2011. A Methodology using Simulation Results for Test and Evaluation.
3. Talafuse, R. March 2011. Empirical Characterization of Ballistic Impact Flash.
4. Peyton, D. March 2012. Ballistic Flash Characterization of Entry Side Flash.
5. Wooddell, D. A. March 2012. Probabilistic Model for Laser Damage to the Human Retina.
6. Wolf, S. E. March 2012. Modeling Small Unmanned Aerial System Mishaps using Logistic Regression and Artificial Neural Networks.
7. Storm, S. M. March 2012. Evaluating Aerial Refueling Simulator Validation Test Designs y Extending Response Surface Methodology to Analyze Time History Responses.
8. Koslow, M. J. March 2012. Ballistic Flash Characterization Penetration and Back-Face Flash.
9. Chamberlain, C. March 2012. Analysis of KC-46 Live-Fire Risk Mitigation Program Testing.
10. Lednicky, E. September 2011. Analytically Quantifying Gains in the Test and Evaluation Process Through Capabilities-Based Analysis.
11. Cappelline, R. September 2011. Error Propagation Through Hierarchical Combat Models.

Dissertations
(completed and in-progress)

1. Kensler, J. July 2012. Analysis of Reliability Experiments with Random Blocks and Resampling. Virginia Tech.
2. Gutman, A. In Progress. Construction, Analysis, and Data-Driven Augmentation of Supersaturated Designs. Air Force Institute of Technology.
3. Stone, B. In Progress. Screening Design Selection, Analysis and Construction. Arizona State University.
4. Hassler, E. In Progress.
5. Dougherty, S. In Progress. Second-Order Response Surface Screening Designs. Air Force Institute of Technology.

Appendix E: Professional Presentations

Douglas C. Montgomery, “Experiments with Physical and Resource Constraints”, Invited presentation at the 2012 Joint Statistical Meetings, 28 July – 2 August, San Diego, CA.

Jennifer Kensler, “Reliability Experiments with Random Blocks and Subsampling,” Quality and Productivity Research Conference, Long Beach, CA, June, 2012.

Jennifer Kensler, “Reliability Experiments with Random Blocks and Subsampling,” Joint Statistical Meetings, San Diego, CA, July, 2012.

Geoff Vining, “Issues in Planning Experiments for Highly Constrained Regions,” 12th Annual Conference of the European Network for Business and Industrial Statistics, Ljubljana, Slovenia, September, 2012.

Raymond R. Hill, “Statistical Engineering Applied to Air Force Problems” 2012 Joint Statistical Meetings, 28 July – 2 August, San Diego, CA.

Raymond R. Hill, “An Empirical Model Development and Data Validation Effort for Missile Fragment Flash Characterization” Quality and Productivity Research Conference, Long Beach, CA, June, 2012.

Raymond R. Hill, J. J. Pignatiello, Jr. and S. Storm, “Advancing Statistical Methods for Examining Flight Test Performance Data” Defense Analysis Seminar XVI, Seoul, South Korea, 23-25 April 2012.

Douglas Hodson, Raymond Hill and Alex Gutman, “Using LVC Simulations for Systems Analysis – Experimental and Software Design Issues” Defense Analysis Seminar XVI, Seoul, South Korea, 23-25 April 2012.

Raymond R. Hill, D. P. Peyton, M. Koslow and C. Chamberlain “Developing and Validating an Empirical Model of Missile Fragment Flash Characterization”, 80th Military Operations Research Society Symposium, United States Air Force Academy, Colorado Springs, CO, June 2012.

Douglas Hodson, Raymond Hill and Alex Gutman “Using LVC Simulations for Systems Analysis – Experimental and Software Design Issues”, 80th Military Operations Research Society Symposium, United States Air Force Academy, Colorado Springs, CO, June 2012.

Brian Stone and Douglas C. Montgomery “An Expected Cost Methodology for Screening Design Selection,” *Quality and Productivity Research Conference (QPRC)*. Long Beach, CA, 6 June 2012.

Brian Stone and Douglas C. Montgomery "An Expected Cost Methodology for Screening Design Selection," *Military Operations Research Society Symposium (MORSS)*. Colorado Springs, CO, 14 June 2012.

Brian Stone, "An Expected Cost Methodology for Screening Design Selection," *Operations Analyst (OA) Forum*. 20 August 2012 (webinar)

Bibliography

Coleman, D. E. and D. C. Montgomery. 1993. A Systematic Approach to Planning for a Designed Industrial Experiment. Technometrics, 35(1): 1-12.

Freeman, L. J. and G. G. Vining. 2010. Reliability Data Analysis for Life Test Experiments with Subsampling. Journal of Quality Technology, 42(3): 233-241.

Freeman, L. J. and G. G. Vining. 2012. "Reliability Data Analysis for Life Test Designed Experiments with Sub-Sampling," Quality and Reliability Engineering International, to appear. DOI: 10.1002/qre.1398.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
02-10-2012	**Research Report**	Mar 2011 – Oct 2012

4. TITLE AND SUBTITLE
Science of Test Research Consortium: Year Two Final Report

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Raymond R. Hill, Douglas C. Montgomery, G. Geoffrey Vining, Rachel T. Silvestrini

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Street, Building 641
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/EN/TR-13-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

OSD DOT&E -	Test Resource Management Center (TRMC)
Dr. Catherine Warner	Mr. George Rumford
1700 Defense Pentagon	1225 South Clark Street
Washington DC 20301	Arlington VA 22202

10. SPONSOR/MONITOR'S ACRONYM(S)

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

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The Science of Test (SOT) Research Consortium is sponsored and funded by both Office of Secretary of Defense (OSD) Director, Operational Test and Evaluation (DOT&E) and by the Test Resource Management Center (TRMC) within the OSD Director, Developmental Test and Evaluation. The consortium members include the Air Force Institute of Technology, Department of Operational Sciences, Arizona State University, School of Computing, Informatics, and Design Systems Engineering, Virginia Tech, Statistics Department and Naval Postgraduate School, Operations Research Department. The SOT research effort commenced in early CY 2011. This report summarizes the accomplishments through the second year of the research consortium. The consortium collectively contributes to the theory of the statistical rigor of test with contributions to the body of knowledge and to the practice of test with applied methods and consultations.

15. SUBJECT TERMS
Science of Test, Statistical Rigor in Test, Test and Evaluation, Design of Experiments

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Raymond R. Hill, USAF (ENS)
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	32	(937) 255-6565, ext 7469; e-mail: Raymond.Hill@afit.edu