ARMY RESEARCH LABORATORY

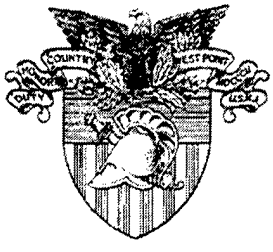
Proceedings of the Fourth Annual U.S. Army Conference on Applied Statistics, 21–23 October 1998

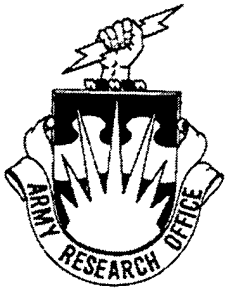
Barry A. Bodt
EDITOR

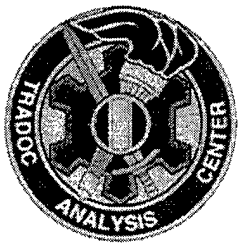
Hosted by:
TRADOC ANALYSIS CENTER-WSMR

Cosponsored by:
U.S. ARMY RESEARCH LABORATORY
U.S. MILITARY ACADEMY
U.S. ARMY RESEARCH OFFICE
WALTER REED ARMY INSTITUTE OF RESEARCH
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
TRADOC ANALYSIS CENTER-WSMR

ARL-SR-84 November 1999







The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5067

ARL-SR-84 **November 1999**

Proceedings of the Fourth Annual U.S. Army Conference on Applied Statistics, 21–23 October 1998

Barry A. Bodt, Editor
Information Science and Technology Directorate, ARL

Hosted by:
TRADOC Analysis Center-WSMR

Cosponsored by:
U.S. Army Research Laboratory
U.S. Military Academy
U.S. Army Research Office
Walter Reed Army Institute of Research
National Institute of Standards and Technology
TRADOC Analysis Center-WSMR

Abstract

The fourth U.S. Army Conference on Applied Statistics was hosted by the U.S. Army Training and Doctrine Command (TRADOC) Analysis Center - White Sands Missile Range (TRAC-WSMR) during 21–23 October 1998. Two sites were used for the conference. The meeting began at the Corbett Center on the campus of New Mexico State University in Las Cruces and concluded at WSMR. The conference was cosponsored by the U.S. Army Research Laboratory (ARL), the U.S. Army Research Office (ARO), the U.S. Military Academy (USMA), TRAC-WSMR, the Walter Reed Army Institute of Research (WRAIR), and the National Institute for Standards and Technology (NIST). The U.S. Army Conference on Applied Statistics is a forum for technical papers on new developments in statistical science and on the application of existing techniques to Army problems. This document is a compilation of available papers offered at the conference.

FOREWORD

The fourth U.S. Army Conference on Applied Statistics was hosted by the U.S. Army Training and Doctrine Command (TRADOC) Analysis Center - White Sands Missile Range (TRAC-WSMR) during 21–23 October 1998. Two sites were used for the conference. The meeting began at the Corbett Center on the campus of New Mexico State University in Las Cruces and concluded at WSMR. The conference was cosponsored by the U.S. Army Research Laboratory (ARL), the U.S. Army Research Office (ARO), the United States Military Academy (USMA), TRAC-WSMR, the Walter Reed Army Institute of Research (WRAIR), and the National Institute for Standards and Technology (NIST). The U.S. Army Conference on Applied Statistics is a forum for technical papers on new developments in statistical science and on the application of existing techniques to Army problems. The purpose of this conference is to promote the practice of statistics in the solution of these diverse Army problems.

The fourth conference was preceded by a short course, "Bayesian Statistical Inference: Principles, Techniques, and Applications," given by Professor Nozer Singpurwalla of George Washington University. Several distinguished speakers spoke during invited general sessions: James Thompson (keynote), Rice University; Richard Laferriere, TRAC-WSMR; Stephen Robinson, University of Wisconsin-Madison; John Bart Wilburn, University of Arizona; Francisco J. Samaniego, University of California, Davis; Sanford Weisberg, University of Minnesota; and Boris Rozovskii, University of Southern California. Three themes were woven through this year's conference. Inference based on combat simulations was addressed in both a special session organized by the Naval Postgraduate School and in a morning session of simulation tools demonstrations and talks organized by the conference host and held on site at WSMR. Statistical application to natural language processing was featured in two sessions, and Bayesian statistical methods in solving Army problems were explored in the tutorial preceding the conference and in one contributed session. An important moment in the conference was the awarding of the Army Wilks Award to Robert L. Launer of ARO "for his major unique contributions to Army statistics and the profession of statistics, and by the highly effective ways that he has brought together academic statisticians and Army scientists to solve problems important to the nation."

The Executive Board for the conference recognizes Ms. Lounell Southard, TRAC-WSMR, for hosting the conference; Mr. David Webb, ARL, and Mr. Edmund Baur, ARL, for assisting with advertisement; Dr. Edward Wegman, GMU, for fiscal oversite; Dr. Jock Grynovicki, ARL, for chairing the Army Wilks Award Committee, and Dr. Barry Bodt, ARL, for chairing the conference and serving as editor of the proceedings. Special thanks is due Linda Duchow, ARL, who handled many on-site details.

INTENTIONALLY LEFT BLANK.

Table of Contents*

* This Table of Contents contains only the papers that appear in the Proceedings.

FINAL PROGRAM
FOURTH U.S. ARMY CONFERENCE ON APPLIED STATISTICS

19–23 October 1998

Hosted by TRADOC Analysis Center-WSMR

Cosponsored by:

U.S. Army Research Laboratory
U.S. Military Academy
U.S. Army Research Office
Walter Reed Institute of Research
National Institute of Standards and Technology
TRADOC Analysis Center-WSMR

Monday, 19 October 1998

0800 - 0900	REGISTRATION (Corbett Center)
0900 - 1200	TUTORIAL (Room 315)
	BAYESIAN STATISTICAL INFERENCE: PRINCIPLES, TECHNIQUES, AND APPLICATIONS Nozer D. Singpurwalla, George Washington University
1200 - 1330	Lunch
1330 - 1600	TUTORIAL

Tuesday, 20 October 1998

0800 - 1200	TUTORIAL
1200 - 1330	Lunch
1330 - 1600	TUTORIAL
1800 - 2000	SOCIAL (Holiday Inn)

Wednesday, 21 October 1998

0800 - 0845 REGISTRATION (Corbett Center, Level II)

0845 - 0930 CALL TO ORDER: (Auditorium)

Chairman: Barry Bodt, U.S. Army Research Laboratory (ARL)

Host: Lounell Southard, U.S. Training and Doctrine Command (TRADOC) Analysis Center-WSMR (TRAC-WSMR)

OPENING REMARKS:

Gary Cunningham, Vice-President of Research, New Mexico State University

John Lyons, Director, ARL

Roy Reynolds, Director, TRAC-WSMR

0930 - 1145 GENERAL SESSION I (Auditorium)

Chair: Barry Bodt, ARL

0930 - 1030 KEYNOTE ADDRESS

DATA, MODELS, REALITY: STATISTICIANS IN THE NEW AGE
James R. Thompson, Rice University

1030 - 1045 Break

1045 - 1145 SCENARIO ANALYSIS IN U.S. ARMY DECISION MAKING
Richard R. Laferriere, TRAC-WSMR
Stephen M. Robinson, University of Wisconsin-Madison

1145 - 1300 Lunch

1300 - 1430 CONTRIBUTED SESSION I (Auditorium)

Chair: Major Andy Napoli, U.S. Military Academy (USMA)

THE USE OF COGNITIVE PROCESSING ADAPTIVE TO DECISION MAKING IN THE JWARS PROJECT
Charles R. Leake, Concepts Analysis Agency & JWARS Office

RAPID FORCE PROJECTION INITIATIVE ADVANCED CONCEPT TECHNOLOGY DEMONSTRATION (REPI ACTD)–THE EXPERIMENTAL PATH
Paul J. Deason, TRAC-WSMR
Greg B. Tackett, RFPI PMO, Redstone Arsenal

CONTRIBUTED SESSION I (Auditorium - Continued)

EMPIRICAL PERFORMANCE OF SOME TESTS OF HYPOTHESIS ON CASTFOREM
Patrick D. Cassady, TRAC-WSMR

1300 - 1430 CONTRIBUTED SESSION II (Dona Ana Room)

Chair: Douglas Tang, Walter Reed Army Institute of Research (WRAIR)

AN APPLICATION OF MIXED MODELS FOR COMPARING ACCURACY FROM TWO TYPES OF FIRING PLATFORMS
David W. Webb, ARL
Thomas Mathew, University of Maryland - Baltimore County

MINIMUM-PERCENTAGE-ERROR REGRESSION UNDER ZERO-BIAS CONSTRAINTS
Stephen A. Book and Norman Y. Lao, The Aerospace Corporation

GAUSS'S ADJUSTMENT ERRORS IN 1799
Aivars Celmins, ARL

1300 - 1430 SPECIAL SESSION I: STATISTICAL ISSUES IN NATURAL LANGUAGE (Senate Chambers)

Organizers: Clare Voss and Ann Brodeen, ARL

AN OVERVIEW OF STATISTICAL APPROACHES IN NATURAL LANGUAGE PROCESSING
Philip Resnik, University of Maryland, College Park

CONSTRUCTING BAYESIAN NETWORKS FROM WORDNET FOR WORD-SENSE DISAMBIGUATION: REPRESENTATIONAL AND PROCESSING ISSUES
Janyce Wiebe and Tom O'Hara, New Mexico State University
Rebecca Bruce, University of North Carolina - Asheville

STATISTICAL TECHNIQUES IN MULTILINGUAL DOCUMENT IMAGE ANALYSIS
Judith Hochberg, Los Alamos National Laboratory (LANL)

1430 - 1500 Break

1500 - 1630 CLINICAL SESSION I (Auditorium)

Chair: W. Jay Conover, Texas Tech University

Discussants:
Nozer Singpurwalla, George Washington University
Tapas Kanungo, University of Maryland, College Park
EVALUATION OF EMBEDDED MACHINE TRANSLATION SYSTEMS
Clare Voss, ARL

ISSUES IN LANGUAGE/ENCODING IDENTIFICATION FROM CORPORA
Jerry Geisler, Department of Defense (DOD)
Florence Reeder, Mitre Corporation

1500 - 1630 SPECIAL SESSION II: INFERENCE BASED ON COMBAT SIMULATIONS (Senate Chambers)

Organizer: David Olwell, Naval Postgraduate School

Discussants:
Don Gaver, Naval Postgraduate School
Jim Hodges, University of Minnesota
Rick Kolb, USMA

THE ADVANCED WARFIGHTING EXPERIMENTS (AWE): WHERE DO WE GO FROM HERE?
Tom Lucas, Naval Postgraduate School

USING EXPERIMENTAL DESIGN TO MINIMIZE RUNS FOR VERIFICATION AND VALIDATION OF LARGE-SCALE SIMULATIONS
Tom Curry, LOGICON

MODELING AND SIMULATION TECHNIQUES USED FOR ANALYZING DIGITIZATION OF THE BATTLEFIELD
Kevin Young, TRAC-WSMR

ASSESSING THE UTILITY OF DIGITIZATION OPTIONS FOR THE ARMY: PROBLEMS AND OPPORTUNITIES
Philip Beaver, USMA

1500 - 1700 CONTRIBUTED SESSION III (Dona Ana Room)

Chair: Bernard Harris, University of Wisconsin - Madison

ASSESSING TRADEOFFS BETWEEN SENSOR DWELL TIMES AND FREQUENCY OF REVISIT
Don Barr, USMA

Wednesday, 21 October 1998 (Continued)

CONTRIBUTED SESSION III (Dona Ana Room - Continued)

RELIABILITY TEST DESIGN & TEST-DESIGN EVALUATION TOOL
Michael Cushing and John Sereno, U.S. Army Material Systems Analysis Activity (AMSAA)

A PANSOPHIC APPROACH FOR RELIABILITY TESTING OF COMPLEX SYSTEMS
Mohammad (Mike) H. Danesh, PATRIOT Project Office

1830 - Wilks Award Banquet

Thursday, 22 October 1998

0830 - 1030 GENERAL SESSION II (Auditorium)

Chair: Edward Wegman, George Mason University

EXPLORING A SYNTACTICAL STRUCTURE OF FEATURES IN IMAGERY
John Bart Wilburn, University of Arizona

LINEAR DATA FUSION
Francisco J. Samaniego, Duane L. Steffey, and Hien Tran, University of California, Davis

1030 - 1045 Break

1045 - 1215 CONTRIBUTED SESSION IV (Auditorium)

Chair: Jayaram Sethuraman, Florida State University

HUMAN FACTORS ISSUES EVALUATED DURING THE 4TH INFANTRY DIVISION XXI ADVANCED WARFIGHTING EXPERIMENT (DAWE)
Jock O. Grynovicki, Kragg P. Kysor, Michael Golden, Madeline Swann, Shamim Rashid, and Tony Ward, ARL

A FRAMEWORK FOR MODEL VALIDATION
Robert G. Easterling, Sandia National Laboratories

1045 - 1215 CONTRIBUTED SESSION V (Senate Chambers)

Chair: Paul Deason, TRAC-WSMR

MODELING ERROR IN PROBABILITY-OF-KILL TABLES
Alyson Gabbard Wilson, Cowboy Programming Resources, Inc.

A BAYESIAN APPROACH TO TARGET VALUE ANALYSIS
Douglas H. Frank, Indiana University of Pennsylvania
Ann Brodeen, ARL

Thursday, 22 October 1998 (Continued)

CONTRIBUTED SESSION V (Senate Chambers - Continued)

SELECTION OF A PRIOR FOR RARE-EVENT PROBABILITY ASSESSMENT
Robert Launer, U.S. Army Research Office (ARO)

1045 - 1215 CONTRIBUTED SESSION VI (Dona Ana Room)

Chair: David Webb, ARL

DIRECTIONAL SPC: A MULTIVARIATE NONPARAMETRIC SPC TECHNIQUE FOR DETECTING DIRECTIONAL PROCESS CHANGES
Burak Birgoren and Russell Barton, Penn State University
COMBINING DATA FROM MULTIPLE SOURCES TO COMPUTE PREDICTION AND TOLERANCE BOUNDS
Jonathan F. Binkley, The Aerospace Corporation

1215 -1330 Lunch

1330 - 1430 GENERAL SESSION III (Auditorium)

Chair: Carl Russell, Joint National Test Facility

SOME GRAPHICS FOR LOGISTIC REGRESSION
Sanford Weisberg, University of Minnesota

1430 - 1445 Break

1445 - 1545 CONTRIBUTED SESSION VII (Senate Chambers)

Chair: David Cruess, Uniformed Services University of the Health Sciences

USING LOGISTIC REGRESSION TO EVALUATE GENDER RELATED EFFICACY DIFFERENCES TO MOSQUITO REPELLENT DATA
Claudia F. Golenda and J. Robert Burge, WRAIR

TOBIT ANALYSIS FOR HYPOTHESIS TESTING WITH LEFT-CENSORED DATA
Joan U. Clarke, Waterways Experiment Station

1445 - 1545 CONTRIBUTED SESSION VIII (Dona Ana Room)

Chair: James Thompson, Rice University

CONFORMATION IN METRIC PATTERN THEORY
Ulf Grenander, Brown University
Jayaram Sethuraman, Florida State University

Thursday, 22 October 1998 (Continued)

CONTRIBUTED SESSION VIII (Dona Ana Room - Continued)

THE COMPARATIVE EFFICACY OF SOME COMBINATORIAL TESTS FOR DETECTION OF CLUSTERS AND MIXTURES OF PROBABILITY DISTRIBUTIONS
Bernard Harris, University of Wisconsin, Madison;
Erhard Godehardt, Heinrich Heine Universitat, Dusseldorf, Germany

1545 - 1600 Break

1600 - 1700 GENERAL SESSION IV

Chair: Robert Launer, ARO

STATE ESTIMATION IN HIDDEN MARKOV MODELS WITH DISTRIBUTED OBSERVATION
B. L. Rozovskii and S. Kligys, University of Southern California

Friday, 23 October 1998

0800 - DEPART TO TRAC-WSMR

0915 - 0930 *WELCOME*
Roy Reynolds, Director, TRAC-WSMR (Main Conference Room)

0930 - 1200 TOOLS OF THE TRADE: TRAC-WSMR MODELS AND SIMULATIONS
(Main Conference Room)

0930 - 1000 SOLDIER STATION
John Galloway, TRAC-WSMR

1000 - 1030 JANUS
Mel Parish, TRAC-WSMR

1030 - 1100 OT-VIS
Joe Bebbs, TRAC-WSMR

1100 - 1130 CASTFOREM
Doug Mackey, TRAC-WSMR

1130 – 1200 COMBAT XXI
David Durda, TRAC-WSMR

1000 - 1200 SOLDIER STATION AND JANUS DEMONSTRATIONS

1200 ADJOURN

INTENTIONALLY LEFT BLANK.

Data, Models, Reality: Statisticians in the New Age

James R. Thompson
Department of Statistics, Rice University
Houston, Texas 77251-1892

ABSTRACT

The traditional role of statisticians has been the testing and modification of models in the light of data. Fast computing has tempted many to replace modeling by a variety of empirical procedures, from visualization to neural networks, nonparametric function estimation, and other forms of smoothing. If current "model-free" trends continue, it is possible that Statistics will simply become sublimated into other disciplines, such as Computer Science. On the other hand, the fact is that fast computing actually enhances our ability to build and modify deep models. We consider the possibility that, instead of collapsing, Statistics may have a new birth, based on an awareness that stochastic process modeling is now a real possibility.

INTRODUCTION

Data can provide a basis for inference into the underlying system(s) which produced them. A set of data, as a standalone, is usually a poor substitute for an understanding of a system. Models and simulations, unstressed by data, can lead to the formulation of actions and policies based on wishful thinking. Ideally, there should be a continuing dynamic between data, modeling and the production of simulations based on both.

In statistics, the advent of cheap high speed computing has not introduced the degree of interaction between data and the building of deep models that one might have wished. Exploratory Data Analysis [17] has led into a variety of data visualization techniques which frequently strive to be "model free." The maxim that "EDA lets the data speak to us without the interference of models" is taken quite seriously by many. Data is rotated, projected, transformed, etc., in order to allow the power of the human visualization system make judgments about it. The data becomes almost *sui generis*. We frequently do not ask about the system which generated the data.

On the other hand, models of extreme complexity, for example models of brigade combat, are built in almost microscopic detail, down to the level of the individual soldier. Historically, aggregation in war games is from larger units, say companies or regiments, to divisions. Almost all military people find this a more natural way to think than aggregating from performances of individual soldiers to divisions. That the dominant DOD war games are aggregates from the individual soldiers to the brigade is impressive computationally but may not be as practical from a modeling standpoint as one might wish. Basically, I believe the "high resolution"individual soldier level is utilized because the speed of the modern computer enables us to carry out computations at such an atomistic level. Computational speed drives the modeling process. It is true that another motivation is that these models are used to compare weapons systems more than to emulate actual battlefield combat.It is certainly true that if the purpose of the game is to compare one weapons system with another, then significant departures of the game from reality may not change the results very much. There is no doubt, however, that the high resolution games have not been tested very much on historical data, whereas the much simpler company level aggregate games now out of favor were extensively validated. Perhaps we should at least consider the possibility of using company level aggregate games as complements to the high resolution ones and comparing the results.

We should be able to use combat models, for example, to help explain why Mladic's campaign against the Bosnians was largely successful, whereas Rokhlin's campaign against the Chechens was a failure. Similar contestants on both sides with similar terrain. Why did the Serbs succeed while the Russians failed? We should be able to take combat models and use them with data from real combats and get, on the whole, the same results as those in reality.

Before computers dominated war gaming, kriegspiel-like board games were built which had very good comparisons in their outcomes when compared to real historical combats. Such games can, of course, be put on the computer, with the additional utilization of stochastic Lanchester rules. I used to assign such projects in my Rice model building class. Around 200 person hours were required for building games quite flexible

in terms of terrain, mobility and armaments. And validation on historical data was encouraging. (See [11], pp.54-71). I think it would be a good idea if a certain amount of effort were spent on looking at computer models which aggregate from, say, the company rather than the individual. By looking almost exclusively at the high resolution games, it would seem that we may be unnecessarily putting all our eggs in one basket.

The "data without models, models without data" syndrome has, no doubt, many causes. Certainly one of these is the fact that the digital computer proceeds digitally, whereas the human modeler proceeds analogistically. By leaving the data to stand alone, we can use the power of the digital computer to perform an almost limitless number of interesting spins, projections and transformations. And, in so doing, we are essentially working to accommodate the computer rather working in the human friendly analogue mode.

Similarly, when we build a "parameter rich" model (and sometimes the number of parameters is in the thousands), because the computer can indeed give time forecasts in a twinkling, we fail to note that we have essentially lost identifiability, so that estimating parameters from a data base is not readily doable. Again, we have fallen into the trap of letting the computer dominate the conversation.

The tendency of letting the computer dysfunctionalize our inference rather than enhancing it is perhaps the greatest danger to the science of Statistics. Not to utilize the computer would be, of course, nonsensical. The transition from analogue to digital computation is a forty years old *fiat accompli.* But a proper approach will enable us, essentially, to make the digital computer emulate analogue behavior when such is appropriate.

The topics proposed here are all computer intensive. But they are oriented toward friendliness to humans rather than friendliness to computers. They rely on the bedrock foundation of statistics, namely, logical inference based on facts. The topics are covered briefly here, with references given to readily available source materials.

VIEWING DATA IN THE LIGHT OF A MODEL: THE FIRST WORLD AIDS EPIDEMIC

At the low end of computer utilization, there is the humble spreadsheet. In Figure 1, I show the rather amazing results one gets when graphing the new AIDS case incidence per hundred thousand for a variety of First World Countries divided into that for the United States.

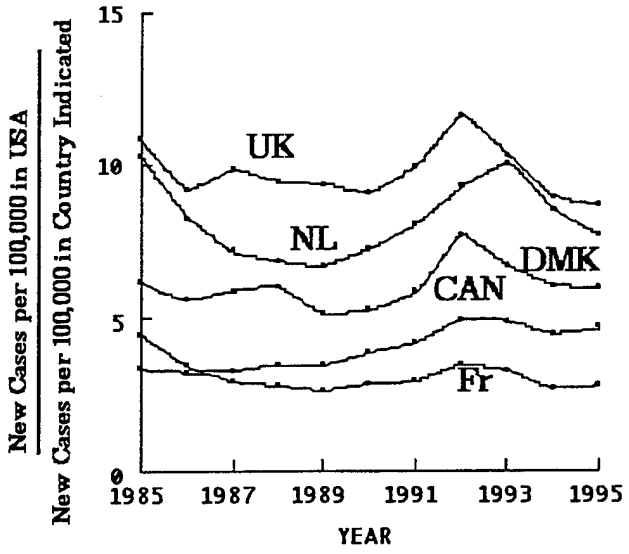


Figure 1. Comparative New Case Rates.

Now let us consider a simple kinetic equation for the growth of an epidemic.

$$\frac{dy_A}{dt} = k_A(t) y_A \tag{1}$$

We show in Figure 2 estimates for $k(t)$ rates on a year by year basis using [18]

$$k_A(t) \approx \frac{\text{new cases per year}}{\text{cumulative cases}} \tag{2}$$

Rates for Canada, Denmark, France, Netherlands, United Kingdom, United States

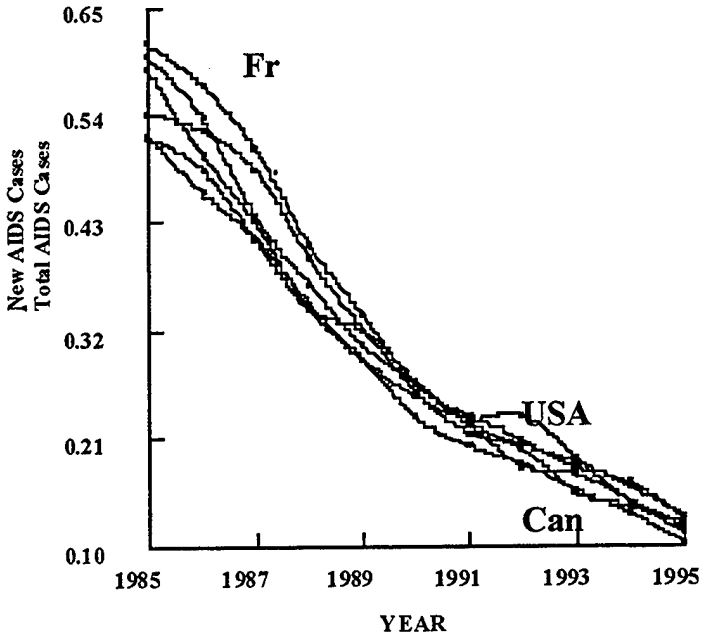


Figure 2. $k_{Country}(t)$ **Values.**

Surprisingly, the $k_{Country}(t)$ values are essentially the same across the First World. The AIDS rates per hundred thousand are seven times higher in the United States than in the rest of the First World. A rather detailed argument is given in [16] to the effect that it is the United States AIDS epidemic, facilitated by inexpensive air travel, which renders the AIDS "epidemics" possible in the other First World countries possible. That is to say, without the AIDS epidemic in the United States, there would be none in Canada and Europe.

SIMEST and SIMDAT

We shall consider here two ways of simulating a pseudo-data set. The first, SIMDAT [9], [11], [12] is model free. It essentially samples an actual data point at random, examines the mean and covariance structure of the data in the neighborhood of the sampled point, and draws a sample from a normal distribution with this mean and covariance.

The front end of the second simulation algorithm, SIMEST, is purely model based. Given a model $\mathcal{M}(X|\Theta)$, assume given a data set of size n from a p-dimensional variable X, $\{X_i\}_{i=1}^n$. Assume that we have already rescaled our data set so that the marginal sample variances in each vector component are the same. For a given integer m, we can find, for each of the n data points, the $m-1$ nearest neighbors. These will be stored in an array of size $n \times (m-1)$.

SIMDAT

Suppose we wish to generate a pseudo-sample of size N. Note that there is no reason to suppose that n and N need be the same (as is the case generally with the bootstrap). To start the algorithm, we sample one of the n data points with probability $1/n$ (just as with the bootstrap). Starting with this point, we recall it and its $m-1$ nearest neighbors from memory, and compute the mean of the resulting set of points:

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i. \tag{3}$$

Next, we subtract from each of the data points the local mean $\bar{X}$, thus achieving zero averages of the transformed cloud:

$$\{X'_j\} = \{X_j - \bar{X}\}_{j=1}^{m}. \tag{4}$$

Although we go through the computations of sample means and coding about them here as though they were a part of the simulation process, the operation will be done once only, just as with the determination of the $m-1$ nearest neighbors of each data point. The $\{X'_j\}$ values as well as the $\bar{X}$ values will be stored in an array of dimension $n \times (m+1)$.

Next, we generate a random sample of size m from the one-dimensional uniform distribution:

$$U(\frac{1}{m} - \sqrt{\frac{3(m-1)}{m^2}}, \frac{1}{m} + \sqrt{\frac{3(m-1)}{m^2}}). \tag{5}$$

We now generate our centered pseudo-data point X', via

$$X' = \sum_{l=1}^{m} u_l X'_l. \tag{6}$$

Finally, we add back on $\bar{X}$ to obtain our pseudo-data point X:

$$X = X' + \bar{X}. \tag{7}$$

These, then, are the nuts and bolts of SIMDAT. As m and n get large, the SIMDAT procedure gives results very much like those one would expect when sampling from normal densities centered at each of the n nearest neighbor clouds.

SIMEST

Turning to SIMEST, we first observe that stochastic process modeling has not had the impact on science that one might have hoped. A major part of the reason for this fact is that, since the time of Poisson, we have axiomitized time indexed phenomena by such forwards statements as

> The probability that a metastasis will be generated in $[t, t+\Delta t]$ is proportional to the mass of the tumor.
> The probability a metastasis will be discovered in $[t, t+\Delta t]$ is proportional to the mass of the metastasis.

All very well, but we are in practise confronted with times of discovery of tumors and their sizes, and from this information wish to estimate the kinetic parameters of the tumor system. A metastasis discovered at a particular time could have been generated from the primary or from another metastasis at a variety of times. The writing down of the likelihood essentially requires a backwards argument involving all possible sources of the generation of the discovered metastasis. And this is almost always, nontractable in the extreme.

SIMEST enables us to assume a set of parameters and then simulate times and volumes of simulated metastasis discovery. The differences the simulated data and the actual data enables us ready measures of assessing the quality of our parametric assumptions.

The problem with the classical likelihood approach in the present context is that it is a backwards look from a data base generated in the forward direction. To scientists before the present generation of fast, cheap computers, the backwards approach was, essentially, unavoidable unless one avoided such problems (a popular way out of the dilemma). However, we need not be so restricted.

Once we realize the difficulty when one uses a backwards approach with a forwardly axiomitized system, a way out of our difficulty is indicated. We need to analyze the data using a forward formulation. The intuitively most obvious way to carry this out is to pick a guess for the underlying vector of parameters, put this guess in the micro-axiomitized model and simulate many times of appearance of secondary tumors. Then, we can compare the set of simulated quasi-data with that of the actual data. The greater the concordance, the better we will believe we have done in our guess for the underlying parameters. If we can quantitize this measure of concordance, then we will have a means for guiding us in our next guess. One such way to carry this out would be to order the secondary occurrences in the data set from smallest to largest and divide them

into k bins, each with the same proportion of the data. Then, we could note the proportions of quasi-data points in each of the bins. If the proportions observed for the quasi-data, corresponding to parameter value Θ, were denoted by $\{\pi_j(\Theta)\}_{j=1}^k$, then a Pearson goodness of fit statistic would be given by

$$\chi^2(\Theta) = \sum_{j=1}^{k} \frac{(\pi_j(\Theta) - \frac{1}{k})^2}{\pi_j(\Theta)}. \tag{8}$$

The minimization of $\chi^2(\Theta)$ provides us with a means of estimating Θ.

Typically, the sample size, n, of the data will be much less than N, the size of the simulated quasi-data. With mild regularity conditions, assuming there is only one local maximum of the likelihood function (which function we of course do not know), Θ_0, as $n \to \infty$, then as $N \to \infty$, as n becomes large and k increases in such a way that $\lim_{n\to\infty} k = \infty$ and $\lim_{n\to\infty} k/n = 0$, the minimum χ^2 estimator for Θ_0 will have an expected mean square error which approaches the expected mean square error of the maximum likelihood estimator. This is, obviously, quite a bonus. Essentially, we will be able to forfeit the possibility of knowing the likelihood function, and still obtain an estimator with asymptotic efficiency equal to that of the maximum likelihood estimator. The price to be paid is simply a computer swift enough and cheap enough to carry out a very great number, N, of simulations, say 10,000. This ability to use the computer to get us out of the "backwards trap" is a potent but, as yet seldom used, bonus of the computer age.

The SIMEST algorithm, developed by Thompson and his associates [1], [2], [10], [11], [12], [13] at Rice University and at the University of Texas M.D. Anderson Cancer Center, has enabled the development of deep models of cancer progression, which defy utilization by classical means. The algorithm has been utilized in an economic setting by Bridges, Ensor and Thompson [5].

There are, of course, many ways to modify the evaluation of the criterion function at various points in the parameter space so that the standard procedures of optimization theory can be utilized. Let us consider a modified Box-Hunter rotatable design ([15], pp. 238-245). To carry out this approach, we evaluate the criterion function at several points in the parameter space and fit a smooth parametric function in such a way as to minimize, say, the least squares fit of the parametric function to the pointwise evaluations of the criterion function. For example, we can approximate the goodness of fit between the data and the simulated data via the local quadratic model:

$$J(\Theta) = \beta_0 + \sum_{i=1}^{p} \beta_i \Theta_i + \sum_{i=1}^{p} \sum_{j=i}^{p} \beta_{ij} \Theta_i \Theta_j + \epsilon. \tag{9}$$

We shall assume that we are standing at the current best guess for that value of Θ, say Θ_n which gives the minimum value for the goodness of fit statistic. We have carried out several numerical experiments to evaluate χ^2 for Θ values around this current best guess. Having fit the coefficients to the data via least squares, i.e.,

$$\beta_{n+1} = (\Theta_n^T \Theta_n)^{-1} \Theta_n^T \chi_n^2, \tag{10}$$

we can move to our next best guess by taking the partial derivatives of $J(\Theta_n)$ and setting them equal to zero.

A means of selecting the experimental Θ points around the last best guess may be achieved by a variation of the rotatable design formulation (see [15]) we will first take our last fit and linearly transform the $\Theta's$ to obtain

$$J(\Theta) = \beta_0 + \sum_{i=1}^{p} \tilde{\beta}_i \tilde{\Theta}_i{}^2. \tag{11}$$

Then, we pick the design unit τ depending on

$$\|\Theta_{\text{opt},n+1} - \Theta_{\text{opt},n}\| = \sqrt{\frac{1}{p} \sum_{i=1}^{p} (\theta_{\text{opt},i,n+1} - \theta_{\text{opt},i,n})^2}. \tag{12}$$

Returning to the goodness of fit criterion of Karl Pearson, we have

$$S(\theta) = \sum_{j=1}^{k} \frac{(\hat{p}_{sj} - \hat{p}_{oj})^2}{\hat{p}_{oj}} \tag{13}$$

where p_{sj} is the proportion of simulated tumor discoveries and volumes falling into the jth bin, and p_{oj} is the proportion of actual tumor discoveries and volumes falling into the jth bin. In this case, the bins deal with two-dimensional data.

Considering systems where the dimensionality of the observed variables is greater than two, it is generally inefficient to use bins constructed according to Cartesian tiling, for then many, frequently most, bins will be empty. We require the construction of bins which are based on the real data set itself. With nearest neighbor tiling, consistency proofs become rather tedious. However, recently, Schwalb [8] has been able to show that for a very wide class of nearest neighbor based binning, SIMEST has the same convergence properties as would have been obtained if we actually knew the likelihood in closed form.

Combining SIMEST with SIMDAT

In many cases, it will be possible to employ a procedure using a criterion function. Such a procedure has proved quite successful in another context (see 275-280 of [15]). First, we transform the data $\{X_i\}_{i=1}^n$ by a linear transformation such that for the transformed data set $\{U_i\}_{i=1}^n$ the mean vector becomes zero and the covariance matrix becomes I.

$$U = AX + b. \tag{14}$$

Then, for the current best guess for Θ, we simulate a quasidata set of size N. Next, we apply the same transformation to the quasidata set $\{Y_j(\Theta)\}_{j=1}^N$, yielding $\{Z_j(\Theta)\}_{j=1}^N$. Assuming that both the actual data set and the simulated data set come from the same density, the likelihood ratio $\Lambda(\Theta)$ should increase as Θ gets closer to the value of Θ, say Θ_0, which gave rise to the actual data, where,

$$\Lambda(\Theta) = \frac{\prod_{i=1}^n \exp[-\frac{1}{2}(u_{1i}^2 + \ldots + u_{pi}^2)]}{\prod_{i=1}^N \exp[-\frac{1}{2}(z_{1i}^2 + \ldots + z_{pi}^2)]}. \tag{15}$$

As soon as we have a criterion function, we are able to develop an algorithm for estimating Θ_0. The closer Θ is to Θ_0, the smaller will $\Lambda(\Theta)$ tend to be.

The procedure above which uses a single Gaussian template will work well in many cases where the data has one distinguishable center and a falling off away from that center which is not too taily. However, there will be cases where we cannot quite get away with such a simple approach. For example, it is possible that a data set may have several distinguishable modes and/or exhibit very heavy tails. In such a case, we may be well advised to try a more local approach. Suppose that we pick one of the n data points at random—say x_1—and find the m nearest neighbors amongst the data. We then treat this m nearest neighbor cloud as if it came from a Gaussian distribution centered at the sample mean of the cloud and with covariance matrix estimated from the cloud. We transform these $m+1$ points to zero mean and identity covariance matrix, via

$$U = A_1 X + b_1. \tag{16}$$

Now, from our simulated set of N points, we find the $N(m+1)/n$ simulated points nearest to to the mean of the $m+1$ actual data points. This will give us an expression like

$$\Lambda_1(\Theta) = \frac{\prod_{i=1}^{m+1} \exp[-\frac{1}{2}(u_{1i}^2 + \ldots + u_{pi}^2)]}{\prod_{i=1}^{N(m+1)/n} \exp[-\frac{1}{2}(z_{1i}^2 + \ldots + z_{pi}^2)]}. \tag{17}$$

If we repeat this operation for each of the n data points, then we will have a set of local "likelihood ratios" $\{\Lambda_1, \Lambda_2, \ldots, \Lambda_n\}$. Then one natural measure of concordance of the simulated data with the actual data would be

$$\Lambda(\Theta) = \sum_{i=1}^{n} \log(\Lambda_i(\Theta)) \tag{18}$$

We note that this procedure is not equivalent to one based on density estimation, since the nearest neighbor ellipsoids are not disjoint. Nevertheless, we have a level playing field for each of the guesses for Θ and the resulting simulated data sets. Since computing is now close to free, we need to start sacrificing speed of algorithms for robustness and ease of use. In the case of SIMEST, that leads us to consider the following binning strategy

1. From the real data base of size n, select a point at random.
2. Find the smallest volume ellipsoid containing the point and its $m-1$ nearest neighbors.
3. Picking a vector parameter characterizing the system, say $\hat{\Theta}$ and, using this $\hat{\Theta}$, find a simulated sample of size N.
4. From the simulated data set, find the number—say M— of simulated points within this ellipsoid
5. The $|m/n - M/N|$ gives a measure of the concordance between the data and the pseudo-data simulated.
6. Repeat NN times to give a pooled measure of concordance.
7. This pooled measure then gives a means of using nonlinear optimization software to move toward a good guess for the true value of Θ.

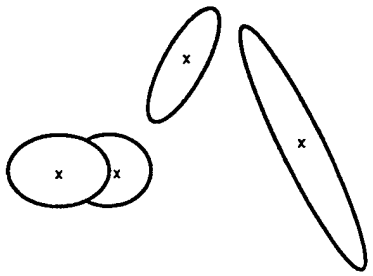


Figure 3. Random Binning for SIMEST.

Because of the fact that the bins so obtained in the above tiling frequently overlap, we need to modify the Schwalb consistency machinery to effect its applicability to the random binning case.

If one is seeking a stopping rule for terminating the estimation process, it is possible to use (18) on two different sets of "pseudo-data." We can use SIMDAT, say 5000 times and obtain 5000 Λ_{SD} values. Then, for a particular guess of Θ, we could compute 5000 $\Lambda(\Theta)$ values. Constructing a Wilcoxon-Mann-Whitney test on the two comparisons, the first set of "pseudorealities" dependent only on the data, the second, dependent only on the model and our parameter guess for Θ might give us a way to know when to stop changing our estimates of Θ. The advantages of such a procedure include robustness and ease of programming. The major disadvantage is computer intensity, which advantage is greatly ameliorated by the speed of current and future generations of digital computers.

ANALYSIS OF DATA IN DIMENSIONS HIGHER THAN THREE

The power of the modern digital computer enables us realistically to carry out analysis for data of higher dimensionality. Since the important introduction of Exploratory Data Analysis in the 1970s, a great deal of effort has been expended in creating computer algorithms for visual analysis of data. One major advantage of EDA, when compared to classical procedures, is a diminished dependency on assumptions of normality. However, for the higher dimensional situation, visualization has serious deficiencies, since it tends to involve projection into two or three dimensions.

What are typical structures for data in high dimensions? This is a question the answer to which is only very imperfectly understood at the present time. Some possible candidates are:

1. Gaussian-like structure in all dimensions.

2. High signal-to-noise ratio in only in one, two, or three dimensions, with only noise appearing in the others. Significant departures from Gaussianity

3. System of solar systems. That is, clusters of structure about modes of high density, with mostly empty space away from the local modes.

4. High signal to noise ratio along curved manifolds. Again the astronomical analogy is tempting, one appearance being similar to that of spiral nebulae.

For Structure 1, classical analytical tools are likely to prove sufficient.
For Structure 2, EDA techniques, including nonparametric function estimation and other nonparametric procedures will generally suffice. Since human beings manage to cope, more or less, using procedures which are no more than three or four dimensional, it might be tempting to assume that Structure 2 is somehow a natural universal rule. Such an assumption would be incredibly anthropomorphic, and we do not choose, at this juncture, to make it.
For Structure 3, the technique investigated by Thompson and his students [2], [6], [7] is the finding of modes, utilizing these as base camps for further investigation.
For Structure 4, very little successful work has been done. Yet, the presence of such phenomena as diverse in size spiral nebulae and DNA shows that such structures are naturally occurring. One way in which the astronomical analogy is deceptively simple is that astronomical problems are generally concerned with relatively low dimensionality. By the time we get past four dimensions, we really are in *terra incognita* insofar as the statistical literature is concerned. One hears a great deal about "the curse of dimensionality." The difficulty of dealing with higher dimensional non-Gaussian data is currently a reality. However, for higher-dimensional Gaussian data, knowledge of data in additional dimensions provides additional information. So may it also be for non-Gaussian data, did we but understand the underlying structure.

The main emphasis for the proposed research is concerned with Structure 3. The finding of modes is based on the Mean Update Algorithm [4], [6], [7], [12]:

Mean Update Algorithm
Let $\hat{\mu_1}$ be the initial guess
Let m be a fixed parameter;
$i = 1$;
Repeat until $\hat{\mu_{i+1}} = \hat{\mu_i}$;
Begin
Find the sample points $\{X_1, X_2, \ldots, X_m\}$ which are closest to μ_i;
Let $\hat{\mu_{i+1}} = \frac{1}{m} \sum_{j=1}^{m} X_j$;
$i = i + 1$;
end.

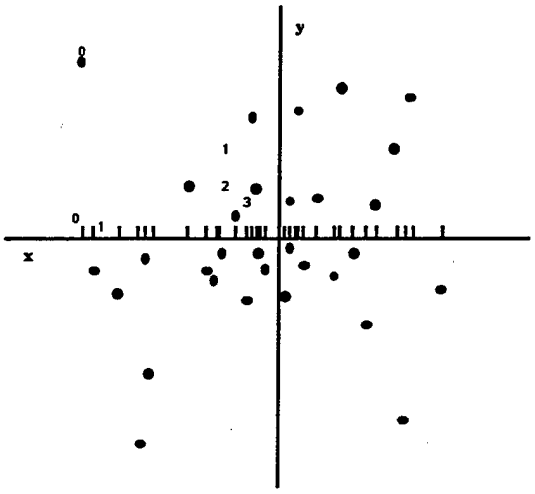


Figure 4. Mean Update Estimation of Mode.

Let us consider a sample from a bivariate distribution centered at (0,0). The human eye easily picks the (0,0) point as a promising candidate for the "location" of the distribution. Such a Gestaltic visualization analysis is not as usable in higher dimensions. We will be advocating such an automated technique as the Mean Update Algorithm. One hears about the "curse of dimensionality" a great deal. Let us examine Figure 4. Suppose that we have only one dimension of data. Starting at the projection of **0** on the x-axis. Let us find the two nearest neighbors on the x-axis. Taking the average of these, brings us to the **1** on the x-axis. And there the algorithm stalls, at quite a distance from the origin.

On the other hand, if we use the full two dimensional data, we note that the algorithm does not stall until point **3**, a good deal closer to the origin. So, increased dimensionality need not be a curse. Here, we note it to be a blessing.

Let us take this observation further. Suppose we are seeking the location of the minor mode in a data set which (unbeknownst to us) turns out to be

$$f(x) = .3\mathcal{N}(x; .051, I) + .7\mathcal{N}(x; 2.4471, I) \tag{19}$$

If we have a sample of size 100 from this density and use the Mean Update Algorithm, we can measure the effectiveness of the MUA with increasing dimensionality using the criterion function

$$\text{MSE}(\hat{\mu}) = \frac{1}{p}\sum_{j=1}^{p}(\hat{\mu}_j - \mu)^2 \tag{20}$$

Below, we consider numerical averaging over 25 simulations, each of size 100.

Table 1. Mean Square Errors.

p	m	MSE
1	20	.6371
3	20	.2856
5	20	.0735
10	20	.0612
15	20	.0520

We note how, as the dimensionality increases, essentially all of the 20 nearest neighbors come from the minor mode, approaching the idealized MSE of .05 as p goes to ∞. So far from being a curse, an increasing dimensionality can be an enormous blessing. We really have no very good insights yet as to the what happens in, say, 8-space. This examination of higher dimensional data is likely to be one of the big deals in statistical analysis for the next fifty years. If we force ourselves, as is currently fashionable, to deal with higher dimensional data by visualization techniques (and hence projections into 3-space) we pay an enormous price and, quite possibly, miss out on the benefits of high dimensional examination of data. Mean date algorithms seem to show great promise for exploratory purposes. Our multidimensional mode finding algorithms actually improve with increasing dimensionality in many situations.

ACKNOWLEDGEMENTS

This research was supported in part by the Army Research Office (Durham) under ARO-DAAH04-95-1-0665.

REFERENCES

[1] Atkinson, E.N., Brown, B.W., and Thompson, J.R. (1990). "Parallel algorithms for fixed seed simulation based parameter estimation." *Computing Science and Statistics: Interface 89.* K. Burke and L. Malone, eds.

[2] Atkinson, E.N., Bartoszyński, Robert, Brown, B.W., and Thompson, J.R. (1983). "Simulation techniques for parameter estimation in tumor related stochastic processes." *Proceedings of the 1983 Computer Simulation Conference.* New York: North Holland. 754-757.

[3] Bartoszyński, Robert, Brown, B.W., McBride, C.M., and Thompson, J.R. (1981). "Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process." *Annals of Statistics* 9: 1050-1060.

[4] Boswell, S.B. (1983). *Nonparametric Mode Estimation for Higher Dimensional Densities.* Doctoral Dissertation (Rice University).

[5] Bridges,Eileen, Ensor, K.B., and Thompson, J.R. (1992) "Simulation based estimation of market structure dynamics." (1992) *Decision Sciences*, pp. 467-477.

[6] Elliott, M.N. and Thompson, J.R.(1993), "The nonparametric estimation of probability densities in ballistics research." *Proceedings of the Twenty-Sixth Conference of the Design of Experiments in Army Research Development and Testing*, 309-326.

[7] Elliott, M, N. (1995). *An Automatic Algorithm For The Estimation of Mode Location and Numerosity in General Multidimensional Data*, doctoral dissertation, Rice University.

[8] Schwalb, O. (1999). *Practical and Effective Methods of Simulation Based Parameter Estimation for Multidimensional Data*, doctoral dissertation, Rice University.

[9] Taylor, M.S. and Thompson, J.R. (1986)."A data based algorithm for the generation of random vectors." *Computational Statistics and Data Analysis.* 4: 93-101.

[10] Thompson, J.R., Atkinson, E.N., and Brown, B.W. (1987). "SIMEST: an algorithm for simulation based estimation of parameters characterizing a stochastic process." *Cancer Modeling.* Eds. Thompson, J. and Brown, B. New York: Marcel Dekker. 387-415.

[11] Thompson, J.R. (1989) *Empirical Model Building.* New York: John Wiley and Sons, 242 pages.

[12] Thompson, J.R. and Tapia, R.A. (1990). *Nonparametric Function Estimation, Modeling and Simulation*, Philadelphia: Society for Industrial and Applied Mathematics, 304 pages.

[13] Thompson, J.R., Stivers, D.N., and Ensor, K.B. (1991). "SIMEST: a technique for model aggregation with considerations of chaos." in *Mathematical Population Dynamics*, Arino, Axelrod and Kimmel, eds., New York: Dekker, pp. 483-510.

[14] Thompson, J.R. (1999)*Simulation: A Modeler's Approach*, New York: John Wiley & Sons, 1999, 315 pages.

[15] Thompson, J.R. and Koronacki, J. (1993) *Statistical Process Control for Quality Improvement*, (with Jacek Koronacki), Chapman & Hall, 391 pages.

[16] Thompson, J.R. (1998) "Is the United States Country Zero for the First-World AIDS Epidemic?" Presented at the Fifth International Conference on Population Dynamics, Zakopane, Poland, June 21-25, 1998, ftp://ftp.stat.rice.edu/pub/thomp/zakopanepaper.pdf.

[17] Tukey, J.W. (1977). *Exploratory Data Analysis*, Reading, Mass.: Addison-Wesley, 503 pages.

[18] West, R. Webster and Thompson, James R. "Models for the Simple Epidemic" (1998) in *Mathematical Biosciences* 141:29-39.

SCENARIO ANALYSIS IN U. S. ARMY DECISION MAKING[1]

Richard R. Laferriere
U. S. Army TRADOC Analysis Center
White Sands Missile Range, NM 88002-5502

Stephen M. Robinson
Department of Industrial Engineering
University of Wisconsin-Madison
1513 University Avenue
Madison, WI 53706-1572, USA

ABSTRACT

This paper describes *scenario analysis,* a technique for modeling and optimizing decisions under uncertainty. The description is informal and is oriented toward working analysts rather than theoreticians. Its purpose is to convey a sense of the scope and some of the applications of the method, so that a potential user can determine if the method might be of benefit in a particular application. The paper includes a description of the successful implementation of this method by the U. S. Army TRADOC Analysis Center, White Sands Missile Range, NM, in a decision support system that has had extensive use in Army analysis.

INTRODUCTION

Scenario analysis is a method for optimizing under uncertainty, in which the possible "states of the world" are represented by a finite number of scenarios, each having a fixed, known probability. As we explain later, by appropriate reformulation this problem can be reduced to a large deterministic optimization problem, which can be solved by efficient implementations of linear programming (LP) such as are found in optimization modeling languages.

The technique uses basic ideas common in stochastic programming, including stages and nonanticipativity, which we explain below. Formal proposals for such methods appeared in the early 1990s: for example, the case of a convex problem, separable by scenarios except for nonanticipativity, appears in the work of Rockafellar and Wets [4]. Convex problems with more general constraints were considered by Robinson [3]. A decomposition approach to computational solution for very large problems was introduced by Chun and Robinson [1].

The basic reason for using a method like scenario analysis is that it allows us to compute hedged decisions. Under uncertainty the best overall decision might not be best for any individual scenario, so we have to look at all scenarios simultaneously to optimize. In many operational situations a hedged approach is essential for making good decisions. An example is military force design. A force may have to be employed in many different environments, and it must not fail in any of these. However, there may not be enough resources available to build a force that will perform brilliantly in each. Therefore we try to design a force that will do fairly well in all scenarios, as well as can be done with the resources at our disposal. This methodology also lets us make tradeoffs within the hedged decision paradigm, considering such questions as:

- How much performance can we afford?
- If we give up some lethality, can we gain survivability?
- What is the tradeoff between cost and survivability at a given level of lethality?

Analyis of this kind has numerous applications. We discuss military analysis below, emphasizing in particular the extensive use of scenario analysis by the Army. Many nonmilitary application areas have also been developed, including finance (multistage investment models with scenarios) and policy analysis. Extended versions can also be applied in quality improvement contexts such as design centering.

The following section explains the need for a method like scenario analysis to optimize in an uncertain environment. After that, we describe the mathematical technique, first giving a general overview of the

[1] Approved for public release; distribution is unlimited

method and then concentrating on the central issue of scenario bundling to satisfy nonanticipativity. We then discuss how the method has been successfully implemented by the U. S. Army TRADOC Analysis Center – White Sands Missile Range (TRAC-WSMR), and has been used in a number of Army studies. The final section contains a brief summary, and references follow.

DYNAMIC OPTIMIZATION UNDER UNCERTAINTY

We begin with a standard problem of optimization over time: we are given time periods $1, \ldots, T$ and in each time period we have to make certain decisions. Some measure of cost or benefit is given, and this is to be optimized with respect to the decisions available in the various periods. Of course, decisions taken in earlier periods may affect those available in later periods.

A familiar example of this kind of situation is a time-staged linear programming problem. The decisions at each stage are modeled by a linear program, and the constraints of these are (often loosely) connected by the interdependence of decisions in different time periods. This kind of problem is well understood, though not always easy to solve.

Now consider a variant of this situation, in which during the first time period any one of a finite number of different alternative situations may occur (each with a fixed, known probability). The portion of the optimization problem (linear program, for example) representing the first-stage decisions will be *different* for each of these alternatives. For each of the first-stage alternatives, there is then a finite set of alternatives that may happen in the second period, and so on. If we start at the beginning and go through a particular alternative at the first stage, a particular alternative at the second, and so on through all T stages, we obtain a single realization, or sample path, of the random process just described. Such a realization is called a *scenario.*

In this probabilistic model, each scenario is a time-staged optimization problem of the sort originally described. However, now there are many of these scenarios, and of course we do not know in advance which will occur. This introduction of multiple scenario situations immediately poses a problem: how are we to combine the performance measures of different scenarios? That is, how are we to deal with the fact that, for example, a certain set of decisions might perform very well against one group of scenarios, but poorly against another?

For the purpose of this paper we will assume that the use of expected performance (in the probabilistic sense) is satisfactory. That is, we will accept as a measure of performance the expected value, or average, of the performances against different scenarios, when the probabilities assigned to those scenarios are taken into account. This is not the only measure that could be used, but it is probably the simplest to deal with, and it fits current practice in many areas.

Note that this use of an expected value performance measure is not at all the same thing as the common use of "expected-value models" in which a single run, essentially deterministic, simulation is made in which stochastic elements are individually and systematically replaced by their expected values. That procedure is invalid as a method for modeling anything, since the outcome cannot be reliably related to the average of the outcomes under the individual scenarios, or to any other quantity of interest. Rather, the expected value performance measure that we are using corresponds to use of a Monte Carlo simulation process, but (as we shall see below) with a certain degree of increased structure.

A well known difficulty of Monte Carlo simulation is the large number of individual runs that need to be made in order to take into account adequately the variation in many different parameters. This problem of dimensionality is compounded if in the process of simulation we also wish to optimize, as we are assuming here. Even moderate numbers of variables and modest amounts of variation can then lead to enormous amounts of computing.

It is this problem of dimensionality that the technique of scenario analysis was designed to overcome. In essence, it does this by employing finite distributions (perhaps approximations or estimates of the actual distributions, if the latter are continuous), and by clever organization of the sources of variation so that the overall problem can either be solved directly using large-scale mathematical programming methods, or can be suitably decomposed into many smaller, independent problems. These can then be solved in parallel, or sequentially if a parallel machine is not available; the results of the individual solutions are then recombined according to certain rules to yield an approximate solution of the overall problem. This sequence of steps is repeated in an iterative process, until a solution of adequate quality has been found.

It should be clear from the above discussion that scenario analysis has the potential to contribute significantly in many applications in which Monte Carlo type analysis is desired but in which dimensionality is a problem. Because of this importance, we describe scenario analysis in its general form in the next two sections.

OVERVIEW OF SCENARIO ANALYSIS

This section establishes notation and explains the general procedure of scenario analysis. It serves as an introduction to the next section, which explains in more detail the concept of nonanticipativity and the consequent necessity for scenario bundling.

To begin with, we assume that the uncertainty in the model can be adequately described by a finite (possibly large) set of scenarios. These scenarios are to be understood as descriptions of the environment. They incorporate those things that cannot be changed by the decisions made in the course of the optimization, whereas the things that can be changed are modeled as part of the optimization problem.

Each scenario is understood to evolve over a fixed (finite) number of time periods. This number of time periods is the same for all scenarios. Within each time period certain decisions can be made by the actors. We index the scenarios by the letter s (running from 1 up to S), the time periods by the letter t (running from 1 up to T), and the decisions made in scenario s at time t by the vector x_{st}, whose dimensionality could depend on both s and t. The collection of vectors $x_{s1}, \ldots, x_{sT}$ will be denoted by x_s, and we interpret it as a larger vector. This is the complete sequence of decisions made in the single scenario s, for time periods 1 up to T.

Each scenario is given a fixed, positive probability p_s of occurrence, and these p_s sum to 1. We assume that they are fixed at the beginning of the analysis, and are known to the decision makers. An alternative interpretation of the p_s is as measures of the importance of each scenario to a decision maker, but we do not pursue that interpretation in this paper.

We also assume that once the decisions x_s have been made, there is a measure of overall cost or loss, given by $f(s, x_s)$; the first index s is used to indicate that the cost measure might well be different in different scenarios. Of course, one might as well use a measure of gain or merit if desired, and this would just be the negative of f. For reasonable results to be obtained with the method, these cost functions should be at least lower semicontinuous and convex in the variable x. These are not heavy requirements: in fact, in many cases of interest the functions f will be linear in x (as they are in the linear programming formulation implemented by TRAC-WSMR).

The overall expected cost, given the decisions x_s in each of the scenarios, will be

$$\sum_{s=1}^{S} p_s f(s, x_s),$$

representing the cost incurred in each scenario weighted by the probability that the scenario will occur. Therefore this overall cost is an expected, or average, cost given the decisions made. If we denote the collection of all scenario decisions x_s by the vector x, then we can write the overall cost as $f(x)$, where this is to be understood as the weighted sum just described.

Finally, the decision x_s to be made in scenario s is not completely arbitrary; we suppose that there is a closed, bounded convex set C_s in which x_s has to lie. Decisions x_s outside of this set are not allowable; decisions in it are called *feasible*. In the linear programming case, C_s would be represented by a finite collection of linear equations and linear inequalities.

With this background, we can describe the problem faced by the decision maker in the following way: choose the vector x (that is, the entire collection of decisions x_{st} for s running from 1 to S and t running from 1 to T), in such a way that x_s belongs to C_s for each s and, among all such feasible decisions, the value of $f(x)$ is least. In words: choose actions that are feasible and that yield the least expected cost among all possible feasible actions.

This description expresses quite well the object of the decision maker. However, it fails to deal with a critical problem that we can expect to face in real situations: namely, if we are making a decision in scenario s at time 1 we cannot expect to know all of the information that will be disclosed to us as the

scenario evolves through times 2,...,T. This imposes a complex constraint upon the decisions that we can make. We deal with that matter in the next section.

NONANTICIPATIVITY AND SCENARIO BUNDLING

The evolution of a scenario over time, and the decision maker's ignorance of what will happen in the future, bring about important constraints on freedom of action in choosing decisions. In the modeling, we express this information problem by allowing "branching" of the environment as time evolves.

For example, suppose that there are three time periods. Depending on random factors, during the first time period one of four different possible environments could occur; depending on which of these four occurred, we could have in the second time period respectively 3, 5, 2, or 6 different environments. That is, the first possible environment in time period 1 could branch in time period 2 into three possibilities, the second possible environment could branch into five, and so on.

This means that we are really dealing with 16 $(= 3+5+2+6)$ different scenarios. In time period 1 the first three of these are identical, as are the next 5, the next 2, and the final 6; in time period 2 all could be different.

Now, here is the difficulty: when the decision maker chooses an action at the beginning of time period 1, that action must not depend upon the (yet unknown) future evolution of the situation. Therefore, in the 16 scenarios just described, the actions chosen for time period 1 must all be identical. Similarly, the actions to be taken in time period 2 must be identical in the first three scenarios, in the next five, and so on, because at the beginning of time period 2 the decision maker knows which of the four possibilities has occurred in time period 1, but does not yet know what will happen in time period 2.

We can express this restriction in a simple way by saying that if two scenarios are identical up to time t, then the actions chosen for those two scenarios must also be identical up to time t. This is the so-called principle of nonanticipativity, and it expresses the logical fact that we cannot expect to use information now that will only become available to us in the future. Decisions that comply with this restriction are called *implementable;* this attribute is independent of the feasibility condition introduced earlier, and the decisions to be selected must comply with both restrictions.

This principle also introduces an enormous amount of complexity into the problem, since the need to enforce implementability leads to a complex system of constraints tying together decisions in different scenarios at the same time period. As we saw above, one scenario with some rather trivial branches in time led already to sixteen different scenarios, and it will be clear that many realistic problems may have large numbers of scenarios to be accounted for.

Fortunately, if the number of time stages is not too large, then with the aid of modern optimization tools such as modeling languages one can often model and solve such problems as large-scale linear programming (LP) problems by using a good commercial LP solver. This is the approach followed by TRAC-WSMR, using the GAMS modeling language with the CPLEX solver. For example, one early brigade model analyzed by TRAC had approximately 3,800 constraints (not counting bounds) and 3,300 variables; this problem is readily manageable with GAMS.

Problems too big for direct solution using a modeling language can be handled with decomposition methods. The classical method of this class is that of Dantzig and Wolfe, but more recently investigators have considered regularized Lagrangian decomposition methods. In these methods one dualizes with respect to the coupling constraints; the problem then splits into many small problems. These are solved independently (perhaps in parallel), and then the results are combined to yield an improved estimate of the dual variables. At optimality, one can recover the primal variables from regularization (bundle) parameters. See [1] for information on this approach, and numerous references.

IMPLEMENTATION AND USE BY TRAC-WSMR

The U. S. Army TRADOC Analysis Center, White Sands Missile Range (TRAC-WSMR), has implemented this methodology as part of a sophisticated analysis capability to develop optimum policies for design of forces and associated equipment based on specific Army requirements. For a description of some of the implementation considerations, see [2]. The capabilities of the TRAC-WSMR analysis methodology include

- Determining families of combat-effective systems
- Providing means to conduct comparative analyses of force structure
- Identifying resources needed to man, maintain, and staff the resulting forces

The methodology consists of embedding the basic mathematical technique of scenario analysis in a sophisticated decision support system. At the "front end" of the system, exploratory data analyses are run, using high-resolution combat models such as CASTFOREM. These analyses identify alternatives by scenario, provide means for grouping like-capability systems, and identify significant contributors to the force by scenario. They also provide data inputs to follow-on analysis. The scenario analysis optimization then provides families of combat effective systems, identifies competing high valued systems, and provides alternative families of systems and unit costs. Finally, a "back end" decision support system provides lists of alternatives based on combat capability, and facilitates presentation of results to decision makers through visualization devices such as Pareto (efficient frontier) diagrams based on combat effectiveness and cost. The result is a quick and flexible tool with which current decisions can be re-evaluated against new information (altered planning horizons, changes in priorities, new combat tactics, doctrine, systems, or new threat capabilities). New decisions can be suggested based on the changed information, and in some cases it has been possible to provide real-time response in face-to-face conferences with senior decision makers.

Studies in which this methodology has been used include

- Apache Procurement Strategy Analysis - 1990*
- An Armor Anti-Armor Mix Methodology – 1990
- Tank Fleet Mix Analysis – 1990
- Scenario Analysis for Combat Systems – 1992*
- Early Entry Analysis: Division Ready Brigade – 1993 *
- Guardian Task Force Fleet Mix Analysis – 1994
- Analysis of Amphibious Assault Fire Support Requirements –1995
- Antiarmor Resource Requirements Study – 1996 *
- Techniques for Increasing Efficiency and Accuracy of Data for Mix Analysis – 1998

The studies marked with asterisks in the above list won the Dr. Wilbur B. Payne Memorial Award for Excellence in Analysis, presented by the Deputy Under Secretary of the Army (Operations Research).

SUMMARY

This paper has described scenario analysis, a practical method for modeling costs and benefits of decisions in a time-phased, uncertain environment. We gave general motivation and developed the idea of probabilistic scenarios as a way of obtaining information about expected costs and benefits, then explained the scenario analysis method and described the computational problem posed by the requirement of implementability. Finally, we described the successful implementation of this method by the U. S. Army TRADOC Analysis Center, White Sands Missile Range, NM, in a decision support system that has had extensive use in Army analysis.

ACKNOWLEDGMENTS

This paper is based upon research sponsored by the US Army Research Office under Grant No. DAAG55-97-1-0324 and predecessor awards.

References

[1] Chun, B. J. and S. M. Robinson, "Scenario Analysis Via Bundle Decomposition," Annals of Operations Research, vol. 56, p. 39, 1995.

[2] Laferriere, R. R., "Scenario Analysis for Combat Systems," Report No. TRAC-WSMR-TR-92-023, U. S. Army TRADOC Analysis Command, White Sands Missile Range, NM, August 1992.

[3] Robinson, S. M., "Extended Scenario Analysis," Annals of Operations Research, vol. 31, p. 385, 1991.

[4] Rockafellar, R. T., and R. J-B Wets, "Scenarios and Policy Aggregation in Optimization Under Uncertainty," Mathematics of Operations Research, vol. 16, p. 119, 1991.

THE USE OF COGNITIVE PROCESSING ADAPTIVE TO DECISION MAKING IN THE JWARS PROJECT

Charles R. Leake, Ph. D.
US Army Concepts Analysis Agency Representative to the JWARS Office
1555 Wilson Blvd.
Arlington, VA 22209

Abstract

JWARS is a closed-form event-based constructive simulation of operational level warfare designed to address a range of analytical needs including operational and weapon system analysis. It is required to represent all elements in a joint operation covering a major regional contingency. It must be easily reconfigurable and run fast.

The JWARS prototype is being built to resolve the many technical issues associated with satisfying the JWARS requirements. The prototype is event-based and incorporates an event filtering position management scheme to minimize the computational overhead associated with object interactions. It is object-oriented to facilitate internal and external reuse. Object orientation helps satisfy the requirement to be easily reconfigurable and to operate at different levels of detail.

One of the goals of a model of this complexity is to contain within the structure some capability for self-organization. This goal has been somewhat demonstrated with the intelligence fusion process. Data is collected from sensors based upon an initial plan. Based upon the model's interpretation of this data a new collection plan is formulated and executed which in turn leads to a new collection plan, and so on and so forth. There are statistical methods used in this process. The statistics are used as part of a recursive process. This is however not a single self-organizing process, but is one that is being developed in terms of a perceived gestalt. Information is fed into this gestalt and compared with it. Based on these comparisons other decisions are made. As mentioned before this is done using a closed form object-oriented model.

The interesting aspect of this is that a theory of learning that contains a subset isomorphic to Piaget's is a possible fall-out of this prototype development. This theory has implications beyond the scope of the JWARS prototype. The explanation of this theory and how it relates to the JWARS prototype fusion process will be explained. In order to do this there will be a brief explanation of the JWARS prototype followed by a discussion of the fusion process and how it relates to the proposed theory of learning. The statistics used to determine such items as deciding an enemy's (COA) will be

provided. Finally given the theory of learning, how this theory fits into self-organizing adaptive decision making as well as the implications of the theory beyond the prototype will be presented.

1. CONCEPT OF A CLOSED-FORM EVENT-BASED SIMULATION

There are 2 main categories of war fighting models [Bracken, et. al, 1995]. One is a simulation and the other is a wargame. A simulation is a model run on a computer that simulates warfighting. It is also closed-form in the sense that after the initial inputs, the user of the model allows the model to run to completion without human intervention. Although there are some simulations that allow human intervention, these are in the exception. There are several categories of simulations. They could be either stochastic or deterministic or is as usually the case contain a mix of both methods.

Assuming that a model fits one of those three categories, a model could then be time stepped or event-based. A time-stepped model requires that the model is brought to state at specified time intervals and calculations of events occurring during or before the time step period are made. Then the model moves forward to the next time step and proceeds forward in the same manner until it reaches its predetermined stopping point.

An event-based model uses a queue in which to schedule events. The concept is not new, but the implementation is enhanced by the use of object-oriented technology (OOT) (See paragraph 3) which provides a new paradigm to model developers. An object schedules an event based on information that is either internally based or passed to it from an other object in the form of a message. The scheduling of events is filtered using a position manager that reduces the computations that would be necessary from a purely geometrically induced interaction. The event manager passes information to the objects which are involved. Depending upon what each object's concerns are the object makes a decision as to whether or not the event is or of interest to them. If it is an event that concerns the object, the object then sends a message to the event manager to schedule an event at some future time. The event manager then enters the event in the event queue and determines its order in the queue. This requires the conversion of a partially ordered set into a totally ordered one. The purpose of this method is to reduce the n^2 computations that would be required without such a method.

A wargame is usually played in front of a map or more recently a battery of computers and humans conduct the warfighting activities. Most wargames are supported by computers for the calculations which are made for the adjudication of events that occur during the play of the game.

JWARS as mentioned in the abstract is a closed-form event-based simulation. This implies that the user of the model will be allowed to input into the model at the beginning of the run. Any subsequent changes to the inputs will require that the model go back to start before the model will run.

2. OBJECT-ORIENTED TECHNOLOGY (OOT)

OOT is not a new concept [Taylor, 1991; Coad & Yourdan, 1992], but its use has remained dormant until recently. What it entails is the creating of object classes which

can be instantiated and used as little programs within a structure containing other objects [Rumbaugh, 1990; Martin & Odell, 1992, Tkach & Puttick, 1994]. The interaction effects between the objects occur with messages being passed between objects requesting that the receiving object react to the message. Messages bare a similarity to call statements [Lewis, 1995; Derr, 1995].
The JWARS prototype was an attempt to use this new technology with which to design a warfighting simulation.

3. THE CONCEPT OF SELF ORGANIZATION

The concept of self-organization is a relatively new concept that was made popular through the works of Gleick and Mandelbrot [Mandelbrot, 1983; Gleick, 1987]. The initial concept was mathematical using the concept of self-similarity as in the Sierpinski Triangle. Barnsley and Devaney, [Barnsley, 1998; Devaney, 1992] in their works on fractals elaborated on the mathematics associated with fractals providing us with a better understanding of what is the mathematics of fractals. Such concepts as iterations, orbits, and attractors are clearly defined mathematically. However, avoiding the strict mathematical interpretation, one might also consider the possibility of expanding the concept for use in software development.

As will be used in this paper self organization will be defined in terms of a events that are similar in the mathematical sense of requiring only a scaling factor to go from one state to another. This implies a set of code that once initialized, based upon these inputs into the algorithm, analyzes the results of running the code, and generates the requirements for the inputs into the next run of the code. This process continues and tends toward an "attractor" in the sense of Mandlebrot. It is this "attractor" which can be used for decision-making purposes.

Decision-making has relied a great deal on the set of beliefs which were and still are very popular in model development. Three which come to mind are:

1. Nature obeys the Law of Least Squares.
2. Given an event there exists a set of n statements such that the addition to these statements of k statements will not appreciably alter the information provided by the n statements.
3. Given an event there exists a probability distribution which can be used to predict the event.

To these three beliefs we might add: There are events which are self-organizing. These events tend to an "attractor" or chaos. This "attractor" can be an orbit. This belief was what was substituted in the JWARS prototype instead of any of the three previously mentioned beliefs.

4. COGNITIVE PROCESSES AS SELF-ORGANIZING

Cognitive processes are related to perception. Certain types of perception are self-organizing and others apparently are not. Piaget [Piaget, 1963; Flavell, 1963] provides us with an excellent insight into this aspect of perception. For example, Piaget's four stages of cognitive development introduce us into the concept of self-organization. With the last two, operational and formal stages one might observe what one would call a self-organizing process. This is one that enables one at the formal stage to interpret any similarity to a learned set of perceptions and respond to these perceptions by assimilating them and acting upon them in what might be called a "mature" manner. At other stages such as the sensori-motor and pre-operational stages this perception-based process is not achieving a "mature" interpretation of its perceptions. However, there are also self-organizing aspects of this behavior as well. Piaget and Flavell in [Piaget, 1963; Flavell, 1963] provide examples of the self-organizing properties of the "immature" behavior. This behavior is recursive until it reaches a bifurcation point and then moves on to the next level of maturity. Piaget interprets the stages of perception occurring in relation to the development of cognitive structures, as one would relate assimilation in an amoebae in accordance with the available structures with which to assimilate entities which the amoebae encounters. Whether or not this accurately portrays what he observed is a matter of conjecture, but his observations have been experientially verified.

An alternate way of explaining the same processes in the thought pattern of adults would be [Leake, 1997]:

1. Fumbling Stage
2. Pieces Stage
3. Groups of Pieces Stage
4. Aha! Stage.

These stages are isomorphic to Piaget's four stages of cognitive developmental psychology. Moreover, if one considers the approach taken to solve a jigsaw puzzle, they are also experientially verified. When these stages are related to a goal, perception of the goal becomes apparent at the last stage. Processes 1 through 4 are part of an iterative process that takes place internally. Gradually as the process iterates, the process tends toward either an "orbit" or an "attractor." It is this process which JWARS attempted to simulate and relate to decision making processes. This process is a new one that subsumes our present belief that: Given an even there is a probability distribution which explains that event. Presently we would use the probability distribution with which to predict the event under consideration [Leake, 1996, 1997].

5. SIMULATION OF REASONING PROCESSES IN JWARS

One of the processes simulated in JWARS that take place in a command center is that of fusion. The process called fusion in JWARS represents the bringing together of

the data which has been collected. Then comparing it with previously collected data to try to make sense out of it. This is done through the use of mathematics to determine how well data matches known items or events. Fusion is a process that begins with intelligence processing of the battlefield (IPB). This IPB is the commander's gestalt. In this IPB is a determination of an enemy course of action (COA). Decisions must be made by the commander as to how to deploy his forces to meet this perceived enemy course of action.

In simulating fusion, which is referred to as a generation based problem solving [Antony, 1995], several key factors come to mind. They are:

1. Target tracking;
2. Target Classification; and
3. Path Planning.

Each of these factors is a part of the recursive process used in JWARS. Target tracking is done through sensor tracking and reports. Target classification is done during the analysis of sensor reports by comparing them to order of battle matrices of known enemy units. Finally path planning is done through the initial IPB and continual updating of the situation map (SITMAP). Additionally there is a collection management plan developed to direct and control the activities of the sensors which are a limited resource. This is all coordinated in the fusion process in JWARS.

Usually there are several possible COA's and the commander must rely on his perception of the battlefield as to which of these COA's is the best given the situation. In order to assist his perception the commander attempts to gain intelligence of the enemy's intentions through the use of various sensing devices including visual observations from aircraft and soldiers on the ground. The commander however does not have an unlimited set of these sensors so he must allocate them where he thinks they will provide him with the information that he is seeking. This is called the collection plan.

The sensor allocation is made in accordance with the collection plan and the information provided by the sensings is analyzed to determine which of the COA's appears most likely. In the beginning, one can only fumble (Stage 1). As the process continues, pieces of the puzzle seem to fit together (Stage 2). These pieces grow larger (Stage 3). Ultimately, it becomes apparent what the enemy's COA is in the perception of the friendly force commander. This information is then correlated with the commander's own situation.

This recursive process is a self-organizing process where one step leads to the next and iterates itself based upon its previous perceptions. Based upon the perceived COA ("attractor"), the commander makes a decision as how to deploy his forces. This is the process simulated in JWARS. So far in over 10 demonstrations of the prototype, the correct COA has been determined from 3 possible COA's indicating that the model is able to conduct fusion based upon initial inputs.

This means that the model must determine its next collection plan based upon an analysis of the results of the initial collection plan. Next it must then conduct the next iteration based upon its analysis and continue this process until a COA develops which can be used to make decisions. These decisions are also made internally in the model based upon the perceived COA.

The mathematics used in this process are quite simple. To characterize an enemy unit sensor reports are correlated with known enemy order of battle matrices using the Pearson Product Moment Coefficient of correlation. Based on the resultant coefficient possible characterizations are categorized into three classes: detected, recognized, and identified. In the event that the best characterization is either detected or recognized, those units in that category are posted as an unknown on the SITMAP. Their location is provided to the collection management process for possible inclusion in the next collection plan. Those that are identified are provided to the air tasking order (ATO) generator for possible targeting.

Following the correlation process, the SITMAP is consulted to determine how many of the units predicted to be located in time and space in the IPB for each predicted COA. The numbers derived from each of these comparisons is compared against what would occur by chance using a z-score. The highest z-score is then the predicted COA. These computations are repeated through several time periods. Ultimately as one positive z-score begins to dominate the others, the COA related to that z-score is determined as the enemy COA and force dispositions and other related activities are then coordinated in accordance with the enemy COA.

Is this process guaranteed to produce the correct COA? History shows us that this is not always the case and it is possible to perceive the wrong COA such as happened to Admiral Halsey at Leyte Gulf in World War II. However, decisions are made based upon the perceptions. It remains to be seen in the next iteration of the JWARS development process whether or not the model will create wrong or less desirable decisions based upon an incorrectly perceived COA.

6. CONCLUSIONS

The experiment in JWARS with the fusion process simulation has led to an interesting result which offers a new way of looking at recursive processes. Recursive processes such as the fusion process in JWARS can be used a tool with which to make predictions and hence base decisions such as how to simulate the commander's deployment of his forces upon. In addition it offers the opportunity to test out the concept of self-organizing process that might enable the modeler to develop a process that can scale up or down to units of varying sizes and composition. The JWARS prototype used this process for fusion, but the concept of self-organization has other applications such as for communication [Leland, et. Al, 1993].

7. INQUIRIES

If you have any questions about the preparation or submission of your paper, please contact:

Dr. Charles R. Leake
Tel: (703)-602-2918
Fax: (703)-602-3388
email: cleake@nova.umuc.edu
charlie.leake@osd.pentagon.mil

ACKNOWLEDGEMENTS

I wish to express my thanks to LTC Terry Prosser for the insights and guidance which he provided me on how an Intelligence Unit conducts its operations.

REFERENCES

Antony, R.T, Principles of Data Fusion Automation, Artech House, Inc., Norwood, MA, 1995.

Barnsley, M., Fractals Everywhere, Academic Press, Inc., San Diego, CA, 1988]

Booch, B., Object Solutions Managing the Object-Oriented Project, Addison-Wesley Publishing Co., Inc., Reading, MA., 1996.

Bracken, J., et. al., Warfare Modeling, John Wiley & Sons, Inc., 1995

Coad, P. & Yourdan, E., Object-Oriented Analysis, Yourdan Press, Inc., 1992.

Devaney, R.L., A First Course In Chaotic Dynamical Systems, Addison-Wesley Publishing Co., Inc., 1992.

Derr, K.W., Applying OMT A Practical Step-by-Step Guide to Using the Object-Modeling Techniques, SIGS Books, New York, NY, 1995.

B. Eckel, B., Thinking in C ++, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1995.

Flavell, J. H, The Developmental Psychology of Jean Piaget, D. Van Nostrand, Co., Inc., Princeton, NJ, 1965.

Gleick, J., Chaos: Making a New Science, Viking Press, Inc., New York, NY, 1987.

Leake, C.R., unpublished presentation to the Washington Evolutionary Society, Washington, D.C.,1996.

Leake, C.R., unpublished presentation to the Virginia Education Research Association, Richmond, VA, 1997.

Leland, et. al., "On the Self-Similar Nature of Ethernet Traffic", presented at ACM SIGComm '93, San Francisco, CA, September, 1993.

Lewis, S., The Art and Science Of SMALLTALK, Prentice Hall, New York, NY, 1995.

Mandelbrot, B., The Fractal Geometry of Nature, W. H. Freeman, Inc., New York, NY, 1983.

Martin, J. & Odell, J.J., Object-Oriented Analysis and Design, Prentice Hall, Englewood Cliffs, NJ, 1992.

Piaget, J., Psychology of Intelligence, Littlefield Adams & Co., Patterson, NJ, 1963.

J. Rumbaugh, *Object-Oriented Modeling and Design*, Prentice Hall, Englewood Cliffs, NJ, 1990.

Taylor, D.A., Object-Oriented Technology, A Manager's Guide, Addison-Wesley Publishing Co., Inc., Reading, Ma, 1991.

Tkach, D.& Puttick, R., Object Technology in Application Development, Benjamin Cummings Publishing Co., Redwood City, CA, 1994.

Voss, G., Object-Oriented Programming: An Introduction, McGraw-Hill, New York, NY, 1991.

Rapid Force Projection Initiative Advanced Concept Technology Demonstration (RFPI ACTD) – The Experimental Path

Paul J. Deason, Ph.D.
USA TRAC-WSMR
(505) 678-1610 DSN 258-1610 deasonp@trac.wsmr.army.mil

Greg B. Tackett
RFPI Project Managers Office, Redstone Arsenal, AL
(256) 842-9501, DSN 788-9501 gtackett@redstone.army.mil

Advanced Concept Technology Programs

The formal acquisition process, as directed by DoD Instructions 5000.1 and 5000.2, is the primary mechanism for the procurement of new systems and the introduction of new capabilities via new or upgraded systems. The Advanced Concept Technology Development (ACTD) Programs process is a pre-acquisition stage providing an important mechanism for the warfighter to evaluate proposed solutions to important military needs. ACTDs exploit mature advanced technologies to solve important military problems, to rapidly transition technology from the developer to the user. ACTDs are structured to address the needs of the warfighter, to provide needed capabilities, address deficiencies, and reduce costs and manpower requirements Each ACTD is aimed at one or more warfighting objectives, and is reviewed by the Services, Defense Agencies, and the Joint Staff. The focus of the ACTD is to react in response to critical military needs. This requires intense user involvement through the specification of the identified military needs. The ACTD program reacts to these stated needs and, through a process that exploits mature technologies places new equipment into the hands of the user, the warfighter. The warfighter may then conduct realistic and extensive exercises to evaluate systems utility and gain operational experience. The user is then the basis for evaluating and refining operational capabilities of these advanced technology systems, understanding their utility, before an acquisition decision is made based upon the systems potential or projected effectiveness.

When the user/warfighter conducts demonstrations of the capability of the advanced concept systems, the user defines measures of effectiveness and measures of performance (MOEs/MOPs) for the systems. He provides or approves the planned operational exercises, demonstrates and develops concepts, to include the new concepts of operation, tactics, and doctrine. The ACTD provides the means to develop, refine, and optimize new warfighting concepts, and to prepare the systems and the units for the transition into acquisition. To facilitate the refinement of requirements, the ACTD must provide a residual capability to the user to further refine the concept of operations and permit continued use, to include combat, or define additional needed capability prior to formal acquisition. The purchase of additional capability beyond the residuals provided by the ACTD, where appropriate, is accomplished through a formal acquisition program.

Each ACTD is managed by a lead Service or Agency developer, driven by the principal user sponsor, usually a Unified Commander. User and development organizations are represented on an oversight panel chaired by the Deputy Under Secretary of Defense (Advanced Technology) (DUSD (AT)), who defines guidelines and provides oversight, support, and evaluation. Final review is provided by Deputy Under Secretary of Defense (Acquisition and Technology) (DUSD (A&T)) and the Vice-Chairman, Joint Chiefs of Staff. The Office of the Army Deputy Chief of Staff for Research, Development, and Acquisition (DCSRDA) has been restructured to support ACTDs. Funding for the ACTD is typically provided from participating technology programs supplemented as needed from the DUSD (AT) ACTD funding line. Funding provides for systems integration, for multiple copies of system elements if needed by the user in his evaluation process, and for technical support of the residual capability for two years beyond the completion of the ACTD.

The following agencies make up the Rapid Force Projection Initiative (RFPI) ACTD User/Developer Integration community. These groups are responsible to the Secretary of Defense and to Congress for the oversight of the program, and for the contributions in the systems design, systems integration, systems assessment and evaluation. The Joint ACTD Management Group provides the operational level management interface between the oversight at the DoD and DA level and the materiel developers, the warfighters, and the analysts. This management group is comprised of the RFPI Technical Program Management Office, Redstone Arsenal, AL, representing the Army Materiel Command; the Commander's Office, XVIII Airborne Corps, Ft. Bragg, NC representing FORSCOM, and the Dismounted Battlespace Battle Laboratory, Ft. Benning, GA representing TRADOC. For the purpose of analysis, evaluation and assessment, the Army Test and Evaluation agencies AMSAA, OPTEC, TECOM, joined the TRADOC representative TRAC-WSMR, and make up an Analysis Steering Committee

Rapid Force Projection Initiative (RFPI)

The RFPI ACTD is the largest ACTD yet, with an operating budget approaching $800M. The RFPI ACTD provides a new capability for the Army and a model for future Early Entry Forces. RFPI is a sensor-weapons-C4I concept that allows light forces to fight the majority of the battle out of contact using non-line-of-sight killers. The use of U.S. forces in Contingency Operations requires the ability to respond quickly to unanticipated challenges to our interests around the globe.

The RFPI ACTD demonstrates a highly lethal, survivable, and airlift constrained enhanced power projection capability through the development and evaluation of new technologies and tactics for early entry forces. The lift-constrained environment requires the new forces to be as deployable as current forces. The Rapid Force Projection Initiative also demonstrates real time targeting from forward sensors to standoff killer weapon systems with the capability to engage high value targets, including heavy armor, beyond traditional direct fire range. Target transfer is facilitated by tactical digital data transfer systems being developed as part of the U.S. Army Battle Command System (ABCS). This synchronization of dispersed forces results in increased force lethality and survivability. This ACTD also provides a tool for further exploration of emerging warfighting concepts and doctrine.

An integral element of the RFPI ACTD is the provision of developmental ACTD items to the participating warfighting unit as a residual operating capability, with suitable technical support for at least two years. The unit which participated in the RFPI Field Experiment, conducted in summer, 1998 was the 2nd Brigade/101st Airborne Division (Air Assault), which is in turn a part of the XVIII Airborne Corps. This unit is still designated as the recipient unit of the RFPI equipment for the two-year residual period, although the possibility exists for the transference of the equipment to another unit of the XVIII Corps.

The RFPI System of Systems is designed to supplement or replace systems currently in the inventory of the Experimental Brigade (the Brigade Modified Table of Organization and Equipment. (MTOE)) See Table 1 for a listing of the brigade's weapon systems, as well as a listing of the RFPI ACTD program systems. The systems, designated the Residual or Leave-behind systems, are comprised of a variety of near-acquisition and brass-board systems that can be used today by the warfighter. Scheduled for initial unit production in the 2001-2003 timeframe, they are at a level of development sufficient for them to be retained by the experimental unit for the two-year residual period following the August 1998 Ft. Benning Field Experiment. The unit will train with them, and deploy with them in case of a warfighting need. The unit also can specify which of the Residual systems it chooses to retain, which needs further work and development to meet the units operational tempo and parameters adequately, and which systems it chooses not to retain at all. The advanced concept systems are more conceptual, with an expected initial production in the 2006-2007 timeframe. These could only be evaluated through the means of interactive and constructive simulation, and were not ready to go into the hands of the warfighter. As shown below, these Technology Demonstration and Advanced Technology Demonstration (TD and ATD) systems are designed to supplement or replace systems currently in the unit's MTOE, or replace the residual systems. Again, the key is to improve the lethality, improve the survivability of the light unit while maintaining the same logistics burden.

Utility and the further exploration of emerging warfighting doctrine is being accomplished through a series of TRADOC sponsored Battle Lab Warfighting Demonstrations and Experiments. The RFPI ACTD builds on ongoing activities of the RFPI Technology Program (TP) and its supporting ATDs and TDs. Refer to Figure 1, which shows the process of taking the predictive performance of the systems from the TD and ATD program managers, and the developers, from limited field tests and limited system of system examinations. These are then focused through the lens of simulation in preparation for the delivery of the residual systems to the experimental unit for retention, and as the process continues on through the acquisition process.

RFPI ACTD System of Systems

	Hunters	Battle Command (C4I)	Standoff Killers
Current Systems	OH-58D, GBS, IREMBASS, AH-64D/LB, JTUAV, Predator,TPQ36/37 COLT, TACP, LRSD, RECON, FO/FIST, GBSC, etc.	ATTCS, other BOS C2 systems, SINCGARS	AH-64C/D, TOW-ITAS, Javelin, 105 HOW, 155 HOW, 60mm & 81mm Mortar, MANPADS, Avenger, 155 SADARM, ER MLRS
Residuals/Leave behind	Hunter Sensor Suite, Remote Sentry, FO/FAC, IAS (HE & ADAS)	Light Digital TOC, Distributed Automated C2, Digital Commo	HIMARS, EFOGM, 155mm AutoHOW
Advanced Concepts/ Systems Programs	Tech Demos: LOSAT, PGMM, RAPTOR IMF, Guided AIS, ASSI Tech Programs: Smart 105mm TGP, LCCM, LOCAAS Acquisition Programs: MSTAR, AVTOC, ER 155mm Comanche, UGV, PI SADARM, FOTT, 155mm LW LRAS3, ATACMS II (not used)		

Table 1. The RFPI ACTD System of Systems

The following is a list of the RFPI Residual/Leave Behind Systems:

Hunters (Advanced Sensors):
Hunter Sensor Suite, Remote Sentry, Forward Observer/Forward Aerial Controller,
Integrated Acoustic System
Battle Command C4I:
Light Digital Tactical Operating Center (LDTOC)
Standoff Killers/ Munitions:
High Mobility Agility Rocket System (HIMARS), Extended Fiber Optic Guided Munition (EFOGM),
155mm Howitzer Automatic Fire Control System (AFCS)

The following is a list of the RFPI Advanced Concepts Systems:

ATDs and TDs
Hunters (Advanced Sensors):
Aerial Scout Sensor Integration

Standoff Killers/Munitions:
120mm Mortar Fire Control System (MFCS) & Precision Guided Mortar Munition (PGMM), Improved Mine Field (IMF), Line-of-Sight Anti-tank (LOSAT), Autonomous Intelligent Submunition (Multiple Launch Rocket System (MLRS) Smart Tactical Rocket (MSTAR) Candidate)

Other Science and Technology Programs
Standoff Killers/Munitions:
Smart 105mm Munition (Terminally Guided Projectile; TGP), Low Cost Competent Munition (LCCM), Low Cost Autonomous Attack System (LOCAAS) (MSTAR candidate)

Advanced Concepts – Acquisition Programs (These are materiel programs which were advanced in the acquisition process independently of the ACTD program, but were included because of the potential for accelerating the systems acquisition timelines, thereby producing the systems for the warfighter much sooner.)
Hunters (Advanced Sensors):
Unmanned Ground Vehicle (UGV), Comanche helicopter
Battle Command C4I:
Aviation Tactical Operations Center (AVTOC)
Standoff Killers/Munitions:
Follow-on to TOW (FOTT), 155mm Lightweight Automatic Howitzer (ATCAS); 155mm Extended Range munition, Search and Destroy Armor Pre-planned Product Improvement (SADARM P3I), Guided MLRS, BAT P3I (MSTAR candidate) ATACMs Blk III

RFPI Milestones

What follows is a listing of the early RFPI events and milestones which served to define what was to be expected in the relationship of the selected TD and ATD systems, particularly keying on showcasing the EFOGM system and the digital communications and advanced sensors which would provide targeting information. The EFOGM system, along with the additional artillery systems, the HIMARS and 155mm ATCAS Howitzer answers the warfighter's interest in the RFPI ACTD to provide precision armor defeating capability beyond the infantry systems' direct fire range, extending the infantry brigade's area of influence to 60km to 100km, or into the divisional area of responsibility. This is due to findings from Desert Storm regarding the vulnerability of the airborne and air assault units defending against an armored attack. The exercise JRTC 94-02 (OOTW) (Operations Other than War) Nov 93 (before the initiation of the ACTD) showed the potential of the application of EFOGM and digital communications technology to the light force. The Infantry Commanders Conference, May 94 was a showcase to the collective Army leadership of the potential of the RFPI Hunter Standoff Killer concept. The Redstone Arsenal Early Version Demonstration, Sep/Oct 94 was the first opportunity to show the Army leadership the potential of the live/virtual representation of the battlefield in the Distributed Interactive Simulation (DIS) environment. This was done using the Battlefield Environment Weapons Systems Simulation (BEWSS) and other simulations and display systems in Redstone Arsenal Building 5400. Warrior Focus JRTC 96-02 November 95 offered another experiment regarding the potential of digital communications and targeting for the Army. The RFPI Initial Systems Mix approved by the U.S. Army Senior Advisory Group in December 94. The final U.S. Army approval of the RFPI Management Plan in occurred in March 95. The 101st ABN (AASLT) was designated by FORSCOM as the ACTD Experiment Force in July 95 for the purpose of pre-field experiment preparation. As far as experimentation is concerned, as differentiated from demonstrations were the April 94 Anti-armor Advanced Technology Demonstration (A2ATD) Battlefield Distributed Simulation -Developmental (BDS-D), which used the BEWSS constructive simulation to first investigate the RFPI system of systems concept in a modified version of the Caribbean TRADOC scenario HRS 33.7. A major BDS-D demonstration was the RFPI Experiment 6, scheduled for June 95 and finally finished in Jan 96.TRAC-WSMR was asked for a CASTFOREM scenario to derive the ModSAF scenario, which matched the BEWSS 33.7 scenario being used by RFPI. Due to new BEWSS DIS capability, BEWSS was linked to ModSAF and the CASTFOREM scenario was not necessary except for the purpose of beginning the process of implementing RFPI representations into CASTFOREM for use in what-if analyses. The scenario in CASTFOREM was dropped in Apr. 96. The need for TRAC-WSMR's involvement in the RFPI assessment process increased dramatically in 1996, with DUSA (OR) blessing. This was to investigate the

performance of combinations of the RFPI residual and advanced concept systems in CASTFOREM and Janus, and to present to DUSA (OR) using the 1999 operational capabilities of a Division Ready Brigade (DRB) of the 82nd Airborne, and a DRB of the 101st Air Assault operating in approved TRADOC scenarios in Southwest Asia and Northeast Asia. The purpose was to indicate which of the residual and advanced concept systems have potential in adding capability to the warfighter, and so therefore should be continued in acquisition consideration. These results were presented in October 97.

Finally, as the climax to the RFPI experimentation was a large scale, free play live/virtual field experiment conducted July-August 1998 at Ft. Benning, GA to support the evaluation of the value added by inserting these new technologies into the force structure of an existing unit. This was also to examine the unit as it developed tactics, techniques, and procedures (TTPs) based on the Ft. Benning-developed operational techniques as to how they would train and fight the RFPI system of systems as they entered into their MTOE as residuals. The developmental equipment was delivered to the artillery units of the XVIII Corps and the 2nd Brigade/ 101st AASLT for training starting in November 1997, and will be retained for two years after the field experiment. This is dependent on the ability of the developers to deliver and maintain the systems, and the unit's interest in retaining the equipment in their go-to-war inventories.

Leading up to this climax live-virtual field experiment were a large number of battle lab warfighting experiments (BLWE), 2nd/101st command post exercises (CPX) and field training exercises (FTX). These alternated with constructive simulation and interactive simulation experiments. These simulation experiments had the results of refining the TTPs and operational concepts for Ft. Benning DBBL and the experimental unit, to investigate the placement of data collection instrumentation and battle flow by representing the field experiment on a digital representation of Ft. Benning. Also, the constructive and interactive simulations were used to prepare the interactive simulation scenarios used to drive the live/virtual DIS field experiment. This back-and-forth between simulations and experiments is portrayed in Figure 2 representing the Model Experiment Model process. The final analytic products of the RFPI ACTD will be a set of five assessments: the OPTEC Assessment; the FORSCOM User Assessment; the TRADOC User Assessment; the Engineering Assessment comprised of the reports from the TD and ATDs, accumulated by the RFPI PMO; and the TRAC-WSMR Performance Assessment based on operational analysis events and constructive simulation. These analytic products, which are due in 4Q FY00, will provide the final and definitive measure of the RFPI Hunter Standoff Killer (HSOK) effectiveness.

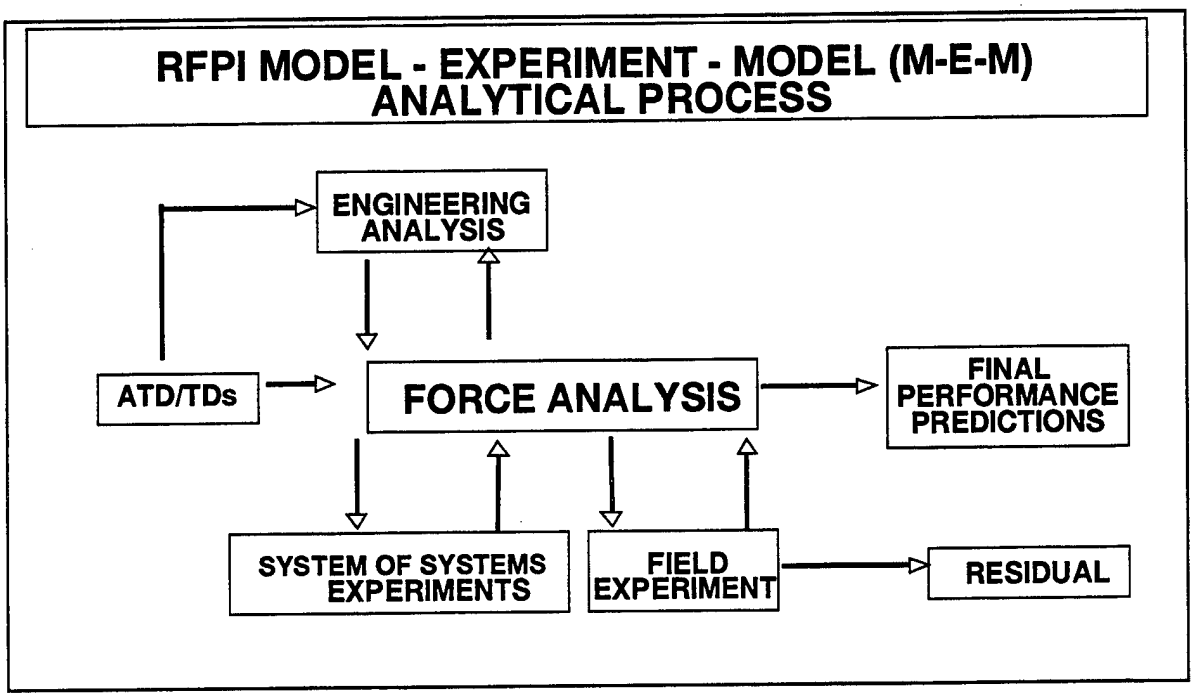


Figure 1 the Model Experiment Model Process

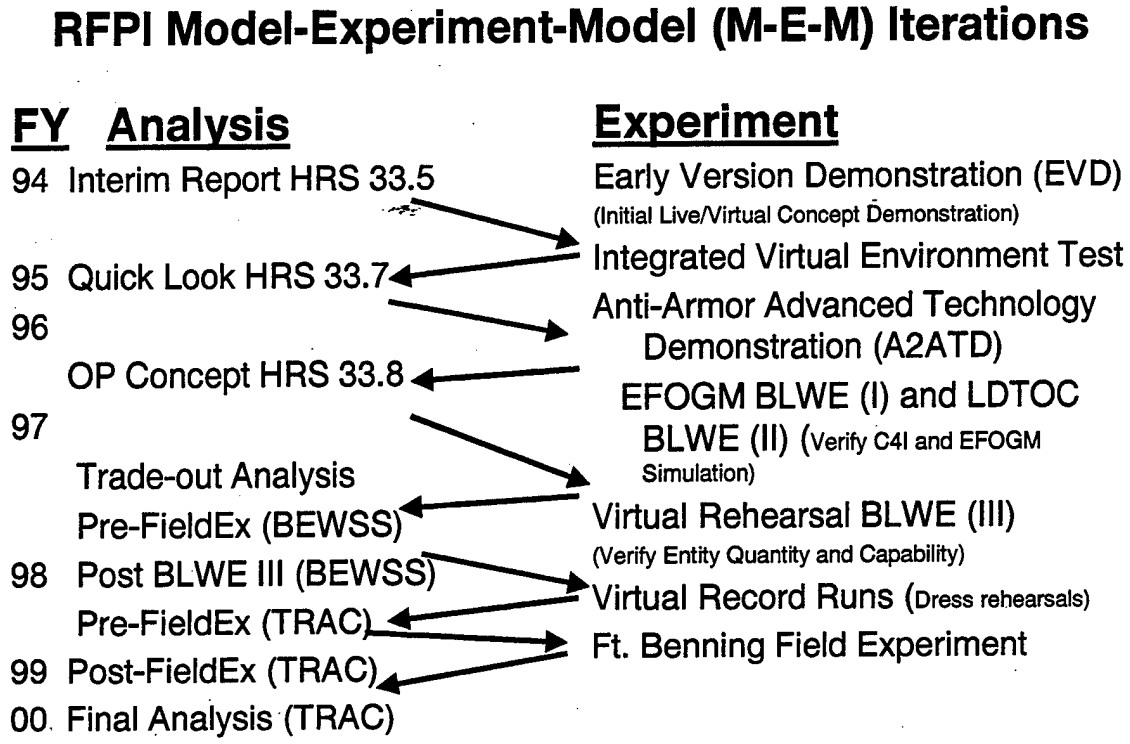


Figure 2 Model Experiment Model Iterations

The two principle components for the definition of success were the proof of the success of the live virtual DIS experiment. Secondly, success was in the completeness and applicability of the analysis and assessment of the equipment comprising the residual and advanced systems. This is in preparation for the accelerated fielding to the warfighter of the RFPI system of systems.

The Live/Virtual Field Experiment and the Distributed Interactive Simulation

The purpose of the live/virtual field experiment is to expand the live fight to the full Blue experimental force, a Division Ready Brigade versus the experimental operational force representing a Red division, and to represent the entire compliment of live and virtual entities in virtual domain. This will enable interaction of live and virtual entities, represent all munitions firing, detonations, and casualties in virtual domain; Inform live entities of their damage status; reflect direct-fire miles casualties in virtual domain; synchronize live and virtual target acquisitions and battlefield damage assessment. In addition, the technology created for the Ft. Benning Field Experiment allows the transition one virtual battalion of Opposition Forces (OPFOR) to live OPFOR at the Ft. Benning range boundary, and interface with live OPFOR voice networks.

In the process of simulating/stimulating the brigade command, control, communication, computers, and intelligence (C4I), critical virtual operational facilities are represented so that they can participate on

tactical voice networks and tactical VMF networks. The Army Tactical Command and Control System (ATTCS) is stimulated to the degree supported by existing stimulation tools. Additionally, the C4I network was used to support exercise control, data collection, and analysis, to interface the virtual environment observer/controllers to the live Ocs through a simulated live/virtual voice network. The digital communications network also facilitated the accumulation and display of battle views and statistics, and the integration with experimental control and instrumentation control via voice and digital nets. All of this coordination and communication was accomplished during the field experiment in real time!

Simulation capabilities were developed enabling the Rapid Force Projection Initiative (RFPI) live / virtual experiments, which interfaced live instrumentation to the distributed interactive simulation (DIS) backbone, allowing the DIS integration of over 1500 entities, both live and virtual. The "shadow server" concept was developed to allow live/virtual interactions, with the shadow server allowing virtual to live transitions in real-time across the Ft. Benning range boundaries (the live ground vehicles could not operate beyond the bounds of the military reservation.) The virtual C4I systems stimulated digital networks modeled after the Task Force XXI (TFXXI) communications system. The field experiment demonstrated the first and only DIS air assault scenario, and demonstrated portability to provide the DIS battle to remote facilities. In order to facilitate the many types of systems represented in the live/virtual simulation, multiple models (ModSAF, TAFSM, IDEEAS, FireStorm, ITEMS) were made interoperable, sharing information about firings, impacts, unit location and status over a common shared information backbone. This information was communicated between the field units, the simulators at Ft. Benning, and the ModSAF and IDEEAS suites at Redstone Arsenal over a high-speed data network. At the same time, monitoring the status of the experiment in real-time were analysis tools collecting information pertaining to the status of the experiment, and collecting data in order to answer the pre-assigned Measures of Effectiveness (MOEs). These answers were then made available for discussion at an After Action Review, conducted directly following the end of each exercise portion.

Field Experiment Accomplishment

RFPI has just executed the most highly interactive live/virtual/constructive simulation exercise ever achieved or attempted. In this Blue brigade versus Red division fight, all combinations of live or virtual system interactions were allowed. Blue C4I systems had completely seamless stimulation of digital and voice traffic for Air Defense Artillery, maneuver, intelligence, and fire support. Virtual entities translated to live elements automatically, with contiguous translation and correlation. No aggregation was used in this 1500+-entity fight, including the first ever live/virtual air assault mission. Finally, RFPI has shown that Distributed Interactive Simulation works for experimentation.

Constructive Simulation - Less than Successful

The intent of the constructive simulation was to specify that the items selected for the Residual Systems fielding to the 2nd BDE/101st AA Division were contributory on the combined arms battlefield, and that even in a brassboard state they could contribute to the performance of the warfighter. Performance data that had been provided by project managers and developers had been used in simulation studies for years. The reality of the situation from the viewpoint of the field experiment was how unprepared some of the systems were when put into the hands of the warfighter. Pretesting of several systems was not possible due to developer delivery slippage. No amount of political posturing or wishful thinking would allow several of the key Residual Systems to perform to the expected level. Even several systems already on the fast track to Army acquisition were not as well prepared as had been expected. However, the systems which were ready and were selected for continuation in the residual period of the ACTD will benefit from the experience, especially as they transition to acquisition.

So, the conundrum is that simulation is valuable in analysis and assessment. But, limitations due to using data that may be inadequate, less than objective, or fanciful may serve to compound problems, especially as doctrinal and systems' integration issues are being planned in the system of systems concept of operations. However this performance expectation sandcastle can readily come apart when actual performance falls far short of predicted data, and the expectations to the warfighter experimental unit are not realized. A case in point is the performance of the acoustic sensor making up the Integrated Acoustic System (IAS) and the Air

Delivered Acoustic System. In simulation, this system was key for early warning of approaching enemy, for direction and classification as to type and number of approaching vehicles, and to pinpoint artillery, all at extended ranges. This information was to be used to cue attack helicopter launch and interception, artillery and EFOGM fire, and maneuvering of reserve elements if a strong enemy thrust was detected. This sensor was also an integral element in the Improved Minefield System, or RAPTOR. Simulation showed the many and various benefits of this system. In actuality, the system did not perform to the level of expectation, and so was eventually dropped from consideration as a residual system. There is a place for speculative data of a system's performance – much of the combat simulation work speculates on performances of systems 10 to 15 years out. But when an actual delivery date was at hand, and actual performance information was required to feed simulation, there is no tolerance for wishful thinking.

So, the results of the constructive simulation experiments, in light of the real performances of the systems in the hands of the experimental force cannot be considered a successful representation to date. The results of the constructive CASTFOREM and Janus simulations were useful for preparation of the field experiment, and have use in investigating the individual RFPI systems as they progress toward acquisition, through Analysis of Alternatives and Cost and Operational Effectiveness Analyses. However, several systems which proved very valuable to the force in simulation, and while the performance of these systems was taken from the input from engineering analysis and data from the proponents, they in reality were not ready for delivery as residual systems. Nor were the performances in the hands of troops that which was anticipated.

Empirical Performance of Some Tests of Hypothesis on CASTFOREM Output

Patrick D. Cassady
TRADOC Analysis Center (TRAC)-White Sands Missile Range (WSMR)
White Sands, NM 88002

ABSTRACT

The Combined Arms and Support Task Force Evaluation Model (CASTFOREM) is a brigade level, force on force, stochastic simulation model. This model is widely used by the US Army in material acquisitions studies to compare the combat effectiveness of alternative weapons systems. Combat effectiveness is often measured by enemy (red) losses, friendly (blue) losses, or their ratio, the loss exchange ratio (LER). Since a single replication of a CASTFOREM production scenario may take several hours to run efficient statistical analysis procedures are necessary. Empirical distributions of losses for two alternatives were developed from 500 independent replications of CASTFOREM. The empirical performance of some tests of hypothesis was assessed using these distributions.

INTRODUCTION

The Combined Arms and Support Task Force Evaluation Model, CASTFOREM, is a stochastic, discrete event, force on force, combat simulation model. This model is widely used in Army analyses to compare the combat effectiveness of alternative weapons systems in the context of brigade or smaller organizations. Resolution of the model is to the individual weapon systems, for example, tank or helicopter. CASTFOREM provides detailed representations of maneuver, communications, search, weapon system engagement, terrain, weather, and obscurants. A specific configuration of CASTFOREM to represent combat forces, tactics, terrain, weather, and weapon system performance is termed a scenario.

Analyses with CASTFOREM begin with the delineation of alternatives and the selection of scenarios that capture a broad range of potential use of the alternatives. A typical analysis involves a dozen alternatives and several scenarios. Combat effectiveness of alternatives is often measured by enemy (red) losses, our (blue) losses, or their ratio, the loss exchange ratio (LER). In addition to these force level measures of effectiveness CASTFOREM provides many other measures by which alternatives may be compared. The overarching approach to statistical analysis is one of hypothesis testing that is implemented by a oneway layout. The analysis proceeds by generating samples of independent replications of CASTFOREM for each alternative and for each scenario. Standard statistical tests, for example, t-test, F-test, Mann-Whitney, Kruskal-Wallis test, Scheffe's multiple comparison procedure, are then applied.

Approved for public release; distribution is unlimited.

Efficient and effective statistical analysis procedures are necessary. A single replication of CASTFOREM may take several hours of computer time to run so sample sizes are limited. Force level measures of effectiveness are discrete and finite. Many of the standard tests of hypothesis are derived under the assumption of continuously distributed data or normally distributed data. Consequently these standard tests may not perform well on force level output data from CASTFOREM. This study investigated the question of "how in practice do some standard statistical tests perform with regard to a particular set of CASTFOREM force level output data?"

ASSESSMENT PROGRAM

Performance of tests of hypothesis was assessed in the following manner. First two distinct alternatives and a single scenario from a recent analysis were chosen. Next CASTFOREM was run 500 times for each alternative. The two samples of 500 replications were then treated as two distinct empirical distributions. These empirical distributions may be thought of as surrogates of the CASTFOREM processes from which they were generated. Next various test of hypothesis were performed on these empirical distributions by repeatedly drawing from them random samples. Because generating random samples from the two empirical distributions is much easier than generating random samples from CASTFOREM various "what if" type investigations such as varying sample size could be accomplished quite easily. The operational assumption is that the performance of a test on the empirical distributions is similar to its performance on the CASTFOREM processes themselves.

Performance with regard to type one error was assessed as follows. Two independent random samples were drawn from the first empirical distribution, which was generated from alternative 1 output. A test of hypothesis, with the null hypothesis that the distributions are the same, was performed. The result of the test, reject the null or accept the null, was noted. The test was then repeated 100 times. The number of times that the test rejected the null, i.e. the number of type one errors, was compared to the theoretical performance of the test.

Performance with regard to type two error was assessed similarly. Independent random samples were drawn from each of the two empirical distributions. A test of hypothesis, with the null hypothesis that the distributions are the same, was performed. The result of the test, reject the null or accept the null, was noted. The test was then repeated 100 times. The number of times that the test did not rejected the null, i.e. the number of type two errors, was compared to the theoretical performance of the test.

Experience has shown that general characterizations of CASTFOREM force level output are difficult. Examples exist showing various degrees of skewness, kurtosis, and heteroscedasticity in the force level data for alternatives of a particular analysis. For this study one scenario and two alternatives from a recent analysis were chosen. To add practicality to the study it was desired to work with a scenario that is frequently used in current analyses and to use alternatives from an actual analysis. However, no claim is made for the representativeness of the data used nor is any inference made from the results of this study to CASTFOREM results in general. In the scenario studied a blue armored brigade attacks a defending red brigade during

the night in desert terrain. The scenario is quite large and complex with over 500 systems on the attacking side and 400 systems defending. Alternative 2 was derived from alternative 1 by replacing 50 combat systems of a certain type with systems having improved survivability and sensor performance characteristics. These two alternatives were chosen because, based on the sample size used in the original analysis, their standard deviations were approximately equal.

RESULTS

Table 1 lists some standard statistics for the two empirical distributions. As illustrated in the table alternative 2 shows improvement in mean of about 10 red losses, 19 blue losses, and

Table 1. Force Level Statistical Summary of the Empirical Distributions

	Red Losses	Blue Losses	LER
Alternative 1 (n = 500)			
Mean	217.10	208.52	1.053
StdDev	12.49	18.95	0.137
Median	218.00	208.00	1.049
Minimum	164.00	159.00	0.648
Maximum	249.00	270.00	1.456
Kurtosis	0.28	-0.02	-0.06
Skewness	-0.30	0.32	0.14
Alternative 2 (n = 500)			
Mean	227.27	189.72	1.212
StdDev	10.87	18.36	0.154
Median	229.00	189.00	1.204
Minimum	189.00	143.00	0.768
Maximum	253.00	264.00	1.720
Kurtosis	0.23	0.50	0.48
Skewness	-0.40	0.32	0.32

0.16 in LER. A t-test on the LER has P value < 0.0005. Levene's test for equality of variances for LER has P value 0.068. The Pearson product correlation of red losses with blue losses is -0.55 for alternative 1 and -0.50 for alternative 2. Such negative correlations are indicative of the general characteristic that the better one force does the worse the other. The coefficients of sknewness and kurtosis (scaled so that the coefficient of the normal distribution is 0) illustrated in Table 2 indicated that the distributions are rather symmetric and neither heavy nor light in the tails.

Figure 1 gives a scatterplot of red losses versus blue losses for alternative 1. Figure 2 gives a similar scatterplot for alternative 2. Scatterplots are a convenient devise for screening force level output. In such a plot the LER for a particular blue loss and red loss pair is simply the slope of the ray from the origin to the point representing the pair.

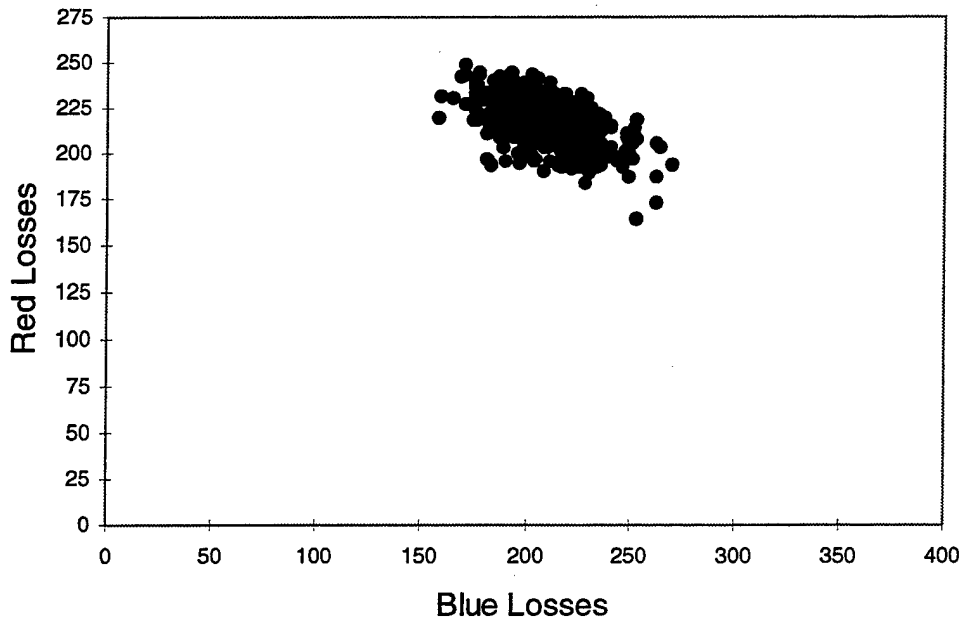


Figure 1. Scatterplot of Losses for Alternative 1

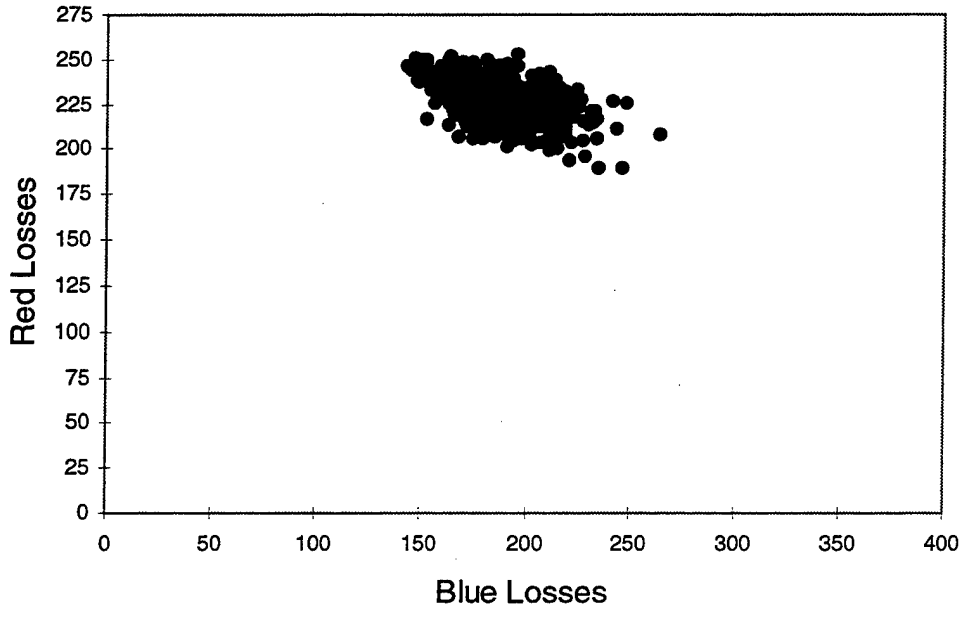


Figure 2. Scatterplot of Losses for Alternative 2

Figure 3 gives a quantile-quantile plot for the LER of empirical distribution of alternative 1 versus the standard normal distribution. Figure 4 gives a similar plot for alternative 2. Both these plots show general agreement of the empirical distributions with the normal distribution except for slight deviations in the tails.

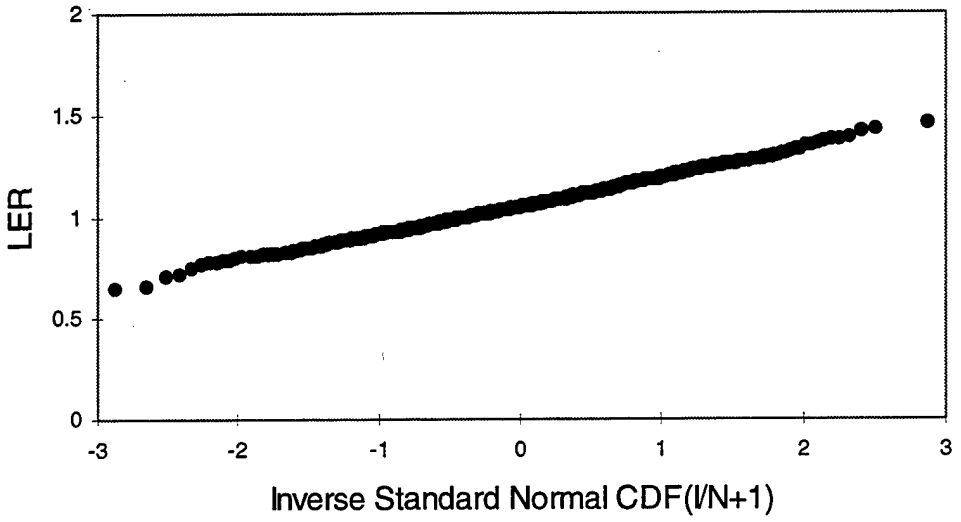


Figure 3. Quantile-Quantile Plot of LER for Alternative 1

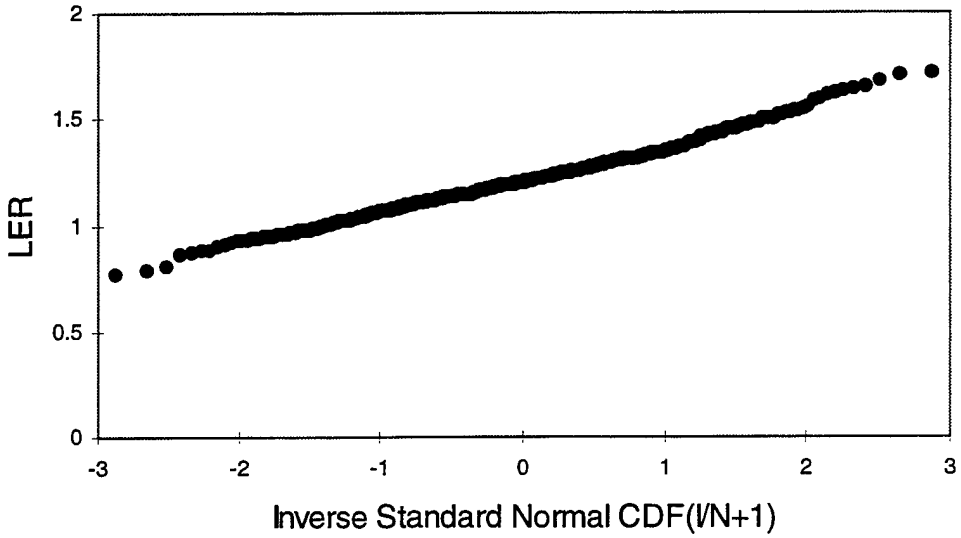


Figure 4. Quantile-Quantile Plot of LER for Alternative 2

Table 2 gives the results of the performance assessment for the t-test and the Mann-Whitney test. The tests were two-sided and conducted at the 5% level.

Table 2. Hypothesis Testing Errors

	n = 7		n = 17		n = 27	
	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2
t-test						
Theoretical	0.05	0.60	0.05	0.20	0.05	0.05
Computed	0.04	0.50	0.04	0.11	0.05	0.01
Mann-Whitney						
Theoretical	0.05	NA	0.05	NA	0.05	NA
Computed	0.05	0.50	0.04	0.11	0.06	0.01

The theoretical type two values are from *Pocket Book of Statistical Tables*, Odeh et al., Marcel Dekker, NY, 1977 and assume normal distributions of the same shape with means differing by one standard deviation. With regard to type one error the tests performed as expected. The tests for each of the three sample sizes had a type two error less than the theoretical values.

SUMMARY

In summary for the two empirical distributions of LER considered in this study the t-test and Mann-Whitney tests performed close to their theoretical values. In conducting this study a Bayesian approach to the statistical analysis was suggested as a potential area for future research. As the scatterplots illustrate, blue and red losses may be represented as points with integer coordinates in the first quadrant. The probabilities of the possible outcome points for any particular alternative can be modeled with a Dirichlet prior distribution. The CASTFOREM output of the alternative can be modeled with a multinomial likelihood function over this set of points. The Dirichlet prior, the natural conjugate prior for a multinomial likelihood, combines with the likelihood to give a Dirichlet posterior distribution for the probabilities of the possible outcomes of the alternatives. The LER can be treated as a utility function on the probabilities of the possible outcome points. By integrating the LER utility function with their posterior distributions, alternatives can be compared by their expected utility. A difficulty with this approach is the large number of parameters required of the Dirichlet distributions.

"An Application of Mixed Models for Comparing Accuracy from Two Types of Firing Platforms"

David W. Webb, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD
Thomas Mathew, University of Maryland Baltimore County, Baltimore, MD

Introduction

In 1997, the Program Manager for Tank Main Armament Systems (PM-TMAS) at Picatinny Arsenal, New Jersey commissioned a study to determine if the type of firing platform used affects the accuracy of the M1 tank main gun. More specifically, the initial test objectives were to evaluate platform effects on average impact location and target impact dispersion (TID).

It was desirable for the study to be conducted under a wide array of firing conditions, including different lots of ammunition, ammunition temperatures, and gun tubes. The Army Test Center at Aberdeen Proving Ground, Maryland formulated an experimental design to conduct the study. Consultants from both the University of Delaware Department of Mathematics and the U.S. Army Research Laboratory (ARL) agreed that the test plan was adequate to satisfy the initial test objectives. Logistics precluded a completely randomized study, and the test was conducted using the following sequence of factor combinations:

Factor Combination	**Factor 1** Lot	**Factor 2** Ammunition Temperature	**Factor 3** Gun Tube	**Factor 4** Platform	**Date of Firing**
1	1	Cold	A	1	Wed 28 May
2	1	Cold	A	2	Wed 28 May
3	1	Ambient	B	2	Thu 29 May
4	1	Ambient	B	1	Thu 29 May
5	1	Hot	C	1	Fri 30 May
6	1	Hot	C	2	Fri 30 May
7	2	Ambient	C	1	Mon 25 Aug
8	2	Ambient	C	2	Mon 25 Aug
9	2	Hot	A	2	Tue 26 Aug
10	2	Hot	A	1	Tue 26 Aug
11	2	Cold	B	1	Wed 27 Aug
12	2	Cold	B	2	Wed 27 Aug
13	3	Hot	B	1	Wed 17 Sep
14	3	Hot	B	2	Wed 17 Sep
15	3	Cold	C	2	Thu 18 Sep
16	3	Cold	C	1	Thu 18 Sep
17	3	Ambient	A	1	Mon 22 Sep
18	3	Ambient	A	2	Mon 22 Sep

Table 1. Firing Sequence

For each factor combination, a total of 10 rounds of ammunition were fired at fixed target positions downrange, and the impact locations were recorded using a Cartesian coordinate system. However, because of equipment failures at the test facility, factor combinations had less than 10 observations.

Impact locations from the study are presented in Figure 1. Due to classification restrictions, units of scale have been omitted from both scatterplots. However, both scatterplots are shown on the same scale to allow for visual comparison of the overall pattern of impact locations.

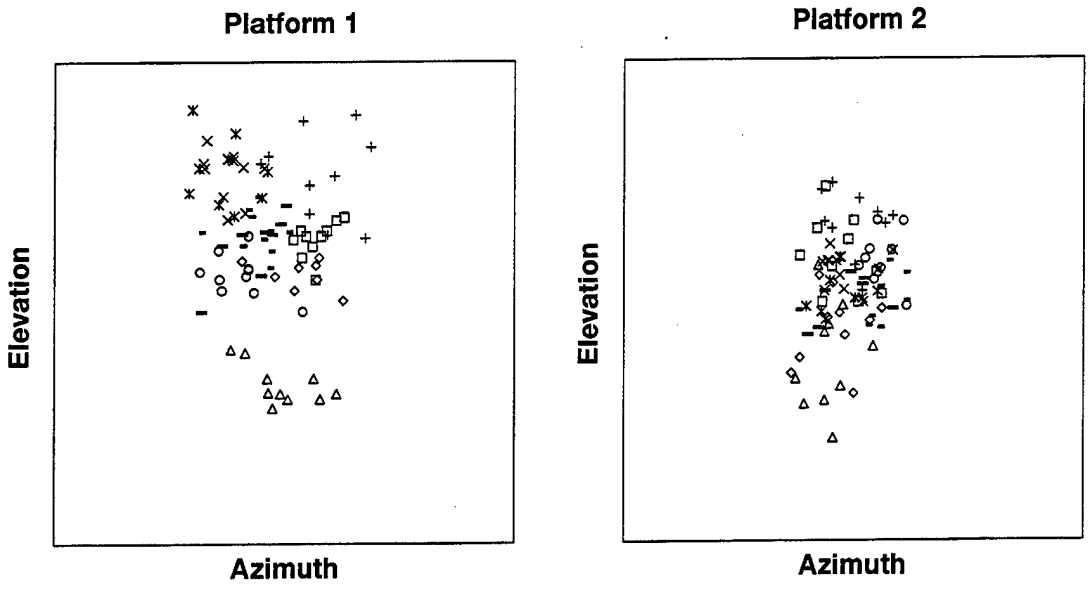


Figure 1. Impact Locations for Each Platform Type

The experimental design can be viewed as two Latin squares. Figure 2 shows the arrangement of the factor combinations and includes the number of observations per factor combination in the upper-right corner of each cell.

Consultants from the University of Delaware Department of Mathematics used analysis of variance (ANOVA) to determine the effects of each factor upon mean impact. Historically, the impact data for azimuth and elevation directions have shown to be independent. Therefore, separate univariate analyses were performed on the data. The ANOVA showed that the Platform 1 rounds landed to the left of ($p=0.03$) and higher than ($p<0.01$) Platform 2 rounds.

To determine if there was a difference in the TIDs obtained for each type of firing platform, ARL conducted simple independent, two-tailed, two-sample hypothesis tests on the pooled estimates of variance. These tests showed no platform effect on TID in the azimuth ($p=0.34$); however, in the elevation, the TID from Platform 1 was found to be somewhat lower than that from Platform 2 ($p=0.07$).

	P_1		
	L_1	L_2	L_3
F_1	10 T_1	10 T_2	10 T_3
F_2	8 T_2	10 T_3	10 T_1
F_3	10 T_3	9 T_1	8 T_2

	P_2		
	L_1	L_2	L_3
F_1	10 T_1	10 T_2	9 T_3
F_2	10 T_2	10 T_3	10 T_1
F_3	10 T_3	10 T_1	6 T_2

Key: P = Platform L = Lot
F = Temperature T = Gun Tube

Figure 2: Test Design Viewed as Two Latin Squares

Presented with these conclusions, PM-TMAS engineers then asked ARL to evaluate as a follow-on question "Is the variability in centers of impact different between the two platform types?" Although the experimental design was not drafted with this question in mind, ARL, in cooperation with the University of Maryland Baltimore County (UMBC) Department of Mathematics and Statistics, had conducted similar comparisons in previous studies and agreed to provide the analysis necessary to answer this question.

Analysis

The type of analysis needed to determine if the type of platform has an effect on the variability of centers of impact depends upon the assumption of homoscedasticity. The initial analyses showed that for the elevation (vertical) data there was a significant difference in TID between platform types. Hence, these data will be analyzed using a generalized p-value approach [see Weerahandi 1995; Khuri, et. al. 1998]. However, azimuth (horizontal) data were homoscedastic. Therefore, a mixed-model ANOVA was used for the analysis of azimuthal impacts. Only the mixed-model approach will be discussed henceforth in this report.

Each observation, denoted by x_{ijkmn}, represents a wind-corrected impact location along the azimuthal axis. The mathematical model for the ANOVA is

$$x_{ijkmn} = \mu + l_i + f_j + t_{km} + p_m + e_{ijkmn},$$

where x_{ijkmn} is the n^{th} observation corresponding to treatment combination $L_iF_jT_kP_m$, for i, j, k = 1, 2, 3 and m = 1, 2. Other terms in the model are

μ,	the common mean;
l_i,	the effect due to the i^{th} lot;
f_j,	the effect due to the j^{th} temperature;
t_{km},	the effect due to the k^{th} tube mounted on the m^{th} platform;
p_m,	the effect due to the m^{th} platform; and,
e_{ijkmn},	the random error.

The terms l_i, t_{km}, and e_{ijkmn} are all independent, normally distributed random variables whereby $l_i \sim N(0, \sigma_l^2)$, $t_{km} \sim N(0, \sigma_{tm}^2)$ for $m = 1, 2$, and $e_{ijkmn} \sim N(0, \sigma_e^2)$. The terms f_j and p_m are fixed effects satisfying $f_1 + f_2 + f_3 = 0$, and $p_1 + p_2 = 0$.

We also assume that $\mathrm{Cov}(t_{k1}, t_{k2}) = \rho_t \geq 0$ for $k = 1, 2, 3$. A justification for assuming $\rho_t \geq 0$ is that t_{k1} and t_{k2} correspond to the same tube. Furthermore, if the effect due to the tube is independent of the platform, so that we can write $t_{km} = t_k$, then ρ_t is the variance of t_k and hence must be nonnegative.

Testing for platform-type differences in the variability of centers of impact is paramount to testing the hypotheses H_0: $\sigma_{t1}^2 = \sigma_{t2}^2$ versus H_1: $\sigma_{t1}^2 \neq \sigma_{t2}^2$.

We proceed by letting

$\bar{x}_{1111}$ = average of the 10 observations for $L_1F_1T_1P_1$,
$\bar{x}_{2311}$ = average of the 9 observations for $L_2F_3T_1P_1$,
$\bar{x}_{3211}$ = average of the 10 observations for $L_3F_2T_1P_1$, and
$\bar{x}_{..11}$ = average of the $\bar{x}_{1111}$, $\bar{x}_{2311}$, and $\bar{x}_{3211}$,

and similarly define $\bar{x}_{..21}$, $\bar{x}_{..31}$, $\bar{x}_{..12}$, $\bar{x}_{..22}$, and $\bar{x}_{..32}$. Then,

$$\mathrm{E}(\bar{x}_{..11}) = \mathrm{E}(\bar{x}_{..21}) = \mathrm{E}(\bar{x}_{..31}) = \mu + p_1, \text{ and}$$
$$\mathrm{E}(\bar{x}_{..12}) = \mathrm{E}(\bar{x}_{..22}) = \mathrm{E}(\bar{x}_{..32}) = \mu + p_2.$$

Additionally,

$$\mathbf{Var}(\bar{x}_{..11}) = \frac{\sigma_l^2}{3} + \sigma_{t1}^2 + \frac{\sigma_e^2}{9}\left(\frac{1}{10} + \frac{1}{9} + \frac{1}{10}\right),$$

$$\mathbf{Var}(\bar{x}_{..21}) = \frac{\sigma_l^2}{3} + \sigma_{t1}^2 + \frac{\sigma_e^2}{9}\left(\frac{1}{8} + \frac{1}{9} + \frac{1}{8}\right),$$

$$\mathbf{Var}(\bar{x}_{..31}) = \frac{\sigma_l^2}{3} + \sigma_{t1}^2 + \frac{\sigma_e^2}{9}\left(\frac{1}{10} + \frac{1}{10} + \frac{1}{10}\right),$$

$$\mathbf{Var}(\bar{x}_{..12}) = \frac{\sigma_l^2}{3} + \sigma_{t2}^2 + \frac{\sigma_e^2}{9}\left(\frac{1}{10} + \frac{1}{10} + \frac{1}{10}\right),$$

$$\mathbf{Var}(\bar{x}_{..22}) = \frac{\sigma_l^2}{3} + \sigma_{t2}^2 + \frac{\sigma_e^2}{9}\left(\frac{1}{10} + \frac{1}{10} + \frac{1}{6}\right),$$

$$\mathbf{Var}(\bar{x}_{..32}) = \frac{\sigma_l^2}{3} + \sigma_{t2}^2 + \frac{\sigma_e^2}{9}\left(\frac{1}{10} + \frac{1}{10} + \frac{1}{9}\right).$$

The covariances are given by,

$$\mathrm{Cov}(\bar{x}_{..11}, \bar{x}_{..21}) = \mathrm{Cov}(\bar{x}_{..11}, \bar{x}_{..31}) = \mathrm{Cov}(\bar{x}_{..21}, \bar{x}_{..31}) = \frac{\sigma_l^2}{3},$$

$$\mathrm{Cov}(\bar{x}_{..12}, \bar{x}_{..22}) = \mathrm{Cov}(\bar{x}_{..12}, \bar{x}_{..32}) = \mathrm{Cov}(\bar{x}_{..22}, \bar{x}_{..32}) = \frac{\sigma_l^2}{3},$$

$$\mathrm{Cov}(\bar{x}_{..11}, \bar{x}_{..12}) = \mathrm{Cov}(\bar{x}_{..21}, \bar{x}_{..22}) = \mathrm{Cov}(\bar{x}_{..31}, \bar{x}_{..32}) = \frac{\sigma_l^2}{3} + \rho_l, \text{ and}$$

$$\mathrm{Cov}(\bar{x}_{..k1}, \bar{x}_{..k'2}) = \frac{\sigma_l^2}{3}, \; k \neq k'.$$

To circumvent the difficulties imposed by the unbalanced nature of the data, suppose the number of observations had been the same per factor combination. In this case, define

$$\sigma_0^2 = \frac{\sigma_e^2}{9} \times \left(\frac{1}{10} + \frac{1}{10} + \frac{1}{10}\right)$$

and

$$\delta_1^2 = \sigma_{t1}^2 + \sigma_0^2, \text{ and } \delta_2^2 = \sigma_{t2}^2 + \sigma_0^2.$$

Then our testing problem is equivalent to testing

$$H_0\colon \delta_1^2 = \delta_2^2 \quad \text{versus} \quad H_1\colon \delta_1^2 \neq \delta_2^2.$$

We can show that

$$\sum_{k=1}^{3}[\bar{x}_{..k1}-\bar{x}_{...1}]^2/(\rho_t+\delta_1^2)\sim\chi^2,$$

and

$$\sum_{k=1}^{3}[\bar{x}_{..k2}-\bar{x}_{...2}]^2/(\rho_t+\delta_2^2)\sim\chi^2,$$

where the chi squares have two degrees of freedom. Each of these distributions does not depend on σ_t^2, so we assume $\sigma_t^2=0$. (In the above, $\bar{x}_{...1}$ denotes the average of $\bar{x}_{..11}$, $\bar{x}_{..21}$, and $\bar{x}_{..31}$; $\bar{x}_{...2}$ is similarly defined.)

However, the two chi-square random variables are not independent. Nevertheless, it is reasonable to reject H_0 when the ratio

$$F=\frac{\sum_{k=1}^{3}[\bar{x}_{..k1}-\bar{x}_{...1}]^2}{\sum_{k=1}^{3}[\bar{x}_{..k2}-\bar{x}_{...2}]^2},$$

is too large, or too small. But F does not have an F-distribution, since the two chi-squares are not independent.

Suppose, as an approximation, we carry out the test using an F-distribution with two degrees of freedom in both the numerator and denominator. If we use a 5% significance level, we reject H_0 when either

$$F\le 0.0256 \quad \text{or} \quad F\ge 39.$$

Since F does not have an exact F-distribution, we have to check via simulation what the actual type I error probability will be for the above test. We proceed by letting $\delta_1^2=\delta_2^2=\delta^2$ (under H_0). It can be shown that the null distribution of F depends on the ratio ρ_t/δ^2. Given below are the simulated Type I error probabilities (based on 10,000 simulations) of the test for various values of the ratio ρ_t/δ^2. The simulations show that when carrying out the test with a 5% significance level, the actual type I error probability is 5% or less.

ρ_t/δ^2	0	.1	.5	1	10	100
Type I error	0.0491	0.0493	0.0423	0.0362	0.0084	0.0009

Table 2. Type I error estimates for $\alpha=5\%$ and various values of ρ_t/δ^2.

The above derivations are based on the assumption that the quantities $\bar{x}_{..11}$, $\bar{x}_{..21}$, and $\bar{x}_{..31}$ have a common variance as do the quantities $\bar{x}_{..12}$, $\bar{x}_{..22}$, and $\bar{x}_{..32}$. In other words, the coefficient of $\sigma_e^2/9$ is the same among the variances of these quantities.

However, in our data set, this is not the case, since we have unequal numbers of observations within the various factor combinations.

In order to make the coefficient of $\sigma_e^2/9$ the same for all the variances, suppose we make the number of observations in each group equal to 10, by imputing artificial observations in the following manner:

1. Compute the mean and variance of the available data for a factor combination.

2. Randomly generate observations from a normal distribution having this mean and variance, so that the number of observations per factor combination becomes 10.

With these additional imputed values, the F statistic and corresponding p-value were obtained. However, instead of drawing conclusions from a single data set containing artificial values, the process was repeated 1000 times to obtain an estimated distribution of the p-value.

Conclusion

Figure 3 shows the distribution of the 1000 P-values corresponding to the 1000 imputated data sets. We see that a majority (nearly 91%) of the P-values are 5% or less. Because of this large percentage and the fact that the F test used is conservative, we feel comfortable in concluding that $\delta_1^2 \neq \delta_2^2$, and hence that the platform types have different variations in their centers of impact.

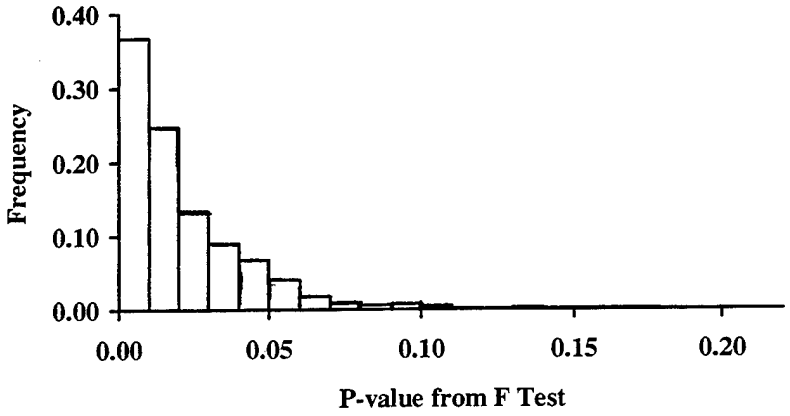


Figure 3. Histogram of P-Values for Test of H_0: $\delta_1^2 = \delta_2^2$ versus H_1: $\delta_1^2 \neq \delta_2^2$.

References

Khuri, A., T. Mathew, and B. Sinha. *"Statistical Tests for Mixed Linear Models."* Wiley, New York, 1998.

Weerahandi, S. *"Exact Statistical Methods for Data Analysis."* Springer-Verlag, New York, 1995.

MINIMUM-PERCENTAGE-ERROR REGRESSION UNDER ZERO-BIAS CONSTRAINTS

Stephen A. Book and Norman Y. Lao
The Aerospace Corporation
El Segundo, CA 90245

ABSTRACT

Classical least-squares regression imposes severe requirements on analysts who want to derive functional relationships between dependent y and independent x variables, forcing the analyst to model the error as additive when the relationship is linear ($y = a+bx$) or logarithmic ($y = a + blogx$) but as multiplicative when the relationship is exponential ($y = ax^b$) or power ($y = ab^x$). This severely restricts his or her ability to optimally model natural phenomena. "General-error regression," taking advantage of modern computing capability and advanced numerical analysis techniques, offers the analyst the choice of minimizing additive or multiplicative error regardless of the functional form of the relationship. It turns out, though, that relationships derived by minimizing percentage (i.e., multiplicative) error contain significant positive bias (i.e., they tend to overestimate the actual values of the dependent variable). In the recent past, the method of iteratively reweighted least squares has been applied to yield zero-bias relationships at some cost in the magnitude of the standard error. In this report the general-error regression problem is instead formulated as a constrained nonlinear optimization problem, with percentage standard error of estimation optimized (i.e., minimized), subject to percentage bias being zero. Naturally, the percentage error will be somewhat larger than it would be if the bias were unconstrained (one cannot serve two masters!), but in general not as large as given by iteratively reweighted least squares, so zero bias is paid for by a small increase in standard error.

INTRODUCTION

Cost-estimating relationships (CERs) comprising Version 7 (August 1994) of the Air Force's *Unmanned Space Vehicle Cost Model* (USCM-7) have been statistically derived from historical cost data using "general-error" regression. Such CERs are usually expressed in the form of linear or curvilinear regression equations that predict cost (the dependent variable) as a function of one or more "cost drivers" (independent variables). Because the range of cost data behind a given CER will span one or more orders of magnitude, the correct choice of error model is "multiplicative" (a percentage of the estimate) rather than "additive" (a specific number of dollars). Unfortunately, when classical least-squares regression, or "ordinary least squares" (OLS), is used to derive functional relationships between dependent y and independent x variables, the analyst must model the error as additive when the relationship is linear ($y = a+bx$) or logarithmic ($y = a + blogx$) but as multiplicative when the relationship is exponential ($y = ax^b$) or power ($y = ab^x$). In the pre-computer age, when explicit formulas were used in the linear case to calculate the coefficients a and b from the data, the latter two curvilinear relationships were derived suboptimally (i.e., with a larger standard error of the estimate than necessary) by applying ordinary least-squares regression to the logarithms of the data points. In addition to inducing a larger than required error of estimation, this technique tended to yield relationships having significant negative bias (i.e., they underestimated the actual value of the dependent variable). Furthermore, when the coefficients of nonlinear forms are derived by taking logarithms of both sides and reducing the formulation to $log(y) = log(a) + b\ log(x) + log(E)$, the error of estimation is expressed in meaningless units ("log dollars"), so that the quality of the nonlinear form cannot easily be compared with the quality of the linear form, whose error of estimation is expressed in "dollars."

Using modern computing capability and advanced numerical analysis techniques in place of applying ordinary least squares to logarithmically-transformed data, The Aerospace Corporation developed "general-error regression" in order to derive functional relationships having optimal (i.e., minimum possible) error of estimation, while allowing the analyst to choose to minimize additive error or multiplicative error regardless of whether the functional relationship turns out to be linear or nonlinear. An additional advantage turned out to be that previously unavailable functional forms (most prominently $y = a+bx^c$) can be fit to the data when appropriate. Unfortunately, as was shown in 1993 by Tecolote Research Inc., functional forms derived by minimizing percentage (i.e., multiplicative) error act with significant positive bias (i.e., they tend to overestimate the actual values of the dependent variable). As a solution to the problem of bias, Tecolote suggested the technique of "iteratively

---

reweighted least squares" (IRLS). See Reference 11 and Appendix C of Reference 15 for details. This report, on the other hand, proposes to formulate the general-error regression problem as a constrained nonlinear optimization problem, the constraint being that the percentage bias of the functional relationship be zero. In particular, percentage standard error of estimation is optimized (i.e., minimized), subject to the percentage bias being zero, with the resulting standard percentage error somewhat larger than it would be if the bias were unconstrained, but in general somewhat smaller than given by IRLS.

General-error regression can be implemented in a number of commercial software packages, including Microsoft's Excel spreadsheet using the Excel Solver routine, which handles complex nonlinear problems by building a worksheet with multiple changing cells. Several numerical examples are provided to illustrate the relative magnitudes of standard error and bias. In any specific case, the analyst has the option to select the minimum-percentage-error relationship (typically with positive bias) or the zero-bias relationship (generally with suboptimal error of estimation).

WHY PERCENTAGE ERROR?

In USCM-7 (Reference 15), all errors of estimation, both standard errors and bias errors, are expressed in percentage terms, not in dollar values. There are two practical benefits of this that accrue to the cost estimator, for whose use the document was produced. The first benefit of expressing cost-estimating error in percentage terms is stability of meaning across a wide range of programs, time periods, and estimating situations. A percentage error of, say 30%, retains its meaning whether a $10,000 component or a $10,000,000,000 program is being estimated. A standard error expressed in dollars, say $59,425, is an extremely huge error when estimating a $10,000 component, but is much less significant when reported in connection with a $10,000,000,000 program. Even in cases that are not so extreme, a standard error expressed in dollars quite often makes a CER virtually unusable at the low end of its data range, where relative magnitudes of the estimate and its standard error are inconsistent.

While "standard error of the estimate" is the root-mean-square (RMS) of all percentage errors made in estimating points of the data base (a "one-sigma" number that bounds probable cost within an interval surrounding the estimate), "net-percentage bias" is the algebraic sum, including positive values and negative values, of all percentage errors made in estimating points of the data base. Net percentage bias is a measure of balance between percentage overestimates and underestimates of data-base actuals. The second practical benefit is the fact that a constant dollar-value expression of bias would not be as informative as estimating the error in percentage terms, because a particular amount of dollars of bias would not have the same meaning at every point of the cost range.

THE MULTIPLICATIVE-ERROR MODEL

OLS regression, either linear or nonlinear, has been applied in the past to historical-cost data in order to derive CERs. A fundamental assumption of OLS is that the error be additive. More precisely, each observed value of cost is assumed to be a function of cost-driving parameters plus a random error term that does not depend on the parameters. Unfortunately, this assumption is not always valid. A case in point is where the values of "actual" costs in the data base change by an order of magnitude or more as a function of the parameters, in which case the random error is more realistically considered to be proportional to the magnitude of the cost, thereby effectively depending on the parameters. In such a case it is often more realistic to assume a multiplicative error model. This type of situation has been dealt with in the past by taking logarithms of both sides and then applying additive-error linear regression. An alternate "ad-hoc" method is described in Reference 9; however, both are suboptimal in the least-squares sense. Other discussions of theoretical and practical difficulties of working with the logarithmic-transformation method can be found in References 3, 6, 9, 10, 13, and 18. This procedure also unnecessarily binds one to a specific class of regression-equation forms (see Reference 2), and it is far from clear that the appropriate forecasting error is the one that is being minimized. Reference 5 reports on a Monte Carlo study of the general question of additive vs. multiplicative error.

General-error regression is designed to fill this gap. It allows the user to specify, given historical-cost data, whether an additive or multiplicative error model is to be used in deriving the least-squares CERs. And it allows the user to select an appropriate functional form of the CER independently of his or her choice of error model. In the past, straight-line OLS regression forced the choice of an additive-error model, while particular curvilinear forms (namely, $y = ax^b$ and $y = ab^x$) required the assumption of a multiplicative-error model. In general-error regression, the choice of functional form is essentially unrestricted. As well the OLS-compatible forms $y = ax^b$ and $y = ab^x$,

now available are forms such as $y = a+bx^c$, $y = a + bc^x$, $y = a+bx+c\sqrt{x}$, etc. The range of available forms is unlimited.

Now that any of a wide range of functional forms may be combined with either of the two error models (additive or multiplicative), it is incumbent upon the cost analyst to choose the best pairing of functional form and error model that is consistent with engineering economics and historical-cost data. The decision made in the case of USCM-7 was to use the multiplicative-error model throughout the analysis and to let the choice of functional form be dictated by engineering and data considerations. It is felt that the multiplicative model, incorporating uniform percentage error of estimation across the entire cost range, reflects reality better than does a uniform dollar amount of error across that range. In cases where the dollar range is sufficiently narrow as to make the uniform-dollar-error assumption tenable, the percentage-error assumption is also adequate to model the reality.

Details of only the two-dimensional case are described in this report, but the procedures can be (and have been) easily generalized to higher dimensions, such generalizations having been used in USCM-7 to derive higher-dimensional CERs where appropriate. In the two-dimensional case, each observation consists of a deterministic cost-driving parameter (x) and a stochastic estimated cost (y). Both linear and nonlinear fits are considered, the theory being impervious to any specific form of the regression equation. The error definition is as follows:

Multiplicative Error = (Actual - Predicted) ÷ Predicted

The following are inputs to the mathematical computations:

n = Number of data points (observations) in sample
x_i = Value of cost-driving parameter for each data point, $i = 1, \ldots, n$
y_i = Observed value of cost for each data point $i = 1, \ldots, n$
m = Number of numerical coefficients in model $(m < n)$.

The term "coefficient" in this context includes numerically constant exponents, as well as numerically constant multiplicative factors. Then

$$y = f(x, \underline{a})$$

is the regression function, y of x, to be fit to the historical cost data, and

$$\underline{a} = (a_1, \ldots, a_m)$$

is the coefficient vector to be determined by mathematical optimization. If $m = n$, the parameter vector $\underline{a}$ is determined exactly ("interpolation"), regardless of the form of $f(x,\underline{a})$. If $m > n$, the parameter vector cannot be determined uniquely.

The multiplicative error model on which USCM-7 CER-development is based has the probabilistic structure

$$Y_i = f(x_i, \underline{a})\, \varepsilon_i, \quad i = 1, \ldots, n,$$

where ε_i is a random error such that

$$E(\varepsilon_i) = 1,$$
$$Var(\varepsilon_i) = \sigma_M^2,$$

and σ_M^2 represents a constant (independent of x_i) multiplicative-error dispersion around 1. Otherwise, as in the additive-error model, the probability distribution of ε_i is arbitrary. Mean and variance of Y_i are, respectively,

$$E(Y_i) = f(x_i, \underline{a})\, E(\varepsilon_i) = f(x_i, \underline{a})$$
$$Var(Y_i) = f^2(x_i, \underline{a})\, Var(\varepsilon_i) = f^2(x_i, \underline{a})\, \sigma_M^2$$

Note that, while the expected values are the same as those of the additive-error model, the variance in the multiplicative-error model depends on the cost-driving parameter, growing in magnitude as the dependent variable (in our context, cost) grows.

MINIMIZING PERCENTAGE ERROR OF ESTIMATION

In the multiplicative-error model, as before one sample observation y_i of Y_i corresponds to each x_i, but in this case the sample error ε_i equals the ratio of y_i to $E(Y_i)$. Thus.

$$\varepsilon_i = \frac{y_i}{E(Y_i)} = \frac{y_i}{f(x_i, \underline{a})}$$

where $\varepsilon_i = 1$ for all i indicates no prediction error. In this case, the least-squares problem can be formulated to find the parameter vector $\underline{a}$ that minimizes the sum of squared relative deviations from the predictions:

$$\mathrm{SSD}_M^2 = \sum_{i=1}^{n} (\varepsilon_i - 1)^2 = \sum_{i=1}^{n} \left(\frac{y_i}{f(x_i, \underline{a})} - 1 \right)^2 = \sum_{i=1}^{n} \left(\frac{y_i - f(x_i, \underline{a})}{f(x_i, \underline{a})} \right)^2 \quad (1)$$

This operation minimizes the sum of squares of the percentage errors for the multiplicative-error model. The far-right-hand representation expresses SSD_M^2 in terms of "relative error" and allows the least-squares multiplicative error model to be interpreted as minimizing the sum of squares of relative errors.

To solve the least-squares problem, we could use Equation (1) to calculate the m partial derivatives of SSD_M^2 with respect to each component $a_j, j = 1, \ldots, M$, of the parameter vector $\underline{a}$ and set them all equal to zero. If possible, we solve the system of simultaneous so-called "normal equations" for $\underline{a}$ to minimize SSD_M^2, but even if $f(x,\underline{a})$ is linear in $a_j, j = 1, \ldots, m$, the normal equations are not necessarily linear in the a_js. For nonlinear normal equations, numerical-analysis methods are usually necessary (unless the equations can fortuitously be solved in closed form analytically). The multidimensional Newton-Raphson method is a good technique (Reference 8). In general, the solution to a nonlinear system of equations is not often unique, because the function being minimized may have several "peaks" and "valleys." Unreasonable MPE solutions must be excluded, and the solution that is most plausible "physically" selected.

To estimate dispersion around the least-squares fit, the standard error of estimate for the multiplicative error model can be defined as follows:

$$SEE_M = \sqrt{\frac{1}{n-m} \sum_{i=1}^{n} \left\{ \frac{y_i}{f(x_i, \underline{a}_0)} - 1 \right\}^2} = \sqrt{\frac{1}{n-m} \sum_{i=1}^{n} \left\{ \frac{y_i - f(x_i, \underline{a}_0)}{f(x_i, \underline{a}_0)} \right\}^2}, \quad (2)$$

where this time $\underline{a}_0$ is the value of $\underline{a}$ that minimizes SSD_M^2. The far-right expression for SEE_M leads to interpretation of SEE_M as a measure of percentage error made in using the multiplicative error regression formula as a predictor. It would make sense to interpret SEE_M times 100% as the "one-sigma" percentage error made in using $f(x,\underline{a}_0)$ as an estimate of the cost corresponding to the cost-driving parameter x_i.

BIAS

Experience has shown that minimum-percentage-error (MPE) CERs resulting from minimizing the expression in Equation (1) have positive net percentage bias, defined by

$$B_M = \sum_{i=1}^{n} \left(\frac{f(x_i, \underline{a}) - y_i}{f(x_i, \underline{a})} \right) \quad (3)$$

The reason for this is not yet fully understood (by us), but we believe it may have something to do with the fact that, for the same absolute difference between y_i and $f(x,\underline{a})$, a smaller value of SSD_M^2 will result from choosing $f(x,\underline{a})$ above rather than below y_i. This is due to the fact that $f(x,\underline{a})$ appears in the denominators, and larger denominators lead to lower values of SSD_M^2. Nevertheless, the magnitude of the net percentage bias is not large for most CERs, typically being around 8%.

REDUCING BIAS BY ITERATIVELY REWEIGHTED LEAST SQUARES

Tecolote Research, Inc. (Reference 11 and Appendix C of Reference 15), the Air Force's prime contractor for development of USCM-7, suggested a method based on IRLS to reduce the bias of MPE CERs at a small cost in standard error. (See also References 1, 12, 16, and 17.) Tecolote's method, referred to in USCM-7 as the "minimum unbiased percentage error" (MUPE) technique, calls for computation of a sequence of CER parameter vectors $\underline{a}_1$, $\underline{a}_2$, ... converging to a parameter vector $\underline{a}_0$, that may or may not be optimal with respect to any appropriate criterion other than zero bias. If the functional form $f(x,\underline{a})$ is specified, then successive sequential CER candidates are defined as follows:

$$f\left(x_i, \underline{a}_{j+1}\right) = Min \sum_{i=1}^{n} \left(\frac{y_i - f\left(x_i, \underline{a}\right)}{f\left(x_i, \underline{a}_j\right)} \right)^2 . \tag{4}$$

Notice that only the parameter vector $\underline{a}$ in the numerator is subject to optimization; the denominator is constant with respect to the optimization process, having been selected in the previous iteration.

The operative element of Expression (4) can be rewritten in the following way:

$$\sum_{i=1}^{n} \left(\frac{y_i - f\left(x_i, \underline{a}\right)}{f\left(x_i, \underline{a}_j\right)} \right)^2 = \sum_{i=1}^{n} \frac{1}{f^2\left(x_i, \underline{a}_j\right)} \left(y_i - f\left(x_i, \underline{a}\right)\right)^2 \tag{5}$$

What is curious about Expression (5) is that the MUPE/IRLS technique is really an additive-error technique. It minimizes a weighted sum of additive squared errors. This apparently is how it manages to reduce the bias to zero. MUPE/IRLS is not truly a multiplicative-error technique. Nevertheless, once a solution is found, the percentage error of estimation and the bias can be calculated and compared with the corresponding statistics for MPE CERs. As noted earlier, the percentage error of a MUPE/IRLS CER will naturally be larger, but its bias will be less, exactly zero in the case of a linear functional form and apparently near zero in other cases.

MPE CERs constitute Section 5 of the USCM-7 document (Reference 15). As a perusal of Section 5 will reveal, MPE CERs tend to have positive average percentage bias (ranging from a low of 1% to a high of 29% at the extremes, with 8% as most typical). What mathematics guarantees about these CERs is that their standard percentage error is as small as possible, consistent with data-base applicability and appropriate technical relationships between cost and cost drivers.

MUPE/IRLS CERs constitute Section 6 of Reference 15. Yet, as the next section of this report will show, the standard errors of MUPE/IRLS CERs are not exactly minimized among the class of unbiased CERs. Perusal of Section 6 will show that these CERs tend to have standard errors somewhat greater than those in Section 5 (ranging from 0% to 23% greater at the extremes, with 6% greater most typical). Percentage bias of MUPE/IRLS CERs does indeed equal zero to the accuracy reported.

TRUE ZERO-BIAS CERs DERIVED BY CONSTRAINED OPTIMIZATION

Obtaining zero percentage bias by the MUPE/IRLS method is intellectually unsatisfying in one major respect: It is not clear (to us) exactly what quantity, quality, or characteristic, if any, is being optimized by the IRLS procedure. This difficulty is what inspired the ZPB/MPE (zero percentage bias, minimum percentage error) constrained optimization solution. The ZPB/MPE method finds coefficients such that the resulting CER has smallest possible percentage error, ***subject to the constraint that*** its percentage bias be zero. We know that this *cannot* lead to any *larger* percentage error than that provided by the MUPE/IRLS method, but it *may* lead to possibly *smaller* error. At worst, the error and bias will be the same as that given by MUPE/IRLS; at best, the error will be smaller and the bias will be closer to zero.

FIGURE 1. EXCEL SPREADSHEET BEFORE OPTIMIZATION.

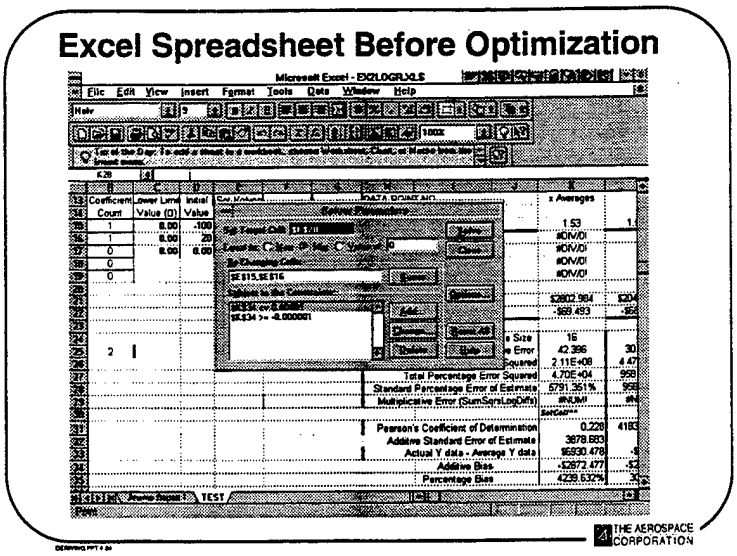
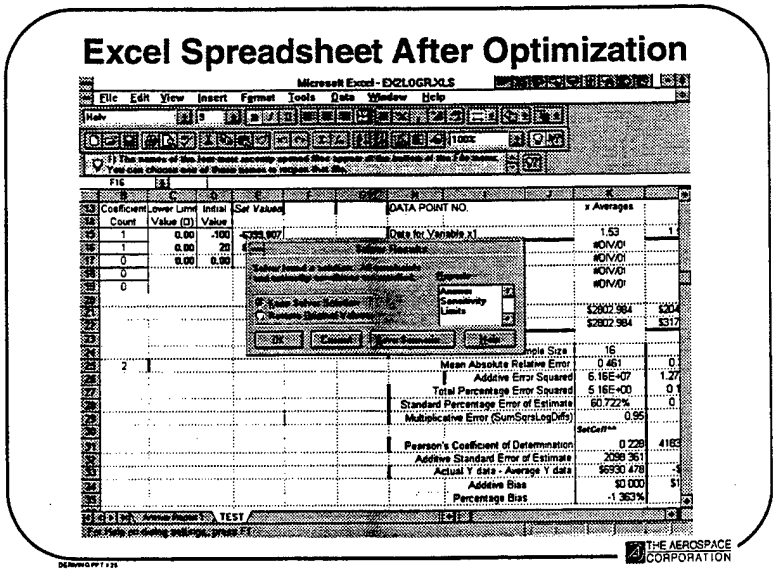


FIGURE 2. EXCEL SPREADSHEET AFTER OPTIMIZATION.



As illustrated in Figure 1, the Excel SOLVER routine has the capability to produce a constrained optimization solution. Note that the bias cell is constrained to zero, while the standard percentage error cell is minimized to calculate the coefficients of the ZPB/MPE CER.

Figure 2 illustrates the Excel screen after SOLVER has produced a constrained optimization solution. Compare the magnitudes of the standard percentage error and bias cells, respectively, in Figures 1 and 2.

Results of several case studies on the relationships between MPE, MUPE/IRLS, and ZPB/MPE CERs are reported in the final section of this report below. Note that, in every case, the ZPB/MPE solution is at least as good as the MUPE/IRLS solution in both the standard percentage error and the percentage bias categories, but falls short of the percentage error available with the unconstrained MPE solution. A understanding of the relative magnitudes of the errors being traded off is also provided by the various tables of results.

CASE STUDIES

Figure 3 lists three sets of data points that we will use to illustrate the standard error of the estimate and the bias of candidate CERs derived by all three methods described above. These three data sets are examples of the type that support the USCM-7 model. However, CER quality would be higher for actual USCM-7 CERs because they are not restricted to one cost driver as we are in this report. It is important to remember that the results provided here are not to be considered typical of USCM-7 CERs, but serve only to illustrate the relative quality obtainable by applying the ZPB/MPE procedure instead of the MPE and MUPE/IRLS techniques.

FIGURE 3. THREE SAMPLE DATA SETS.

Example 1 Data		Example 2 Data		Example 3 Data	
y	*x*	*y*	*x*	*y*	*x*
357.79	6.90	2045.42	38.59	134.96	4.18
823.70	11.79	1619.62	28.92	2.05	0.32
652.31	10.23	2079.58	23.30	5.35	0.57
278.81	6.74	918.85	21.11	64.64	2.34
1066.73	16.70	1231.13	17.54	32.85	0.50
437.44	8.05	3641.96	27.60	95.42	2.70
1219.83	23.46	1314.85	16.20	66.22	4.54
368.38	16.50	1128.39	34.89	112.23	4.42
		3989.48	46.61	29.24	0.55
		3130.08	65.90	123.09	0.79
		376.47	14.63	28.66	0.20
		9028.31	50.10	16.93	0.80
		2786.09	38.10	218.20	2.40
		2497.71	73.21		
		2051.06	64.81		
		7008.74	41.60		

Figures 4 and 5 list, respectively, standard percentage errors of the estimate and bias of each of six CERs of different forms optimally selected by each of the three methods we have been discussing using Example 1 data.

FIGURE 4. PERCENTAGE ERRORS OF CER FORMS FIT TO EXAMPLE 1 DATA.

FUNCTION	MPE	MUPE (IRLS)	ZPB/MPE	ZAB/MPE
$y = bx$	27.814%	28.806%	28.806%	29.143%
$y = a + bx$	29.539%	31.040%	30.555%	31.464%
$y = a + b \log x$	26.884%	28.268%	27.644%	28.204%
$y = bc^x$	34.135%	35.793%	35.732%	35.904%
$y = bx^c$	29.932%	31.270%	30.992%	31.430%
$y = a + bx^c$	29.839%	30.438%	29.994%	31.341%

Note that the ZPB/MPE percentage error never exceeds the corresponding MUPE/IRLS percentage error and is sometimes quite a bit smaller. Take note also of the nonzero bias of the MPE CERs and the increase in standard error required to bring the bias to zero. For purposes of comparison only, we have calculated and listed in the far-right column the percentage error and bias of the MPE CER that is constrained to zero *additive* bias, the so-called ZAB/MPE CER. This CER will always have larger percentage error than the ZPB/MPE CER, as well as nonzero percentage bias.

FIGURE 5. PERCENTAGE BIAS OF CER FORMS FIT TO EXAMPLE 1 DATA.

FUNCTION	MPE	MUPE (IRLS)	ZPB/MPE	ZAB/MPE
$y = bx$	6.770%	0.000%	0.000%	-1.090%
$y = a + bx$	6.551%	0.000%	0.000%	-1.439%
$y = a + b \log x$	5.415%	0.000%	0.000%	-1.128%
$y = bc^x$	8.739%	0.000%	0.000%	0.393%
$y = bx^c$	6.720%	0.000%	0.000%	-0.479%
$y = a + bx^c$	5.501%	0.000%	0.000%	-1.177%

Figures 6 and 7 list, respectively, the standard percentage error of the estimate and the bias of each of six CERs of different forms optimally selected by each of the three methods we have been discussing using Example 2 data.. As before, the ZPB/MPE percentage errors never exceed the corresponding MUPE/IRLS percentage errors and are sometimes quite a bit smaller. Note also the nonzero bias of the MPE CERs and the increase in standard error required to bring the bias to zero.

FIGURE 6. PERCENTAGE ERRORS OF CER FORMS FIT TO EXAMPLE 2 DATA.

FUNCTION	MPE	MUPE (IRLS)	ZPB/MPE	ZAB/MPE
$y = bx$	53.851%	63.109%	63.109%	63.866%
$y = a + bx$	52.258%	61.390%	59.902%	65.638%
$y = a + b \log x$	53.275%	60.978%	58.181%	60.722%
$y = bc^x$	56.791%	71.589%	67.032%	74.124%
$y = bx^c$	53.321%	64.011%	61.519%	66.108%
$y = a + bx^c$	53.125%	61.270%	60.461%	63.478%

FIGURE 7. PERCENTAGE BIAS OF CER FORMS FIT TO EXAMPLE 2 DATA.

FUNCTION	MPE	MUPE (IRLS)	ZPB/MPE	ZAB/MPE
$y = bx$	27.197%	0.000%	0.000%	-1.182%
$y = a + bx$	23.892%	0.000%	0.000%	-3.891%
$y = a + b \log x$	19.080%	0.000%	0.000%	-1.364%
$y = bc^x$	28.222%	0.001%	0.000%	-0.491%
$y = bx^c$	24.876%	0.000%	0.000%	-1.212%
$y = a + bx^c$	22.864%	0.010%	0.000%	-1.671%

Figures 8 and 9 list, respectively, the standard percentage error of the estimate and the bias of each of six CERs of different forms. optimally selected by each of the three methods we have been discussing using Example 3 data.. Again, the ZPB/MPE percentage errors never exceed the corresponding MUPE/IRLS percentage errors and are sometimes quite a bit smaller. Continue to note the nonzero bias of the MPE CERs and the increase in standard error required to bring the bias to zero.

FIGURE 8. PERCENTAGE ERRORS OF CER FORMS FIT TO EXAMPLE 3 DATA.

FUNCTION	MPE	MUPE (IRLS)	ZPB/MPE	ZAB/MPE
$y = bx$	69.711%	93.878%	93.878%	134.220%
$y = a + bx$	68.176%	88.529%	87.527%	82.492%
$y = a + b \log x$	63.393%	78.072%	78.034%	73.032%
$y = bc^x$	69.578%	95.767%	90.554%	85.440%
$y = bx^c$	65.260%	83.208%	81.599%	76.095%
$y = a + bx^c$	66.426%	87.059%	82.786%	76.605%

FIGURE 9. PERCENTAGE BIAS OF CER FORMS FIT TO EXAMPLE 3 DATA.

FUNCTION	MPE	MUPE (IRLS)	ZPB/MPE	ZAB/MPE
$y = bx$	44.866%	0.000%	0.000%	-36.971%
$y = a + bx$	39.347%	0.000%	0.000%	6.785%
$y = a + b \log x$	34.018%	0.000%	0.000%	7.564%
$y = bc^x$	40.963%	0.000%	0.000%	5.962%
$y = bx^c$	36.037%	0.000%	0.000%	8.070%
$y = a + bx^c$	33.940%	0.000%	0.000%	7.557%

SUMMARY

General-error regression separates the problem of whether estimating error should be additive (expressed as a uniform dollar value across the board) or multiplicative (expressed as a percentage of the estimate) from the problem of whether the functional relationship is linear or nonlinear. It turns out, though, that relationships derived by minimizing percentage (i.e., multiplicative) error contain significant positive bias (i.e., they tend to overestimate the actual values of the dependent variable). Because functional relationships cannot be optimized with respect to more than one criterion, the analyst must decide whether to insist upon minimum possible percentage error or to accept an increase in percentage error in trade for a reduction in bias. The method of iteratively reweighted least squares appears to be useful in this respect, but the fact that it does not seem to be optimal in any particular respect leaves room for a more intellectually satisfying solution. In this report the general-error regression problem is formulated as a constrained nonlinear optimization problem, with percentage standard error of estimation optimized (i.e., minimized), subject to percentage bias being zero. Naturally, the percentage error turns out to be somewhat larger than it would be if the bias were unconstrained, but in general not as large as that given by iteratively reweighted least squares. In short, zero bias is achieved with the smallest possible increase in standard error.

ACKNOWLEDGMENTS

The authors would like to thank Jonathan F. Binkley for his detailed study of the mathematics of bias vs. standard error and his review of the iteratively reweighted least squares method. We would also like to thank Richard H. Lucas for his development of a PC-compatible C++ stand-alone piece of software that implements general-error regression and gives us a validation capability independent of Excel. We would also like to thank Marie Calamaria and Sandra E. McCarthy for their excellent work in typing this technical material.

REFERENCES

1. Bickel, P.J. and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, Inc., 1977 pages 132-141.
2. Book, S.A. and P.H. Young, "Optimality Considerations Related to the USCM-6 'Ping Factor'," ICA/NES National Conference, Los Angeles, CA, 20-22 June 1990, 40 briefing charts.
3. Bradu, D. and Y. Mundlak, "Estimating in Lognormal Linear Models," *Journal of the American Statistical Association*, Vol. 65 (March 1970), pages 198-211.
4. Draper, N.R. and H. Smith, *Applied Regression Analysis*, (2nd Edition), New York: John Wiley, 1981, pages 43-45, 90-98.
5. Eskew, H.L. and K.S. Lawler, "Correct and Incorrect Error Specifications in Statistical Cost Models," *Journal of Cost Analysis*, Spring 1994, pages 105-123.
6. Eskew, H.L., "Tutorial on Log-Linear Regression," *National Estimator*, Spring 1994, pages 10-13.
7. Farnum, N.R., "Improving the Relative Error of Estimation," *The American Statistician*, Vol. 44 (November 1990), pages 288-289.
8. Gallant, A.R., *Nonlinear Statistical Models*, John Wiley & Sons, 1987, page 473.
9. Heien, D.M., "A Note on Log-Linear Regression," *Journal of the American Statistical Association*, Vol. 63 (September 1968), pages 1034-1038.
10. Hu, S.P. and A.R. Sjovold, "Error Corrections for Unbiased Log-Linear Least Square Estimates," Tecolote Research, Inc., Report No. TR-006/1, October 1987, page 51.
11. Hu, S.P. and A.R. Sjovold, "Multiplicative Error Regression Techniques," Tecolote Research, Inc., 1994, 31 pages. (Also available as Appendix C of Reference 17.)
12. Jennrich, R.I. and R.H. Moore, "Maximum Likelihood Estimation by Means of Nonlinear Least Squares," American Statistical Association, *Proceedings Statistical Computing Section*, 1975, pages 57-65.
13. Miller, D.M., "Reducing Transformation Bias in Curve Fitting," *The American Statistician*, Vol. 38 (May 1984), pages 124-126.
14. Neyman, J. and E.L. Scott, "Correction for Bias Introduced by a Transformation of Variables," *Annals of Mathematical Statistics*, Vol. 31 (September 1960), pages 643-655.
15. Nguyen, P., N. Lozzi, W. Galang, et al., *Unmanned Space Vehicle Cost Model, Seventh Edition*, U.S. Air Force Space and Missile Systems Center (SMC/FMC), 2430 E. El Segundo Blvd., Suite 2010, Los Angeles AFB, CA 90245-4687, August 1994, xvi + 452 pages.
16. SAS Institute Inc., *SAS/Statistical User's Guide, Volume 2*, 1990, pages 1135-1193.
17. Seber, G.A.F, and C.J. Wild, *Nonlinear Regression*, John Wiley & Sons, 1989, pages 37, 46, 86-88.

GAUSS'S ADJUSTMENT ERRORS IN 1799

Aivars Celmiņš
U.S. Army Research Laboratory
Aberdeen Proving Ground, Maryland 21005-5067

ABSTRACT

One of the major controversies about priority claims in the mathematical community of the 19th century was about the invention of the method of least squares. The method was first published and named by Legendre in 1805. On the other hand, in a publication in 1809, Gauss asserted that he had used the method since 1795 and insisted on his claim in spite of not having earlier publications to prove it. We have, however, two sets of data that were published in 1799 and that Gauss had adjusted by what he then called "my method." In this paper, the adjustments of the data are repeated to determine whether Gauss's "my method" was indeed the least-squares method. The results are found inconclusive because Gauss's published adjustments apparently contain arithmetical errors.

INTRODUCTION

The controversy about the discovery of the method of least squares is described in detail by Plackett,[1] Sprott,[2] Stigler,[3, 4] and Stewart.[5] It might be summarized as follows. In 1805, Legendre (1752-1833) published a memoir *Nouvelles méthodes pour la determination des comètes* in which he introduced and named the method of least squares. In 1809, Gauss (1777-1855) published in 1809 a book *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*[6,7] where he discussed the method of least squares and, mentioning Legendre's work, stated that he himself had used the method since 1795. Legendre was offended by Gauss's statement and protested, first privately and then, in 1820, publicly, stating that claims of priority should not be made without proof by previous publications. Gauss did not have such a publication, but repeated his claim in 1821, additionally claiming that he had used the method almost daily, especially since 1801 (see Gauss,[5] p. 180). In 1831 Schumacher wrote to Gauss about a 1799 publication that contains data and adjustment results by Gauss. Schumacher suggested repeating the calculations and thereby demonstrating that the method of least squares was indeed used by Gauss in 1799. Gauss's answer was that he would not permit a recalculation, and that he furthermore opposed any more public testimony on his behalf; his word should be enough.

Schumacher's suggestion to repeat Gauss's calculations was taken up by Stigler.[4] He obtained the data in question and tried least-squares adjustments on them. He could not reproduce Gauss's results and hypothesized that Gauss might have used a constraint that is more accurate than the linearized one-term expansion of the constraint equation that was used by Stigler. In the present paper, we review the adjustment and conclude that the results published by Gauss certainly are not obtained by a minimization of observational errors in a least-squares sense nor by any other approach mentioned by Gauss. This raises the intriguing question as to what method or principle did Gauss use when calling it "my method."

THE ADJUSTMENT PROBLEM

The data in question are from the measurement of a meridian arc of the Earth. The measurements were made on behalf of the French Academy of Sciences with the purpose of establishing a standard for a new length unit, "*metre*," as one 10^7th part of the quadrant of the meridian arc of the Earth. The measurements consisted of astronomical determinations of latitudes along a meridian from Dunkirk to Barcelona and land surveying between the latitude observations. The results of the measurements are listed in Table 1. Originally, the data were published in *Allgemeine Geographische Ephemeriden* 4 (1799), page XXXV. Gauss reported his results in the same publication (p. 378) and

---

Approved for public release; distribution is unlimited.

Table 1. Original Data

No.	Location	S_i, modules	δ_i, degrees	Φ_i
1	Barcelona to Carcassone	52 749.48	1.852 66	42°, 17′, 20″
2	Carcassone to Evaux	84 424.55	2.963 36	44°, 41′, 48″
3	Evaux to Pantheon	76 145.74	2.668 68	47°, 30′, 46″
4	Pantheon to Dunkirk	62 472.59	2.189 10	49°, 56′, 30″
Totals		275 792.36	9.673 80	—

Note: S_i are the distances between the indicated locations, δ_i are the corresponding differences in latitudes, and Φ_i are the latitudes of the midpoints of the distances. The distance S_3 between Evaux and Pantheon was due to a printer's error originally given as 76 545.74 modules.

added a comment in the Corrections to Vol. 4 of the *Allgemeine Geographische Ephemeriden* (1800) (p. 193). The originally published data contained a printer's error, and Gauss asserted that he had used his method (that was not explained) on both sets, with and without the error. His values for the ellipticity f of the meridian ellipse and the length Q of the quadrant are as follows.

- Data without error ..... f = 1/187 and Q = 2 565 006 modules.
- Data with printer's error ..... f = 1/50.

(One module = 1/1000 league ≈ 3.898 m.)

Gauss also reported in his comment that the ellipticity found by French surveyors was f = 1/150. (The method used by the French is not known.) Using the exact constraint equations and a simultaneous adjustment of all observations one obtains the following least-squares results.

- Data without error ..... f = 1/152 and Q = 2 564 897 modules.
- Data with printer's error ..... f = 1/79 and Q = 2 568 230 modules.

Assuming that the meridian is an ellipse and that the latitude is defined by the elevation angle of the normal to the ellipse, one has the following relation between an arc length S and the latitudes Λ_S and Λ_E of its end points:[8]

$$S = A \int_{\Lambda_S}^{\Lambda_E} (1 - B \sin^2\phi)^{-3/2} \, d\phi \, , \tag{1}$$

where A and B are constants. These constants can be determined by a least-squares adjustment of the data listed in Table 1 with equation (1) as constraint. After determination of the values of A and B, the ellipticity and the length of the quadrant can be computed as follows. Let a and b be the semimajor and semiminor axis of the ellipse, respectively, and let f be its ellipticity. Then $A = b^2/a$, $B = 1 - (b/a)^2$, and

$$f = (a - b)/a = 1 - \sqrt{1 - B} \, . \tag{2}$$

The length, Q, of the quadrant is given by the integral

$$Q = A \int_0^{\pi/2} (1 - B \sin^2\phi)^{-3/2} \, d\phi \, . \tag{3}$$

The integral in equation (1) cannot be evaluated in closed form. Using modern computers, this poses no problem since the integral can be computed numerically. (We used a Romberg quadrature algorithm for our numerical calculations.) But in the 1790's, an approximate expression of equation (1) was likely used in the adjustment process. Because, in our case, $B \ll 1$ and $\Lambda_E - \Lambda_S < 3°$, a good approximation of S is

$$S \approx (\Lambda_E - \Lambda_S)\, A \left[1 + \frac{3}{2} B \sin^2\left((\Lambda_S + \Lambda_E)/2\right) \right] . \tag{4}$$

This one-term approximation of equation (1) is also suggested by the form in which the data were published. The entries in Table 1 are the arc lengths S; the differences $\delta = \Lambda_E - \Lambda_S$ between the latitudes of the end points; and the midpoint latitudes $\Phi = (\Lambda_S + \Lambda_E)/2$. If one treats the midpoint latitudes Φ as fixed parameters, then the equation (4) is a linear constraint equation for the observations S and δ.

The exact constraint of equation (1) can be approximated also by more sophisticated formulas than equation (4) and Stigler,[4] after finding by numerical experimentation that Gauss did not use the linearized form of equation (4), suggested that Gauss had a better approximation to the exact constraint. If this were true, then an adjustment based on the exact constraint of equation (1) should be closer to Gauss's solution than to an adjustment with the approximate constraint of equation (4). We shall test this property of the solution by computing several variants of adjustments based on exact constraints. We need several variants because, even with a given constraint equation, one can adjust, for instance, only the surveyed arc lengths S or only the observed latitudes Λ or both with appropriate weights.

Also missing are estimates of data accuracies that might have been used by Gauss for the computation of adjustment weights. In particular, one would normally assume that the standard deviations of the arc lengths S_i are proportional to $\sqrt{S_i}$, but we are not at all sure that Gauss made such an assumption. Moreover, if one simultaneously adjusts the arc lengths S_i as well as the latitudes Λ_j then one needs prior estimates of the standard deviations of all data. Fortunately, assumptions about data accuracies are not essential for the present investigation because they do not greatly influence the values of the fitted constants A and B.

PROBLEM FORMULATIONS

In this section, we describe three formulations of the adjustment problem that were used in our calculations. The corresponding numerical solutions were obtained with utility routines described in Celmiņš.[9] Those routines solve constrained least-squares problems that are defined as follows.

Minimize

$$W = \sum_{i=1}^{s} c_i^T P_i^{-1} c_i \tag{5}$$

subject to

$$F_i\ (X_i + c_i\,;\, T) = 0, \quad i = 1, \ldots, s\,, \tag{6}$$

where X_i are observed vectors with dim $(X_i) = n_i$, c_i are the corresponding least-squares corrections, P_i are estimated variance-covariance matrices of the observations X_i, T is a free model parameter vector with dim $(T) = p$, and F_i are constraint functions with dim $(F_i) = r_i$. The unknowns of the problem are the corrections c_i of the observations X_i and the parameter vector T. It is assumed that the constraint or model functions F_i are twice differentiable with respect to all their arguments and that

$$\sum r_i - \sum n_i < p < \sum r_i\,. \tag{7}$$

If the constraint functions F_i are scalar ($r_i \equiv 1$) then the utility routine COLSAC, ibid, can be used. If the constraints $F_i = 0$ contain sets of simultaneous equations for the c_i ($r_i > 1$), then the more complicated routine COLSMU must be used.

We have tried adjustments of the arc lengths as well as of the latitude observations. The adjustment of the latter can be somewhat simplified by expressing the constraints in terms of the end-point observations Λ_i themselves rather than in terms of the differences δ_i and midpoint latitudes Φ_i that are given in Table 1 because the differences and midpoint values are interdependent. (They are constrained by the condition that adjacent arcs must have common end points after adjustment.) We therefore reconstructed the observed end-point latitudes from the data in Table 1. The result is shown in Table 2, which also contains *a priori* estimates of the standard deviations of the observations. Such estimates are necessary for the joint adjustment of arc lengths and latitudes and they can be obtained by preliminary adjustments as described in Celmiņš.[8]

Table 2. Reconstructed Latitude Data

No.	Location	Latitude, degrees	Arc, modules	
		Λ_i	S_i	e_{Si}
1	Barcelona	41.362 42		
			52 749.48	22.16
2	Carcassone	43.215 08		
			84 424.55	28.03
3	Evaux	46.178 44		
			76 145.74	26.62
4	Pantheon	48.847 12		
			62 472.59	24.11
5	Dunkirk	51.036 22		

Note: The estimated standard errors of the latitudes are $e_\Lambda = 5.005 \cdot 10^{-4}$ degrees. The estimated standard error for the erroneous distance $S_3 = 76\,545.74$ is $e_{S3} = 26.69$ modules.

We now describe the adjustment processes for which we distinguish three cases.

CASE 1: ADJUSTMENT OF ARC LENGTHS

In this case, the adjustable data are the surveyed arc lengths S_i whereas the latitude observations Λ_j are treated as fixed nonadjustable constants. In terms of the problem formulation of equations (5) and (6) we have, therefore, the data (regressand variables)

$$X_i = S_i, \; i = 1, 2, 3, 4, \tag{8}$$

with the variance estimates (from Table 2)

$$P_i = e_{Si}^2, \; i = 1, 2, 3, 4 \; . \tag{9}$$

From the exact relation, equation (1), we obtain the following constraint equations for $i = 1, 2, 3, 4$:

$$F_i(S_i + c_{Si}; A, B) = S_i + c_{Si} - A \int_{\Lambda_i}^{\Lambda_{i+1}} (1 - B \sin^2\phi)^{-3/2} d\phi = 0, \tag{10}$$

where the Λ_j are fixed constants (regressor variables). Corresponding linearized constraints are for $i = 1, 2, 3, 4$:

$$L_i(S_i + c_{Si}; A, B) = S_i + c_{Si} - A \, \delta_i \left(1 + \frac{3}{2} B \sin^2 \Phi_i\right) = 0 \tag{11}$$

where $\delta_i = \Lambda_{i+1} - \Lambda_i$, and $\Phi_i = (\Lambda_{i+1} + \Lambda_i)/2$ are fixed constants (regressor variables). The parameters of the adjustment problem are the free constants A and B. The condition (7) is satisfied with $\sum r_i = 4$, $\sum n_i = 4$, and $p = 2$. Because the constraints are scalar, this problem can be solved using the utility program COLSAC with either the exact constraints (10) or the linearized constraints (11).

CASE 2: ADJUSTMENT OF LATITUDES

In this case, we adjust the latitude observations Λ_j and treat the surveyed arc lengths S_i as fixed numbers. Because the Λ_j enter the model equation (1) as limits of the arc-length integrals, the adjustable latitude observations Λ_2, Λ_3, and Λ_4 appear each in two of the four constraint equations, and the constraints must be treated as a single equation system $F_1 = 0$ of four simultaneous equations.

To cast the adjustment problem into the form of equations (5) or (6), we define the adjustable data (the regressand variables) as a single vector X_1 of five observations. That is, in equation (5) $s = 1$, and the data vector is

$$X_1^T = (\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4, \Lambda_5) \ . \tag{12}$$

The data variance matrix P_1 is a diagonal (5×5)-matrix with the diagonal elements e_Λ^2. The single constraint function $F_1 (X_1 + c_1 ; A, B)$ has four components f_i. If the exact relation (1) is used, then the components $f_i = 0$ of the constraint equation $F_1 = 0$ are as follows:

$$f_i = S_i - A \int_{\Lambda_i + c_{\Lambda,i}}^{\Lambda_{i+1} + c_{\Lambda,i+1}} (1 - B \sin^2\phi)^{-3/2} d\phi = 0, \ i = 1, 2, 3, 4. \tag{13}$$

In linearized form, the constraint equation has the components

$$l_i = S_i - (\Lambda_{i+1} + c_{\Lambda,i+1} - \Lambda_i - c_{\Lambda,i}) A (1 + \frac{3}{2} B \sin^2\Phi_i) = 0, \ i = 1, 2, 3, 4. \tag{14}$$

The arc lengths S_i and the midpoint latitudes Φ_i [in the linearized constraints (14)] are assumed to be fixed nonadjustable constants (regressor variables). The condition (7) is satisfied with $r_1 = 4$, $n_1 = 5$, and $p = 2$. This type of problem (with constraints in form of simultaneous equations) can be solved using the utility program COLSMU.

CASE 3: ADJUSTMENT OF ARC LENGTHS AND LATITUDES

In this case, all observations, the surveyed arc lengths S_i as well as the latitude observations Λ_j are adjusted simultaneously. The problem can be solved by treating the arc lengths S_i in the constraint equations (13) or (14) as adjustable observations and using the utility program COLSMU for constraints in the form of simultaneous equations. Then the corresponding vector of observations X_1 would have nine components (five Λ_j and four S_i). The variance matrix P_1 of the single observation vector X_1 would be a diagonal (9×9) matrix, and the constraint $F_1 = 0$ would again be a system of four simultaneous equations. However, the numerical treatment and the coding of the problem can be simplified by introducing nonessential parameters[10] that render the problem separable and transform the constraints into a set of nine independent scalar equations. For details of the application of the technique to this problem see Celmiņš.[8]

LEAST-SQUARES RESULTS

The results of adjustments using the least-squares method are listed in Tables 3 and 4 and shown in Figures 1 and 2. Table 3 lists six adjustment results (Cases 1, 2, and 3, with exact and linearized constraints, respectively) giving the values of the arc length Q; the inverse ellipticity $1/f$; their corresponding estimated standard deviations e_Q and $e_{1/f}$, respectively; and an estimate of the correlation coefficient between Q and $1/f$. For the adjustments involving only the arc lengths S_i (Case 1), we assumed that their standard deviations are proportional to $\sqrt{S_i}$. The adjustments of the Λ_i only (Case 2) were made assuming that the standard deviations of the data are all equal. The adjustment weights in Case 3 were calculated from the standard error estimates listed in Table 2.

Table 3. Least-Squares Results From Correct Data Set

Constraint	Observation	Q	e_Q	$1/f$	$e_{1/f}$	Correlation
Exact	S	2 564 891	501	155.3	31.3	0.338 477
	S and Λ	4 897	544	151.8	36.8	0.380 900
	Λ	4 909	198	148.7	20.2	0.576 044
Linear	S	5 100	479	152.7	31.3	0.174 293
	S and Λ	5 116	513	149.3	36.8	0.192 755
	Λ	5 138	170	146.2	20.2	0.299 832

Table 4. Least Squares Results From Data Set With Error

Constraint	Observation	Q	e_Q	$1/f$	$e_{1/f}$	Correlation
Exact	S	2 568 000	491	87.8	9.9	0.281 788
	S and Λ	8 230	528	79.3	10.0	0.303 993
	Λ	8 523	183	72.7	4.8	0.457 031
Linear	S	8 682	472	85.0	9.9	−0.043 977
	S and Λ	9 067	507	76.6	9.9	−0.116 073
	Λ	9 525	171	70.0	4.8	−0.320 064

Figure 1 shows the results of adjustments for the correct data set. The figure displays the six pairs of values of Q and $1/f$ that are listed in Table 3. To illustrate the accuracy of the adjustment, we have also plotted one standard deviation error ellipses for the three adjustments with exact constraints. The dashed curve corresponds to distance adjustments, the dot-dash curve corresponds to latitude adjustments, and the solid curve represents the standard deviation in the case where all data are adjusted simultaneously. Gauss's result is about one standard deviation apart from all of our results. The difference is not important statistically, but it indicates that the values reported by Gauss are not obtained by a least-squares adjustment, regardless of whether or not the linearized constraint or the exact constraint has been used.

Next, we consider the data set that contains the printer's error. The adjustment results are listed in Table 4 and displayed in Figure 2. One observes that differences among the six results are larger than in Figure 1, but the overall situation is about the same as shown in Figure 1. In this case, Gauss did not report a value for the quadrant length Q and we can only compare the line $1/f = 50$ (Gauss's value) with our results. The line is well below any of our results.

Figure 3 is a combined display of all least-squares results. The error ellipses correspond to one standard deviation, as before, and are for the simultaneous adjustments of all data. The figure shows that the quadrant length and the inverse ellipticities that were reported by Gauss are not obtained by least-squares adjustments. A printer's error in the results reported by Gauss is not likely because Gauss's ellipticity values were published twice, in two different issues of the journal. Thus, we are left with the question as to whether or not Gauss used a different adjustment principle or made an arithmetical error.

ADJUSTMENTS USING DIFFERENT PRINCIPLES

Candidates for adjustment principles that might have been used by Gauss are the minimization of the sum of nth powers of the absolute values of residuals, Boscowich's (1711-1787) method, and a minimization of the maximum deviation. (Boscowich's method consisted of a minimization of the sum of absolute values of the residuals under the condition that the sum of the residuals should be equal to zero.) Gauss discusses all these methods in Article 186 of *Theoria Motus*[6] and suggests to use least squares on grounds of numerical expediency.

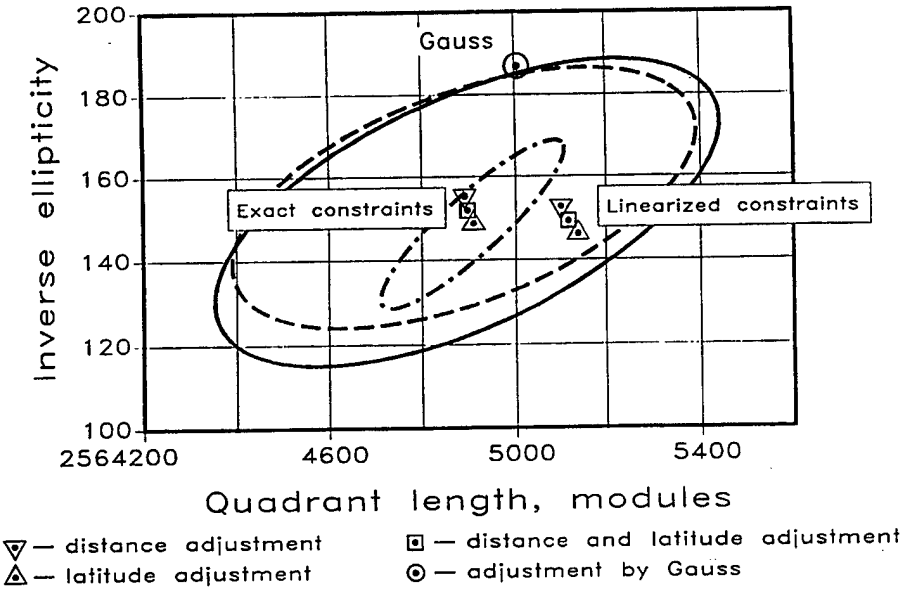


Figure 1. Adjustments of correct data.

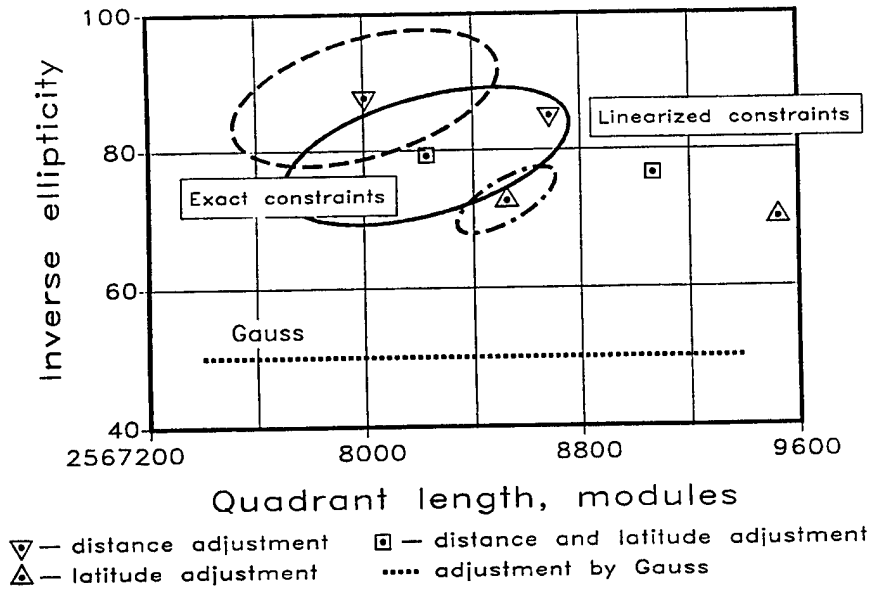


Figure 2. Adjustments of data with error.

A further method that might have been used by Gauss is suggested by Sheynin.[11] In that method, a least-squares technique is used to minimize an objective function in the parameter space. The method has an *ad hoc* nature and it is not considered by Gauss in *Theoria Motus*,[6] but Sheynin asserts that the method has been widely used in land surveying during the past two centuries.

To get an idea about the range of results that can be obtained with these different adjustment methods, we carried out a number of adjustments of the arc length measurements S_i equally weighted, and using the linearized constraint (11). The solutions were obtained by a numerical search for the minimum of the respective objective function. Some

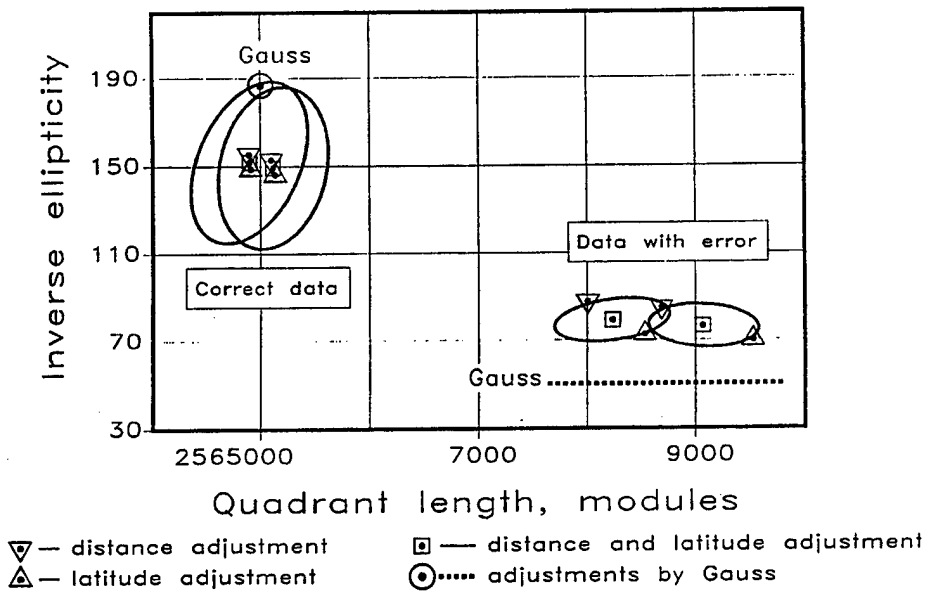


Figure 3. All least-squares adjustments.

typical results are displayed in Figure 4. The results for the correct data show that the quality and consistency of the data are so good that the adjustment method does not matter: all methods produce very similar results and all are different from Gauss's result. On the other hand, adjustments of the data with the printer's error produce parameters that vary over a large range as the power n of the residuals varies between unity and infinity. Boscowich's method produces solutions that are, in both cases, close to the corresponding least-squares solutions, and so does the least-squares technique suggested by Sheynin (not shown in Figure 4). Gauss's results are found to be different from all other results.

DISCUSSION

The numerical results presented in the two previous sections suggest that Gauss's results are not consistent with any obvious and reasonable adjustment of observational errors nor with a least-squares adjustment in the parameter space. This leaves three possible explanations for the strange values.

(1) Gauss used a relation different from equation (1) as a basis for his analysis.

(2) Gauss made an error in simplifying the exact constraint in equation (1).

(3) Gauss's computations contain arithmetical errors.

We now discuss these possibilities in turn.

A relation different from equation (1) is obtained if the latitude is differently defined, for instance, as the elevation angle of the plumb line to a solid ellipsoid, or as the elevation angle of the ray from the center of the ellipsoid. Corresponding constraint equations are derived in Celmiņš.[8] Let f_C, f_N, and f_P be the ellipticities that correspond to latitude definitions in terms of the center ray, the normal to the ellipsoid, and the plumb line to the ellipsoid, respectively. Then, one obtains, with a linearized unweighted least-squares adjustment of the arc lengths S_i the results in Table 5. If one uses the center-ray definition of the latitudes, then the inverse ellipticity $1/f_C$ is less than $1/f_N$ for both data sets. If one uses the plumb-line definition, then $1/f_P$ is larger than $1/f_N$ for both data sets. Gauss's value is higher than $1/f_N$ for the correct data and lower for the erroneous data set. Hence a change of the definition of latitudes that reduces the

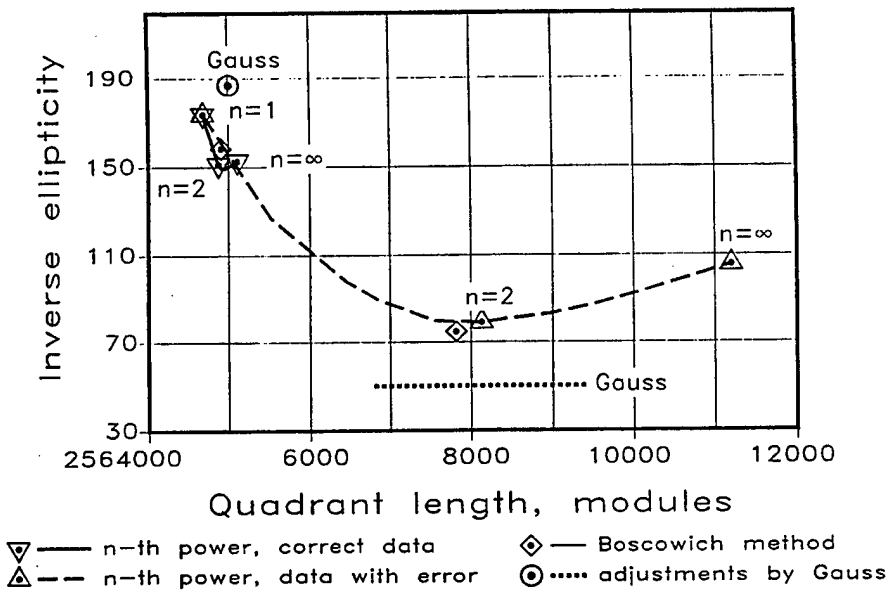


Figure 4. Adjustments by different methods.

difference between Gauss's values and our $1/f_N$ for one data set increases the difference for the other data set. Therefore, neither of the two alternative definitions considered can explain the discrepancies.

Table 5. Results for Different Latitude Definitions

Data Set	$1/f_C$	$1/f_N$	$1/f_P$	Gauss
Correct Distances	49.2	148.7	168.6	187.0
Distances with Error	25.2	76.5	86.8	50.0

An error in the simplification of the exact constraint equation (1) cannot be excluded, except for the reason that the equation and corresponding analyses are so simple that it is difficult to make an error.

Arithmetical errors seem, at first, unlikely because both results by Gauss are erroneous, suggesting at least two errors. However, this need not be the case. The calculations by Gauss were done manually, writing down intermediate results, such as the values of trigonometric functions and logarithms. Then, as the problem was solved again with a corrected value of the distance S_3, only those parts had to be recalculated that directly involved the new datum. An error in a quantity that was not recalculated would influence both results. We tested this possibility by assuming that one of the four values of sin Φ_i was in error. By a proper choice of the value of sin Φ_3, we obtained $1/f = 187$ for the correct data set and a corresponding $1/f = 64$ for the data set with printer's error. This does not exactly duplicate Gauss's result, but it does show that a single error can indeed increase the ellipticity in one case and reduce it for the other data set.

We conclude from these considerations that the results published by Gauss likely contain arithmetical errors. Hence Gauss's publication neither supports nor falsifies his claim that he used the method of least squares before 1800. As Gauss suggested, we have to trust his word.

REFERENCES

1. Plackett, R. L. "The Discovery of the Method of Least Squares." Biometrica, vol. 59, pp. 239–251, 1972.

2. Sprott, D. A. "Gauss's Contributions to Statistics." Historia Mathematica, vol. 5, pp. 183–203, 1978.

3. Stigler, S. M. "An Attack on Gauss Published by Legendre in 1820." Historia Mathematica, vol. 4, pp. 31–35, 1977.

4. Stigler, S. M. "Gauss and the Invention of Least Squares." The Annals of Statistics, vol. 9, pp. 465–474, 1981.

5. Gauss, C. F. Theoria Combinationis Observationum Erroribus Minimis Obnoxiae. Original 1821–1826, Reprint with translation and afterword by G. W. Stewart, SIAM, Philadelphia, PA, 1995.

6. Gauss, C. F. Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium. Perthes und Besser, Hamburg, 1809.

7. Gauss, K. F. Theory of the Motion of the Heavenly Bodies Moving About the Sun in Conic Sections. Translation of Gauss's Theoria Motus by C. H. Davis, Little, Brown, and Company, 1857, reprinted by Dover Publications, New York, NY, 1963.

8. Celmiņš, A. "The Method of Gauss in 1799." Statistical Science, vol. 13, 123–135, 1998.

9. Celmiņš, A. "A Manual for General Least Squares Model Fitting." BRL-TR-02167, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, 1979.

10. Celmiņš, A. "Least Squares Optimization With Implicit Model Equations." Mathematical Programming With Data Perturbations II, pp. 131–152, A. V. Fiacco (editor), Marcel Dekker, New York, NY, 1982.

11. Sheynin, O. "On the History of the Principle of Least Squares." Archive for History of Exact Sciences, vol. 46, pp. 39–54, 1993.

Constructing Bayesian Networks from WordNet for Word-Sense Disambiguation: Representational and Processing Issues *

Janyce Wiebe and **Tom O'Hara**
Department of Computer Science and
Computing Research Laboratory
New Mexico State University
Las Cruces, NM 88003
wiebe, tomohara@cs.nmsu.edu

Rebecca Bruce
Department of Computer Science
University of North Carolina at Asheville
Asheville, NC 28804-3299
bruce@cs.unca.edu

Abstract

This paper describes a probabilistic model that is formed from the integration of an analytical and empirical component. The analytical component is a Bayesian network derived from WordNet, and the empirical component is composed of compatible probabilistic models formulated from tagged training data. The components are integrated in a formal, uniform framework based on the semantics of causal dependence. The paper explores various representational issues that must be addressed when formulating a Bayesian network representation of lexical information such as that expressed in WordNet. These issues are essential to the design of such a network and they have not been previously explored. We describe two choices for the representation of lexical items and two choices for the representation lexical relations. The effect of each combination of choices on evidence propagation in the network is discussed.

INTRODUCTION

There is a long tradition in AI of resolving interdependent lexical ambiguities through spreading activation, from Quillian's (1968) seminal work on semantic networks, through Hirst's work (1988) on Polaroid words, to more recent work by Voorhees (1993) and Veronis and Ide (1990) on large-scale disambiguation. This research investigates a probabilistic realization of spreading activation to resolve interdependent word-sense ambiguities. The core idea is to exploit belief propagation in Bayesian networks: Words are mapped to nodes, lexical relations are mapped to edges, and evidence is propagated from word senses to other related word senses.

The lexical relations are derived from an existing knowledge source, because this information cannot be automatically extracted from training data with existing techniques. The knowledge source we use is the WordNet *is-a* hierarchy, i.e., the *hypernym/hyponym* taxonomy (Miller, 1990). Although this hierarchy was developed for other purposes, it has been frequently applied to word-sense disambiguation (Resnik, 1995; Sussna, 1993). In this work, we investigate various approaches to constructing a Bayesian network representation of the *is-a* hierarchy for use in word-sense disambiguation. As this work continues, other relations such as part/whole and entailment relations will also be included in the network.

Another contribution of our work is a novel proposal for integrating symbolic and statistical information for the purpose of performing NLP tasks. Statistical approaches to word-sense disambiguation have had the most success to date, when evaluated on unseen test data. The "analytical" Bayesian network component of our method is actually built on top of "empirical" probabilistic classifiers induced statistically from training data. In particular, an empirical classifier is induced for each word in the current sentence to be disambiguated (i.e., for each *target word*). Each empirical classifier is developed independently of the empirical classifiers for other target words. A Bayesian network is constructed from the segment of the WordNet *is-a* hierarchy that is connected to the target words. The results of the empirical classifiers are fed as evidence into the Bayesian network, thus initiating belief propagation. All of the information is represented in a formal, uniform framework: a probabilistic model embodying conditional independence relationships among the variables that form the joint distribution. Conditional independence relationships simplify the formulation of the joint distribution making it possible to work with a large number of variables. Further, models

* Approved for public release; distribution is unlimited. This is a reprint of Wiebe, Janyce, O'Hara, Tom, and Rebecca Bruce (1998). Constructing bayesian networks from WordNet for word sense disambiguation: representation and processing issues. In *Proc. COLING-ACL '98 Workshop on the Usage of WordNet in Natural Language Processing Systems*, Association for Computational Linguistics, Montreal, Canada, August 16, 1998. This research was supported in part by the Office of Naval Research under grant number N00014-95-1-0776.

that characterize conditional independence relationships have desirable computational properties (e.g., see the discussion on decomposable models in (Pearl, 1988)). These properties form the basis of the evidence propagation scheme used for Bayesian networks discussed in the section labeled "Edge Direction and Belief Propagation". We also make use of these properties in formulating the empirical classifiers as described in (Bruce and Wiebe, 1994). Bayesian networks are a very rich and complex representational framework. They support easy integration of diverse information sources and form the basis for much of the current work on reasoning under uncertainty (Pearl, 1988).

This paper explores the representational issues that must be addressed when mapping the lexical information in WordNet to a Bayesian network. The implications of the various choices are analyzed in depth. In the next section, we introduce the basic concepts and illustrate them with an example in the following section, which also includes a brief description of the empirical component. The Bayesian network representations of lexical items and lexical relations are then discussed. Then, we describe the integration of the empirical component into the Bayesian network, and the process of sense disambiguation. Finally, we discuss related work and conclude.

BAYESIAN NETWORKS: BACKGROUND

Bayesian networks model dependencies among nodes through the use of conditional probabilities. Specifically, if a node ($Cause2$) is considered as a cause for another node ($Symptom1$), then the second node is defined relative to the first (i.e., $P(Symptom1|Cause2)$). Some nodes don't have associated causes, so they are just defined via unconditional probabilities (e.g., $P(Cause2)$). Taken together, the set of all the conditional and unconditional probabilities determine a joint distribution for all the nodes being modeled (e.g., $P(Symptom1, ..., SymptomN, Cause1, ...CauseM)$). Such global distributions are usually difficult to assess directly; hence, the Bayesian network provides a convenient formalism for specifying the same distribution via local distributions, under conditional independence assumptions. Furthermore. without the conditional independence relations, the full joint distribution for cases with hundreds of senses would be infeasible to process—the independence assumptions are key. Pearl (1988) presents an in-depth coverage of the theory of Bayesian networks and provides an efficient algorithm for evaluating them.

In a Bayesian approach to statistical inference, we distinguish between *prior* and *posterior* probabilities. Prior probabilities express the beliefs that we hold about the likelihood of events *prior* to being given any evidence, *posterior* probabilities express our beliefs in the likelihood of events given all the evidence that is currently known. Thus, the posterior probability of an event changes as new evidence is learned. The conditional and unconditional probabilities mentioned above are the prior probabilities. The posterior probabilities are calculated using the Bayesian network propagation algorithm each time new evidence is added.

We discuss propagation in greater detail in the section labeled "Edge Direction and Belief Propagation". Intuitively, the posterior probability of a node, say the node GATHERING#1 (switching to a word-sense disambiguation example), is a combination of the beliefs received from its children and the beliefs received from its parents. Once a node has calculated its own belief, it calculates outgoing messages to send to its parents and to its children, which enable them, in turn, to calculate their posterior probabilities. In this way information is propagated throughout the network.

AN EXAMPLE

In this section, we illustrate how a simple Bayesian network can be constructed to model the interdependencies among words. This identifies the basic steps in the overall process and helps to motivate the representational issues discussed later.

Suppose that the words "community" and "town" appear in a single sentence, and that their correct senses in that context are COMMUNITY#1 and TOWN#2, respectively. Our task is to assign the correct word senses to both of them, considering information automatically derived from the corpus and gathered individually for each word, as well as information derived from the WordNet *is-a* hierarchy and represented in a Bayesian network. The basic strategy is to add the corpus-derived information to the Bayesian network representations of "community" and "town," in such a way that it initiates propagation.

Let us consider this process in more detail. The words "community" and "town" have the following senses in WordNet:

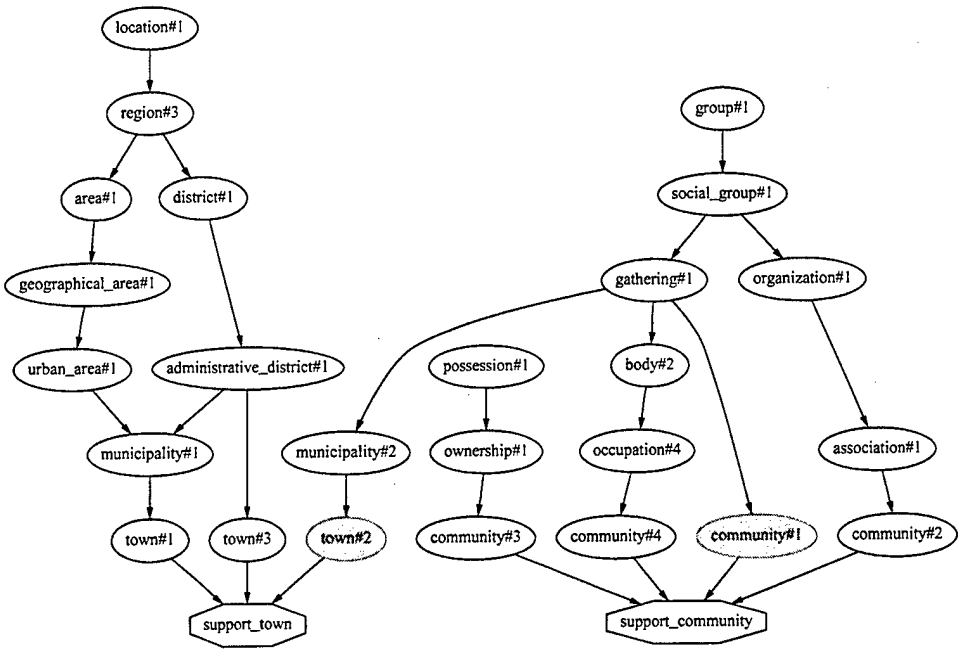


Figure 1: Sense per node Bayesian network with hypernym→hyponym links

community:
1. people living in a particular local area
2. an association of people with similar interests
3. common ownership
4. the body of people in a learned occupation

town:
1. an urban area with a fixed boundary that is smaller than a city
2. the people living in a municipality smaller than a city
3. an administrative division of a county

These senses are represented as sets of synonyms, or *synsets*. In the *is-a* hierarchy, each synset is linked to its *hypernym*, i.e., the synset representing its conceptual parent. For example, the synset corresponding to

{occupation, vocation, occupational group}

is the hypernym of the synset corresponding to

{profession, community}.

A new Bayesian network is created for each sentence. It includes all of the synsets for the target words in the sentence, together with all of the synsets reachable from them in the WordNet *is-a* hierarchy. Extracting this information from WordNet is straightforward.

Figure 1 illustrates one way that the Bayesian network for the example sentence containing "town" and "community" can be constructed. In this representation, each word sense is mapped to a node in the network, and there is an edge from X to Y iff word sense X is a hypernym (i.e., a superordinate) of word sense Y (please ignore the octagonal nodes at the bottom for now). Notice that the relation between COMMUNITY#1 and TOWN#2 is mediated by GATHERING#1, a type of GROUP#1. Our goal is for the contextual evidence provided by the empirical classifiers to propagate along this path in such a way that the correct senses of the target words reinforce one another.

After the topology of the network has been established, the conditional probability tables required for each node must be defined. As will be discussed later, we can make independence assumptions that make estimating the necessary probabilities more easier.

Next, an empirical classifier is developed for each ambiguous word, in this case, "town" and "community".

Each classifier defines a probability distribution describing the likelihood of each sense of the targeted word given the automatically derived features of the context. An example of the type of feature used is the part-of-speech of the word to the right; see (Bruce and Wiebe, 1994) for the other ones we use.

The distributions determined by the empirical classifiers are added as evidence to the Bayesian network, initiating belief propagation. Once the network reaches equilibrium, the posterior probabilities of the nodes for "town" and "community" determine the senses assigned to each ambiguous word.

REPRESENTING LEXICAL ITEMS: WHAT DOES A NODE MEAN?

There are two basic approaches to representing WordNet synsets in a Bayesian network. Since the lexical relations are among synsets and not words, a natural approach is to represent the synsets as nodes. Alternatively, one node could be used to represent all senses of a word.

THE ONE NODE PER WORD APPROACH

When nodes correspond to words, the possible values for each node are *sense*0 through *sense*N, where N is the number of WordNet synsets representing senses of the target word. *Sense*0 represents the composite of all other meanings, i.e., of all meanings that are not represented by WordNet synsets. Figure 2 shows the graph for the Bayesian network when word nodes are used for the relations. It also illustrates the use of logical links, which are described in the next section. This involves more than just a change in link direction.

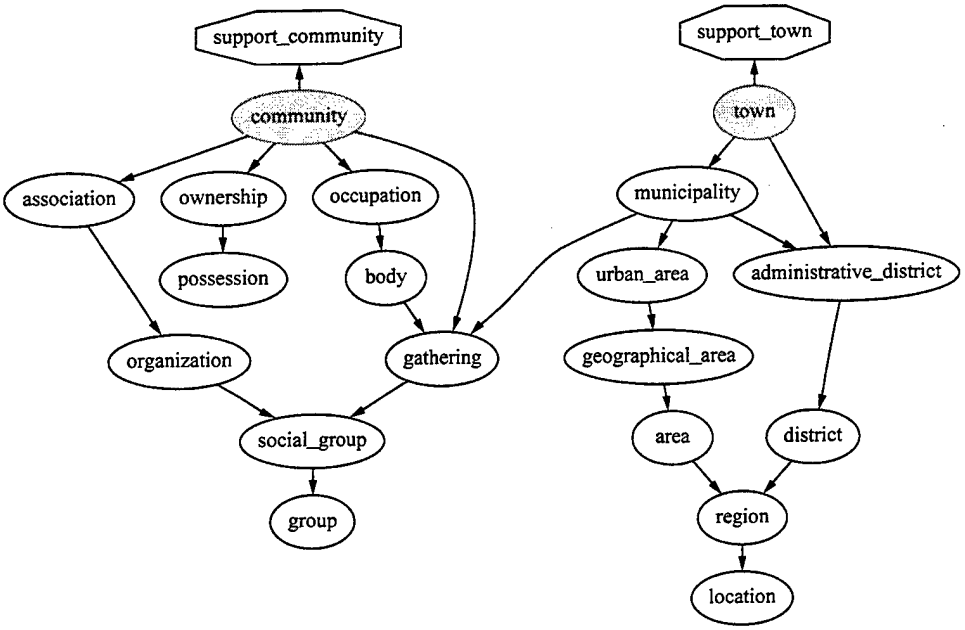


Figure 2: Word per node Bayesian network with hyponym→hypernym links

THE ONE NODE PER SENSE APPROACH

Figure 1 illustrates the approach in which each synset (each sense) is mapped to a node. An important advantage of using the node per sense approach is that it facilitates handling dependencies among the senses of a word. In the node per word approach, single node cycles are produced when modeling the dependencies of words that have a meaning that is defined in terms of other meanings for that same word.

A disadvantage of this approach is that modeling mutual exclusion among the senses of a single word becomes more difficult. The most straightforward approach modeling mutual exclusion is to create a dependency from each sense node to a separate node with a CPT enforcing mutual exclusion. But since the table must have 2^N entries, this approach becomes impractical for words with a large number of senses. To get

around this problem, two levels of mutual-exclusion dependencies could be introduced: one at which mutual exclusion among small groups of senses is enforced, and another enforcing mutual exclusion of the groups.

REPRESENTING LEXICAL RELATIONS: WHAT DOES AN EDGE MEAN?

Here, we address issues concerning the representation of WordNet *is-a* relationships as causal dependencies. The two primary issues to be addressed are: (1) expressing the Hypernym/Hyponym relationship as a causal dependency, and (2) quantifying the causal dependencies with conditional probability distributions.

HYPERNYM→HYPONYM REPRESENTATIONS.

The Hypernym → Hyponym Representation was illustrated above: there is an edge from node X to node Y iff X represents a hypernym of node Y in the WordNet *is-a* hierarchy. Consider the node per sense representation (see figure 1). Suppose $Hyper$ is a synset that is a hypernym of synsets $Hypo_1 \cdots Hypo_k$. Then, the relevant part of the Bayesian network expresses the following:

$$Hyper \rightarrow Hypo_1 \vee \cdots \vee Hypo_k$$

As such, we are making a closed world assumption. If, for example, there is a synset ANIMAL#1 with three hyponyms DOG#1, CAT#1, and MOUSE#1, we are assuming that these three are the only kinds of ANIMAL#1's there are.

When using this link representation with either of the node per sense or the node per word representations, the roots of the network are the most superordinate synsets reachable from the target words, and the target words are typically (but not necessarily) the leafs of the network.

We now turn to defining the CPT. We discuss this with respect to the node per sense representation (in figure 1) because it is easier to discuss and similar conditional probabilities must be defined under the node per word representation.

To define the CPT for each child node in the Bayesian network, where each child node corresponds to a hyponym node in WordNet, we assign the conditional probability $P(hyponym|hypernym)$ to be inversely proportional to the number of children that the hypernym has. For instance, MUNICIPALITY#1 has two children in WordNet, so we assign the following conditional probability for TOWN#1 given this hypernym.

P(town#1 | municipality#1)

municipality#1	P(town#1)
F	$0.000 + \epsilon$
T	0.500

In so doing we are: (1) considering each hyponym of a given hypernym to be equally likely, and (2) maintaining the closed world assumption by requiring that these conditional probabilities sum to one. In all CPTs, we add a small positive probability ϵ to all zero probability values in order to allow the realization of all possible configurations of node values (e.g., to handle inconsistent evidence). In future work, we will consider using frequency of occurrence information in tagged training data to define these CPTs.

For the root nodes, which represent the most superordinate concepts, prior probabilities must be specified. With no evidence to the contrary, uniform prior distributions are assigned to the root nodes; the empirical classifiers are relied upon to provide contextual support (through the leafs of the network).

HYPONYM→HYPERNYM REPRESENTATIONS

Under the Hyponym → Hypernym Representation, there is an edge from node X to node Y iff X represents a hyponym of node Y in the WordNet *is-a* hierarchy. Consider the node per sense representation (see figure 2). The Bayesian network represents the following:

$$(Hypo_i = s_i \rightarrow Hyper_j = s_j) \wedge \cdots \wedge (Hypo_n = s_n \rightarrow Hyper_m = s_m)$$

Under the semantics of the WordNet *is-a* hierarchy, all instances of a hyponym are instances of its hypernym. So, a typical CPT for this representation is as follows:

P(municipality#1 | town#1)

town	P(municipality#1)
F	0.0 + ϵ
T	1.0 - ϵ

Note that this case is not illustrated in the graphs shown: these only cover two of the four main possibilities.

Interestingly, in this representation, the root nodes represent the target words. Thus, the root nodes are the sites where evidence from the empirical classifiers is added to then network. In the absence of this evidence, these nodes take on their prior probabilities. As above, we assign uniform distributions as the priors. Recall that, in the case of multiple parents, CPTs must specify the conditional distribution of the child node given the values of all of its parent nodes. The issues involved in working with multiple parent nodes are discussed below.

CPT ENTRIES WHEN MULTIPLE PARENTS: CAUSAL INDEPENDENCE

If a node has multiple parents, say n parents, then specifying all of the entries in the CPT for that node can be prohibitive. If no additional independence assumptions are made regarding the interactions among the parent nodes, then the number of probabilities that must be specified is exponential in n, and probabilistic inference is made correspondingly more complex (Heckerman and Breese, 1994). To overcome this problem, the *noisy-OR* model (Pearl, 1988) is often adopted. Under this model, certain independence assumptions are made regarding the interactions among the parent nodes, with the effect that the number of probabilities that must be specified is linear in n. Basically, one need only specify the conditional probabilities of the child and each parent individually.

As presented in (Pearl, 1988), the noisy-OR model assumes that all of the variables are binary. Heckerman and Breese (1994) present a generalization of the noisy-OR model, *causal independence.* In this model, the parents are assumed to be independent causes for the child. This allows us to formulate a CPT from the specification of only the following conditional probabilities: $P(c|p_{ij})$, where c ranges over the values of the child, and p_{ij} ranges over the values of parent P_i. These values are combined via the constraints of the model to produce the CPT for the child node.

We assign the probabilities using a causal independence model which specializes to the noisy-OR model when applied to binary nodes. First consider that the inclusive-or connective can be viewed as outputting a true value iff none of the inputs is false:

$output = \neg((\neg v_1) \wedge \cdots \wedge (\neg v_n))$

where each v_i is a logical-valued input variable. The extension to the case where probabilities are associated with each input is relatively straightforward:

$child = \neg((\neg v_1) \wedge \cdots \wedge (\neg v_n))$

$P(child|V_1 = v_1, ..., V_n = v_n) = 1.0 - \Pi(1.0 - P(child|v_i)), \qquad \forall_{V_i} v_i = T.$

When extending to the general case, the relationship between the value of the child node and the values of its parent nodes is not necessarily defined by a truth function. But, the probabilities are assigned analogously:

$P(Child = c|V_1 = v_1, ..., V_n = v_n) = 1.0 - \Pi(1.0 - P(child = c|V_i = v_i))$

$\forall_{V_i} P(child = c|V_i = v_i) > \epsilon.$

In their work on plan recognition, Charniak and Goldman (1993) use the noisy-OR model, specifically for representing the dependencies of observed actions on the potential plans that could explain them.

INTEGRATING EMPIRICAL AND ANALYTICAL INFORMATION: VIRTUAL EVIDENCE NODES

Due to space limitations, we consider just one method for integrating the empirical and analytical components. In this technique, support from the empirical classifiers is added to the Bayesian network using *virtual evidence nodes* (Pearl, 1988). The usual way to add evidence to a Bayesian network is to instantiate a node to a particular value (called "clamping"); the influence of this evidence is then propagated through the network. However, that method is not appropriate for our task, because we do not know the sense of any word (so there is no node in the Bayesian network that can be initially instantiated). Virtual evidence nodes provide a way to specify uncertain evidence, in the form of a distribution over node values (i.e., the probability of each node value). They are represented by the octagonal nodes in figures 1 and 2. There is one for each of the target words to be disambiguated. These nodes represent the support for each sense that was derived from the corpus by the empirical component. Each virtual evidence node is implemented as a binary-valued node whose parent is the node for which evidence is being provided. The evidence distribution determines the conditional probability table.

sense	before	after
community#1	.20	**.70**
gathering#1	.55	.87
municipality#2	.25	.33
town#2	.25	**.33**
community#4	.20	.10
body#2	.20	.10
location#1	.50	.67
municipality#1	.25	.33
town#1	.25	**.33**

Table 1: Propagation w/ hyponym→hypernym links

EDGE DIRECTION AND BELIEF PROPAGATION

There is a very important implication of the choice between the hypernym → hyponym and the hyponym → hypernym representations. In a Bayesian network, suppose that evidence is added to a node (either by clamping or by virtual evidence nodes). This evidence will propagate to its ancestors in the Bayesian network, **and also to the children of its ancestors**. For example, in figure 1, evidence introduced at node SUPPORT_COMMUNITY will propagate, among other places, back to COMMUNITY#1, back to GATHERING#1, and then down to MUNICIPALITY#2, and so on. Thus, this representation, hypernym → hyponym, supports the kind of propagation described in this paper.

On the other hand, consider the hyponym → hypernym representations (figures 1 and 2). In these representations, the targeted words are the roots of the Bayesian network, so the evidence is added to the roots of the network. This evidence will **not** propagate from, say, COMMUNITY#1 to TOWN#2 in figure 1. Information propagates between such nodes **only if evidence were added to their mutual descendents**. As Pearl says, "evidence gathered at a particular node does not influence any of its spouses until their common child gathers diagnostic support" ((Pearl, 1988), p. 182). Thus, if evidence is only added at the virtual evidence nodes in figure 2, evidence will not propagate from COMMUNITY=1 to MUNICIPALITY=2 (so it will not propagate further to TOWN=2). The corresponding nodes are spouses, but their child (GATHERING) has not received diagnostic support, by which Pearl means evidence propagated from below.

However, there are many other possibilities for adding evidence to the network, under which desired propagation would occur. Thus, our discussion of the hyponym → hypernym representations is not just a cautionary tale. For example, one might use Yarowsky's (1992) unsupervised method for assigning words to thesaural categories to add evidence to a node representing a superordinate concept in the WordNet *is-a* hierarchy. (Virtual evidence nodes could be used for this purpose too.) In the hyponym → hypernym representations, this superordinate concept (say GATHERING#1 or SOCIAL-GROUP#1) is a **descendent** of the nodes representing the targeted words. It would thus provide the needed diagnostic support to enable propagation from one target word to another. Note that the hyponym → hypernym representation is conceptually appealing, since its semantics is based directly on the semantics of the WordNet *is-a* hierarchy.

As an illustration, consider applying sample evidence of (.70, .10, .10, .10) for the senses of "community" (with no evidence for town). Table 1 shows the posterior probabilities before and after applying this evidence.

As can be seen, the high evidence for COMMUNITY#1 increases the support for the hypernym GATHERING#1 (as well as for the other ancestors in the same path not shown). However, no support is reaching MUNICIPALITY#2.

If the hypernym → hyponym representation is used instead (as in figure 1), an appropriate propagation does take place. The propagation occurs in two phases. First, the high evidence for COMMUNITY#1 is propagated "upstream" to the hypernym node. Then, the increased support for this synset is propagated "downstream" to increase the likelihood of the value for the appropriate sense of "town". Table 2 shows the posterior probabilities in this case.

ATTENUATION OF SPREADING ACTIVATION

An important aspect of spreading activation approaches is that the strength of the evidence being propagated is attenuated the further the evidence spreads from the original source. Traditional spreading activation

sense	before	after
community#1	.054	**.562**
gathering#1	.126	.631
municipality#2	.063	.312
town#2	.060	**.290**
community#4	.020	.030
body#2	.064	.296
location#1	.500	.500
municipality#1	.068	.063
town#1	.053	**.050**

Table 2: Propagation w/ hypernym→hyponym links

schemes have used various heuristics to model this attenuation, often incorporating a distance factor in terms of number of links. By using probabilistic propagation, we can account for both length of path and fan-out at the nodes along the path (i.e., how many children they have). The length of the path is taken into account by the propagation algorithm. Intuitively, when a node calculates its posterior distribution, it calculates a distribution taking into account all possibilities (e.g., gathering#1=1, municipality#2=1; gathering#1=1, municipality#2=0; and so on). As the evidence is dispersed among the various possibilities at subsequent nodes, the evidence for any single possibility tends to decrease. This is so for either edge direction.

COMPARISON TO RELATED WORK

Spreading activation schemes have been common in various forms, starting with Quillian's (Quillian, 1968) work on semantic memory. Quillian used spreading activation to identify paths between concepts for the purpose of comparison and contrast. To construct the semantic networks, dictionary definitions were manually encoded in the form a graph.

Hirst (1988) also used spreading activation to perform word-sense disambiguation. The approach relies on the identification of paths between interdependent word meanings. To avoid extraneous connections, constraints were introduced; for instance, a limit on path length was introduced, and *is-a* links were normally not traversed in reverse direction. Furthermore, heuristics were used to give preference to shorter paths and to avoid connections through nodes with many out-going arcs.

There have been several approaches that have relied upon word-overlap in dictionary definitions to resolve word-sense ambiguities in context, starting with (Lesk, 1986). Cowie et al. (1992) extend the idea by using simulated annealing to optimize a configuration of word senses simultaneously in terms of degree of word overlap.

Veronis and Ide (1990) developed a neural network model to overcome the limitation of addressing only pairwise dependencies in word-overlap approaches. Using dictionary definitions, they constructed a network containing links from each word node to the nodes for each of its senses and links from each of the sense nodes to the nodes of the words used in the definition.

Sussna (1993) produces a semantic network based on several different WordNet relations. His disambiguation method minimizes the pairwise distance among senses via a weighting scheme that accounts for both fan-out and depth in the hierarchy. Of the approaches we have surveyed, his is most similar to our analytical component.

Voorhees (1993) describes an unsupervised approach that exploits the WordNet hypernym taxonomy. In particular, the hierarchy for a given word is automatically partitioned so that the words occurring in the synsets of a partition (or *hood*) only occur with one of the senses for the word. Disambiguation is based on the selecting the hood which has the highest estimated relative frequency for the context relative to training text.

Resnik (1995) also describes an unsupervised approach that is based on estimating synset frequencies. As with Voorhees, the estimated frequency of a synset is based on the frequency of the word plus the frequencies of all its descendant synsets in a large corpus. Therefore, the top-level synsets have the highest frequencies and thus the highest estimated frequency of occurrence. For each pair of nouns from the text to be disambiguated, the *most-informative-subsumer* is determined by finding the common ancestor with the highest information content, where information content is inversely related to frequency. Then each noun is

disambiguated by selecting the synset that receives the most support (i.e., information content) from the all of the most-informative-subsumers.

Eizirik et al. (1993) also describe a Bayesian network model for word-sense disambiguation, which includes syntactic disambiguation as well as lexical information. However, their networks are not automatically constructed.

CONCLUSIONS

This paper explores various representational issues that must be addressed when formulating a Bayesian network representation of lexical information such as is expressed in WordNet. We describe two choices for the representation of lexical items and two choices for the representation lexical relations. The effects on evidence propagation in the network is also discussed.

References

1. Bruce, R., and Wiebe, J. (1994), "Word-sense disambiguation using decomposable models", in *Proc. of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL-94)*, pp 139–146.
2. Charniak, E., and R. Goldman (1993), "A Bayesian Model of Plan Recognition", *Artificial Intelligence* 64:53-79.
3. Cowie, J., J. Guthrie, and L. Guthrie (1992), "Lexical Disambiguation Using Simulated Annealing", *Proc. COLING-92*, pp. 359-365.
4. Eizirik, L. , V. Barbosa, and S. Mendes (1993), "A Bayesian-Network Approach to Lexical Disambiguation", *Cognitive Science*, 17:257-283.
5. Heckerman, D. and J. Breese (1994), "Causal Independence for Probability Assessment and Inference Using Bayesian Networks", Technical Report MSR-TR-94-08, Microsoft Research, (Revised October, 1995).
6. Hirst, G. (1988), "Resolving Lexical Ambiguity Computationally with Spreading Activation and Polaroid Words", in *Lexical Ambiguity Resolution*, S. Small, G. Cottrell, and M. Tanenhaus (eds), San Mateo, CA: Morgan Kaufmann Publishers, pp. 73-107.
7. Lesk, M. (1986), "Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone", in Proc. SIGDOC, Toronto.
8. Miller, G., (1990), "WordNet: An On-line Lexical Database", *International Journal of Lexicography* 3(4).
9. Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems*, San Mateo, CA: Morgan Kaufmann.
10. Quillian, M.(1968), "Semantic Memory", in *Semantic Information and Processing*, M. Minsky, ed., Cambridge, MA: MIT Press.
11. Resnik, P. (1995), "Disambiguating Noun Groupings with Respect to WordNet Senses", in *Proc. Third Workshop on Very Large Corpora*, Cambridge, MA, June 1995.
12. Sussna, M. (1993), "Word Sense Disambiguation for Free-text Indexing Using a Massive Semantic Network", in *Proc. Second International Conference on Information and Knowledge Management (CIKM-93)*, Arlington, Virginia.
13. Veronis, J., and N. Ide (1990), "Word Sense Disambiguation with Very Large Neural Networks Extracted from Machine Readable Dictionaries", in *Proc. COLING-90*, Helsinki, August 1990.
14. Voorhees, E. (1993) "Using WordNet to Disambiguate Word Senses for Text Retrieval", in *Proc. 16th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, Pittsburgh, pp. 171-180.
15. Yarowsky, D. (1992), "Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora", in *Proc. COLING-92*, Nantes, Aug 23-28, pp. 454-460.

INTENTIONALLY LEFT BLANK.

Assessing the Utility of Digitization Options for the Army: Problems and Opportunities

P. Beaver

Mathematical Sciences Center of Excellence
Department of Mathematical Sciences
U. S. Military Academy, West Point, New York

Sponsor: Army Digitization Office

ABSTRACT

The Army is digitizing the battlefield. Equipment is already being fielded in the first digitized division, and the first digitized corps will follow shortly. There are numerous unanswered questions concerning digitization that cover a wide spectrum of areas. These include questions of interoperability, training and doctrine, testing and evaluation, simulation, network architecture and design, and survivability of the system. Many of these questions are already being addressed by the appropriate Army agencies; however, there are many important questions that are not yet being addressed. Some of these questions are discussed within the appropriate frameworks for answering them, for example, marketing, modeling and simulation, and network architecture, and references are made to works already in progress. The goal is to generate discussion of some of the problem areas and to identify possible solution techniques.

INTRODUCTION

The Army is digitizing the battlefield. One division is already fielding the new equipment, and the first digitized corps is soon to follow. However, digitization involves more than introducing 109 new line items into our tactical units. As with any new system, the digitization effort raises questions of interoperability, training and doctrine, and testing and evaluation, among others. The Army Digitization Office (ADO) has been charged with oversight of this effort, and in this paper we explore some of the digitization issues which the Army and ADO face. Our goal is to generate some dialogue as to how we may address some of these problems.

The questions of interoperability that digitization presents to the Army are unique to this technology. Certainly, ground combat systems should be compatible with those of our allies, but the digital exchange of information not only creates information exchange problems for combined warfighting, but also for joint operations, operations within the Army (such as compatibility between ground and aviation), and even between digitized and non-digitized units. Additionally, there are problems with linking the new digitization technology with previously automated systems, such as field artillery and air defense systems, to name a few.

A problem with digitization that the Army rarely encounters is in the area of development and fielding. With most weapons systems, the Army is at the "cutting edge" of the technology; however, with digitization, military technology is lagging the civilian sector. Consequently, the Army has adopted the "spiral development process" as opposed to its traditional equipment development and fielding procedures, and this new approach presents many challenges.

Since the new equipment is being fielded concurrently with the test and evaluation process, this presents challenges for the first digitized units. They need to learn the new technology, develop doctrine and techniques, tactics and procedures (TTPs), and evaluate the effectiveness of the equipment. However, at the same time they must maintain combat readiness since they are still tactical (and not T&E) units.

Most of the questions facing ADO are closely related, for example simulation and testing and evaluation, but for the purpose of organization we classify them into major categories. The categories mentioned above (interoperability, training and doctrine, and testing and evaluation) are addressed

peripherally in terms of where they might overlap with our main focus. We concentrate our discussion on questions of modeling and simulation, and network architecture and design. In addition to these areas, ADO has identified a problem that is not as often encountered in the fielding of new systems: marketing digitization. It is this issue that we first address.

MARKETING DIGITIZATION

The digitization effort will be expensive, and those paying for it are going to want to know what the payoff is in terms of combat effectiveness. This can be answered by simply generating a response surface that shows combat effectiveness as a function of the level of digitization and its effects. That is, we isolate the effects of digitization (perhaps in a simulation) include all of the other variables with which digitization effects may interact, and run enough simulations (or live exercises) to create a field of data points sufficient to fit a combat effectiveness surface over the domain.

Even the most cursory attempt at this approach demonstrates its inherent difficulties. Running an armor platoon-on-platoon scenario on JANUS and increasing the situational awareness level of the tanks did, in fact, show a statistically significant increase in combat effectiveness, measured in loss exchange ratios. Other simulation experiments, however, have failed to show significant results (see Krahn, et. al., (1998)). Furthermore, the main result of these experiments was the realization that our current collection of simulations is woefully inadequate in its representation of the effects of digitization (see, for example, Barr, (1995) or Sherrill, (1998)).

As an example, one way to represent the effects of digitization is to place a person in the loop on a simulation who can take advantage of the enhanced ability to modify a course of action due to increased situational awareness. This can be done with an applique that provides total situational awareness to the player. However, it is difficult to use this to compare a digitized force to our current forces because JANUS does not reflect current technology. Currently the heads-up display for the player in the loop already provides complete situational awareness, and not just the information a commander might have through the current reporting system. In order to get a reasonable comparison to current technology, we would have to superimpose our existing communications nets onto JANUS, where the person in the loop receives data comparable to that provided with current technology. Since our simulation effort should be focused on developing new models that reflect evolving technologies, fixing JANUS to reflect old technologies is not a useful effort.

Attempts to validate digitization with our current suite of models will run into similar problems. Each model has its own input variables, and these do not necessarily intersect with the effects of digitization. Furthermore, placing a person in the loop does not necessarily let us replicate the decision-making process of a tactical operations center, since there is always the tendency for the person to play the "gamesmanship" of the simulation and not truly select the best military course of action. The response surface becomes even more elusive under these considerations.

Similarly, testing the effects of digitization with live exercises has its own collection of problems. The cost of the Advanced War-fighting Experiments (AWEs) is too great to conduct a large number of exercises, making the design of the experiments quite tricky. It is difficult to draw statistical inferences from a single data point, and equally as difficult to isolate the effects of digitization and their interactions while at the same time using the exercise to develop and test doctrine. . We certainly need a baseline with which to compare our current experiments, and twelve years of National Training Center (NTC) rotations may or may not be the best candidate. Lucas (1998) discussed ways to design an AWE to derive the most information from the experiment while preserving its other functions (i.e., development of digitization doctrine) but it is clear that the elusive response surface is not going to drop out of an AWE.

It seems intuitive that increased situational awareness and enhanced command and control would lead to increased combat effectiveness. However, supporting this conclusion (or not) with sound statistical analysis is the first of the problems we present. Even if this can be used to generate a response surface, we need to translate it into terms that can be understood by those paying for the digitization effort. As we have already alluded to issues of modeling and simulation, we next discuss some problems in these areas.

MODELING AND SIMULATION

We have identified a primary question of digitization, which asks "What is its utility?" Related to this question are several others, for example, what is the marginal increase in effectiveness for digitizing various units (i.e., an air defense unit that is already almost fully digitized versus a light infantry unit which is not)? How do we digitize different types of units (light, heavy, or cav), and what is the appropriate level for digitization? We must also address the use of digitization technology for not only conventional operations, but for operations other than war and for a variety of environments. These are questions of doctrine that are too important to be answered via simulation. Commanders shudder at the thought of being issued equipment and taught doctrine based on a computer's insight of how the equipment will perform. However, the basis for field testing can come from analysis derived from simulations, and simulations can lay the groundwork for developing TTPs and identifying problems with the equipment. Furthermore, simulations remain an invaluable training tool for battle staffs when used correctly. These considerations show it is essential that our ground combat models accurately reflect the effects of digitization.

The first issue, which is critical when simulations are used in the test and evaluation role, is what measure of effectiveness should be used to analyze results. Typically, loss exchange ratios or some subjective judgment of mission accomplishment have been used (see, for example, Young (1998), Cassady (1998), or Grynovicki, et. al. (1998)), but an *a priori* selection of a measure is essential to accurate analysis.

Whether simulations are going to be used in a test and evaluation role or strictly to aid training, it is important that we accurately represent the effects of digitization in our combat models. Do we overlay a communications network onto our current models that replicates the new situation awareness and command and control technology; do we simply identify the toggles in our current models which may be switched to reflect the effects of digitization; or do we re-write all of our models based on digitization technology?

Even with digitization accurately reflected in our combat models, there are numerous other questions to consider. With the increased amount of data available at every level, to avoid a person in the loop making decisions, can we automate this procedure? An artificial intelligence application may be employed to represent the decision making process at each level, and a course of action may be selected without a person in the loop. Leake (1998) is already investigating this question as it applies to JWARS, and his applications should be generalizable to other ground combat models. However, automating the decision-making process, if only for representing digitization effects in our simulations, is a hard problem. Furthermore, there is the human factors consideration of how the decision making process may need modification based on the increased volume of data. An artificial intelligence application as described above may be employed to assist commanders in their decision making process, although the computing power required to support such an application (i.e., something like Deep Blue of computer chess fame) is still perhaps a few generations removed from our ground combat units.

In addition to the possibility of this new amount of information overwhelming our tactical decision-makers, there are other human factors problems to consider. There is the danger of too much reliance on the technology, which will keep commanders from making decisions as they wait for the "complete picture." There is the further danger in the possibility of this technology becoming a crutch, leaving our commanders crippled when the system goes down. Furthermore, there will be commanders who fail to make decisions because they know their commanders have the same information, and they do not want to get caught making the "wrong" decisions. While this technology can make good commanders better, it has the potential to make bad ones worse.

We next address the equipment itself and discuss the network architecture supporting digitization.

NETWORK ARCHITECTURE

The equipment used to digitize the battlefield serves two primary purposes: situational awareness and command and control. The concept of a tactical internet over which messages are sent in bursts of 576 bytes raises numerous questions of system design which range from stochastic graph theory to queuing

theory. The optimal level of connectivity given a probability of failure for each node is an important question for system survivability. The solution on a graph of this stochastic optimization problem can certainly be integrated into the architecture of the tactical internet. Task reorganization (combining networks) raises further question along these lines, which should be solved prior to fielding the equipment.

With the current architecture (messages stacking in a queue until a "handshake" is established with the server) questions of tagging messages with a priority, degradation of messages as they become obsolete, and the need to back up a server in case a node is lost before messages are relayed are important considerations. These problems of dynamic network theory and queuing theory are critical to the effectiveness of this new technology.

The time between transmission of situational awareness data can be increased if predictive algorithms are employed to locate elements within a prescribed tolerance, and if updates are sent only when this tolerance is exceeded. Such predictive algorithms can be employed by both the sending unit to determine when to update the server as well as by the higher element to maintain a reasonably accurate situational picture. The increased range allowed by this (over that achieved with continuous transmissions) in addition to the reduced electronic footprint of the unit, are clear advantages of employing such algorithms.

The flexibility of the technology to be adaptable to future applications is an important consideration as well. With bursts of 576 bytes, messages of up to a few Kbytes can currently be sent, but this need can conceivably increase dramatically in the future. If messages with increased volume (logistics requests, intelligence updates, orders, images, or video) are to be sent on these same nets, data compression algorithms may be needed to handle the additional traffic. These represent only a few of the issues with network architecture.

The survivability of the system and its resistance to "digitization operations" is an important question of architecture, employment, evaluation, and doctrine. Like any tactical communications net, digitization technology needs to be resistant to intrusions such as jamming, hacking, bugs, false information, and loss of nodes. These factors need to not only be represented in our models, but they need to be present in our exercises and experiments if we are going to be successful in employing this technology. Additionally, we are going to have to maintain these new systems. Training qualified system administrators for our tactical units will be a challenge, and retaining these qualified system administrators will be a major challenge.

SUMMARY

We have presented problems of digitization of the battlefield from the areas of marketing digitization, modeling and simulation, and network design. Many of these are closely related to questions of interoperability, training and doctrine, and test and evaluation, among others. The questions call on various disciplines such as statistical analysis, human factors, artificial intelligence, graph and queuing theory, dynamic programming, and perhaps simple common sense. We hope this serves to generate some discussion about these problems that may be the beginning of some of their solutions.

REFERENCES

Barr, D. R. and Sherrill, T. E., "Estimating the Operational Value of Tactical Information," United States Military Academy Department of Systems Engineering Technical Report, West Point, NY, July 1995.

Cassady, P. D., "What to do with 1000 Replications of CASTFOREM," U.S. Army Conference on Applied Statistics, Las Cruces, New Mexico, 21-23 October 1998.

Grynovicki, J. O., Golden, M. G., and Kysor, K. P., "Performance-based Metrics to Assess the Effects of Digitization during the DIVISION XXI Advanced War-fighting Experiment," U.S. Army Conference on Applied Statistics, Las Cruces, New Mexico, 21-23 October 1998.

Krahn, G., et. al., "Exploring the Value of Information on the Digital Battlefield," United States Military Academy Department of Mathematical Sciences Technical Report, West Point, NY, 18 June 1998.

Leake, C. R., "The Use of Cognitive Processing Adaptive to Decision Making in the JWARS Project," U.S. Army Conference on Applied Statistics, Las Cruces, New Mexico, 21-23 October 1998.

Lucas, T., "The Advanced Warfighting Experiments: Where do we go from here," U.S. Army Conference on Applied Statistics, Las Cruces, New Mexico, 21-23 October 1998.

Sherrill, T. E. et. al., "Quantifying Situation Awareness in JANUS," United States Military Academy Operations Research Center Technical Report, West Point, NY, March, 1998.

Young, K., "Modeling and Simulation Techniques used for Analyzing Digitization of the Battlefield," U.S. Army Conference on Applied Statistics, Las Cruces, New Mexico, 21-23 October 1998.

INTENTIONALLY LEFT BLANK.

Exploring a syntactic structure of features in imagery[1]

John Bart Wilburn
University of Arizona, Optical Sciences Center
Bart.wilburn@opt-sci.arizona.edu
Ph: (520)621-6995

Abstract: We explore the structure of a language model of features that would link images directly to a logical model for an automated interpretive transform of imagery. The syntax of the language model is based on the two dimensional (2-D) solution of the "cross" median window filter (MW) constrained by a predicate of the data to extract features in terms of fixed-point (FP) roots of the filter. Experimental results to date on feature extraction using this method are presented. An FP root of a MW is shown to be an object in itself having a property that the data composing it are related and are one of a finite set of distinct 2-D locally monotonic patterns of relationships. This relational property of FPs, and that they are a finite set, results in a grammar for the co-joining or juxtaposition of the root patterns. We represent this notion of a grammar of features in a syntactical structure in terms of FP root patterns of features and show that it extends to a language model having semantic content satisfying a non-classical propositional language system.

1.0 INTRODUCTION

This essay combines recent developments in median window theory with the fundamental requirements of logical models for an interpretive transform of imagery, or object recognition. Models for object recognition, however, are not the subjects of this paper except insofar as concerns the compatibility of their fundamental constraints with the extraction of features using the methods described herein. What we are exploring here is a way of directly connecting an image-processing scheme for the extraction of features to models of the meaning-content of object features. This work is a step in the direction of addressing the classical "frame problem" of machine intelligence. In addition, the method of feature extraction described has utility in its own right, apart from any model of object recognition. The essay is organized in three parts: (A) Feature extraction. (B) A language model of features. (C) Validation of the language model to be a legitimate, non-classical, propositional language system.

2.0 BACKGROUND

The primary hypothesis of this work is that the connecting link between feature extraction and the interpretive transform of imagery is a language with a common basis of semantic content in the image and in a model of object recognition. The method of feature extraction by machine employed in this work is a median window filter constrained by a predicate of the data to detect only fixed-point (FP) roots of the filter in two dimensions (2-D). The median window filter is one of the ranked-order filters and has the characteristic of having fixed-point and oscillating roots. The roots of the filter are data that satisfy relational patterns that pass through the filter unchanged in value and position. The fixed-point roots are a set of contiguous multivalued data, and oscillating roots are a set of oscillating binary data. The secondary hypothesis is that data satisfying the constraints of 2-D FP roots of the filter are patterns of data that form features of many objects of interest. Experience has shown the secondary hypothesis to be a valid assumption for most manmade objects and some natural objects. We will show that extracting features in this way results in a finite representation of features constrained by syntax, and as such it constitutes a structure of feature representation based on the image itself rather than on a neo-Kantian presupposition of it. This apparent structure of a representation of features by FPs is what has led to the notion that we may be able to develop a syntactic structure for the semantic content of imagery, or a language of imagery, based on naturally occurring features. The connection of the primary hypothesis with models for object

[1] Approved for public release.

recognition is that such a model is a logical model expressed in language for the deduction of an assertion of the object.

2.3 Models

The introduction above implies a logical structure of syntactic deduction and semantic entailment. That is, a logical structure based on true propositions - sentences - whereby a hypothesis is judged true or false as a logical consequence of a set of sentences. Consider, for example, the recognition of a numbered object such as a navigation marker. The meaning of such an object is determined by what *sort*[5] of object it is, e.g., a navigation marker, and the present tense of the context it and the observer are in. The act of recognizing the object is the interpretation of it by what it means to us. This is an act of judgement of the context including the object to be harmonious, i.e., logically consistent, with the relevant experience of the observer. We may think of recognition of an object, then, as the 'functioning' of an experienced observer embedded in the present tense, or actual, world. Furthermore, as language is the mode of interpreting the actual world by a conscious observer into his experience of it, e.g., "S is p" - that object is a desk - a logical model based on sentences is most compatible with the interpretation of an image by a conscious observer.

We may consider modelling the experience of the observer involved in recognition as the content of possible worlds of the observer that represent his experience of other situations in the form of sentences, or propositions. Such possible worlds are indexed by the objects represented in them, events, time, etc., and further by accessibility relations to other possible worlds. In the case of object recognition, the possible worlds are those relevant to the present tense actual world of the object. Such possible worlds include the representation of the intensions of other objects, and any sentences describing application, relevant to the object in the present tense actual world. In the actual world of real people, the collection of relevant experience is a natural filtration of all experience over the experience of the present tense actual world. In a logical model, the experience of an event is a world represented by propositions, or sentences, and a filtration is an equivalence class of worlds with respect to a set of sentences Φ. This means that whatever else any of the possible worlds in that class contain, they contain sentences and subsentences that are semantically equivalent to those in Φ. We may entertain the notion of *model* based on this concept of filtration.

A logical model contains the set of all sentences, W, which can, at least in principle, be infinite, but they are denumerably infinite, thus each of them is of finite length, in order to have semantic value. The set Φ must satisfy semantic deduction for finite consequences, which is our concern here, thus the set of sentences, Φ, must be finite and closed under a finite set of subsentences including atomic sentences. Atomic sentences are those sentences that are indivisible in semantic value. The fundamental requirement of a model based on filtration of sentences, or worlds, is that the filtration is over a finite set of sentences closed under sub sentences. If we denote the present tense world as α, and regard it as a set of sentences, or propositions, the result of a filtration over Φ_α is a finite collection of worlds that contain sentences semantically equivalent to the closure set Φ_α and relevant to the present tense world α. For our purposes in this essay, α is the event of an image of an object, or set of objects.

We may define a model more formally as $\mathcal{M}=\langle W, R, V\rangle$, where W is a non-empty set of all possible worlds, $w\in W$, R is a dyadic relation over the members of W, and V is a truth value assignment to sentences in w, $w\in W$, $V(\beta,w)\in\{T=1, F=0\}$. Given a model, $\mathcal{M}$, we may form a sub-model, $\mathcal{M}_\alpha$, from a set of sentences (well-formed formulas: wff), Φ_α. A subset A of W is an equivalence class in W with respect to Φ_α iff A is non-empty and there is some subset Λ of Φ_α such that for every $w\in W$, $w\in A$ iff for every $\beta\in\Lambda$, $V(\beta,w)=1$ and for every $\gamma\in\Phi_\alpha - \Lambda$, $V(\gamma,w)=0$. The point here is that a model such as this is a *finite* model based on sentences having semantic value, and selects from W those w that contain wffs semantically equivalent with respect to Φ_α. This is the basic form of a logical model of object recognition, and, as we shall see, it is entirely compatible with the form of a language model, and well it should be as the logical model is defined on sentences. We may defer the mathematics of this for now and illustrate the need for finiteness by supposing that you were asked to understand a sentence that was infinitely long, or even a finite set of finitely long sentences composed from an infinite alphabet. It is unlikely you could make any sense of either. I will not engage in a detailed discussion of a possible worlds model of object

recognition[17] here because that would depart from our purpose which is to explore the idea of a language of imagery compatible with a language of such a model.

The problem we have had with models for interpretation of imagery in the past is that we have had to rely on the observer to detect and interpret features in an image into sentences and atomic sentences. The problem derives from there being no uniform procedure for each observer to decompose an object into a linguistic description of features, particularly in the case of a partially obscured object. Furthermore, algorithmic feature extraction schemes using extensional data of imagery fail in satisfying the finite requirement. This is because extensional data, such as raw image data or statistical measures of imagery, are samples from a population that is certainly indefinably large and mathematically infinite. Some extensional methods have been employed that have the finite property, but they either fail in neo-Kantian presuppositions of morphology[6], or they fail to autonomously detect features. These shortcomings define our requirements for a feature extraction scheme to be able to detect and translate features in an image directly into a language that satisfies the structure of a logical system - a machine - directly linking it to the present tense context of the object. In the course of this paper, we will show that a linguistic transform of features must have the properties of consistency, completeness, finitary entailment, and that it is an intensional representation of objects. The intensional representation is necessary in order that features can be semantically related to any possible world - image - where they, or their sort, may exist.

3.0 FEATURE EXTRACTION

The MW has been around for some time and that it has roots has also been known. The characterization of the roots, however, has been slow in coming, and in particular the characterization of 2-D roots has eluded solution until very recently. The following is a review of the predicate constrained solution to the 2-D MW with results showing that it seems to work well for extraction of features, and that it can represent the intensions of features in terms of FPs. In addition to 2-D root characterization, the essential feature of this solution to the MW is that it enables control of the filter for selective detection of roots. A simple description of the FP roots to the MW might suffice for the development of a language model based on them, but in order to judge the claim that the model is based on an intensional representation of natural features, it is necessary to understand the basis of the FP representation. For that reason, the solution of the MW for FP roots (MW(FP)) is presented. Following the MW(FP) solution, a language model incorporating the MW(FP) satisfying a non-classical language system is proposed for an interpretive transform of imagery in a later development. The terms "MW filter" and "filter" are used interchangeably in this section referring to the median window filter. The context of use should preclude any confusion about the term "filter" as used above with the same term used in section 4.0, referring to a logical filter. I should also point out that the MW is not necessarily the only scheme for feature extraction that may satisfy the requirements given above. As other methods come along, or are discovered to have the necessary properties, they may be incorporated into this rationale.

3.1 Solution of 2-D Median window for feature extraction

The notion for this approach to the 2-D median window (MW) filter came about from knowing that the basis of the functioning of the MW filter was the relationship of data values rather than the absolute value of the data. Furthermore, as remarked above, the MW filter has the characteristic of patterns of data known as roots of the filter defined by the relationship of the data within the pattern and with the neighboring data of the roots. These observations suggested that mathematical logic was a tool likely to have success in solving the filter problem of defining the properties of data that qualify as FP roots of the filter. The application of the 2-D MW to feature extraction derived from the assumption that there were a significant number of objects of interest that were composed of features satisfying the characteristics of FP roots to a MW. Our experience to date seems to verify this assumption for manmade objects, and to a lesser extent for natural objects. The following is a discussion of the solution to the filter, with applications to feature extraction, and a discussion of the 2-D FP roots comprising a finite set of data patterns constrained by rules of combination, or grammar. The results shown here are drawn from a prior paper discussing the MW filter as a member of the class of ranked-order filters[18].

3.1.1 The formalism

The MW is remarkably simple in construction, and understanding how it operates is almost trivially simple. Considered in 1-D, the MW is simply a window of length N applied to a set of data selecting N data and sorting them – thus the term ranked-order – and the median value is selected as the output. It has been a source of bafflement that understanding what the filter does is easy, but understanding what the filter would do, i.e., what kinds of data patterns are roots of the filter, has been difficult. As it happens, the solution is remarkably simple and turns on the representation of the data, and on the distinction between functions and predicates of the data. The solution is developed for a filter in 1-D of length N: $\varphi_{med}(N)$, N=2k+1, and then shown for 2-D for a 'cross' filter of size L: $\vartheta_{+}(L)$ L= 2N-1 (See ref. 18. The square filter is a false filter, and the hexagonal filter is insufficiently developed for this application).

We begin with a description of the image data u_n as being composed of terms designating individual entities having property u. Please note that this differs considerably from the conventional notion of data being simply a set of numbers. In our case, the individual entities are "pixels" – acronym for: "picture element" – and they have the properties of value, or intensity, and position. Let us take note that if $v^1,\ldots,v^n$ are terms, and a function, f, is n-ary, then $fv^1,\ldots,v^n$ is a term. The data set, u, may be considered to be of this form where u is the symbol that is n-ary such that $uv^1,\ldots,v^n$ is a term. The number, n, is a natural number determined by u and is the *index* of u. We may use this form to associate with every u_n in a sequence, u, a term v^n designating its position in that sequence. This convention enables separation of the index of a variable from the value of the variable so that we may construct functions of its position in a sequence independently of its value, or functions of its value independently of its position. The notational convention adopted is: $\mathfrak{u}=(\ \mathfrak{u})v^i$, $i\leq n$ such that for every n-tuple $\mathfrak{u}=(u_1,\ldots,u_n)$ are associated functions $((u_1),\ldots,(u_n))$ and $\langle v^1,\ldots,v^n\rangle$ where the (u_i) are the value of the data element in the position indicated by v^i.

We may define a sub-sequence of $\mathfrak{u}$: $\mathfrak{u}_i$, by

$$\mathfrak{u}_i\colon (\mathfrak{u}_i) = ((u_i), (u_{i+1}),\ldots,(u_{i+N-1})),\ \mathfrak{v}^i = \langle v^i, v^{i+1},\ldots,v^{i+N-1}\rangle$$

Further, we may define a recursive function : $\beta(\mathfrak{a}^i, j)$, $\mathfrak{a}^i=(a^{i+1},\ldots, a^{i+Q})$:

$$\beta(\mathfrak{a}^i, j)= a^{i+j}{}_j,\ j=0,N-1$$
$$\beta((\mathfrak{u}_i), j)=(\mathfrak{u}_i)_j$$
$$\beta(\mathfrak{v}^i, j)= \mathfrak{v}^i{}_j$$

$$(\mathfrak{u}_i)_j=((u_i)_0,\ldots,(u_{i+N-1})_{N-1})$$
$$\mathfrak{v}^i{}_j=\langle v^i{}_0,\ldots,v^{i+N-1}{}_{N-1}\rangle.$$
$$\mathfrak{u}_i = (\mathfrak{u}_i)_j\ \mathfrak{v}^i{}_j$$

This construction defines $\mathfrak{u}_i =(\mathfrak{u}_i)_j\mathfrak{v}^i{}_j$ and preserves all necessary information to recover $\mathfrak{u}$ from $\mathfrak{u}$ by logical addition of $\mathfrak{u}_i$: $\mathfrak{u}_i\oplus\mathfrak{u}_{i+1}=u_i,\ u_{i+1}$. This construction allows $(\mathfrak{u}_i)_j$ to denote the value of the j^{th} term of $\mathfrak{u}_i$ separable from its position in an ordered sequence: $\langle v^i{}_0,\ldots, v^i{}_{2k}\rangle$. This construction further defines a window, $\mathfrak{u}_i$, of length N that selects N of $\mathfrak{u}$ and associates an index with them of j =0,N-1 for every increment of i; i=1,n-(N-1). Finally, this construction constitutes the *sampling* function, $\mathfrak{u}_i$, of $\mathfrak{u}$ and forms a set:

$$\mathfrak{u}= (\mathfrak{u}_1,\ldots, \mathfrak{u}_{n-2k})$$

where every $\mathfrak{u}_i$ is: $\mathfrak{u}_i = (u_i, u_{i+1},\ldots,u_{i+2k})$; i=1,n-2k. The set $\mathfrak{u}$ is a power set of $\mathfrak{u}_i$, and

$$\mathfrak{u} = \bigcup_{i=1}^{n-2k} \mathfrak{u}_i.$$

These functions will be employed to construct the filter function applied to a data sequence, $\mathbf{u}$. We must now define the logical basis for the filter as a function.

The structure of the logical development of the median filter is in terms of functions having the recursive properties of: $K_<$, '=', and $F(\mathbf{a}) = \mu x(M(\mathbf{a}, x) = 0)$ where $\mu x(\ldots x \ldots)$ is the μ-operator defined as the minimum x such that ...x... is true. We further define a recursive predicate $\Re(\mathbf{a})$ and its representing function, $K_{\Re}(\mathbf{a})$:

$$K_{\Re}(\mathbf{a}) = 0 \text{ if } \Re(\mathbf{a})$$
$$= 1 \text{ if } \neg\, \Re(\mathbf{a}), \text{ Def. '}\neg\text{' is negation,}$$

And we define a function $\wp(\mathbf{a})$ subject to the predicate being true as:

$$\wp(\mathbf{a}) \widetilde{\leftrightarrow} \Re(\mathbf{a}),$$

to mean both $\Re$ and $\wp$ are defined for the same argument $\mathbf{a}$, or both are undefined. We may use these basic notions to develop a general filter function: $\Im(\mathbf{u}_1,\mathbf{u}_1')$. To do this, we generalize the notion of $\wp(\mathbf{a}) \widetilde{\leftrightarrow} \Re(\mathbf{a})$ to[8]: $\wp(\mathbf{a}_1) \widetilde{\leftrightarrow} \Re(\mathbf{a}_1)$, $\wp(\mathbf{a}_2) \widetilde{\leftrightarrow} \Re(\mathbf{a}_2),\ldots, \wp(\mathbf{a}_q) \widetilde{\leftrightarrow} \Re(\mathbf{a}_q)$. We use this construction to apply the median filter, $\varphi_{med}(N)$, to data represented by the sampling function in the set $\mathbf{u}$, and the result is the filtered output: $\mathbf{u}'$, $\mathbf{u}' = ((u')_1\nu^1,\ldots, (u')_n\nu^n)$ represented by a filter function $\Im(\mathbf{u}, \mathbf{u}')$ defined as follows:

$$\Im(\mathbf{u}_1,\mathbf{u}_1') = \wp(\mathbf{u}_1, \mathbf{u}'_1) \text{ iff } \Re(\mathbf{u}_1)$$
$$\vdots$$
$$\Im(\mathbf{u}_{n-2k},\mathbf{u}_{n-2k}') = \wp(\mathbf{u}_{n-2k}, \mathbf{u}'_{n-2k}) \text{ iff } \Re(\mathbf{u}_{n-2k}),$$

where

$$\Re(\mathbf{u}_i) \widetilde{\leftrightarrow} \mu x(K_{\Re}(\mathbf{u}_i, x) = 0).$$

Now we define the function $\wp(\mathbf{u}_i,\mathbf{u}'_i)$ more explicitly as:

$$\wp(\mathbf{u}_i,\mathbf{u}'_i) = \mu x[H(G(\mathbf{u}_i, x))],$$

and introduce two functions: A selection function: $\pi_i^Q\left(\left\langle x_1,\cdots,x_Q\right\rangle\right) = x_i$ and an ordering function: $O(\mathbf{x}) = \langle (x)_\alpha \eta_0^\alpha, (x)_\beta \eta_1^\beta, \cdots, (x)_j \eta_Q^j \rangle$; $(x)_\alpha \geq (x)_\beta \geq, \ldots, (x)_j$.

For the ordering function, we make use of the separability of the value, $(u)_i$, from its designator, ν^i, and assign a different designator, $\eta^i_\ell = \ell$, ℓ=0,N-1. We must take note that the superscript index of η^x_ℓ is determined by $(u_y)_x\nu^y_x$. Expressed in terms of the sampling function, $\mathbf{u}_i$, the ordering function is:

$$O(\mathbf{u}_i) = \langle (\mathbf{u}_i)_j \eta_0^j, \cdots, (\mathbf{u}_i)_j \eta_{N-1}^j \rangle$$

Now we may construct the filter by defining G in terms of the selection function and the ordering function as:

$$G(\mathbf{u}_i, (u'_i)_m\nu^{i'}_m) = \mu x[x = \pi_m^{N-1}(O(\mathbf{u}_i)) \,\&\, (u'_i)_m\nu^{i'}_m = x],$$

where m is the rank selected by the filter, e.g., for $\varphi_{med}(N)$, m: m=k. (As remarked, we shall be concerned with the median filter, thus m=k.) The function, H, is defined as a writing function:

$$H(G(\mathbf{u}_i, (u'_i)_k\nu^{i'}_k)) = \mu x_{x<N}[(u'_i)_x\nu^{i'}_x = (u'_i)_k\nu^{i'}_k \vee (u'_i)_x\nu^{i'}_x = (0)_x\nu^{i'}_x],$$

so that:

$$\wp(\mathbf{u}_i,\mathbf{u}'_i) = ((0)_0 v^{i'}_0,\ldots,(0)_{k-1} v^{i'}_{k-1}, (u'_i)_k v^{i'}_k,(0)_{k+1} v^{i'}_{k+1},\ldots(0)_{N-1} v^{i'}_{N-1}).$$

By the properties of the sampling function as a member of a power set and of the representing function of the predicate, $K_{\Re}(\mathbf{a})$, we may construct the total filter function $\Im(\mathbf{u},\mathbf{u}')$ as:

$$\Im(\mathbf{u},\mathbf{u}') = \wp(\mathbf{u}_1, \mathbf{u}'_1)\cdot K_{\Re}(\mathbf{u}_1) \oplus,\ldots,\oplus \wp(\mathbf{u}_q, \mathbf{u}'_q)\cdot K_{\Re}(\mathbf{u}_q),$$
$$q=n-2k,\ k=(N-1)/2,$$

or

$$\Im(\mathbf{u},\mathbf{u}')=\bigcup_{i=1}^{n-2k} \wp(\mathbf{u}_i, \mathbf{u}'_i)\cdot K_{\Re}(\mathbf{u}_i). \qquad (1)$$

3.2 The FP solution

With this representation of the filter by $\Im(\mathbf{u},\mathbf{u}')$, we may address the solution for a fixed-point root. The notion of a fixed-point root is that there is a sequence of either multivalued or binary data of some continuous length, **r**, in **u** that is invariant to the filter. This means that if every value of **r** is selected by the filter preserving its value *and its identity in the sequence*, then every number of **r** must be the in the k^{th} position of $O(\mathbf{u}_i)$, i.e., that it is the median value of $\mathbf{u}_i$ *and* is coincident with the k^{th} position in $\mathbf{u}_i$ for each increment of i of a median filter of length N: N=2k+1. This requirement for the output of the median data value to be coincident with the k^{th} data position in $\mathbf{u}_i$ defines a condition of

$$\text{c: } (\mathbf{u}_i)_j v^i_j=(\mathbf{u}_i)_j \eta^j_e \qquad (2)$$
$$\text{k=(N-1)/2; } \Rightarrow v^i_{j=k} = \eta^k_{l=k} = k^i \quad \text{at median}$$

as necessary for $\Re_c(\mathbf{u}_i)$ to be true. This is a *forcing condition* in the predicate. We must note that the use of the representation of data by $\mathbf{u}_i = \langle(\mathbf{u}_i)_j \mathbf{u}^i_j\rangle$ in $(\mathbf{u}_i)_j v^i_k=(\mathbf{u}_i)_j \eta^j_k$ results in the condition of *identity* of the pixel in the k^{th} position of the sampling function and the pixel having the median value of the ordering function. This asserts the coincidence of the $(u_i)_k v^j_k$ to be the median value of $\mathbf{u}_i$. The result of this condition is an ordered sequence of data in the ordering function:

$$O(\mathbf{u}_i) =\langle(\mathbf{u}_i)_j \eta^j_0 \geq\ldots\geq(\mathbf{u}_i)_j \eta^j_{k-1}\geq(\mathbf{u}_i)_k v^k_k\geq(\mathbf{u}_i)_j \eta^j_{k+1}\geq\ldots\geq(\mathbf{u}_i)_k \eta^j_{N-1}\rangle,$$
$$(\mathbf{u}_i)_k v^i_k=(\mathbf{u}_i)_k \eta^k_k \Rightarrow v^i_k=\eta^k_k.$$

If we increment to $\mathbf{u}_{i+1}$ and maintain the condition of coincidence, then we have: $(\mathbf{u}_i)_k v^i_k=(\mathbf{u}_{i+1})_{k-1} v^{i+1}_{k-1}$; $v^{i+1}_{k-1} = \eta^{k-1}_{k+1}$ *or* $v^{i+1}_{k-1} = \eta^{k-1}_{k-1}$ which we may incorporate into the $O(\mathbf{u}_{i+1})$ sequence to result in:

$$O(\mathbf{u}_{i+1}) =\langle(\mathbf{u}_{i+1})_j \eta^j_0 \geq\ldots\geq(\mathbf{u}_{i+1})_j \eta^j_{k-1}\geq(\mathbf{u}_{i+1})_k v^{i+1}_k \geq(\mathbf{u}_{i+1})_{k-1} v^{i+1}_{k-1}\geq\ldots\geq (\mathbf{u}_{i+1})_j \eta^j_{N-1}\rangle,$$

or

$$O(\mathbf{u}_{i+1}) =\langle(\mathbf{u}_{i+1})_j \eta^j_0 \geq\ldots\geq(\mathbf{u}_{i+1})_{k-1} v^{i+1}_{k-1}\geq(\mathbf{u}_{i+1})_k v^{i+1}_k\geq(\mathbf{u}_{i+1})_j \eta^j_{k+1}\geq\ldots\geq (\mathbf{u}_{i+1})_j \eta^j_{N-1}\rangle,\ j>k. \quad (3)$$

We can see now that the forcing condition of: cor($c_{fp}(\mathbf{u}_i)$): $v^i_k=\eta^k_k$ *for i=i, i+k,* results in a predicate representation of the data, $\mathbf{u}_i$, that is a k+1 sequence of data that are either monotonically non-decreasing or non-increasing as a condition for $\Im(\mathbf{u},\mathbf{u}')$ to be a fixed-point of a median window filter of length N. Furthermore, each datum in the k+1 sequence is a member of a k+1 monotonic sequence of data. This defines a data sequence to be locally monotonic of order k+1 denoted as: LOMO(K+1). This means that for a 1-D filter along the x-axis, an object that is a root must be symmetrical in terms of local

monotonicity such that the sequence of data constituting the data object is locally monotonic on entry as well as on exit by the filter.

The predicate $\Re(\mathbf{u}_i)$ fulfills the role of representing the form and geometry of the data such that $\Im(\mathbf{u},\mathbf{u}')$ is defined and true with respect to the *intended use* of $\Im(\mathbf{u},\mathbf{u}')$. The predicate accomplishes these distinctions by *forcing conditions*[9], c, to be correct, cor(c), conditions imposed on $\Re(\mathbf{u}_i)$ for $\Re(\mathbf{u}_i)$ to be true such that: $\mathbf{u}'=\bigcup_{i=1}^{n-2k} \mathbf{u}'_I$. The sampling function, $\mathbf{u}_i$, and any forcing conditions, c, are defined by the predicate:

$$\mathbf{u}_i(u_i,\ldots,u_{i+2k}) = [x|\ \Re_c(x, u_i,\ldots,u_{i+2k}, \text{cor}(c)]$$

to mean: The set of all x such that $\Re_c$ is true. The subscript, c, indicates correct conditions, cor(c), that force $\Re_c(\mathbf{u}_i)$ reflecting the filter design and intent. The forcing condition of (3) for i=i,i+k, is tantamount to the data being LOMO(k+1), and together constitute a biconditional forcing condition of the predicate. This forcing condition is hereafter referred to for convenience as the *coincidence condition.*

3.3 The structure of FP roots

The FP root is an object in itself having a *property* that the data composing it are related and describable as one of several distinct types of locally monotonic patterns of relationships. In the case of a 1-D filter, we see the following patterns:

(a) A monotonically increasing sequence of k+1 terms, e.g., an up-ramp, or step up, of length k+1, referred to as an α pattern. The α pattern is followed by a 'ragged' plateau of k data all greater than any of the monotonic data in the ramp – a k^+ pattern - and following a similar pattern of data all less than any in the k+1 sequence – a k^- pattern. This denoted by: $k^-\alpha k^+$

(b) A monotonically decreasing sequence of k+1 terms, e.g., a down-ramp, or step down, referred to as an β pattern. The β pattern is preceded by a k^+ pattern, and followed by a k^- pattern, denoted by: $k^+\beta k^-$.

(c) A monotonic sequence of k+1 terms related by '=' rather than '<' or '>', i.e., a flat pulse, referred to as an χ pattern. The leading and trailing sequence of k terms may be independent derived from the neutrality of '=' , allowing for a 'pulse' of k+1 binary data, denoted by: $(k^+/k^-)\chi(k^+/k^-)$.

An important point here is that these FPs are *patterns* independent of the absolute pixel values.

The characteristic of a fixed-point pattern being a monotonic relationship of each of the $(u_i)_j$ to all other $(u_i)_j$, j=0,k in a window i-k to i+k and buffered by k^+ and k^- patterns suggests that there is a definable *grammar* for the composition of FPs, to comprise a complex fixed-point set, or *sentence* of data. For example, a complex of $\alpha\chi$ or $\beta\chi$, or their converse, is allowed but $\alpha\beta$ or $\beta\alpha$ are disallowed combinations. Thus follows, that the intent to discover such special phenomena as fixed-point or oscillating roots for a 1-D or 2-D filter become discoveries of $\mathbf{u}$ and cor(c) in the representation of the data such that $\wp(\mathbf{u}_i,\mathbf{u}'_i)$ iff $\Re_c(\mathbf{u}_i)$. The predicate and the functions $\wp(\mathbf{u}_i,\mathbf{u}'_i)=\mu x[H(G(\mathbf{u}_i, x), \mathbf{u}^j)]$ and $\Im(\mathbf{u},\mathbf{u}')$ constitute a conceptual framework of analysis wherein the predicate is the defining condition of the filter as remarked above, and the functions consist of an individual filtering function $\wp(\mathbf{u}_i,\mathbf{u}'_i)$ comprised of a selection function, G, of the output datum, a writing function, H, of $\mathbf{u}'_i$, and a composing function, $\Im(\mathbf{u},\mathbf{u}')$, of $\mathbf{u}'$ from $\mathbf{u}'_i$. We may extend these patterns to 2-D and see that the number of relational patterns is very limited and the superposition or co-joining of any of $\Re_c(\mathbf{u}_\zeta)$ must satisfy a grammar of patterns of relationships.

3.3.1 2-D FPs

For the case of the 2-D filter, we will analyze the 'cross' filter: $\vartheta_+(L)$, L=2q+1& L=2N-1, N=2k+1, described below. The data are assumed to be an array in X, Y with the origin in the upper left. The set of data variables, u: $\mathbf{u}=\{u_\zeta\}$, are defined in X×Y:

$$\mathbf{X}\times\mathbf{Y}=[\zeta|\delta\in\mathbf{X}\ \&\ \gamma\in\mathbf{Y}\ \&\ \zeta\in\langle\delta,\gamma\rangle]$$

so that, when expressed in the form of designator terms, the data u are: $\mathbf{u}=(\mathbf{u})\nu^\zeta$ where the index, ζ, is an ordered pair $\zeta=\langle\delta,\gamma\rangle$, $\zeta=\langle\langle1,1\rangle,\ldots,\langle n,n\rangle\rangle$; thus, $\mathbf{u}=(u_{11}, u_{12},\ldots,u_{nn})$. The $\vartheta_+(L)$ sampling function is $\mathbf{u}_\zeta=(\mathbf{u}_\zeta)_\omega\mathbf{v}^\zeta{}_\omega$; $\omega=\langle\rho,\sigma\rangle$:

$$\mathbf{u}_\zeta=\forall\rho((\mathbf{u})\,\nu^{\langle\delta+\rho,\,\gamma+k\rangle}))_{\rho<N}\ \&\ \forall\sigma((\mathbf{u})\nu^{\langle\delta+k,\,\gamma+\sigma\rangle})_{\sigma<N};$$
$$(\mathbf{u}_\zeta)_\omega=((u_\zeta)_{00},\ldots,(u_{\zeta+\omega})_\omega)$$
$$\mathbf{u}^\zeta{}_\omega=\langle\nu^\zeta{}_{00},\ldots,\nu^{\zeta+\omega}{}_\omega\rangle.$$

$$L=2N-1,\ N=2k+1,\ \text{and}\ \omega=\langle\langle0,0\rangle,\ldots,\langle2k,2k\rangle\rangle. \qquad (4)$$

This sampling function defines a 2-D filter, $\vartheta_+(L)$ shown below for L=9:

$$\vartheta_+(9)=\begin{bmatrix} & & \circ & & \\ & & \circ & & \\ \circ & \circ & \otimes & \circ & \circ \\ & & \circ & & \\ & & \circ & & \end{bmatrix}$$

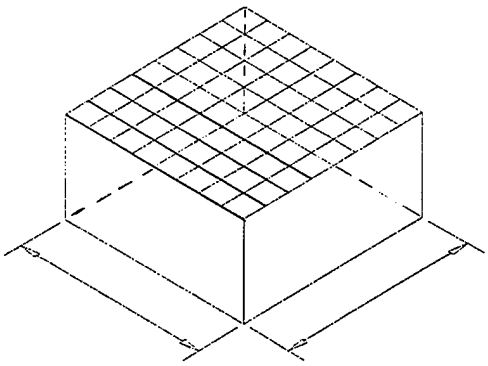
This sampling function results in a double sequence (5) representing the intersection of the two arms of the "cross" filter, $\vartheta_+(L)$, shown in figure 6.

$$\mathbf{u}_\zeta=(\langle(\mathbf{u}_\zeta)_{0k}\nu^\zeta{}_{0k},\ldots,(\mathbf{u}_\zeta)_{2k,k}\nu^\zeta{}_{2k,k}\rangle,\ \langle(\mathbf{u}_\zeta)_{k0}\nu^\zeta{}_{k0},\ldots,(\mathbf{u}_\zeta)_{k,2k}\nu^\zeta{}_{k,2k}\rangle) \qquad (5)$$
$$\mathbf{u}_\zeta=\mathbf{u}'_{\gamma k}\cup\mathbf{u}'_{k\delta};\quad \mathbf{u}'_{\gamma k}\cap\mathbf{u}'_{k\delta}=(\mathbf{u}_\zeta)_{kk}\nu^{\zeta'}{}_{kk}.$$

If we go through the 2-D analog of the 1-D analysis above for FP roots of $\vartheta_+(L)$ with $\mathbf{u}_\zeta$, we will discover that the FP roots are orthogonal strings of data with a common pixel at the intersection, each data string being a locally monotonic structure of length k+1. The FP roots detected by $\vartheta_+(L)$ in an image containing an object are represented by the predicates of the roots, $\Re_c(\mathbf{u}_\zeta)$; cor (c)=: $\mathbf{u}_\zeta$ satisfies the *coincidence condition* ($\mathbf{u}^{\delta\gamma}{}_{kk}=\eta^{kk}{}_{\lambda=q}$) *& is a member of a set of (k+1)×(k+1) data*, such that: $\Re_c(\mathbf{u}_\zeta)$; $\zeta=\langle\delta,\delta+k;\,\gamma,\gamma+k\rangle$.

The following four figures (figure 8.a-d) illustrate some the FP roots of $\vartheta_+(L)$. There are 14 fundamental FP root patterns to $\vartheta_+(L)$ and they are connectable according to a syntax allowing some combinations and disallowing others to form complex FP root structures. The root patterns are formed by a monotonic constraint in X and Y, but not necessarily on the diagonal, connecting any of the three 1-D patterns, α, β and χ on the edges. Each FP pattern is constrained to be piece-wise continuous at the vertices and buffered by the appropriate k^+ and k^- patterns. These FP root patterns, then, are a square set of data defined by the two 1-D components of $\mathbf{u}_\zeta$ that are simultaneously monotonic of order k+1 about $\omega=kk$, or LOMO(k+1, k+1). The examples are shown as follows:

A uniform pulse of data:



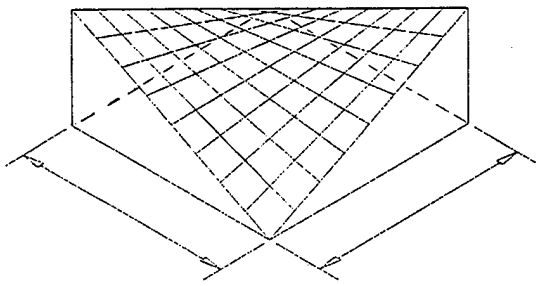
The associated example data set is:

				0	0	0	0	0				
				0	0	0	0	0				
				0	0	0	0	0				
				0	0	0	0	0				
0	0	0	0	5	5	5	5	5	0	0	0	0
0	0	0	0	5	5	5	5	5	0	0	0	0
0	0	0	0	5	5	5	5	5	0	0	0	0
0	0	0	0	5	5	5	5	5	0	0	0	0
0	0	0	0	5	5	5	5	5	0	0	0	0
				0	0	0	0	0				
				0	0	0	0	0				
				0	0	0	0	0				
				0	0	0	0	0				

(a)

Figure 1

A saddle pattern:



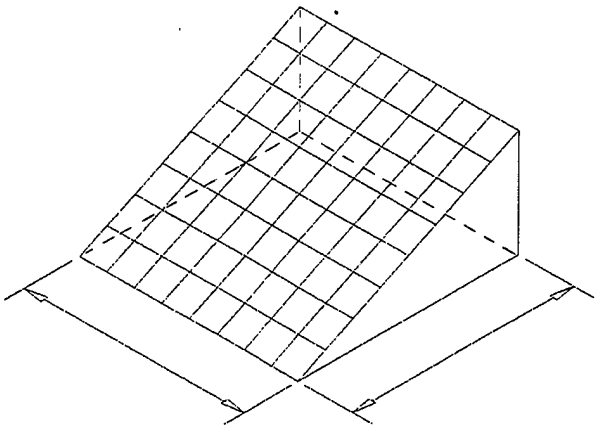
The associated example data set for this pattern, again using arbitrary data values 0,...,5, is:

				5	5	0	0	0				
				5	5	0	0	0				
				5	5	0	0	0				
				5	5	0	0	0				
5	5	5	5	5	4	3	2	1	0	0	0	0
5	5	5	5	4	4	3	2	2	0	0	0	0
0	0	0	0	3	3	3	3	3	0	0	0	0
0	0	0	0	2	2	3	4	4	5	5	5	5
0	0	0	0	1	2	3	4	5	5	5	5	5
				0	0	0	5	5				
				0	0	0	5	5				
				0	0	0	5	5				
				0	0	0	5	5				

(b)

Figure 1.

A wedge pattern:



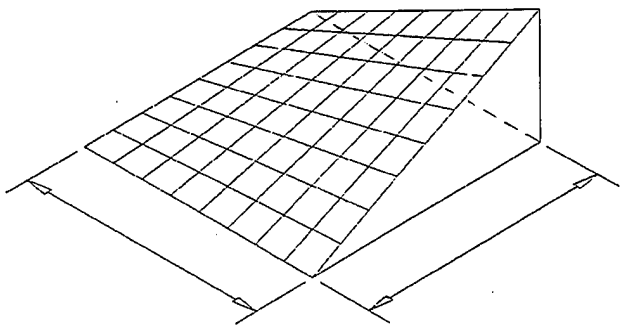
The associated data set is:

				0	0	0	0	0				
				0	0	0	0	0				
				0	0	0	0	0				
				0	0	0	0	0				
0	0	0	0	1	2	3	4	5	5	5	5	5
0	0	0	0	1	2	3	4	5	5	5	5	5
0	0	0	0	1	2	3	4	5	5	5	5	5
0	0	0	0	1	2	3	4	5	5	5	5	5
0	0	0	0	1	2	3	4	5	5	5	5	5
				0	0	0	0	0				
				0	0	0	0	0				
				0	0	0	0	0				
				0	0	0	0	0				

(c)

Figure 1.

A diagonal wedge pattern:



The associated example data set is:

				5	5	5	5	5				
				4	4	4	4	5				
				3	3	3	5	5				
				2	4	4	4	5				
0	0	0	0	1	2	3	4	5	5	5	5	5
0	0	0	0	1	2	3	4	4	5	4	5	5
0	0	0	0	1	2	3	3	3	4	5	4	4
0	0	0	0	1	2	2	2	2	3	4	4	3
0	0	0	0	1	1	1	1	1	0	0	0	0
				0	0	0	0	0				
				0	0	0	0	0				
				0	0	0	0	0				
				0	0	0	0	0				

(d)

Figure 1.

We can see quite readily now how some of these few patterns in figure 8 can be combined into a complex FP pattern. The following example, shown in figure 2, involving the wedge, pulse and diagonal wedge, is a FP pattern buffered by a k^- of "0". This pattern could be further combined with $\pi/2$ rotations of the wedge and diagonal wedge on matching edges, and buffered by k^- of "0", to form a 2-D image of a truncated pyramid. Suffice to say it could also be combined with $\pi/2$ rotations of itself to form a similar image, and other more complex and extensive FP surfaces could be imagined involving the other patterns. All in all, there are fourteen fundamental FP patterns to the $\vartheta_+(L)$ filter including the four shown in figure 1. The important fact is that there are *only* fourteen patterns.

1	2	3	4	5	5	5	5	5	5
1	2	3	4	5	5	5	5	5	5
1	2	3	4	5	5	5	5	5	5
1	2	3	4	5	5	5	5	5	5
1	2	3	4	5	5	5	5	5	5
1	2	3	4	5	5	5	5	5	5
1	2	3	4	4	4	4	4	4	4
1	2	3	3	3	3	3	3	3	3
1	2	2	2	2	2	2	2	2	2
1	1	1	1	1	1	1	1	1	1

Figure 2.

3.4 Feature extraction logic

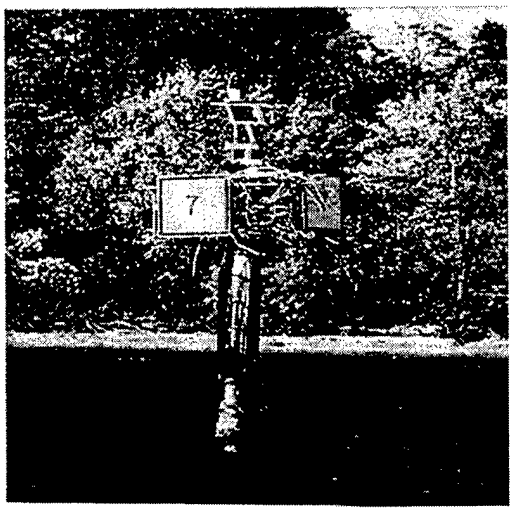
The application of the RO filter to feature extraction may be realized by enforcing a condition on the predicate such that the data satisfy the correct conditions of a FP root for the filter function to be not necessarily zero. The implementation of this logic is described by (1); $\Im(\mathbf{u},\mathbf{u}')=\wp(\mathbf{u}_1,\mathbf{u}'_1)\cdot K_{\neg\Re}(\mathbf{u}_1)\oplus,\ldots,$

$\oplus \wp(u_{n-2k}, u'_{n-2k}) \cdot K_{\neg \Re}(u_{n-2k})$ where $K_{\neg \Re}(u)=1$ for $\Re(u)=\Re_c(u_i)$ and $\neg \Re_c(u_i)$ is false; i.e., that the data satisfy cor(c) and $\Re_c(u_i)$ is true. If $\Re_c(u_i)$ is false, i.e., $\neg \Re_c(u_i)$ is true, then $K_{\neg \Re}(u)=0$ and the associated $\wp(u, u')$ contributes 0 to $\Im(u,u')$. A collection of juxtaposed, thus grammatically compatible, $\Re_c(u_\zeta)$ represents a predicate of an object in an image composed of FP roots and is denoted as:

$$R_c(u_{\zeta'}); \; R_c(u_{\zeta'}) = \bigcup_{\zeta}^{\zeta'} \Re_c(u_\zeta); \; \zeta' = \zeta + n$$

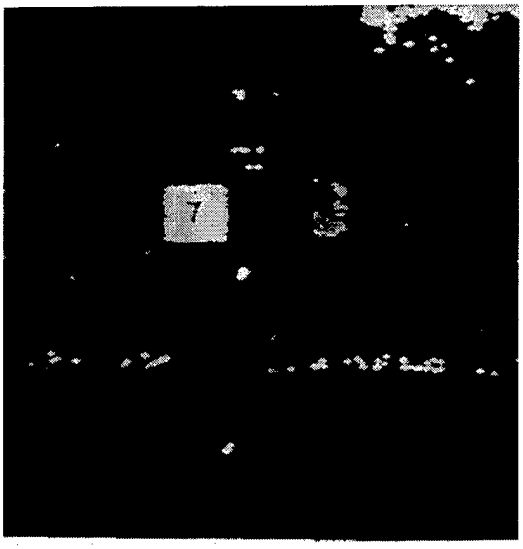
3.4.1 Preliminary results

The application of RO filters for feature extraction assumes that objects of interest are composed of a set of related features distinguished from features not comprising objects of interest by their monotonic structure, and that the composition of features to constitute the object is a relational structure. This assumption clearly does not apply to all images and objects of interest, but may apply to many cases such as objects embedded in clutter. The potential of the RO filter for feature extraction of objects embedded in clutter is illustrated as preliminary results in figure 3. Figure 3 shows the application of $\vartheta_+(L)$ iff $\Re_{c=FP}(u_i)$, L=13 constrained by the condition of coincidence and inclusion in (k+1)×(k+1) sets denoted by $\vartheta_+(13)$ iff $\Re_{c=FP\pm}(u_i)$, and also for the inclusion of the precursor and trailing 'ragged' data which define the complete root denoted by $\vartheta_+(13)$ iff $\Re_{c=FP\pm}(u_i)$. These results are shown as: (a) Original image of a navigation marker, (b) Filtered image subject to constraints for selecting LOMO(q+1) FP structures, (c) Filtered image of complete FP roots, i.e., including the k^+ and k^- data. These results demonstrate that feature extraction by this method is independent of shape and gray level of the object. The dimension of the input image was 840×1024 pixels and required approximately 100sec. Of computation by a Sun Microstation to produce figure 3(c).



(a)
Original image

(b)
$\vartheta_+(13)$ iff $\Re_{c=FP}(u_i)$



(c)
$\vartheta_+(13)$ iff $\Re_{c=FP\pm}(u_i)$
Chesapeake Navigation Marker
Figure 3

The extraction of features as shown above for use in a model of object recognition is not to suppose that this kind of representation of features in imagery is all there is to a model of human object recognition by any means. It is simply pursued as an approach to a functional model of some practical utility based on human object recognition. The human observer may interpret raw, or hyletic, data with respect to some complex intentional purpose in other ways than by the relatedness of feature morphology and texture. For example, variation in color of an image understood by an observer to be a natural scene[20] may be interpreted with respect to a complex non-veridical object determined by purpose, e.g., to walk in the cool shade. (N.B. We do not, however, want to confuse a logical model of perception with a physical model of vision.) Interpretation of a complex of objects to be a single complex object is a common mode of recognition. This is the recognition of a complex object whose features are not just those determined by the shape or texture of independent parts, or objects, but also includes features determined by the collection and, most importantly, the relationship of independent parts[4,7,19], or objects, in the image. An example of complex object recognition based on context and relationship is the recognition of the words on this page, and further, their combination into sentences.

The FPs detected by $\vartheta_+(L)$ are, as shown above, a pattern of data determined by a monotonic structure within themselves and also satisfy a relation with neighboring k data as all "$\geq$" or "$\leq$" to be a finite set of possible types of roots. If we restrict the predicates of the monotonic structures of the roots to "$=$" and "$<$" or "$>$", for $(k+1)\times(k+1)$ sets, or 'tiles', we have potentially 81 possible types of FP patterns. This number is reduced to 54, however, by the constraint of being connected at the vertices of the FP, and further reduced to 14 fundamental patterns by disregarding patterns equivalent under rotation. Thus, for our language, we have an alphabet of 14 fundamental patterns: e^β, $\beta=1,14$; $e^\beta{}_\zeta \widetilde{\leftrightarrow} \Re_{c=FP}(u_\zeta)$.

The relations required between the monotonic data and neighboring k data, or 'ragged' plateaus, result in the roots being constrained by rules of combination so that the co-joining or partial superposition of roots satisfy a kind of grammatical compatibility. That the $\{e^\beta\}$ are a set of relational patterns independent of absolute pixel value means that they represent the property of relation in that region of the image independently of the extension of the image, thus the $\{e^\beta\}$ are an intension of the image in that region. An image of a feature represented by an ordered sequence of e^β: $\langle e^\beta{}_\zeta,\ldots,e^\beta{}_{\zeta'}\rangle$ is then an intensional representation of that feature.

As a practical matter, we cannot say that *any* image of that feature will have the same representation by $\langle e^{\beta}_{\zeta},\ldots,e^{\beta}_{\zeta'}\rangle$. We can say, however, that any image from a set of images of that feature obtained within the limits of linear exposure, equivalent range, perspective and resolution, and digitized at the same level will have the same intensional representation. Furthermore, as individual objects are members of a sort of object, features are members of a sort of feature and images of them form families of images over range and perspective. The result is usually a fairly large set of possible images for which the intensional property of the feature holds true in practice.

4.0 LANGUAGE MODEL BASED ON FP ROOTS

The notion of a language based on an alphabet of e^{β} is that a feature, d, composed of e^{β}_{ζ} is an ordered sequence Δ_d: $\langle e^{\beta}_{\zeta},\ldots,\widehat{e^{\beta}_{\zeta'}}\rangle$; $\Delta_d(e^{\beta}_{\zeta}) \widetilde{\leftrightarrow} \mathbb{R}_c(\mathbf{u}_{\zeta'})$ that satisfies a syntax and has semantic content. The sets $\Delta_d(e^{\beta}_{\zeta})$ that have semantic content in the sense of being a true or false valuation, i.e., $\nu(\Delta_d(e^{\beta}_{\zeta}))\in\{T,F\}$, in the vocabulary of a semantic model are atomic sentences closed under subsentences e^{β}_{ζ}. Sets $\Delta_d(e^{\beta}_{\zeta})$ that do not have this kind of atomic semantic value, i.e., $\nu(\Delta_d(e^{\beta}_{\zeta}))\notin\{T,F\}$, are molecular expressions. Thus, if any $\Delta_d(e^{\beta}_{\zeta})$ belongs to a vocabulary of features it is atomic, and if not, then it is molecular. N.B. that the $\Delta_d(e^{\beta}_{\zeta})$ are closed under atomic sentences, perhaps better understood as here as well-formed formulas (wffs), which are the individual e^{β} and their combinations by the logical connectives. A sentence A: $\nu(A)\in\{T,\ F\}$ is a wff.

4.1 Logical model structure

In the simplest terms, a semantic language is a model of a propositional syntactic system (PCS)[10]. The system is a triple $\langle$A, L, S$\rangle$ where:

A: A set of denumerable atomic sentences.

L: A set of logical symbols, e.g., { $\neg$, $\rightarrow$, $\wedge$, $\vee$, (,)}. Respectively: '$\neg$': negation; read as: "not", '$\rightarrow$': conditional or entailment; read as: "if...then...", '$\wedge$': conjunction; read as: "and", '$\vee$': disjunction; read as: "or", and '(,)': parentheses.

S: The smallest set of sentences including A such that if A, B $\in$S, then so are $\neg$A and (A$\wedge$B).

A language model of this system includes the concept of valuation of sentences composed by the logical connectives of atomic sentences, thus a language model includes the semantics of the calculus.

Let us describe a language with the construction[1] $\mathcal{L}_o$; $\mathcal{L}_o(\gamma_0,\ \gamma_1,\ldots,\ \gamma_n;\ p_0,\ p_1,\ldots)$. The p's are propositional variables and in the case of this language they are substituted with e^{β}. The $\gamma_{i(j)}$ are i(j)-ary connectives of the i^{th} type connecting j variables, e.g., $\gamma_{i(r)}(p_0,\ p_1,\ldots,p_r)$,. We will show later that the number of γ_i and p_i must be finite. In our case of the e^{β}, we recall that connection of any two may or may not be allowable depending on which types they are. Further, we now see that any logically allowable combination of the e^{β} must also be in the vocabulary of the language for it to have semantic value. In this case, we say the γ_i are truth-and-relation functional connectives if there is a truth table for determining the truth value of $\gamma_i(p_0,\ p_1,\ldots,p_r)$. The truth value is determined according to the truth value of the individual p_i, and if the p_i are related by a relationship $R_{\gamma(i)}(p_0,\ p_1,\ldots,p_r)$ governing γ_i permitted by the language. The relationship $R_{\gamma}(p_0,\ p_1,\ldots,p_r)$ introduces the relatedness component of the logical system for a language model to be a semantic structure of non-classical logic. This form of non-classical logic is distinguished from classical logic that considers only form and truth-value neglecting content and relationship. The language $\mathcal{L}_o$ is based on a non-classical logic.

Non-classical logic is defined by two types of relational structure, a set-assignment semantics and a relations-based semantics, although in the end, the sentences defined by either are semantically equivalent. A complete model must consider both, but our immediate concern here with the structure of the language will focus on the relation-based type. The set assignment type simply states that a logical

connective of two or more wffs is constrained by commonality of content. In the set assignment approach, we assume there to be a set , **S**, of contents, content here being any set of named veridical or non-veridical objects. The content of a wff A is s(A) and similarly for wff B, s(B). The valuation of a connective γ of wffs A and B, $v(\gamma(A,B))$ =T iff $R_c(A, B)$ is true; $R_c(A, B)$ =s(A)$\cup$s(B). In a similar manner, the connective itself may be subject to a relation and that defines a relation-based semantics. The relation-based semantics is the case described here for a structure based on e^{β}.

In the case of a relations-based semantics, we have for A and B :

$$v(\gamma(A,B)) = \text{T iff } R_\gamma(A, B) \text{ is true, i.e., } \gamma(A,B) \overset{\sim}{\leftrightarrow} R_\gamma(A, B).$$

A relations-based semantics model for $\mathcal{L}_o$ may be described in rather general terms[2] as:

$$\mathcal{L}_o(\gamma_0, \gamma_1,\ldots, \gamma_n; e^1, e^2,\ldots, e^n).$$

$$\downarrow \text{ realization}$$

$$\Delta_1(e^{\beta}{}_{\zeta}), \Delta_2(e^{\beta}{}_{\zeta}) ,\ldots \text{ complex propositions composed using } \gamma_i$$

$$\downarrow \text{ } v, R \text{ and truth tables}$$

$$\{T, F\}$$

This model can be simplified conceptually, skipping the intermediate step 'realization', by including the connectives into the truth tables with a truth function of the connectives, f_i: $f_0, f_1, \ldots, f_n$ for $\gamma_0, \gamma_1,\ldots, \gamma_n$. The relation $R_\gamma(A, B)$ is simply the relationship governing allowable combinations of A and B. The truth-and - relation function f_i is the calculation of the truth value of the combination of A and B given their individual existence as T or F. Suppose that $v(e^1)$ =T or F and $v(e^2)$ =T or F, then for a simple logical operation of γ_i: "$\wedge$" , $f_\wedge(v_j(e^1), v(e^2))$ is:

$$f_\wedge(T, T) = T,\ f_\wedge(T, F) = F,\ f_\wedge(F, T) = F,\ f_\wedge(F, F) = F,$$

if and only if the combination of e^1 and e^2 by $\wedge$ is an allowed combination according to $R_\gamma(e^1, e^2)$=T. Conversely, $f_\wedge(v_j(e^1), v(e^2))$=F if $R_\gamma(e^1, e^2)$=F. Thus we have $f_\wedge(v(e^1), v(e^2)) \in \{T, F\}$ iff $R_\gamma(e^1, e^2)$=T. We may combine these notions to get a complete statement of truth conditions as:

$$v(\gamma_j(e^1, e^2)) = T \text{ iff } R_\gamma(e^1, e^2)=T \ \&\ f_\wedge(v(e^1), v(e^2))=T.$$

A formal relation-based model, M, for $\mathcal{L}_o$ is given then by

$$M = \langle v, R_\gamma, \gamma_0, \gamma_1,\ldots, \gamma_n; e^1, e^2,\ldots, e^n; \Delta_1(e^{\beta}{}_{\zeta}), \Delta_2(e^{\beta}{}_{\zeta}) ,\ldots\rangle \qquad (5)$$

where:

$\Delta_1(e^{\beta}{}_{\zeta}), \Delta_2(e^{\beta}{}_{\zeta}) ,\ldots$ complex propositions composed using γ_i
v is the valuation, $v(p) \in \{T, F\}$
$R_\gamma \subseteq \text{Sub}(\text{wffs}(i_j))$ is the relation governing the truth table for γ allowed by logical compatibility and also by the vocabulary of the language.

The valuation, v, is applied to the complex propositions $\Delta_i(e^{\beta}{}_{\zeta})=\gamma_j(e^1, e^2,\ldots, e^n)$ by:

$$v(\gamma_j(e^1, e^2,\ldots, e^n)) = T \text{ iff } \left\{ \begin{array}{c} R_j(e^1,\cdots,e^n) \\ \textit{and} \\ f_j(v(e^1),\cdots,v(e^n)) = T \end{array} \right\}. \qquad (6)$$

In our general case, the logical functions f_i and connectives γ_i are in 2-D and are more complex than $\neg$, $\wedge$, $\vee$ as we have cases of partial superposition and co-joining along the four edges of a FP pattern and suffer rotation. The logical connectives γ_i allowed by R_γ in the language based on e^β will likely be determined by a representation of a symmetric group, and a feature may be represented by complex propositions, or sentences, in a hierarchy of f_i as a schema, θ: $\theta_k(f_i, R_c, R_\gamma)$, and also closed under sub-wffs. The general model for the language $\mathcal{L}_\sigma$ then is:

$$M = \langle v, \theta_k(f_i, R_c, R_\gamma), \gamma_0, \gamma_1,\ldots, \gamma_n; e^1, e^2,\ldots, e^n, \Delta_1(e^\beta\zeta), \Delta_2(e^\beta\zeta), \ldots\rangle.$$

This completes the description of the language model based on an alphabet of the FPs of $\vartheta_+(L)$ defining structure, terms and functions. As remarked, the 2-D representations of the γ_I, f_i and R_γ are the subject of further research, but for our purposes here we may assume the representation of γ_i, f_i and R_γ and proceed to examine the structure of this syntax with respect to the semantic logic of the system based on $v(\gamma_j(e^1, e^2,\ldots, e^n))$. Our purpose is to determine that the model M of $\mathcal{L}_\sigma$ at least allows the properties of a logical system we can use, i.e., that it is a propositional language system. To do this, we will first analyze the properties of a propositional language system, and then compare it to $\mathcal{L}_\sigma$.

4.2 General language model properties

The intent of what follows is to examine the requirements for a language in terms of structural and logical properties necessary for distinguishing features, and to relate those requirements to the language $\mathcal{L}_\sigma$ described above. I must stress that it is important to keep in mind what we are and are not attempting to do. We are not attempting to design a model for object recognition; we are exploring the plausibility of a language for such a model, a language based on a direct, algorithmic interpretation of an image.

We need to look at the properties a language may have and understand them with respect to our purpose, i.e., properties that a language of imagery must have. To do this, we need to define a few concepts of classical logic systems incorporating the notions of syntax and semantics. The property that stands out in importance is the property of *finitary entailment.* Finitary entailment can be illustrated by the theorems for transitivity and semantic-syntactic deduction for finite consequences as follows:

Transitivity: $\Gamma\cup\{A_1,\ldots, A_n\} \models B$ iff $\Gamma \models A_i$, i=1.n, then $\models B$

Semantic consequences: $\Gamma\cup\{A_1,\ldots, A_n\} \models B$ iff $\Gamma \models (A_1\wedge A_2\wedge\ldots\wedge A_n)\rightarrow B$,

where '$\models$' means 'validates', e.g., $A \models B$: A validates B, or B is a semantic consequence of A. The case of $\models A$ means: A is a tautology, or is true in every model. A wff that is always true in every model is a *valid* wff. This theorem is an analog to the case of finite syntactic consequences:

$$\Gamma\cup\{A_1,\ldots, A_n\} \vdash B \text{ iff } \Gamma \vdash A_1\rightarrow(A_2\rightarrow(\ldots\rightarrow(A_n\rightarrow B)\ldots)$$

or

$$\Gamma\cup\{A_1,\ldots, A_n\} \vdash B \text{ iff } \Gamma \vdash\!\!\!\vdash((A_1\wedge A_2\wedge\ldots\wedge A_n)\wedge\neg B).$$

In this expression, '$\vdash$' is read as ' is deducible from', e.g., $\Sigma \vdash B$: B is deducible from Σ in a logical proof. In the case of $B\in\{\Sigma\}$, then we refer to B as a theorem of the system Σ and may denote it as $\vdash_\Sigma B$. In this case, we may also refer to a theory Σ closed under a rule, e.g., *modus ponens,* denoted by $Th(\Sigma) = \{A: \Sigma \vdash A\}$.

We can understand these two theorems as two forms of expressing a logic: semantic and syntactic, semantic being in terms of the truth of propositions and syntactic in terms of the theoremhood of propositions. These theorems are founded on the notions of consistency and completeness of a system, and include the important property of compactness.

I.

(a) A system is *consistent* if all of its theorems are valid well-formed-formulas (wffs), meaning always true, in all models, i.e., all theorems are tautologies: If $\vdash A$, then $\models A$.

(b) A system is *complete* if all valid wffs are also theorems of the system: If $\models A$, then $\vdash A$.

(c) A system that is both consistent and complete is *strongly complete.*: $\vdash A$ iff $\models A$.

(d) A system is *compact* in either a semantic or syntactic sense of : $\Gamma \models A$ iff there is a finite $\Delta \subseteq \Gamma$ such that $\Delta \models A$, and similarly for '$\vdash$'. Thus, compactness incorporates the notion of finitenesss.

The notions of consistency, completeness and compactness imply some other important properties of systems. A compact system has a finite model of it. Furthermore, it is without contradiction (consistent), and is a model of all its elements (complete). Succinctly stated[3]:

II.

(a) Consistency means that for every A, $\Gamma \vdash A$ or $\Gamma \vdash \neg A$, but not both.
(b) Completeness means that for every A, A or $\neg A$ is in Γ.
(c) If a system Σ is consistent and there is a D: $\Sigma \nvdash D$ (using '$\nvdash$' to mean: D is not deducible from Σ), then there is a strongly complete Γ such that $D \notin \Gamma$ and $\Sigma \subseteq \Gamma$.
(d) Every consistent set of wffs has a model.
(e) if Γ is strongly complete, then Γ has a model and every finite subset of Γ has a model.
(f) If Γ is consistent, then there is some A: $\Sigma \nvdash A$.

These definitions and theorems serve to introduce the basic notions of semantic and syntactic entailment, consistency, completeness, and compactness as properties of language systems. To explore the structure of a non-classical language based on these properties, however, we need to change our approach to a more powerful, or flexible, metalanguage.

We introduce this metalanguage by recalling the notion of valuation v given above as $v \in \{T, F\}$ to be a mapping of sentences into {T, F}. If a valuation maps all sentences of a language into {T, F}, the language is a *bivalent* language, in this case specifically a bivalent propositional language. A valuation is an *admissible* valuation if it is a member of a set of points, VL; $v \in VL$, associated with the closure set of wffs of a language $\mathcal{L}$. The truth valuation space of a wff A is $H(A) = \{ v \in VL; v(A) = T\}$. H(A) is the truth set of A, meaning the set of all point in VL where A is true, or more logically stated: the set of points where A is *satisfied.*. The valuation space of the language L is: $H = \langle VL, \{H(A); A \in \mathcal{L}\}\rangle$. We have several useful definitions that follow from this concept[11]:

III.

(a) A wff is a *valid* wff, $\models A$, in $\mathcal{L}$ iff every admissible valuation in H of $\mathcal{L}$ satisfies A, i.e., $v(A) = T$ for all v, $v \in VL$.
(b) A set of wffs X is *unassailable* if every admissible valuation v of $\mathcal{L}$ satisfies some member of X.
(c) A set X of $\mathcal{L}$ *semantically entails* A, $X \models A$, in $\mathcal{L}$ iff every v of $\mathcal{L}$ that satisfies X also satisfies A.
(d) The set $H(X) = \bigcap_{A \in X} H(A)$ is the *elementary* class of X. A union of elementary classes that span H is the *cover* of H.

These results are a kind of re-statement of those given in **II.** above following from consistency and completeness, but defined here in set-theoretic terms. With this approach, we may proceed to develop a set of constructs that enable us to describe a language in set theoretic terms necessary to understand its semantic structure. We may begin by summarizing the above in conclusions as follows (Def. : $\varnothing =_{def}$ the null set.)

IV.

(a) $\models A$ iff H(A) =H.

(b) X is unassailable iff $\bigcup_{A \in X} H(A) = H$, i.e., if the truth set of all of X is the cover of H.

(c) X is satisfiable iff $\bigcap_{A \in X} H(A) \neq \varnothing$, i.e., if the elementary class of X is not empty.

(d) $B \models A$ iff $H(B) \cap H(A) \neq \varnothing$.

(e) $X \models A$ iff $\bigcap_{B \in X} H(B) \subseteq H(A)$, i.e., if the elementary class of X is a subset of truth set of A.

We may now re-define compactness in terms of intersection and union, leading in turn to the notion of *convergence* useful later in the discussion of filters (logical sense). We can see in **IV** above, particularly in **IV.** (d) and **IV.** (e), the genesis of relatedness logic, and we will relate the properties of the classical logic structure of a filter to the non-classical relatedness logic of the model M of $\mathcal{L}_o$.

Compactness is described in two forms[12]: I-compact (intersection), and U-compact (union).

V.

(a) A language $\mathcal{L}$ and its valuation space H is *I-compact* iff for any set X of wffs in $\mathcal{L}$, $\bigcap_{A \in X} H(A) = \varnothing$ only if $\bigcap_{A \in Y} H(A) = \varnothing$ for any finite subset Y of X. This is the same as saying that the property of I-compactness means that any set in $\mathcal{L}$ is satisfiable iff all of its finite subsets are satsifiable.

(b) A language $\mathcal{L}$ and its valuation space is *U-compact* iff for any set X of $\mathcal{L}$, $\bigcup_{A \in X} H(A) = H$ only if $\bigcup_{A \in Y} H(A) = H$ for some finite subset Y of X. This is same as saying that the property of U-compactness means that any set in $\mathcal{L}$ is unassailable only if it has a finite unassailable subset.

(c) A language that is both I-compact and U-compact is *compact.*

Finitary semantic entailment is definable in these terms now as a property of a language: $X \models A$ iff for any X of $\mathcal{L}$ and a wff A of $\mathcal{L}$, $H(X) \subseteq H(A)$ only if $H(Y) \subseteq H(A)$ for some finite subset Y of X.

The language we have discussed in connection with the model M, is as we have said, a bivalent language. A bivalent language has the inherent property of *exclusion negation.* A language has this property if for every wff A of $\mathcal{L}$, there is an A* of $\mathcal{L}$ such that: H(A*) =H - H(A). A basic theorem is given by van Fraasen[13] connects compactness and finitary entailment for a language having exclusion negation:

Theorem 4.2-A: If a language $\mathcal{L}$ has exclusion negation, then:

(a) $\mathcal{L}$ is I-compact.

(b) $\mathcal{L}$ is U-compact.

(c) $\mathcal{L}$ is compact.

(d) $\mathcal{L}$ has finitary entailment.

This theorem result for finitary entailment is conditional, however, and not necessary. For finitary semantic entailment to be necessary, we need the property of convergence that is supplied by the construction of a filter.

Let me first digress for a moment to the subject of finite axiomatizability, referring the reader to van Fraasen[14] for a detailed discussion. A *system* in a language $\mathcal{L}$ is a set X such that any A in $\mathcal{L}$ is semantically entailed by X. If X is such a system, then a set Y is a set of axioms for X if all of X are semantically entailed by Y. To be applied generally for any set of sentences in a language that may have an infinite set of complex sentences, we need to invoke the notion of semantic equivalence. Any two sets X and Y are semantically equivalent if H(X)=H(Y). A set X is finitely axiomatizable in $\mathcal{L}$ iff X *can be* semantically equivalent to some finite set of sentences Y in $\mathcal{L}$.

4.3 Filtration and $\mathcal{L}_o$

We saw that for $\mathcal{L}$ having exclusion negation and compactness implies finitary entailment, but not necessarily. The missing condition is the convergence of a filter. The notion of filters has found application in logic as a tool for proving compactness and finitary entailment, and that is a use here as well, but we also appeal to filters as a tool for distinguishing features by finitary semantic entailment, i.e., semantic deduction for finite consequences. To do this, we must understand filters and show that the language $\mathcal{L}_o$ is compatible with the structure of filters. Finitary entailment is intuitively necessary in order to make a deducible assertion in a language based on finite evidence. Convergence is also intuitively necessary in order that an assertion is consistent, or unambiguously understandable in that language. Even if the assertion is a disjunction, e.g., S is P or Q or ..., it must be the same disjunction given same the evidence for the assertion.

A filter, $\Im$, is defined on a set of sentences X in terms of the valuation space H(X) to be the set X in $\Im$ such that :

VI.

(a) $\varnothing \notin \Im$.
(b) If $Y \in \Im$ and $Y \subseteq Z \subseteq X$, then $Z \in \Im$.
(c) If $Y \in \Im$ and $Z \in \Im$, then $Y \cap Z \in \Im$.

This definition in terms of H on $\mathcal{L}$ leads to $\mathcal{L}(\Im)=\{A: H(A) \in \Im\}$. From this follows that if for i=1,n; $\{A_i\} \in \mathcal{L}(\Im)$, then if $\{A_i\} \models B$ in $\mathcal{L}$, then if $B \in \Im$ $\mathcal{L}$ has finitary entailment and $\mathcal{L}(\Im)$ is a system. A filter $\Im$ may contain sub-filters, or *filter bases*, $\mathcal{B}$, that can generate filter $\Im$ such that $\Im$ contains $\mathcal{B}$. A filter $\Im$ on X is an *ultrafilter* if there is no filter on X that contains $\Im$ as a proper part, i.e., that it is a maximal element as the basis for including all of X subject to H(X); $v(X)=T$. That is:

Ultrafilter $\Im$: {A: H(A); $v(X)=T$ for every $A \in X$ in $\mathcal{L}$},

and every filter base is contained in an ultrafilter.

The notion of maximal element in itself implies finiteness, but more rigorously, X must be finite since the system is defined for A, B, $\neg A$ and $(A \wedge B)$, and we cannot have a sentence that is a maximal element of an infinite conjunction. (N.B. in this case, we have the system defined for the truth-functionally complete $\mathcal{L}(\neg, \wedge)$.) This construction of the system also defines the filter to be defined on H over the closure set of X and it is finite, thus the union of the elementary classes is the closure set of X.

VII.

If $\Im$ is an ultrafilter on X, then[15]:
(a) $Y \cup Z \in \Im$ iff $Y \in \Im$ or $Z \in \Im$, for all $Y, Z \subseteq X$.

(b) for every $Y \subseteq X$, either $Y \in \Im$ or $X\text{-}Y \in \Im$.

We need to show that if $\mathcal{L}(\Im)$ is bivalent, compact and convergent, then $\mathcal{L}(\Im)$ has finitary semantic entailment. First, we need to establish convergence in terms of compactness. A filter on H is:

VIII.

(a) *U-convergent* to v in H iff every elementary class containing v belongs to $\Im$.
(b) *I-convergent* to v iff every elementary class in $\Im$ contains v.
(c) *Convergent* to v iff an elementary class belongs to $\Im$ iff it contains v.

I refer the reader to van Fraasen[16] for proofs of his theorems as follows establishing finitary semantic entailment if $\mathcal{L}(\Im)$ is bivalent and compact, and then move on to the compatibility of $\mathcal{L}(\Im)$ and $\mathcal{L}_o$.

Theorem 4.3-A: If every ultrafilter on H is I-convergent(U-convergent) then H is I-compact(U-compact).
Theorem 4.3-B: If every ultrafilter on H of $\mathcal{L}$ converges, then $\mathcal{L}$ has finitary semantic entailment and is compact.
Theorem 4.3-C: If H is the valuation space of a bivalent propositional language $\mathcal{L}$, then every ultrafilter on H is convergent.

We have established that language $\mathcal{L}_o$ is finite and bivalent, thus by theorem 4.2-A, it is compact and has finitary entailment. By theorems 4.3-A-C, we have that an ultrafilter on a $\mathcal{L}$ is convergent and has finitary semantic entailment, i.e., that $\mathcal{L}(\Im)$ is a system. It may have seemed that we could have gotten this far with just theorem 4.2-A, but we needed the concept of filters to distinguish features in $\mathcal{L}_o$, thus we need to show that $\mathcal{L}_o(\Im)$ is a system to have these properties. What we have not shown is the truth functional completeness of the connectives γ_j, their truth functions f_i and R_γ. If we may assume a satisfactory development of the γ_j, R_γ and f_i, then we can most easily show how they are incorporated into this structure as $\mathcal{L}_o(\Im)$ by example.

4.3 Example $\mathcal{L}_o(\Im)$

Let us imagine a simple language of atomic sentences p, q and r, and the truth functionally complete syntax of $\mathcal{L}_o(\neg, \wedge)$. Further, let us imagine the complex connectives γ_j to be simply represented by the conventional connective '$\wedge$' subject to the relatedness logic predicate of R(A, B). Now then, suppose the set of sentences p, q, r are subject to R(p, q)=1, R(p, r)=0, and R(q, r)=1. We may imagine this language as analogous to the 1-D MW FP roots with p & r corresponding to the two kinds of ramps, and q corresponding to the flat pulse. We may now compose truth table for a partial listing of sentences as: p, q, r, (p$\wedge$q), (p$\wedge$r) and (q$\wedge$r). With these three atomic sentences (wffs), we have $H=2^3$, H=8 for $v_1,\ldots,v_8$. The truth table is shown as follows (T=1, F=0):

	p	q	r	$\neg$p	$\neg$q	$\neg$r	p$\wedge$q: R(p, q)	p$\wedge$r: R(p, r)	q$\wedge$r: R(q, r) ...
v_1	1	1	1	0	0	0	1	0	1
v_2	1	1	0	0	0	1	1	0	0
v_3	1	0	1	0	1	0	0	0	0
v_4	1	0	0	0	1	1	0	0	0
v_5	0	1	1	1	0	0	0	0	1
v_6	0	1	0	1	0	1	0	0	0
v_7	0	0	1	1	1	0	0	0	0
v_8	0	0	0	1	1	1	0	0	0

Table 1
Truth table of $\mathcal{L}_o(\neg, \wedge, R, p, q, r)$

We can demonstrate the properties of $\mathcal{L}_f$ using the tools developed above. We may construct a filter base $\mathcal{B}_0$ on a subset Y_0 of X; Y_0: p, q as: $\mathcal{B}_0$ = H(p), H(q), and H(p∧q).

$$\mathcal{B}_0 = \{v_1, v_2, v_3, v_4\}, \{v_1, v_2, v_5, v_6,\}, \{v_1, v_2\}.$$

The filter $\mathcal{B}_0$ is a base because Y_0: p, q, and p∧q is satisfiable, i.e., $\bigcap_{A \in Y_0} H(A) = \{v_1, v_2\}$, which is the elementary class of Y_0. We may also observe that the filter base $\mathcal{B}_0$ is the valuation space of the closure set of p∧q. The filter base $\mathcal{B}_0$ generates a filter $\mathfrak{I}_1$ over the complete subset $Y \supseteq Y_0$ that contains $\mathcal{B}_0$, $\mathcal{B}_0 \subseteq \mathfrak{I}_1$: [N.B. (a∨b)=¬(¬a∧¬b).]

$$\mathfrak{I}_1 = \text{H(p), H(q), H(p∧q), H((p∧q)∨(¬p∧¬q)), H(¬(¬p∧¬q)), H(¬(¬p∧q)), H(¬(p∧¬q)), H(p∨¬p).}$$

This corresponds to:

$$\mathfrak{I}_1 = \{v_1, v_2, v_3, v_4\}, \{v_1, v_2, v_5, v_6,\}, \{v_1, v_2\}, \{v_1, v_2, v_7, v_8\},$$
$$\{v_1, v_2, v_3, v_4, v_5, v_6\}, \{v_1, v_2, v_3, v_4, v_7, v_8\},$$
$$\{v_1, v_2, v_5, v_6, v_7, v_8\}, \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8\}.$$

This is a more interesting filter than $\mathcal{B}_0$. We see that the tautology, H(p∨¬p) = H, and the contradiction, H(p∧¬p) =∅. Note that exclusion negation accounts for not including H(¬(p∧q)). Further, that $\bigcap_{A \in Y} H(A) = \{v_1, v_2\}$ and $\bigcup_{A \in Y} H(A) = H$, and that $\bigcap_{A \in Y} H(A) \subseteq H(p \rightarrow q)$, i.e., $\{v_1, v_2\} \subseteq \{v_1, v_2, v_5, v_6, v_7, v_8\}$, thus $Y \models q$ [N.B. p→q=¬(p∧¬q)]. This shows that $\mathcal{L}_f$ is I-compact and U-compact for a finite subset Y of X, thus for every X in $\mathcal{L}_f$, $\mathcal{L}_f$ is compact and has finitary semantic entailment including finite axiomatizability. The filter $\mathfrak{I}_1$, however, is not an unltrafilter because it is itself contained in a filter $\mathfrak{I}$ over X: p, q, r, and its family of conjunctions. Adding r to the subset Y to be the set X results in

$$\mathfrak{I} = \text{H(p), H(q), H(r), H(p∧q), H((p∧q)∨(¬p∧¬q)), H(¬(¬p∧¬q)), H(¬(¬p∧q)), H(¬(p∧¬q)),}$$
$$\text{H(q∧r), H((q∧r)∨(¬q∧¬r)), H(¬(¬q∧¬r)), H(¬(¬q∧r)), H(¬(q∧¬r)), H(p∨¬p) .}$$

The filter $\mathfrak{I}$ shows that the language is I-convergent to $\{v_1\}$ - **VIII**(b) - and U-convergent to $v \in H$ - **VIII**(a) -, and has finitary semantic entailment of all sentences containing r, but not (p∧r). Furthermore, the filter $\mathfrak{I}$ is an ultrafilter as it contains $\mathfrak{I}_1$; $\mathcal{B}_0 \subseteq \mathfrak{I}_1 \subseteq \mathfrak{I}$ and there is no filter that contains $\mathfrak{I}$ as a proper part, thus $\mathcal{L}_f(\mathfrak{I})$ is a non-classical propositional language system. The effect of R(p, r)=0 is to reduce the number of truth sets in $\mathfrak{I}$, and it provides the structure of sub-filters, but has no effect on the compactness and convergence of $\mathcal{L}_f(\mathfrak{I})$. If we were to substitute $(e^1, e^2, \ldots, e^n)$ for (p, q, r,...,s), and the γ_j and f_1, were found to be truth functionally complete, then $\mathcal{L}_f$ is $\mathcal{L}_o$ and we can see that $\mathcal{L}_o$ would satisfy the requirements for $\mathcal{L}_o(\mathfrak{I})$ to be a satisfactory language structure for a linguistic transform of features in imagery. This is the principle result:

IX.

If $f_0, f_1, \ldots f_n$ for connectives $\gamma_j(e^1, e^2, \ldots, e^n)$, $\gamma_0, \gamma_1, \ldots, \gamma_n$; $R_\gamma(e^1, \ldots, e^n)$ in $\mathcal{L}_o$ are truth-functionally complete, then $\mathcal{L}_o(\mathfrak{I})$ is a non-classical propositional language system that is compact, bivalent and convergent.

5.0 CONCLUSION

We conclude that with suitable, and expected, development of R_γ, γ_j and f_i, the language $\mathcal{L}_o$ can be $\mathcal{L}_\mathcal{J}(\mathfrak{I})$, a non-classical propositional language system. The utility of $\mathcal{L}_\mathcal{J}(\mathfrak{I})$ for a logical model of objects is derived from the fact that $\mathfrak{I}$ is closed under superset formation as we have shown with $\mathfrak{B}_0 \subseteq \mathfrak{I}_1 \subseteq \mathfrak{I}$. The closure under superset formation allows for the definition of subsets of $\mathcal{L}_\mathcal{J}(\mathfrak{I})$ to select various $\Delta_d(e^\beta)$ subject to $f_i,(v_i(e^1),\ldots, v_i(e^n))$; $R_j(e^1, e^2,\ldots, e^n)$ in a propositional language of a logical model. The limitation for developing a language model of features based this language is the distinction of actual objects by features sampled by the MW digital filter, $\vartheta_+(L)$, and that the intensional representations of features by $\Delta_d(e^\beta{}_\mathcal{L})$ are semantically entailed by an intensional property, $\Delta^i_d(e^\beta)$, of the features true in any practical image presenting them. In other words, the practical limits of the theory presented above to be a tool for interpreting reality must be validated by experience and refinement of practice. We may be encouraged that the theory is supported by our experience that we can recognize a few FP images of objects, e.g., figure 1, and that this method is a model of our perception.

Given $\mathcal{L}_\mathcal{J}(\mathfrak{I})$ and reasonable success in sampling of objects, such as shown in figure 1, this language would permit the direct machine interpretation of an image feature by $\vartheta_+(L)$ into a sentence expressing it as a feature of an object that is a member of a *sort* of object indexed by a noun[5]. The notion guiding further research is that the noun is a filter contained in the ultrafilter of $\mathcal{L}_\mathcal{J}(\mathfrak{I})$. The principal risk in this program is in establishing the distinction of sorts based on $v(\gamma_j(e^1, e^2,\ldots, e^n))$ iff $R_\gamma(e^1,\ldots, e^n)$. Theoretical development and refinement of practice will mitigate this risk. There will also be statistical problems to address in designing the optimum acquisition and conditioning of the imagery for a given filter dimension. The most likely first application of this research would be in character recognition as the simplest case with more complex object recognition to follow.

Establishing the viability of this language will enable us to seriously consider connecting the disciplines of image acquisition and processing to the disciplines of artificial intelligence. If we can develop the automated interpretation of an object in an image to result in a noun expression of its sort, then it is reasonable to further consider the plausibility of interpreting it as an object of meaning. This would be accomplished by associating its noun expression with a possible world(s) of experience - our experience expressed in propositions - in a logical model implemented by a computer. We should observe a cautionary note, however, that the phenomenon of human language is much more complex than any logical model of a language system. Human language is the nexus of all that is conscious life; it is our mode of interpreting reality to ourselves and, most importantly, to each other constituting a community. That said, what we are doing here is only attempting to emulate a human function of cognition; we are not trying to emulate consciousness through some sort of simulation. The results presented here describe how we may recognize a word, or to assign a name to an object, that is in the vocabulary of a language. The audacious suggestion is that we may consider the more complicated task of interpreting what that word means in the context of a sentence, or group of objects. The cautionary note is to beware that it is one thing to understand a sentence such as to be able to answer a question, but it is quite another thing to ask a question in fulfillment of a purpose. The desired outcome of this research is simply a useful tool that we may employ in our endeavors.

6.0 ACKNOWLEDGMENTS

The photograph in Figure 1(a) is reprinted by courtesy of *Chesapeake Bay* magazine and Michael Wooten, MCW Photography.

This research received partial support from a grant (R35CA 538701 to Peter H. Bartels) from the National Institutes of Health to the National Cancer Institute. The author expresses sincere appreciation to Hubert B. Bartels for the encoding of the RO filters and the production of the images for figure 1. The content of this paper is solely the responsibility of the author and does not necessarily represent the official views of the National Cancer Institute.

REFERENCES

1. Epstein, R. L., *The Semantic Foundations of Logic,* Oxford University Press, 1995, ch. IV.H.
2. IBID. , p 104.
3. IBID., pp. 72-25.
4. Fine, K., *Part-Whole,* in *The Cambridge Companion to Husserl,* Ed. B. Smith & D. W. Smith, Cambridge University Press, 1995, pp.463-485.
5. Gupta, A., *The logic of common nouns,* Yale University Press, 1980.
6. Heidigger, M., 1975, *Basic Problems in Phenomenology,* Trans. Albert Hofstader, 1988, Indiana University Press, p. 201.
7. Husserl, E. *Logishe Untersuchunge,* 1st ed. Halle: Neimeyer, 1900-01; 2nd ed. 1912-13. English translation, *Logical Investigations*, J.N. Findlay, London, Routledge, 1970.
8. Shoenfeld, J. R., *Mathematical Logic,* Addison-Wesley, 1967, ch. 6 & 7.
9. IBID., pp. 282-292.
10. van Fraasen, B. C., *Formal Semantics and Logic*, Macmillan Company, 1971, pp.28-29.
11. IBID., Ch. II.3, pp.31-34.
12. IBID., p.36.
13. IBID., p.37.
14. IBID. pp. 40-51
15. IBID. p.54.
16. IBID., pp. 56-59.
17. Wilburn, J. B., *A possible worlds model of object recognition*, Synthese (Forthcoming; accepted 25 Aug 98).
18. Wilburn, J. B., 1998, *Developments in generalized ranked-order filters*, JOSA A, **15**(5), pp. 1084-1099.
19. Willard, D., *Wholes, Parts and the Objectivity of Knowledge* in *Parts and Moments: Studies in Logic and Formal Ontology*, Ed., B. Smith, Philosophia Verlag GmbH, München, 1982, pp. 379-400.
20. Zaidi, Q., 1998, *Identification of illuminant and object colors: heuristic based algorithms,* JOSA A, **15**(7), pp. 1767-1776.

LINEAR DATA FUSION[1]

Francisco J. Samaniego, Duane L. Steffey[2] and Hien Tran
University of California, Davis

ABSTRACT

Selected literature on existing procedures for combining information from disparate sources is reviewed. Special emphasis is given to empirical Bayes (EB) and hierarchical Bayes (HB) methods. The particular challenges of the problem of combining information from developmental and operational tests in the context of the Department of Defense's acquisitions program are then discussed in detail. It is noted that the traditional EB and HB approaches tend to be inapplicable in the latter problem due to the conspicuous absence of exchangeability of the separate experiments involved. A more flexible framework, which relaxes the usual exchangeability assumption, is proposed. The feasibility and efficacy of linear Bayes estimation is demonstrated in this new framework, and is shown to yield promising results in a specific formulation of the DT/OT combination problem. Several possibilities for extending these preliminary modeling and inference ideas into a general theory for treating data from "related", though nonexchangeable, experiments are discussed.

1. INTRODUCTION

A great many statistical investigations have as their starting point a set of data drawn under fixed experimental conditions. Given these data, the statistical analysis is aimed at making inferences about the general parameters of the process or population from which the data were generated. The possibility of exploiting various forms of auxiliary information, be it empirical (e.g., data from a related experiment) or subjective (e.g., input from an expert), in order to improve these inferences is one that has intrigued statisticians for decades. Indeed, the fields of Bayesian statistics, empirical Bayes methods and meta-analysis each focus on particular prescriptions for appropriately combining information from disparate sources, and can each be thought of as a way of exploiting information auxiliary to the experiment of current interest. The report of the National Academy of Sciences Panel on Statistical Issues and Opportunities for Research in the Combination of the Information (see Gaver et al. (1992)) presents an excellent overview of statistical approaches to combining information, and contains a plethora of references to related work. Recent monographs in this general area include Hedges and Olkin's (1985) tome on meta-analytic techniques and Maritz and Lwin's (1989) treatise on empirical Bayes methods. Among

[1]Approved for public release; distribution is unlimited
[2]On sabbatical leave from San Diego State University

the notable contributors to this literature are Fisher (1932), Cochran (1937), Savage (1954), Robbins (1955), Glass (1978) and Deely and Lindley (1981).

When one narrows the scope of the inferential problems of interest to focus on the estimation of a parameter in the "current" experiment, and when one is particularly concerned with making proper use of data from one or more related experiments, one finds that, among extant approaches to combining information, the empirical Bayes (EB) approach (in either its frequentist or Bayesian forms) is the only one which formally fits the problem. We briefly review the EB approach, both for the sake of clarity in our subsequent developments and in order to motivate the need for a more general theory.

Let us assume that we are presented with data from a sequence of $k+1$ "similar" experiments, and that all but one are viewed as past experiments that may or may not be useful for the task at hand. Our goal is to estimate the parameter θ_{k+1} in the current experiment. Under the assumptions of the EB approach, which strictly defines the meaning of the word "similar", Robbins (1955) demonstrated that one could borrow strength from the past k experiments in formulating good estimators (indeed, asymptotically optimal estimators) of θ_{k+1}. More specifically, if (X_i, θ_i) represents the datum available in the ith experiment and the parameter of the distribution of X_i in that experiment (where one or both may be vector valued), then the EB approach assumes that the pairs $(X_1, \theta_1), ..., (X_k, \theta_k), (X_{k+1}, \theta_{k+1})$ are independent, with

$$\theta_1, ..., \theta_{k+1} \overset{iid}{\sim} G \tag{1.1}$$

and

$$X_i|\theta_i \sim F_{\theta_i},\ i = 1, ..., k+1. \tag{1.2}$$

Robbins took G in (1.1) to be completely unknown. In typical applications, F_{θ_i} is taken to belong to a particular parametric family. In such settings, Robbins demonstrated that is was often feasible to construct an estimator $\hat{\theta}_{k+1} = \hat{\theta}_{k+1}(X_1, ..., X_k, X_{k+1})$ which, at least asymptotically (as $k \to \infty$), could match the performance of the optimal (Bayes) estimator $\hat{\theta}_G$ of θ_{k+1} with respect to the unknown prior G and a squared error loss criterion. Indeed, if $r(G, \hat{\theta})$ represents the Bayes risk of the estimator $\hat{\theta}$ relative to G, that is, if

$$r(G, \hat{\theta}) = E_{\underset{\sim}{\theta}} E_{\mathbf{X}|\underset{\sim}{\theta}} \left(\hat{\theta} - \theta_{k+1}\right)^2, \tag{1.3}$$

then Robbins (1955) and Johns (1957) showed that EB estimators $\hat{\theta}_{k+1}$ could be constructed for which $r(G, \hat{\theta}_{k+1}) \to r(G, \hat{\theta}_G)$ as $k \to \infty$.

Important modifications of the EB approach were introduced by Deely and Lindley (1981) and by Morris (1983). The first of these works introduced hierarchical modeling into the EB framework, suggesting that a Bayes empirical Bayes approach was possible. Specifically, Deely and Lindley modeled the unknown prior G as a random distribution drawn from a family $\mathcal{G} = \{G(\cdot|\eta)\}$, where the hyperparameter η had distribution H. The second of these modifications, now referred to as parametric EB analysis, concentrated on fully parametric versions of the family $\mathcal{G}$ above, and advocated the estimation of the hyperparameter η above from the available data.

The striking power and utility of the empirical Bayes approach to estimation is especially apparent in the problem of estimating the mean $(\theta_1, ..., \theta_{k+1})$ of a $(k+1)$ dimensional normal distribution (with $\Sigma = I$) relative to generalized squared error loss. While the focus of that problem (i.e., estimating all the θs simultaneously) differs from the EB problem we have described, it is nonetheless an excellent example of the notion that borrowing strength from related experiments is efficacious. As is well known, the James-Stein (1956, 1961) estimator, which was shown to be an EB estimator by Efron and Morris (1973), shrinks the sample mean vector toward an arbitrary point and improves on the sample mean vector uniformly as an estimator of θ. When expanded, via hierarchical modeling, to include a capacity for incorporating both subjective and empirical information from sources separate from the current experiment, EB methods clearly represent a flexible and powerful tool for combining information in estimation problems.

The above notwithstanding, EB methodology is no panacea. There is an important class of multi-experiment problems in which EB methods are basically inapplicable. The crux of this limitation is the tacit assumption of exchangeability of the experimental data. The "similarity" of experiments encapsulated in the i.i.d. assumption for $\theta_1, ..., \theta_{k+1}$ is clearly a mixed blessing. When it is deemed a reasonable assumption, its imposition facilitates analytical work and justifies the blending of clearly relevant information. When, on the other hand, the assumption is untenable, application of EB methods will be, at best, misleading, and can in fact result in conclusions and decisions that are blatantly incorrect. The problems on which we will focus in the sequel lie squarely in the latter camp. These problems are characterized by knowledge which precludes an exchangeability assumption. Instead, it will

be acknowledged that current and past experiments are different, (i.e., not "similar") but are nonetheless related or linked in some specific fashion. In addressing such problems, we will need to replace an existing theory for treating information from similar experiments with a new theory for treating information from related experiments. Although there are a wide variety of situations in which such a new theory might be applied, our own motivation and interest has been sparked by a data-combination problem which arises in the context of military acquisitions processes. A brief description of that context follows.

2. DEVELOPMENTAL AND OPERATIONAL TESTING

The processes of developmental testing (DT) and operational testing (OT) within the context of the Department of Defense (DoD) acquisitions program is well described in Cohen, Rolph and Steffey (1998). In the course of the development of a new system (e.g., a communications device, a tank or an airplane), there are two distinct phases during which prototypes are tested. Roughly speaking, developmental testing refers to experimentation done while the prototype is being built and "perfected", while operational testing refers to independent experiments on the completed prototypes which are aimed at determining the new system's effectiveness and suitability. Under protocols now in force, it is standard practice to keep DT and OT data separate, and to make decisions regarding effectiveness and suitability based on OT data alone. Among the questions we wish to address in this context is whether, and how, estimates of a system's performance parameters can be improved using methods of data-combination. In seeking to answer such questions, we begin by acknowledging that DT and OT experiments are rarely, if ever, exchangeable. Indeed, it is generally the case that operational tests are run in a wider variety of environments and often involve higher stress and demands than those occurring in the DT setting. It is thus not uncommon to see somewhat poorer performance under OT than under DT. Given that, one would clearly not wish to apply standard EB methods to the estimation of OT parameters.

To make this discussion more concrete, consider the data shown below, simulated under two separate exponential distributions.

Table 1: Simulated Life Testing Data

DT DATA		OT DATA
28.7335	18.005	13.4764
21.7593	1.5495	18.6327
6.0077	35.5350	4.5435
46.6829	22.0601	23.5081
7.5756	2.5790	5.3412
11.2651	20.8876	8.3927
16.0805	7.1455	39.9724
8.0645	10.1876	7.7885
9.9661	67.0262	33.1363
41.6639	7.7921	6.1353

SUMMARY	SUMMARY
MODEL: EXP (θ_1)	MODEL: EXP (θ_2)
$n_1 = 20$	$n_2 = 10$
$\bar{X}_{1.} = 19.63$	$\bar{X}_{2.} = 16.09$

Although the data above are simulated, they are realistic in two particular ways. First, they capture the fact that system lifetime tends to be higher under DT than under OT. Secondly, the OT data is somewhat more sparse, a common occurrence given the expense involved in operational testing in realistic use environments and the fact that OT is often, in practice, an underfunded activity. Because of this sparseness, it can be especially important to determine ways of combining information from DT and OT which promise to yield improved estimates of OT parameters. The need and motivation to do so is strongly underscored by the following excerpt from the report of the National Academy of Sciences Panel on Statistical Methods for Testing and Evaluating Defense Systems (see Cohen, et al. (1998, p. 119)).

> "In the later stages of developmental testing, when the process of developing a reliable system prototype has become relatively stable, there will be experimental data that can be both relevant and quite useful to the operational tester. The use of that data in combination with the data

collected in designed operational tests may offer some important benefits, including the possibility of an early resolution regarding a system's suitability. There should be greater openness to the selective (and supported) use of statistical methods for combining data from developmental and operational testing, as well as other relevant information, including subjective inputs from scientists with appropriate expertise and commercial or industrial data on related components or systems."

"Recommendation 7.7: Methods of combining reliability, availability, and maintainability data from disparate sources should be carefully studied and selectively adopted in the testing processes associated with the Department of Defense acquisition programs. In particular, authorization should be given to operational testers to combine reliability, availability, and maintainability data from developmental and operational testing as appropriate, with the proviso that analyses in which this is done be carefully justified and defended in detail."

While the commentary above might be read as an unequivocal call to arms for researchers to address this data-combination problem, the matter of how to do this is far from clear. In the next two sections, we describe our preliminary work in this direction.

3. A GENERAL FRAMEWORK FOR TREATING RELATED EXPERIMENTS

Let us adopt a set of assumptions that are suitably broader than those characterizing EB work. As we shall see, we will be able to accommodate a theory for treating related experiments within the framework we describe below. Suppose that

$$\{(X_i, \theta_i)\,,\ i = 1, ..., k+1\} \text{ are independent;} \tag{3.1}$$

$$\theta_i \sim G_i, \quad i = 1, ..., k+1; \tag{3.2}$$

$$E\theta_i = \mu_i, \text{ and } V(\theta_i) = \sigma_i^2 < \infty; \tag{3.3}$$

$$X_i|\theta_i \sim F_{\theta_i},\ \ i = 1, ..., k+1; \tag{3.4}$$

$$E\left(X_i|\theta_i\right) = \theta_i, \text{ and } V\left(X_i|\theta_i\right) < \infty; \tag{3.5}$$

and finally,

$$V_i = E(V\left(X_i|\theta_i\right)) < \infty. \tag{3.6}$$

The key difference between the EB framework and the assumptions above is embodied in (3.2); this assumption allows the $k+1$ experiments to be dissimilar. Without further structure, there is rather little that one can say about combining the information in these $k+1$ experiments in seeking to estimate θ_{k+1}. In the next section, we'll add sufficient structure to obtain some preliminary (but promising!) results. Yet even in the generality above, some progress can be made.

We are about to restrict attention to linear estimators of θ_{k+1}, that is, to estimators of the form

$$\hat{\theta}_{k+1} = \sum_{i=1}^{k+1} c_i X_i.$$

The notable success of "best linear unbiased" estimators in linear model theory and elsewhere (see, e.g., Rao (1973)), and the worthiness of linear Bayes estimators in many problems (see Hartigan (1969) and Ericson (1970)) give the class of linear estimators substantial credibility. In two recent papers on empirical Bayes estimation, Samaniego and Neath (1996) and Samaniego and Vestrup (1999) show that linear EB estimators can provide improved performance over standard estimators based on the current experiment alone, even with just a single past experiment available. Taken together, this work suggests that linear estimators might prove to be useful vehicles for combining data from related experiments. Indeed, even without specifying any particular relationship between the experiments governed by (3.1)-(3.6), that is, even without stating explicitly how the distributions $\{G_i\}$ might be related, one can obtain results such as the following:

Theorem: Assume that conditions (3.1)-(3.6) hold. Then for arbitrary constants $c_1, ..., c_{k+1}$, the Bayes risk of $\hat{\theta}_{k+1} = \sum_{1}^{k+1} c_i X_i$ as an estimator of θ_{k+1} under squared

error loss is given by

$$r(\underset{\sim}{G}, \hat{\theta}_{k+1}) = E_{\underset{\sim}{\theta}} E_{\mathbf{X}|\underset{\sim}{\theta}} \left(\sum_{i=1}^{k+1} c_i X_i - \theta_{k+1} \right)^2$$

$$= \sum_{i=1}^{k+1} c_i^2 \left(V_i + \sigma_i^2 \right) + \left(\sum_{i=1}^{k+1} c_i \mu_i \right)^2$$

$$-2 \sum_{i=1}^{k+1} c_i \mu_i \mu_{k+1} + \mu_{k+1}^2 + \left(1 - 2c_{k+1} \right) \sigma_{k+1}^2. \tag{3.7}$$

Since the application of special interest involves but two experiments (DT and OT), we record the important special case:

Corollary: Let $k = 1$, and assume that conditions (3.1)-(3.6) hold. Then the Bayes risk of $\hat{\theta}_2 = c_1 X_1 + c_2 X_2$ as an estimator of θ_2 under squared error loss is given by

$$r(\underset{\sim}{G}, \hat{\theta}_2) = c_1^2 \left(V_1 + \sigma_1^2 \right) + c_2^2 \left(V_2 + \sigma_2^2 \right) + \left(c_1 \mu_1 + c_2 \mu_2 \right)^2$$

$$-2c_1 \mu_1 \mu_2 + \left(1 - 2c_2 \right) \left(\mu_2^2 + \sigma_2^2 \right). \tag{3.8}$$

Note that $r(\underset{\sim}{G}, \hat{\theta}_2)$ above represents the squared error of $\hat{\theta}_2$ averaged over the distributions of X_1, X_2, θ_1 and θ_2. From (3.8), it is easy to verify that the linear function of X_1 and X_2 that minimizes $r(\underset{\sim}{G}, \hat{\theta}_2)$ has coefficients

$$c_1^* = \frac{V_2 \mu_1 \mu_2}{\left(V_1 + \sigma_1^2 + \mu_1^2 \right) \left(V_2 + \sigma_2^2 + \mu_2^2 \right) - \mu_1^2 \mu_2^2} \tag{3.9}$$

and

$$c_2^* = 1 - \frac{V_2 \left(V_1 + \sigma_1^2 + \mu_1^2 \right)}{\left(V_1 + \sigma_1^2 + \mu_1^2 \right) \left(V_2 + \sigma_2^2 + \mu_2^2 \right) - \mu_1^2 \mu_2^2}. \tag{3.10}$$

The identification of the "linear Bayes rule" via (3.9) and (3.10) does not necessarily solve the problem of interest. Why this is so may not be immediately apparent. The difficulty with $\hat{\theta}_2^* = c_1^* X_1 + c_2^* X_2$ as an estimator of θ_2 in the two-experiment problem is that, for many versions of the modeling framework in (3.1)-(3.6), the optimal coefficients will depend on unknown parameters of the distributions of θ_1 and θ_2. Thus, $\hat{\theta}_2$ will not, in fact, be a bona-fide estimator in such situations. The potential nonetheless exists for estimating θ_2 via the mechanics above. The results in the next section provide an example of how this might be done.

4. MODELING AND INFERENCE FOR RELATED LIFE TESTING EXPERIMENTS

Consider, now, the data in Table 1. Suppose we have accepted the assumption that these data represent independent samples from exponential distributions with means θ_1 and θ_2. Letting T_i be the total time on test from the ith experiment, and denoting the Gamma distribution with shape parameter α and scale parameter β as $\Gamma(\alpha, \beta)$, we have that

$$T_i \sim \Gamma(n_i, \theta_i), \quad i = 1, 2. \tag{4.1}$$

In building a hierarchical Bayes model for the DT/OT data, we now add the following assumption about the parameters θ_1, θ_2 (i.e., about the distributions G_1 and G_2 in (3.2)): assume that

$$\theta_i \sim \Gamma(K_i, \frac{\mu_i}{K_i}) \quad i = 1, 2, \tag{4.2}$$

where μ_i is the (unknown) mean of θ_i and K_i is the known shape parameter. We are, in essence, modeling the uncertainty about θ_i by a one-parameter Gamma distribution with unknown mean μ_i. The parameter K_i thus governs the dispersion in the model, with large K_i corresponding to quite precise priors for θ_i and small K_i corresponding to rather diffuse priors for θ_i. Finally, we model the linkage between the two experiments by the assumption

$$\mu_2 = \lambda \mu_1 \tag{4.3}$$

for some fixed constant λ.

We begin by taking λ to be known, but we will show in the sequel that this assumption is by no means essential. Because of (4.3), we may rewrite (4.2) as

$$\theta_1 \sim \Gamma(K_1, \frac{\mu}{K_1}) \text{ and } \theta_2 \sim \Gamma(K_2, \frac{\lambda\mu}{K_2}). \tag{4.4}$$

For the model above, it is easy to check that

$$\sigma_1^2 = \frac{\mu^2}{K_1} \text{ and } \sigma_2^2 = \frac{\lambda^2\mu^2}{K_2} \tag{4.5}$$

and

$$V_1 = \left(\frac{K_1+1}{K_1}\right)\frac{\mu^2}{n_1} \text{ and } V_2 = \left(\frac{K_2+1}{K_2}\right)\frac{\lambda^2\mu^2}{n_2}. \tag{4.6}$$

Letting

$$r_i = \frac{K_i+1}{K_i} \text{ for } i = 1, 2, \tag{4.7}$$

the coefficients of the linear Bayes rule relative to the models in (4.4) are given by

$$c_1^* = \frac{n_1 r_2 \lambda}{r_1 r_2 (n_1+1)(n_2+1) - n_1 n_2} \tag{4.8}$$

and

$$c_2^* = 1 - \frac{r_1 r_2 (n_1+1)}{r_1 r_2 (n_1+1)(n_2+1) - n_1 n_2}. \tag{4.9}$$

If standard practices were applied to the analysis of the information in Table 1, the parameter θ_2 would be estimated by the mean $\bar{X}_{2\cdot}$ of the OT data, that is, by

$$\hat{\theta}_2 = 16.09. \tag{4.10}$$

As an example of a contrasting analysis, suppose λ were known to be .75, and suppose that our prior modeling specified the constants K_i as

$$K_1 = 50 \quad \text{and} \quad K_2 = 100.$$

These latter assumptions correspond to standard errors for θ in the $1.5 - 3.0$ range. With these choices, the coefficients of the optimal estimator are given by

$$c_1^* = .3989 \quad \text{and} \quad c_2^* = .4303,$$

and the linear Bayes estimator of θ_2 is

$$\hat{\theta}_2^* = 14.7593. \tag{4.11}$$

We now reveal the parameter values under which the data of Table 1 were generated: the true values of the $\theta' s$ are

$$\theta_1 = 20 \quad \text{and} \quad \theta_2 = 15. \tag{4.12}$$

The example above suggests that a rather striking reduction in the error of estimation is possible when data from DT and OT are linearly combined. This example is no accident; we have been able to establish analytically that, under conditions (3.1)-(3.6),

$$r(\underset{\sim}{G}, \hat{\theta}_2^*) = c_2^* r(\underset{\sim}{G}, \bar{X}_{2.}), \tag{4.13}$$

showing that, in the hierarchical scenario specified above, one will realize, on the average, about a 60% improvement in the precision of the estimator, as measured by the Bayes risk criterion.

While having some knowledge concerning λ in a DT/OT setting may be a reasonable expectation, the assumption that λ is known might well be considered unduly heroic. Fortunately, it is really quite unnecessary. Suppose, instead, that one takes λ and μ to be random with prior distributions L and M, respectively, where λ has known finite first and second moments, given by $E\lambda = L_1$ and $E\lambda^2 = L_2$, and μ has finite second moment M_2. It is easy to show that the Bayes risk of a linear estimator $\hat{\theta} = c_1\bar{X}_{1.} + c_2\bar{X}_{2.}$ depends on L and M only through L_1, L_2 and M_2, and that the optimal coefficients c_i^*, which are independent of M_2, are given as follows:

$$c_1^* = \frac{n_1 r_2 L_1 L_2}{r_1 r_2 (n_1+1)(n_2+1)L_2 - n_1 n_2 L_1} \tag{4.14}$$

and

$$c_2^* = 1 - \left[\frac{r_1(n_1+1)}{n_1 L_1}\right] c_1^*. \tag{4.15}$$

It is thus possible to devise reasonable linear estimators, and achieve substantial improvement over the estimator $\hat{\theta}_2 = \bar{X}_{2.}$, without making unduly stringent assumptions regarding the rescaling parameter λ. Our simulations to date show that the optimal estimator $\hat{\theta}_2^*$ in this latter setting tends to offer improvement over $\bar{X}_{2.}$ for a rather broad range of priors.

5. DISCUSSION

It is not unreasonable to surmise, at this juncture, that the developments in sections 3 and 4 raise a good many questions. The results obtained thus far have been offered as evidence that a new theory for treating related experiments is feasible. In work currently in progress, we are seeking to expand our modeling and inference

in two specific ways, moving beyond the tight parametric assumptions utilized in Section 4 and relaxing the linearity constraint on our estimators. We briefly discuss these two research directions below.

While the exponential distribution is widely used in reliability and life testing applications, the need to treat other, less restrictive, models in such applications is also well understood (see, e.g., Samaniego and Chong (1996)). Among other parametric models of interest, the normal model no doubt arises in general inference problems with the greatest frequency. It is our intention to replicate the analysis of section 4 in a suitable normal context, and indeed, to explore the possibility of extensions to both discrete and continuous exponential families in general. The normal and binomial cases are, of course, the most important due to the central and recurring role of these models in common inferential settings. But it is also our intention to explore various formulations of "related life testing experiments", generalizing our treatment of exponential life testing to the gamma, Weibull and lognormal models. Versions of exponential life testing that are less restrictive than the framework in (4.1)-(4.3) are also being explored.

The appeal of linear-Bayes rules such as $\hat{\theta}_2$ in (3.11) is that they represent closed-form estimators of the parameter of interest, they clearly utilize data from related experiments in estimating the current parameter, and they can provide a substantial reduction in the mean squared estimation error over estimators which only utilize data from the current experiment. At the same time, one must recognize that best linear estimators are compromises; that is, they represent the best estimators within a restricted but convenient class. Among the questions that remain are: can we obtain, analytically, the (unrestricted) Bayes estimator of the current parameter? If not, is it feasible to obtain it numerically? How much further reduction in Bayes risk is possible? As we have seen, the derivation of linear Bayes estimators in the life testing example of section 4 does not require a full specification of the hierarchical model, since the probability distributions of λ and μ need to be specified only up to their first two moments. A fully Bayesian treatment requires a complete specification of these models. Ultimately, we expect that Bayes estimators will be obtained under a variety of prior models for the parameters (e.g., for θ_1, θ_2, μ and λ in the exponential problem). In the exponential example, we have already obtained results when λ and μ are modeled as inverse gamma variables (and obtained the interesting distributional result that θ_1 and θ_2 are, marginally, independent scaled F variables).

Our main goal in moving beyond linearity is to carry out a careful comparative study of possible estimators of the parameter(s) of the current experiment, with a view toward giving concrete advice regarding the circumstances under which combining experiments from related experiments yields improvement over estimators based on current data alone. We anticipate being able to obtain results of the sort found in Samaniego and Reneau (1994), where specific conditions on particular families of prior distributions are identified as necessary and sufficient for Bayes or linear Bayes estimators to outperform classical ones in an exponential family setting. There are some imposing computational issues that arise in doing a hierarchical Bayes analysis in the problems of interest. Closed form representations of posterior quantities will be unachievable in most cases. We are confident, however, that approximate methods such as those used by Kass and Steffey (1989) will provide reliable results.

REFERENCES

Cochran, W.G. (1937), "Problems arising in the analysis of a series of similar experiments", Journal of the Royal Statistical Society (Suppl.) 4, 102-118.

Cohen, M.L., Rolph, J.E. and Steffey, D.L. (Editors), (1998), Statistics, Testing and Defense Acquisition: New Approaches and Methodological Improvements. Washington D.C.: National Academy Press.

Deely, J. and Lindley, D. (1981), "Bayes empirical Bayes", Journal of the American Statistical Association, 76, 833-841.

Efron, B. and Morris, C. (1973), "Stein's estimation rule and its competitors - an empirical Bayes approach", Journal of the American Statistical Association, 68, 117-130.

Ericson, W.A. (1970),"On the Posterior mean and variance of a population mean", Journal of the American Statistical Association, 65, 649-652.

Ferguson, T.S. (1973), "A Bayesian analysis of some nonparametric problems", Annals of Statistics, 1, 209-230.

Fisher, R.A. (1932), Statistical Methods for Research Workers (4th Ed.), London: Oliver and Boyd.

Gaver, D.P., Draper, D., Goel, P.K., Greenhouse, J.B., Hedges, L.V., Morris, C.N., and Waternaux, C., "Combining Information: Statistical Issues and Opportunities for Research", Washington, D.C.: National Academy Press.

Glass, G.V. (1978), "Integrating findings: the meta-analysis of research" in Review of Research in Education, Shulman, L.S. (Ed.), 5, 351-379, Itasca, Ill.: F.E. Peacock.

Hartigan, J.A. (1969), "Linear Bayesian methods", Journal of the Royal Statistical Society, B, 31, 446-454.

Hedges, L.V. and Olkin, I. (1985), Statistical Methods for Meta-Analysis, New York: Academic Press, Inc.

James, W. and Stein, C. (1961), "Estimation with quadratic loss", in Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, 361-379, Berkeley: University of California Press.

Johns, M.V. Jr. (1957), "Non-parametric empirical Bayes procedures", Annals of Mathematical Statistics, 28, 649-669.

Kass, R.E. and Steffey, D.L. (1989), "Approximate Bayesian inference in conditionally independent hierarchical models", Journal of the American Statistical Association, 84, 717-726.

Maritz, J.S. and Lwin, T. (1989), Empirical Bayes Methods (2nd Ed.), London: Chapman and Hall.

Morris, C.N. (1983), "Parametric empirical Bayes inference: theory and applications", Journal of the American Statistical Association, 78, 47-55.

Rao, C.R. (1973), Linear Statistical Inference and Its Applications, New York: John Wiley and Sons.

Robbins, H. (1955), "An empirical Bayes approach to statistics", in Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, Berkeley: University of California Press.

Samaniego, F.J. and Reneau, D.M. (1994), "Towards a reconciliation of the Bayesian and frequentist approaches to point estimation", Journal of the American Statistical Association, 89, 947-959.

Samaniego, F.J. and Neath, A.A. (1996), "How to be a better Bayesian", Journal of the American Statistical Association, 91, 733-742.

Samaniego, F.J. and Chong, Y.S. (1996), "On the performance of Weibull life tests based on exponential life testing designs", Proceedings of the First Annual U.S. Army Conference on Applied Statistics, ARL-SR-43, Aberdeen Proving Ground, Maryland: Army Research Laboratory.

Samaniego, F.J. and Vestrup, E.M. (1999), "On improving upon standard estimates via linear empirical Bayes methods", Stat. and Prob. Letters, to appear.

Savage, L.J. (1954), The Foundations of Statistics, New York: John Wiley and Sons.

Stein, C. (1956), "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution", in Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, 197-206, Berkeley: University of California Press.

Human Factors Issues Evaluated During the 4th Infantry Division XXI Advanced Warfighting Experiment (DAWE)

Dr. Jock O. Grynovicki, Kragg P. Kysor, Michael Golden, Madeline Swann,
Shamim Rashid, Tony Ward
U.S. Army Research Laboratory
Human Research and Engineering Directorate
Aberdeen Proving Ground, Maryland 21005-5425

ABSTRACT

The focus of the Human Factor Issue (ALO1) within the DIVISION XXI Advanced War-fighting Experiment (DIV XXI AWE) was on the Maneuver Control System's (MCS) interface role in enhancing commander and staff performance. Usability characteristics and the ability of the MCS to provide the required functionality to the Battle Staff (i.e., planning, information management, decision making, situational awareness, & control of the battle-space) were investigated. The findings indicated that the MCS computer's usability and functionality assisted the soldiers and command staff in performing command and control; however, further improvements in both these areas are recommended.

INTRODUCTION

Human Factors (HF) Issue Focus for the DAWE. The focus of the HF Issue within the DIVISION XXI Army War-fighting Experiment (DAWE) was to develop an analytical understanding of how the commander and the battle staff use and interface with the Army Tactical Command and Control Systems (ATCCS). Specifically for ARL's DAWE efforts, the HF Issue analysis was centered on the Maneuver Control System (MCS) human computer interface (HCI) "usability" characteristics as well as the ability of the MCS to provide the required functionality to the Battle Staff for planning, information management, decision making, and control of the battle-space.

Human Factors (HF), one of the seven domains of manpower personnel integration (MANPRINT), is concerned with the role of humans in complex systems, the design of equipment and facilities for human use, and the development of environments for safe operations. The Division Advanced War-fighting Experiment (DAWE) HF initiative was intended to help the Army leadership assess the impact of digitization on individual soldier and staff performance by studying iterative ATCCS soldier-system interface designs and functionality as they support command decision making processes and related Battlefield Operating System (BOS), Command and Control (C2), and information management operations. The lack of emphasis on the human component in the design and integration of

automation can result in significant performance degradation, increased training requirements, and a lack of system acceptance by the soldier. The analysis of MCS interface issues will be incorporated into the overall analysis of FORCE XXI operational concepts.

Objective MCS System Description. The MCS is one of five C2 systems that constitute the Army Tactical Command and Control System (ATCCS). The objective MCS will support commanders of all maneuver elements at corps through battalion squadron level and selected companies with a single command and control system. Computer displays will allow commanders and staffs to access databases for situation reports (SITREPs), enemy assessments, friendly force status, and maneuver control. The MCS is being developed and fielded in an evolutionary fashion. The MCS will utilize Mobile Subscriber Equipment (MSE), the Army Data Distribution System (ADDS), and Combat Net Radio (CNR) communication systems. The MCS software will operate on the Common Hardware and Software-2 (CHS-2). The MCS V12.3 has the full Operational Requirements Document capability, operating on CHS-2, Large Scale Printer Plotter (LSPP), Large Screen Display (LSD), Tactical Scanner (TACSCAN) and housed in Standardized Integrated Command Post Systems (SICPS). The MCS will extend the automated command and control capability below the brigade level within the Maneuver Function Area (MFA) to include selected Engineer, Military Police, Signal, Aviation, Chemical, Armor, and Infantry units.

Objective System Interoperability. The MCS shall be inter-operable, using direct computer-to-computer data exchanges or standardized message text formats with Army, Joint and Coalition command and control systems, as appropriate to the current fielding status of these systems, in accordance with valid user interface requirements. The minimum acceptable direct data exchange integrity parameter is 95%; that is, automatic database updates are to be free of any errors so that resultant database entries are exactly as the data transmitted 95% of the time. Each datum is defined to be the information context of a single field in a single record. Graphics that are stored as bit map image files shall count as a single datum.

MCS software will provide for the integration of the five Battlefield Functional Area Control Systems (BFACS) into the objective Force Level Control (FLC). As defined in the ATCCS Required Operational Capability (ROC), FLC is the process by which the commander and staff of a combined arms team integrate and synchronize the efforts of the five Battlefield Functional Areas (BFAs) to support attainment of the unit mission. Integration and synchronization are effected primarily through the management, manipulation and assessment of information from across the five BFAs and the development of tactical plans and orders. Figure 1 depicts the objective C4I architecture which includes all projected ATCCS, Army Battle Command Systems (ABCS), and Continental U.S. (CONUS) systems.

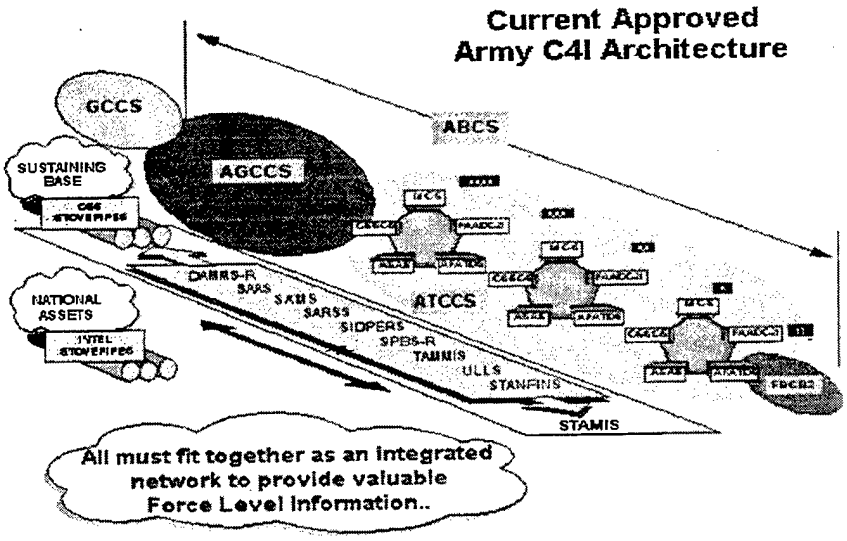


Figure 1. Current Approved U.S. Army C4I Architecture including ATCCS, ABCS, and CONUS systems.

Concept of Employment DIV XXI. The MCS V12.3 was the central digitized command and control platform used in DIVXXI. The objective MCS supported commanders of all maneuver elements of the 4th ID with a single command and control system. The MCS allowed access to databases for situation reports (SITREPs), enemy assessments, friendly force status, and maneuver control.

The MCS computers consisted of baseline battle command system software running on commercial Sun computer platforms. This system provided a common picture of the battlefield overlaid on Defense Mapping Agency (DMA) digital maps as shown in Figure 2. The system provided a capability to synchronize the battle plan based on near-real-time information and assessments from staff and subordinate commanders. The MCS had the capability to convey current information about military unit locations, strength, and other pertinent information about friendly and enemy forces such as the ability to:

- Define unit task organization
- Receive enemy and friendly position feeds
- Build and manipulate databases
- Generate and display reports
- Present situation awareness
- Build graphical overlays to maps
- Create operations plans/operations orders (OPLANs/OPORDs)
- Send and receive information and briefs.

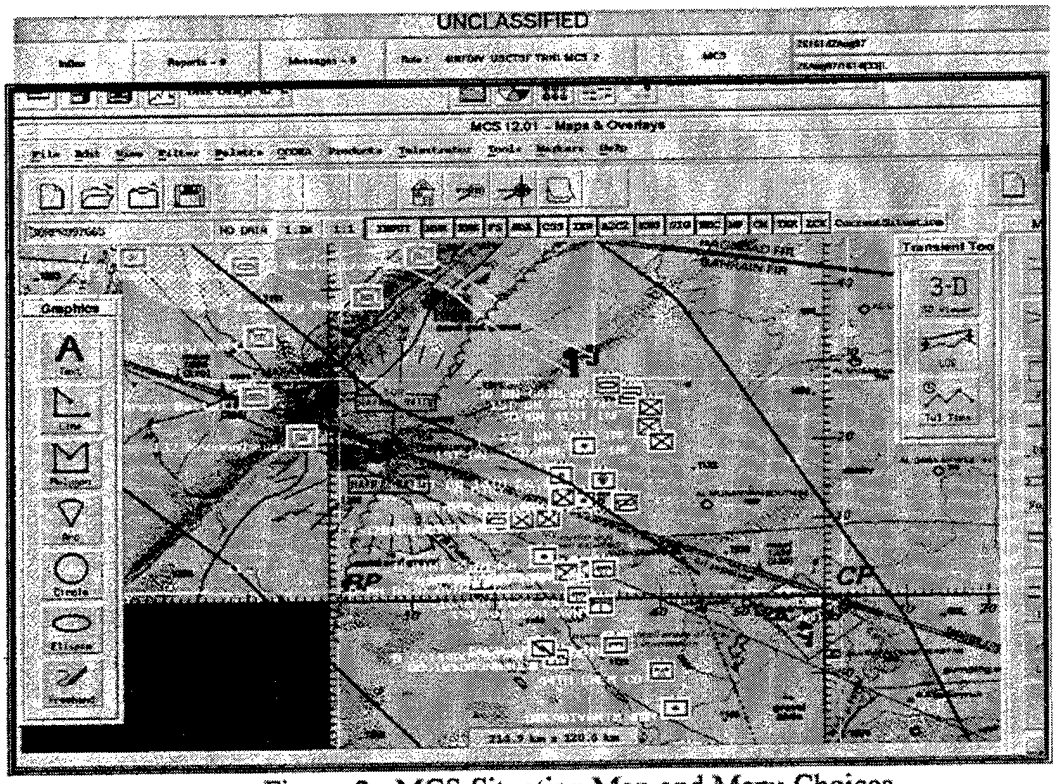


Figure 2. MCS Situation Map and Menu Choices

METHODOLOGY and DATA SOURCES

As the lead agency responsible for the DAWE Human Factors Issue, the ARL provided the following resources and assessment materials in support of DIV XXI: (1) an Issue Proponent Manager, five HFE Subject Matter Expert (SME) observers and one analyst to serve throughout the Simulation Exercises (SIMEXs) and the AWE. (2) Of the total 105 Military SMEs participating as observers during the DAWE, 26 were assigned by TEXCOM to support ARL in collecting HF related observations. To assist in orienting these military SMEs to the HF Issue, ARL developed an HF Observer's Guide and provided the 26 SMEs with training just prior to the AWE start of experiment (STARTEX). The 5 ARL HF SMEs joined with 21 military SMEs in conducting HF-focused observations throughout the AWE which were recorded on laptop computers for daily downloading to the TEXCOM DAWE data base repository which were later deposited into the Center for Army Lessons Learned (CALL) Collection Plan and Observer Management System (CALLCOMS) data base. (3) The ARL executed analysis oversight of the HF Observations input to the CALLCOMS data base including the resolution of anomalous observations. (4) ARL developed two HF-focused questionnaire surveys. The first survey (i.e., "Usability Survey") was administered to all 4thID TOC MCS operators and addressed MCS software interface "usability." The second survey (i.e., "Functionality Survey") was administered to the 4thID Command Staff and selected BOS cell Officers in Charge (OICs), and addressed MCS functionality as it supported critical staff tasks associated with decision-making and maneuver control. These two

surveys were administered to the 4thID personnel by the Test and Experimentation Command (TEXCOM) following the DAWE end of experiment (ENDEX). The analysis of all HF Issue data sources listed above (SME Observations and 4thID Staff Surveys), subsequently formed the basis of the MCS human factors assessment presented in this report. Detailed descriptions of the Observer Guide and Surveys are provided as follows:

ARL's SME Observer Guide. Developed as an outline for the collection of the DAWE HF Sub-Issue data, the guide provided information for the SME on which to focus personnel and digitized equipment factors (i.e., operator performance or MCS functions) during the DAWE to help answer each associated HF Sub-Issue. The content of the Observer's guide is outlined in Table 1.

Table 1: Human Factors Issue Observer Guide for the Subject Matter Experts (SMEs) with Focus on the Maneuver Control System (MCS)

Sub-Issue	Description
Sub-Issue: HF AL 01 A.1 How effective is the MCS operator-system interface design?	Consider such factors as graphics, flexibility, intuitiveness, consistency of computer processes, system feedback to the user, recovery from errors, shortcuts, speed, and how the system reflects doctrine as well as consistency between the different Battlefield Operating Systems.
Sub-Issue: HF AL 01 G.1Evaluate the functionality of MCS as it supports staff decision making.	Consider the effect of the MCS in helping the staff perform mission analysis, course of action (COA) development and analysis, orders preparation, time analysis, and issuing orders.
Sub-Issue: HF AL 01 C.1 How well does the MCS help the staff obtain the Commander's Critical Information Requirements (CCIR) based on information from digitized text messages, graphics, and data bases?	Consider the ability of the staff to use the MCS to find and filter critical task relevant information from distributed text and graphical data base repositories (e.g., Battlefield Operating Systems, higher echelon sources).
Sub-Issue: HF ALO1 D.1 How well did the MCS contribute to a shared, real-time situation awareness of the battlespace.	Consider the "push/pulls" of data and graphics between the ATCCS Systems within the TOC and between higher and lower echelons. Also consider such factors as the ability to filter text messages and produce graphic visualizations to help the Staff develop a timely "critical event" awareness and accurate interpretation of these events.
Sub-Issue: HF AL 01 E.1 Compared to manual methods, how well did the MCS help to distribute the workload among staff members during continuous operations of low and high activity so that some staff members could rest while others were able to maintain critical staff functions?	Automation should facilitate the Battlestaff to distribute command and control workload burdens during continuous operations such that selected staff members can rest while others take over and maintain critical functions. In addition, in dealing with military complexity, stress and sleep deprivation, automation should provide a user-computer interaction framework that reduces the Commander and Staff's mental workload associated with developing timely and accurate "shared mental models" of situation awareness.
Sub-Issue: HF AL 01 F.1 How well did the MCS's graphics and drawing tools help the Staff in developing templates and overlays on maps?	Consider the following: (1) Were the staff members able to easily draw boundaries, radar fans, restricted air corridors, etc. using MCS graphic tools? (2) Were the Staff able to place and rotate icons, markers, arrows, and lines?
How well did the MCS/P provide the TOC Staff with continuous and reliable automated war fighting capabilities?	Consider the following factors: (1) Consider the number of times per hour the MCS/P system crashed. (2) The number and type of observed modules, menus, or functions that were not operational.
Sub-Issue: HF AL 01 H.1 How well did MCS support the development and maintenance of a coordinated, timely, and accurate Relevant Common Picture (RCP) ?	How well did the graphics developed and distributed by the Division Main Information Cell provide a visual slice of the information space as it was scaled to the specific mission requirements of commanders at various echelons?
Sub-Issue: HF AL 01 I.1 Describe the use of the resource status reports (i.e., the Chicklet, Gumball, Mercedes, and 18-Wheeler reports):	Resource status reports, tables, and charts provide the commander and staff with the latest status of resources in graphical and tabular form. The REPORTS application provides detailed information about specific friendly military units and their readiness.
Sub-Issue: HF AL 01 J.1 Describe the utility of the "Message Handler" software:	Consider such factors as: (1) ease of creating messages using autofill. (2) ease of previewing, editing, sending, accessing, and reviewing messages. (3) ability of the header to allow the staff to distribute information to higher and lower echelons.

Table 1 (continued): Human Factors Issue Observer Guide for the Subject Matter Experts (SMEs) with Focus on the Maneuver Control System (MCS)

Sub-Issue: HF AL 01 K.1 Describe the use of the "Unit Task Organization" (UTO) software:	Consider the following: (1) Does the UTO function allow the commander to reorganize units to best accomplish the mission? (2) Does the UTO graphical display allow for easy identification of the units name and status? (3) Could the operator change the command relationship of a unit?
Sub-Issue: HF AL 01 L.1 Describe the ability of the MCS to provide the information required to support the synchronization of combat operations.	Consider the information available for developing a synchronization matrix related to: (1) accurately locating enemy units on the terrain. (2) reducing the uncertainty associated with different courses of action.
Sub-Issue: HF AL 01 M.1 How well did the MCS/P facilitate the Staff members in preparing desired reports (WARNORDs, SITMAPs, OPORDs, FRAGOS, etc.) ?.	MCS should provide the Staff collaboration tools and standard pre-formatted messages and report displays with both menu selections and free-text data fields
Sub-Issue: HF AL 01 N.1 Did the MCS support the interoperable data exchange of text, graphical information and functions between the various Battlefield Operating Systems (BOSs)?	Data connectivity for DIV XXI should be provided by the Army Battle Command System which supports each battlefield functional area (Maneuver, Military Intelligence, Fire Support, etc.) in providing commanders and their staff members with timely and accurate information to support effective command and control. Information and ATCCS functions should be able to be automatically ported between the Intel (ASAS), Air Defense (FAADC2), and DIVARTY (AFTADS) cells, and the command staff (MCS) cells. The MCS should be able to use "Netscape" software to view data and reports provided by the AFATDS, CSS and the ASAS as well as provide MCS produced data and reports to the other BOSs.

ARL's MCS Usability Survey. The first survey developed by ARL focused on the usability of the soldier-MCS system interface. The "usability factor" has a direct impact on staff performance because shortcomings in system usability lead to underlying error patterns, attentional fatigue and excessive workload which can be linked to inappropriate decisions and priorities, serious delays in operational tempo, and failures in effective staff coordination and communications. This MCS Usability Survey's focus was guided by 13 usability issues that have been defined in the research literature as reflecting hardware and software design with good interface usability. Subsequently, these design characteristics were used by ARL-HRED to focus on rating staff tasks on interface usability and graphics issues. An example of the tasks and issues is shown in Table 2. The 13 usability characteristics include: whether the computer system contains simple and natural dialogue, applications reflect doctrine, "speaks" the user language, minimizes user memory load, remains consistent between different modules and across applications, provides user feedback, provides clearly marked exits, provides process shortcuts, and prevents errors. Additionally, questions regarding contrast, symbol color, screen layout, etc., were also included in the survey. It should be noted here that ARL developed a standardized ATCCS version of this MCS usability survey which was subsequently administered by TEXCOM to all ATCCS operators who participated in the DIV XXI AWE operations. The results of the standardized ATCCS survey (which canvassed FAADC3I, ASAS, AFATDS, and CSSCS operators), are being published separately, by the TRADOC Analysis Center (TRAC).

Table 2: Usability Characteristics Index versus Maneuver Staff Tasks

Usability Issues	Maneuver Staff Tasks
Tempo	Displaying & Manipulating Maps
Utility	Plotting & Manipulating Units
Flexibility in use	Building Overlays Templates
Prevent Fatigue	Creating, Editing Updating Data Bases
Mirror Doctrine	Building Friendly & Enemy Order of Battle
Provide process Short Cuts	Building & Modifying Synchronization Matrix
Consistency between Modules	Preparing Task Organizations
Minimize demand on Memory	Computing Force Ratios
Provide Feedback	Preparing Briefings
Good Error Recovery	Preparing Operation Orders
Process Shortcuts	Building & Displaying Alarms
Common framework	Battlefield Functional Area Software Consistency
Intuitiveness	Sending & Receiving Information

In the MCS Usability Survey's application, a heuristic methodology was used (i.e., a method of usability analysis in which users are presented with an interface design and then requested to comment on it). For the DAWE, the 4th ID MCS operators were asked to rate each usability characteristic (i.e., sub-issue item) on a scale from 1 to 5 to rate the MCS software design as it attempts to support effective execution of identified critical Maneuver Staff tasks.

ARL's MCS Functionality Survey. Utilizing the DOD Universal Joint Task List (UJTL) for C2 as a foundation, ARL's "Functionality Survey" focused on the interrelationship between the division staff functions or processes required for effective maneuver control and decision making as supported by MCS software. ARL's survey metrics methodology, evolved as part of an Army level Science and Technology Objective (STO), established a cross-linking of FM 101-5 Decision Making Processes (DMP) with the MCS software modules believed to support critical command and staff task execution. The Army's field manual (FM 101-5) states that a staff supports the "Science of Control" in four primary ways: (1) gathers and provides information to the commander, (2) makes estimates of the set of actions required, (3) prepares plans and orders, and (4) measures organization behavior. To perform this type of support, the staff and commanders use various time-dependent "Decision Making and Information Management" processes which require extensive staff coordination between and within echelons. Shortcomings in C3I automation functionality can lead to serious tactical failures such as inadequate battle plans, inadequate reporting, lack of coordination, and inadequate situation awareness which can lead to fratricide, for example. It was assumed that the MCS system software function capabilities were developed to support these human-centered command and control processes and avoid errors in judgment and timing.

ARL-HRED developed eight key behavior-based "functionality" dimensions to assess the ability of the MCS to support the DMP and associated critical Command staff coordination. These eight dimensions, listed in Table 3 below, were further defined in terms of sub-dimensions and specific, operationally relevant, staff-related task behaviors. Subsequently, for each critical sub-dimension and task behavior, behavior-anchored rating scales (BARS) were formulated to provide a conceptual (cognitive) framework to guide both the military SME observers as well as the 4ID survey respondents. Here, descriptive definitions were provided as examples of what was considered to be "hindered," "borderline," or "facilitated" task performance. The descriptive examples for each level of performance were assigned a value of 1 through 5 (i.e., 1=hindered performance, 3=same performance as manual method, 5=facilitated performance) to serve as anchors for the five-point scale.

Table 3: Behavior Evaluation Dimensions

Evaluation Dimension	Sub Dimension	Behavior Anchor Focus
Decision Making	Impact on Mission Analysis	Automated information being readily available and assessable to facilitate horizontal and parallel planning
Decision Making	Impact on COA Development	Coordinated input into the developing COAs of key staff perspectives
Decision Making	Impact on COA Analysis	Staff simultaneously analyzing alternative COAs by maintaining a shared common understanding of mission intent, joint identification of COA problems, branch contingencies, etc.
Information Assimilation	Assimilation of digitized messages	Finding, reviewing, and assimilating information from text messages to obtain CCIR
Information Assimilation	Assimilation of digitized graphics	Finding, reviewing, and assimilating information from graphical display to obtain CCIR
Generation of Messages and Reports	Enhance ability to prepare orders and reports	Supporting the staffs' ability to prepare and send desired messages and reports
Situational Awareness	Real-time asses to data sources at all echelons for effective CCIR-based push/pulls?	Staff maintaining a shared, real-time awareness of the battlespace which is formulated into a coordinated RCP. Selective filtering and assimilation of situation-based information.
Situational Awareness	Facilitate effective monitoring of critical events and receipt of critical messages	How digitization assisted the battle staff in keeping each element aware and informed of critical events and factors.
The Relevant Common Picture	Facilitate development and maintenance of a coordinated relevant common picture?	The formulation of the RCP graphic visualizations and initial information dissemination. Staff automatic situation information monitoring. Automated graphic aids for timely RCP and follow-on distribution?
The Relevant Common Picture	Facilitate distribution of the relevant common picture updates to all battle command elements?	Timely distribution of the RCP graphic visualizations and information updates. Automated situation monitoring. Automated graphic aids for timely RCP updating and follow-on distribution?
Span of Control	Facilitate synchronization of operations and coordination between echelons	Timely and accurate information management and coordination reflecting the CCIR,
Interoperability	Facilitate accurate exchange of information between BFAs	Data connectivity should be provided by the ATCCS which supports the horizontal and vertical exchange of information.
Workload Distribution	Appropriately distribute mission tasks between staff	Mission task prioritization and workload distribution.

R ESULTS and DISCUSSION

The focus of the Human Factor Issue (ALO1) within the DIVISION XXI Advanced Warfighting Experiment (DIV XXI AWE) was on the Maneuver Control System's (MCS) interface role in enhancing commander and staff performance. Usability characteristics as well as the ability of the MCS to provide the required functionality to the Battle Staff for planning, information management, decision making, and control of the battle-space were investigated. This section provides results and findings based on analysis of Subject Matter Expert's (SMEs) observations and data from the Usability and Functionality Surveys developed by ARL and administered by the Test and Experimentation Command (TEXCOM).

Analysis of the Human Factors Issue: HFALO1- How effective was the soldier and staff information BOS (MCS) interface in enhancing soldier and staff performance?

1. HF Sub-Issue: How effective is the MCS operator-system interface design?

1.a MCS Software Design. In general, the SME observations and Usability Survey results suggested that the MCS human-computer interface (HCI) has improved over past versions. The MCS software mirrored doctrine and reduced the time it took for the user to perform some tasks. The systems demand on the soldier's memory was not a burden. For the majority of the operators, the MCS was the same or less fatiguing than standard analog methods, automated processes and sequences were intuitive and mirrored doctrine, and its process shortcuts had improved over past versions. The operators felt that the abbreviations, acronyms, and instructions were easy to understand, matched standard military terminology and that the screen display was readable in all light conditions. Feedback has improved. The MCS on-screen "Help" functions are timely and easy to understand but need to be expanded.

Results and Discussion. In general, the SME observations and Usability Survey results regarding the MCS operator-system interface design suggested that MCS version 12.3 usability has improved over past releases, but some software design areas need further improvement. A summary of the results is provided here.

Positive Findings:

(1) The results of the Usability Survey statistical analysis suggest that the MCS software mirrored doctrine. Statistical analysis of the majority of the 4th ID MCS users responded significantly that the MCS mirrored doctrine (65%, Chi Square = 23.29, $p < 0.05$).

(2) Process shortcuts have improved over past versions. Almost three-fourths of the operators rated the MCS process shortcuts capabilities as "fair," "good," or "excellent" with the majority rating them "good."

(3) The MCS automated processes and sequences were intuitive. All of the MCS operators rated the MCS as at least fair regarding the intuitiveness of sequences (Chi Square = 55.9, $p < .05$). At least 80 % of the 4th ID MCS operators felt that the abbreviations, acronyms, and instructions were easy to understand (Chi Square = 68.1, $p < .05$) and the screen display was readable in all light conditions. (Chi Square = 42.5, $p < .05$)

(4) Over half (53%) of the MCS operators rated the demand on their memory as fair, while over 40% rated the demand as low (Chi Square = 18.19, $p < 0.05$).

(5) The MCS reduced the time it took for the soldier to perform tasks. Sixty-six percent of the users rated the speed of performing the tasks on MCS as faster than manual (Chi Square = 42.5, $p < 0.05$).

(6) Sixty-seven percent of the MCS operators felt that on-screen "Help" functions were timely and understandable (Chi Square =29.25, $p < 0.05$), but should be expanded.

(7) The methods for navigating around the screen were not difficult. No operator responded that they were difficult (Chi Square =34.7, $p < 0.05$). The operator may view a different area of the SITMAP by two methods; that is, inputting grid coordinates or clicking in the "up, down, left, right" cursor keys.

Less than Positive Findings:

(1) The statistical analysis indicated that error recovery could be improved. Forty-four percent rated error recovery as poor with an additional 29.5% rating it only fair (Chi Square = 9.61, $p< 0.05$).

(2) The system lacked flexibility. More than 60% of the users felt that the software was not flexible (Chi Square = 33.84, $p< 0.05$).

(3) The operator's opinions of the system's feedback were mixed. Thirty percent rated it poor, while 43% rated it at least good (Chi Square =23.24, $p < 0.05$).

(4) Fifty-percent of the operators had trouble finding the required information on the data displays (Chi Square = 43.27, $p< 0.05$). While over one-third of respondents felt that the automated processes and sequences were consistent, certain menu designs were seen as lacking consistency and were complex. Additionally, the inconsistent application of both UNIX and Windows techniques in the same automated tool was seen as increasing the complexity.

(5) The percent of the MCS staff tasks performed on the MCS versus analog methods varied. Forty-one percent of the soldiers used analog methods more to perform their task while 34% used the MCS more (Chi Square = 1.84, $p > .05$).

1.b Screen Size, Configuration and Design. The MCS 17-inch display contained the information the soldier needed to perform effective command and control. However, users noted that they sometimes could not find the information on the screen while, at the same time, they could not see enough of the surrounding area. The relatively small size and inadequate resolution of the 17-inch diagonal MCS computer screen, along with several other factors, were not able to adequately replace the size and detail provided by the TOC paper maps and acetate overlays. The display had a cluttered appearance because icons and menus were overlaid and crowded. Other limiting factors were the lack of digitized maps at appropriate scales and the lack of computer speed and memory to provide smooth scrolling and rapid zooming of map and overlays.

The one large screen display (LSD) at the Division Main Plans cell was the only display that allowed the 4th ID to view the entire area of operation with enough detail to conduct plans and operations without using map boards. The MCS displays were viewed as passive (i.e., non-interactive) information devices that result in a low level of cognitive involvement by the command group and other Battlefield Functional Area (BFA) managers within a TOC. For the close fight, the small screen display and poor resolution forced the BDEs to rely on paper wall maps and acetate overlays to track the battle and develop their plans.

2. HF Sub-Issue: How effective is the functionality of MCS as it supports staff decision making?

Observations and the results of the Functionality Survey indicated that the use of the Maneuver Control System (MCS) by the staff in supporting the military decision process varied across echelons and was a function of the digitized tools available at each location. A digitized system like the Battlefield Planning and Visualization tool (BPV) was the primary system utilized for mission analysis while the MCS was used for mission execution.

Division Main (DMAIN): The MCS, BPV tool, and MCS-WIN laptop computers with a large 7 ft. by 14 ft. display screen system were used by the DMAIN Plans Cell. These technologies proved to be valuable planning and rehearsal tools to enhance COA development, war-gaming and synchronization as the staff worked collaboratively at a large conference table.

DTAC: Here the DAWE offered a glimpse of future digitization for force effectiveness and decisive operations. The Quick Decision-Making Process (QDMP) was routinely conducted at the DTAC utilizing the Joint Surveillance Target Attack Radar System (JSTARS), Unmanned Aerial Vehicle (UAV), and Air and Missile Defense Workstation (AMDWS) "live" feeds combined with the command level Video Telephone Conference (VTC) discussions. These near real-time visualizations, on-the-fly Course of Action (COA) development, and VTC planning and execution processes increasingly negated the utility of the DMAIN Plans Cell Deliberate Decision-Making Process-based products and raised the QDMP (characterized by VTC-based troop-leading procedures),

to new levels of utility as it became the primary process for decisively exploiting the technology-based opportunities being presented. The MCS played a minor role in this military decision-making process (MDMP) effort, only being used to project the current SITMAP for the VTC. The MCS was subsequently used to draft the resulting FRAGO for issue to DMAIN for final drafting and dissemination.

BDEs: While some BDEs utilized MCS-based M&O (SITMAP) in the Close Battle COA planning process, other BDEs relied on paper maps due to poor size and resolution of the MCS display.

Results and Discussion. Statistical analysis of the Functionality Survey indicated that the majority of the staff (58%) responded that the MCS facilitated or greatly facilitated mission analysis (Chi Square = 5.24, $p > 0.05$). An additional 21% felt that it offered borderline support. At least 70% of the staff responded that the MCS at least borderline supported the development and analysis of the COAs (Chi Square = 10.1, $p<0$ .05). SME observations suggest, however, that MCS itself did not facilitate a rapid decision making process with the exception of the ability to instantly disseminate products once completed. In fact, the use of the MCS by the staff in supporting the military decision making process varied across echelons and was a function of the digitized tools available at each location.

3. HF Sub-Issue: How well does the MCS support "information assimilation" and the Commander's Critical Information Requirements (CCIRs)?

This sub-issue addresses how automation supported finding, retrieving, and assimilating critical task information from distributed textual data base repositories for efficient development of CCIR related information products, such as the Essential Elements of Friendly Information (EEFI), Friendly Force Information Requirements (FFIR), and Priority Intelligence Requirements (PIR). The majority of the respondents stated that the MCS only marginally assisted their efforts in obtaining the CCIR. The text messages and graphics did provide a greater understanding of Corps and Division intent and helped to develop and monitor the CCIR. Also, information assimilation between units was more accurate using the MCS than using FM radio communications. However, the majority of the information was distributed by voice because the process was faster. In addition, there was no standard format for developing and sending EEFI, FFIR or PIR. Finally, the reduced reliability of the MCS and its inability to keep pace with battle requirements and the CCIRs prevented the staff from relying on the MCS.

Results and Discussion. The opinion of the majority of the Functionality Survey respondents was that the MCS only marginally assisted their efforts to obtain the commander's critical information requirements (CCIR) . The MCS text messages and graphics (Chi Square = 14.22, $p< .05$) did promote a greater understanding of Corps and Division intent and helped the development and monitoring of the CCIR. Information assimilation regarding messages and reports between units using MCS appeared more accurate than using FM radio communications (Chi Square = 2.57, $p > .05$). However,

the majority of the information was distributed by voice, but digital orders were performed using Microsoft PowerPoint software.

4. HF Sub-Issue: How well does the MCS contribute to a shared, real-time situation awareness of the Battle-Space?

The level of situation awareness (SA) seen in the DAWE was impressive. The Functionality Survey results suggested that MCS facilitated the staff in maintaining situation awareness. Observations, however, suggest that the unprecedented level of SA was a function of the entire suite of technology (including MCS, JSTARS, UAV, AMDWS, RAPTOR) available to the 4th ID command.

Situation awareness from the MCS was delayed by as much as an hour because of the time-consuming manual input of enemy and friendly military unit location data required from the paper map boards and acetate overlays into the MCS. Once this information was input at the brigade level, the brigade database was digitally transmitted to the next higher echelon where the data was manually fused into an RCP and re-distributed as a common overlay on the MCS. One reason for the need of manual data input involved interoperability problems of the various systems. The lack of a single database system between the echelons also contributed to unit location discrepancies in friendly and enemy information between the MCS and the TOC map boards.

Results and Discussion. The DIV XXI AWE offered a glimpse of the future and how digitization can achieve real-time situation awareness (SA) and facilitate decisive operations. The results of the MCS Functionality Survey (Table 22) indicated that 62% of the respondents felt that the MCS "facilitated" or "greatly facilitated" the staff in maintaining situation awareness (Chi Square = 16.65, $p < .05$).

5. HF Sub-Issue: Compared to manual methods, how well does the MCS help to distribute the task workload among staff members and reduce "mental" workload during continuous operations of low and high activity?

The results of the MCS Functionality Survey and SME observations indicated that changes need to be made to improve the MCS workload distribution. Manual methods still had to be performed in parallel with the MCS methods because (1) the MCS data was not as current or reliable as data obtained from other sources and (2) the MCS might not be available because of power failure or computer problems. Another workload distribution problem resulted from the lack of enough MCS computers to simultaneously perform all the needed TOC functions. Instead "bottlenecks" occurred while some tasks were performed but other tasks had to wait their turn. Some problems also occurred from trying to distribute the workload among TOC members who had different levels of expertise on the various ATCCS. This was especially difficult as personnel changes occurred between shifts.

Results and Discussion. The results of the MCS Functionality Survey indicated that the opinions of the respondents were mixed. Sixty-one percent of the respondents felt that the MCS offered borderline support, hindered, or seriously hindered the commander to distribute mission tasks among staff members. Similarly, the consensus of the subject matter experts was that the MCS did not help to reduce or distribute the workload (Chi Square = 8.94, $p > .05$).

A variety of factors interacted that resulted in an additional workload for the TOC staff personnel. The first was the slowness of obtaining information from the MCS which was usually an hour behind. Secondly, the staff questioned the reliability of the RCP in a situation where the MCS RCP graphic showed no enemy units in the area but telephone communication with some friendly units noted that they were surrounded or in contact with the enemy (Chi Square = 4.75, $p > .05$).

Regarding mental workload, the results of the Functionality Survey indicated that although not statistically significant, more than 69% of the staff felt that the MCS, at least slightly, reduced the commander's and staff's mental workload by helping them to identify and interpret key events and shared mental models of battlefield events (Chi Square = 2.61, $p > 0.05$). To this end, automation facilitated attaining an accurate match between the commander's timely event awareness and accurate interpretation of event meaning. The MCS assisted the commander and staff to become aware that a key event was about to occur, was occurring or had occurred. Given key event awareness, the staff then accurately interpreted the meaning or implications of the key event using all the analog and digital tools available.

6. HF Sub-Issue: How well do the MCS graphics and drawing tools help the Staff in developing templates and overlays on maps ?

In general, observations and survey results indicated that the drawing tools were easy to use and that it was easy to display other segments of large area maps and overlays that could not be shown simultaneously on one MCS screen display monitor. However, while it may have been easy to execute individual MCS graphics functions, the overall MCS system response was not fast enough to complete the many steps needed to achieve an integrated situation map or an RCP graphic display at the modern pace of wartime battles.

7. HF Sub-Issue: What is the MCS reliability and its impact on Staff Performance?

The stability of the MCS has improved over past versions. However, the MCS still exhibited stability and reliability problems. The system frequently crashed, especially during message handling, creation of markers, and loading of digital maps and overlays. The process of restarting the system was lengthy, requiring at least 15 minutes. If the MCS operating system crashes or the SITMAP module is slow in responding, the commander's ability to control the battle is likely to be severely impaired.

8. HF Sub-Issue: How well does the MCS support the development and maintenance of a coordinated, timely and accurate relevant common picture (RCP)?

Generally, the SME observations and Functionality Survey results suggested that the "timeliness" shortcoming associated with the RCP graphics product was affected full utility of the MCS RCP tools. This shortcoming resulted from a com bination of the MCS graphics tool complexity and the human-based coordination and analysis processes that fed the development of the RCP on the MCS computer. Consequently, the 15-60 minute time lag affected the utility of the RCP in fighting the close battle, whereas JSTARS, UAV, AMDWS, etc., provided the DMAIN, DTAC, and BDE commanders with current data that could contribute to the development of an automated RCP.

Results and Discussion. The results of the MCS Functionality Survey statistical analysis indicated that 54% of the respondents felt that the MCS "facilitated" or "greatly facilitated" the staff in the development of a RCP graphic display (Chi Square = 7.73, $p > .05$). Furthermore, observations suggested that the MCS-based RCP current SITMAP "distribution" process was one of the success stories of the ATCCS suite at the AWE. More than 90% of the respondents felt that the MCS supported the staff with adequate information exchanges between echelons (Chi Square = 22.61, $p < 0.05$). This process, the Staff TTPs, and the MCS functionality that supports it, have been evolutionary and have culminated in a product that generally supported the development and maintenance of "general" situation awareness across all echelons of the 4ID. Unfortunately, SME observations also indicated that the "timeliness" shortcoming associated with the RCP product raised an issue regarding full utility. Specifically, the RCP lacked reliability and timeliness in reporting of the location of enemy units. This shortcoming resulted from a combination of the MCS graphics tool complexity and the human-based coordination and analysis processes that fed development of the RCP on the MCS. Consequently, the 15-60 minute time lag affected the utility of the RCP in fighting the close battle, whereas JSTARS, UAV, AMDWS, etc., provided the DMAIN, DTAC, and BDE commanders with the true, up-to-the minute situation awareness of the battlespace.

9. HF Sub-Issue: What is the utility of the "Resource Reports" software?

The majority of the staff did not use the Resource Status Reports module on the MCS. Most staff members relied directly upon text messages and phone calls. One reason was that the data was missing or outdated by the time it was available from the MCS. Therefore, updates were confirmed by voice, while laptop computers with Microsoft MS PowerPoint products were used to develop graphics for reporting resource status during Battlefield Update Briefings (BUBs).

Results and Discussion. The Resource Status Reports on the MCS were not used by the majority of the staff. Fifty-three percent of the 4th ID MCS operator responded that they never used the resource reports and an additional 25% responded that they almost never used this module even though they were aware that the reports exist. At the DMAIN, the G-3, which has the responsibility for tracking current SLANTs (operational vs. authorized

strength of key combat systems), did not use the Resource Status Reports for this, relying instead on text messages and phone calls directly from the DMAIN Sustainment cell. One reason given was that the data was missing or outdated by the time it was available to the MCS.

10. HF Sub-Issue: What is the utility of the "Message Handler" software?

The message handling application was a highly used capability of the MCS that supported the creation, reviewing, editing, sending, and receiving of free-text and standard U.S. Army format messages. Positive message handling aspects noted by SMEs involved the easy creation of a "message distribution list" and the time saved from the "auto-forwarding" of messages to a group of recipients. The "message alarm" system was also observed to be working well. However, some aspects of message handling need immediate improvement. It was generally found that there were too many steps involved in executing message processes. For example, each message received had to be acknowledged individually and messages that were not needed could not be deleted in groups but, instead, had to be deleted individually. Another time-consuming procedure was the necessity to search for a desired message header in a list that was very long. Especially frustrating was the lack of an automatic word-wrapping feature in the message input window where only the words showing in the window were printed.

Results and Discussion. Based on SMEs' observations, the MCS worked well for receiving and forwarding text information (Chi Square = 15.30, $p < .05$). More than 80% of the 4th ID MCS users responded on the questionnaire that the creation of a distribution list was easy. However, the software could be improved. There are too many steps that are required to create, review, edit, and retrieve a message. It can take more than 20 steps to create and error check a message. It was noted by an MCS operator in the DMAIN G3 OPERATIONS cell that the "Message Handler" still requires the user to delete messages one message at a time. This is very inefficient and time-consuming and should be a high priority improvement consideration because a Brigade TOC typically received as many as 1000 messages a day during the DAWE. In contrast, modern word processors generally allow the highlighting of items that are desired to be processed in the same manner. Then the modern processors allow the execution of the appropriate software command to be applied to the highlighted items. This software grouping procedure is needed for sets of messages that are to be processed with the same software command (e.g., "DELETE").

11. HF Sub-Issue: What is the utility of the MCS "Unit Task Organization" (UTO) MCS module?

In most cases the observations and survey results indicated that the "Unit Task Organization" MCS module was not used. It was considered rather cumbersome and very inefficient in sending and receiving information. A unit could not drag and drop a military unit and automatically change the addressing of messages.

12. HF Sub-Issue: What is the ability of the MCS to provide the information to support the synchronization of combat operations?

The MCS did not support synchronization. Observations and survey results indicated that at the DMAIN G3 Plans cell, the staff preferred laptop computer-based Microsoft Excel spreadsheets for creating synchronization matrices. Using MCS-WIN, these distributed computer systems allowed each BOS Representative to work simultaneously on COA-related synchronization matrices and orders. At the brigade levels, planning staffs used a pre-formatted table from MS Excel to build and distribute the product.

13. HF Sub-Issue: How well does the MCS support the staff to generate messages and reports?

The MCS Functionality Survey and SME observations indicated that one of the most valuable features of the MCS was its ability to rapidly distribute orders once they were created. However, there was less acceptance of the MCS's ability to support the staff members in the preparation of reports (e.g., WARNORDs, SITMAPs, OPORDs, FRAGOS, etc.). Consequently, rather than using the MCS word-processing and graphics software, almost half of the staff used commercial software (e.g., Microsoft WORD and MS PowerPoint) on laptop computers to produce reports and then paste them into the MCS free-text format for distribution.

Results and Discussion. While many operators felt that the most important feature of the MCS system in the reports process was its ability to rapidly distribute orders once they were created, there was divided opinion on the ability of the MCS to assist in the report preparation, itself. The 50/50 split of opinions between the MCS operators regarding how well the system facilitated the staff members in preparing desired reports (WARNORDs, SITMAPs, OPORDs, FRAGOS, etc.) showed that there was a lack of a consensus here (Chi Square = 8.81, $p > 0.05$). About half the operators said that the pre-formatted templates and free-text data fields of the provided report displays were easy to navigate and saved time and effort for battle captains. However, the other operators preferred to use commercial software, such as Microsoft Word and MS PowerPoint, on a laptop computer to produce reports and then pasted them onto the MCS free-text format rather than directly using MCS capabilities. In many cases it took too long to pull up the formats, and it was not easy to manipulate the pre-formatted templates. One user found it difficult to locate a text-file that he had previously started, saved unfinished, and stored in the system. One group actually admitted that analog radio "SPOT" reports and manual map updates were used during the DAWE rather than the digital MCS report capabilities.

FRAGOS, OPORDs, etc., were cumbersome to load and send, with the result that this information was not distributed in a timely manner. Immediate FRAGOS still had to be sent via traditional analog communications links. Likewise, clienting into other systems, or pulling information from other systems was not easily done.

14. HF Sub-Issue: How well does the MCS inter-operate with other Battlefield Operating Systems (BOS)?

The results of the MCS Functionality Survey and SME observations indicated that the Army Tactical Command and Control Systems (ATCCS) need improved interoperability between the different Battlefield Functional Area computer systems. The battlefield geometry message was a major interoperability problem for all the ATCCS. Information exchange between ASAS and the MCS was especially problematic. At the Brigade levels, difficulty in passing current enemy situations from the ASAS to the MCS was seen. The FTP tool, however, was used effectively throughout the exercise to distribute information to horizontally- and vertically-organized military units.

Results and Discussion. The general consensus of the 4th ID Battlestaff was that interoperability between the different BFAs needs to improve. While interoperability between MCSs was improved, 50% of the Battlestaff responded that the limited interoperability between the other BOSs and the MCS greatly hindered the accurate and timely exchange of information as well as the horizontal integration of information. An additional 25% responded that the BOSs' interoperability was borderline which resulted with the MCS computer's marginal rating in facilitating the exchange of information.

SUMMARY

Perhaps the most notable aspect of the DAWE was the TOCs' continued use of paper maps and acetate overlays in an experiment that provided a multitude of digitized information processing systems. The MCS computer offered some digitized support for human-computer interfacing map and overlay processes but the relatively small MCS display screen, inadequate map scrolling and zooming features, and generally slow response times, were not able to keep up with the pace of modern battlefield operations.

The Functionality and Usability Survey respondents felt that the MCS software had improved since previous AWEs but some areas need further improvement as follows:

1. Increase speed of error recovery.
2. Improve software consistency for display formats.
3. Provide movable mouse for right and left-handed users.
4. Embed training into the MCS.
5. Pursue large screen interactive computer display technology.
6. Automate map overlay data transfer from paper map to the MCS computer.
7. Provide more MCS displays and workstations for improved workload distribution.
8. Improve collaborative decision making processes by incorporating BPV capabilities.
9. Provide a standard format for developing and sending CCIR information.
10. Improve reliability of MCS hardware, software, and training systems.
11. Increase speed of MCS reporting of locations of friendly and enemy military units.
12. Improve efficiency of message handling.
13. Improve interoperability of the various BFA computer systems.

A FRAMEWORK FOR MODEL VALIDATION

Robert G. Easterling
Sandia National Laboratories
Albuquerque, NM 87185

ABSTRACT

Computational models have the potential of being used to make credible predictions in place of physical testing in many contexts, but success and acceptance require a convincing model validation. In general, model validation is understood to be a comparison of model predictions to experimental data, but there appears to be no standard framework for conducting this comparison. This paper gives a statistical framework for the problem of model validation that is quite analogous to calibration, with the basic goal being to design and analyze a set of experiments to obtain information pertaining to the 'limits of error' that can be associated with model predictions. Implementation, though, in the context of complex, high-dimensional models, poses a considerable challenge for the development of appropriate statistical methods and for the interaction of statisticians with model developers and experimentalists. The proposed framework provides a vehicle for communication between modelers, experimentalists, and the analysts and decision-makers who must rely on model predictions.

INTRODUCTION

Mathematical models of phenomena, processes, products, and their performance, accompanied with high-performance computing, have the potential of reducing the amount, or changing the nature, of the physical testing required to design and produce complex components and systems, predict their performance in various environments, and certify their safety and reliability. This is the premise behind a tremendous amount of national and international research and code development and is manifested in DOE (Department of Energy) plans for science-based stockpile stewardship and such programs as ASCI (the Advanced Scientific Computing Initiative). Similarly, modeling and simulation of weapon systems and military operations are becoming increasingly important in DoD acquisitions and decision-making. The value of such models (existing or future) is realized when model calculations, at points in the parameter space that have not been or cannot be physically tested to provide confirmation, can be trusted for the purpose of drawing conclusions or making important decisions at these untested points. (For example, the DOE may be required to certify a new component's performance when subjected to radiation fields only achievable in underground nuclear tests, now precluded.) This trust comes when physical phenomena are well-understood and expressed mathematically, then accurately converted to code, and when empirical support is obtained at some points in the model's parameter space.

Developing this empirical support, the process of which is generally and generously termed 'model validation,' is well-recognized as critical to the success of modeling and simulation. Yet, there does not appear to be any consistent approach to model validation, no overarching guidelines or framework for linking model objectives with validation efforts. The expense and difficulty of testing, and the absence of clear validation objectives, can lead to ad hoc approaches to validation, which in turn can be unconvincing to potential users of the model. A more formal approach to validation is needed.

Model-based prediction is a combination of scientific and statistical inference, because it is based on both theory and the accompanying model-building or validation data. Determining the nature and amount of validation testing is a problem in statistical experimental design -- a systematic approach to determining a suite of tests that can efficiently estimate or predict characteristics of interest with predetermined precision. Hence, the problem of model validation is to a major extent a statistical problem. Statistics provides methods and approaches for stating the validation objectives and determining the data requirements for meeting those objectives. These methods have typically been applied to model validation problems only in fairly simple mathematical situations. Extending statistical methods to models of complex interactions and

geometries, captured, e.g., in finely meshed, finite element, massively parallelized codes, is a major challenge.

Statistical methods also pertain to the analysis of data that result from model validation experiments and the corresponding model calculations. Model-based predictions are uncertain and the user or decision-maker who will rely on the predictions needs some idea of credible 'limits of error' to associate with a prediction. Thus, the objective pursued here is to design and analyze model validation experiments in ways that permit any model-based prediction to be accompanied by a credible 'limit of error.' This is an extension beyond typical model validation analyses which overlay computations with experimental results and invite the judgment: close enough, or not. The model validation situation is comparable to calibration, in which measurements from two methods of measuring physical phenomena are compared, the objective being to be able to use one method to predict the other and to characterize how good the predictions are.

MODEL VALIDATION IN THE NEWS

A 1996 news item pertaining to model validation stated something like:

> DoD says comparison of computer simulations versus live-fire tests of the effect of gunfire o helicopter blades shows that even the most sophisticated computer models cannot accurately mirror real life. On a scale of 1 to 10, the models scored a 7 in predicting how the shell would penetrate the blade, a 3 in predicting the destruction of the blade, and a 2 in predicting the loss of a helicopter.

This brief item demonstrates several important facts about modeling and model validation. First, note the modeling chain: phenomenon, component, system. Accurate modeling gets more difficult as one moves up the chain from single phenomena to multiple, interacting phenomena. The physics of shell penetration can be modeled fairly accurately, but predicting whether that penetration will destroy the blade is more difficult and predicting whether the helicopter will crash as a result is more difficult still. This example also points out a problem in resource allocation. Where should future resources be spent? Improving the shell model to a 9? Improving the helicopter model to a 4? And, how to decide? Within an overall budget, how should resources be divided between modeling and testing? What is the most cost-effective and scientifically sound way to develop and certify a survivable helicopter?

Modern advances in modeling and computing capabilities, plus pressure to reduce the amount of money spent on test facilities and testing, have led present decision-makers to lean toward the modeling and simulation side. The cost of developing these tools, particularly in situations in which important, real decisions are going to ride on them, is being found to be substantial, though. These situations also call for the most care in model validation testing and, as will be illustrated here, this, too, can be a substantial undertaking. There is a clear need for careful analyses of the evolution of testing and modeling and simulation lest we lose confidence in the decisions that have to be made on their combined basis as this evolution progresses.

This example uses scoring on a 1 to 10 scale as a means of measuring agreement of computer models and 'real life.' Doing so is a useful communication device, but it does point to the problem of developing definitive, informative measures of this agreement.

TERMINOLOGY

If 'validation' is interpreted as meaning 'to establish the truth of,' then it is clear that such a goal is unattainable. A model is, after all, an approximation. Even if it is found to closely approximate reality in selected circumstances, there is no guarantee that it will do so in all circumstances. This is particularly true when the model has been calibrated, or 'tuned,' to match test results. Oreskes et al.[1] discuss these and related issues in depth and provide a healthy antidote to glib claims of validated models.

The most prominent definition in use today appears to be that adopted by the DoD's Defense Modeling and Simulation Office[2]:

> Validation. The process of determining the degree to which a model or simulation is an accurate representation of the real-world from the perspective of the intended uses of the model or simulation.

Thus, 'determining/estimating the degree of accuracy' recognizes that the goal is not the establishment of agreement, but the measurement of disagreement between the model and the situation it is approximating. This definition leads naturally to the above-stated goal of being able to derive credible limits of error from validation testing and data analysis.

Model validation is sometimes set up as a test of the hypothesis: MODEL = NATURE. Then, based on the test data, the decision is made either to reject or accept this hypothesis. The alternative view here is that the basic problem is one of estimation: What is the magnitude of (MODEL – NATURE) within the set of situations in which it is desired to use the model to approximate nature? With that information, the user can decide whether the approximation is adequate for "the intended uses."

MATHEMATICAL SET-UP

"All models are wrong, but some are useful,[3]" is a statement by George Box, University of Wisconsin, that succinctly captures the essence of the preceding section. This statement and the problem of model validation can be expressed mathematically, as follows:

Model.

$y^* = h(x)$, where
$h(x)$ is the (computer) model of the phenomenon of interest,
x = model input (in general, a vector),
y^* = model output, a prediction of a characteristic, y (possibly a vector)

Nature.

$y = g(x, w)$, where
w = additional parameters or variables that influence nature's outcome,
$g(x, w)$ is nature's function (generally unknown),
y = nature's outcome, at x,w.

In words, the computer model, h(x), takes input x and calculates output y*. I assume here that the computer model is fully specified – grid size, convergence criteria, internal parameters, etc. – and that the vector x describes, in general, a physical entity and the environment to which it is subjected. The predicted outcome is y*. The vector x is an abstraction of nature, so nature's variables at work in the simulated situation will generally involve variables beyond what the modeler has chosen. These are the w's. The functional relationship is consequently different, and might be different even if there were no w's. For example, nature may be nonlinear where linearity is assumed in h(x). Also, nature's relationship might not even involve some x's the modeler has chosen. Nature's function, g(x,w) is unknown and some or all of the w's are also unknown. An example, given below, may help clarify these ideas.

ERROR

Consider now the difference between model and nature:

$$\text{error}(x,w) = y^* - y$$

The degree-of-accuracy of the model is characterized by error(x,w), abbreviated as e(x,w) below, as x and w range over spaces, X and W. Trust in a model prediction comes from knowing how large e(x,w) might be, i.e., how far y* might be from nature's result, and knowing that the decision to be made is unaffected within this range. The only way to learn anything about e(x,w) is to run tests, or collect data, at selected x's, allowing nature to generate w's at will. There is no substitute.

EXAMPLE

Chambers et al.[4] developed a model for the stresses and strains produced in a structure by the curing of an encapsulant. A comparison of model and experiment was obtained from instrumented tubes that were filled with a resin, then cured. The 2-D model assumed the tube was a perfect cylinder, the geometry of which is specified by three parameters: diameter, length, and wall thickness. These are x's in the model. Experimental results, generated as a check on the model, showed deviations from the predicted strains and these deviations were found to be due to tube out-of-roundness. Thus, actual tube dimensions, which could take a 3-D map of hundreds of variables to characterize, are nature's w's in this example, and they produced noticeable prediction errors. Whether such errors would preclude use of the model for certain situations would be a matter to investigate.

Enhanced computing capabilities make it possible to develop a 3-D model. A heuristic that helps describe this situation is that advanced code development would move some of the w's into the x's for a more detailed computer model. Validation testing would be used to check to see if the remaining w's, and modeling error, lead still to appreciable prediction errors. It is possible that the code might predict well for some types of out-of-roundness, poorly for others. To validate that the code adequately predicted tube strains for a wide variety of geometries might take a considerable amount of experimentation.

Alternatively, one could treat the 2-D model as an expected value model -- the input parameters would be the average tube diameter and wall thickness -- and test a random sample of instrumented tubes to provide a statistical characterization of deviations from expected strain. One would not have to take detailed dimensional measurements on the tubes – only assure that they were representative of the tubes for which predicitons are desired.

Either of the preceding options is appropriate if the objective of the model and experimentation is the prediction of the range of strains that might be produced in encapsulated tubes. Alternatively, if one were just interested in choosing the best curing temperature, comparative calculations using the original 2-D model might provide a valid basis for making that process decision, even though the model doesn't capture the effect of tube out-of-roundness, an effect that might largely cancel out in a comparison of curing temperatures. Establishing such validity, though, brings us back to the basic problem of model validation with respect to predicting differences. If validity of predicted differences is established, a further inferential question to consider, though, is whether one can also infer that the selected temperature is also (near) optimum for other geometries, such as a more complex shape. Other refinements could also be pursued. For example, resin is not necessarily homogeneous within a tube and its properties can vary from lot to lot and pour to pour. The effect of these variabilities can either be incorporated into the model, or estimated experimentally.

This example points out an important conundrum that will have to be resolved. The modeler believes that by putting more physics into the model, which generally means putting more x's into the model, model predictions will improve and the error can be driven to negligible. This is science. The experimentalist, however, recognizes that the more parameters are included in the model, the more difficult it will be to carry out validation tests that adequately explore the parameter space and characterize prediction ability. The potential user wants some of both – adequately sophisticated models adequately supported by data. The sponsor cannot afford unconstrained modeling or testing. The proposed mathematical framework offers at least the start of a context within which this battle of the x's and w's can be debated and resolved.

STATISTICAL CHARACTERIZATION OF ERROR

Because not all of nature's w's are known, or measured, in a validation experiment, it is not possible to observe e(x,w) directly. For any situation described by x, nature may generate a variety of outcomes because of variability in the w's that also come into play in a situation the modeler describes by x. This situation will be described statistically: the random variable, e_x, has a probability distribution indexed by x. That is, at any point x, in the X-space, there is a distribution of possible errors – deviations between nature and y* at x. The nature of that distribution could, and is likely to, depend on x; predictions are apt to be more accurate in some parts of X than in others. An ideal situation, beloved by statisticians, would be for e_x to be normally distributed, with mean zero and constant standard deviation throughout X. In general, though, one can consider a mean function, $\mu(x)$, and a standard deviation function, $\sigma(x)$, as characterizing error. The goal of validation testing is to shed some useful light on these two functions –.we can't hope to know them well over a high-dimensional, complex X-space, and we can't hope to say anything definitive about the nature of the distribution of e_x.

VALIDATION EXPERIMENTS

EXPERIMENTAL CONDITIONS

Designing a validation experiment, in the framework of the preceding section, means to select some points in the X-space, then conduct tests at those points. Because our goal is inference about how nature would respond at those x-conditions, it is important that the experiment be conducted in such a way that nature's variability in the w's is given full rein. For example, if the tube-strain tests had been conducted with high-precision tubes, the results might have been better in terms of how well the computational results matched the test results, but we would not have a good estimate of how much variability in strain rates was incurred by standard issue tubes and therefore would not have obtained data on which to base credible limits of error for predictions about that population.

This statistical perspective is contrary to a common view of validation testing, which is that well-controlled, subscale lab experiments can be used to validate models that will be applied to full-scale systems under use conditions. While certain physical laws may scale, it is also necessary to know how residual, or error, variability will scale, if validation testing is done at a subscale level. The main point to make is that the nature of the experimentation, not just the points in X-space, needs to be addressed in planning model validation experiments.

EXPERIMENTAL DESIGN

Objectives. Now, consider the selection of test points in the X-space. The number and location of those points depends on objectives and resources. At one extreme, if validation is desired at a single x-point, and experimentation is possible at that point, then the obvious thing to do is to test at that point. The only statistical issue would be to decide how precisely the bias (mean error) and/or the sigma at that point need to be estimated and run the appropriate number of tests, resources permitting. Of course, in this situation, the model is somewhat superfluous – the test results themselves can be used directly to predict performance at the selected x-point. A slight expansion is the case in which subject-matter knowledge enables one to say that error variability exhibited at the selected x also applies in some specified region about x, so that error limits for subsequent model predictions in that region are subject to the same limits of error as determined at the test point. At the other extreme, validation is desired throughout a high-dimensional X-space, thus requiring experimental 'coverage' of that space, to some degree.

To illustrate the sort of objectives that might drive validation testing, consider a simple linear model: Theory, and the computational model built on that theory, says that a single response variable, y, is a linear function of a single input variable, x. Thus, $y^* = \alpha + \beta x = h(x)$ is the computer model. Theory might provide the coefficients, α and β, or theory might say that they are functions of materials or environments,

in which case experimentation would be required to estimate them, as well as 'validate' the linear model. Possible testing objectives in this situation include:

a. test the linearity assumption
b. estimate the model coefficients
c. test the agreement of nature with theoretical model coefficients
d. test that the slope is positive
e. estimate the error standard deviation, $\sigma(x)$, as a function of x.
f. some subset or all of the above

Statistical Power. Because of nature's variability – induced by the w's not captured in the model – the sorts of questions embodied by the preceding list of objectives can be answered only with some uncertainty. That is, there is some chance that with limited data, chance variations could mislead us, for example, by indicating that the slope is positive when it really is negative. Statistical power is the method by which risks can be controlled. For example, we could design a test such that there is a 90% probability of detecting that y* at a particular x-point differs from nature's mean outcome at that point by more than 20%. Or, design a test such that there is only a 5% chance of concluding the slope is positive when in fact it is really negative by some specified amount. Solutions to these sorts of design problems generally depend on a prior estimate or assumption of error-variability. Preliminary experimentation may be required to provide this information.

Uncertainty Analysis. Computer models can contain physical constants, such as transfer coefficients and material properties, that are estimates from limited data and so are uncertain approximations of these constants. In the set-up here, these estimated constants are considered to be part of the model, h( ), not the model input, x. That is, these constants have to be specified to run the model, just as grids and convergence criteria have to be specified. A common analysis in this situation is to assume 'uncertainty distributions' that probabilistically represent the uncertainty of the estimated constants, then, by Monte Carlo or other methods, propagate these distributions through h(x) to generate an uncertainty distribution of y*. There is a temptation to interpret this uncertainty distribution in terms of how well y* predicts nature, but such is not justified. The uncertainty distribution of y* just characterizes how close the chosen model might be to the best model, the one with the correct physical constants in it. It does not, and cannot, capture the effects of nature's w's that are not even in the model and which contribute to prediction error.

Conceptual Outcome of a Validation Experiment. Table 1 illustrates the conceptual outcome of a suite of validation experiments. In general, n points in X-space are selected for testing and ri tests are done at the ith point. As discussed, the design issues are how many x points to select, where should they lie in the X-space, and how many replications of the experiment should be run at the selected x-points. Figure 1 shows the conceptual outcome for the simple case of a single x-variable, and tests conducted at two levels of this variable. This figure provides a glimpse into how one might model e(x) as a function of x and to quantify prediction errors at untested x, within physically-supported reason. That is, errors appear to be centered on zero, but with a standard deviation that increases with increasing x. Statistical analysis could characterize this pattern and also characterize the uncertainty associated with estimates based on such limited data.

Table 1. Conceptual Model-Validation Experiment:
Design and Outcomes

x	Model	Experiment	Errors
x_1	y^*_1	$y_{11}, y_{12}, \cdots, y_{1r1}$	$e_{11}, e_{12}, \cdots, e_{1r1}$
x_2	y^*_2	$y_{21}, y_{22}, \cdots, y_{2r2}$	$e_{21}, e_{22}, \cdots, e_{2r2}$
...	...	...	...
...	...	...	...
x_n	y^*_n	$y_{n1}, y_{n2}, \cdots, y_{nrn}$	$e_{n1}, e_{n2}, \cdots, e_{nrn}$

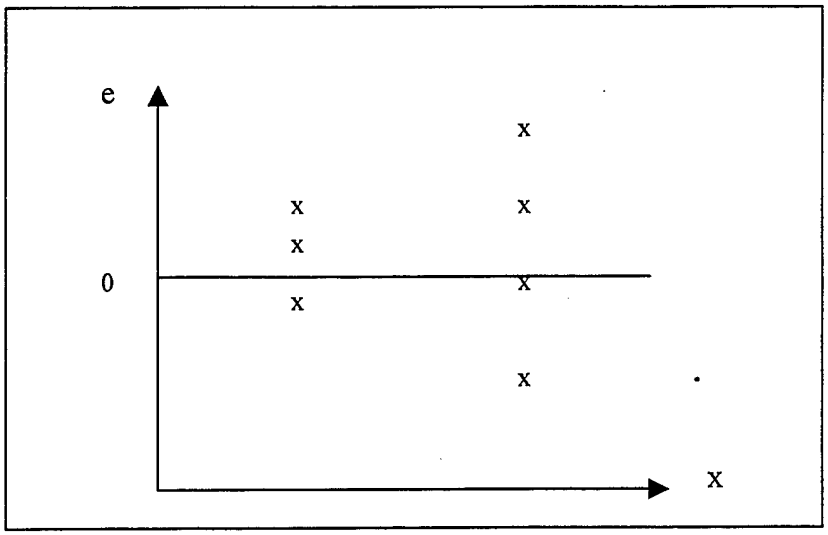


Figure 1. Conceptual Outcome of Validation Test: One-Dimensional X

To conduct validation experiments, as just described, it is necessary to be able to control the x's experimentally. For high-dimensional models of complex phenomena, such experimental control may not be possible. The model may contain x's that cannot be controlled, or even measured in a test environment. As will be discussed, this poses problems to a validation test data-analysis and will likely lead to more uncertain predictions than would be achievable if the x's can be controlled. There is a clear need for communication between the model developer and experimentalist. Building the most sophisticated model possible, then throwing it over the wall for the experimentalist to validate will not lead to a model with the best predictability.

Uncontrollable x's. In many contexts, even with the best efforts of modeler and experimentalist, some x's will be uncontrollable, either directly or by controlling other variables that determine particular x's. Suppose the input vector, x, is separated into its controllable and uncontrollable variables, x_C and x_U. The experiment will then consist of selecting points in X_C and conducting the experiment so that nature can randomly deal x_U's as well as the w's. On the computational side, one could generate a set of y*'s at a given x_C by drawing Monte Carlo samples of x_U, from an assumed or estimated conditional distribution of x_U given x_C. Thus, rather than the point-to-point comparison of test results and computations, as illustrated in Table 1, the analysis will involve comparisons of collections of test results (samples) and computations at the selected x_C points. It is beyond the scope of this paper to work out analysis details by which one would obtain credible limits of error in this situation, but clearly there are more sources of variability in the data in this case than above, so broader error limits, than would be obtained if all the x's are controllable, are the likely result.

Uncontrollable x's are also a problem for the potential user of a model. Suppose one has a model of a production process and wants to use it to optimize the design of the process – determine the times, rates, and temperatures, e.g., at which the process will be run. If key variables in process control are uncontrollable x's, the knowledge is of little use to the process engineer. In general, the model developer, experimentalist, and model user all need to be involved in developing meaningful, usable, validated models. While this may be know at least in the abstract, the framework developed here illustrates the importance of such communication.

Shortcuts. If one adopts it that a goal of model validation testing is the development of a statistical characterization of the error function, e(x), over a specified X-space, then it is clear that for high-

dimensional X a considerable amount of testing would be required – on the same order of the amount of testing required to build an empirical model, h(x). In general, this is grossly impractical and defeats the purpose of developing models that can be used in place of testing. Either this goal has to be abandoned, or scaled back. Some possible shortcuts are:

a. reduce dimensionality. Focus on a subset of the x's and set the remaining x's at bounding values. The selection of the subset could be based on a sensitivity analysis of h(x), the underlying assumption being that the code is adequate for this purpose.

b. focus on a sub-space of interest. Use h(x) to help find interesting subspaces

c. do worst-case testing. Through subject-matter knowledge, identify points or regions in the X-space where the model/test disagreement is expected to be maximized.

d. test at the sub-model or single phenomenon level. This approach would be justified in cases in which the interactions and interfaces between phenomena are known not to be sources of prediction error.

CONCLUDING COMMENTS

There is no single validation problem to be worked, but rather a collection of problems. The focus here has been on point predictions, y*, but there are contexts in which 'integral' predictions are of interest. For example, in the tube encapsulant example, one may be interested in predicting the extreme strains in some population of tubes whose geometry, as characterized by x, varies. Assumed probability distributions for x could be propagated through h(x) to predict a parameter such as the 99th percentile of strain, or the probability that strain exceeds a defined failure threshold. Testing of a sample of tubes would lead to estimates of the same parameters. Differences between these estimates could either reflect modeling error, such as the effect of w's not included in the model, or erroneous assumed distributions of the x's. The possibility of either compounding or offsetting errors is apparent.

It is one thing to write about model validation in the abstract; quite another to apply the concept. Just as a computer model needs to be tested against reality, theories of model validation need to be tested in real applications. It can be expected that such testing will lead to refinements and improvements, and an increased appreciation of the difficulty of the task. The nature of both modeling and validation testing may change as a result of this testing.

REFERENCES

1. Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994) "Verification, Validation, and confirmation of Numerical Models in the Earth Sciences," *Science*, v. 263, 641-646.

2. Glossary of Modeling and Simulation (M&S) Terms, (January, 1998) Defense Modeling and Simulation Office, available at http://www.dmso.mil/.

3. Box, G. E. P. "Robustness in the Strategy of Scientific Model Building," (1979) *Robustness in Statistics*, Launer, R. L., and Wilkinson, G. N., ed., Academic, NY, London, 201-236.

4. Chambers, R. S., Lagasse, R. R., Guess, T. R., Plazek, D. J., and Bero, C. "A Cure Shrinkage Model for Analyzing the Streses and Strains in Encapsulated Assemblies," (1992) *Proceedings of the Joint ASME/JSME Advances in Electronic Packaging Conference*, vol. 1, G0660A, 537 – 542.

DIRECTIONAL SPC: A MULTIVARIATE NONPARAMETRIC SPC TECHNIQUE FOR DETECTING DIRECTIONAL PROCESS CHANGES*

B. Birgoren and R. R. Barton
Department of Industrial and Manufacturing Engineering
The Pennsylvania State University, University Park
State College, PA 16802

We discuss a new multivariate SPC technique, directional SPC, which monitors process shifts in prescribed directions simultaneously. The monitored statistics are the inner products of a random quality vector with prescribed directions. Directional SPC controls the overall error rate at a prescribed level; marginal error rates are also controlled with respect to prescribed ratios denoting the relative importance of directions.

INTRODUCTION

A restrictive assumption in multivariate statistical process control (SPC) as well as in univariate SPC is the assumption of normality. It has been believed that a process variable is usually under the additive influence of many different factors, therefore exhibiting a central limit effect. The application of the central limit theorem to subgroup means has also been given as further justification for normality assumption. On the other hand, Willemain and Runger [1] pointed out that as control charts are used more widely, more nonnormal data is encountered. They also argued, based on their experience with manufacturers, that the use of individual X-charts is increasing. Charting individual measurements might be necessary when either the production rate is too slow to conveniently allow subgroup sizes greater than one, or when repeated measurements differ only because of laboratory or analysis error, as in many chemical industries [2]. Even when SPC is applied using subgroups, the distribution of the sample means may converge slowly to a normal distribution. One solution for monitoring a nonnormal variable is to transform the variable into a normal one. However, use of a transformed variable may result in difficulties in interpretation of control charts [3].

There have been relatively few developments in the area of multivariate nonparametric SPC. Liu [4] and Hayter and Tsui [5] provided two different approaches. Liu [4] provided a set of nonparametric multivariate SPC procedures based on a *reference sample*, a set of random measurements of product quality vectors produced by an in-control process, and the concept of *data depth*. Data depth is a measure of the centrality of an observation with respect to a reference sample. Like Hotelling's T^2, this approach provides little insight as to the nature of the problem when an out-of-control signal is generated.

Hayter and Tsui [5] developed a multivariate SPC procedure that produces practically interpretable out-of-control signals. Their (parametric) procedure assumes that the monitored random vector follows a multivariate normal distribution, and calculates a set of simultaneous two-sided confidence intervals for the means of the components of a new observation. The process is declared out-of-control when any of these intervals does not contain the in-control mean value; the components whose intervals cause the out-of control signal can be identified as those responsible for the signal. Hayter and Tsui [5] proposed a nonparametric version of their procedure. This version operates the same way except that it uses a reference sample directly instead of a simulated sample. Hayter and Tsui's procedure may not properly detect out-of-control situations when certain assignable causes shift the process mean of several components simultaneously.

In this paper, we describe a new nonparametric SPC technique, *Directional SPC*, for monitoring an m-variate random vector $\mathbf{Y}$ for location shifts in k ($k \geq 1$) directions of interest, $\{\mathbf{u}_i: \mathbf{u}_i \in \mathbf{R}^m, \mathbf{u}_i \neq 0, 1 \leq i \leq k\}$. Monitoring process shifts in specified directions for a multivariate normal vector $\mathbf{Y} \sim N(\mu, \Sigma)$ has been previously discussed by Healy [6], Hawkins [7], and Runger [8]. Hawkins [7] and Runger [8] pointed out that in certain processes assignable causes may be known to shift the process mean in certain directions; in these situations control schemes specifically designed to detect these shifts provide more powerful diagnostic than those based on Hotelling's T^2. Barton and Gonzalez-Barreto [9] also identified relations between directions and assignable causes, and proposed a regression-based multivariate diagnostic method.

---

* Approved for public release; distribution is unlimited

The test statistic $T = \mathbf{u}'\mathbf{Y}$ is proposed in directional SPC; the projection of $\mathbf{Y}$ onto the vector $\mathbf{u}$ is given by $(\mathbf{u}'\mathbf{Y})\mathbf{u}$, therefore a location shift in $\mathbf{Y}$ in the direction $\mathbf{u}$ causes the location of T to increase for a general class of multivariate distributions that are characterized in the following section. Directional SPC computes simultaneous one-sided confidence intervals $(-\infty, L(\mathbf{u}_i)]$ for $T_i = \mathbf{u}_i'\mathbf{Y}$ using either a simulated sample from the distribution of $\mathbf{Y}$ or a reference sample in the absence of any information about the distribution. The control intervals are computed by controlling the overall error rate at a specified level α. A process is declared out of control when any of the intervals does not contain its respective value T_i. Note that two-sided sided control intervals for T_i can be implemented by constructing two one-sided intervals for the directions $\mathbf{u}_i$ and $-\mathbf{u}_i$.

In the following sections, we first discuss the directional SPC methodology and then introduce an algorithm for calculating the multivariate control region from a reference sample. Finally, application of this algorithm to nonparametric and parametric situations are considered.

METHODOLOGY FOR DIRECTIONAL SPC

We consider the following multivariate control situation. Assume that a product is produced in a manufacturing process, and m characteristics of the product are monitored to determine the quality of the product. Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)'$ be a random vector whose components represent the m quality characteristics. When the process is in control it is assumed that the quality characteristics follow a prescribed multivariate probability distribution F, that is, $\mathbf{Y} \sim F$. Let $S_n = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n)$ be a sample of n random observations from F. S_n is considered as a simulated sample from F when F is known, and as a reference sample when F is unknown. Directional SPC uses S_n to calculate the control intervals for T_i in an empirical fashion.

We assume that F has support over $\mathbf{R}^m$ and has a density function, and we assume that the overall error rate α is between 0 and 1. Let $\mathbf{Y} = \mathbf{Y}_{n+1}$ be a new observation, and suppose that the real distribution of $\mathbf{Y}$ is G. Then a general SPC problem is to test the hypothesis H_0: $G = F$, vs. H_a: $G \neq F$. Directional SPC is designed to capture departures in the form of location shifts from F to G, hence the particular problem becomes testing the following multiple hypotheses:

$$\begin{aligned} &H_0(\mathbf{u}_i)\text{: There is no location shift in } F \text{ in the direction } \mathbf{u}_i \text{ vs.} \\ &H_a(\mathbf{u}_i)\text{: There is a location shift in } F \text{ in the direction } \mathbf{u}_i, \ 1 \le i \le k. \end{aligned} \tag{1}$$

When G and F belong to the same location family such that G has shifted in the direction $\mathbf{u}_i$, ie. $g(\mathbf{y}) = f(\mathbf{y} - \mathbf{u}_i)$ where g is the pdf associated with G, it can be easily shown that the location of the distribution of T_i increases under assumptions on F. Therefore, comparing new observations of T_i against an upper control limit $L(\mathbf{u}_i)$ is a sensible test for the hypothesis $H_0(\mathbf{u}_i)$: $H_0(\mathbf{u}_i)$ is rejected when T_i is greater than $L(\mathbf{u}_i)$. The rejected hypotheses directly indicate the assignable causes to be investigated when the directions are associated with certain assignable causes.

The overall error rate α for (1) is defined as the probability of falsely rejecting at least one hypothesis $H_0(\mathbf{u}_i)$ when $\mathbf{Y} \sim F$. Accordingly we define marginal error rates $\alpha(\mathbf{u}_i)$ as the probability of falsely rejecting a single hypothesis $H_0(\mathbf{u}_i)$ when $\mathbf{Y} \sim F$. If the hypotheses are equally important, then $\alpha(\mathbf{u}_i)$'s are all equal; $\alpha(\mathbf{u}_i) = \gamma, \ \forall i$, $1 \le i \le k$. In this case γ is a function of α and F. Instead, suppose that a quality engineer is more interested in detecting changes in some particular directions than some others, and is able to weight the corresponding hypotheses according to a ratio of marginal error rates. A weight $w(\mathbf{u}_i)$ is assigned for each direction $\mathbf{u}_i$ such that $w(\mathbf{u}_i)$ takes values between 1 and r $(1 \le r < \infty)$, $\min(w(\mathbf{u}_i)) = 1$, $\max(w(\mathbf{u}_i)) = r$, where r denotes the highest importance and 1 denotes the lowest. Then $\alpha(\mathbf{u}_i)$'s can be set to reflect the relative importance of each hypothesis:

$$\alpha(\mathbf{u}_i) / w(\mathbf{u}_i) = \gamma, \ 1 \le i \le k. \tag{2}$$

In (2), γ is the marginal error rate of the hypothesis with the lowest importance since $\min(w(\mathbf{u}_i)) = 1$. In this case γ is a function of α, F and $\{w(\mathbf{u}_i): 1 \le i \le k\}$. This weighting scheme was suggested by Westfall and Young [10] in the context of weighted p-values for multiple hypotheses.

A control interval for $T_i = \mathbf{u}_i'\mathbf{Y}$, $CI(\mathbf{u}_i) = \{\mathbf{u}_i'\mathbf{y}: \mathbf{u}_i'\mathbf{y} \le L(\mathbf{u}_i)\}$, can be viewed as a directional control region for $\mathbf{Y}$, $CR(\mathbf{u}_i) = \{\mathbf{y}: \mathbf{u}_i'\mathbf{y} \le L(\mathbf{u}_i)\}$, which defines a closed half-space on $\mathbf{R}^m$. Since the events $\{\mathbf{u}_i'\mathbf{Y} \in CI(\mathbf{u}_i)\}$ and $\{\mathbf{Y} \in CR(\mathbf{u}_i)\}$ are equivalent, their probabilities are the same. Using this information, the directional control regions, $CR(\mathbf{u}_i)$, and the overall control region, CR, can be set simultaneously as given in (3) and (4).

$$CR(\mathbf{u}_i) = \{\mathbf{y}: \mathbf{u}_i'\mathbf{y} \le L(\mathbf{u}_i)\}$$
$$\text{s.t. } P(CR(\mathbf{u}_i)) = P(\mathbf{Y} \in CR(\mathbf{u}_i)) = 1 - w(\mathbf{u}_i)\,\gamma \text{ when } \mathbf{Y} \sim F \tag{3}$$

In (3), the value of γ and $L(\mathbf{u}_i)$ are determined by the coverage probability of the overall control region CR, $1 - \alpha$.

$$CR = \{\mathbf{y}: \mathbf{u}_i'\mathbf{y} \le L(\mathbf{u}_i), \forall \mathbf{u}_i, 1 \le i \le k\}$$
$$\text{s.t. } P(CR) = P(\mathbf{Y} \in CR) = 1 - \alpha \text{ when } \mathbf{Y} \sim F \tag{4}$$

Since $\{CR(\mathbf{u}_i), 1 \le i \le k\}$ are closed half-spaces and CR is the intersection of these half-spaces, CR is a polyhedron. In connection with (3) and (4), overall and directional rejection regions, RR and $RR(\mathbf{u}_i)$, $1 \le i \le k$, can be defined as the complementary sets of CR and $CR(\mathbf{u}_i)$ with $P(RR) = \alpha$ and $P(RR(\mathbf{u}_i)) = w(\mathbf{u}_i)\,\gamma$.

Example. Consider a bivariate SPC situation: $\mathbf{Y} = (Y_1, Y_2)'$, $\mathbf{Y} \sim F$, and four directions are specified as $\mathbf{u}_1 = (1, 0)'$, $\mathbf{u}_2 = (-1, 0)'$, $\mathbf{u}_3 = (0, 1)'$, and $\mathbf{u}_4 = (1, 1)'$ with the respective weights 1, 1, 2, and 2. Setting these directions corresponds to specifying an upper and a lower control limit for Y_1, an upper control limit for Y_2 and an upper control limit for $\mathbf{u}_4'\mathbf{Y} = Y_1 + Y_2$. The weights indicate that detecting a change in Y_1, Y_2 and in the direction $\mathbf{u}_4$ are equally important. Given overall error rate α, the overall control region CR, and the rejection regions $RR(\mathbf{u}_i)$, $1 \le i \le 4$, for this situation are illustrated in Figure 1. The coverage probabilities are such that $P(CR) = 1 - \alpha$, and $2 \cdot P(RR(\mathbf{u}_1)) = 2 \cdot P(RR(\mathbf{u}_2)) = P(RR(\mathbf{u}_3)) = P(RR(\mathbf{u}_4))$.

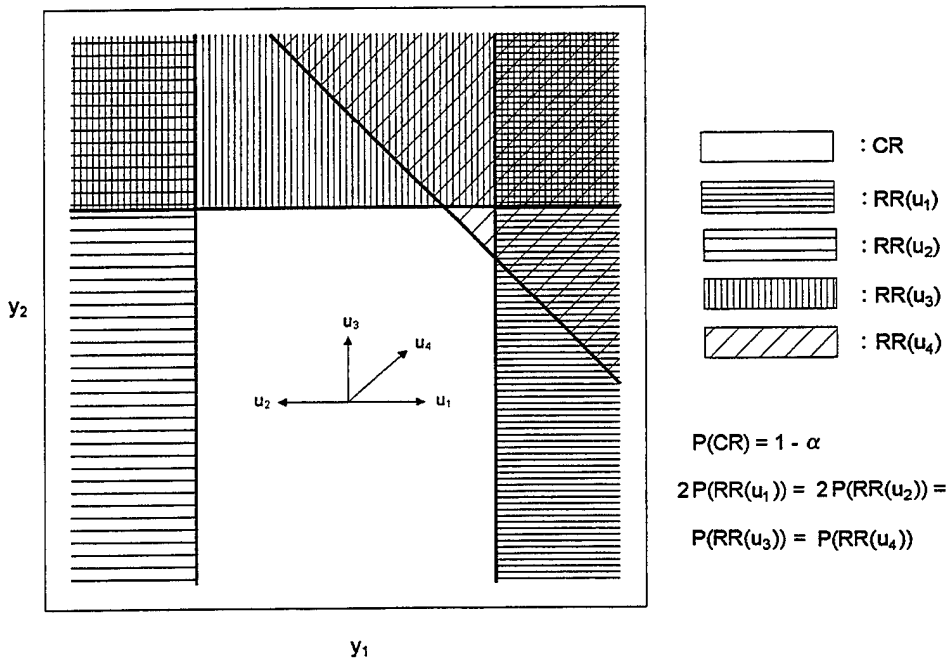


Figure 1. Control and Rejection Regions for a Bivariate Situation

Finally, we provide without proof Theorem 1, which guarantees the existence and uniqueness of directional and overall control regions under the assumptions given above. We first provide Lemma 1, which is a result proved by Beran [11]. It is used to prove Theorem 1.

Lemma 1. Let $\mathbf{u}$ be a direction and F_u be the cdf of $\mathbf{u}'\mathbf{Y}$, $\mathbf{Y} \sim F$. If the assumptions 1 and 2 hold for F, then F_u is continuous and strictly monotone increasing in $\mathbf{u}'\mathbf{Y}$.

Theorem 1. Consider prescribed directions $\mathbf{u}_i$, $1 \le i \le k$. For any choice of α, $0 < \alpha < 1$, there exist a unique overall control region CR and k unique directional control regions $CR(\mathbf{u}_i)$, $1 \le i \le k$, such that $P(CR) = 1 - \alpha$, and $P(CR(\mathbf{u}_i)) = 1 - w(\mathbf{u}_i)\gamma$, where γ is determined by α and $0 < \gamma < 1$.

HOTELLING'S T^2 CHARTS AS A SPECIAL CASE OF DIRECTIONAL SPC

In many multivariate SPC applications, the underlying distribution F is assumed to be multivariate normal with a mean vector μ and covariance matrix Σ, where μ and Σ are unknown and estimated from a reference sample S_n. In such applications the most commonly used control regions are based on Hotelling's T^2 statistic. An ellipsoidal control region based on this statistic with an estimated $1 - \alpha$ coverage probability is defined as

$$CR_n = \{\mathbf{y} : (\mathbf{y} - \hat{\mu})' \hat{\Sigma}^{-1} (\mathbf{y} - \hat{\mu}) \le d_n(\alpha)\}, \tag{5}$$

where $d_n(\alpha)$ is the αth quantile of the $F(m, n-m)$ distribution, with m and $n-m$ degrees of freedom, multiplied by a factor $(1+1/n)m(n-1)/(n-m)$, and $\hat{\mu}$ and $\hat{\Sigma}$ are the sample mean vector and the sample covariance matrix. Tracy et al. [12] can be seen for a detailed discussion on Hotelling's T^2 charts.

A sphere can be defined by a set of uncountably many tangent hyper-planes, each with a corresponding normal $\mathbf{u}$. Define the set $U = \{\mathbf{u}: \mathbf{u} \in \mathbf{R}^m, \ |\mathbf{u}| = 1\}$. Beran [11] used this concept to show that CR_n in (5) is equivalent to the intersection of uncountably many directional control regions

$$CR_n(\mathbf{u}) = \{\mathbf{y}: \mathbf{u}'\mathbf{y} \le \mathbf{u}'\hat{\mu} + s_{n,u} d_n^{1/2}(\alpha), \ \mathbf{u} \in U\} \tag{6}$$

where $s_{n,u} \ge 0$ and $s_{n,u}^2 = \mathbf{u}'\hat{\Sigma}\mathbf{u}$. The control regions $CR_n(\mathbf{u})$ have equal weights, because the estimated coverage probability of each $CR_n(\mathbf{u})$ is $\Phi\left\{\left[\chi_m^{-1}(\alpha)\right]^{1/2}\right\}$, where Φ denote the standard normal cdf, and χ_m denote the cdf of the chi-squared distribution with m degrees of freedom. Given CR_n, $CR_n(\mathbf{u})$ can be produced by finding the supporting hyper-plane of CR_n with a normal vector $\mathbf{u}$; $CR_n(\mathbf{u})$ is the associated half-space containing CR_n.

Many authors expressed the difficulty in interpreting out-of-control signals from SPC charts using Hotelling's T^2 statistic [13]. This inherent difficulty can be explained through the equivalence of (5) and (6); constructing a control region based on Hotelling's T^2 statistic is in fact equivalent to specifying all possible directions of shift, characterized by the set U, and constructing equally weighted directional control regions for all these directions [11]. Let $\mathbf{w}$ be an observed value of the monitored vector $\mathbf{Y}$. If $\mathbf{w}$ is outside of CR_n then it is also outside of a set of directional control regions $W = \{CR_n(\mathbf{u}): \mathbf{w} \notin CR_n(\mathbf{u})\}$. Let $U_w \subset U$ denote the corresponding set of directions: $U_w = \{\mathbf{u}: \mathbf{w} \notin CR_n(\mathbf{u}), \mathrm{u} \in U\}$. Then we conclude that the process has shifted in all the directions in U_w. For any choice of $\mathbf{w}$, U_w contains uncountably many directions and this gives rise to difficulty in interpreting what kind of process shifts $\mathbf{w}$ indicates.

CONSTRUCTION OF CONTROL REGIONS

We have so far seen the advantages of using finite number of directions in terms of interpretability. The question remains of how to construct corresponding control intervals in a nonparametric setting. In this section, we propose an algorithm to calculate the upper control limits $L(\mathbf{u}_i)$, or equivalently $CR(\mathbf{u}_i)$ from a sample S_n. The proposed algorithm works by changing control limits for each direction in incremental steps alternating from one direction to the next in a cyclic loop, hence it is named *the Alternating Direction algorithm* (*AD Algorithm*). The AD algorithm resembles a procedure proposed by Beran [11] when the weights $w(\mathbf{u}_i)$ are equivalent.

Let the overall and directional control regions estimated by the AD algorithm be ECR_n and $ECR_n(\mathbf{u}_i)$, and let $EL(\mathbf{u}_i)$ be the estimated control limit for the direction $\mathbf{u}_i$. Following the definitions (3) and (4), $ECR_n(\mathbf{u}_i) = \{ \mathbf{y}: \mathbf{u}_i'\mathbf{y} \le EL(\mathbf{u}_i)\}$ and $ECR_n = \{ \mathbf{y}: \mathbf{u}_i'\mathbf{y} \le EL(\mathbf{u}_i), \forall \mathbf{u}_i, 1 \le i \le k \}$. Now denote the actual coverage probabilities of $ECR_n(\mathbf{u}_i)$ and ECR_n under F by $1 - \hat{\alpha}(\mathbf{u}_i)$ and $1 - \hat{\alpha}$. Then

$$\begin{aligned} &P(ECR_n(\mathbf{u}_i)) = P(\mathbf{Y} \in ECR_n(\mathbf{u}_i)) = 1 - \hat{\alpha}(\mathbf{u}_i), \\ &P(ECR_n) = P(\mathbf{Y} \in ECR_n) = 1 - \hat{\alpha} \quad \text{when } \mathbf{Y} \sim F. \end{aligned} \tag{7}$$

We require that, for large reference samples, $\hat{\alpha}(\mathbf{u}_i) \cong \alpha(\mathbf{u}_i) = w(\mathbf{u}_i)\gamma$ and $\hat{\alpha} \cong \alpha$. This condition can be stated asymptotically as $\lim_{n\to\infty} \hat{\alpha}(\mathbf{u}_i) = w(\mathbf{u}_i)\gamma$ and $\lim_{n\to\infty} \hat{\alpha} = \alpha$. The AD algorithm gives asymptotically correct control regions in this sense as will be discussed after the algorithm is introduced.

Given the observations $\mathbf{Y}_s$, $1 \le s \le n$, in the reference sample S_n, order $T_{is} = \mathbf{u}_i'\mathbf{Y}_s$ from lowest to highest. Let $T_{i[l]}$ be the lth highest value of $\{ \mathbf{u}_i'\mathbf{Y}_s: 1 \le s \le n \}$. Then $T_{i[l]}$ is the lth order statistic of T_i. Also denote by $\mathbf{Y}(\mathbf{u}_i, T)$ the observation $\mathbf{Y}_s$ such that $\mathbf{u}_i'\mathbf{Y}_s = T$. Note that $\mathbf{Y}(\mathbf{u}_i, T)$ is defined only for those values of T such that there exists an observation $\mathbf{Y}_s$ satisfying $\mathbf{u}_i'\mathbf{Y}_s = T$. This notation is used in the description of the AD algorithm.

The algorithm assumes that α is a multiple of $1/n$ and the weights associated with directions are integer-valued, that is, $1 \le w(\mathbf{u}_i) \le r$ and $w(\mathbf{u}_i)$ is an integer $\forall \mathbf{u}_i$, $1 \le i \le k$. It searches for an overall control region containing $(1 - \alpha)n$ of the reference sample points by setting the control limits to order statistics $T_{i[n-j]}$ such that j is approximately equal to $cw(\mathbf{u}_i)$, where c is an integer coefficient, which is reduced sequentially until the target overall control region is found. This way the ratios between the marginal error rates, which are driven by $w(\mathbf{u}_i)$s, are empirically preserved.

Given a sample S_n, the *empirical probability* of a set A is defined as the number of points of S_n that are inside A divided by n [14]. In this respect, the algorithm searches for an overall control regions with $1 - \alpha$ empirical probability, and at the same time, balances the empirical probabilities of the directional rejection regions, the empirical marginal error rates, with respect to $w(\mathbf{u}_i)$'s.

In the following algorithm CL[i] is the control limit set for T_i, and INSIDE[s] indicates whether $\mathbf{Y}_s$ is inside the control region set by CL[i], $\forall i$, $1 \le i \le k$. INSIDE[s] is set in Step 0 of the algorithm, but it is not used in the following steps. It will be used when an improved version of the algorithm is introduced later in this section that does not require any change in Step 0 and Step 2.

The AD Algorithm

Step 0
Given S_n, order $T_i = \mathbf{u}_i'\mathbf{Y}$ to obtain order statistics $T_{i[l]}$, $1 \le l \le n$
Set CL[i] = $T_{i[n]}$ $\forall i$, $1 \le i \le k$
Set INSIDE[s] = 1 $\forall s$, $1 \le s \le n$
Set counters $c = 0$, $p = n$
Set Terminate = FALSE
Step 1
Repeat{ $c = c + 1$
Set $i = 0$
Repeat{ $i = i + 1$
Set $j = 0$
Repeat{ $j = j + 1$
CL[i] = $T_{i[n-(c-1)w(u_i)-j]}$
$p = n$
For s from 1 to n {
if $\mathbf{u}_t'\mathbf{Y}_s >$ CL[t] $\exists\, t$, $1 \le t \le k$, then $p = p - 1$ }
if $p \le (1 - \alpha)n$, then Terminate = TRUE}
Until (Terminate = TRUE or $j = w(\mathbf{u}_i)$)
Until (Terminate = TRUE or $i = k$)}

Until (Terminate = TRUE)}
Step 2
$EL(\mathbf{u}_i) = \text{CL}[i],\ \forall \mathbf{u}_i,\ 1 \le i \le k$
$ECR_n(\mathbf{u}_i) = \{\ \mathbf{y} : \mathbf{u}_i'\mathbf{y} \le EL(\mathbf{u}_i)\ \},\quad \forall \mathbf{u}_i,\ 1 \le i \le k$
$ECR_n = \{\ \mathbf{y} : \mathbf{u}_i'\mathbf{y} \le EL(\mathbf{u}_i),\ \forall \mathbf{u}_i,\ 1 \le i \le k\ \}$

The logic of the algorithm can be explained as follows. The algorithm initially sets the control limits CL[i] to the highest order statistics $T_{i[n]}$. Then all CL[i] are sequentially lowered to the order statistics $T_{i[n-w(u_i)]}$. This constitutes the outermost repeat loop in the algorithm. This is accomplished in the following fashion: For each T_i, CL[i] is lowered one order statistic at a time in the innermost repeat loop, that is, from $T_{i[n]}$ to $T_{i[n-1]}$, from $T_{i[n-1]}$ to $T_{i[n-2]}$, etc. At each time CL[i] is lowered one order statistic, the number of points that fall inside the newly established overall control region is computed. When the outermost loop completes one full iteration, it starts again from CL[i] = $T_{i[n-w(u_i)]}$ at the beginning of the next iteration, and sequentially lowers CL[i] to $T_{i[n-2w(u_i)]}$. The algorithm stops whenever there is $(1\text{-}\alpha)n$ or fewer points inside the overall control region.

The following theorems establish some characteristics of the AD algorithm. Theorem 2 ensures that the algorithm starts with an initial overall control region containing all the points in S_n. Theorem 3 explains the changes in the overall control region when the innermost repeat loop completes one iteration. The results in Theorem 3 will be used later to motivate a new version of the algorithm with improved efficiency. Finally, Theorem 4 formally states the bounds of the errors incurred by the algorithm in finding an overall control region with $(1\text{-}\alpha)n$ points and in balancing the ratios between empirical marginal error rates.

Theorem2. When initially CL[i] = $T_{i[n]}$ $\forall i,\ 1 \le i \le k$, all points in S_n are in the overall control region ECR_n defined by CL[i].

Proof of Theorem2. Suppose not, that is, suppose that $\exists\ \mathbf{Y}_t \in S_n \ni \mathbf{Y}_t \notin ECR_n$. Then $\exists\ \mathbf{u}_j,\ 1 \le j \le k,\ \ni \mathbf{Y}_t \notin ECR_n(\mathbf{u}_j)$, or equivalently $\mathbf{u}_j'\mathbf{Y}_t > \text{CL}[j] = T_{j[n]}$. On the other hand, $T_{j[n]} = \max(\mathbf{u}_j'\mathbf{Y}_s),\ \forall \mathbf{Y}_s \in S_n$, by definition, and the inequality $\mathbf{u}_j'\mathbf{Y}_t > T_{j[n]}$ contradicts with this fact. □

Theorem 3. Let ECR_n^{I} and ECR_n^{II} denote the control regions before and after the innermost repeat loop of the algorithm lowers CL[i] from $T_{i[l]}$ to $T_{i[l-1]}$. Let also $\mathbf{Y}_s^{I} = \mathbf{Y}(\mathbf{u}_i, T_{i[l]})$, $\mathbf{Y}_s^{II} = \mathbf{Y}(\mathbf{u}_i, T_{i[l-1]})$, and $\mathbf{Y}_t \in S_n \ni \mathbf{Y}_t \ne \mathbf{Y}_s^{I}$ and $\mathbf{Y}_t \ne \mathbf{Y}_s^{II}$.

(i) $\mathbf{Y}_t \in ECR_n^{I}$ if and only if $\mathbf{Y}_t \in ECR_n^{II}$.

(ii) The difference between the number of points that fall in ECR_n^{I} and ECR_n^{II} can be 0 or 1.

Proof of Theorem 3 - part (i). Consider the directional control regions $ECR_n^{I}(\mathbf{u}_i)$ and $ECR_n^{II}(\mathbf{u}_i)$ defined by the control limits $T_{i[l]}$ and $T_{i[l-1]}$: $ECR_n^{I}(\mathbf{u}_i) = \{\mathbf{y}: \mathbf{u}_i'\mathbf{y} \le T_{i[l]}\}$ and $ECR_n^{II}(\mathbf{u}_i) = \{\mathbf{y}: \mathbf{u}_i'\mathbf{y} \le T_{i[l-1]}\}$. The points $\mathbf{Y}_s^{I}$ and $\mathbf{Y}_s^{II}$ satisfy the inequalities in the definition of these regions as equalities: $\mathbf{u}_i'\mathbf{Y}_s^{I} = T_{i[l]}$ and $\mathbf{u}_i'\mathbf{Y}_s^{II} = T_{i[l-1]}$. Therefore the hyper-planes that define $ECR_n^{I}(\mathbf{u}_i)$ and $ECR_n^{II}(\mathbf{u}_i)$ pass through $\mathbf{Y}_s^{I}$ and $\mathbf{Y}_s^{II}$ respectively. Since $\mathbf{u}_i'\mathbf{Y}_s^{I}$ and $\mathbf{u}_i'\mathbf{Y}_s^{II}$ are two successive order statistics of T_i, there are no points of S_n in the region between these two hyper-planes, that is, no $\mathbf{Y}_s \in S_n$ satisfies the inequality $\{\mathbf{y} : T_{i[l-1]} < \mathbf{u}_i'\mathbf{y} < T_{i[l]}\}$. This implies that when the control limits for $\mathbf{u}_i$ is lowered from $T_{i[l]}$ to $T_{i[l-1]}$, $\mathbf{Y}_s^{I}$ and $\mathbf{Y}_s^{II}$ are the only points of S_n that might be in ECR_n^{I} but not in ECR_n^{II}, or vice versa. Consequently, either $\mathbf{Y}_t \in ECR_n^{I}$ and $\mathbf{Y}_t \in ECR_n^{II}$, or $\mathbf{Y}_t \notin ECR_n^{I}$ and $\mathbf{Y}_t \notin ECR_n^{II}$, which is summarized as an if-and-only-if statement in the theorem.

Proof of Theorem 3 - part (ii). First note the following relationships that derive directly from the definitions of $\mathbf{Y}_s^{I}$, $\mathbf{Y}_s^{II}$, $ECR_n^{I}(\mathbf{u}_i)$, and $ECR_n^{II}(\mathbf{u}_i)$: $ECR_n^{II} \subset ECR_n^{I}$, $ECR_n^{II}(\mathbf{u}_i) \subset ECR_n^{I}(\mathbf{u}_i)$, $\mathbf{Y}_s^{I} \in ECR_n^{I}(\mathbf{u}_i)$, $\mathbf{Y}_s^{II} \in ECR_n^{II}(\mathbf{u}_i)$, $\mathbf{Y}_s^{I} \notin ECR_n^{II}(\mathbf{u}_i)$, and $\mathbf{Y}_s^{II} \in ECR_n^{I}(\mathbf{u}_i)$. Now consider all the possible cases about $\mathbf{Y}_s^{I}$ and $\mathbf{Y}_s^{II}$'s being inside or outside of ECR_n^{I} and ECR_n^{II}:

Case 1: $\mathbf{Y}_s^{I} \in ECR_n^{I}$ Case 2: $\mathbf{Y}_s^{I} \notin ECR_n^{I}$ Case 3: $\mathbf{Y}_s^{I} \in ECR_n^{II}$ Case 4: $\mathbf{Y}_s^{I} \notin ECR_n^{II}$
Case 5: $\mathbf{Y}_s^{II} \in ECR_n^{I}$ Case 6: $\mathbf{Y}_s^{II} \notin ECR_n^{I}$ Case 7: $\mathbf{Y}_s^{II} \in ECR_n^{II}$ Case 8: $\mathbf{Y}_s^{II} \notin ECR_n^{II}$

Among these cases, Case 3 cannot happen since $\mathbf{Y}_s^{I} \notin ECR_n^{II}(\mathbf{u}_i)$, thus $\mathbf{Y}_s^{I} \notin ECR_n^{II}$. This implies that Case 4 always holds. Moreover, since $ECR_n^{II} \subset ECR_n^{I}$, Case 7 implies Case 5 and Case 6 implies Case 8. Therefore it suffices to consider cases 1, 2, 6 and 7. These cases can occur in the following four combined situations:

Situation a: Case 1 & Case 6 Situation b: Case 1 & Case 7
Situation c: Case 2 & Case 6 Situation d: Case 2 & Case 7

Let $\Delta(j)$ denote the difference between the number of points that fall in ECR_n^I and ECR_n^{II} when the situation j happens, j = a, b, c, d. Considering the part (i) of the theorem, it is easy to see that $\Delta(a) = 1$, $\Delta(b) = 1$, $\Delta(c) = 0$, and $\Delta(d) = 0$.□

Theorem 4 Let $EP(ECR_n)$ and $EP(ECR_n(\mathbf{u}_i))$ denote the empirical probabilities of the control regions ECR_n and $ECR_n(\mathbf{u}_i)$ that the algorithm produces.

(i) $EP(ECR_n) = 1 - \alpha$.

(ii) There exists a constant γ such that $EP(ECR_n(\mathbf{u}_i)) = 1 - (\gamma w(\mathbf{u}_i) + \Delta_i)$ where $0 \le \Delta_i \le w(\mathbf{u}_i)/n$, $\forall i$, $1 \le i \le k$.

Proof of Theorem 4 - part (i). This proof directly follows from part (ii) of Theorem 3. The algorithm computes the number points that fall inside ECR_n at each iteration of the innermost repeat loop, and Theorem 3 states that the number of points inside ECR_n may decrease by at most one point at each of these iterations. Therefore the algorithm always stops with $(1 - \alpha)n$ points inside ECR_n. Dividing this value by n gives the result in the theorem.

Proof of Theorem 4 - part (ii). Suppose that the algorithm stops at some point of the innermost repeat loop when $i = i'$, $j = j'$, and $c = c'$, where $1 \le i' \le k$ and $1 \le j' \le w(\mathbf{u}_i)$. Then $CL[i] = T_{i[n-c'w(u_i)]}$ if $1 \le i < i'$, $CL[i] = T_{i[n-(c'-1)w(u_i)-j']}$ if $i = i'$, and $CL[i] = T_{i[n-(c'-1)w(u_i)]}$ if $i' < i \le k$. Then $ECR_n(\mathbf{u}_i)$ contains $n - c'w(\mathbf{u}_i)$ points if $1 \le i < i'$, $n - (c' - 1)w(\mathbf{u}_i) - j'$ points if $i = i'$, and $n - (c' - 1)w(\mathbf{u}_i)$ points if $i' < i \le k$. Dividing these values for the three different cases by n, and letting $\gamma = (c'-1)/n$ produces the desired result. □

Theorem 4 shows that the overall control region produced by the proposed algorithm has an exact empirical overall error rate α, hence, the accuracy is very high. On the other hand, the empirical marginal error rates are equal to $\gamma w(\mathbf{u}_i) + \Delta_i$ such that $0 \le \Delta_i \le w(\mathbf{u}_i)/n$, hence the error is $w(\mathbf{u}_i)/n$, which can be quite high for high values of $w(\mathbf{u}_i)$. As a result, the required ratio between the empirical error rates may be significantly distorted. Nevertheless, reasonable choices such as 1, 2, or 3 for $w(\mathbf{u}_i)$ produce quite a good accuracy.

A drawback of the AD algorithm is that it processes all the points in S_n to compute the number of points inside the overall control region at each iteration of the innermost repeat loop. However, part (i) of Theorem 3 indicates that the only points to be kept track of are $\mathbf{Y}_s^I$ and $\mathbf{Y}_s^{II}$, in the notation of the theorem. The proposed algorithm can be improved to take advantage of this property. The improved version involves changes only in the innermost repeat loop of the algorithm:

Step 1 – Innermost Repeat-Loop Improved

Repeat{ $j = j + 1$
$\mathbf{Y}_s = \mathbf{Y}(\mathbf{u}_i, CL[i])$
if INSIDE[s] = 1, then {INSIDE[s] = 0, $p = p - 1$}
$CL[i] = T_{i[n-(c-1)w(u_i)-j]}$
$\mathbf{Y}_s = \mathbf{Y}(\mathbf{u}_i, CL[i])$
if $\mathbf{u}_t'\mathbf{Y}_s > CL[t]$ $\exists$ t, $1 \le t \le k$, then INSIDE[s] = 0
if $p \le (1 - \alpha)n$, then Terminate = TRUE}
Until (Terminate = TRUE or $j = w(\mathbf{u}_i)$)

Now, we show the consistency of the AD algorithm by examining the asymptotic behavior of the coverage probabilities of the generated control regions. As discussed at the beginning of this section, the AD algorithm can be considered consistent if $\lim_{n\to\infty} \hat{\alpha}(\mathbf{u}_i) = w(\mathbf{u}_i)\gamma$ and $\lim_{n\to\infty} \hat{\alpha} = \alpha$, where $\hat{\alpha}$ and $\hat{\alpha}(\mathbf{u}_i)$ are as defined in (7). By Theorem 4, $\lim_{n\to\infty} EP(ECR_n) = 1 - \alpha$ and $\lim_{n\to\infty} EP(ECR_n(\mathbf{u}_i)) = 1 - w(\mathbf{u}_i)\gamma$. However, the empirical distribution function F_n, which is constructed from a reference sample of size n, converges to F with probability 1 [15] as $n \to \infty$. Therefore, the empirical probabilities, which are computed under F_n, are asymptotically equivalent to actual probabilities, which are computed under F, as $n \to \infty$: $\lim_{n\to\infty} EP(ECR_n) = P(ECR_n) = 1 - \alpha$ and $\lim_{n\to\infty}$

$EP(ECR_n(\mathbf{u}_i)) = P(ECR_n(\mathbf{u}_i)) = 1 - w(\mathbf{u}_i)\gamma$. Since $P(ECR_n) = 1 - \hat{\alpha}$ and $P(ECR_n(\mathbf{u}_i)) = 1 - \hat{\alpha}(\mathbf{u}_i)$, $\lim_{n\to\infty} \hat{\alpha} = \alpha$ and $\lim_{n\to\infty} \hat{\alpha}(u_i) = w(\mathbf{u}_i)\gamma$.

APPLICATION TO A NONNORMAL SITUATION

This section includes application of the AD algorithm to a bivariate reference sample taken from a nonnormal distribution. A sample of size 200 is generated by the *Johnson Translation System* following the formulas provided by Johnson [16]. The generated random vector $\mathbf{Y} = (Y_1, Y_2)'$ has left-skewed components with means 3 and 5, and standard deviations 0.1 and 0.2, and correlation 0.6594.

First consider setting lower and upper control limits for Y_1 and Y_2. This form of multivariate control regions is advocated by Hayter and Tsui [5] since it clearly indicates which variable went out-of-control in an out-of-control situation. Setting lower and upper control limits requires specifying four directions $\mathbf{u}_1 = (1, 0)'$, $\mathbf{u}_2 = (-1, 0)'$, $\mathbf{u}_3 = (0, 1)'$, and $\mathbf{u}_4 = (0, -1)'$; also consider equal weights for these directions, $w(\mathbf{u}_i) = 1\ \forall i,\ 1 \le i \le 4$, and an overall error rate $\alpha = 0.1$.

Figure 2 includes a plot of the reference sample with a drawing of the rectangular overall control region produced by the AD algorithm. The control limits pass through the points that are indicated with a circle. Now suppose that a particular assignable cause increases both Y_1 and Y_2, therefore, it is desired to monitor process shifts that are due to this cause as well as monitoring shifts in Y_1 and Y_2. The direction $\mathbf{u}_5 = (1, 1)'$ is selected to monitor such shifts and a high weight, $w(\mathbf{u}_5) = 3$, is assigned to $\mathbf{u}_5$ due to frequent shifts in this direction. Figure 3 illustrates the pentagonal overall control region produced by the algorithm. Also directional rejection regions, $ERR(\mathbf{u}_i)$ are indicated in this figure.

It can be noticed in Figure 3 that there are 20 points outside the overall control region, and there are 5, 5, 4, 4, and 12 points in $ERR(\mathbf{u}_1)$, $ERR(\mathbf{u}_2)$, $ERR(\mathbf{u}_3)$, $ERR(\mathbf{u}_4)$ and $ERR(\mathbf{u}_5)$ respectively; hence, the algorithm attempts to balance the number of points with respect to the weights. Note that these numbers agree with the results in Theorem 4. As a different scenario, specifying a lower weight $w(\mathbf{u}_5) = 1$ would lead to an $ERR(\mathbf{u}_5)$ whose boundary line would not intersect with the rectangular control region that would be formed for $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ and $\mathbf{u}_4$. In this case, the rectangular control region would be equal to the one in Figure 2; therefore, adding $\mathbf{u}_5$ would not have any impact on the existing control region.

Now we compare the overall control regions in Figure 2 and Figure 3; Figure 4 superimposes these two regions. Clearly, adding $\mathbf{u}_5$ decreases the directional sensitivity of the first four directions. On the other hand, the pentagonal region would declare observations that fall inside the region B in Figure 4 as out of control but the rectangular region would not.

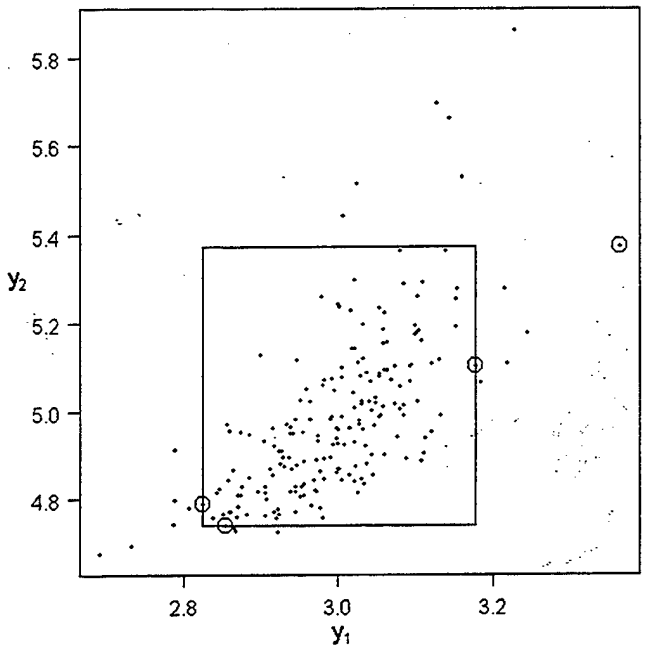


Figure 2. Overall Control Region for directions u_1, u_2, u_3, *and* u_4

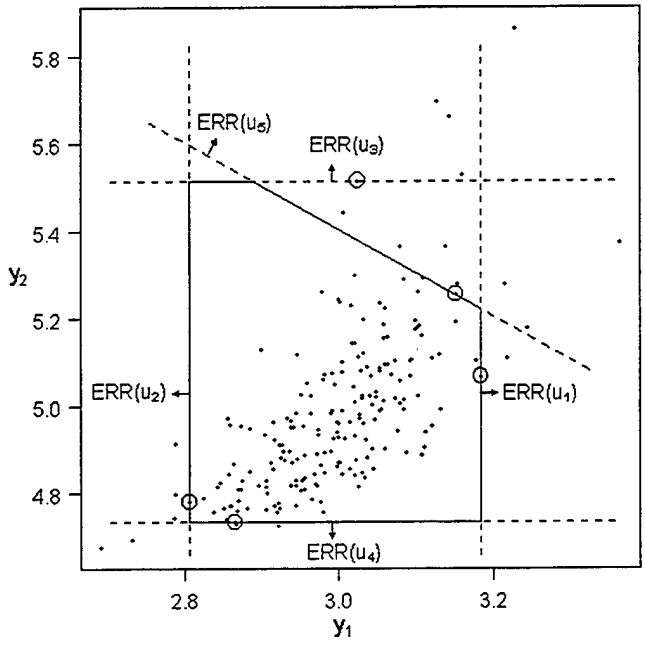


Figure 3. Overall Control Region for directions u_1, u_2, u_3, u_4, *and* u_5

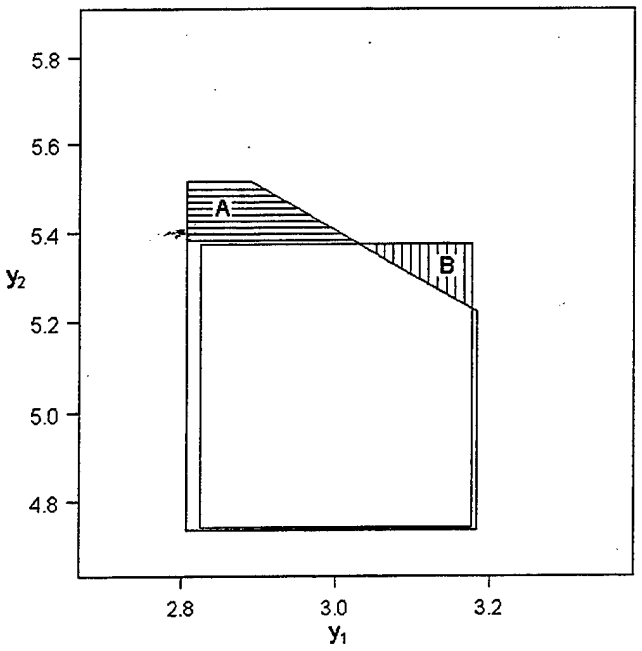


Figure 4. Comparison of the Controls Regions in Figures 3 and 4

APPLICATION TO PARAMETRIC SITUATIONS

The AD algorithm can be applied to situations when the distribution F is known by simulating data from F. For instance, a common practice in multivariate SPC is to assume multivariate normality and estimate the distribution parameters from a reference sample. If historical evidence and the reference sample confirm the assumption of normality, then it is better to use the fitted normal distribution for construction of control regions rather than to use the reference sample directly. The fitted distribution can be used to simulate a large sample in this case, and the AD algorithm can be applied to this sample.

Directional SPC proposes selecting directions along hypothesized process shifts, however these directions do not provide the greatest power under normality assumption. Assuming F is $N(\mu, \Sigma)$ where the distribution parameters μ and Σ are known, Healy [6] showed that the logarithm of the likelihood-ratio test statistic for testing the single hypothesis H_0: $\mu = \mu_0$ vs. H_a: $\mu = \mu_0 + \lambda\mathbf{u}$, where $\lambda > 0$ is a given constant, is given by

$$Z = \lambda\mathbf{u}'\Sigma^{-1}\mathbf{Y} - \lambda\mathbf{u}'\Sigma^{-1}\mu_0 - .5\lambda\mathbf{u}'\Sigma^{-1}\lambda\mathbf{u}. \qquad (8)$$

Since Z depends only on $Z^* = \mathbf{u}'\Sigma^{-1}\mathbf{Y}$ and λ, and is a nondecreasing function of Z^* for $\lambda > 0$, a one-sided test using Z^* is the uniformly most powerful test for testing H_0: $\mu = \mu_0$ vs. H_a: $\mu = \mu_0 + \lambda\mathbf{u}$, for any $\lambda > 0$. Therefore, one might be interested in monitoring $Z_i^* = \mathbf{u}_i'\Sigma^{-1}\mathbf{Y}$ instead of T_i for increased power: the statistics Z_i^* can be easily implemented in directional SPC by replacing the monitored directions $\mathbf{u}_i$ by the directions $\Sigma^{-1}\mathbf{u}_i$.

On the other hand, Z^* cannot be used when Σ is ill-conditioned, thus Σ^{-1} cannot be computed. This might be the situation if a process is driven by the probability model $\mathbf{Y} = \mathbf{Az} + \varepsilon$, where the columns of $\mathbf{A}$ correspond to shift directions due to assignable causes, and ε is the common-cause variability in the process. Therefore, components of the $\mathbf{z}$ vector represent the level of shifts in each direction. This model was introduced by Barton and Gonzalez-Barreto [9]. When the variability in $\mathbf{Az}$ is much larger than the variability in ε, and $\mathbf{A}$ has fewer columns than rows, the

underlying multivariate normal distribution does not have the full support of $\mathbf{R}^m$, hence Σ becomes ill-conditioned. In contrast, the statistic T can be safely used in such situations because it does not involve computing Σ^{-1}.

Another disadvantage of using Z^* could be the interpretation of the projections onto the directions $\Sigma^{-1}\mathbf{u}_i$ from an engineering viewpoint. While checking projections of $\mathbf{Y}$ onto the suspected shift directions $\mathbf{u}_i$ make sense to an engineer, the idea of projections onto $\Sigma^{-1}\mathbf{u}_i$ might not, particularly when the resulting overall control region as in Figure 1 is presented to an engineer.

For known underlying distributions that are nonnormal, the uniformly most powerful tests may not exist, or may be very difficult to compute while tests based on $T = \mathbf{u}'\mathbf{Y}$ remain as easily applicable.

CHOICE OF REFERENCE SAMPLE SIZE

The AD algorithm is based on empirical coverage probabilities, and the true coverage probabilities of the control regions produced by the algorithm get closer to the desired ones as the sample size increases. Therefore, using relatively small samples might lead to overall control regions with coverage probabilities that are significantly different from the target level 1 - α. This might in turn cause deviations from the target ARL of a directional SPC scheme.

While an overall error rate $\alpha = 0.1$ is used previously for illustration, the target values are usually much lower in industry, for instance a traditional choice has been $\alpha = 0.0027$. Such low α levels restrict the application of the AD algorithm since the algorithm requires that α should be a multiple of 1/(sample size). Considering the sample of size 200, even a choice of $\alpha = 0.005$ would leave only one point outside the overall control region setting the control limits mostly to the highest order statistics of T_is. In general, small samples will lead to situations that will set the control limits to order statistics with very high orders.

This situation has two disadvantages. First, the sample might have outliers; in this case the control limits might be set to the outlier points. Second disadvantage is related to the distribution of the order statistics of T_is. The variance of the rth order statistic, $T_{i[r]}$, where $1 \le r \le n$, is $\text{var}(T_{i[r]}) = r(n - r + 1)/(n+1)^2(n+2)$, which is a distribution-free result [17]. This is a quadratic function in r which attains its minimum at $r = n/2$ when considered continuous in r. Therefore, the variability in $T_{i[r]}$ increases as r increases for $r > n/2$. In other words, the uncertainty associated with the control limits increases as higher order statistics are used as control limits. An immediate consequence is an increased uncertainty in the coverage probabilities of the control regions. While the uncertainty in the coverage probability can be evaluated using results from the theory of order statistics for univariate SPC [1], no immediate result is available to quantify the uncertainty of the coverage probability of an overall control region when the control limits are set to $L(\mathbf{u}_i) = T_{i[u_i]}$, which is the case in the AD algorithm.

In conclusion, a reference sample should be formed by collecting as many sample points from a process as possible, hence, maximizing the sample size. As an ad-hoc method, a sample size that sets the control limits for each T_i to the order statistics that are lower than the highest ones can be considered large enough. Many modern production processes allow collecting large reference samples in short periods of time by the help of automated quality inspection systems using CMM machines or vision systems. For instance, consider die alignments in wafers for inkjet printer production. At the in-process level, every die for every wafer (300 dies/wafer) is inspected for alignment, and a daily production rate of 500 wafers experienced in certain industries allow collection of 500 multivariate reference sample points in a day.

CONCLUSIONS

Directional SPC has two major advantages as compared with the current multivariate SPC techniques. The significant shift directions are immediately available when an out-of-control signal is received, because each hypothesis in (1) is linked with a direction; when the directions are associated with assignable causes, significant directions also provide diagnostic information. Although not proved here, we have also seen that higher sensitivities are achieved when smaller numbers of directions are monitored. Limiting the number of monitored directions to a minimum number provides a better ARL performance when the process is out of control, since shorter control intervals for T_i will result in quicker detection of out-of-control situations. Hotelling's T^2 charts are considered as the

worst case since they correspond to monitoring uncountably many directions, and out-of-control signals are difficult to interpret.

Most of the current multivariate SPC tools are based on normality assumption. Directional SPC, on the other hand, can be applied to a reference sample in a nonparametric fashion. Based on simulation studies, Westfall and Young [10] commented that the effect of nonnormality is amplified in multiple hypothesis testing applications. Therefore subgrouping with a limited sample size such as 4 or 5 can still lead to simultaneous control regions with a significantly wrong overall coverage probability. A better approach is to use a reference sample, or fit a nonnormal multivariate distribution family to the reference sample. Directional SPC can be applied in both of these cases.

ACKNOWLEDGMENTS

Part of this Research was supported by grants from IBM, the Turkish government and by NSF Grant DMI-9700330.

REFERENCES

1. Willemain, T. R., and G. C. Runger. "Designing Control Charts Using an Empirical Reference Distribution," Journal of Quality Technology, 28, p.31, 1996.

2. Tracy, N. D., Young, I. C., and R. L. Mason. "Multivariate Control Charts for Individual Observations," Journal of Quality Technology, 24, p.88, 1992.

3. Seppela, T., Moskowitz, H., Plante, R., and J. Tang. "Statistical Process Control via Subgroup Bootstrap," Journal of Quality Technology, 27, p.139, 1995.

4. Liu, R. Y. "Control Charts for Multivariate Processes," Journal of the American Statistical Association, 90, p.1380, 1995.

5. Hayter, A. J., and K. L. Tsui. "Identification and Quantification in Multivariate Quality Control Problems," Journal of Quality Technology, 26, p.197, 1994.

6. Healy, J. D. "A Note on Multivariate CUSUM Procedures," Technometrics, 29, p.409, 1987.

7. Hawkins, D. M. "Multivariate Quality Control Based on Regression-Adjusted Variables," Technometrics, 33, p.61, 1991.

8. Runger, G. C. "Projections and the U^2 Multivariate Control Chart," Journal of Quality Technology, 28, p.313, 1996.

9. Barton, R. R., and D. R. Gonzalez-Barreto. "Process Oriented Basis Representations for Multivariate Process Diagnostics," Quality Engineering, 9, p.107, 1996.

10. Westfall, P. H., and S. S. Young. Resampling Based Multiple Testing: Examples and Methods for p Value Adjustments, New York: John Wiley, 1993.

11. Beran, R. "Probability-Centered Prediction Regions". The Annals of Statistics, 21, p.1967, 1993.

12. Tracy, N. D., Young, I. C., and R. L. Mason. "Some Aspects of Hotelling's T^2 Statistic for Multivariate Quality Control," in Statistics of Quality. Ghosh, S, Schucany, W. R., Smith, W. B. (eds), New York: Marcel Dekker, p.77, 1997.

13. Lowry, C. A., and D. C. Montgomery. "A Review of Multivariate Control Charts," IIE Transactions, 27, p.800, 1995.

14. Efron, B. An Introduction to the Bootstrap, New York: Chapman & Hall, 1993.

15. Serfling, R. J. Approximation Theorems of Mathematical Statistic, New York: John Wiley, 1980.

16. Johnson, M., E. Multivariate Statistical Simulation, New York: John Wiley, 1987.

17. Gibbons, J. D., and S. Chakraborti. Nonparametric Statistical Inference (3rd ed.), New York: Marcel Decker, 1992.

INTENTIONALLY LEFT BLANK.

Some Graphics for Logistic Regression

Sanford Weisberg*
University of Minnesota
Department of Applied Statistics
St. Paul, Minnesota USA

Abstract

Plots of a response variable versus predictors are a standard starting point for regression analysis. In problems with a single predictor, this graph provides a complete summary of a regression, and with many predictors, Cook and Weisberg[3] show how graphs can be used as the basis of a complete analysis. When the response variable in binary, the standard plot of a response versus a predictor is much less useful, since the response can assume only one of two values. In this article, a number of useful graphs are described for binary regression problems.

Key Words: Logistic regression, graphical methods, transformations.

Introduction

Graphs can play an important role in regression analysis. In problems with a continuous response y and a single continuous predictor x, a graph like Figure 1a provides a reasonably complete summary of the regression. We can use the data points on the plot to obtain a visual lack of fit test for a model, such as the simple linear regression straight-line mean function shown on the plot. In this instance, we can judge that the mean function does not match the data; the mean increases with x, but not linearly. Similarly, in Figure 1a the variance appears to increase from left to right.

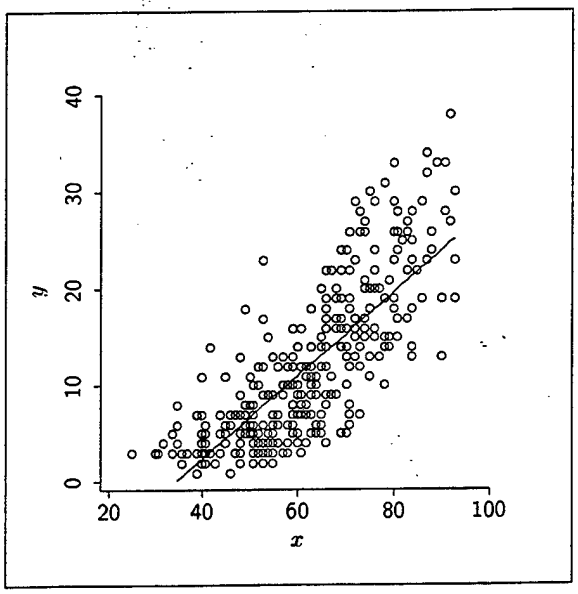
When the response is binary, by convention equal to either zero or one, the graph of the response versus the predictor is of much less value. A typical graph is shown as Figure 1b. The data fall onto two horizontal bands for $y = 0$ and $y = 1$, and because of this the points provide little information. For example, we cannot tell if the fitted logistic curve shown in Figure 1b is appropriate for these data or not.

In this article, we present some simple graphs that are better suited for use with a binary response. We first briefly review the logistic regression model, then discuss one-predictor problems, and finally generalize to many predictors with emphasis on the $p = 2$ predictor case. We conclude with a brief discussion. Throughout, we will use an example from Härdle and Stoker.[5] These data are the result of an experiment to study side-impact vehicle crashes using crash dummies. The response y is one if the effects on the crash dummy are judged to have been fatal, and zero otherwise. The predictors are the velocity *Vel* of the vehicle, the *Age* of the dummy, presumably determined by it size, and the acceleration *Acl* measured on the dummy's abdomen shortly before impact. The sample size is $n = 58$. Figure 1b is a plot of y versus *Age*.

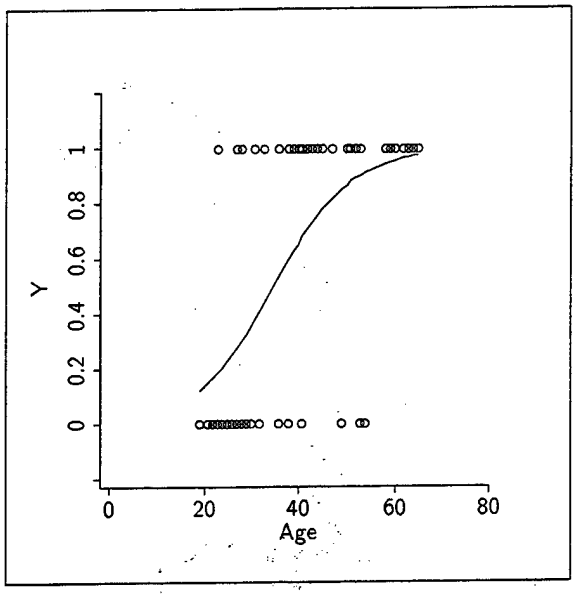
Review of Logistic Regression

With a binary response y and a p-dimensional predictor x, the conditional distribution of y given x, $F(Y|x)$, is a Bernoulli distribution, which is completely characterized by $\Pr(y = 1|x) = \mathrm{E}(y|x) = p(x)$. The

*Work supported by National Science Foundation grant DUE 9652887.



a. Continuous response.



b. Binary response.

Figure 1: Plots of response versus a continuous predictor. Smooth curves are estimated mean functions from (a) simple linear regression and (b) logistic regression.

regression problem is to determine $p(x)$ as x is varied. Logistic regression is based on two additional assumptions about $p(x)$. First, we assume that the dependence of $p(x)$ on x is only through some linear combination of $\beta' x$, so that

$$p(x) = M(\beta_0 + \beta' x) \tag{1}$$

where the function M is called a *kernel mean function* (Cook and Weisberg[3]), and is bounded on $[0, 1]$. Logistic regression specifies M to be the logistic function, $M(\beta' x) = (1+\exp[-(\beta_0+\beta' x)])^{-1}$. Equivalently by solving for $\beta_0 + \beta' x$,

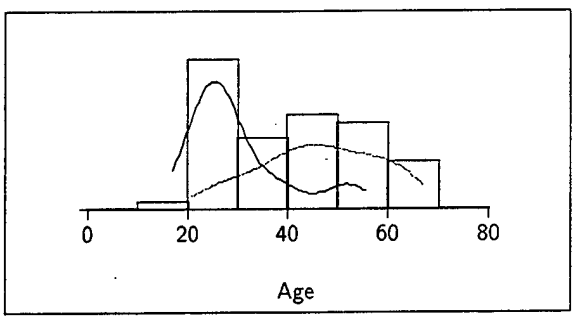
$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta' x \tag{2}$$

This equation says that the log-odds of success, called a logit, is equal to a linear combination of the predictors. Equation (2) gives the *link function* equivalent to the logistic kernel mean function.

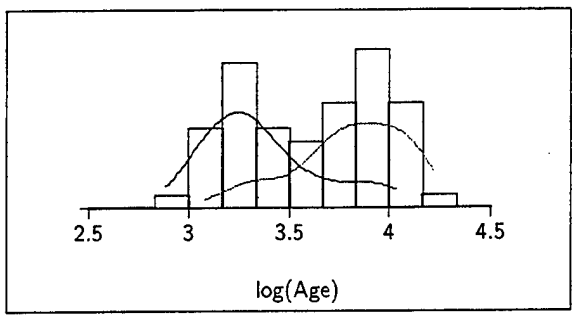
Given these assumptions, maximum likelihood can be used to get an estimate for (β_0, β), and for the mean function $p(x)$. Details are given, for example, in McCullagh and Nelder,[7] and most standard statistical packages provide software for fitting logistic regression. Of interest in building logistic regression models is whether or not the assumptions hold. In particular, the smooth curve in Figure 1b is the estimate of $p(x)$ assuming the logistic regression model and maximum likelihood estimation. Can we use graphs to tell if this function matches the data?

Log-density Ratio

Because y is discrete, we cannot really *see* in Figure 1b whether or not the fitted mean function is appropriate; the residuals from the fit don't help because of patterns in the residuals caused by the discrete response. However, the discrete response is an advantage if we examine the *inverse* problem $F(x|y)$ because we need consider only the two distributions $F(x|y = 0)$ and $F(x|y = 1)$. These can be studied using density estimates. For example, Figure 2a shows density estimates for *Age* for the collision data. The means



a. *Age*.



b. log(*Age*).

Figure 2: Density estimates of *Age* and log(*Age*) for the collision data. These histograms are easier to compare using color; here the dark lines are for the deaths and the light lines are for the survivors. Throughout this paper, densities are estimated using a Gaussian kernel density estimate with bandwidth equal to 0.9 of the normal reference rule.

$\mathrm{E}(x|y=0)$ and $\mathrm{E}(x|y=1)$ are well-separated, with most of the survivors corresponding to "younger" dummies. The density for deaths appear to have a larger variance as well. We can summarize by saying that the density $f(x|y=0)$ appears to be different from $f(x|y=1)$, or equivalently, the *log-density ratio* $h(x) = \log(f(x|y=1)/f(x|y=0))$ is not constant.

How does information about the log-density ratio translate into information about $p(x)$? Using Bayes' theorem, the log-density ratio is:

$$\begin{aligned} h(x) &= \log\left(\frac{f(x|y=1)}{f(x|y=0)}\right) \\ &= \log\left(\frac{\Pr(y=1|x)f(x)/\Pr(y=1)}{\Pr(y=0|x)f(x)/\Pr(y=0)}\right) \\ &= \log\left(\frac{p(x)}{1-p(x)}\right) - \log\left(\frac{\Pr(y=1)}{1-\Pr(y=1)}\right) \end{aligned}$$

The second term on the right side of the last equation is independent of x and is a constant d whose value depends on the sampling plan used to generate the data (in case-control studies, d is fixed by design, and in randomized studies d is the log-odds of success in the sampled population). Rearranging terms, we get

$$\log\left(\frac{p(x)}{1-p(x)}\right) = d + h(x) \tag{3}$$

If the log-density ratio $h(x)$ is known, then according to (3) the logistic regression model is appropriate, but the predictor is any non-zero constant times $h(x)$.

One Predictor

Of course $h(x)$ is generally unknown. When $p = 1$, we can estimate $h(x)$ using data, perhaps using the two density estimates as in Figure 2a directly. This was suggested by Silverman,[8] but directly estimating $h(x)$ in this way is very inefficient. If a nonparametric approach is to be used, ideas from generalized additive models (Green and Silverman[4]) provide a direct estimate of $h(x)$ via regression.

We describe here a parametric approach. Suppose that u is a $k \times 1$ vector of *terms*, such that each of the k elements of u is a function of x. For example, with a one-dimensional predictor x, u might consist of $(1, x, x^2, x^3)$, giving $k = 4$, or u might consist of $(1, \log(x))$ for $k = 2$. Suppose further that, at least

to a reasonable approximation, $h(x) = \eta' u$ for some η. Comparing (2) to (3), we now see that the *logistic regression model is appropriate, but with the predictor x replaced by the terms u.*

The next task is to determine u. Suppose for example that the densities $f(x|y = j), j = 0, 1$ are normal, $\mathrm{N}(\mu_j, \sigma_j^2)$. We can now calculate $h(x)$ exactly, and we find, Kay and Little,[6]

$$\log\left(\frac{p(x)}{1-p(x)}\right) = d_0 + \left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_0}{\sigma_0^2}\right)x + \frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)x^2 \tag{4}$$

where

$$d_0 = \log(d) + \log(\sigma_0/\sigma_1) + (\mu_0^2/2\sigma_0^2 - \mu_1^2/2\sigma_1^2)$$

is a constant that does not depend on x. From (4), we can read off $u = (1, x)$ if $\sigma_0 = \sigma_1$, but that otherwise we will need $u = (1, x, x^2)$.

The normal distribution provides a useful target for continuous predictors in binary regression. If the density of $x|y$ appears to be reasonably normal for each value of y, then the normality results can be applied to tell us the appropriate terms to use in the mean function. In Figure 2a the density estimates do not resemble normal distributions, and so the logistic fit shown in Figure 1b is not appropriate for these data. Perhaps, however, we can *induce* normality through transformation of the predictor $x = Age$. We can use the Box and Cox[1] procedure to find the transformation that makes the within-group distributions as close to normal as possible. For the data of Figure 2a, this leads to replacing Age by $\log(Age)$, giving the density estimates shown in Figure 2b. While normality is not certain here—both estimates are slightly asymmetric due to a small number of extreme points—normality with common variance is at least plausible. This suggests using logistic regression with terms $u = (1, \log(Age))$. This technique of transforming predictors toward normality is very useful, and often leads quickly to useful results.

Many Predictors

With $p > 1$, equation (3) continues to hold, except that now x is a vector rather than a scalar. Simple expressions are available for the log-density ratio in the multivariate case only for the multivariate normal (Kay and Little citekl), where the terms $(1, x)$ are required if the within-group covariance matrices are equal; if the covariance matrices are not equal, then quadratics and interactions may be required. This suggests a general goal: when all the predictors are continuous, seek transformations to make the within-group distributions as close to multivariate normal with common covariance matrix. This requires an extension of the Box and Cox procedure to multivariate data (Velilla[9]). The transformed predictors are then candidates for terms in the logistic mean function.

The choice of terms u is potentially more complex if some of the predictors are binary. To show all the possibilities, we consider the case of $p = 2$ predictors in detail. The goal is to obtain an expression for the log-odds ratio as a function of $f(x_1, x_2|y)$. We can write, with a slight abuse of notation,

$$\begin{aligned}
\log\left(\frac{\Pr(y=1|x_1,x_2)}{\Pr(y=0|x_1,x_2)}\right) &= \log\left(\frac{f(y=1,x_1,x_2)/f(x_1,x_2)}{f(y=0,x_1,x_2)/f(x_1,x_2)}\right) \\
&= \log\left(\frac{f(x_1|x_2,y=1)f(x_2|y=1)\Pr(y=1)}{f(x_1|x_2,y=0)f(x_2|y=0)\Pr(y=0)}\right) \\
&= \log\left(\frac{\Pr(y=1)}{\Pr(y=0)}\right) + \log\left(\frac{f(x_2|y=1)}{f(x_2|y=0)}\right) + \log\left(\frac{f(x_1|x_2,y=1)}{f(x_1|x_2,y=0)}\right) \qquad (5)
\end{aligned}$$

The first term in the last expression in (5) does not depend on the predictors, and is therefore unimportant. To understand the other terms, we assume that both $f(x_2|y)$ and $f(x_1|x_2, y)$ have exponential family distributions, written as (McCullagh and Nelder,[7] p. 28)

$$\begin{aligned}
f(x_2|y) &= \exp\{[x_2\theta(y) - b(\theta(y))]/\phi(y) + c(x_2, \phi(y))\} \\
f(x_1|x_2,y) &= \exp\{[x_1\theta_1(x_2,y) - b_1(\theta_1(x_2,y))]/\phi_1(x_2,y) + c_1(x_1,\phi_1(x_2,y))\}
\end{aligned}$$

Table 1: Values of c, b and the canonical parameter $\theta(z)$ for the distribution of $y|z$ for three standard exponential family distributions, from McCullagh and Nelder,[7] p. 30.

Distribution	b	c	$\theta(z)$
Normal	$\theta^2/2$	$-.5(y^2/\phi + \log(2\pi\phi))$	z
Bernoulli	$\log(1+\exp(\theta))$	1	$\log[p(z)/(1-p(z))]$
Poisson	$\exp(\theta)$	$\log(1/y!)$	$\log(z)$

$\theta(y)$ and $\theta_1(x_2, y)$ are the *canonical parameters* of the distributions, b and b_1 are called *cumulant functions*, c and c_1 are essentially normalizing constants, and ϕ and ϕ_1 are scale factors. Expressions for b, c and θ for the three standard distributions we use are shown in Table 1. Substituting these into the (5), we get seven terms:

$$\begin{aligned}
\log\left[\frac{\Pr(y=1|x_1,x_2)}{\Pr(y=0|x_1,x_2)}\right] &= \log\left[\frac{\Pr(y=1)}{\Pr(y=0)}\right] \\
&\quad + \left[\frac{c(x_2,\phi(1))}{c(x_2,\phi(0))}\right] && (6)\\
&\quad + x_2\left[\frac{\theta(1)}{\phi(1)} - \frac{\theta(0)}{\phi(0)}\right] && (7)\\
&\quad - \left[\frac{b(\theta(1))}{\phi(1)} - \frac{b(\theta(0))}{\phi(0)}\right] && (8)\\
&\quad + \left[\frac{c_1(x_1,\phi_1(x_2,1))}{c_1(x_1,\phi_1(x_2,0))}\right] && (9)\\
&\quad + x_1\left[\frac{\theta_1(x_2,1)}{\phi_1(x_2,1)} - \frac{\theta_1(x_2,0)}{\phi_1(x_2,0)}\right] && (10)\\
&\quad - \left[\frac{b_1(\theta_1(x_2,1))}{\phi_1(x_2,1)} - \frac{b_1(\theta_1(x_2,0))}{\phi_1(x_2,0)}\right] && (11)
\end{aligned}$$

We will now consider four important special cases.

Both Predictors Normal

Suppose first that $(x_1, x_2)|y$ is bivariate normal, possibly with different covariance matrices for each value of y. For constants $\alpha_1, \alpha_2, \alpha_{12}$, we can therefore write

$$\begin{aligned}
x_2|y &\sim \mathrm{N}(\mu_y, \phi(y)) \\
x_1|x_2, y &\sim \mathrm{N}\left(\alpha_0 + \alpha_1 y + \alpha_2 x_2 + \alpha_{12} y x_2, \phi_1(x_2, y)\right)
\end{aligned}$$

This form is completely general. If the two within-group covariance matrices are equal, then $\alpha_{12} = 0$ and $\phi_1(x_2, 0) = \phi_1(x_2, 1)$, and the regressions of x_1 on x_2 within-group are linear with constant variance. This can be checked graphically or using a test statistic.

Substituting from Table 1 into (6)–(11) gives:

$$\begin{aligned}
(6) &= \frac{1}{2}x_2^2\left(\frac{1}{\phi(0)} - \frac{1}{\phi(1)}\right) + \frac{1}{2}\log\left(\frac{\phi(0)}{\phi(1)}\right) \\
(7) &= x_2\left(\frac{\mu_1}{\phi(1)} - \frac{\mu_0}{\phi(0)}\right)
\end{aligned}$$

$$
\begin{aligned}
(8) &= -\frac{1}{2}\left(\frac{\mu_1^2}{\phi(1)} - \frac{\mu_0^2}{\phi(0)}\right) \\
(9) &= \frac{1}{2}x_1^2\left(\frac{1}{\phi_1(x_2,0)} - \frac{1}{\phi_1(x_2,1)}\right) + \frac{1}{2}\log\left(\frac{\phi_1(x_2,0)}{\phi_1(x_2,1)}\right) \\
(10) &= x_1\left(\frac{\alpha_0+\alpha_1+\alpha_2x_2+\alpha_{12}x_2}{\phi_1(x_2,1)} - \frac{\alpha_0+\alpha_2x_2}{\phi_1(x_2,0)}\right) \\
(11) &= \frac{1}{2}\left(\frac{(\alpha_0+\alpha_1+\alpha_2x_2+\alpha_{12}x_2)^2}{\phi_1(x_2,1)} - \frac{(\alpha_0+\alpha_2x_2)^2}{\phi_1(x_2,0)}\right)
\end{aligned}
$$

As previously suggested, if $\alpha_{12} = 0$, essentially equivalent to within-group covariance matrices equal, then the log-density ratio depends only on the terms $(1, x_1, x_2)$. The term x_1^2 must be added only if $\phi(0) \neq \phi(1)$. A quadratic in x_2 and possibly an x_1x_2 interaction may be required whenever $\alpha_{12} \neq 0$.

Normal/Binary Case

Suppose that $x_2|y$ is normal, and $x_1|x_2, y$ is binary. A desirable procedure would be (1) transform x_1 (or, in general, all the continuous predictors) for normality; and (2) then add the binary term(s), and possibly binary by predictor interactions to the mean function and begin examining logistic models. We examine if this procedure is justifiable. We have

$$
\begin{aligned}
x_2|y &\sim \mathrm{N}(\mu_y, \phi(y)) \\
x_1|x_2, y &\sim \text{Bernoulli}(p_1(x_2, y))
\end{aligned}
$$

where

$$
\log\left(\frac{p_1(x_2,y)}{1-p_1(x_2,y)}\right) = \alpha_0 + \alpha_1 y + \alpha_2 x_2 + \alpha_{12}x_2y
$$

Given this model, the six terms of interest are:

$$
\begin{aligned}
(6) &= \frac{1}{2}x_2^2\left(\frac{1}{\phi(0)} - \frac{1}{\phi(1)}\right) + \frac{1}{2}\log\left(\frac{\phi(0)}{\phi(1)}\right) \\
(7) &= x_2\left(\frac{\mu_1}{\phi(1)} - \frac{\mu_0}{\phi(0)}\right) \\
(8) &= -\frac{1}{2}\left(\frac{\mu_1^2}{\phi(1)} - \frac{\mu_0^2}{\phi(0)}\right) \\
(9) &= 0 \\
(10) &= x_1\left(\alpha_1 + \alpha_{12}x_2\right) \\
(11) &= -\log\left\{[1+\exp(\alpha_0+\alpha_1+(\alpha_2+\alpha_{12})x_2)]/[1+\exp(\alpha_0+\alpha_2x_2)]\right\}
\end{aligned}
$$

Terms (6)–(8) are as in the normal/normal case, requiring terms $(1, x_2)$ g (7) $\neq 0$, and x_2^2 if $\phi(0) \neq \phi(1)$. From term (10), x_1 is needed if $\alpha_1 \neq 0$, and x_1x_2 is needed if $\alpha_{12} \neq 0$. The condition $\alpha_{12} \neq 0$ implies that the regression functions for $x_2|x_1, y$ are not parallel in logit scale, which can be checked via graphs or with a test.

Term (11) is the problem term. As long as this term is approximately linear in x_2, then (11) does not introduce any additional terms in the logistic mean function. In general, (11) is not linear in x_2. If we assume that $\alpha_{12} = 0$, so no interaction is needed, we get

$$
(11) = \log\left\{[1+\exp(\alpha_0+\alpha_1+\alpha_2x_2)]/[1+\exp(\alpha_0+\alpha_2x_2)]\right\}
$$

For $\alpha_2 > 0$, this quantity approaches α_1 for large positive x_2, and zero for large negative x_2. Unless the range of x_2 is very large, this function will be close to linear, and terms $(1, x_1, x_2, x_2^2, x_1x_2)$ are enough to approximate $h(x)$. When the range of x_2 is large and α_{12} is clearly non-zero, then using these terms will not give an adequate description of $h(x)$, and the process of transforming continuous predictors ignoring the binary variables will not provide an adequate solution.

Binary/Normal

The next special case fits the binomial and the normal in the opposite order. This is less desirable in practice because we need to transform for normality in the four groups determined by x_2 and y (or in general even more groups determined by all the binary predictors and the response). The set up is

$$\begin{aligned} x_2|y &\sim \text{Bernoulli}(p(y)) \\ x_1|x_2, y &\sim \text{N}(\alpha_0 + \alpha_1 y + \alpha_2 x_2 + \alpha_{12} y x_2, \phi(x_2, y)) \end{aligned}$$

The six terms are:

$$\begin{aligned}
(6) &= 0 \\
(7) &= x_2\,(p(1) - p(0)) \\
(8) &= -\log\left[(1 + \exp(p(1)))/(1 + \exp(p(0)))\right] \\
(9) &= \frac{1}{2}x_1^2\left(\frac{1}{\phi_1(x_2,0)} - \frac{1}{\phi_1(x_2,1)}\right) + \frac{1}{2}\log\left(\frac{\phi_1(x_2,0)}{\phi_1(x_2,1)}\right) \\
(10) &= x_1\left[\alpha_0\left(\frac{1}{\phi_1(x_2,1)} - \frac{1}{\phi_1(x_2,0)}\right) + \frac{\alpha_1}{\phi_1(x_2,1)}\right. \\
&\quad \left. + x_2\left(\frac{\alpha_2}{\phi_1(x_2,1)} - \frac{\alpha_2}{\phi_1(x_2,0)} + \frac{\alpha_{12}}{\phi_1(x_2,1)}\right)\right] \\
(11) &= \frac{1}{2}\left(\frac{(\alpha_0 + \alpha_1 + \alpha_2 x_2 + \alpha_{12}x_2)^2}{\phi_1(x_2,1)} - \frac{(\alpha_0 + \alpha_2 x_2)^2}{\phi_1(x_2,0)}\right)
\end{aligned}$$

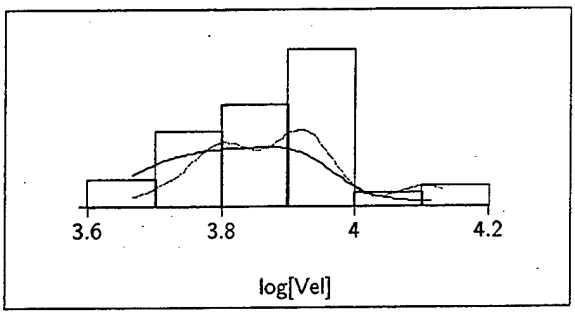
Terms (6)–(8) require only x_2 in the mean function. Term (9) requires a quadratic in x_1 only if the two variances are unequal. From term (10), an interaction is generally needed if $\phi(x_2, 1) \neq \phi(x_2, 0)$, and is always needed if $\alpha_{12} \neq 0$. Term (11) can be neglected because x_2 is binary.

Normal/Poisson Case

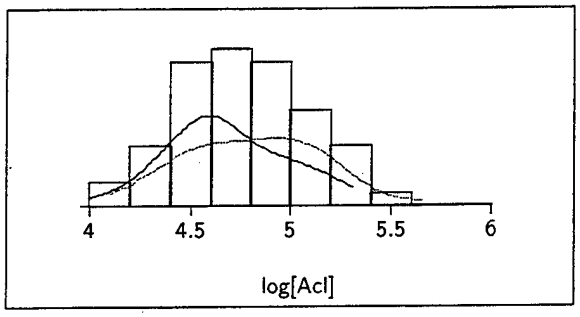
This is similar to the normal/binomial case, except we assume that $x_1|x_2, y$ is Poisson rather than Bernoulli. The only term that is different from the normal/binomial case is

$$\begin{aligned}
(11) &= \exp(\alpha_0 + \alpha_1 + \alpha_2 x_2 + \alpha_{12} x_2) - \exp(\alpha_0 + \alpha_2 x_2) \\
&= \exp(\alpha_0 + \alpha_2 x_2)\,[\exp(\alpha_1 + \alpha_{12} x_2) - 1]
\end{aligned}$$

Under the assumption that α_{12} is zero, so there is no x_1x_2 interaction, then transformation of x_1 is not required as long as $\exp(\alpha_0 + \alpha_2 x_2)$ is approximately linear in x_2.



a. log(*Vel*).



b. log(*Acl*).

Figure 3: Density estimates of transformations of two predictors.

Collision Data

The collision data has three predictors. Applying the results for the normal/normal case above, we used a multivariate generalization of the Box and Cox method to select power transformations toward normality. The maximum likelihood estimates of the suggested transformations are $(0.04, -1.5, 0.25)$ for (*Age*,*Vel*, *Acl*), respectively. The standard error of the estimate for *Vel* is very large, and a likelihood ratio test that all transformations are zero (for log transformations) has a p-value of 0.4, suggesting that using log transformations is adequate. Figure 2b displays the density estimates for $\log(Age)$; the density estimates for the other transformed predictors are shown in Figure 3. Normality is uncertain in each of these plots. Figure 2b appears slightly skewed, while Figure 3a appears possibly bimodal. However, these are density estimates based on only about 25 or so points in each population, far too few for reliable density estimation. We conclude that beginning with log transformed predictors as terms is a reasonable starting point for analysis of this problem. Figure 4 shows a scatterplot matrix of the three transformed predictors, with deaths marked with an $\times$ and survivors with a $\circ$. The plot of $\log(Age)$ versus $\log(Vel)$ is particularly interesting, as there is a straight line that can separate the deaths from the survivals with almost no error. Since the separating line is straight, the logistic regression model is appropriate with these two terms. Similar comments can be made about $\log(Age)$ and $\log(Acl)$; of course *Acl* and *Vel* are highly correlated. See Cook,[2] Chapter 5 for more on this use of graphs with a binary response.

To illustrate the results with binary variables, suppose we replace *Age* by a binary predictor *Age1* that is one if *Age* is 40 or more, and zero otherwise. Consider the regression with predictors *Age1* and *Acl*. The marginal transformation of *Acl* toward within group normality is the log transform. According to the results above, if the regressions of *Age1* on $\log(Acl)$ given y are linear in the logit scale, and if the range of $\log(Acl)$ is not too large, then we can fit the logistic regression with terms $(1, Age1, \log(Acl))$. A test for parallel within-group regression has deviance of 0.41 with 1 d.f., providing no evidence against parallel within-group regressions. Since the range of $\log(Acl)$ is quite narrow, this approach to fitting a logistic regression model is sustained.

Conclusion

With one predictor and a binary response, examination of the densities of $x|y$, and in particular the log-density ratio, provides guidance on how to select terms for a logistic regression, and a visualization of the logistic fit. One useful approach is to transform the predictor toward normality, and then use the appropriate terms in the logistic mean function.

With many predictors, the methodology generalizes, but the results are more complex. With all continuous predictors, beginning with a multivariate transformation toward normality makes sense, and if successful

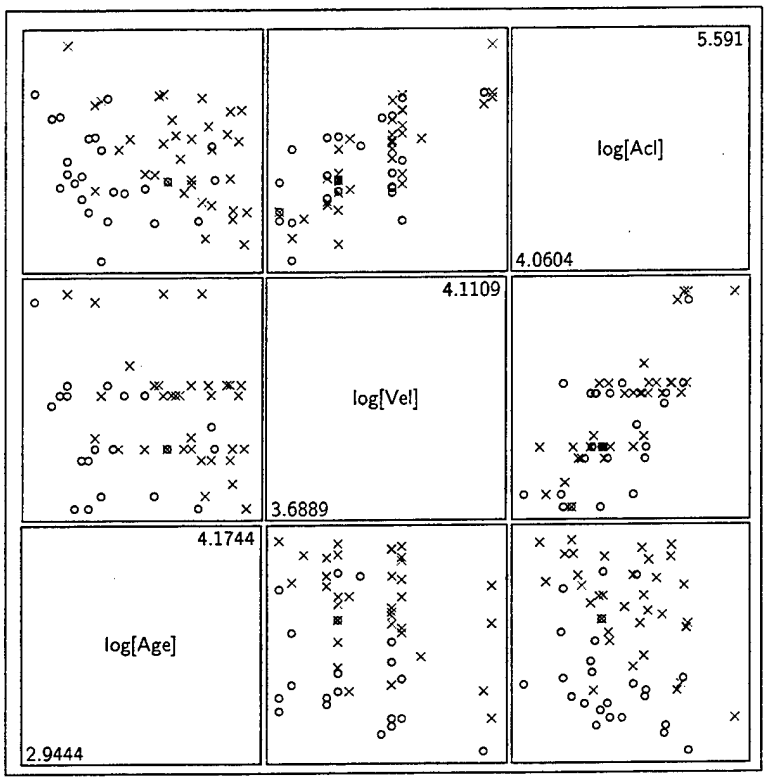


Figure 4: Scatterplot matrix of the transformed predictors in the collision data.

then fitting logistic regression is straightforward. When some predictors are binary, methodology is harder, as derived in this paper for the case of two predictors.

All the methods here are for a binary response, either zero or one. If the response is binomial, say y successes in m trials, then the problem can be converted into the Bernoulli case by allocating y cases to the "success" population, and $m - y$ cases to the "failure" population. This can also be done using weights rather than actually repeating observations.

All the computations in this paper were done using the program **Arc**, which is described in Cook and Weisberg,[3] and can be downloaded for free from `www.stat.umn.edu/arc`.

*

References

[1] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26, 211-246.

[2] Cook, R. D. (1998). *Regression Graphics: Ideas for Studying Regressions Through Graphics*. New York: Wiley.

[3] Cook, R. D. and Weisberg, S. (in press). *Applied Regression Including Computing and Graphics*. New York: Wiley.

[4] Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. London: Chapman & Hall.

[5] Härdle, W. and Stoker, T. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84, 986–995.

[6] Kay, R. and Little, S. (1987). Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika*, 74, 495–501.

[7] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Second Edition. New York: Chapman & Hall.

[8] Silverman, B. W. (1978). Density ratios, empirical likelihood and cot death. *Applied Statistics*, 27, 26-33.

[9] Velilla, S. (1993). A note on the multivariate Box-Cox transformation to normality. *Statistics and Probability Letters*, 17, 259–263.

[τ]Using Logistic Regression to Evaluate Gender Related Efficacy Differences to Mosquito Repellent Data

Claudia F. Golenda and J. Robert Burge
Walter Reed Army Institute of Research
Washington, DC 20307-5100

ABSTRACT

A clinical trial (n=120, 60 males and 60 females) was conducted to assess the efficacy of an EDTIAR (Extended Duration Tropical Insect/Arthropod Repellent) topical formulation of N, N-diethyl-m-toluamide (deet). The increased opportunity for women to serve in field positions and the paucity of information on the efficacy and metabolism of deet in females prompted research to examine the validity of label directions in regard to application dose and efficacy for the military issue repellent (EDTIAR). Volunteers were observed over a 12-hr time period to determine if the label claim indicating at least 95% protection against mosquitoes for 12 hr under normal use conditions could be substantiated. The results of a binary logistic regression analysis demonstrated that females experienced significantly less protection over time than did males. Conclusion: female soldiers using the EDITAR formulation in accordance with label directions can expect diminished protection against biting mosquitoes compared to male soldiers.

INTRODUCTION

In many situations, repellents are a first line of defense in protecting against diseases transmitted by the bites of arthropods. Vector-borne diseases, notably malaria and scrub typhus, have plagued military campaigns throughout history (1). Service members' field exposures to arthropod vectors that can transmit disease, and nuisance bites that can undermine individual and unit performance (2) have resulted in a military doctrine that stresses personal protection by the use of military-issue repellent (EDTIAR) on exposed skin (3).

Deet (N, N-diethyl-m-toluamide) is the most common active ingredient in repellent products (4) and is available in commercial solutions, lotions, gels, creams, aerosol sprays, sticks, or impregnated towelettes (5). In studies using human volunteers designed to determine the protective efficacy, duration, and biodistribution of deet formulations, the participants have been primarily males. For example, in efficacy studies conducted using deet formulations against mosquitoes (6,7), ticks (8), and sand flies (9), males comprised more that 80% of the participants . The 1994 repeal of the Department of Defense Risk Rule significantly changed the assignment policy for females in the U.S. military. It opened 260,000 additional positions in combat aviation, combatant naval vessels, and ground assignments to women. The increased opportunity for women to serve in field positions and the paucity of information on the efficacy and metabolism of deet in females lead to this clinical study to assess the validity of the label directions in regard to application dose and efficacy for the military issue repellent (EDTIAR). Volunteers were observed over a 12-hr time period to determine if the label claim indicating 95% or greater protection against mosquitoes for 12 hr or more under normal use conditions could be substantiated.

MATERIALS AND METHODS

Adult volunteers (between the ages 18-50) were recruited from the Washington, DC community and instructed in the methodology that included a high probability of them being bitten by mosquitoes. Human experimentation guidelines of the National Institutes of Health and those of the Walter Reed Army Institute of Research were followed in conducting this study. Volunteers completed a questionnaire to ensure that they did not have a history of allergic reaction to insect bites, stings or repellents.

Chemicals. The EDTIAR formulation of deet (National Stock Number 6840-01-284-3982) contained: N, N-diethyl-m-toluamide 31.58%, other isomers 1.75%, inert ingredients 66.67% (Personal

[τ] Approved for public release: distribution is unlimited.

Care Products, 3 M Consumer Specialties Division, St. Paul, MN). The same release batch of EDTIAR was used on all volunteers.

Application of Repellent. The ventral portions of both of the volunteers' forearms were wiped with alcohol pads saturated with 70% isopropanol. An area equivalent to the shape of a trapezoid was marked on the left forearm and measured between the wrist and elbow. At 7 AM repellent was applied within this area in accordance with the label directions: "Squeeze into the hand 2.5 ml of repellent, a strip equal in length and width to the diagram on the side of the tube. Rub hands together and apply thoroughly in a thin layer to both forearms". Since the EDTIAR was applied only to the venter of the left forearm rather than both forearms, the dimensions of the diagram on the side of the tube were reduced by 75%. The volunteers actively participated by removing a ribbon of repellent equivalent to the reduced diagram and applied it using their gloved, right hand. The tube of repellent was weighed before and after removal of the repellent and the application dose calculated (mg/cm^2) for each volunteer. The repellent was allowed to dry for 5 minutes. Volunteers did not cover or wash the area, remained indoors for the duration of the 12 hr test and did not exercise.

Exposure to Mosquitoes. Adult, female, laboratory-reared 3-to-7-day old Anopheles stephensi Liston mosquitoes were used. Mosquitoes were provided water but no sugar for 14-28 hr prior to exposure to volunteers. Feeding challenges occurred at 1PM, 4 PM, and 7 PM so duration of protection could be monitored at 6, 9, and 12 hr post repellent application. Cages fitted with a plastic slide apparatus that could be removed to expose the forearm skin via 5 equidistant circles (2.8 cm diameter) were used. Every volunteer underwent three challenge periods. At 6, 9 and 12 hr following repellent application, two cages, each confining 12 mosquitoes, were attached to the EDTIAR-treated and untreated forearms and slides simultaneously removed to expose mosquitoes to the volunteer for up to 5 minutes. On each arm the number of feeding mosquitoes was recorded. To limit discomfort to the volunteers, a cage was removed as soon as 12 mosquitoes fed. Cages were thoroughly washed with soap and water and dried before reuse.

Statistical Analysis: The data were analyzed using logistic regression. Minitab statistical software was used to obtain maximum likelihood estimates of the modeled parameters through an iterative-reweighted least squares algorithm (10). The odds ratios for repellent failure associated with gender, amount of administered repellent (dose), and challenge period (duration of repellent effect) were estimated. Ninety-five percent confidence intervals were based on the standard errors of the coefficients and normal approximation (11).

RESULTS

Table 1 presents the total number of mosquitoes bites received by female and male volunteers on control and repellent-treated arms at each of the challenge periods. As expected, the data indicate:

(i.) Mosquitoes fed equally on female and male control arms;
(ii.) Repellent protection decreased over time; and
(iii.) The higher the dose the better the protection.

Table 1. Total number of mosquitoes that fed by time of challenge, sex , and administered dose level.

Challenge: Hour=6

Gender / Dose	Dose=0	Below Median	Above Median	6 hr Totals
Male Volunteers	617	13	0	630
Female Volunteers	642	35	8	685
Total	1259	48	8	1315

Challenge: Hour=9

Gender / Dose	Dose=0	Below Median	Above Median	9 hr Totals
Male Volunteers	626	36	12	674
Female Volunteers	627	98	40	765
Total	1253	134	52	1439

Challenge: Hour=12

Gender / Dose	Dose=0	Below Median	Above Median	12 hr Totals
Male Volunteers	657	102	54	813
Female Volunteers	652	149	79	880
Total	1309	251	133	1693

Binary logistic regression analysis was employed to assess the gender differences in repellent efficacy over time. Dose was treated as a categorical predictor with three levels: (0) – control arm; (1) – low dose for below the median dose of 3.4 mg/cm^2; and (2) – high dose for above the median dose. The other two variables sex and hour were also modeled as factors. Table 2 provides the results and includes odds ratio estimates and associated 95% confidence intervals.

Table 2. Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	Lower 95% CL	Upper 95% CL
Constant	1.89268	0.09185	20.61	0.000			
sex F=1	0.09549	0.09542	1.00	0.317	1.1002	0.91	1.33
hour							
9	-0.0374	0.1117	-0.33	0.738	0.9549	0.77	1.20
12	0.3623	.1213	2.99	0.003	1.4366	1.13	1.82
dose							
1	-5.0326	0.1931	-26.07	0.000	0.00652	0.00	0.01
2	-6.8334	0.3838	-17.80	0.000	0.00108	0.00	0.00
INT(dxs)							
D1xF	0.7531	0.1510	4.99	0.000	2.12	1.58	2.86
D2xF	0.6689	0.1892	3.54	0.000	1.95	1.35	2.83
INT(hxd)							
H9xD1	1.2212	0.2108	5.79	0.000	3.39	2.24	5.13
H9xD2	1.9812	0.4000	4.95	0.000	7.25	3.31	15.88
H12xD1	1.7006	0.2093	8.13	0.000	5.48	3.63	8.26
H12xD2	2.6655	0.3883	6.86	0.000	14.38	6.72	30.77

The Hosmer-Lemeshow goodness-of-fit test judged the fit to be adequate, Chi-Square(df=5) = 1.386 and associated p-value = 0.926 .

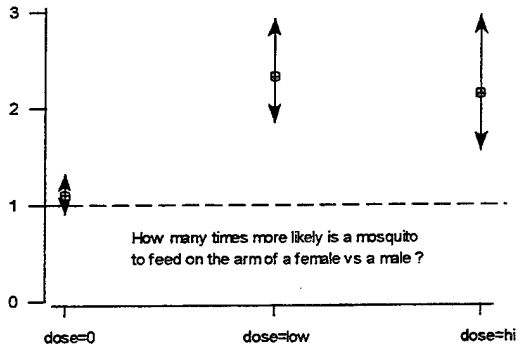
At the heart of a logistic analysis is the underlying statistical structure that the probability of an event (mosquito bite) described by a logistic function of the independent variables can be transformed to a linear function of the independent variables by converting to the log-odds. Hence, for the additive logistic model, it follows that the overall odds ratio comparing two observations is a product of a series of adjusted odds ratios. That is, since the influences of the adjusted odds ratios are multiplicative, the overall odds ratio comparing the i^{th} and j^{th} observation is a product of a series of adjusted odds ratios. Consequently, the question of whether the likelihood of "repellent failure" is the same for a female soldier as it is for a male can be examined by placing a confidence interval around the relative odds of the event. If the interval excludes 1.0, the association between the probability of repellent failure and gender is judged to be significant. If the interval overlaps 1.0, one concludes there is no association. These ideas are illustrated in Table 4 which contrasts the likelihood of repellent failure between a female versus a male soldier when both are tested nine hours after administering a high protective dose of deet.

Table 4. Computational display for comparing the overall odds ratio of two observations as the product of an associated series of adjusted odds ratios.

Predictor	Beta	Odds Ratio		X vector Female	X vector Male	Covariate Difference	Adjusted Odds Ratios
Constant	1.89268			1	1	1	
sex F1	0.09549	**1.1002**		1	0	**1**	***1.1002***
hour							
9	-0.0374	0.9633		1	1	0	
12	0.3623	1.4366		0	0	0	
dose							
1	-5.0326	0.00652		0	0	0	
2	-6.8334	0.00108		1	1	0	
INT(dxs)							
D1xF	0.7531	2.1236		0	0	0	
D2xF	0.6689	**1.9521**		1	0	**1**	***1.9521***
INT(hxd)							
H9xD1	1.2212	3.3912		0	0	0	
H9xD2	1.9812	7.2514		1	1	0	
H12xD1	1.7006	5.4772		0	0	0	
H12xD1	2.6655	14.3751		0	0	0	

The log-odds is an additive function of the predictor variables, so the influences of the adjusted odds ratios are multiplicative (i.e., ***1.1002x1.9521 = 2.1477***). Hence, a mosquito is **2.15** times more likely to feed on repellent protected arms 9 hours post application of a high dose if the 'soldier' is female versus male. A similar comparison between gender at the low dose raises the likelihood of female-to-male repellent failure to **2.34**. A picture of what the analysis has accomplished is provided in Figure 1.

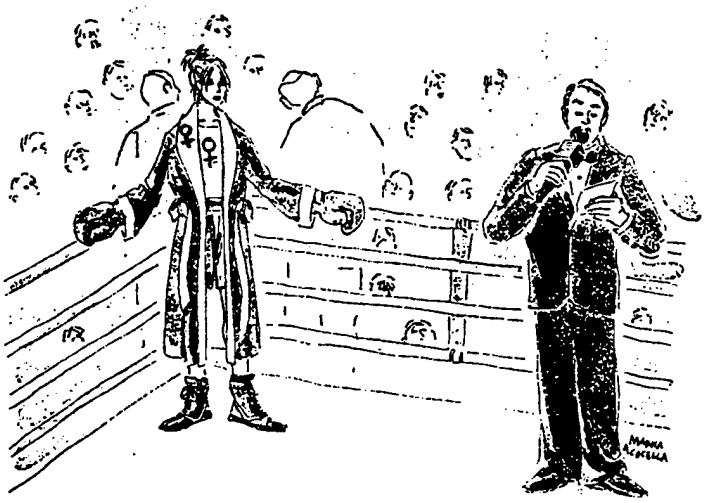
Figure 1. 95% Confidence Intervals on Odds Ratios on Repellent Failure for Females vs Males.



Following the modeling procedure, regression diagnostic plots were constructed to further assess model validity. Problems with the model were not detected after examining Pearson residuals and deviance residuals through graphical displays.

DISCUSSION

The distribution of doses of EDTIAR applied by the volunteers was similar for females and males, demonstrating that when the label directions were followed, consistent and comparable amounts were applied. In previous studies, conducted on primarily male participants, deet was reported to provide >95% protection against An. Stephensi for 10-12 hr under 3 different climatic conditions (12). In our study, by 6 hr in females and 9 hr in males the protective efficacy decreased below 95%. While there was no association between gender and the probability of repellent failure at the zero dose (i.e., on the control arms), the odds ratio described remarkable differences in the likelihood of repellent failure between females and males at both the low and high doses. Logistic regression allowed us to convincingly contrast the likelihood of repellent failure between the sexes. Furthermore, the estimated odds ratios and associated confidence intervals provided a natural way to assess gender differences in repellent over time.
Conclusions: This study contradicts the claim on the label of the Army issued EDTIAR product that indicates it: provides 95% or greater protection against mosquitoes for 12 hr or more under normal use conditions. Based on our laboratory study, females who use the EDITAR formulation in accordance with label directions can expect a significantly diminished protection against biting mosquitoes compared to male users. Logistic regression analysis provided an insightful alternative to the traditional analysis of variance approach for evaluating protective efficacy of repellents. Given the increasing military operational opportunities for females, we recommend that further studies be conducted with other military relevant mosquito species and biting arthropods.



" And in this corner, weighing five pounds more than she'd like and experiencing less than half the insect repellent efficacy of a man ..."

REFERENCES

1. Gabriel, RA and Metz, KS. A History of Military Medicine. Vol 1., Greenwood Press, NY, 30-31.

2. Mehr, ZA, Rutledge LC, Echano NM, Gupta RJ. "U.S. Army Soldiers' Perception of arthropod pests and their effects on military operations", Mil Med 12:804-807.

3. "Personal Protective Techniques Against Insects and Other Arthropods of Military Significance", 1996 Technical Information Memorandum No 136, Army Forces Pest Management Board, Forest Glen Section, Walter Reed Army Medical Center, Washington DC.

4. Osimitz TG, Grothaus RH, "The present safety and assessment of Deet", J Am Mosquito Control Assoc, 1995 11:274-278.

5. Robbins PJ, Cherniack MG, "Review of the biodistribution and toxicity of the insect repellent N, N-diethyl-m-toluamide (Deet)", J Toxicol Environ Health, 1986 18:503-525.

6. Mehr ZA, Rutledge LC, Morales EL, Meixsell VE, Korte DW. "Laboratory evaluation of controlled-release insect repellent formulations", J Am Mosquito Control Assoc, 1985 1:143-247.

7. Frances SP, Eikarat N, Sripongsa B, Eamsila C. "Response of Anopheles dirus and Aedes albopictus to repellents in the laboratory", J Am Mosquito Control Assoc, 1993 9:474-476.

8. Solberg, VB, Klein TA, McPherson KR, Bradford BS, Burge JR, Wirtz RA. "Field Evaluation of Deet and a piperidine repellent (AI3-37220) against Ambloyommna americanum (Acari:Ixodidae)", J Med Entomol, 1995 32:870-875.

9. Coleman RE, Robert LL, Roberts LW, Glass JA, Seeley DC, Laughinghouse A, Perkins PV, Wirtz RA. "Laboratory evaluation of repellents against four anopheline mosquitoes (Diperta: Culicidae) and two phlebotomine sand flies (Diptera: Psycodidae)", J Med Entomol, 1993 30:499-502.

10. McCullagh P and Nelder JA, Generalized Linear Models, 1992 Chapman & Hall.

11. Schlesselman JJ. Case-Control Studies: Design, Conduct, Analysis, 1982 Oxford University Press, p. 247.

12. Gupta RK, Rutledge LC. "Controlled-release formulations on human volunteers under three climatic regimens", J Am Mosquito Control Assoc, 1991 4:233-236.

THE COMPARATIVE EFFICACY OF SOME COMBINATORIAL TESTS FOR DETECTION OF CLUSTERS AND MIXTURES OF PROBABILITY DISTRIBUTIONS

Bernard Harris[1], Erhard Godehardt[2]

[1]University of Wisconsin, Madison, Wisconsin, U.S.A.

[2]Heinrich Heine Universität, Düsseldorf, Germany

Abstract

Assume that n k-dimensional data points have been obtained and subjected to a cluster analysis algorithm. A potential concern is whether the resulting clusters have a "causal" interpretation or whether they are merely consequences of "random" fluctuation. In previous reports, the asymptotic properties of a number of potentially useful combinatorial tests based on the theory of random interval graphs were described. In the present work, comparisons of the asymptotic efficacy of a class of these tests are provided. As a particular illustration of potential applications, we discuss the detection of mixtures of probability distributions and provide some numerical illustrations. Due to space limitations, much of the mathematical details will be described elsewhere.

1. Introduction and Summary

Let $F_X(x)$ be a cumulative distribution function on E_k, k-dimensional Euclidean space, $k \geq 1$. We assume that $F_X(x)$ is absolutely continuous with respect to k-dimensional Lebesgue measure and denote the corresponding probability density function by $f_X(x)$. Assume that a random sample of size n has been obtained from $F_X(x)$ and denote the realizations by $x_1, x_2, \ldots x_n$. In cluster analysis, similar objects are to be placed in the same cluster. We will interpret similarity as being close with respect to some distance on E_k. The relationship between graph theory and cluster analysis has been described in the books by Bock (1974) and Godehardt (1990). Mathematical results related to those used here are given in Eberl and Hafner (1971), Hafner (1972), Godehardt and Harris (1995), Godehardt and Harris (1998) and Maehara (1990).

In order to proceed, we need to introduce some notions from graph theory.

2. Graph Theoretic Concepts.

A graph $G_n = (V, E)$ is defined as follows. V is a set with $| V | = n$ and E is a set of (unordered) pairs of elements of V. The elements of V are called the vertices of the graph G and the pairs in E are referred to as the edges of the graph G. With no loss of generality,

we can assume V = {1,2,...,n}. For the purposes at hand, we choose a distance ρ on E_k and a threshold d >0. Then for i $\neq$ j, place (i,j) $\in$ E if $\rho(x_i,x_j) < d$. Since $x_1,x_2,...,x_n$ are realizations of random variables, the set E is a random set and the graph G is a random graph. In particular, these graphs are generalizations of interval graphs. Specifically, if $I_1,I_2,...,I_n$ are intervals on the real line, then the interval graph $G(I_n)$ is defined by V={1,2,...,n} and (i,j) $\in$ E if $I_i \cap I_j \neq \emptyset$, $1 \leq i<j \leq n$. Thus, for the model under consideration, if k=1, the intervals I_i, i=1,2,...,n, are the intervals $[x_i-d/2, x_i+d/2]$. Let $V_m \subset V$ with $|V_m| = m < n$. $K_{m,d}$ is a complete subgraph of order m, if all $\binom{m}{2}$ pairs of elements of V_m are in E. If m=1, then $K_{1,d}$ is a vertex, if m=2, then $K_{2,d}$ is an edge and if m=3, then $K_{3,d}$ is called a triangle. A vertex has degree ν, ν=0,1,2,...,n-1, if there are exactly ν edges incident with that vertex. If ν=0, then that vertex is said to be an isolated vertex.

3. Probability Distributions for Characteristics of Real Interval Graphs.

We now describe the probability that a specified set of m vertices form a $K_{m,d}$ With no loss of generality, we can assume that these vertices are {1,2,...,m}. Then,

$$P\{\max_{1\leq i\leq m} X_i - \min_{1\leq i\leq m} X_i \leq d\}$$

$$= m \int_{-\infty}^{\infty} [F(x+d) - F(x)]^{m-1} f(x)\, dx . \qquad (1)$$

The probability that a specified vertex has degree ν , ν = 0,1,...,n-1 is

$$P\{\text{vertex 1 has degree } \nu\}$$

$$= \int_{-\infty}^{\infty} \binom{n-1}{\nu} [F(x+d) - F(x-d)]^{\nu} \{1-F(x+d) + F(x-d)\}^{n-\nu-1} f(x)\, dx. \qquad (4)$$

To obtain asymptotic approximations to the above distributions, some assumptions concerning the behavior of the probaility density function $f_X(x)$ are needed. Hence we will assume that the probability density function is uniformly continuous on every compact subset of the carrier set for X and let $f'_X(x)$ exist and be uniformly bounded on the carrier set of X.

4. Asymptotic Behavior of Probability Distributions of Properties of Random Interval Graphs

In this section, we examine the asymptotic behavior of the probability distributions introduced in the preceding section, under the conditions $n\to\infty$ and $d\to 0$ and alsoassuming the regularity conditions for $f_X(x)$ given above . The asymptotic probability ($d \to 0$) that the vertices {1,2,...,m} form a $K_{m,d}$ is

$$m\, d^{m-1} \int_{-\infty}^{\infty} f^m(x)\, dx . \qquad (5)$$

Also, the asymptotic means and variances of the number of complete subgraphs of order m are given by

$$E\{ | K_{m,d} | \} \sim \binom{n}{m} P\{K_{m,d}\}.$$

and

$$Var\{ | K_{m,d} | \} \sim \binom{n}{2m-1} (2d)^{2m-2} \int_{-\infty}^{\infty} f^{2m-1}(x)dx.$$

5. Rates of convergence for asymptotic normality.

From the theory of U-statistics, the number of complete subgraphs of order m that will be observed in a graph on n vertices has an asymptotically normal distribution whenever $n^m d^{m-1} \to \infty$ and $n^{2m-1} d^{2m-2} \to c > 0$, as $n \to \infty$ and $d \to 0$. Hence, suppose m=2, that is, we are counting the number of edges that are observed and presuming that $n^2 d$ is "large" and $n^3 d^2$ is "moderate".. If the above rates apply, then the expected number of triangles is of the order $n^2 d(nd)$ and tends to a positive constant and the variance of the number of triangles tends to zero. Simlarly, for larger values of m, under this limiting process, asymptotically degenerate random variables will be obtained. This suggests very strongly that m=2 provides more information for a given value of n than larger values of m.

6. Detection of mixtures of probability distributions.

A problem closely related to detection of clusters is the detection of mixtures of probability distributions. A mixture of two probability density functions is a probability density function of the form

$$f(x) = \sum_{i=i}^{k} \alpha_i f_i(x),$$

and let $\alpha_i > 0$ satisfy $\sum_{i=1}^{k} \alpha_i = 1$. In order to avoid some mathematical complications, we require that the probability distributions whose probability density functions are $f_i(x)$ be mutually absolutely continuous and that they are distinct (that is, there is some set A(i,j) of positive Lebesgue measure for which $\int_A f_i(x)dx \neq \int_A f_j(x)dx$ for $i \neq j$, $1<i<j<k$). Let X_1, X_2, ..., X_n be a random sample from f(x) and let $d > 0$, be a specified threshold. The asymptotic approximations to be used here are valid under the assumption that $d \to 0$ and $n \to \infty$ so that $n^m d^{m-1} \to \infty$. Then, for example, the asymptotic probability that m randomly selected realizations result in a complete subgraph of order m is

$$P\{K_{m,d}\} = md^{m-1}\sum_{i=1}^{k}\frac{m!}{\prod_{i=1}^{k} j_i!}\prod_{i=1}^{k}\alpha_i^{j_i}\int_{-\infty}^{\infty} f_i^{j_i}(x)\,dx,$$

where $j_i \geq 0$, $\sum_{i=1}^{k} j_i = m$. We propose proceeding as follows. First, we postulate that there is no mixture, that is, k=1, α_1=1. We estmate the expected value and variance of $K_{m,d}$ under this assumption and compare the observed number of complete subgraphs of order m with this mean. If it differs substantially (relative to the variance) from the estimated expected value, then we conclude that there is a non-trivial mixture. We can repeat this, postulating that there is a mixture of two distributions and observe whether the data is compatible with this assumption, and so forth. Naturally, this process results in tests that are not stochastically independent. However, much practical statistical data analysis is carried out in essentially this manner.

7. Mixtures of exponential distributions.

The following application of these methods is apparently not of great significance in cluster analysis, but arises naturally in reliability theory and risk analysis. Nevertheless, this can serve as a demonstration of the methodology. Here, we consider detecting mixtures of exponential distributions using tests based on the theory of complete subgraphs of order m. A natural way that might be proposed is to use the likelihood ratio test and the corresponding asymptotic theory. Unfortunately in the case of mixtures, the required regularity conditions for the asymptotic theory of the likelihood ratio are not satisfied. Therefore, the present technique may be a suitable alternative.

Consequently let

$$f_i(x) = \lambda_i e^{-\lambda_i x},\ \lambda_i>0,\ x>0,\ i=1,2,\ldots,k,$$

Hence,

$$P\{K_{m,d}\} = m\,d^{m-1}\sum_{j_1 j_2,\ldots j_k}\left[\frac{m!}{j_1!j_2!\ldots j_k!}\prod_{i=1}^{k}\alpha_i^{j_i}\lambda_i^{j_i}\Big/\sum_{i=1}^{k} j_i\lambda_i\right].$$

Two specific cases are considered. In order to simplify the computations both illustrations utilize k=2 and m=2.

Example 1. Let X be the random lifetime of some device. It has been assumed that X has the exponential distribution with known intensity λ. Subsequently, it is suspected that a second production source with a different intensity rate has been introduced. Therefore some of the data may be coming from the first production source and some from the second production source and thus the observed data may be from a mixture of two

exponential distributions. With no loss of generality, we can assume λ=1. To provide a numerical illustration, 300 observations were simulated, 183 from an exponential distribution with λ_1=1 and 117 from the exponential distribution with λ_2=2, hence α= .61. The simulated lifetime data has a sample mean $\overline{X}$ = .830 and a sample variance s^2=.708. The threshold d was set at .01 resulting in an observed number of edges of 563. In addition, the sample second moment of the lifetimes is m_2 = 1.397 and the sample third moment m_3= 3.402. For this specific mixture, the theoretical mean lifetime μ = .805 and the theoretical variance σ^2 = .767. Hence the simulated data has reasonable agreement with the theory.

For example 1,we would assume that λ_1 =1 is known and that λ_2 is unknown. The natural null hypothesis H_0 is α=0, that is, that there is no mixing. If this is true, then the asymptotic value for the expected number of edges is 448.5. Similarly, the asymptotic value for the variance of the number of edges under the null hypothesis is 594.01 giving a standard deviation of 24.4. One notes that the observed number of edges is sufficiently large that the null hypothesis is untenable and hence the null hypothesis is rejected. For the specific mixture employed, the expected number of edges is 587.8. Thus, it is clear that for Example 1, detection of a mixture, using the number of edges , can be accomplished. Should one wish to estimate α from the data, there are several procedures that one might use. Typically,one would utilize the lifetime data and employ one of the standard statistical estimation techniques, such as the method of maximum likelihood or the method of moments. Because of its computational simplicity, the method of moments has been utilized, resulting in the estimates α = .25 and λ_2 = 1.29. These values are compatible with the data, since the expected number of edges for the mixture is somewhat greater than the actual number of edges obtained.

Example 2. In this example, data has been obtained which may be from a mixture of two exponential distributions, but in contrast to the Example 1, none of the putative exponential parameters is known. Hence we wish to determine if the data is from a single exponential distribution or from a mixture of two exponential distributions. To illustrate, we will use the same data as in example one. Since our intent is to demonstrate the use of these combinatorial techniques, we equate the number of edges observed to the expected number of edges and solve for the exponential parameter λ, obtaining $\hat{\lambda}$ = 1.255. Using the lifetime data, the maximum likelihood estimate under this assumption is $\hat{\lambda}$=1.204. If λ = 1.204, then the expected number of edges is 540.0 and the variance of the number of edges is 861.1 giving a standard deviation of 29.3 and there is insufficient evidence using the number of edges to support the presence of a non-trivial mixture.

If you postulate that the data is in fact from a mixture of two exponential distributions, then using the lifetime data and the method of moments, the following parameter estimates are obtained. The estimate of the mixing parameter, $\hat{\alpha}$ = .54, $\hat{\lambda}_1$ = 1.09 and $\hat{\lambda}_2$= 1.38, which is compatible with the estimate 1.204 obtained under the assumption of homogeneity.

8. Concluding Remarks.

The authors are continuing their investigation of these and related methods. At present, an extensive examination of the methods introduced by R. A. Fisher and, in particular, an analysis of the famous iris data of R. A. Fisher (1936), (1938), (1940) is under way to provide further information about the efficacy of these methods and to provide useful comparisons with other techniques.

References

BOCK, H. H. (1974) : Automatische Klassifikation. Vandenhoek & Ruprecht, Göttingen.

EBERL, W. and HAFNER, R. (1971), Die Asymptotische Verteilung von Koinzidenzen. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 18, 322-332.

FISHER, R. A. (1936), The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7, 179-188.

FISHER, R. A., (1938), The statistical utilization of multiple measurements. Annals of Eugenics, 8, 376-386.

FISHER, R. A. (1940), The precision of discriminant functions. Annals of Eugenics, 10, 422-429.

GODEHARDT, E. (1990): Graphs as Structural Models, 2nd Edition. Viewog, Braunschweig .

GODEHARDT, E. and HARRIS, B.(1995): Asymptotic Properties of Random Interval Graphs and Their Use in Cluster Analysis. University of Wisconsin Statistics Department Technical Report (Submitteed for Publication).

HAFNER, R. (1972): Die Asymptotische Verteilung von Mehrfachen Koinzidenzen. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 24, 96-108.

JAMMALAMADAKA, S. R. and JANSON, S. (1986): Limit theorems for a triangular scheme of U-staistics with applications to inter-point distances. Annals of Probability, 14 , 1347-1358.

JAMMALAMADAKA, S. R. and ZHOU, X. (1990): Some goodness of fit tests in higher dimensions based on interpoint distances. In: Proceedings of the R. C. Bose Symposium on Probability, Statistics and Design of Experiments. Wiley Eastern, New Delhi..

MAEHARA, H. (1990): On the intersection graph of random arcs on a circle. In: M. Karonski, J. Jaworski, A. Rucinski (eds.): Random Graphs, John Wiley & Sons, Ltd., Chichester.

MATCHED FILTERS AND HIDDEN MARKOV MODELS WITH DISTRIBUTED OBSERVATION

B.L. Rozovsky and S. Kligys
Center for Applied Mathematical Sciences
University of Southern California
Los Angeles, CA 90089-1113

ABSTRACT

The 3D matched filter proposed by Reed et al.[7] provides a powerful processing technique for detecting moving low observable targets. This technique is a centerpiece of various track-before-detect systems. However, the 3D matched filter is designed for constant velocity targets and its applicability to more complicated patterns of target dynamics is not obvious.

In this paper we demonstrate that the 3D matched filtering can be cast into a general framework of optimal spatio-temporal nonlinear filtering for hidden Markov models with distributed observation. A real time algorithm for detection and tracking of low observable agile targets is presented. The proposed algorithm is sequential (not only in time but also spatially) and allows for multi-resolution processing.

It is shown that the approximate scheme based on the algorithm converges to the optimal filter and the error of the approximation is computed.

INTRODUCTION

Filtering of a signal with distributed observation is an important and at the same time very challenging problem of signal and image processing. A distinctive feature of this particular problem is that the observation is a sequence of random fields rather then a random process. One important practical motivation for the distributed observation setting is the problem of tracking a dim target moving in a plane or in 3D space, using a sequence of noisy images of the region of the space in which it evolves (see e.g.[4]). If the signal to noise ratio (SNR) is low the target could not possibly be localized on a single image. In this case one has to align successive frames judiciously. If the alignment is done properly the signals of the various images would add up and produce a "spike" with a sufficiently large SNR while the noises would cancel out. This approach to detection of a dim target is usually referred to as "tracking before detection" (TBD).

Unfortunately the alignment of successive frames necessary for TBD is extremely difficult in the case of acutely maneuvering noncooperative target.

The 3D matched filter proposed by Reed et al.[7] is currently a technique of choice in various track-before-detect (TBD) systems. However, the 3D matched filter is designed for constant velocity targets and its applicability to more complicated patterns of target dynamics is questionable. In this paper we demonstrate that the 3D matched filtering for TBD can be cast into a general framework of optimal spatio-temporal nonlinear filtering. This allows to extended the matched filtering algorithm to the case of target dynamics modeled as an arbitrary Markov process.

Key words and phrases. Hidden Markov models, distributed observation, target tracking, optimal nonlinear filtering, matched filter.

The optimal nonlinear filtering is not an obvious candidate for applications to practical tracking problems. In fact this approach used to be notorious for difficulty of its practical implementation. However in the last decade a very substantial progress has been made in the development of fast and effective numerical algorithms for nonlinear filtering (see e.g.[1,2,5] and references therein). This together with the explosive development of the data processing hardware has changed the situation quite dramatically.

In this paper we extend one of the aforementioned algorithms, specifically the separation of parameters and observations scheme[1,2,3] , to the case of distributed observation processes. We derive an asymptotically optimal nonlinear filter for a Markov state process and a spatially distributed (possibly multiresolutional) observation process. The proposed algorithm is sequential (not only in time but also spatially). The latter means that if additional spatial measurements become available later in time, they can be incorporated into the Fourier coefficients of the filtering density (at all previous time moments) without complete recomputing of these coefficients. The algorithm facilitates optimal fusion of sensor measurements and prior information regarding target dynamics.

It is shown that the approximate scheme based on the algorithm converges to the optimal filter and the error of the approximation is computed.

Due to space limitation we concentrate on the theoretical aspects of the problem. The interested reader is referred to[10] for reading on simulations and application of this approach to practical problems of target tracking.

HIDDEN MARKOV MODEL WITH DISTRIBUTED OBSERVATION

Let X_t be a continuous homogeneous Markov process in $\mathbb{R}^d$ with the transition probability density $P^t(x,x') := P(X_{s+t} = x' | X_s = x)$ and the marginal probability density function $P^t(x) := P(X_t = x)$.

Assume that the state process X_t is not fully observable. Specifically, the observations Z_k are made at discrete time moments $t_k = k\Delta$, $k = 0, 1, 2, \dots$,, and are given by

$$Z_k(x) = h(X_{t_k}, x) + V_k(x), \quad x \in \mathbb{R}^d, \tag{1}$$

where $h(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $\{V_k(x)\}, x \in \mathbb{R}^d$, $k = 1, 2, \dots$ is a Gaussian system so that $\mathbf{E}V_k(x) = 0$, and $\mathbf{E}V_k(x)V_n(y) = \delta_{k,n}q(x,y)$ where $\delta_{k,n}$ is the Kronecker symbol and the function $q(x,y) \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$. We will consider the problem of estimating X_{t_k} given observation Z.

Let us fix a function $\phi : \mathbb{R}^d \to \mathbb{R}$ so that $\mathbf{E}\phi(X_{t_k})^2 < \infty$ for all k. Denote by $\hat{\phi}_k$ the best mean square estimate for $\phi(X_{t_k})$ based on $\mathcal{Z}_k = \{Z_k^l, l \geq 1\}$; $\hat{\phi}_k$ is usually referred to as the *optimal (mean square) filter* for $\phi(X_{t_k})$. It is a standard fact that $\hat{\phi}_k = \mathbf{E}[\phi(X_{t_k})|\mathcal{Z}_k]$, however efficient computation of the latter conditional expectation usually presents substantial problem.

The main goal of this paper is to present efficient numerical approximations for $\mathbf{E}[\phi(X_{t_k})|\mathcal{Z}_k]$.

The spatial covariance operator of the signal noise is given by $Qf(x) = \int_{\mathbb{R}^d} q(x,y)f(y)dy$. Throughout what follows we will assume that the trace of Q is finite, i.e.

$$traceQ = \int_{\mathbb{R}^d} g(x,x)\,dx < \infty \tag{2}$$

Let $\{e_k(x)\}_{k\in\mathbb{N}}$ be an orthonormal basis in $L^2(\mathbb{R}^d)$. The spatial covariance operator Q is fully characterized by the (infinite-dimensional) matrix $(Q^{ij})_{i,j\geq 1}$ where $Q^{ij} = \int q(x,y)e_i(x)e_j(y)dxdy$. This matrix is symmetric and positively definite. Thus one can choose the basis $\{e_k(x)\}_{k\in\mathbb{N}}$ so that $\sum_j Q^{ij}e_j(x) = \lambda_i e_i(x)$,

$\lambda_i \geq 0$, and $traceQ = \sum_i \lambda_i < \infty$. Obviously in this basis $(Q^{ij})_{i,j\geq 1}$ is diagonal and $\lambda_i, i = 1, 2, ...$ are the diagonal entries. In addition, everywhere below we will assume that the matrix Q is invertible ($\lambda_i > 0$) and

$$\sup_y \int_{\mathbb{R}^d} (Q^{-1}h(y,x))^2\, dx < \infty.$$

Of course in applications it is practical to assume that only finite number of statistics based on the observation field $Z_\cdot(x)$ are known, e.g. measurements of $Z_l(x)$ on some finite grid or a finite number of spatial Fourier coefficients of $Z_\cdot$, etc. In this paper we will specialize in the later case.

Projecting (1) on the basis $\{e_l(x)\}_{l\in\mathbb{N}}$, we can rewrite the observation process in the coordinate form:

$$Z_k^l = h^l(X_{t_k}) + V_k^l, \quad l = 1, 2, \ldots, \quad k = 1, 2, \ldots, K, \tag{3}$$

where

$$Z_k^l = \int_{\mathbb{R}^d} Z_k(x) e_l(x)\, dx,$$
$$h^l(y) = \int_{\mathbb{R}^d} h(y,x) e_l(x)\, dx,$$
$$V_k^l = \int_{\mathbb{R}^d} V_k(x) e_l(x)\, dx.$$

Obviously, V_k^l are zero mean Gaussian random variables with covariance $\mathbf{E}V_k^l V_n^m = \delta_{k,n}\delta_{l,m}\lambda_l$.

It is a standard fact that $\hat{\phi}_k^N = \mathbf{E}[\phi(X_{t_k})|Z_k^l, l \leq N]$ is the best mean square estimate of $\phi(X_{t_k})$ based on $Z_k^l, l \leq N$. In the future we refer to $\hat{\phi}_k^N$ as an *optimal projection filter*. Write $Q_N = (Q^{ij})_{i,j\leq N}$. Obviously Q_N is the covariance matrix for the random vector $\{Z_k^l, l = 1, ..., N\}$.

Theorem 1. *The optimal projection filter $\hat{\phi}_k^N$ is given by*

$$\hat{\phi}_k^N = \int_{\mathbb{R}^d} \phi(x) P_k^N(x) dx \Big/ \int_{\mathbb{R}^d} P_k^N(x) dx, \tag{4}$$

where the unnormalized filtering density (UDF) $P_k^N(x)$ is a solution of the following equation

$$P_0^N(x) = \exp\{\textstyle\sum_{l=1}^N \lambda_l^{-1} h^l(x) Z_0^l - \frac{1}{2}\sum_{l=1}^N \lambda_l^{-1} h^l(x)^2\} P_0(x)$$

$$P_k^N(dx) = \exp\{\textstyle\sum_{l=1}^N \lambda_l^{-1} h^l(x) Z_k^l \tag{5}$$

$$-\tfrac{1}{2}\textstyle\sum_{l=1}^N \lambda_l^{-1} h^l(x)^2\} \int_{\mathbb{R}^d} P^\Delta(x,x') P_{k-1}^N(x') dx', k \geq 1.$$

The optimal filter $\hat{\phi}_k = \lim_{N\to\infty} \hat{\phi}_k^N$, moreover $\hat{\phi}_k$ satisfies (4) and (5) where N is replaced by ∞.

Proof of this Theorem is somewhat involved and for the sake of shortness will be omitted. The interested reader is referred to Kligys[8].

Remark. Obviously the optimal posterior (filtering) densities with respect to the observation $\{Z_k^l, l \leq N\}$ is given by $\pi_k^N(x) = P_k^N(x) / \int_{\mathbb{R}^d} P_k^N(x) dx$ for all finite N and for $N = \infty$.

Approximation of the optimal filter by the projection filters is discussed in more detail in the last Section.

MATCHED FILTERS

Let $G_y(x)$ be a deterministic function in $\mathbb{R}^d \times \mathbb{R}^d$. Assume that $G_y(x)$ evolves along some unknown continuous curve Y_t in $\mathbb{R}^d$. Let us assume also that the observation of $G_{Y_{t_l}}(x)$ is given by

$$Z_l(x) = G_{Y_{t_l}}(x) + V_l(x), l \leq k \tag{6}$$

Suppose that we want to estimate the path of Y. from the observation given by (6).

One can think of $G_y(x)$ as the image (signature) of an object (target) moving along the path Y_t. In this context, Y_t can be thought of as the trajectory of some reference point of the image, e.g. the center of mass.

One important particular case often considered in target tracking (see e.g.[7]) is

$$G_y(x) = G(x-y), G(z) = 0 \text{ if } |z| > R \text{ and } G(z) = 0 \text{ if } |z| \leq R. \tag{7}$$

In this case the target signature $G(x-y)$ of radius $R/2$ evolves along the path Y_t without rotation. Write $\hat{G}_k(\tilde{Y}) = \sum_{l\leq k} \int_{\mathbb{R}^d} G_{\tilde{Y}_{t_l}}(y) Q^{-1} Z_l(y) dy$. One possible estimate of the true path Y_{t_l} is the function $\hat{Y}_{t_l}$ so that $\hat{G}_k(\hat{Y}) = \max_{\tilde{Y}} \hat{G}_k(\tilde{Y})$. This estimate is quite popular in engineering literature and is usually referred to as a matched filter estimate (see Pratt[6], Reed[7], etc.). Note that the curve $\{\hat{Y}_{t_l}\}_{l\leq k}$ "matches" the data Z to the evolution of the image $G_{\tilde{Y}_{t_l}}$ by maximizing their Q- normalized correlation $\hat{G}_k(\tilde{Y})$. This explains the origins of the term "matched filter". For the reasons that become clear later, we will call this filter *maximum correlation matched filter* (MCMF). The MCMF is optimal in that it maximizes the output signal to the output noise ratio.

Now we shall introduce a modified matched filter. Let $\{\bar{Y}_{t_l}\}_{l\leq k}$ be a sequence in $\mathbb{R}^d$ that maximizes

$$D_k(\tilde{Y}) := \sum_{l\leq k} \int_{\mathbb{R}^d} G_{\tilde{Y}_{t_l}}(y) Q^{-1} Z_l(y) dy - \frac{1}{2} \sum_{l\leq k} \int_{\mathbb{R}^d} G_{\tilde{Y}_{t_l}}(y) Q^{-1} G_{\tilde{Y}_{t_l}}(y) dy.$$

Obviously,

$$\begin{aligned} &\max_{\tilde{Y}} D_k(\tilde{Y}) = \tfrac{1}{2} \textstyle\sum_{l\leq k} \int_{\mathbb{R}^d} Z_l(y) Q^{-1} Z_l(y) dy \\ &-\tfrac{1}{2} \min_{\tilde{Y}} \textstyle\sum_{l\leq k} \int_{\mathbb{R}^d} (Q^{-1/2}(Z_l(y) - G_{\tilde{Y}_{t_l}}(y)))^2 dy \\ &= \tfrac{1}{2} \textstyle\sum_{l\leq k} \left|Q^{-1/2} Z_l\right|^2_{\mathbb{L}^2} - \tfrac{1}{2} \min_{\tilde{Y}} \sum_{l\leq k} \left|Q^{-1/2}(Z_l - G_{\tilde{Y}_{t_l}})\right|^2_{\mathbb{L}^2} \end{aligned} \tag{8}$$

Here and below we use notation $|.|_{L^2}$ and $(\cdot,\cdot)_{\mathbb{L}^2}$ for the norm and the scalar product in $L^2(\mathbb{R}^d)$, respectively. Thus, $\{\bar{Y}_{t_l}\}_{l\leq k}$ "matches" the data, Z, to the evolution of the image $G_{\tilde{Y}_{t_l}}$ by minimizing the distance between $Q^{-1/2} Z_l$ and $Q^{-1/2} G_{\tilde{Y}_{t_l}}$. We refer to this estimation algorithm as the *minimum distance matching filter* (MDMF).

In spite of the fact that the MDMF does not necessarily maximize the output signal to the output noise ratio, for the purpose of estimation of the state process Y_t it is as good as MCMF, and may be even more natural then the latter. Moreover, if

$$\left|Q^{-1/2} G_y\right|_{\mathbb{L}^2} = \left|Q^{-1/2} G_x\right|_{L^2} \text{ for all } x, y \in \mathbb{R}^d, \tag{9}$$

then the estimates of $\{Y_{t_l}\}_{l\leq k}$,given by MCMF and MCMD coincide.

Note that condition (9) holds in many important cases. For example it is satisfied if $Q = I$ and (7) holds.

Now it is clear that the MDMF is a particular case of the Optimal Nonlinear Filter given by (4) and (5). Indeed, let the unobservable state process X_t be a deterministic function so that $X_0 = x$ and $h(y, y') = G_y(y')$. It is readily checked that in this case

$$P_k(x) = \Pi_{i=0}^k \exp\left\{(Q^{-1}G_{X_{t_i}}, Z_k)_{L^2} - \frac{1}{2}\left|Q^{-1/2}G_{X_{t_i}}\right|_{L^2}^2\right\}\delta_{X_{t_k}}(x). \tag{10}$$

Write $\hat{X}_k := \arg\max_x P_k(x)$, i.e. $\hat{X}_k$ is the $P_k(x)$-based maximum likelihood estimate of the vector $X^k = (X_{t_0}, \ldots, X_{t_k})$. Then by (10) we have

$$\hat{X}_k = \arg\max_{\tilde{X}} \Pi_{i=0}^k \exp\left\{(Q^{-1}G_{\tilde{X}_{t_i}}, Z_k)_{L^2} - \tfrac{1}{2}\left|Q^{-1/2}G_{\tilde{X}_{t_i}}\right|_{L^2}^2\right\}$$

$$= \arg\max_{\tilde{X}} \exp\left\{\textstyle\sum_{l\le k}(Q^{-1}G_{\tilde{X}_{t_i}}, Z_k)_{L^2} - \frac{1}{2}\sum_{l\le k}\left|Q^{-1/2}G_{\tilde{X}_{t_i}}\right|_{L^2}^2\right\}$$

$$= \arg\max_{\tilde{X}} \left\{\textstyle\sum_{l\le k}(Q^{-1}G_{\tilde{X}_{t_i}}, Z_k)_{L^2} - \frac{1}{2}\sum_{l\le k}\left|Q^{-1/2}G_{\tilde{X}_{t_i}}\right|_{L^2}^2\right\}.$$

Thus $\hat{X}_k$ coincides with the MDMF estimate.

NUMERICAL APPROXIMATIONS: A SPECTRAL APPROACH

In this section we concentrate on an important particular type of the state process X_t. Specifically we will assume that X_t is a diffusion type process governed by the Itô equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, X_0 = x_0$$

where $b(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma(\cdot) : \mathbb{R}^d \to \mathbb{R}^{d\times d_1}$, and W_t is a d_1-dimensional Brownian motion. It will be assumed that b, h and σ are bounded. In this case, under quite general assumptions

$$u(t,x) := T_t^*\phi(x) := \int_{\mathbb{R}^d} P^t(y,x)\phi(y)dy$$

is a solution of the Fokker-Plank equation

$$\frac{\partial u(t,x)}{\partial t} = \mathcal{A}^* u(t,x), u(0,x) = \phi(x), \tag{11}$$

where

$$\mathcal{A}^* u(t,x) = \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j}\left(a_{ij}(x)u(t,x)\right) - \sum_{i=1}^d \frac{\partial}{\partial x_i}\left(b_i(x)u(t,x)\right).$$

where $a_{ij} = \frac{1}{2}\sum_{\ell=1}^{d_1}\sigma_{i\ell}(x)\sigma_{j\ell}(x)$

In principle, formulas (4), (5) provide an algorithm for computing the estimate $\hat{\phi}_k$. However, the algorithm becomes inefficient when the dimension of the state process is large. The main difficulty of practical implementation of the optimal nonlinear filtering is computational complexity of evaluating $T_\Delta^* P_{k-1}^N(x) = \int_{\mathbb{R}^d} P^\Delta(x,x')P_{k-1}^N(x')dx'$, the "prediction term" in the recursive formula (5) for the posterior distribution of the state process . As indicated above, computing of $T_\Delta^* P_{k-1}^N(x')$ reduces to solving the Fokker-Planck equation (11). If the dimension d of the state space is large, solving this equation in real time is a very difficult task.

We remark also that even though the parameters (the coefficients b, h, σ, and the prior distribution $P_0(x)$) might be known *a priori*, the algorithm (5) does not allow off-line solving of the Fokker-Planck equation. The reason is that on every time step the initial condition, $P_{k-1}^N(x)$, depends on the previous measurements. Moreover on-line computations of the integrals in (4) even for simple functions ϕ may be very time consuming if the dimension of the state process is large. These computations alone can rule out the on-line application of the algorithm.

The objective of this section is to develop a recursive numerical algorithm for computing $\hat{\phi}_k$ in which the on-line part is as simple as possible; in particular, no differential equations are to be solved on-line.

Following the approach introduced in[1,2] we develop a spectral separating scheme (S^3) for solving the nonlinear filtering problem (4), (5).

In the proposed algorithm (see below) both time consuming operations of solving the Fokker-Planck equation and computing the integrals are performed off line, which makes the algorithm suitable for on-line implementation. Since the result is only an approximation of the optimal filter, the error of the approximation is computed.

In this section we assume that the coefficients $a_{ij}(x) \in C_b^2, b_i(x) \in C_b^1, a = \frac{1}{2}\sigma\sigma^*$ is uniformly nondegenerate, and $P^0 \in L_2(\mathbb{R}^d) \bigcap C_2(\mathbb{R}^d)$.

Let $\{\zeta_k(x)\}_{k\in\mathbb{N}}$ be a CONS in $L_2(\mathbb{R}^d)$. Due to our assumptions, for every k, the unnormalized filtering density (UFD) $P_k^N \in L_2(\mathbb{R}^d)$, $\mathbf{P}-a.a.$ (see[9]) and so it admits the following development

$$P_k^N(x) = \sum_l \psi_l^N(k)\zeta_l(x) \tag{12}$$

where $\psi_l^N(k) = \int_{\mathbb{R}^d} P_k^N(x)\,\zeta_l(x)\,dx$ is the lth Fourier coefficient of P_k^N.

Let $\mathcal{J}^N$ be the set of all multiindices $\alpha = (\alpha_1, \alpha_2, \ldots \alpha_N)$ so that all α_i are nonnegative and finite.

Lemma. *The correction term in (5) can be expanded in the series*

$$\exp\{\sum_{l=1}^N \lambda_l^{-1} h^l(x) Z_k^l - \frac{1}{2}\sum_{l=1}^N \lambda_l^{-1} h^l(x)^2\} = \sum_{\alpha\in\mathcal{J}^N} H^\alpha(Z_k)\, G_\alpha(x) \tag{13}$$

where

$$H^\alpha(Z_k) = \Pi_{i=1}^N H_{\alpha_i}(\lambda_i^{-1/2} Z_k^i)/\sqrt{\alpha_i!}$$

$$G_\alpha(x) = \Pi_{i=1}^N (\lambda_i^{-1/2} h^i)^{\alpha_i}/\sqrt{\alpha_i!}$$

and $H_n(y)$ is the nth Hermite polynomial defined by $(-1)^n e^{y^2/2}\frac{d^n}{dy^n}e^{-y^2/2}$.

Proof. It is readily checked that for every θ, the function $\exp\{\theta z - z^2/2\}$ is analytic in z and could be expanded in the series

$$\exp\{\theta z - z^2/2\} = \sum_{n=1}^\infty \frac{z^n}{n!} H_n(\theta).$$

Thus,

$$\exp\{\textstyle\sum_{l=1}^{N} \lambda_l^{-1} h^l(x) Z_k^l - \frac{1}{2} \sum_{l=1}^{N} \lambda_l^{-1} h^l(x)^2\} =$$

$$\textstyle\prod_{l=1}^{N} \sum_{i=1}^{\infty} (\lambda_l^{-1/2} h^l(x))^i H_i(\lambda_l^{-1/2} Z_k^l)/i! =$$

$$\textstyle\sum_{\alpha} \prod_{l=1}^{N} (\lambda_l^{-1/2} h^l(x))^{\alpha_i} H_{\alpha_i}(\lambda_l^{-1/2} Z_k^i)/\alpha_i!.$$

■

Substituting, (13) into (5) and projecting the resulting formula on the subspace generated by $\zeta_l(x)$, we arrive at the following result.

Theorem 2. *The Fourier coefficients $\psi_l^N(k)$ of the UFD $P_k^N(x)$ satisfy the equation*

$$\begin{aligned} \psi_l^N(k) &= \sum_n \Phi_{ln}^N(Z_k) \psi_n^N(k-1), \\ \psi_l^N(0) &= \int_{\mathbf{R}^d} P_0^N(x) \zeta_l(x) dx \end{aligned} \tag{14}$$

where $\Phi_{ln}^N(Z_k) = \sum_{\alpha \in \mathcal{J}^N} \Gamma_{l\alpha n} H^{\alpha}(Z_k)$ and $\Gamma_{l\alpha n} = \int_{R^d} G_{\alpha}(x)\, \zeta_l(x)\, T_{\Delta}^* \zeta_n(x)\, dx$.

Note that in contrast to the original algorithm (5) for UFD, the recursion (14) separates parameters and observations, in that, it allows to shift solving of the Fokker-Planck equations off line. Indeed, the only term in (14) that involves solving the Fokker-Planck equation is Γ, and this term can be precomputed before any measurements Z_k , $k \geq 1$,are obtained.

Obviously, to make the recursion (14) practically applicable, one has to truncate all the involved infinite series. After the truncation of the procedure, the exact equation for the spatial Fourier coefficients of the UFD (14) will turn into an approximate one.

Finally we arrive at the following spectral separating scheme (S^3):

1. Set a cut-off level M for the number of the basis elements $\{\zeta_\ell\}$.
2. Compute recursively

$$\begin{aligned} &\bar{\psi}_m(0) = \psi_m^N(0) \\ &\bar{\psi}_m(k) = \textstyle\sum_{n \leq M}^{N} \Phi_{mn}(Z_k) \bar{\psi}_n(k-1), m \leq M. \end{aligned} \tag{15}$$

3. Compute approximations to the unnormalized posterior density, the unnormalized optimal filter and normalized optimal filter by formulas

$$\begin{aligned} &\overline{P}_k(x) = \textstyle\sum_{\ell \leq M} \bar{\psi}_\ell(k) \zeta_\ell(x), \\ &\bar{\Psi}_k[\phi] = \textstyle\sum_{\ell \leq M} \bar{\psi}_\ell(k) (\phi, e_l)_{L^2} \\ &\overline{\phi}_k = \Psi_k[\phi]/\Psi_k[1]. \end{aligned} \tag{16}$$

The resulting finite dimensional algorithm (15), (16) has the same structure as the exact formulas given in Theorem 1, in particular it is recursive in time.

Since the algorithm involves spatial data, it would be of course very desirable to have an option to feed them into the algorithm also recursively. Such an option might be invaluable in practical situations when the measurements done at the same time become available for processing in stages. Even when all

the measurement are available, it is often convenient to do scalable processing, e.g. to start with lower resolutions and then, if necessary, to add higher ones.

Let us assume now that by the time Δk, additional measurements (Z_k^ℓ), $N < \ell \leq N'$ became available. Write $\psi'_m(n), n \leq k$ for the coefficients computed by the formula obtained from (15) by substituting N' for N in $\Phi_{ln}^N(Z_k)$. Then, obviously,

$$\begin{aligned} &\psi'_m(n) = \bar{\psi}_m(n), n \leq k-1 \\ &\psi'_m(k) = \bar{\psi}_m(k) + \textstyle\sum_{n\leq M} \sum_{\alpha\in\mathcal{J}^{N'}/\mathcal{J}^N} \Gamma_{l\alpha n} H^\alpha(Z_k)\, \bar{\psi}_n(k-1) \end{aligned} \tag{17}$$

It follows from (17) that in the aforementioned situation the Fourier coefficients at the last moment can be recomputed recursively.. Unfortunately, if one would like to incorporate additional measurements made not only in the last moment $k\Delta$ but also in the previous moments $n\Delta, n < k$, the situation becomes more complicated.

Let $\mathcal{J}^{L_0} \subset \mathcal{J}^{L_1}$ be subsets of $\mathcal{J}^N$. Denote by $\psi_m^{L_i}(k), i = 0, 1$, the coefficients computed by the formula obtained from (15) by substituting L_i for L.

Since

$$\begin{aligned} &\Phi_{ln}^{L_1}(Z_k) = \Phi_{ln}^{L_0}(Z_k) + \\ &\textstyle\sum_{\alpha\in(\mathcal{J}^{L_1}/\mathcal{J}^{L_0})} \Gamma_{l\alpha n} H^\alpha(Z_k), \end{aligned}$$

we have

$$\begin{aligned} &\psi_m^{L_1}(0) = \psi_m(0), \\ &\psi_m^{L_1}(k) = \textstyle\sum_{n\leq M} \Phi_{ln}^{L_0}(Z_k)\psi_n^{L_1}(k-1) + \sum_{n\leq M} \Phi_{ln}^{L^c}(Z_k)\psi_n^{L_1}(k-1), \end{aligned} \tag{18}$$

where $\Phi_{ln}^{L^c}(Z_k) = \sum_{\alpha\in(\mathcal{J}_{L_1}\setminus\mathcal{J}_L)} \Gamma_{l\alpha n} H^\alpha(Z_k)$.

Note, that if L_0 is large, which is the case in the image processing problem that motivated the present research, formula (18) allows to substantially reduce the computational complexity of the on-line part of the algorithm. This reduction of complexity is mainly achieved by precomputing and storing the elements of the matrices $\Phi_{mn}^{L_0}(Z_n)$, $n = 1, 2, \ldots, k$. Of course, one has to consider the trade-off between the computational complexity and storage capacity.

In conclusion, we remark that if a, b, h, p are smooth enough, then for each $K \in \mathbb{N}$, $\gamma \in \mathbb{R}_+$, there is a constant C depending only on K and the parameters of the model such that

$$\max_{k\leq K} |P_k^N - \sum_{m\leq M} \bar{\psi}_m(k)\zeta_m| \leq C(\Delta + M^{-\gamma}\Delta^{-1/2}).$$

This statement can be proved in complete analogy with[2] and we leave it to the interested reader.

ACKNOWLEDGEMENTS

Supported in part by the U.S. Office of Naval Research grant N00014-95-1-0229 and the U.S. Army Research Office grant DAAG55-98-1-0418.

REFERENCES

[1] Lototsky, S., Mikulevičius, R., and Rozovskii, B.L. "Nonlinear Filtering Revisited. A Spectral Approach." SIAM J. Opt. and Control, 1997.

[2] Lototsky, S.V. and Rozovskii, B.L. "Recursive Nonlinear Filter For A Continuous-Discrete Time Model." IEEE Trans. Automatic Control, 1998.

[3] Lototsky, S.V. and Rozovskii, B.L. "Recursive Multiple Wiener Integral Expansion For Nonlinear Filtering Of Diffusion Processes." In: A Festschrift in Honor of M.M. Rao, Marcel Dekker, 1996.

[4] Haddad, Z.S. and Simanca, S.R., "Filtering Image Records Using Wavelets And The Zakai Equation." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 1, No. 11, pp. 1069-1078, 1995

[5] Lototsky, S.V., Rao, C., and Rozovskii, B.L. "Fast nonlinear filter for continuous-discrete time multiple models." Proc. 35th IEEE Conf. on Decision and Control, Kobe, Japan, pp. 4060-4064, 1996.

[6] Pratt, W.K. Digital Image Processing. New York: John Wiley, 1991.

[7] Reed, I.S., Gagliardi, R.M., and Shao, H.M. "Application Of Three-Dimentional Filtering To Moving Target Detection." IEEE Trans. AES, vol. 19, No. 6, pp. 897-905, 1983.

[8] Kligys, S., "Some Problems In Statistics Of Random Processes With Applications To Nonlinear Filtering And Image Processing." PhD Thesis, Department of Mathematics, University of Southern California, 1998.

[9] Rozovskii,B.L. Stochastic Evolution Systems. Dordrecht, Boston: Kluwer Academic Publ., 1990.

[10] Kligys, S., Rozovsky, B.L., and Tartakovsky, A.G. "Detection Algorithms And Track Before Detect Architecture Based On Nonlinear Filtering For Infrared Search And Track Systems." Technical Report CAMS–98.9.1, Center for Applied Mathematical Sciences, University of Southern California, 1998.
(Available at `http://www.usc.edu/dept/LAS/CAMS/usr/facmemb/tartakov/preprints.html`).

INTENTIONALLY LEFT BLANK.

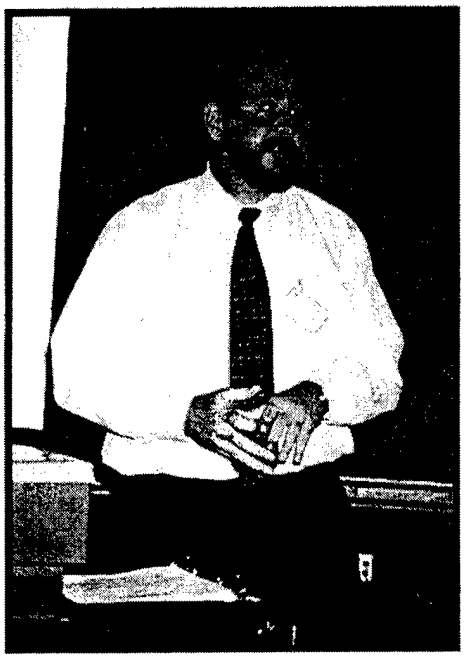
Appendix:

Conference Snapshots

INTENTIONALLY LEFT BLANK.

The Conference

Nozer Singpurwalla leads off with the tutorial. Attending, left to right: Jagdish Chandra, David Cruess, and Barnard Bissinger.

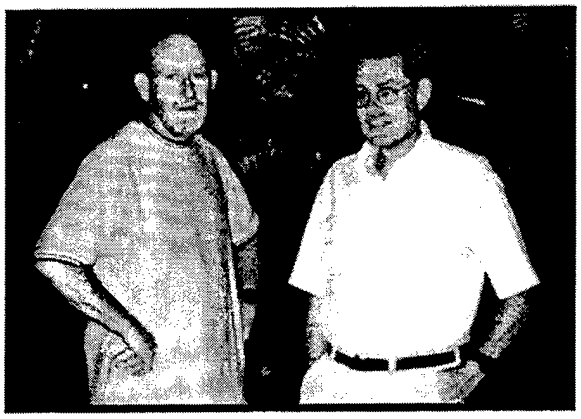


Joan Clarke (left) and Tom Lucas (right) during their presentations.

Lounell Southard, conference host, outlines the events of the day.

After Hours

Left to right: Jock Grynovicki, Robert Launer, and Carl Russell



Tom Walker (left) and Bob Burge (right)

Roy Reynolds (left) and Barry Bodt (right)

Left to right, Army Wilks Award winners in attendance: Bernie Harris, James Thompson, Robert Launer, Jay Conover Nozer Singpurwalla, Douglas Tang, Jayaram Sethuraman

ATTENDANCE LIST

U.S. ARMY CONFERENCE ON APPLIED STATISTICS
21–23 OCTOBER 1998

Donald Barr
Department of Systems Eng.
U.S. Military Academy
West Point, NY 10996

MAJ Philip Beaver
Department of Math Sci.
U.S. Military Academy
West Point, NY 10996

John Binkley
3531 White Court
Torrance, CA 90503

Burak Birgoren
Department of Ind/Man Eng.
Penn State University
State College, PA 16802

Barnard Bissinger
281 W. Main Street
Middletown, PA 17057

Barry Bodt
U.S. Army Research Laboratory
ATTN: AMSRL-IS-CI
Aberdeen Proving Ground, MD
21005-5067

Stephen Book
The Aerospace Corp.
P.O. Box 92957, M4/921
Los Angeles, CA 90009-2957

Ann Brodeen
U.S. Army Research Laboratory
ATTN: AMSRL-IS-CI
Aberdeen Proving Ground, MD
21005-5067

Barbara Broome
U.S. Army Research Laboratory
ATTN: AMSRL-IS-CI
Aberdeen Proving Ground, MD
21005-5067

Robert J. Burge
7134 Smooth Path
Columbia, MD 21045

Patrick Cassady
U.S. Army TRAC-WSMR
ATTN: ATRC-WAC
WSMR, NM 88002-5502

Aivars Celmins
U.S. Army Research Laboratory
ATTN: AMSRL-IS-CI
Bldg 1116A
Aberdeen Proving Ground, MD
21005-5067

Jagdish Chandra
U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783

Joan Clarke
Waterways Experiment Station
ATTN: CEWES-ES-F
3909 Halls Ferry Road
Vicksburg, MS 39180

W. J. Conover
College of Bus. Admin.
Texas Tech University
Lubbock, TX 79409

Paul Cox
2930 Huntington Drive
Las Cruces, NM 88011

David F. Cruess
6123 Camel Back Lane
Columbia, MD 21045

Gary Cunningham
New Mexico State University
Las Cruces, NM 88003

Thomas F. Curry
105 E Vermijo Suite 450
Colorado Springs, CO 80903

Mike Cushing
U.S. Army Materiel Systems Analysis Activity
ATTN: AMXSY-A
Aberdeen Proving Ground, MD 21005-5001

Mike Danesh
126 Kitty Hawk Lane
Harvest, AL 35749

Paul Deason
U.S. Army TRAC-WSMR
ATTN: ATRC-WAD
WSMR, NM 88002-5502

Gene Dutoit
3544 Statler Drive
Columbus, GA 31907

Robert G. Easterling
Sandia National Laboratories
MS 0419
Albuquerque, NM 87185

Bruce Gafner
U.S. Army TRAC-WSMR
ATTN: ATRC-WAD
WSMR, NM 88012

Doug Frank
Department of Math
Indiana University of Penn
233 Stright Hall
Indiana, PA 15705

Duke Gard
2213 Laramie Drive
Las Cruces, NM 88011

Donald Gaver
Operations Research Department
Naval Postgraduate School
Monterey, CA 93943

Paul Girard
10260 Campus Point Drive
San Diego, CA 92121

Jock Grynovicki
U.S. Army Research Laboratory
ATTN: AMSRL-HR-SD
Aberdeen Proving Ground, MD 21005-5425

Bernard Harris
Department of Stat.
University of Wisconsin-Madison
Madison, WI 53706

Judith Hochberg
Los Alamos National Laboratory
MS B265
Los Alamos, NM 87545

John Hosford
The Boeing Company
1800 Satellite Boulevard DL3M
Duluth, GA 30097

Scott Huxel
UNISYS Corp., Suite 300
1401 Wilson Boulevard
Arlington, VA 22209-2306

Tapas Kanungo
University of Maryland
College Park, MD 20742

Barbara J. Kaschenbach
U.S. Army Aberdeen Test Center
ATTN: STEAC-SL-B
Aberdeen Proving Ground, MD
21005-5059

Pat Kinney
U.S. Army TRAC-WSMR
WSMR, NM 88002-5502

Kragg Kysor
PO Box 685
Edgewood, MD 21040

Richard Laferriere
U.S. Army TRAC-WSMR
WSMR, NM 88002-5502

Robert Launer
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC
27709-2211

Charles Leake
8101 Ainsworth Avenue
Springfield, VA 22152-2431

Tim Linn
U.S. Army TRAC-WSMR
WSMR, NM 88002-5502

Tom Lucas
Operations Research Department
Naval Postgraduate School
Monterey, CA 94943

John Lyons
U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783

John F. Lyons
Applied Physics Laboratory
Bldg. 24
Laurel, MD 20723

Thomas Mathew
Department of Math/Stat.
UMBC
1000 Hilltop Circle
Baltimore, MD 21250

MAJ Kent M. Miller
4513 Foxwood Road
Chester, VA 23831

Linda Moss
U.S. Army Research Laboratory
ATTN: AMSRL-SL-BE
Aberdeen Proving Ground, MD
21005-5068

MAJ Andy Napoli
Department of Math Sci.
U.S. Military Academy
West Point, NY 10996

LTC David Olwell
Operations Research Department
Naval Postgraduate School
Monterey, CA 93943

Kathy Nau
Box 444
WSMR, NM 88002

Mike Prather
U.S. Army EAC
ATTN: CSTE-EAC-AV/CS
Aberdeen Proving Ground, MD
21005-3013

Tony Quinzi
1703 Salina Drive
Las Cruces, NM 88011

Florence Reeder
MS W640
Mitre Corp.
1820 Dolley Madison Boulevard
McLean VA 22102

Philip Resnick
1401 Marie Mount Hall
University of Maryland
College Park, MD 20742

Roy Reynolds
U.S. Army TRAC-WSMR
WSMR, NM 88002-5502

Stephen Robinson
Department of Ind. Eng.
U. of Wisconsin-Madison
Madison, WI 53706-1572

Robin Rose
5875 N. Jornada Road
Las Cruces, NM 88012

B. L. Rozovskii
University of Southern California
Los Angeles, CA 90089-1113

Carl Russell
JNTF/SE 730 Irwin Avenue
Schriever AFB, CO 80912-7300

Francisco Samaniego
Div. of Stat.
University of California
One Shields Avenue
Davis, CA 95616

Jayaram Sethuraman
Dept of Stat.
Florida State University
Tallahassee, FL 32306-4330

Nozer Singpurwalla
Department of Operations Research
707 22nd Street N.W.
George Washington University
Washington, DC 20052

Lounell Southard
Director
TRAC-WSMR
ATTN: ATRC-WE
WSMR, NM 88002

Marie K. Stapp
TRAC-WSMR
ATTN: ATRC-WGB
WSMR, NM 88002-5504

Douglas Tang
10809 Meadow Hill Road
Adelphi, MD

James R. Thompson
Department of Statistics
Rice University
Houston, TX 77251-1892

Clare Voss
U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783

Thomas Walker
U.S. Army Aberdeen Test Center
ATTN: STEAC-PC
Aberdeen Proving Ground, MD
21005-5059

David Webb
U.S. Army Research Laboratory
AMSRL-WM-BC
Aberdeen Proving Ground, MD
21005-5066

Edward J. Wegman
Ctr. for Comp. Stat., MS4A7
George Mason University
Fairfax, VA 22030

Sandford Weisberg
2154 Folwell Avenue
Saint Paul, MN 55708

Janyce Wiebe
New Mexico State University
Las Cruces, NM 88003

John Bart Wilburn
Optical Sciences Center
Universtiy of Arizona
Tucson, AZ 85721

Alyson Wilson
2209 Dakota Drive
Las Cruces, NM 88011-9080

Kevin Young
U.S. Army TRAC-WSMR
WSMR, NM 88002-5502

INTENTIONALLY LEFT BLANK.

NO. OF COPIES ORGANIZATION

2 DEFENSE TECHNICAL
INFORMATION CENTER
DTIC DDA
8725 JOHN J KINGMAN RD
STE 0944
FT BELVOIR VA 22060-6218

1 HQDA
DAMO FDQ
D SCHMIDT
400 ARMY PENTAGON
WASHINGTON DC 20310-0460

1 OSD
OUSD(A&T)/ODDDR&E(R)
R J TREW
THE PENTAGON
WASHINGTON DC 20301-7100

1 DPTY CG FOR RDE HQ
US ARMY MATERIEL CMD
AMCRD
MG CALDWELL
5001 EISENHOWER AVE
ALEXANDRIA VA 22333-0001

1 INST FOR ADVNCD TCHNLGY
THE UNIV OF TEXAS AT AUSTIN
PO BOX 202797
AUSTIN TX 78720-2797

1 DARPA
B KASPAR
3701 N FAIRFAX DR
ARLINGTON VA 22203-1714

1 NAVAL SURFACE WARFARE CTR
CODE B07 J PENNELLA
17320 DAHLGREN RD
BLDG 1470 RM 1101
DAHLGREN VA 22448-5100

1 US MILITARY ACADEMY
MATH SCI CTR OF EXCELLENCE
DEPT OF MATHEMATICAL SCI
MAJ M D PHILLIPS
THAYER HALL
WEST POINT NY 10996-1786

NO. OF COPIES ORGANIZATION

1 DIRECTOR
US ARMY RESEARCH LAB
AMSRL DD
J J ROCCHIO
2800 POWDER MILL RD
ADELPHI MD 20783-1145

1 DIRECTOR
US ARMY RESEARCH LAB
AMSRL CS AS (RECORDS MGMT)
2800 POWDER MILL RD
ADELPHI MD 20783-1145

3 DIRECTOR
US ARMY RESEARCH LAB
AMSRL CI LL
2800 POWDER MILL RD
ADELPHI MD 20783-1145

ABERDEEN PROVING GROUND

4 DIR USARL
AMSRL CI LP (305)

NO. OF COPIES	ORGANIZATION
1	USMA DEPT OF SYSTEMS ENGNG D BARR WEST POINT NY 10996
1	USMA DEPT OF MATH SCI MAJ P BEAVER WEST POINT NY 10996
1	J BINKLEY 3531 WHITE CT TORRANCE CA 90503
2	PENN STATE UNIV DEPT OF IND/MAN ENGNG B BIRGOREN STATE COLLEGE, PA 16802
1	B BISSINGER 281 W MAIN ST MIDDLETOWN PA 17057
2	THE AEROSPACE CORP PO BOX 92957 S BOOK M4/921 LOS ANGELES CA 90009-2957
4	R J BURGE 7134 SMOOTH PATH COLUMBIA MD 21045
1	USA TRAC-WSMR ATRC WAC P CASSADY WSMR NM 88002-5502
1	US ARMY RESEARCH LAB 2800 POWDER MILL RD J CHANDRA ADELPHI MD 20783
1	WES CEWES-ES-F 3909 HALLS FERRY RD J CLARKE VICKSBURG MS 39180

NO. OF COPIES	ORGANIZATION
1	TEXAS TECH UNIV COLLEGE OF BUS ADMIN W J CONOVER LUBBOCK TX 79409
1	P COX 2930 HUNTINGTON DR LAS CRUCES NM 88011
1	D F CRUESS 6123 CAMEL BACK LN COLUMBIA MD 21045
1	NEW MEXICO STATE UNIV G CUNNINGHAM LAS CRUCES NM 88003
1	T F CURRY 105 E VERMIJO STE 450 COLORADO SPRINGS CO 80903
1	M DANESH 126 KITTY HAWK LN HARVEST AL 35749
4	USA TRAC WSMR ATRC WAD P DEASON WSMR NM 88002-5502
5	G DUTOIT 3544 STATLER DR COLUMBUS GA 31907
1	SANDIA NATL LABS R G EASTERLING MS 0419 ALBUQUERQUE NM 87185
1	INDIANA UNIV OF PENN DEPT OF MATH 233 STRIGHT HALL D FRANK INDIANA PA 15705
1	USA TRAC WSMR ATRC WAD B GAFNER WSMR NM 88012

NO. OF COPIES	ORGANIZATION
1	D GARD 2213 LARAMIE DR LAS CRUCES NM 88011
1	NAVAL POSTGRAD SCHOOL OR DEPT D GAVER MONTEREY CA 93943
1	P GIRARD 10260 CAMPUS POINT DR SAN DIEGO CA 92121
1	UNIV OF WISCONSIN-MADISON DEPT OF STAT B HARRIS MADISON WI 53706
1	LANL J HOCHBERG MS B265 LOS ALAMOS NM 87545
1	THE BOEING COMPANY 1800 SATELLITE BLVD J HOSFORD DL3M DULUTH GA 30097
1	UNISYS CORP 1401 WILSON BLVD STE 300 S HUXEL ARLINGTON VA 22209-2306
1	UNIV OF MARYLAND T KANUNGO COLLEGE PARK MD 20742
1	USA TRAC WSMR P KINNEY WSMR NM 88002-5502
1	K KYSOR PO BOX 685 EDGEWOOD MD 21040
1	USA TRAC WSMR R LAFERRIERE WSMR NM 88002-5502

NO. OF COPIES	ORGANIZATION
5	US ARO PO BOX 12211 R LAUNER RESEARCH TRIANGLE PARK NC 27709-2211
1	C LEAKE 8101 AINSWORTH AVE SPRINGFIELD VA 22152-2431
1	USA TRAC WSMR T INN WSMR NM 88002-5502
1	NAVAL POSTGRAD SCHOOL OR DEPT T LUCAS MONTEREY CA 93943
1	APPLIED PHYSICS LAB BLDG 24 J F LYONS LAUREL MD 20723
1	UMBC DEPT MATH/STAT 1000 HILLTOP CIRCLE T MATHEW BALTIMORE MD 21250
1	MAJ K M MILLER 4513 FOXWOOD RD CHESTER VA 23831
2	USMA DEPT OF MATH SCI MAJ A NAPOLI WEST POINT NY 10996
3	NAVAL POSTGRAD SCHOOL OF DEPT LTC D OLWELL MONTEREY CA 93943
1	K NAU BOX 444 WSMR NM 88002

NO. OF COPIES	ORGANIZATION
1	T QUINZI 1703 SALINA DR LAS CRUCES NM 88011
1	MITRE CORP F REEDER MS W640 1820 DOLLEY MADISON BLVD MCLEAN VA 22102
1	UNIV OF MARYLAND 1401 MARIE MOUNT HALL P RESNICK COLLEGE PARK MD 20742
1	USA TRAC WSMR R REYNOLDS WSMR NM 88002-5502
1	UNIV OF WISCONSIN-MADISON DEPT OF IND ENGNG S ROBINSON MADISON WI 53706-1572
1	R ROSE 5875 N JORNADA RD LAS CRUCES NM 88012
2	B L ROZOVSKII UNIV OF SOUTHERN CA LOS ANGELES CA 90089-1113
5	JNTF/SE 730 IRWIN AVE C RUSSELL SCHRIEVER AFB CO 80912-7300
3	UNIV OF CALIFORNIA DIV OF STAT ONE SHIELDS AVE F SAMANIEGO DAVIS CA 95616
1	FLORIDA STATE UNIV DEPT OF STAT J SETHURAMAN TALLAHASSEE FL 32306-4330

NO. OF COPIES	ORGANIZATION
1	GWU DEPT OF OR 707 22ND ST NW N SINGPURWALLA WASHINGTON DC 20052
3	DIRECTOR ATRC WE L SOUTHARD WSMR NM 88002
1	USA TRAC WSMR ATRC WGB M K STAPP WSMR 88002-5504
5	D TANG 10809 MEADOW HILL RD ADELPHI MD
1	RICE UNIVERSITY DEPT OF STATISTICS J R THOMPSON HOUSTON TX 77251-1892
1	US ARMY RESEARCH LABORATORY 2800 POWDER MILL RD C VOSS ADELPHI MD 20783
3	GEORGE MASON UNIV CTR FOR COMP STAT E J WEGMAN MS4A7 FAIRFAX VA 22030
1	S WEISBERG 2154 FOLWELL AVE ST PAUL MN 55708
3	NEW MEXICO STATE UNIV J WIEBE LAS CRUCES NM 88003
1	UNIVERSITY OF ARIZONA OPTICAL SCIENCES CTR J B WILBURN TUCSON AZ 85721

NO. OF
COPIES ORGANIZATION

1 A WILSON
2209 DAKOTA DR
LAS CRUCES NM 88011-9080

1 USA TRAC WSMR
K YOUNG
WSMR NM 88002-5502

ABERDEEN PROVING GROUND

16 USARL
AMSRL IS CI
B BODT (5 CPS)
A BRODEEN
B BROOME
A CELMINS
AMSRLHR SD
J GRYNOVICKI (5 CPS)
AMSRL WM BC
D WEBB (3 CPS)

2 USA ATC
STEAC SL B
B J KASCHENBACH
STEAC PC
T WALKER
APG MD 21005-5059

1 AMSAA
AMXSY A
M CUSHING
APG MD 21005-5001

INTENTIONALLY LEFT BLANK.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1999	Final, 21-23 October 1998

4. TITLE AND SUBTITLE
Proceedings of the Fourth Annual U.S. Army Conference on Applied Statistics, 21-23 October 1998

5. FUNDING NUMBERS
1L162618AH80

6. AUTHOR(S)
Barry A. Bodt, Editor

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: AMSRL-IS-CD
Aberdeen Proving Ground, MD 21005-5067

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-SR-84

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

10.SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The fourth U.S. Army Conference on Applied Statistics was hosted by the U.S. Army Training and Doctrine Command (TRADOC) Analysis Center - White Sands Missile Range (TRAC-WSMR) during 21–23 October 1998. Two sites were used for the conference. The meeting began at the Corbett Center on the campus of New Mexico State University in Las Cruces and concluded at WSMR. The conference was cosponsored by the U.S. Army Research Laboratory (ARL), the U.S. Army Research Office (ARO), the U.S. Military Academy (USMA), TRAC-WSMR, the Walter Reed Army Institute of Research (WRAIR), and the National Institute for Standards and Technology (NIST). The U.S. Army Conference on Applied Statistics is a forum for technical papers on new developments in statistical science and on the application of existing techniques to Army problems. This document is a compilation of available papers offered at the conference.

14. SUBJECT TERMS
applied statistics, experimental design, statistical inference

15. NUMBER OF PAGES
214

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

INTENTIONALLY LEFT BLANK.

USER EVALUATION SHEET/CHANGE OF ADDRESS

This Laboratory undertakes a continuing effort to improve the quality of the reports it publishes. Your comments/answers to the items/questions below will aid us in our efforts.

1. ARL Report Number/Author ARL-SR-84 (Bodt, Editor) Date of Report November 1999

2. Date Report Received _____

3. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which the report will be used.) _____

4. Specifically, how is the report being used? (Information source, design data, procedure, source of ideas, etc.) _____

5. Has the information in this report led to any quantitative savings as far as man-hours or dollars saved, operating costs avoided, or efficiencies achieved, etc? If so, please elaborate. _____

6. General Comments. What do you think should be changed to improve future reports? (Indicate changes to organization, technical content, format, etc.) _____

CURRENT ADDRESS

_____
Organization

_____
Name E-mail Name

_____
Street or P.O. Box No.

_____
City, State, Zip Code

7. If indicating a Change of Address or Address Correction, please provide the Current or Correct address above and the Old or Incorrect address below.

OLD ADDRESS

_____
Organization

_____
Name

_____
Street or P.O. Box No.

_____
City, State, Zip Code

(Remove this sheet, fold as indicated, tape closed, and mail.)

(DO NOT STAPLE)