2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

A History of Experimental Design (Keynote)

J. Stuart Hunter
Princeton University

A brief review of the many personalities and episodes that established the arts of experimental design.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

New Directions for Test & Evaluation: Design of Experiments in Defense Acquisition

Catherine Warner
Office of the Secretary of Defense, DOT&E

Design of Experiments (DOE) is a formal process and a scientific tool for conducting rigorous and comprehensive tests. In November 2009, Dr. Gilmore, Director, Operational Test & Evaluation (DOT&E), in a memorandum to his staff, outlined four initiatives for DOT&E:
(1) Field new capabilities rapidly,
(2) Early engagement of testers in requirements determination,
(3) Implement integrated testing, and
(4) Improve suitability.
Previously, the Operational Test Agencies and DOT&E entered into a Memorandum of Understanding that endorsed DOE as a discipline to improve integrated testing. This talk provides insights as to how DOE can be implemented in test & evaluation as a best practice for achieving the initiatives of DOT&E. Several past case studies are examined retrospectively to investigate the impact of including DOE methodologies into the test planning process. The talk concludes with a discussion on the benefits of including statistical rigor, including but not limited to DOE, into the test and evaluation process.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

A Framework for Using Design of Experiments Across and Within Test Events for DOD Systems Acquisition Testing

(Special Session on Design of Experiments for Integrated Testing)

Nancy Dunn
US Army Evaluation Center

What creates a Robust and Rigorous test methodology across test events and within test event? To some, robustness and rigor mean the same thing and they both imply more realistic scenarios and threat play; to others it means the application of formal Design of Experiments (DoE) to determine the experiment trials with more control over test trial execution. Test robustness and rigor includes all of these and more. This presentation proposes a comprehensive framework for understanding the components of test robustness and test rigor, their interrelationships, and techniques for increasing both.

A robust test examines all of the important factors and conditions that could impact the system under test (SUT) across the full operational environment that the system will be expected to operate. DoE techniques are focused on designing one test event at a time to maximize the number of factors and levels that can be examine with the most economical set of test runs to support a statistical model that characterizes the impacts on system performance. I will present one method of showing the robustness of the testing across all test events.

Test rigor, on the other hand, characterizes the credibility of the test to produce evidence to support the formal DoE analysis model. A comprehensive view of test rigor can best be explained by four (4) basic test validity requirements. This basic rigor framework is then used to organize test-design best practices as “fixes" to these validity threats. Organizations that conduct tests can employ this framework to organize and apply their own lessons-learned and best-practices to design rigorous tests.

This presentation emphasizes traditional DoE textbook methods which are critical to determining test robustness and that the lesson-learned by experienced testers in overcoming all the things that can go wrong in test execution that are critical to test rigor. Both robustness and rigor are necessary for credible Department of Defense Test and Evaluation.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Integrated Test via a Grid of Well-Designed Experiments
(Special Session on Design of Experiments for Integrated Testing)

Jim Simpson
53rd Wing, Eglin Air Force Base

The discipline of experimental design took root in agriculture early in the 20th century and has since gained widespread acceptance across various aspects of most other communities including the industrial, government, social, and business sectors. The Department of Defense has recently endorsed the application of the design of experiments (DOE) as standard practice in operational test and evaluation, and is keenly interested in the potential for DOE to enhance the integration of developmental and operational (both initial and follow on) test. Textbooks on experimental design and analysis (e.g. Box, Hunter, and Hunter, and Montgomery) primarily address the single objective, single system variant scenario. Sequential testing in this context refers to a series of experiments over a short time horizon for the purpose of leveraging system knowledge to ultimately test efficiently and effectively. So a relevant question is how should DOE be applied for complex systems undergoing several stages of development then operational test? Obviously, a number of scoped experimental designs should be developed, so the focus of this paper is to provide some of the keys for piecing these experiments together into a cohesive, integrated plan for weapon system test and evaluation. The mosaic of experiments includes first the vertical dimension - groups of experiments across capabilities and requirements at a given level of system maturity. This series of experiments are linked tying capabilities/requirements to specific test objectives. The second, horizontal dimension of the DOE grid requires testing across the lifecycle of development to employment. Here the sequence of plan-design-execute-model for one variant must feed seamlessly into the next series of experiments leveraging the insight from the earlier stage while adapting for the evolution of the system. Examples will be provided illustrating these ideas.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Evolutionary Test Design for Enhanced Binary Response

(Special Session on Design of Experiments for Integrated Testing)

Donald P. Gaver and Patricia A. Jacobs
Naval Postgraduate School

A large number of military items are judged on binary (Detect/Not-Detect, Hit/Miss, Operate/Fail, Kill/Not-Kill,…) metrics. Early Developmental Tests and later Operational Tests often use binary success counts to evaluate effectiveness and suitability of military platforms. However, simple success-ratio summaries of performance may be less cost-effective than records of actual happenings. Miss distances of projectiles, e.g. air-launched small diameter bombs, should be accounted for and analyzed as well as simply scored Hit, or Miss. System reliability and even human survivorship data should be analyzed to suggest useful warnings of impending failure, or vulnerability.

This paper revisits classical "design of experiments" (DOE) lore, in particular that pertaining to quantitative understanding of sequential Up and Down methods of selecting factor types and levels. It seems clear that examination of concomitant measures, additional to primitive Hit/Miss, Yes/No,…, will in many cases add to test insights and considerably reduce costs throughout the acquisition process.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Network Reliability: A Fresh Look at Some Basic Questions

Francisco Samaniego
University of California, Davis

Network centric operations are becoming increasingly dominant in U.S. Army doctrine and practice in current communications applications. The Proceedings of the 2009 Conference of the ARL Collaborative Technology Alliance in Communications and Networks features dozens of research advances in the design and performance of communication networks, with connectivity, speed, cost and security issues carefully considered, individually and in tandem. Various forms of approximate optimality are examined. In this paper, we address fundamental questions about network topologies. We examine the utility of analytical tools such as "dominations" and "signatures" in assessing the reliability of a given network and in comparing two competing network designs. The traditional approach to treating the question of whether there exists a uniformly optimal network design of a given size (i.e., with v vertices and n edges for fixed v and n) is examined, and certain weaknesses in the approach are noted. An alternative approach is proposed and is shown to provide rather striking results in a context in which the traditional approach had been shown to fail. A scenario which treats the tension between performance and cost is also considered, and the optimality of a network under a "performance per unit cost" criterion is studied. An agenda for further research in this area is outlined.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Thinking Outside the Box: Toward an Enhanced Role for Statistical Modeling in Army Decision Processes

James R. Thompson
Rice University

Statisticians are most effective when they are brought into the picture at the exploratory phase of a project and continue working on the project during its construction and after its completion to provide analysis of the effectiveness of the enterprise and modifications, which might be implemented on the “finished” product. Unfortunately, in many cases statisticians are not an integral part of the process. Some examples are given where the non–involvement of statisticians has led to flawed weapons and flawed policies.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Educating Army Leaders in the Use of Statistical Methods to Enable Counter IED Operations

(Special Session on Case Studies in Army Applications of Statistics)

LTC Paul Kucik
United States Military Academy

Over the last several years, improvised explosive devices (IEDs) have been the insurgents' weapon of choice against US and coalition forces. In order to counter this threat, the Counter-IED Operations Integration Center (COIC) fuses multi-source intelligence, providing quality information and analytical products to decision makers to enable them to make data-based decisions. The CIOC continually generates highly sophisticated statistical analysis products, which may be difficult for unit leaders to fully understand, given the leaders' urgent need to deal with tangible life-threatening situations on a daily basis. The CIOC has begun an initiative to educate unit level commanders on statistical methods employed in counter IED operations in order to improve their ability to integrate analysis into their planning and execution timelines and to better understand the uncertainty involved in statistical estimates. This presentation will address the contributions to this effort by the Operations Research Center within the Department of Systems Engineering at West Point.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Exact Customer Sojourn Time Distributions for the M/M/s Queue
(Special Session on Case Studies in Army Applications of Statistics)

LTC William Kaczynski
United States Military Academy

The exact distribution of the nth customer's sojourn time in an M/M/s queue with k customers initially present is derived. Algorithms for computing the covariance between sojourn times for an M/M/1 queue with k customers present at time zero are also developed. Maple computer code is developed for practical application of transient queue analysis for many system measures of performance without regard to traffic intensity (i.e., the system may be unstable with traffic intensity greater than one). Applications are provided regarding military air traffic control and military personnel processing stations.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Modeling Officer Manpower

(Special Session on Case Studies in Army Applications of Statistics)

LTC Tina Hartley, LTC Scott Nestler, and LTC Andrew Hall
United States Military Academy

Current officer modeling efforts have aggregated at level of Competitive Category. Modeling at the Competitive Category Level of aggregation masks data available at lower levels that could potentially increase modeling accuracy. Our examination of the availability of and predictive power of additional data illuminates potential shortfalls within the existing modeling environment. In particular, we have developed refined models for predicting losses at the specialty level in the near-term (2 years), and then allow the model to dictate losses further out in time; this is a change from only using aggregated historical rates. We have modeled time in grade, as opposed to commissioned years of service, in order to provide a more powerful forecasting model. This, in conjunction with exploiting Markovian structures within the officer manpower system, allows us to reduce the model state space dimensionality and permits reasonable model solution times. All of these changes should lead to increased modeling and forecasting fidelity and will allow exploration of an ever-enlarging solution space for alternative officer policies.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

The Percolated Beta-Binomial Distribution

Nozer D. Singpurwalla
The George Washington University

In this talk I will introduce a scenario that arises in several diverse contexts pertaining to the activities of the Department of Defense, business, industry, and social life. The scenario entails a large group of entities that need to pass through a series of hierarchically placed sieves, and interest centers around the probability of successfully transitioning through all the sieves. Assessing this probability results in what I think is a new distribution that I call a percolated beta-binomial distribution. To motivate my work I consider some data obtained under the freedom of information act pertaining to the selection of astronauts for the US Space program. The model is general and my hope is that it may have applications to reliability problems as well.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Complex Failures in Fiber/Epoxy Composites under Tensile Load with an Application of Partition based Dirichlet Priors

Jayaram Sethuraman
Florida State University

Around 1960, B. W. Rosen ran an experiment on a segment of fiber/epoxy composite by slowly increasing the tensile load until complete failure. The data shows cascading failures of among elements of the fibers as the load increases. We postulate a Phase I/II failure model based on a load distributing nearest neighbor rule. The data on the strengths of the elements are now interval censored in some interesting ways. Partition based Dirichlet priors introduced in Sethuraman and Hollander (2009, JSPI) are well suited to analyze such data. We shall show some results of such a Bayesian analysis. Research is sponsored by the Air Force Research Laboratory (AFRL).

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Spectral Approximation of Infinite-Dimensional Black-Scholes Equation with Memory

Mou-Hsiung (Harry) Chang
US Army Research Office

This paper considers the pricing of a European option using a (*B, S)*-market in which the stock price and the asset in the riskless bank account both have hereditary price structures described a system of stochastic functional differential equations. Under the smoothness assumption of the payoff function, it is shown that the infinite dimensional Black-Scholes equation possesses a unique classical solution. A spectral approximation scheme is developed using the Fourier series expansion in the space C[−*h*, 0] for the Black-Scholes equation. It is also shown that the *n*th approximant resembles the classical Black-Scholes equation in finite dimensions.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

3 Systems - 3 DOEs (UFHD / LVSR / CTN)

Swala Burns
Marine Corps Operational Testing and Evaluation Activity

MCOTEA is the independent operational test agency for the Marine Corps in charge of conducting independent operational test and evaluation on systems in the acquisition cycle that the Marine Corps may procure. MCOTEA recently completed three test plans using a Design of Experiments (DOE) approach to support the operational testing of three very unique classes of systems. The three classes of systems include a network command and control device, a logistics vehicle, and a small arms marksmanship measurement device. This presentation will illustrate the test designs chosen, the rationale for the selection of the design, and noted strengths/weakness’ in the designs that could be improved upon in future testing.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Missile Fuze Experiment

William C. Thomas and Chris Schott
Raytheon Missile Systems

A 3 factor mixed level full factorial experiment with 8 replicates was completed on a missile fuze. The objective of the experiment was to identify changes to the fuze design that would cause the fuze to reliably arm at a given time over a range of acceleration levels. The three factors were lever thickness, spring rate, and acceleration. The response was arm time. The three main effects, the two way interactions, the three way interaction and the two quadratic effects were analyzed for the average and the standard deviation. Interactions with the supplier will also be discussed. The resulting model gave the engineer flexibility in what changes to make that would minimize the cost and the impact to the program and the supplier. The fuze design was optimized using the model by selecting a new lever thickness and new spring rate.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Random Contamination of Semiconductor Materials

Bernard Harris
Universities of Wisconsin and Nebraska

In the manufacture of semiconductor materials, undesired contaminating particles may appear. If there are "too many" and/or the contaminants are "too close together", the material is defective. Probability models for this type of contamination are developed and appropriate asymptotic distributions are obtained.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Modeling with "Infinite" Bootstrapping

COL Andrew Glen, MAJ Chris Marks, and CDT Matt Robinson
United States Military Academy

We present some probabilistic modeling techniques for representing stochastic systems where only component failure times are known. The methods produce distribution functions of discrete random variables that model the entire systems' lifetime. The discrete random variables are found with analytic methods that replicate the outcome of a similarly designed bootstrap model that was able to resample an infinite number of times. Analytic results are presented about bias and convergence of the discrete model, compared to the unknown continuous model.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Multiple Testing Using Sequential Permutation P-Values

Dan Nettleton
Iowa State University

Permutation techniques provide reliable tests of significance when the condition of exchangeability holds under the null hypothesis. Such tests are particularly useful when parametric model assumptions are (or may be) substantially discordant with reality. Before the development of modern computing power, permutation tests were often avoided because of computational expense. Now, however, conducting a single permutation test is typically a trivial computational exercise. If the total number permutations is prohibitively large, it is well known that a valid test can be based on a simple random sample from the permutation distribution. In many applications, it is easy and affordable to sample many values from the permutation distribution to obtain reliable permutation-based inference. However, there are some situations where permutation testing can become computationally expensive. In many modern applications, it is necessary to simultaneously test thousands of null hypotheses, and it may be the case that testing each null hypothesis requires the computation of a nontrivial test statistic. In such situations, it is useful to develop efficient algorithms for sampling from the permutation distribution of each test statistic. In this talk, we will consider the problem of testing each of m null hypotheses with a sequential permutation procedure in which the number of draws from the permutation distribution of each test statistic is a random variable. Each sequential permutation p-value has a null distribution that is nonuniform on a discrete support. We show how to use a collection of such p-values to estimate the number of true null hypotheses m_0 among the m null hypotheses tested and how to estimate the false discovery rate (FDR) associated with p-value significance thresholds. The properties of our estimates of m_0 and FDR are evaluated through simulation and illustrated through the analysis of a microarray dataset.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Data Assimilation and Uncertainty Quantification for Complex Computer Models

Brian Williams and David Higdon
Los Alamos National Laboratory

This presentation provides an overview of statistical model calibration for complex computer experiments. Calibration involves solving a probabilistic inverse problem to constrain uncertain physics/engineering inputs so that the resulting model realizations are consistent with observed experimental data. A prior probability distribution for these model parameters is updated using experimental information at the separate and integral effects levels simultaneously. Incorporation of model form uncertainty in rigorous uncertainty quantification analyses is also discussed.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Understanding Afghanistan
(Special Session on Social and Cultural Terrain)

Edward J. Wegman
Center for Computational Data Sciences, George Mason University

The early history of Afghanistan is written in terms of rival tribal leaders of scattered and ethnically diverse populations. The country itself is a barren territory which does not produce enough revenue for a central government to rule it. Thus, a regional tribal form of government naturally developed in order to eke out a subsistence living for the members of the tribes. In the eighteenth century, as neighboring empires collapsed, Afghan tribal leaders seized opportunities to build states by conquering richer areas in the region. Modern Afghanistan was created as a buffer state by the British between colonial India and imperial Russia. Afghanistan has only been organized with a central government by virtue of intervention of outside powers; first Britain, later the Soviet Union, and, most recently, Pakistan, who created the Taliban. Without external interventions, the country tends to revert to tribal/patrilineal regional rulers. This talk will discuss the role of the history, the Pashtuns, and the culture of Pashtunwali in shaping modern Afghanistan. This talk is not a statistics talk, but rather a discussion of the social and cultural context of modern Afghanistan. It is intended to provide the logic underlying the need for a quantitative assessment of social and cultural issues associated with Afghanistan and similar tribal-oriented societies.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Data Acquisition and Data Employment Protocols for Multi-agent Modeling of Conflict
(Special Session on Social and Cultural Terrain)

Armando Geller
Center for Social Complexity, George Mason University

Computational techniques in general and multiagent modeling in particular have gained increasing popularity as analytic approaches in the social sciences. However, a majority of multiagent models of adversarial social interactions, be it organized crime, irregular warfare, stability and support operations and intrastate conflict, remain empirically underspecified, mostly because available data are rarely collected to meet the requirements of multiagent modeling. Therefore, it is imperative to formulate concepts and methods that address data collection, organization and dissemination challenges unique to multiagent modeling. During the talk I will outline the current understanding of multi-agent modeling as an empirical methodology and later compare the state-of-the-art in multiagent modeling with ways of employing data within other modeling traditions. Particular emphasis will be given to qualitative data sources, including narratives and case studies in applications to designing cognition of agents and validating their behaviors.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Inferring Social Network Structure from Incident Size Distribution in Iraq
(Special Session on Social and Cultural Terrain)

Tim Gulden
Center for Social Complexity, George Mason University

The violence in Iraq between 2003 and the present has been multi-faceted and extremely hard to characterize in terms of its motivations, participants and even overall scale. This work identifies a remarkably stable truncated power-law pattern in the size distribution of violent incidents in the Iraq Body Count (IBC) database and seeks to explain it in terms of social networks and the US role in breaking up and repressing major violent groups. While the work is preliminary, it offers a potentially useful metric in assessing the progress of counterinsurgency operations.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Large-scale Experimental Designs for Simulation

Susan M. Sanchez, Thomas W. Lucas, Paul J. Sanchez, and Christopher J. Nannini
Naval Postgraduate School

The department of defense (DoD) regularly makes decisions about how to equip, organize, train, employ, and maintain military forces. Unfortunately, physical experimentation to support these decisions is almost always severely limited due to time, money, and safety considerations. Consequently, DoD relies heavily on models and simulations to support and justify many decisions. These models are often investigated in a haphazard manner. Historically, even if experimental designs have been used, they have typically been applied to only a handful of factors---even though the simulation models may have hundreds or thousands of inputs. This suggests that more modelers and analysts need to be aware of the power of experimental design, especially the recent breakthroughs in large-scale experimental designs that enable us to understand the impact of many factors and their intricate interactions on model outcomes. In this presentation, we review a portfolio of designs we have developed and successfully used to support decision-makers in defense and homeland security. These include single-stage designs appropriate for hundreds of factors, as well as sequential approaches that can be used to screen thousands of factors. We conclude with brief overviews of some recent applications to support Army projects. While our work was motivated by DoD needs and requirements, these technologies should be of equal interest to industry and policy decision makers.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Novel Experimental Design & Analysis Methods for Simulation Experiments Involving Algorithms

Terril N. Hurst, Chatterpaul Joseph, John S. Rhodes, and Keith Vander Putten
Raytheon Missile Systems

MISSILE system algorithms pose challenges for applying traditional design of experiment (DOE) methods. The notion of a "response surface" is stretched to the limit, due to (a) the intrinsically nonlinear, discontinuous nature of the observed response within the factor space and (b) the presence of many categorical factors that interact with numeric factors. Additionally, as in most simulation experiments, algorithm analysis is further complicated by non-constant response variance.

This paper describes how designed simulation experiments were conducted to allocate missile system tolerance requirements in order to achieve optimal performance of the system's target tracking algorithm. A statistical model of the seeker subsystem was used in a reduced-fidelity simulation in order to make Monte Carlo simulation feasible within the allotted time. Initial factor screening was accomplished using the reduced-fidelity simulation and a two-level Resolution V fractional factorial design. A Latin Hypercube Sampling experiment was then conducted in order to build a more complex model of the system response. This model was employed within a novel tool called RAVE, or Raytheon Analysis of Variability Engine. RAVE is based upon the generation of system moments ("delta method") and a novel tolerance allocation algorithm to conduct sensitivity analysis and to allocate system performance margin to design variables.

Since the initial simulation experiments used a reduced-fidelity simulation to allocate tolerances, this allocation required verification using a high-fidelity, scene-based integrated flight simulation (IFS). Each IFS run required several hours, so it was infeasible to run the required number of Monte Carlo replicates for achieving high confidence in the results. Therefore, a second novel tool was used, called PEM, or Point Estimation Method. PEM is a "smart sampling" alternative to Monte Carlo analysis, requiring a small fraction of the number of simulation runs. PEM takes as input the factors' moments (mean, variance, skewness, and kurtosis) and estimates of the first and second partial derivatives of the response. These derivatives are obtained either by differentiating an analytic response function (infeasible in this case) or by numerical estimation, based upon a few one-factor-at-a time simulation runs near the nominal design point. Given these inputs, PEM produces a menu of parsimonious experimental designs, as well as a response prediction before any additional simulation runs are executed as chosen from the design menu.

The RAVE and PEM tools have recently been converted from Excel/VB to MATLAB, which is a more commonly used language within the algorithm developers' community. The tools can be used either via a graphical user interface (GUI) or the MATLAB command line.

The paper describes results of both the reduced-fidelity and high-fidelity experiments. Alternative empirical models are compared in both cases: low-order polynomials, logistic regression, and regression trees. Extensions to the RAVE and PEM tools for categorical factors are also described.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Meta-modeling of Personnel Entry into Collectively-Protected Shelters

Jack Edwards, Jung-Ii Choi, Jeffrey Eischen, Nathan Obringer, NC State University
Harold Barnette and Gaurang Dave, Naval Surface Warfare Center
Roger Gibbs, BSM, Inc.

Entry of personnel into collectively protected shelters constitutes a critical point in which system protections are necessarily breached. A typical entry event starts with doffing of possibly contaminated clothing, followed by entry into an airlock that separates the area external to the shelter from the 'toxic-free area' (TFA) inside the shelter. The airlock operates using principles of over-pressure –air from the TFA or from a separate HVAC unit flows into the airlock, increasing the inside pressure to levels higher than the external air pressure. The air flow is vented out of the airlock, moving either outside or back into the HVAC unit, where it is filtered. The over-pressure helps keep contaminated air from flowing into the airlock, while the vented flow drives contaminated air out of the airlock. As a person enters the airlock, two things happen. First, the over-pressure is relieved as the doors are opened, causing a strong gas jet to flow outward. Second, wakes generated by the motion of the doors and the person cause a net flow of contaminated gas into the airlock. Once the doors close, the over-pressure begins to build up, and the contaminated gas is purged from the airlock. The initial agent concentration within the airlock is a function of the external agent concentration, the wind speed and direction, and specific details of the entry event, such as the walking speed of the person(s) entering, the door opening mechanism, and the over-pressure level within the airlock.

As part of the DoD's Joint Expeditionary Collective Protection shelter system program, NSWC-Dahlgren has been tasked to develop a System Performance Model (SPM) for the shelter system. A prediction of the amount of agent transported into an airlock and the TFA is a required feature of this model. As experimental data for personnel entry is generally unavailable, a computational fluid dynamics (CFD) approach is used to generate time-dependent predictions of agent transport during entry. An outcome of the airlock SPM is a model for the cubic-feet of air (CF) transported versus time. Given the external agent concentration as determined by other techniques, the CF versus time distribution can be used to predict the agent concentration in the airlock for each entry event.

CFD simulations of this type have been performed for entry into a generic airlock using a sequence of trials generated by a Latin hypercube design-of-experiments (DOE) procedure. The trials vary the prevailing wind speed (0 to 5 m/s), the wind direction (0 to 180 degrees), the proximity of a person from the airlock (0 to 3 m), and the walking speed (1 to 1.5 m/s). Based on the CFD data, a meta-modeling approach is used to reduce the CF versus time distributions into a simpler form that correlates the response with the driving parameters of wind speed, wind direction, proximity, and walking speed. The modeling challenge is the regression of the parameters that correlate the distributions as a function of the driving parameters and the fact that the number of trials that can be performed is limited greatly due to the expense of the simulations. The talk will describe the CFD modeling of the ingress events, the design-of- experiments procedure, and the non-linear regression techniques used to obtain the meta-model. Simulations of trials not used in constructing the regression model are used as checkpoints to assess the performance of the meta-model.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Statistical Origins of Allometry

Bruce J. West
US Army Research Office

Allometry is the study of size and its consequences and is a recurrent theme in the study of complex networks including those in the Army's network-of-networks. The most studied of the allometric relationships is that between basal metabolism R and total body mass M and is expressed as $R=aMb$. Most investigators have focused on determining the proper value of b entailed by an appropriate biological model and have dismissed the coefficient a as an irrelevant constant. In order to develop an argument that is applicable to general complex networks we use data to settle on the empirical value of b and find that data gives rise to a random a described by a Pareto distribution. The intermittent fluctuations in a are of the same statistical form as those of speciation found in the punctuated equilibrium theory of biological macroevolution [1]. The theoretical probability density for the time t between bursts of species generation is consequently determined to be inverse power law with index -2 in agreement with the empirical distribution determined by the fossil record [2, 3]. This reasoning also applies to the statistics of innovation in complex networks such as in the Army's Network-centric Operations.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Modeling, Simulating, Measuring, and Understanding Cooperation

Chris Arney, Kristin Arney, and Elisha Peterson
United States Military Academy

The authors utilize the results of simulations for cooperative systems to develop metrics to measure system performance and properties. The framework of cooperative game theory sets the context for understanding cooperation in simulations of various applications including network flow, pursuit and evasion, and operational equi-distribution of assets. A statistical analysis of simulation data from multiple algorithms provides metrics of each algorithm's contribution (altruistic or competitive) to overall system performance. Various cooperative/competitive flow management schemes for the operation of a simple network are simulated and analyzed for their altruistic factor. Altruism is shown to have a positive effect on these networks' performance. Pursuit and evasion games encompass a large class of games in which pursuers attempt to find and/or capture evaders. We discuss insights gained on the efficacy of cooperation in pursuit and evasion in various contexts. One of the fundamental ways that entities cooperate with one another is by sharing of tasks through synchronized movement to balance their geometric load. For example, players of a team defending a goal may be assigned equal-spaced zones to defend or units in a military force may be assigned equal-spaced sectors to defend or control. As the dynamics of the situation unfold and as entities are forced to move, withdraw, or enter the space, the other entities cooperate by adjusting their positions to retain their load balance. The various ways that this geometric cooperation can be accomplished from the perspective of local autonomous control are simulated and analyzed. The authors use the results of these various simulations and metrics to generalize concepts and develop an underlying theory of cooperation. Examples, simulations, metrics, and analysis of the algorithms' cooperative performance are presented and used to explain and establish this new theory of cooperation.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Regression Tree Models for Longitudinal Repeated Measures and Multivariate Response Data

Wei-Yin Loh
University of Wisconsin, Madison

Linear mixed models and GEE models are frequently used to analyze longitudinal and multi-response data. These parametric and semi-parametric approaches are quite satisfactory when the number of explanatory variables is small. When there are very many explanatory variables, with some having no effect and others nonlinear effects on the responses, it can be very difficult to find good models using these techniques. We present an alternative approach based on regression trees that is completely nonparametric as well as completely automatic. We also discuss the goals of the analysis of these types of data.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Discovering Partially Ordered Patterns of Terrorism via Spatio-Temporal Data Mining

James A. Shine and James P. Rogers, US Army ERDC Topographic Engineering Center
Shashi Shekhar and Pradeep Mohan, University of Minnesota

An ongoing military research problem is the discovery of useful patterns in spatio-temporal (ST) data. The Army continues to generate larger and larger amounts of ST data in many different and interconnected databases, and the limited capacity of human analysts to properly analyze all these databases continues to become more problematic. In particular, the relationships between different ST events are vital to increasing knowledge and understanding of military challenges such as discovering and predicting insurgent attack patterns and tactics, modeling cultural geography, asymmetric warfare, and other civil affairs and human terrain paradigms.

The current work grew out of earlier efforts to discover patterns in spatial and ST data. The earliest effort used an extension of the principle of association rules to the spatial realm to create the concept of spatial co-location [1]. Subsequent work extended this concept to consider temporal as well as spatial co-location and led to the discovery of various ST co-occurrence patterns. Our most innovative discovery was mixed-drove co-occurrence patterns (MDCOPs), where two or more different ST object-types were found to have a significant percentage of instances within a spatial and temporal neighborhood of each other [2].

The work presented in this paper, examines cascading or partially ordered ST patterns (CSTPs) and their discovery by computational approaches. As with ST co-occurrence patterns, the object-types or events occur within a defined spatial and temporal neighborhood of each other, but in CSTPs a temporal order is assumed. For example, a sect may blow up a religious shrine of another sect, leading to increased events such as suicide bombings, attacks on civilians, and destruction of infrastructure such as water and oil supplies. The initial event or events is called a generator or attractor, and subsequent events are assumed to at least be influenced by the generator or attractor.

This work requires the creation and testing of new metrics, or interest measures, to determine to frequency and influence of CSTPs. Also, because order is now a variable, the computational issues are even more challenging than on previous efforts. In addition, we are beginning to look at issues of multiple scales; for example, a spatial neighborhood may be measured in meters in a city but miles in a region, and a temporal neighborhood can be measured anywhere from seconds to months depending on the speed of the events being examined. This adds yet another layer of computational challenge and increases the requirement for innovative algorithms and programming heuristics to ensure computational tractability.

We have developed an interest measure for discovering CSTPs called acascade participation index (CPI). This interest measure has both a loose and a tight form, depending on the sensitivity desired in the discovery algorithm. We have also developed two separate filters for pruning the search space: a multi-resolution filter and an upper bound filter. The multi-resolution filter in particular has shown great computational promise. We have tested our

approach on a crime data set from the city of Lincoln, Nebraska and discovered some CSTPs. In particular, bar closings and Saturday nights seem to be linked to a significant increase in the rate of crimes such as burglaries, assaults and vandalism. Al Qaeda, other military data sets?

We are aware of the applications of this concept to non-military patterns such as el Nino effects, global financial meltdowns such as the one in the fall of 2008, and the possible spread of epidemics. We are working on discovering patterns in an el Nino database and hope to have some results by the time of the conference.

REFERENCES

[1] N. Mamoulis, "Co-location patterns and Algorithms", in Encyclopedia of GIS, ISBN 978-0-387-30858-6, Springer 2008, pp. 103-107.

[2] M. Celik, S. Shekhar, J. P. Rogers, J. A. Shine, and J.S. Yoo. Mixed-Drove Spatio-Temporal Cooccurrence Pattern Mining In IEEE Transactions on Knowledge and Data Engineering (TKDE), October 2008.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Design of Experiments for Test and Evaluation: Best Practices and Future Challenges

Laura Freeman
Institute for Defense Analyses

Design of Experiments (DOE) is a scientific tool for conducting rigorous and comprehensive tests. In November 2009, Dr. Gilmore, Director, Operational Test & Evaluation wrote in a memorandum to his staff, "DOE provides the scientific and statistical methods needed to rigorously plan and execute tests and evaluate their results." This talk provides an overview on the current applications of DOE to test and evaluation. Best practices are examined from a statistical standpoint and through current case studies. Future challenges for the test and evaluation community for the robust application of DOE are identified. The talk concludes by outlining research areas for the statistical community that will extend the application of DOE methodologies to test and evaluation.

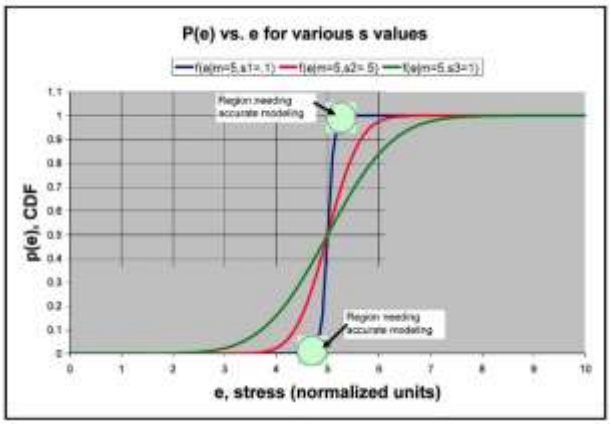
2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Comparison of Methods for Determining Sensitivity of Explosive Devices to Applied Stress

Allan T. Mense and Don Woolley, Raytheon Missile Systems
Jerry Alderman, St. Johns Research

The problem of sensitivity analysis for explosives is as old as time in memorial. Rockets, in military and recreational uses, date back to at least the 13th century [1]. From the time the Chinese invented (discovered) gunpowder through the advent of nuclear weapons there has always accompanied weapons development the task of safe storage and predictable performance.

In this regard in 1943, at the Army’s Bruceton, Pennsylvania Explosives Research Laboratory a technique was developed to find the sensitivity curve of probability of detonation vs. externally applied stress for any explosive [2]. This cumulative probability distribution p(e) was needed in order to determine the “always safe” and “always fires” values for external triggers for explosive devices. Such a curve might take on the form shown below.



The determination of the shape of this curve is the content of this paper. The original Bruceton Method was useful and given that was developed before the age of powerful computers and without the apparent benefit of statistically design experiments it was a good engineering solution to an important military and civilian problem. A number of critics including Neyer [3] have noted that a better approach would be to use test points (stress levels) based upon D-optimal designs. Such designs have the advantage of reducing the variance in the predicted coefficients that describe the sensitivity curve, p(e). Both approaches however make use of the assumption of the curve being a normal distribution [4] thus requiring the determination of only two parameters (m,s), and p(e)=F((e-m)/s). The introduction of asymmetry and even kurtosis to give a better fit to the data is explored in this paper. Originally a probit analysis was performed based on the binary (explosion/no explosion) data obtained from tests. This will be explored and is the subject of some scrutiny based on more modern design of experiments work combined with sequential testing. The importance of this analysis for this conference lies in the use of statistical techniques for exploring sensitivity analyses in the tails of non-normal distributions.

[1] Needham, Joseph (1986), *Science and Civilisation in China*, Cambridge: Cambridge University Press.
[2] W.J. Dixon & A.M. Mood,(1948), "A Method for Obtaining and Analyzing Sensitivity Data," Journal American Statistical Association (JASA), Vol. 43, pp 109-126.
[3] B.T. Neyer, (1994), “A D-optimality-Based Sensitivity Test,” Technometrics, Vol.36, No. 1, pp 61-70.
[4] E. Strømsøe, (1992), " Prediction of Extremely Low and High Explosion Frequencies from Sensitivity Tests," Propellants, Explosives, Pyrotechnics Vol. 17, pp 295-297.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Clay Calibration: Radial Dependence of Calibration Drop Depths for Body Armor Testing

Shane Esola, Barbara Gillich*, Ward Boughers, Sandra Meiselwitz
US Army Aberdeen Test Center

Clay calibration is conducted before every hard and soft body armor test on clay backing to ensure that the ballistic clay backing yields consistent, repeatable backface deformation (BFD) results during body armor testing. It is an indication of the clay's mechanical behavior and part-to-part consistency. An experiment was designed to answer questions concerning the homogeneity of the mechanical properties of the 24" x 24" x 5.5" clay used to measure BFD in body armor testing. The three main objectives of this test were:
1) characterize the relationship between drop depth (response) and radial distance from the center of the clay box;
2) characterize how that relationship may change as the clay box ages; and
3) characterize how those relationships may be affected by the position of the clay box in the temperature conditioning oven.
This presentation will review the design approach and results of the experiment.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

An Information Retrieval Framework for Assessing Intelligent Stare

Barry Bodt
US Army Research Laboratory

Current perception systems in computer vision target entity extraction from developing scenes; that is, they focus on identifying predetermined nouns such as human, truck, and wall. But other than tracking these entities from frame to frame and using trajectories for prediction, there is limited work focused on identifying action. A new DARPA program seeks to develop systems capable of identify verbs in developing scenes. Such a capability would have obvious benefits when applied to constant surveillance systems that currently require eyes on —moving us beyond "Scout TV" to a system of intelligent stare.

In this presentation, I offer, in the spirit of a clinical paper, a framework for how an assessment of the developing technology could be conducted. An initial set of verbs has been constructed and a corpus of videos depicting these verbs is underway. An evaluation plan currently exists in only broad terms; there is still time to influence the final approach. I intend to describe the problem in the first third of my time and then open the floor to discussion.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Meta-Analysis to Determine if Assigning Running Shoes Based on Footprint Shape Reduces Injury Risk in Military Basic Training

(Special Session on Health Science and Statistics in the DoD)

J.J. Knapik, D.W. Trone, L. Brosch, K.G. Hauret, T. Grier, S.H. Bullock, and B.H. Jones
US Army Center for Health Promotion and Preventive Medicine; Naval Health Research Center; Lackland Air Force Base

Popular running magazines and running shoe companies suggest that the shape of the footprint can be used as an indication of the height of the foot arch which can then be used to select individually appropriate types of running shoes. Shoe manufacturers market three types of running shoes designed for individuals with high, normal, and low arches: cushioned, stability, and motion control, respectively. Assignment of appropriate shoe types is hypothesized to reduce injury risk by compensating for presumed differences in running mechanics. This study involved a meta-analysis on published investigations examining whether or not assigning running shoes based on footprint shape influenced injury risk. The literature was searched and three independent studies were found which investigated the shoe assignment procedure in Army, Marine Corps, and Air Force basic training. All three studies used a randomized prospective design involving 2 groups. In the experimental (E) group, subjects were assigned motion control, stability, or cushioned shoes for footprints judged to represent low, medium, or high foot arches, respectively. A control group (C) received a stability shoe regardless of footprint shape. For the meta-analysis, a general variance base method was used that employed hazard ratios and confidence intervals to produce a summary hazard ratio (SRR) and summary 95% confidence intervals (S95%CI) pooling the results of the 3 investigations (Greenland, Epidemiol Rev 9:1-30, 1987). The three studies involved a total sample of 4,987 men and 2,245 women. The meta-analysis technique indicated little difference in injury risk between the E and C groups among men (SHR(E/C) = 1.04, S95%CI = 0.94–1.14) or women (SHR(E/C) = 1.05, S95%CI = 0.95–1.18). When only individuals with footprint shapes representing high and low arches were considered (1,236 men and 558 women) there was modestly higher risk among the E group compared to the C group among both men (SHR(E/C) = 1.20, S95%CI = 0.99-1.44) and women (SHR(E/C) = 1.11, S95%CI = 0.88–1.39). Assigning running shoes based on footprint shape did not reduce injury risk in basic military training.
SHRs were calculated as $\sum(w_i * \ln HR_i)/\sum W_i)$ where ln= natural log, HR_i is the HR for each study 1 through 3, and Wi=1/variance HR_i. The variance $HRi=[\ln (RR_u \div RR_i)/1.96]^2$, Where HR_u and HR_l are the upper and lower bounds, respectively, of the HRs in each study.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Effect of Insecticide-Treated Bed Nets on the Feeding Behavior of Sand Flies in the Laboratory

(Special Session on Health Science and Statistics in the DoD)

G. Zollner, Division of Entomology, Walter Reed Army Institute of Research
D. Brambilla and T. Nolan, RTI International
C. Lim, Walter Reed Army Institute of Research

Leishmaniasis is a current operational military threat and a globally devastating disease caused by a protozoan parasite (*Leishmania* spp.) transmitted via the bite of an infectious phlebotomine sand fly (*Phlebotomus* or *Lutzomyia* spp.). Currently there are no vaccines or prophylaxis available, so sand fly vector control methods are the only means of preventing leishmaniasis transmission. Personal protective methods include the use of insecticide-treated bed nets, but most nets target mosquitoes and not sand flies, which are smaller than mosquitoes and can easily pass through mosquito nets. The goal of this study was to evaluate three new types of long-lasting, insecticidal nets (LLINs) for their ability to prevent sand flies (*P. papatasi*) from passing through the treated netting and biting a host animal (mouse) on the other side of the netting. A series of experiments was designed to investigate the effect of treated netting on sand fly knockdown, mortality and feeding behavior. Assays were conducted using LLINs consisting of different mesh sizes and treated with a pyrethroid insecticide (permethrin, deltamethrin or α-cypermethrin). The outcome variable was the overall proportion of flies in a given treatment at a given time. Results were analyzed using logistic regression models of the probability that a sand fly passed through the netting or fed on the mouse as a function of mesh size, insecticide treatment, presence/absence of insecticide and/or time and their interactions. Linear contrasts were employed for pairwise comparisons of treatments or time points for each statistically significant overall effect. Where necessary, a repeated measures design was employed to account for any correlations among repeated observations on the same units. All analyses were performed in SAS.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

SAS Enterprise Miner and Application to Health Sciences using the DMSS

(Special Session on Health Science and Statistics in the DoD)

Theresa Real
Armed Forces Health Surveillance Center / General Dynamics Information Technology

This presentation will provide a brief overview of SAS Enterprise Miner (EM): its graphic user interface (GUI) for building process flow diagrams; data sampling and partitioning; different types of plots, modeling, and model comparisons available; and discussion of how EM may be used to generate hypotheses. An example of an application of EM using data from the DMSS will be provided. This example will center around Guillain-Barré Syndrome (GBS) and exploratory analysis, including sequential analysis, cluster analysis, and several models.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Adversarial Risk Analysis: A Routing Game

Shouqiang Wang and David Banks
Duke University

Suppose one wants choose a route for a convoy through a road network on which IEDs may have been placed. The Router may have unreliable private information on the location of the IEDs; the Attacker may have unreliable private information on the routing choice. Additionally, both sides may have distinct utility functions that are imperfectly known to their opponents. Classical game theory cannot solve this problem, but Adversarial Risk Analysis, using a Bayesian mirroring argument, can produce reasonable solutions.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Adversarial Risk Analysis: Borel Games

David Banks, Francesca Petralia, and Shouqiang Wang
Duke University

The Borel game, or La Relance, is a simplified game of poker that has been analyzed by some of the giants in game theory---it is technically challenging, and there is not yet a full solution to the general problem. This talk presents a Bayesian analysis, using the mirroring argument from Adversarial Risk Analysis. In many ways, the solution is more satisfactory than is obtained from classical minimax theory.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Adversarial Risk Analysis in Bioterrorism

Juan Vivar
Duke University

This talk applies a "mirroring" technique to a game theory problem, so that one opponent models the decision making of the other, while taking account of the fact that the opponent is performing a symmetric analysis. The approach is Bayesian, and allows one to incorporate probabilistic information. The ideas are illustrated through an application to the 2002 U.S. response to the possibility that smallpox would be used as a bioweapon by terrorists.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

RExcel for Experimental Design and Analysis

Richard Heiberger
Temple University

Many people use Excel, some even use it for their statistical analyses and production of their graphs. Providing superior tools to these users is of utmost importance. The RExcel interface provides a spreadsheet front end to the superior graphics and statistical analyses available in R. By placing any R function inside the spreadsheet paradigm of automatic recalculation, we have placed the full power of R, one of the best programs for statistical analysis and graphical display of data, into a familiar user environment.

I use RExcel with Rcmdr as the primary software tool in introductory courses. I use the RExcel interface to R as a supplementary tool in advanced courses. The Rcmdr package provides a clickable menu into R. RExcel places the Rcmdr menu on the Excel menu bar. Our worksheets designed for teaching---using Excel for problem specification and R for graphics---illustrate the concepts of hypothesis testing, construction and interpretation of confidence intervals, and least squares fitting. We use Excel scrollbars, checkboxes, and change of data values for dynamic control of R graphical displays.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Predictive Modeling with Wide Data: Cross-Validation Model Comparison and Learning Curves

Russ Wolfinger, Pei-Yi Tan, and Padraic Neville
SAS Institute, Inc.

Many military and applied statistics problems in general can be formulated as ones of prediction, in which a dependent variable of interest is to be predicted as a function of many independent variables. Modern measurement technologies often produce orders of magnitude more independent variables than observations, the well-known n << p problem. Overfitting is a serious risk to generalizability in this case, compounded by the typical situation where no single class of models has been demonstrated to be superior across the problem set. We present a comprehensive model screening approach based on honest cross-validation and side-by-side comparisons using common performance metrics. We also discuss learning curves, which aid in determination of sample size for prediction problems. Examples from genomics, QSAR, and clinical trials provide illustration.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Model Calibration through Minimal Adjustment

V. Roshan Joseph and Chia-Jun Chang
Georgia Institute of Technology

The Gaussian process model provides a powerful methodology for calibrating a physical model in the presence of model uncertainties. However, if the data contains systematic experimental errors, then the Gaussian process model can lead to an unnecessarily complex adjustment of the physical model. In this work, we introduce an adjustment procedure that brings the physical model closer to the data by making minimal changes to it. This is achieved by approximating the Gaussian process model by a linear regression model and then applying a simultaneous variable selection of the model and experimental bias terms. Two real examples and simulations are presented to demonstrate the advantages of the proposed approach.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Detection of Outliers in Spatial-Temporal Data

James P. Rogers
US Army ERDC Topographic Engineering Center

Outlier detection is an important data mining task that is focused on the discovery of objects that are exceptional when compared with a set of observations that are considered typical. These objects are important since they often lead to the discovery of exceptional events. Outlier detection can reveal objects that behave anomalously with respect to other observations, and these objects may highlight current or future problems. In other cases, the sudden appearance of a large number of outliers can point to a change in the underlying process that is generating the data. Outliers are observations are inconsistent with the majority of the data and deviate significantly from the normal observations. The goal of outlier detection methodology is to not make assumptions about the data and the data distribution. A robust outlier detection methodology should not assume a distribution of the data.

The objective of my research is to develop a robust method of diagnosing outliers and to extend it to detecting outliers in spatial-temporal data. In spatial-temporal data, observations are tagged with their geographical location and their timestamp. These observations are a vector of attribute values for the features being measured. A spatial-temporal outlier is an observation whose values are significantly different from those of other spatially and temporally referenced objects in its spatial-temporal neighborhood. It represents an object that is significantly different from its neighborhood even though it may not be significantly different from the entire population.

Geographic phenomena, which provide spatial-temporal data, are dynamic, active, and change over time in direction, speed, growth, location, and other variables. Many geographic related phenomena could be changing simultaneously in space and time, and often it is difficult to predict these changes. These phenomena are difficult to analyze using traditional data mining methods. Determining relationships among phenomena as they move and change over time is not possible by means of human analysis of spatial-temporal data streams. Also, the volume of spatial-temporal data being collected is increasing steadily due to the usage of cameras, sensors, and mobile devices (e.g., cell phones) and is too much data for the human to analyze.

Previous outlier detection methods have focused primarily on only one non-spatial numerical attribute and have not successfully dealt with multiple dimensions, so there is a need to better identify outliers in high-dimensional data. Many previous methods assume a Gaussian distribution of the data which is probably a major fallacy in determining outliers for spatial-temporal data. It is strongly desirable to not presume any distribution of the data. Most previous efforts did not provide a statistical confidence measure, but including a confidence measure should improve the detection of outliers. Outlier detection is often complicated by noise in the data, so a good outlier detection methodology should be successful in identifying outliers in noisy data. Global outlier methods calculate a single outlier statistic that summarizes the outliers for the entire geographic area and temporal duration, while local outlier methods calculate a outlier statistic for each feature based on its similarity to its neighbors. Most

previous outlier detection methods have focused on global outliers instead of local outliers, and previous methods have not been able to determine outliers as the vector of attributes, location, and time change. The objective of my research is to devise a methodology to address these problems and challenges.

My method, unlike many detection methods found in the literature, does not require the user to enter the number of outliers to be found or the percentage of outliers to be found and does not assume any distribution of the data (e.g., Gaussian). My method only requires the input of two parameters: the statistical confidence level and the number of nearest neighbors, and only the statistical confidence level is significant. My method allows for different ways to measure the degree of non-conformity and works for high-dimensional data, noisy data, and data with or without clustering information.

The basic outlier detection method was extended to spatial-temporal data that uses kernels for vector of attributes, spatial, and time to provide capability to focus outlier detection on local neighborhoods, and the user is able to input weights for each of the kernels. My method is capable of effectively detecting local outliers, which overcomes one of the main problems of previously developed techniques for outlier detection based on global parameters. Local spatial-temporal outliers are outliers determined within a specific spatial area and time frame, which is a subset of the entire spatial area and total temporal duration. A non-parametric likelihood-based method was created, using Kernel Density Estimation, to improve computational performance, because nearest neighbor computations are costly.

Empirical evaluation was conducted on several datasets with very good results achieved. The datasets increased in complexity and dimensionality. The experiments on these datasets using my method produced results with a high True Positive percentage and a low False Positive percentage.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Fitting pdfs to Small Datasets

Jerry Alderman
St. Johns Research

An Excel based tool tests goodness of fit (gof) of pdfs to data by applying a U Statistic to all possible subsets of small datasets. For a typical dataset of size 30, Excel memory limits subset sizes to about 8 but these 5,852,925 tests would be conducted automatically and the fraction passing compared to a preselected confidence level to establish an overall accept/reject decision for the proposed fit. For datasets as small as 6, up to 63 independent statistical tests can be done so the method leverages the amount of information one can extract from a small sample. The method is illustrated with examples that test the gof of Johnson distributions to small samples drawn from known Raleigh distributions

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Design of Experiments in the Department of Defense

W. Ward Vaughn
Office of the Secretary of Defense

Review of the application of Design of Experiments (DOE) within the Department of Defense (DoD) indicated that DOE is infrequently used and in many cases was inconsistently applied and insufficiently detailed when performed. In February 2010, a team of over 60 DoD DOE professionals was established to evaluate the root causes of these deficiencies and propose solutions. Team observations were:
(1) DOE training programs within each of the Services were widely different and, in some cases, deficient in both clarity and content;
(2) DOE concepts were inconsistently understood and insufficiently communicated during training within the Services;
(3) DOE programs were insufficiently defined and structured, and
(4) some Test & Evaluation was insufficiently defined to prescribe the timely application of DOE to weapon system acquisition programs.

Results of the DoD DOE Lean Six Sigma Project included an improved elaboration of the Integrated Test & Evaluation (T&E) lifecycle to promote appropriate phased consideration, and meaningful inclusion, of DOE. Further results of the DOE Project included elaboration of numerous primary concepts within DOE to promote consistent understanding and use of DOE as well as documentation of an improved DOE training program model.

The Project also succeeded in raising the awareness and use of DOE within development and operational T&E of weapon systems across the DoD by establishing a knowledge base and identifying centers of excellence and DOE contacts within the DoD. Numerous DOE concepts have been defined precisely, in DoD terms, allowing for improved consistency and relevance within DoD DOE training. A curriculum for training has been established to promote relevant and sufficient DOE training across the DoD which will lead to improved weapon system acquisition by reducing their costs and improving their objectivity with actionable and accurate description of the weapon system's developmental and operational status.

2010 "Sixteenth Army Conference on Applied Statistics", sponsored by Interface Foundation of North America

Maximum Clarity: An Alternative to Minimum Aberration for Fractional Factorial Designs

Jose G. Ramirez, W.L. Gore and Associates
Randall D. Tobias, SAS Institute, Inc.

Minimum Aberration (Fries and Hunter, 1980) is a design criterion for which the longest alias string of a given order is as short as possible, so that no effect has many more aliases than any other. For the past 25 years this criterion has been used for designing highly fractionated factorial designs and, more recently, for designing fractional factorial split-plots (Bingham and Sitter, 1999).

In this presentation we discuss an alternative criterion for fractional factorial design, which we call maximum clarity. A clear effect is one which is aliased with no other effect of interest, so that it can be estimated without bias. Our criterion specifies that there should be a maximum number of these. That is, in terms of the alias strings, maximum clarity requires that there be as many as possible of length 1.

We will motivate the need for maximum clarity in industrial experiments, in particular split-plot experiments. Examples of maximum clarity designs, generated using the SAS/QC procedure FACTEX, will be given and contrasted with minimum aberration alternatives.

2010 “Sixteenth Army Conference on Applied Statistics”, sponsored by Interface Foundation of North America

Doing the Right Things: Experimenting So That Warriors Do Not

Greg Hutto
53rd Wing Eglin Air Force Base

No one desires a poorly planned test program; all aspire to excellence! Our USAF Chief, Gen Schwartz, reminded testers that our warriors will pay the price for defective systems on the battlefield. Everyone desires excellence, but how to attain excellence in test? It was Peter Drucker who said: “Management is doing things right; Leadership consists of doing the right things.” This talk will consider the essential role of leadership in transforming the business of test in the DOD from “best efforts” to demonstrably excellent tests. It is easy to make the case that design of experiments (DOE) is a superior experimentation strategy to traditional methods of test including scenario (use case) or varying one factor at a time. It is also a simple matter to recount numerous instances in the literature, dating back to R.A. Fisher in the 1920’s, where the statistical principles of test design have transformed testing from art to science. Why, then, is experimental design not by far the most widely practiced test strategy in the Department of Defense? Simply put, it is because our leaders have not yet successfully led our organizations in the changes necessary to adopt it. In turn, they have not led us because we, in the technical statistical community, have not made them aware of why they should and how they can do so. In this talk, I wish to discuss the right things: helping our leaders to catch the vision of test excellence, establishing the infrastructure to make well-designed tests a reality, and a series of quality metrics to ensure our proposed test programs are, in fact, excellent.