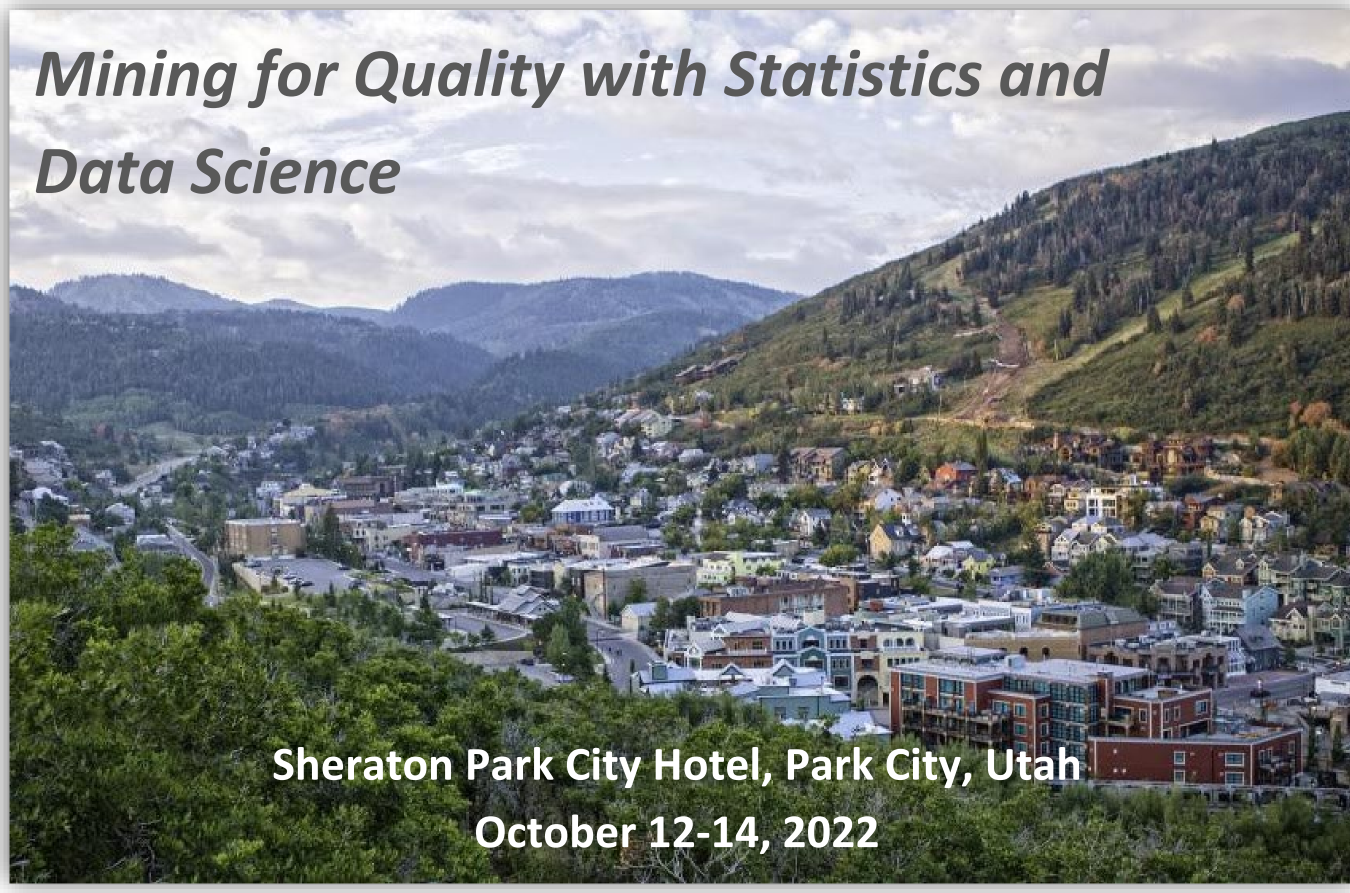
64th ANNUAL FALL TECHNICAL CONFERENCE



Mining for Quality with Statistics and Data Science

Sheraton Park City Hotel, Park City, Utah

October 12-14, 2022

Co-sponsored by:

ASQ
Excellence Through Quality™

Chemical & Process Industries Division
Statistics Division

ASA
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

Section on Physical & Engineering Sciences
Quality & Productivity Section

Dear Attendee,

It is my pleasure to welcome you to Park City, Utah, and the 2022 ASA/ASQ Fall Technical Conference! The Fall Technical Conference brings together researchers and practitioners from academia, industry, and government to discuss ways to more effectively use statistical methods for research, innovation and quality improvement. The conference is co-sponsored by the American Society for Quality (Chemical & Process Industries Division and the Statistics Division) and the American Statistical Association (Section on Physical & Engineering Sciences and the Quality & Productivity Section).

This year's conference theme, ***Mining for Quality with Statistics and Data Science***, highlights the important roles that statistics and data science play in quality. It also alludes to the historical roots of the old mining town of Park City. The technical program committee has put together an outstanding group of sessions, including topics in experimental design, reliability, statistical process control, and machine learning. Although there are plenty of traditional topics in the program, we have also incorporated sessions that are focused on data science and machine learning in quality. The program concludes on Friday afternoon with a SPES special session entitled ***Innovations in Industry 4.0*** and a reception. The technical program and the short courses on Wednesday, October 12, combine to cover an expansive range of topics in statistics and quality.

The plenary and lunch speakers are, as always, high points of the conference. Dr. Peter Parker from NASA will be the Thursday morning plenary speaker. He will discuss how organizations can benefit from ***Statistical Influence***. The Thursday lunch session is dedicated to Professor Raymond Myers for his profound contributions to the community, which will be hosted by Dr. Douglas Montgomery and Dr. William Myers. Dr. Bradley Jones will give the W. J. Youden Address on Thursday afternoon. His talk is titled ***Latin Squares, Youden Squares, Balanced Incomplete Block Designs (BIBDs) and Extensions for Industrial Application***. On Friday, we are privileged to hear from the ASA president, Dr. Kathy Ensor, with remarks on ***News from ASA and Partnerships to Establish a System for Wastewater Epidemiology for Houston***.

I would like to thank everyone involved in this year's conference for all their efforts! Thanks to the FTC Steering Committee for their support and guidance, thanks to the Technical Program Committee led by Yili Hong, and the Short Course Committee led by Steve Schuelka. Thanks as well to the other conference committee members that helped with all the other details: Sharad Prabhu as FTC Treasurer, Shane Bookhultz as Publicity Chair, John Szarka as Exhibitor & Sponsor Chair, Caleb King as Webmaster and Jie Min as the Program Book Editor. Special thanks also to Adam Pintar as the previous FTC chair.

We hope that you have a productive conference and a wonderful time in Park City. Enjoy the great presentations and networking opportunities. Also, please join us for an informal gathering in the Hospitality Suite Wednesday and Thursday nights from 8:00pm – 10:00pm in Room 340 at the Park City Sheraton hotel.

Sincerely,
Richard Warr
2022 FTC General Conference Chair

Acknowledgments

FTC Steering Committee

ASQ CPID: Stephanie DeHart
ASQ STAT: Theresa Utlaut
ASA SPES: Anne Driscoll
ASA Q&P: David Edwards

2022 FTC General Committee

General Chair: Richard Warr
Treasurer: Sharad Prabhu
Publicity: Shane Bookhultz
Exhibitors & Sponsors: John Szarka
Webmaster: Caleb King
Registration: Rick Peterson
Program Book Editor: Jie Min

2022 Technical Program Committee

ASQ CPID: Yili Hong (Chair)
ASQ STAT: Mindy Hotchkiss
ASA SPES: Jon Stallrich
ASA Q&P: Leslie M. Moore

2022 Short Course Committee

ASQ STAT: Steven Schuelka (Chair)
ASQ CPID: Jennifer Van Mullekom
ASA SPES: Matt Pratola
ASA Q&P: Brian Weaver

Conference Sponsors

ASQ Chemical & Process Industries Division

Chair: Jennifer Kensler
Chair-Elect: Amanda Yoder
Secretary: Katherine Brickey
Shewell Award: Brooke Marshall
Wilcoxon & Youden Awards: Edgar Hassler

ASQ Statistics Division

Chair: Gary Gehring
Chair-Elect: Geoff Farmer
Treasurer: Geoff Farmer
Secretary: Karen Hulting

ASA Section on Physical & Engineering Sciences

Chair: Lulu Kang
Chair-Elect: Ying Hung
Secretary & Treasurer: Emily Calseton

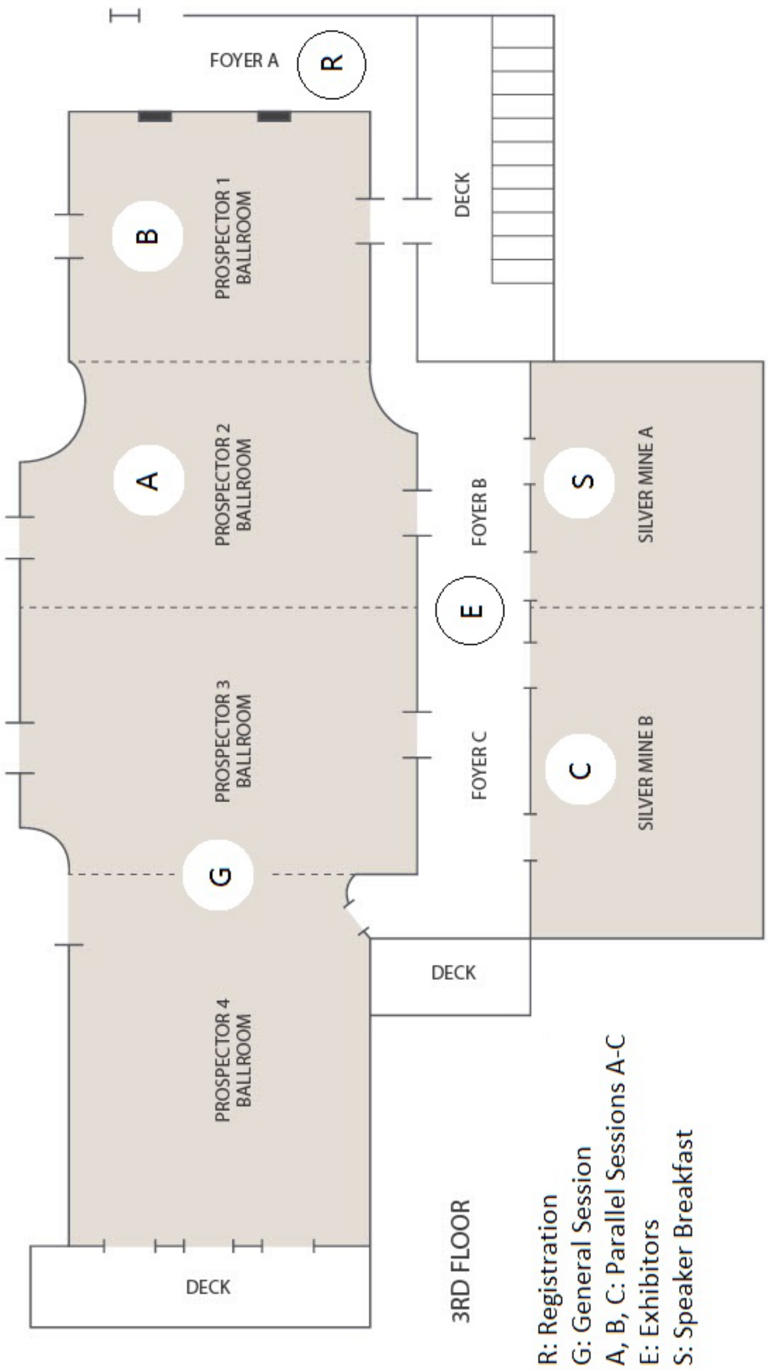
ASA Quality & Productivity Section

Chair: Xinwei Deng
Chair-Elect: Sarah Burke
Treasurer: Sharad S. Prabhu
Secretary: Annie Zangi

Corporate Sponsors

Diamond: JMP
Platinum: Procter & Gamble
Gold: Brigham Young University
Silver: Shell
Bronze: Adsurgo
Exhibitors: Stat-Ease

Facility Map and Meeting Locations



Conference Event Times & Locations

Tuesday, October 11

7:00pm – 8:00pm	Registration	Foyer A

Wednesday, October 12

7:30am – 8:30am	Registration	Foyer A
12:00pm – 12:30pm	Registration	Foyer A
5:00pm – 7:00pm	Registration	Foyer A
7:30am – 8:30am	Continental Breakfast	Foyer B&C
8:30am – 5:30pm	Short Courses	Silver Mine B
5:30pm – 6:30pm	ASQ CPID Board Meeting	Silver Mine B
8:00pm – 10:00pm	Conference Hospitality Suite	Room 340

Thursday, October 13

7:00am – 5:30pm	Registration	Foyer A
7:00am – 8:00am	Continental Breakfast	Foyer B&C
7:00am – 8:00am	Speakers' Breakfast	Silver Mine A
7:30am – 5:30pm	Exhibits	Foyer B&C
8:00am – 9:00am	Welcome & Plenary Session	Prospector 3&4
9:15am – 10:00am	Parallel Sessions 1	A: Prospector 2 B: Prospector 1 C: Silver Mine B
10:00am – 10:30am	Morning Break	Foyer B&C
10:30am – 12:00pm	Parallel Sessions 2	A: Prospector 2 B: Prospector 1 C: Silver Mine B
12:15pm – 1:45pm	Lunch	Prospector 3&4
2:00pm – 3:30pm	Parallel Sessions 3	A: Prospector 2 B: Prospector 1 C: Silver Mine B
3:30pm – 4:00pm	Afternoon Break	Foyer B&C
4:00pm – 5:00pm	W. J. Youden Address	Prospector 3&4
6:00pm – 8:00pm	JQT Editorial Review Board Dinner	Windy Ridge Café
8:00pm – 10:00pm	Conference Hospitality Suite	Room 340

Conference Event Times & Locations

Friday, October 14

7:00am – 10:00am	Registration	Foyer A
7:00am – 8:00am	Continental Breakfast	Foyer B&C
7:00am – 8:00am	Speakers' Breakfast	Silver Mine A
7:30am – 3:00pm	Exhibits	Foyer B&C
8:00am – 9:30am	Parallel Sessions 4	A: Prospector 2 B: Prospector 1 C: Silver Mine B
9:30am – 10:00am	Morning Break	Foyer B&C
10:00am – 11:30am	Parallel Sessions 5	A: Prospector 2 B: Prospector 1 C: Silver Mine B
11:45am – 1:15pm	Lunch	Prospector 3&4
1:30pm – 3:00pm	Parallel Sessions 6	A: Prospector 2 B: Prospector 1 C: Silver Mine B
3:15pm – 5:15pm	SPES Wine & Cheese Reception and Panel Discussion	Prospector 3&4

Technical Program At-A-Glance Schedule: Thursday, October 13

8:00am – 9:00am	**Welcome & Plenary Session** **Gerald J. Hahn Achievement Award** *Statistical Influence* Peter Parker, NASA Langley Prospector 3&4		
	Prospector 2	**Prospector 1**	**Silver Mine B**
9:15am – 10:00am	**1A: Profile Monitoring** *A Non-linear Mixed Model Approach for Detecting Outlying Profiles* Valeria Quevedo, Universidad de Piura **Moderator:** *James Lucas*	**1B: Staggered DOE** *A Staggered-Level Design Case Study: Staple Fiber Cutting Experiment* Peter Goos and Katherine Brickey, KU Leuven and Pfizer **Moderator:** Anne Driscoll	**1C: Text Mining** *Statistical Tools for Survey Data: A Case Study of a First Responders Survey* Adam Pintar, National Institute of Standards and Technology **Moderator:** Arman Sabbaghi
10:30am – 12:00pm	**2A: Q&P Invited** *Additive Manufacturing: A Case Study for Characterizing Variability* Lauren Wilson, Sandia National Laboratories *A Closed-Loop Machine Learning and Compensation Framework for Geometric Accuracy Control of 3D Printed Products* Arman Sabbaghi, Purdue University **Moderator:** *Lyndsay Shand*	**2B: New Designs** *Optimal Designs with Axial Values* Cameron Willden, BD *The State of Supersaturated Design and Analysis* Byran Smucker, Miami University **Moderator:** Jiayi Lian	**2C: Technometrics Invited** *Using BART to Perform Pareto Optimization and Quantify its Uncertainties* Akira Horiguchi, Duke University *Sensitivity Prewarping for Local Surrogate Modeling* Nathan Wycoff, Georgetown University **Moderator:** Roshan Joseph
12:15pm – 1:45pm	**Luncheon** *Professor Raymond H. Myers Memorial Session* Douglas Montgomery and William Myers, Arizona State University, and Procter and Gamble Prospector 3&4		
2:00pm – 3:30pm	**3A: Analytics** *Analytical Problem Solving Based on Causal, Predictive and Deductive Models* Jeroen de Mast, University of Amsterdam *Data Analytics for Decision-Making* Joanne Wendelberger, Los Alamos National Laboratory **Moderator:** Rob Goedhart	**3B: DOE and AI** *New Developments in Space Filling Designs* Christine Anderson-Cook and Lu Lu, Los Alamos National Laboratory and University of South Florida *Constructing the Large-scale Mixture Design with Application to the Evaluation of AI Algorithm* Jiayi Lian, Virginia Tech **Moderator:** Adam Pintar	**3C: Reliability** *Debunking Stress Rupture Theories Using Weibull Regression Plots* Anne Driscoll, Virginia Tech *Spatially Correlated Time-to-Event Model for Titan GPU Failures Data Under Competing Risks* Jie Min, Virginia Tech **Moderator:** Qinglong Tian
4:00pm – 5:00pm	**Youden Address** *Latin Squares, Youden Squares, Balanced Incomplete Block Designs (BIBDs) and Extensions for Industrial Application* Bradley Jones, JMP Prospector 3&4		

Technical Program At-A-Glance Schedule: Friday, October 14

	Prospector 2	Prospector 1	Silver Mine B
8:00am – 9:30am	**4A: STAT Invited** *Covariate Software Vulnerability Discovery Model to Support Cybersecurity Test & Evaluation* Lance Fiondella, University of Massachusetts Dartmouth *Rethinking Software Fault Tolerance* Kishor Trivedi, Duke University **Moderator:** Mindy Hotchkiss	**4B: CPID Invited** *D-optimal Mixture Designs for Ordinal Responses* Rong Pan, Arizona State University *Specifying Prior Distributions in Reliability Applications* Qinglong Tian, University of Waterloo **Moderator:** Yili Hong	**4C: JQT Invited** *Estimating Pure-Error from Near Replicates in Design of Experiments* Caleb King, JMP *Monitoring Proportions with Two Components of Common Cause Variation* Rob Goedhart, University of Amsterdam **Moderator**: Allison Jones-Farmer
10:00am – 11:30am	**5A: SPES Invited** *Data Splitting* Roshan Joseph, Georgia Tech *Active Learning for Deep Gaussian Process Surrogates* Annie Sauer, Virginia Tech **Moderator:** Jon Stallrich	**5B: Response Surface Design & Optimization** *Advances in Orthogonal Minimally Aliased Response Surface (OMARS) Designs: Designs with Two-level Categorical Factors, with a Blocking Factor, Strong OMARS Designs* José Núñez Ares, KU Leuven *Model Robust Response Surface Optimization for Mixed Quantitative and Qualitative Factors* Gautham Sunder, University of Minnesota **Moderator:** Byran Smucker	**5C: QE Invited** *Incorporating Uncertainty for Enhanced Leadership Scoring and Ranking in Data Competitions* Lu Lu and Christine Anderson-Cook, University of South Florida and Los Alamos National Laboratory *Understanding and Addressing Complexity in Problem Solving* Willis Jensen, W. L. Gore & Associates, Inc. **Moderator:** David Edwards
11:45am – 1:15pm	**Luncheon** *News from ASA and Partnerships to Establish a System for Wastewater Epidemiology for Houston* Kathy Ensor, ASA President Prospector 3&4		
1:30pm – 3:00pm	**6A: Statistical Engineering** *Modeling Semiconductor Yield Using Statistical Engineering - A Case Study* Dana Krueger, Roadmap Research Global *Fleet Predictions Using Bayesian Multilevel Models: An Example using Targeting Pod Accuracy* Todd Remund, Northrop Grumman Space **Moderator:** Willis Jensen	**6B: Machine Learning & Optimization** *Tuning Parameter Selection for Variable Selection via an R2 Metric* Jon Stallrich, North Carolina State University *Bayesian Optimization via Barrier Functions* Tony Pourmohamad, Genentech **Moderator:** Jie Min	**6C: SPC** *Applied vs Academic Approaches to Statistical Process Monitoring (SPM)* James Lucas, J. M. Lucas and Associates *A Consistent Data Model for Different Data Granularity in Control Charts* Scott Grimshaw, Brigham Young University **Moderator:** Valeria Quevedo
3:15pm – 5:15pm	**Wine & Cheese Reception with SPES Special Panel Session** *Innovations in Industry 4.0* Arman Sabbaghi (Purdue), Laura Freeman (Virginia Tech), Nathan Wycoff (Georgetown University), and Mindy Hotchkiss (Aerojet Rocketdyne) **Moderator:** Jon Stallrich Prospector 3&4		

Shewell Award Evaluation Process

About the Award

The Shewell Award sponsored by ASQ CPID is based on two broad categories of criteria:

1. Oral presentation (60%) as rated by attendee's evaluation as provided on this form;
2. Written material available at talk (40%), rated by a panel of judges based on quality of subject matter, scientific merit, applicability and originality.

Two Ways to Evaluate

1. **Use your smart phone and the QR Code:** Scan the QR code below with your smart phone. Most iPhone and Android smartphones can navigate you directly to the Form just using your camera. Alternatively, you can use a QR scanning app that you have already installed on your phone.

2. **Navigate to the following web link and bookmark it in your smartphone browser.**

https://forms.office.com/r/g5qJRis1iF

Need help? See one of the FTC Student Grant Winners for assistance.

Corporate Sponsors

Special thanks to our sponsors for helping to continue FTC's great tradition!

Diamond Sponsor

JMP® statistical discovery software is the tool of choice for scientists, engineers and other data explorers worldwide. JMP links dynamic data visualization with powerful statistics, in memory and on the desktop. Interactive and visual, JMP reveals insights that raw tables of numbers or static graphs tend to hide. JMP is a SAS company. Get details at www.jmp.com.

Platinum Sponsor

Procter & Gamble serves consumers around the world with one of the strongest portfolios of trusted, quality, leadership brands, including Always®, Bounty®, Charmin®, Crest®, Dawn®, Downy®, Febreze®, Gain®, Gillette®, Head & Shoulders®, Lenor®, Olay®, Oral-B®, Pampers®, Pantene®, SK-II®, Tide® and Vicks®. The P&G community includes operations in approximately 70 countries worldwide. Visit www.pg.com.

Gold Sponsor

Brigham Young University is a private research university in Provo, Utah sponsored by the Church of Jesus Christ of Latter-day Saints. BYU offers 186 undergraduate, 64 master's, and 26 doctoral programs with a population of approximately 35,000 students. Its undergraduate statistics program is one of the largest in the US.

Silver Sponsor

Shell is a global group of energy and petrochemical companies with 82,000 employees and operations in more than 70 countries. We use advanced technologies and take an innovative approach to help build a sustainable energy future. Visit www.shell.com.

Bronze Sponsor

Adsurgo is a Service Disabled Veteran Owned Small Business who provides applied statistics consulting services, tailored training workshops, and software solutions. Our passion is solving interesting, challenging, and meaningful problems in collaborative engagements with our clients across the medical, bio-pharma, financial services, energy, consumer products, technology, and government industries. Visit www.adsurgo.com.

Exhibitors

Special thanks to our exhibitors for helping to continue FTC's great tradition!

Our exhibitors will be showcasing their products on Thursday and Friday in Foyer B&C. Be sure to stop by to check out their latest offerings!

Stat-Ease, Inc. provides software and training for design of experiments (DOE). Industrial researchers worldwide use Design-Expert® software (DX) to make the most from every experiment. For power users, Stat-Ease® 360 software expands the horizon of DX (all features included) to Python scripting and other advanced tools. Get details at www.statease.com.

Full Day Courses: Wednesday, October 12

Short courses will be held on Wednesday, October 12, from 8:30am to 5:30pm.

Advanced Quality Control Methods: CUSUM and EWMA Procedures

James Lucas
J. Lucas & Associates
JamesM.Lucas@verizon.net

This course will discuss CUSUM (Cumulative Summation) and EWMA (Exponentially Weighted Moving Average) Procedures for Quality Control Applications. The presenter is a consultant and researcher who was an integral part of the largest known implementation of CUSUM procedures. He has guided many successful implementations of these procedures and was the developer of many of the most used enhancements. This will be a “hands on” course with discussions of real problems and workshops to practice using the techniques. You will come from this course knowing how to use these procedures to monitor a process (and to detect problems). The course knowledge will enhance your data analysis skills. Both variables and count data (including procedures for rare events) will be discussed. Comparisons with classical Shewhart procedures will be given throughout the course to show where these advanced procedures have large benefits. The prerequisite for the course is a quality control course or some implementation experience.

Full Day Courses: Wednesday, October 12

Short courses will be held on Wednesday, October 12, from 8:30am to 5:30pm.

Methods for Designing & Analyzing Mixture Experiments

Greg Piepel
MIXSOFT
mixsoft@aol.com

Mixture experiments involve changing the proportions of the components of a mixture that make up a product and then observing the resulting changes in the product's characteristics. The proportions of the components in the mix cannot be varied independently (as in factorial experiments) because they must sum to 1.0 for each run in the experiment. Mixture experiments are very useful in many product development areas, including foods and drinks, plastics, alloys, ceramics and glass, gasoline blending, fertilizers, textile fibers, concrete, drugs, and many others.

The short course will provide an overview of various approaches and methods used in designing mixture experiments and analyzing the resulting data. Designs for simplex-shaped and irregular-shaped regions (the latter resulting from placing additional constraints on the component proportions) will be covered. The various types of mixture models that can be fitted to mixture data will be covered, as will graphical techniques for interpreting component effects. Methods for including process variables and/or a total amount variable in mixture experiments will be discussed. Graphical and analytic methods for developing mixtures with optimum properties will also be covered. Numerous examples will be used to illustrate the topics discussed. The course is designed for anyone (statistician or non-statistician) wanting to know about statistical methods for designing mixture experiments and analyzing the resulting data. Prerequisites are an understanding of elementary statistics concepts and some previous exposure to experimental design and least squares regression.

Half Day Courses: Wednesday, October 12

This short course will be held on Wednesday, October 12, from 8:30am to 12:30pm.

Functional Data Analysis and Its Applications

Pang Du
Virginia Tech
pangdu@vt.edu

This course aims to introduce the modern field of functional data analysis to a general audience with an emphasis on how the relevant techniques can be applied to real examples. As a generalization of the traditional data concepts from numbers and vectors of numbers to curves and surfaces, functional data has attracted much attention from statisticians and found many interesting applications in various fields in the past decades. The course will start with the introduction of real examples of functional data. Based on these examples, common functional data analysis techniques such as function smoothing, functional principal component analysis, and functional linear regression models will be presented. R implementation of these techniques will be introduced and demonstrated.

Half Day Courses: Wednesday, October 12

This short course will be held on Wednesday, October 12, from 1:30pm to 5:30pm.

Non-parametric Approaches to Uncertainty and Hypothesis Tests

Robert Richardson
Brigham Young University
richardson@stat.byu.edu

This course is aimed to highlight several ways that non-parametric methods can be used to derive answers from data where other parametric methods are not readily available. We will cover the basic concepts of bootstrapping and permutation tests and compare them against parametric methods for scenarios where both can be used. A variety of scenarios will be discussed where the non-parametric methods are more feasible. We will show how to use these methods to get confidence intervals and perform hypothesis tests for parameter estimates, predictions, or model-derived statistics. We will also show how to use these methods to perform more advanced hypothesis tests on data such as testing for independence between variables, correlation structures of dependent data, and randomness of missing values.

Thursday, October 13

Welcome & Plenary Session

Presentation of Gerald J. Hahn Achievement Award

Thursday, October 13, 8:00 – 9:00am ASA Q&P Presiding
Prospector 3&4

Statistical Influence

Peter Parker
NASA Langley
peter.a.parker@nasa.gov

Abstract: Advancing the spread and practice of statistics enhances an organization's ability to successfully achieve their mission. While there may be leadership mandates to employ statistical methods, often the spread of statistical concepts flourishes more effectively through the practice of statistical influence. At first glance, the term influence may seem to imply a passive and unenthusiastic posture toward promoting organizational change. However, a classical definition states that, "In a general sense, influence denotes power whose operation is invisible and known only by its effects, or a power whose cause and operation are unseen." This definition articulates the powerful and yet subtle aspects of influence that embodies the theme of this presentation. Stated plainly, powerful statistical concepts become more widely known and engrained primarily through demonstrated impact; a strategy known only by its effects. In this presentation, elements of statistical influence are exemplified through practice at NASA.

Reference:
Webster, Noah (1828) American Dictionary of the English Language, United States, Editorium.

Biography: Dr. Peter Parker is Team Lead for Advanced Measurement Systems at the National Aeronautics and Space Administration's Langley Research Center in Hampton, Virginia. He serves as an Agency-wide statistical expert across NASA's Aeronautics, Exploration, and Science mission directorates to infuse statistical thinking, engineering, and methods. His expertise is in collaboratively integrating research objectives, measurement sciences, modeling and simulation, and experimental test design to produce actionable knowledge that enable rigorous decision-making in aerospace research and development. After eight years in private industry, Dr. Parker joined NASA Langley Research Center in 1997.

He holds a B.S. in Engineering, a M.S. in Applied Physics and Computer Science, and a M.S. and Ph.D. in Statistics from Virginia Tech. He is a licensed Professional Mechanical Engineer in the Commonwealth of Virginia. Dr. Parker is an Associate Fellow of the American Institute of Aeronautics and Astronautics, and Senior Member of the American Society for Quality and the American Statistical Association. Dr. Parker is Editor Emeritus of Quality Engineering, past-Chair of ASQ's Publication Management Board and Journal Editor's Committee. He currently serves as Chair of the International Statistical Engineering Association.

Notes

Session 1A: Profile Monitoring

Thursday, October 13, 9:15 – 10:00am Moderator: James Lucas
Prospector 2

A Non-linear Mixed Model Approach for Detecting Outlying Profiles

Valeria Quevedo
Universidad de Piura
valeria.quevedo@udep.edu.pe
Additional Author: Geoff Vining

Abstract: In parametric non-linear profile modeling, mapping the impact of the model parameters to a single metric is essential. The profile monitoring literature suggests using multivariate T^2 statistics to monitor the stability of the parameters simultaneously. These approaches focus on the estimated parameters of the non-linear model and treat them as separate but correlated quality characteristics of the process. As a result, they do not take full advantage of the model structure. We propose a procedure based on a non-linear mixed model that takes into account the proper variance-covariance structure. This procedure recognizes and uses the correlations among the estimates of the fixed parameters created by the model. The proposed method is based on the concept of the externally studentized residual. We test whether a given profile is sufficiently different from the other profiles based on a non-linear mixed model. Our results from our simulation show that our method appears to perform better than the classical control chart based on the T^2 statistics. We also apply our approach to an aquaculture process on a shrimp farm which monitors its shrimp weight from over 300 pounds each year.

Notes

Session 1B: Staggered DOE

Thursday, October 13, 9:15 – 10:00am Moderator: Anne Driscoll
Prospector 1

A Staggered-Level Design Case Study: Staple Fiber Cutting Experiment

Peter Goos
KU Leuven
peter.goos@kuleuven.be

Katherine Brickey
Pfizer
Katherine.Brickey@pfizer.com

Abstract: Real-world experimental designs often require constraints on randomization. Time, cost, or equipment limitations prevent some factors from being independently reset at frequent intervals. Staggered-level designs provide a flexible structure for experiments with more than one hard-to-change factor by allowing the hard-to-change factor levels to be reset at different points in time throughout the design. In this case study, a staggered-level designed experiment was completed involving five experimental factors. Two hard-to-change factors were reset at staggered intervals during the experiment. This presentation will describe the experimental planning and the analysis of the results of the first documented real-world staggered-level experiment. It will also discuss the advantages of staggered-level designs compared to traditional split and split-split plot designs.

Notes

Session 1C: Text Mining

Thursday, October 13, 9:15 – 10:00am Moderator: Arman Sabbaghi
Silver Mine B

Statistical Tools for Survey Data: A Case Study of a First Responders Survey

Adam Pintar
National Institute of Standards and Technology
adam.pintar@nist.gov
Additional Authors: Kerrianne Buchanan and Yee-Ying Choong

Abstract: This presentation examines a survey conducted by researchers at the National Institute of Standards and Technology (NIST). The survey is of first responders, e.g., firefighters, EMS, police, and 9-1-1 dispatch, and its purpose was to assess first responders' experiences with communication technology, e.g., radios. From the perspective of a statistical analysis, there are some non-standard and interesting aspects.

First, one of the demographic questions requested the respondent's title. The question required an open text response. One research topic of interest is if respondents with chief/managerial roles think differently about technology than those on the frontlines. To consider that topic, it was necessary to code the open text responses into categories, which would typically be done manually by the researchers for each of the more than 7,000 respondents. It was possible to save hours of researcher time by applying "off-the-shelf" tools in natural language processing (word embeddings) and machine learning (random forests). A similar process could be useful for categorizing user or product experience feedback.

A second interesting aspect is a type of data that seem uncommon, traditionally, in engineering and physical science settings, but could be very applicable to new product development. A respondent is presented with a list of *K* choices, and is asked to rank their top *l* selections. Given the rankings of *n* respondents, the goal is to estimate the group preference of each choice. A few probability models are applicable to the problem. The Plackette-Luce model, used in this work, will be reviewed, and its ability to highlight group differences in technology preferences shown.

The presentation will conclude with some key takeaways for the NIST researchers. For example, those with chief/managerial roles do indeed seem to think differently about technology than those on the frontlines, with the price tag of that technology receiving greater scrutiny.

Notes

Session 2A: Q&P Invited Session

Thursday, October 13, 10:30am – 12:00pm Moderator: Lyndsay Shand
Prospector 2

Additive Manufacturing: A Case Study for Characterizing Variability

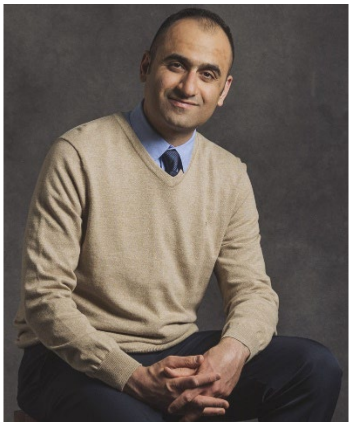
Lauren Wilson
Sandia National Laboratories
lcwilso@sandia.gov
Additional Authors: Manuel Martinez, Leslie Moore, and Joshua Yee

Abstract: Additive manufacturing (AM) provides production advantages with respect to cost, weight, and design complexities. The relative newness of AM compared to conventional methods and its known variability carries non-trivial risk. Building metal AM components with dimensional requirements is of interest for a high-risk application. A statistical experimental design & analysis evaluates the effect of various factors on component acceptance criteria, which demonstrates methodology to provide quantitative evidence in support of an alternative manufacturing process. Analysis includes common data visualization techniques novel to AM process exploration, quantification of margin & uncertainties (QMU) to ensure product quality relative to dimensional limits, and estimation of variance components (within- and between-batch) to inform production sampling and future AM studies.

Notes

Session 2A: Q&P Invited Session

Thursday, October 13, 10:30am – 12:00pm Moderator: Lyndsay Shand
Prospector 2



A Closed-Loop Machine Learning and Compensation Framework for Geometric Accuracy Control of 3D Printed Products

Arman Sabbaghi
Purdue University
sabbaghi@purdue.edu
Additional Author: Wenbin Zhu

Abstract: Additive manufacturing (AM) systems enable direct printing of three-dimensional (3D) physical products from computer-aided design (CAD) models. Despite the many advantages that AM systems have over traditional manufacturing, one of their significant limitations that impedes their wide adoption is geometric inaccuracies, or shape deviations between the printed product and the nominal CAD model. Machine learning for shape deviations can enable geometric accuracy control of 3D printed products via the generation of compensation plans, which are modifications of CAD models informed by the machine learning algorithm that reduce deviations in expectation. However, existing machine learning and compensation frameworks cannot accommodate deviations of fully 3D shapes with different geometries. The feasibility of existing frameworks for geometric accuracy control is further limited by resource constraints in AM systems that prevent the printing of multiple copies of new shapes. We present a closed-loop machine learning and compensation framework that can improve geometric accuracy control of 3D shapes in AM systems. Our framework is based on a Bayesian extreme learning machine (BELM) architecture that leverages data and deviation models from previously printed products to transfer deviation models, and more accurately capture deviation patterns, for new 3D products. The closed-loop nature of compensation under our framework, in which past compensated products that do not adequately meet dimensional specifications are fed into the BELMs to re-learn the deviation model, enables the identification of effective compensation plans and satisfies resource constraints by printing only one new shape at a time. The power and cost-effectiveness of our framework are demonstrated with two validation experiments that involve different geometries for a Markforged Metal X AM machine printing 17-4 PH stainless steel products. As demonstrated in our case studies, our framework can reduce shape inaccuracies by 30% to 60% (depending on a shape's geometric complexity) in at most two iterations, with three training shapes and one or two test shapes for a specific geometry involved across the iterations. Ultimately, our closed-loop machine learning and compensation framework provides an important step towards accurate and cost-efficient deviation modeling and

compensation for fully 3D printed products using a minimal number of printed training and test shapes, and thereby can advance AM as a high-quality manufacturing paradigm.

Notes

Session 2B: New Designs

Thursday, October 13, 10:30am – 12:00pm Moderator: Jiayi Lian
Prospector 1

Optimal Designs with Axial Values

Cameron Wilden
BD
ccwillden@gmail.com
Additional Author: Willis Jensen

Abstract: We introduce a modification to the coordinate-exchange algorithm for generating optimal experimental designs that incorporates off-face axial value runs similar to a central composite design (CCD). This improvement addresses a weakness of current optimal designs relative to classical designs: the superior properties of CCDs with off-face axial values relative to equal-sized optimal designs that are constrained to the cuboidal experimental design region. CCDs tend to have significantly less collinearity among quadratic effects, which result in higher power for quadratic terms, better D-efficiency, and lower average prediction variance (i.e. I-optimality). Optimal designs offer greater flexibility in design size, model specification, and the ability to incorporate categorical factors. By incorporating axial values into an optimal design algorithm, the strengths of both approaches can be combined in a single design that generally outperforms both CCDs and current optimal designs.

Notes

Session 2B: New Designs

Thursday, October 13, 10:30am – 12:00pm Moderator: Jiayi Lian
Prospector 1

The State of Supersaturated Design and Analysis

Byran Smucker
Miami University
smuckebj@miamioh.edu
Additional Authors: Maria Weese, Jon Stallrich, David Edwards, and Kade Young

Abstract: Supersaturated designs (SSDs) try to pull off what seems impossible: Identify important effects when there are more factors than experimental runs. Despite the vast amount of literature on the topic, there is little record of their use in practice. We contend this imbalance is due to conflicting recommendations regarding SSD use in the literature as well as the designs' inabilities to meet practitioners' analysis expectations. In this talk, we learn about some practitioner concerns and expectations via an informal questionnaire, discuss and compare two recent SSDs that pair a design construction method with a particular analysis method, and introduce some new results that provide a way to directly assess the quality of SSDs that will be analyzed using the LASSO.

Notes

Session 2C: Technometrics Invited Session

Thursday, October 13, 10:30am – 12:00pm Moderator: Roshan Joseph
Silver Mine B

Using BART to Perform Pareto Optimization and Quantify its Uncertainties

Akira Horiguchi
Duke University
akira.horiguchi@duke.edu

Abstract: Techniques to reduce the energy burden of an industrial ecosystem often require solving a multiobjective optimization problem. However, collecting experimental data can often be either expensive or time-consuming. In such cases, statistical methods can be helpful. This article proposes Pareto Front (PF) and Pareto Set (PS) estimation methods using Bayesian Additive Regression Trees (BART), which is a nonparametric model whose assumptions are typically less restrictive than popular alternatives, such as Gaussian Processes (GPs). These less restrictive assumptions allow BART to handle scenarios (e.g., high-dimensional input spaces, nonsmooth responses, large datasets) that GPs find difficult. The performance of our BART-based method is compared to a GP-based method using analytic test functions, demonstrating convincing advantages. Finally, our BART-based methodology is applied to a motivating engineering problem. Supplementary materials, which include a theorem proof, algorithms, and R code, for this article are available online.

Notes

Session 2C: Technometrics Invited Session

Thursday, October 13, 10:30am – 12:00pm Moderator: Roshan Joseph
Silver Mine B

Sensitivity Prewarping for Local Surrogate Modeling

Nathan Wycoff
Georgetown University
nathanbrwycoff@gmail.com

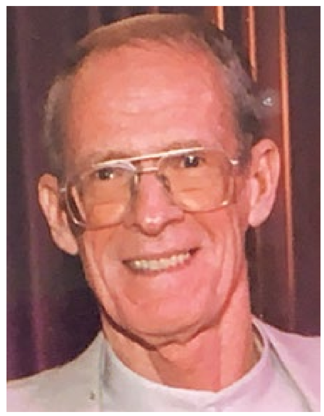
Abstract: In the continual effort to improve product quality and decrease operations costs, computational modeling is increasingly being deployed to determine feasibility of product designs or configurations. Surrogate modeling of these computer experiments via local models, which induce sparsity by only considering short range interactions, can tackle huge analyses of complicated input–output relationships. However, narrowing focus to local scale means that global trends must be relearned over and over again. In this article, we propose a framework for incorporating information from a global sensitivity analysis into the surrogate model as an input rotation and rescaling preprocessing step. We discuss the relationship between several sensitivity analysis methods based on kernel regression before describing how they give rise to a transformation of the input variables. Specifically, we perform an input warping such that the "warped simulator" is equally sensitive to all input directions, freeing local models to focus on local dynamics. Numerical experiments on observational data and benchmark test functions, including a high-dimensional computer simulator from the automotive industry, provide empirical validation.

Notes

Luncheon

Thursday, October 13, 12:15 – 1:45pm ASQ CPID Presiding
Prospector 3&4

Professor Raymond H. Myers Memorial Session



Raymond "Ray" Harold Myers
1937 - 2020

Ray Myers was a member of the Virginia Tech statistics faculty for 44 Years. He made many contributions in the areas of experimental design, response surface methodology, linear models and generalized linear models, authoring many influential research papers and 6 books. He was honored by the American Society for Quality for his body of work with the Shewhart Medal in 1998. He served as an editorial board member for the *Journal of Quality Technology* and as an associate editor for *Technometrics*. He was a Fellow of the American Statistical Association and an elected member of the International Statistical Institute.

Ray is remembered by many as an outstanding, indeed gifted teacher. His classroom lectures were always insightful and brilliant. He received the William E. Wine Award for Teaching Excellence at Virginia Tech in 1980 and was named Virginia Professor of the Year in 1985. He was a member of the Virginia Tech Academy of Teaching Excellence. He was always involved in guiding others in the use of statistics in their own research, serving as the director of the Virginia Tech Statistical Consulting Center.

Ray's classroom teaching was unequaled in the view of most of his students, but he was also an outstanding mentor. He directly advised 42 doctoral students and served on the dissertation committee for many more. There are five Virginia Tech "Hokies" who have received the Shewhart Medal, including Ray, but he was an important mentor to all of the other four. All of the other four recipients readily acknowledge that they would have never received their Shewhart Medals without him.

Ray Myers was a wonderful teacher, mentor, colleague and friend to many. He will be missed by our community.

Presented by:

Douglas Montgomery
Arizona State University
Doug.Montgomery@asu.edu

William Myers
Procter and Gamble
myers.wr@pg.com

Biographies of Presenters: Dr. Douglas C. Montgomery is Regent's Professor of Industrial Engineering and Statistics and ASU Foundation Professor of Engineering at Arizona State University. He holds BSIE, MS and Ph.D. degrees from Virginia Tech. Dr. Montgomery's professional interests focus on industrial statistics, including design of experiments, quality and reliability engineering. Dr. Montgomery is an author of 15 books, including Design and Analysis of Experiments, 10th edition, and Design of Experiments: A Modern Approach. Professor Montgomery has authored over 280 papers that have appeared in many archival journals, such as *Journal of Quality Technology, Technometrics,* and *Management Science*. He is currently one of the Chief Editors of *Quality and Reliability Engineering International* and is a former Editor of the *Journal of Quality Technology*. Dr. Montgomery is a Fellow of the American Statistical Association, an Honorary Member of the American Society for Quality, a Fellow of the Royal Statistical Society, a Fellow of the Institute of Industrial and Systems Engineers, an Elected Member of the International Statistical Institute and an Academician of the International Association for Quality. His recognition awards include the Shewhart Medal, the Distinguished Service Medal, the William G. Hunter Award, the Brumbaugh Award, and the Shewell Award (twice) from the American Society for Quality, the Deming Lecture Award from the American Statistical Association, the George Box Medal from ENBIS (European Network for Business and Industrial Statistics), the Greenfield Medal from the Royal Statistical Society and the Ellis R. Ott Award.

Dr. William (Bill) Myers is a Principal Statistician at the Procter & Gamble Company. Over his 25+ years at Procter & Gamble, he has applied statistics to further innovation, productivity, and quality. This has facilitated solving complex problems for many of Procter & Gamble's Billion Dollar Brands. His areas of expertise are Design and Analysis of Experiments, Predictive Modeling, and Computer Experiments. He is actively involved in collaborative research with academia to develop methods to solve critical business problems at Procter & Gamble. This has resulted in publications in peer-reviewed journals (e.g., *Technometrics* and *Journal of Quality Technology*). Over the years he has been actively involved in the ASA sections – *Quality & Productivity* and *Physical and Engineering Sciences*. Bill leads the Corporate Statistics Training at Procter & Gamble, which involves developing and teaching internal courses (e.g., Applied Statistics, Design & Analysis of Experiments, Data Mining & Machine Learning) and creating curricula for the R&D organization. He is also an Adjunct Faculty at the University of Cincinnati.

Session 3A: Analytics

Thursday, October 13, 2:00 – 3:30pm Moderator: Rob Goedhart
Prospector 2

Analytical Problem Solving Based on Causal, Predictive and Deductive Models

Jeroen de Mast
University of Amsterdam
j.demast@uva.nl
Additional Authors: Wim P. M. Nuijten and Daniel Kapitan

Abstract: Many approaches for solving problems in business and industry are based on analytics and statistical modelling. Analytical problem solving is driven by the modelling of relationships between dependent (Y) and independent (X) variables, and we discuss three frameworks for modelling such relationships: cause-and-effect modelling, popular in applied statistics and beyond, correlational predictive modelling, popular in machine learning, and deductive (first-principles) modelling, popular in business analytics and operations research. We aim to explain the differences between these types of models, and flesh out the implications of these differences for study design, for discovering potential X/Y relationships, and for the types of solution patterns that each type of modelling could support. We use our account to clarify the popular descriptive-diagnostic-predictive-prescriptive analytics framework, but extend it to offer a more complete model of the process of analytical problem solving, reflecting the essential differences between causal, predictive and deductive models.

Notes

Session 3A: Analytics

Thursday, October 13, 2:00 – 3:30pm
Moderator: Rob Goedhart
Prospector 2

Data Analytics for Decision-Making

Joanne Wendelberger
Los Alamos National Laboratory
joanne@lanl.gov

Abstract: Many organizations collect vast amounts of data for various purposes. While data streams are often collected with a specific end goal in mind, they can also provide valuable information for broader initiatives and improved decision-making. Standard analyses can often be enhanced by incorporating additional information from previously underutilized data sources into the modeling process. The impact of individual analyses can be increased by developing reusable analysis processes with associated computational infrastructure and workflows. The integration of heterogeneous data from multiple sources can pose a variety of challenges in terms of data acquisition, storage, processing, analysis, and ongoing delivery of results. This presentation will discuss the development of an integrated analytics capability that leverages the power of statistical methods with access to data from multiple sources. Examples of analysis applications that have been developed to provide dynamic results informed by disparate data streams will also be discussed.

Notes

Session 3B: DOE and AI

Thursday, October 13, 2:00 – 3:30pm Moderator: Adam Pintar
Prospector 1

New Developments in Space Filling Designs

Christine Anderson-Cook
Los Alamos National Laboratory
candcook@gmail.com

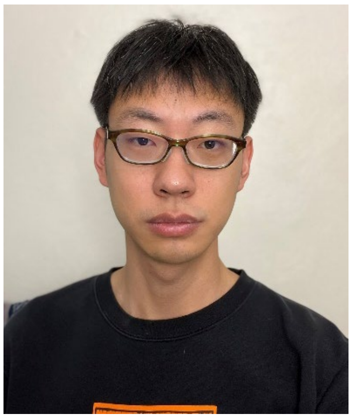
Lu Lu
University of South Florida
lulu1@usf.edu

Abstract: Space filling designs have been broadly used for data collection in computer experiments. Traditional they have focused on uniform spacing throughout the input space. In this talk, we introduce two new recent developments for more flexible construction of these designs. They allow existing information about the study to be strategically incorporated into the construction designs. The non-uniform space filling (NUSF) designs allow the users to customize the regions of emphasis with a varied degree of concentration of points. These designs are desirable when the user wishes to focus on regions of interesting features, larger variability, or larger discrepancy between the computer model and the physical experimental data. The input-response space filling (IRSF) designs allow the consideration of the spread of design points in the input space as well as the estimated response values across its anticipated range. By constructing a Pareto front, the IRSF approach provides a suite of potential designs with varied emphasis of the input and response spaces. The methods and step-by-step implementation process are illustrated with examples to demonstrate their flexibility to match experimental goals.

Notes

Session 3B: DOE and AI

Thursday, October 13, 2:00 – 3:30pm Moderator: Adam Pintar
Prospector 1



Constructing the Large-scale Mixture Design with Application to the Evaluation of AI Algorithm

Jiayi Lian
Virginia Tech
lianjiayi@vt.edu
Additional Authors: L. Freeman, Y. Hong, X. Deng, K. Choi, A. Hu, and B. Veeramani

Abstract: Artificial intelligence (AI) algorithms provide state-of-art performance in a variety of applications. However, the performance of AI is reliant on informative datasets. In other words, these algorithms are vulnerable to data quality issues, such as mislabeled training data, out-of-distribution test data, class imbalance in both training and test data, adversarial attacks. The safety and robustness of AI applications must be evaluated and tested. In this work, we present a new framework based on experimental design to investigate the performance of AI classification algorithms. The performance of AI algorithms can be affected by a variety of factors, including the algorithm's hyperparameters, data types, class proportions in training and test data (unequal proportions of classes), etc. Specifically, a space-filling design based on projection criterion is developed with an efficient construction algorithm. Our approach to construct the design of experiments can have a better capacity to comprehensively reveal the performance of AI algorithms. We gathered experimental AI algorithm performance data and conducted a statistical analysis to better understand how these factors influence the robustness of AI algorithms. Our framework can be used to evaluate the effects of factors on the AI algorithm robustness, sensitivity of the hyperparameters, to compare and assess the performance of AI algorithms against data quality.

Notes

Session 3C: Reliability

Thursday, October 13, 2:00 – 3:30pm Moderator: Qinglong Tian
Silver Mine B

Debunking Stress Rupture Theories Using Weibull Regression Plots

Anne Driscoll
Virginia Tech
adriscoll@vt.edu
Additional Author: Geoff Vining

Abstract: As statisticians, we are always working on new ways to explain statistical methodologies to non-statisticians. It is in this realm that we never underestimate the value of graphics and patience! In this presentation, we present a case study that involves stress rupture data where a Weibull regression is needed to estimate the parameters. The context of the case study results from a multi-stage project supported by NASA's Engineering Safety Center (NESC) where the objective was to assess the safety of composite overwrapped pressure vessels (COPVs). The analytical team was tasked with devising a test plan to model stress rupture failure risk in carbon fiber strands that encase the COPVs with the goal of understanding the reliability of the strands at use conditions for the expected mission life. While analyzing the data, we found that the proper analysis of the data contradicts accepted theories about the stress rupture phenomena. In this talk, we will introduce ways to graph the stress rupture data to better explain the proper analysis of the model and also explore assumptions.

Notes

Session 3C: Reliability

Thursday, October 13, 2:00 – 3:30pm — Moderator: Qinglong Tian
Silver Mine B

Spatially Correlated Time-to-Event Model for Titan GPU Failures Data Under Competing Risks

Jie Min
Virginia Tech
jiem@vt.edu
Additional Authors: Yili Hong and William Q. Meeker

Abstract: Graphics processing units (GPUs) are widely used in high-performance computing (HPC), and the reliability of GPU is of interest for the overall reliability of HPC systems. The Cray XK7 Titan supercomputer was one of the top ten supercomputers in the world. The failure times of more than 19,000 GPUs in Titan were recorded and previous research shows that the failure time of GPU may be affected by the GPU's position inside the supercomputer. In this paper, we conduct in-depth statistical modeling of the effect of positions on GPU failures under competing risks with covariates and spatial correlated random effects. In particular, two major failure types of GPUs in Titan are considered, the positions of GPUs inside each cabinet are modeled as covariates, and the positions of cabinets are modeled as spatially correlated random effects. We use the powered-exponential covariance function to construct the spatial random effects' covariance matrix and estimate the correlation of random effects between two failure modes. The Bayesian method is used for statistical inference, and posterior samples are obtained using the No U-Turn Sampler (NUTS), a Hamiltonian Monte Carlo Method. The proposed model combines competing risks and spatial random effects in modeling the Titan GPU failures data and our results interesting insights in GPU failures in HPC systems.

Notes

W. J. Youden Address

Thursday, October 13, 4:00 – 5:00pm ASQ STAT Presiding
Prospector 3&4

Latin Squares, Youden Squares, Balanced Incomplete Block Designs (BIBDs) and Extensions for Industrial Application

Bradley Jones
JMP Statistical Discovery, LLC
Bradley.Jones@jmp.com

Abstract: Latin Squares are beautifully symmetric designs employing a given number of symbols in the same number of rows and columns. Each row and each column have all the symbols. A Youden Square, attributed to Jack Youden, is a Latin Square from which a number of rows have been removed. The Youden Square is a special case of a balanced incomplete block design (BIBD). All these designs have been around for decades. They support a single categorical factor and one or two blocking factors. They are commonly used in agriculture where the rows and columns are rows and columns of plants and it is desirable to remove any fertility gradients in a field in the analysis.

In industrial settings it is unusual for experiments to be limited to only one categorical factor and one or two blocking factors. However, it could be very useful to use these designs as building blocks for creating experiments in industry with more factors. This talk will show how to extend the above designs when there are continuous factors to explore.

Biography: Dr. Bradley Jones is the Distinguished Research Fellow in the JMP division of SAS where he is responsible for the development of new methods in design of experiments (DOE). He built the JMP Custom Designer, a general and powerful tool for generating optimal experimental designs. He holds a patent on the use of DOE for minimizing registration errors in the manufacture of laminated circuit boards and four other patents for innovations in design of experiments. He is the inventor of the prediction profile plot for interactive exploration of multiple input and output response surfaces.

He is the co-discoverer with Christopher Nachtsheim of Definitive Screening Designs. With Peter Goos, he wrote *Optimal Design of Experiments - A Case Study Approach*, which won the 2012 Ziegel Prize. He is also the coauthor of the book *Design of Experiments: A Modern Approach* with Douglas C. Montgomery.

He received the Brumbaugh Award for the paper making the largest contribution to industrial quality control in 2009, 2011, and 2022. He is the recipient of the Lloyd S. Nelson Award for the article having the greatest immediate impact to practitioners in 2010 and again in 2012. In 2012 he received the Statistics in Chemistry Award from the American Statistical Association for outstanding collaborative work with a chemist. He also won the Youden Prize for the best expository paper in *Technometrics* in 2012. In 2022 he won the Shewhart Medal. He is a former Editor-in-Chief of the *Journal of Quality Technology* and a Fellow of the American Statistical Association.

Notes

Friday, October 14

Session 4A: STAT Invited Session

Friday, October 14, 8:00 – 9:30am Moderator: Mindy Hotchkiss
Prospector 2

Covariate Software Vulnerability Discovery Model to Support Cybersecurity Test & Evaluation

Lance Fiondella
University of Massachusetts Dartmouth
lfiondella@umassd.edu

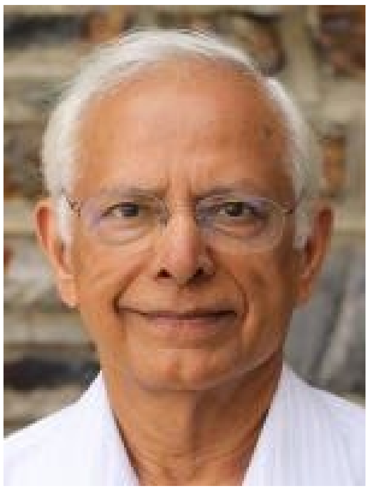
Abstract: Vulnerability discovery models (VDM) have been proposed as an application of software reliability growth models (SRGM) to software security related defects. VDM model the number of vulnerabilities discovered as a function of testing time, enabling quantitative measures of security. Despite their obvious utility, past VDM have been limited to parametric forms that do not consider the multiple activities software testers undertake in order to identify vulnerabilities. In contrast, covariate SRGM characterize the software defect discovery process in terms of one or more test activities. However, data sets documenting multiple security testing activities suitable for application of covariate models are not readily available in the open literature.

To demonstrate the applicability of covariate SRGM to vulnerability discovery, this research identified a web application to target as well as multiple tools and techniques to test for vulnerabilities. The time dedicated to each test activity and the corresponding number of unique vulnerabilities discovered were documented and prepared in a format suitable for application of covariate SRGM. Analysis and prediction were then performed and compared with a flexible VDM without covariates, namely the Alhazmi-Malaiya Logistic Model (AML). Our results indicate that covariate VDM significantly outperformed the AML model on predictive and information theoretic measures of goodness of fit, suggesting that covariate VDM are a suitable and effective method to predict the impact of applying specific vulnerability discovery tools and techniques.

Notes

Session 4A: STAT Invited Session

Friday, October 14, 8:00 – 9:30am Moderator: Mindy Hotchkiss
Prospector 2



Rethinking Software Fault Tolerance

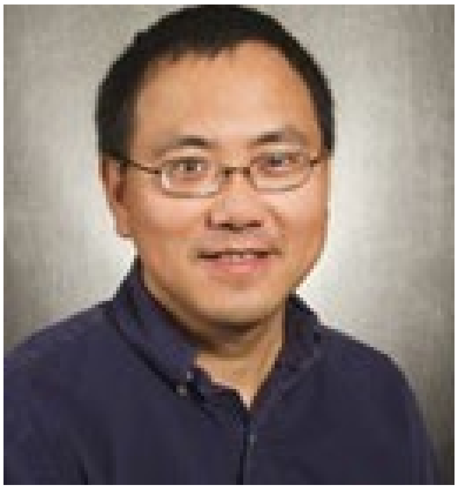
Kishor Trivedi
Duke University
ktrivedi@duke.edu

Abstract: Complex systems in different domains contain significant amount of software. Several studies have established that a large fraction of system outages are due to software faults. Traditional methods of fault avoidance, fault removal based on extensive testing/debugging, and fault tolerance based on design/data diversity are found inadequate to ensure high software dependability. The key challenge then is how to provide highly dependable software. We discuss a viewpoint of fault tolerance of software-based systems to ensure high dependability. We classify software faults into Bohrbugs and Mandelbugs, and identify aging-related bugs as a subtype of the latter. Traditional methods have been designed to deal with Bohrbugs. The key challenge then is to develop mitigation methods for Mandelbugs in general and aging-related bugs in particular. We submit that mitigation methods for Mandelbugs utilize environmental diversity. Retry operation, restart application, failover to an identical replica (hot, warm or cold) and reboot the OS are reactive recovery techniques applied after the occurrence of a failure. They are examples of techniques that rely on environmental diversity. For software aging related bugs, it is also possible to utilize a proactive environmental diversity technique known as software rejuvenation. We discuss environmental diversity both from experimental and analytic points of view and cite examples of real systems employing these techniques.

Notes

Session 4B: CPID Invited Session

Friday, October 14, 8:00 – 9:30am Moderator: Yili Hong
Prospector 1



D-optimal Mixture Designs for Ordinal Responses

Rong Pan
Arizona State University
Rong.Pan@asu.edu

Abstract: Ordered, categorical data in experimental studies are common in industries with sensory programs that typically solicit subjective responses from trained sensory experts. For example, Mancenido, Pan, and Montgomery (2016) described a case study where the quality attribute of interest from a three-chemical experiment was the fragrance intensity of chemical formulations, measured using an ordered, intensity scale from 1 (very weak) to 7 (very strong). Due to the lack of guidelines in designing mixture experiments for categorical responses in literature, the computer-generated optimal designs for continuous, numeric responses are often used in substitution. In our experience, using experimental designs which are ill-adapted for categorical responses could be detrimental to the modeling result. In this talk we will discuss the challenges in finding D-optimal mixture designs for ordinal responses and we will present an exchange algorithm that is efficient in finding a near optimal design. Our methods will be compared with other approaches.

Notes

Session 4B: CPID Invited Session

Friday, October 14, 8:00 – 9:30am Moderator: Yili Hong
Prospector 1

Specifying Prior Distributions in Reliability Applications

Qinglong Tian
University of Waterloo
qltian@iastate.edu

Abstract: Especially when facing reliability data with limited information (e.g., a small number of failures), there are strong motivations for using Bayesian inference methods. These include the option to use information from physics-of-failure or previous experience with a failure mode in a particular material to specify an informative prior distribution. Another advantage is the ability to make statistical inferences without having to rely on specious (when the number of failures is small) asymptotic theory needed to justify non-Bayesian methods. Users of non-Bayesian methods are faced with multiple methods of constructing uncertainty intervals (Wald, likelihood, and various bootstrap methods) that can give substantially different answers when there is little information in the data. For Bayesian inference, there is only one method—but it is necessary to provide a prior distribution to fully specify the model. Much work has been done to find default or objective prior distributions that will provide inference methods with good (and in some cases exact) frequentist coverage properties. This paper reviews some of this work and provides, evaluates, and illustrates principled extensions and adaptations of these methods to the practical realities of reliability data (e.g., non-trivial censoring).

Notes

Session 4C: Journal of Quality Technology Invited Session

Friday, October 14, 8:00 – 9:30am | Moderator: Allison Jones-Farmer
Silver Mine B

Estimating Pure-Error from Near Replicates in Design of Experiments

Caleb King
JMP
Caleb.King@jmp.com
Additional Authors: Thomas Bzik and Peter Parker

Abstract: In design of experiments, setting exact replicates of factor settings enables estimation of pure-error; a model-independent estimate of experimental error useful in communicating inherent system noise and testing model lack-of-fit. Often in practice, the factor levels for replicates are precisely measured rather than precisely set, resulting in near-replicates. This can result in inflated estimates of pure-error due to uncompensated set-point variation. In this presentation, we review different strategies for estimating pure-error from near-replicate, including our own recent work as well as new material brought to our attention after the publication of our original work.

Notes

Session 4C: Journal of Quality Technology Invited Session

Friday, October 14, 8:00 – 9:30am — Moderator: Allison Jones-Farmer
Silver Mine B

Monitoring Proportions with Two Components of Common Cause Variation

Rob Goedhart
University of Amsterdam
R.Goedhart2@uva.nl
Additional Author: William H. Woodall

Abstract: We propose a method for monitoring proportions when the in-control proportion and the sample sizes vary over time. Our approach is able to overcome some of the performance issues of other commonly used methods, as we demonstrate in this paper using analytical and numerical methods. The derivations and results are shown mainly for monitoring proportions, but we show how the method can be extended to the monitoring of count data.

Notes

Session 5A: SPES Invited Session

Friday, October 14, 10:00 – 11:30am Moderator: Jon Stallrich
Prospector 2

Data Splitting

Roshan Joseph
Georgia Tech
roshan@gatech.edu
Additional Author: Akhil Vakayil

Abstract: For developing statistical and machine learning models, it is common to split the dataset into two parts: training and testing. The training part is used for fitting the model and the testing part for evaluating the performance of the fitted model. The most common strategy for splitting is to randomly sample a fraction of the dataset. In this talk, I will discuss an optimal method for data splitting called Support Points-based split (SPlit).

Notes

Session 5A: SPES Invited Session

Friday, October 5, 10:00 – 11:30am Moderator: Jon Stallrich
Prospector 2

Active Learning for Deep Gaussian Process Surrogates

Annie Sauer
Virginia Tech
anniees@vt.edu
Additional Authors: Robert B. Gramacy and David Higdon

Abstract: Deep Gaussian processes (DGPs) are increasingly popular as predictive models in machine learning for their non-stationary flexibility and ability to cope with abrupt regime changes in training data. Here we explore DGPs as surrogates for computer simulation experiments whose response surfaces exhibit similar characteristics. In particular, we transport a DGP's automatic warping of the input space and full uncertainty quantification, via a novel elliptical slice sampling Bayesian posterior inferential scheme, through to active learning strategies that distribute runs non-uniformly in the input space — something an ordinary (stationary) GP could not do. Building up the design sequentially in this way allows smaller training sets, limiting both expensive evaluation of the simulator code and mitigating cubic costs of DGP inference. When training data sizes are kept small through careful acquisition, and with parsimonious layout of latent layers, the framework can be both effective and computationally tractable. Our methods are illustrated on simulation data and two real computer experiments of varying input dimensionality. We provide an open source implementation in the "deepgp" package on CRAN.

Notes

Session 5B: Response Surface Design & Optimization

Friday, October 14, 10:00 – 11:30am Moderator: Bryan Smucker
Prospector 1

Advances in Orthogonal Minimally Aliased Response Surface (OMARS) Designs: Designs with Two-Level Categorical Factors, with a Blocking Factor, and Strong OMARS Designs

José Núñez Ares
KU Leuven
jose.nunezares@kuleuven.be
Additional Authors: Eric Schoen and Peter Goos

Abstract: Response surface designs are a core component of the response surface methodology, which is widely used in the context of product and process optimization. The best known response surface designs are central composite and Box-Behnken designs. The problem with these designs is that the number of tests they require increases rapidly with the number of factors. In this presentation, we present a new family of response surface designs called orthogonal minimally aliased response surface or OMARS designs. The OMARS designs are available for many different run sizes and therefore bridge the gap between the small definitive screening designs and the large classical response surface designs. While the original OMARS designs included only three-level quantitative factors, there are now also mixed-level OMARS designs including both three-level quantitative factors and two-level categorical ones. In addition, we show how to arrange OMARS designs in blocks, how to select the best possible OMARS design for a given experiment, and how to assess the trade-off between the run size and the statistical quality of the design. We end the talk with the presentation of strong OMARS designs, which form a subclass of OMARS designs specially suited for optimization experiments.

Notes

Session 5B: Response Surface Design & Optimization

Friday, October 14, 10:00 – 11:30am Moderator: Byran Smucker
Prospector 1

Model Robust Response Surface Optimization for Mixed Quantitative and Qualitative Factors

Gautham Sunder
University of Minnesota
sunde153@umn.edu
Additional Author: Christopher Nachtsheim

Abstract: Motivated by our joint work with a medical device manufacturer on hyperparameter optimization of deep neural networks, in this study, we propose a model robust response surface optimization (RSO) strategy for the dual goals of model estimation and response optimization in the presence of mixed quantitative and qualitative (QQ) factors. RSO literature prescribes using Classical-RSO (C-RSO) methods when the response surface is assumed to be second-order and Bayesian Optimization (BO) when the response surface is assumed to be complex and nonlinear. However, neither C-RSO nor BO formally validate the assumptions made on the response surface complexity. The proposed model robust RSO strategy is highly efficient when the experimenter is ambiguous about the response surface complexity, second-order or complex and nonlinear. The proposed model robust RSO strategy is initialized with a model robust starting design which is a compromise between space-filling designs and Bayesian D-optimal designs and is supersaturated for a full second-order model. We formally validate the adequacy of a second-order approximation by proposing a modified version of RPtest, a bootstrap goodness-of-fit test for validating a second-order approximation. The proposed model robust RSO strategy transitions to BO methods if the second-order approximation is inadequate. Our simulation study illustrates that the proposed RSO method is highly efficient for estimating a second-order response surface with QQ factors and the goodness-of-fit test proposed has high power to detect the inadequacy of a second-order approximation. Additionally, the proposed RSO strategy has comparable performance to standard BO methods for estimating a complex response surface.

Notes

Session 5C: Quality Engineering Invited Session

Friday, October 14, 10:00 – 11:30am   Moderator: David Edwards
Silver Mine B

Incorporating Uncertainty for Enhanced Leadership Scoring and Ranking in Data Competitions

Lu Lu
University of South Florida
lulu1@usf.edu

Christine Anderson-Cook
Los Alamos National Laboratory
candcook@gmail.com

Abstract: Data competitions have become a popular and cost-effective approach for crowdsourcing versatile solutions from diverse expertise. Current practice relies on the simple leaderboard scoring based on a given set of competition data for ranking competitors and distributing the prize. However, a disadvantage of this practice in many competitions is that a slight difference in the scores due to the natural variability of the observed data could result in a much larger difference in the prize amounts. In this article, we propose a new strategy to quantify the uncertainty in the rankings and scores from using different data sets that share common characteristics with the provided competition data. By using a bootstrap approach to generate many comparable data sets, the new method has four advantages over current practice. During the competition, it provides a mechanism for competitors to get feedback about the uncertainty in their relative ranking. After the competition, it allows the host to gain a deeper understanding of the algorithm performance and their robustness across representative data sets. It also offers a transparent mechanism for prize distribution to reward the competitors more fairly with superior and robust performance. Finally, it has the additional advantage of being able to explore what results might have looked like if competition goals evolved from their original choices. The implementation of the strategy is illustrated with a real data competition hosted by Topcoder on urban radiation search.

Notes

Session 5C: Quality Engineering Invited Session

Friday, October 14, 10:00 – 11:30am Moderator: David Edwards
Silver Mine B

Understanding and Addressing Complexity in Problem Solving

Willis Jensen
W. L. Gore & Associates, Inc.
wjensen@wlgore.com
Additional Authors: Roger Hoerl and Jeroen de Mast

Abstract: Complexity manifests itself in many ways when attempting to solve different problems, and different tools are needed to deal with the different dimensions underlying that complexity. Not all complexity is created equal. We find that most treatments of complexity in problem solving within both the statistical and quality literature focus narrowly on technical complexity, which includes complexity of subject matter knowledge as well as complexity in the data access and analysis of that data. The literature lacks an understanding of how political complexity or organizational complexity interferes with good technical solutions when trying to deploy a solution. Therefore, people trained in statistical problem solving are ill-prepared for the situations they are likely to face on real projects. We propose a framework that illustrates examples of complexity from our own experiences, and the literature. This framework highlights the need for more holistic problem-solving approaches and a broader view of complexity. We also propose approaches to successfully navigate complexity.

Notes

Luncheon

Friday, October 14, 11:45am – 1:15pm ASA SPES Presiding
Prospector 3&4

News from ASA and Partnerships to Establish a System for Wastewater Epidemiology for Houston

Kathy Ensor
President, American Statistical Association
ensor@rice.edu

Abstract: The American Statistical Association is dedicated to the practice and profession of statistics, and I am honored to serve as its 117th President. I will provide updates on ASA and take this opportunity to discuss the expansion of the ASA Leadership Institute, activities in data science and AI, and my vision to promote growth in community analytics. Community analytics requires strong partnerships between local governments and businesses, NGOs, community organizations, and academia. I will embed my community analytics discussion in the story of developing from scratch the wastewater epidemiology program for the City of Houston to aid in the management of the SARS-CoV-2 pandemic. Wastewater surveillance has emerged as a vital tool for city, county, state, and national public health departments and its use will continue beyond the pandemic. For example, by examining SARS-CoV-2 RNA viral load, wastewater provides a strong signal of the extent of the virus present in a community. Measuring and modeling the evolution of the virus in a community is however fraught with many complexities. For Houston, the SARS-CoV-2 viral load is measured weekly at 39 wastewater treatment plants and approximately 100 other locations throughout the city. How to combine all this information into a system that informs decisions is the role that I have played in the multi-disciplinary and multi-institutional team. I will share the concepts behind building the statistical system as well as provide a deep dive into why the partnership succeeded and what is planned for the future.

Biography: Katherine Bennett Ensor is the Noah G. Harding Professor of Statistics at Rice University where she serves as director of the Center for Computational Finance and Economic Systems (cofes.rice.edu) and creator of the Kinder Institute's Urban Data Platform (kinderudp.org). Ensor served as chair of the Department of Statistics from 1999 through 2013 and has shaped data science at Rice as a member of the campus-wide hiring committee. Her research focuses on the development of statistical and data science methods for practical problems. Her expertise is dependent data covering time, space, and dimension with applied interests in finance, energy, environment, health, and risk management. She is a fellow of ASA and AAAS and has been recognized for her leadership, scholarship, and mentoring. Ensor is the 2022 President of the American Statistical Association (ASA). She served as Vice President of ASA from 2016 to 2018 and as a member of the National Academies Committee on Applied and Theoretical Statistics from 2014 to 2020. Ensor holds a BSE and MS in Mathematics from Arkansas State University and a Ph.D. in Statistics from Texas A&M University. Ensor is a member of the Texas A&M College of Science Academy of Distinguished Former Students.

Notes

Session 6A: Statistical Engineering

Friday, October 14, 1:30 – 3:00pm Moderator: Willis Jensen
Prospector 2

Modeling Semiconductor Yield Using Statistical Engineering – A Case Study

Dana Krueger
Roadmap Research Global
dana@roadmapresearchglobal.com
Additional Author: Douglas Montgomery

Abstract: Yield is a key process performance characteristic in the capital-intensive semiconductor fabrication process. In an industry where machines cost millions of dollars and cycle times are several months, predicting and optimizing yield are critical to process improvement, customer satisfaction, and financial success. Semiconductor yield modeling is essential to identifying processing issues, improving quality, and meeting customer demand in the industry. However, the complicated fabrication process, the massive amount of data collected, and the number of modeling approaches available make yield modeling a complex and unstructured problem that crosses potential silos around various process engineering domains, device engineers, defect metrologists, statisticians, and yield analysts. This complexity makes the problem a strong candidate for applying statistical engineering principles. This project first examines the issues of data pedigree and integration as a dataset is created to combine wafer-level process measurements (such as critical dimensions, where process owners use control charts to detect problems), differences found in defectivity scans after each processing layer of the wafer (where defect metrology engineers use scanning electron microscopes to identify problems and look for patterns on the scans), and electrical test data that assess functionality for the wafer (such as threshold voltage and approximately 400 other parameters). The yield data are captured after testing each die, and each is assigned to a bin that is designated as "passing" or "failing." Various failure modes are denoted by different bin numbers. Once the dataset is constructed, a yield modeling approach must be determined. Relying on statistical methods alone limits the effectiveness of this effort. This case study reviews the benefits and limitations of using an iterative approach to improving yield models by first applying generalized linear models (GLMs), then generalized linear mixed models (GLMMs), and finally a combination of classification trees and GLMs. This third approach illustrates how being "tool agnostic" can help the statistical engineer find new, creative solutions to modeling challenges. The results show the strengths and limitations of each modeling approach. The challenge and

benefits of obtaining subject matter expertise on the project throughout the modeling process and ideas for how a developed model could be deployed and sustained will also be discussed.

Notes

Session 6A: Statistical Engineering

Friday, October 14, 1:30 – 3:00pm Moderator: Willis Jensen
Prospector 2

Fleet Predictions Using Bayesian Multilevel Models: An Example using Targeting Pod Accuracy

Todd Remund
Northrop Grumman Space
todd.remund@ngc.com

Abstract: Fleet behavior and structure must be represented in the model, so that the model can represent the fleet. Deeper thinking suggests that fleet behavior and structure must be represented in the data collection effort, so that the model best represents the reality of the fleet. This leads to the statement: Appropriate models represent data, data represents reality—we hope. Fleets make up many populations of operational systems across the aerospace industry, as well as plenty of other engineering industries. Fleets can be made up of subgroups such as batches, material lots, software versions, supplier, and aircraft tail numbers to name a few. With these varying layers of groups comes structure that can heavily dictate the decisions and guidance to operators. Some of the results from modeling and analysis guide operators of systems to make real-time decisions that can have heavy implications such as weapon delivery to a target in the proximity of friendlies. Safety and effectiveness drive the need to seek proper representation of fleets so that indeed these two aspects of operational activity can be preserved to the best extent possible. Multilevel models are not a new approach for data analysis and prediction. Alternate names for this capability are mixed models, random coefficient models, hierarchical models. Textbooks have existed for decades describing both frequentist and Bayesian implementations of this method. The literature and computer applications available offer plenty of avenues to utilize these modeling methods. Within the aerospace and engineering communities there are ample opportunities to apply this approach. Application to tolerance intervals and probabilistic metrics are the focus here with the emphasis on Bayesian implementation. Methods of applying these to the aerospace industry will be given with an example of targeting accuracy in target pod performance. This is a call to seriously consider, or increase, the use of multilevel models and associated data collection to model fleet realities. Methods are offered in a simple example to depict how the Bayesian output from multilevel models can be specifically used to represent fleet realities in performance measures. Where some applications appropriately remove nuisance effects from performance metrics such as tester ID,

many aerospace applications have the opposite need—to account for fleet group-to-group effects instead of removing them. This paper advances the use of the Bayesian multilevel model's posterior distribution to account for these random group-to-group effects within predictions involving tolerance intervals and probabilistic representations of fleet success criteria. An example of notional target pod data will be used to demonstrate these methods.

Notes

Session 6B: Machine Learning & Optimization

Friday, October 14, 1:30 – 3:00pm Moderator: Jie Min
Prospector 1

Tuning Parameter Selection for Variable Selection via an R2 Metric

Jon Stallrich
North Carolina State University
jwstalli@ncsu.edu
Additional Author: Julia Holter

Abstract: Many penalized estimators are capable of performing simultaneous variable selection and estimation, but are burdened by tuning parameter selection. Many tuning parameter selection procedures tend to choose more variables than necessary and are computationally expensive. We propose a tuning parameter selection strategy based on the squared correlations between the observed response and the predicted values of models, rather than squared error loss. Tuning parameters selected under our procedure are shown to better balance predictive capability and model simplicity. The approach is computationally efficient and, in the domain of penalized estimation, competitive with popular tuning parameter selection techniques in its capacity for variable selection. We explore the efficacy of this approach in a project involving optimal EMG placement for a robotic prosthesis controller.

Notes

Session 6B: Machine Learning & Optimization

Friday, October 14, 1:30 – 3:00pm Moderator: Jie Min
Prospector 1

Bayesian Optimization via Barrier Functions

Tony Pourmohamad
Genentech
tpourmohamad@gmail.com
Additional Author: K. H. Lee

Abstract: Hybrid optimization methods that combine statistical modeling with mathematical programming have become a popular solution for Bayesian optimization (BO) because they can better leverage both the efficient local search properties of the numerical method and the global search properties of the statistical model. These methods seek to create a sequential design strategy for efficiently optimizing expensive black-box functions when gradient information is not readily available. In this presentation, we propose a novel BO strategy that combines response surface modeling with barrier methods to efficiently solve expensive constrained optimization problems in computer modeling. At the heart of all BO algorithms is an acquisition function for effectively guiding the search. Our hybrid algorithm is guided by a novel acquisition function that tries to decrease the objective function as much as possible while simultaneously trying to ensure that the boundary of the constraint space is never crossed. Illustrations highlighting the success of our method are provided, including a real-world computer model optimization experiment from hydrology.

Notes

Session 6C: SPC

Friday, October 14, 1:30 – 3:00pm Moderator: Valeria Quevedo
Silver Mine B

Applied vs Academic Approaches to Statistical Process Monitoring (SPM)

James Lucas
J. M. Lucas and Associates
JamesM.Lucas@verizon.net

Abstract: I was an integral part of the largest known implementation of CUSUM procedures. CUSUM was the Statistical Process Monitoring (SPM) procedure used in our company wide quality system. There are differences between what we did when we were implementing CUSUM SPM for our quality system and the recommendations contained many academic articles. This talk will describe the differences and explain why we conducted SPM differently from many academic recommendations. We introduce this talk by giving an overview of our quality system. We introduce CUSUM control procedures and give an implementation example. The example describes the specific CUSUM procedure that was used in a majority of our implementations. This specific CUSUM was used so frequently that I have described it as the 80% solution. A head-start feature was almost always used. We show that a headstart feature should be part of most SPM implementations. We then discuss specifics about our implementation, describe differences from academic recommendations and tell why those different choices were made. Some specific topics discussed are: Does a CUSUM signal cause you to shut down the process? Why or when does it cause a process shutdown? What False Alarm Probability (FAP) or Average Run Length (ARL) should be used? Why. Compare our choices with academic recommendations. How much historical process data is needed to set up an individual CUSUM loop. Academics frequently recommend much more data than we used. Our recommendations can be succinctly described as we recommend that you implement and update the CUSUM procedure rather than wait to implement. A recent review of data aggregation procedures did not describe the aggregation procedures we used. We describe our general process model that uses five variance components. Our modelling structure often used vessels, sub-vessels sub-sub-vessels. We show how this fact determined our aggregation procedures. Reducing process variability is an important goal of the quality system. However almost all of the CUSUM loops were for shifts in mean level. We tell how variability reduction was achieved by using mean-level control of sub-vessels. I conclude by comparing academic and applied approaches to SPM research. I describe the recent work of our research team that improves SPM and makes it easier to

conduct. We have developed Power Guidelines for SPM. The guidelines are succinctly described using a stoplight model. Our power guidelines use an ARL ratio; we recommend an ARL Ratio=ARLic/ARLoc ≥ 20. Much of our recent research concerns SPM for attribute data. Our power guidelines are used to determine shift detection possibilities when count levels low. We describe our algorithmic approach for designing CUSUM for Counts SPM procedures. Table lookup procedures are more difficult to use when there are distributional; changes with changes in sample sizes and mean level. We take "A Closer Look at Geometric and Bernoulli Procedures." We compare steady-state and initial state ARL procedures and tell when each is most appropriate. We tell why CCC-r procedures should seldom if ever be used. This continues recent valuable research criticizing misguided SPM methods.

Notes

Session 6C: SPC

Friday, October 14, 1:30 – 3:00pm Moderator: Valeria Quevedo
Silver Mine B

A Consistent Data Model for Different Data Granularity in Control Charts

Scott Grimshaw
Brigham Young University
grimshaw@byu.edu

Abstract: After a long-running show was canceled, control charts are used to identify if and when viewing drops. The finest granularity daily viewing has high autocorrelation and control charts use residuals from a seasonal ARIMA model. For coarse granularity data (weekly and monthly viewing) an approximate AR model is derived to be consistent with the finest granularity model. With the proposed approach, a longer memory model is used in the granular data control charts that reduces the number of false alarms from control charts constructed treating granular data as a different measurement.

Notes

Wine & Cheese Reception with SPES Special Panel Session

Friday, October 15, 3:15 – 5:15 pm

Moderator: Jon Stallrich

Panel session will begin at 4 pm.

Prospector 3&4

Innovations in Industry 4.0

Panelists:

Arman Sabbaghi
Purdue University
sabbaghi@purdue.edu

Laura Freeman
Virginia Tech
laura.freeman@vt.edu

Nathan Wycoff
Georgetown
nathanbrwycoff@gmail.com

Mindy Hotchkiss
Aerojet Rocketdyne
Mindy.Hotchkiss@rocket.com

Abstract: Industrial manufacturing and logistics have greatly benefited from applications of statistical methodology, particularly experimental design and quality control. Industry 4.0 refers to the fourth industrial revolution sparked by integration of machine learning and artificial intelligence in manufacturing and logistics. Examples of Industry 4.0 include Industrial Internet of Things, cyber physical systems, and cyber manufacturing. The shift towards more automated and fast-paced systems relies on collaborations between experienced data scientists, statisticians, and engineers. This panel represents different perspectives and experiences on the current and future state of Industry 4.0.

Biographies of Panelists: Arman Sabbaghi is an Associate Professor in the Department of Statistics and Associate Director of the Statistical Consulting Service at Purdue University. He became an Elected Member of the International Statistical Institute in 2020. He received his PhD in Statistics from Harvard University in 2014. Dr. Sabbaghi's research interests are in Bayesian data analysis, experimental design, and causal inference. Specific major objectives of his current research are (1) the development of efficient and interpretable statistical frameworks and machine learning algorithms for modeling and quality control in advanced manufacturing systems, (2) the creation of mathematical tools that facilitate the characterization of broad classes of experimental designs for the study and improvement of processes in engineering and the physical sciences, and (3) the development of new causal inference methods for the analysis of Big Observational Data and clinical trials plagued by nonadherence.

Laura Freeman is a Research Associate Professor of Statistics and serves as the Deputy Director of the Virginia Tech National Security Institute. Her research leverages experimental methods for conducting research that brings together cyber-physical systems, data science, artificial intelligence (AI), and machine learning to address critical challenges in national security. She is also the Assistant Dean for Research for the College of Science. Dr. Freeman has a B.S. in Aerospace Engineering, a M.S. in Statistics and a Ph.D. in Statistics, all from Virginia Tech.

Nathan Wycoff is a postdoc at the McCourt School's Massive Data Institute at Georgetown University. His research interests include blackbox optimization, especially in high dimension using Gaussian processes, and variable selection via nonsmooth priors and variational inference applied to human migration.

Mindy Hotchkiss is a Technical Specialist and enterprise-wide Subject Matter Expert (SME) in Statistics for Aerojet Rocketdyne. She is a member of the American Statistical Association (ASA), the Military Operations Research Society (MORS), and the Society of Reliability Engineers (SRE). She is a Senior Member of ASQ, is a Past Chair of the ASQ Statistics Division, and serves on multiple committees. She also represents the ASQ Reliability & Risk Division (RRD) on the Board of Directors for the Reliability and Maintainability Symposium (RAMS) and is currently Secretary of RRD. She is an ASQ-Certified Reliability Engineer (CRE), Quality Engineer (CQE), and Six Sigma Black Belt (CSSBB). She co-chaired the Annual ASQ/ASA Fall Technical Conference (FTC) in 2018. Mindy holds BS degrees in Mathematics and Statistics and an MBA from the University of Florida, and a Masters of Statistics from North Carolina State University. She has over 20 years of experience as a statistical consultant between Pratt & Whitney and Aerojet Rocketdyne, including work supporting technology development across the enterprise.

Notes

Call for Papers

65th Annual Fall Technical Conference

Quality: Rooted in Statistics and Data Science

October 3-5, 2023

Sheraton Raleigh Hotel, Raleigh, North Carolina

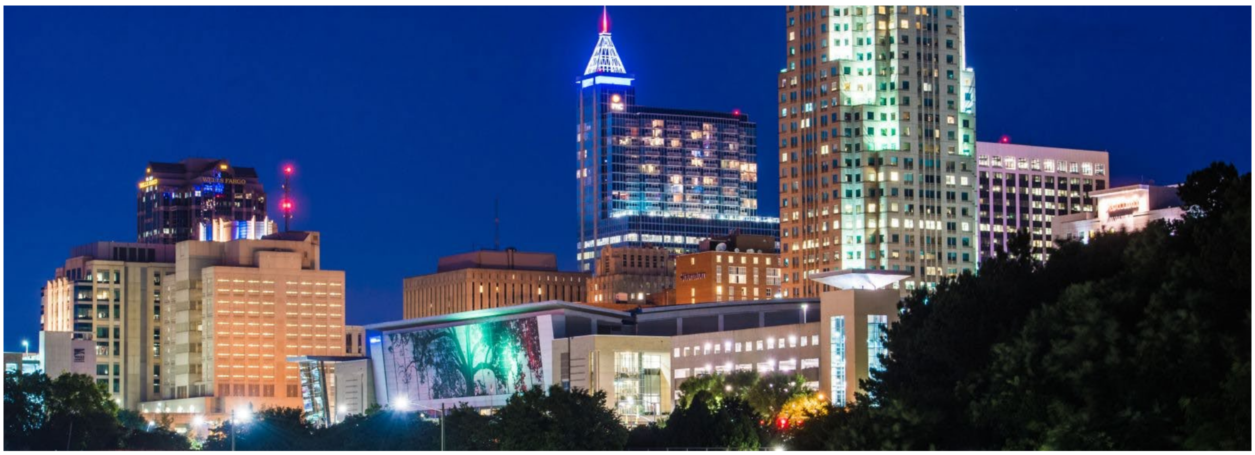


Photo courtesy of visitraleigh.com.

Co-Sponsored By:

American Society for Quality
Chemical and Process Industries Division
Statistics Division

American Statistical Association
Section on Physical and Engineering Sciences
Quality & Productivity Section

We invite you to submit abstracts for presentation at the 65th Annual Fall Technical Conference to be held on October 3-5, 2023 (short courses to be held October 3) at the Sheraton Raleigh hotel in downtown Raleigh, NC. Raleigh is the capital of North Carolina and is known for its universities and its beautiful oak trees. The Fall Technical Conference has long been a forum for both statistics and quality, and is co-sponsored by the American Society for Quality (Chemical and Process Industries Division and Statistics Division) and the American Statistical Association (Section on Physical and Engineering Sciences and Section on Quality and Productivity). The goal of this conference is to engage researchers and practitioners in a dialogue that will lead to the more effective use of statistics to improve quality and foster innovation.

Please submit an abstract online if you are interested in presenting an applied or expository paper in any of the categories of: Statistics, Quality, Experiment Design, Machine Learning, or Tutorial/Case Studies,. The work in a presentation should be strongly justified by an application to a problem in engineering, manufacturing, the process/chemical industry, the physical sciences, or a service industry. The mathematical level of the papers may range from introductory to advanced (e.g., that of the *Journal of Quality Technology* or *Technometrics*). Please note which level of audience is targeted (Introductory, Intermediate, or Advanced) so that the committee can assign papers appropriately and plan a balanced program. The program committee welcomes any suggestions for special session topics or speakers. If you have any suggestions or proposals, please contact one of the program committee members listed below. For more information please visit http://www.falltechnicalconference.org

Program Committee

SPES
Arman Sabbaghi (Chair)
Purdue University
sabbaghi@purdue.edu

Q&P
Anne Driscoll
Virginia Tech
adriscoll@vt.edu

STAT
Mindy Hotchkiss
Aerojet Rocketydyne
Mindy.Hotchkiss@rocket.com

CPID
Yili Hong
Virginia Tech
yilihong@vt.edu

General Conference Chair
Emily H. Griffith
NC State University
919-515-1926
eghohmei@ncsu.edu

Abstract Submission Information

The abstract submission deadline is February 28, 2023.

Online Abstract Submission
www.falltechnicalconference.org/submit-an-abstract

Abstract Content

The abstract should include a session preference, target audience, and the following three components:
1. Motivation or background
2. Description of work done
3. Significance. Are there improvements, applications, new abilities, new points of view, etc.? How will the status quo be changed?

Session Preference (choose one)
___ Statistics
___ Quality
___ Experimental Design
___ Tutorial/Case Study
___ Machine Learning
___ Data Science

Target Audience (choose one)
___ Introductory/Practitioner
___ Intermediate
___ Advanced/Theoretical

One presenter for each talk will receive a 50% discount on conference registration.

2023 SPRING RESEARCH CONFERENCE
May 24 - 26, BANFF CENTER, ALBERTA

ORGANIZING COMMITTEE
Xinwei Deng (Virginia Tech)
Devon Lin (Queen's University)

PROGRAM COMMITTEE
Tirthankar Dasgupta (Chair, Rutgers University)
Emily Casleton (Los Alamos National Lab)
Simon Mak (Duke University)
Nicole Pashley (Rutgers University)
Arman Sabbaghi (Purdue University)
David Stennin (Simon Fraser University)
Nathaniel Stevens (University of Waterloo)

QUESTIONS & INQUIRIES
For more informations, questions or inquiries, please contact devon.lin@queensu.ca.

INTRODUCTION

The SRC is the annual meeting of the Section on Physical and Engineering Statistics (SPES) of the American Statistical Association (ASA). Although historically emphasizing industrial statistics, design of experiments, quality and reliability, the meeting increasingly emphasizes modern methods in statistical and machine learning and high performance computing in statistical methodology, with diverse applications encompassing all areas of applied sciences.

KEYNOTE SPEAKERS

Bianca Maria Colosimo (Politecnico di Milano)
Ron Kohavi (Ex-Airbnb, Ex-Microsoft, Ex-Amazon)

Keynote Panel Discussion
Derek Bingham (Simon Fraser University)
Robert Gramacy (Virginia Tech)
Roshan Joseph (Georgia Tech)

INVITED PROGRAM

Twenty one invited sessions will be organized on various relevant topics with special focus on emerging areas such as New generation simulations: Digital twins, Machine learning in industry 4.0, Data fusion, Statistical modeling in astro, particle and nuclear physics, and experiments on social networks. For a complete list of invited sessions, please check out the conference website.

CONTRIBUTED PROGRAM

The contributed program includes talks and posters. Contributed talks are essentially no different than the invited ones, except that the process is competitive rather than curated by the program committee.

STUDENT SCHOLARSHIPS

Courtesy of our generous sponsors, we are able to offer a limited number of **student travel scholarships**. If you wish to be considered, please additionally provide a short statement of interest in the meeting (e.g., how would it benefit your study?), a CV/resume, and a note from your advisor (on letterhead) confirming your student status and explaining the nature of your role in the research you are proposing to present at the meeting. All funded students must contribute a talk or poster.

VISIT https://sites.google.com/view/src2023

Special thanks to BYU for printing the technical program book.

Photo on front cover courtesy of Park City Chamber/Bureau.