# JOINT RESEARCH CONFERENCE 2024


ASA

UNIVERSITY OF WATERLOO

ims

## STATISTICS IN QUALITY TECHNOLOGY AND INDUSTRY

The joint meeting of the 40th ASA Quality and Productivity Research Conference and the 29th ASA/IMS Spring Research Conference.

JUNE 17 – 20, 2024

WATERLOO, ONTARIO, CANADA

# Table of Contents

## About the Conference

The 2024 Joint Research Conference on Statistics in Quality, Industry and Technology will be held in Waterloo, Ontario, Canada from June 17-20, 2024 at the University of Waterloo. This is a joint meeting of the 29th *Spring Research Conference on Statistics in Industry and Technology* and the 40th *Quality and Productivity Research Conference*. The conference is co-sponsored by the American Statistical Association Section on Quality and Productivity, the American Statistical Association Section on Physical & Engineering Sciences, and the Institute of Mathematical Statistics. Organization of this conference is also in partnership with Virginia Tech.

The goal of this conference is to stimulate interdisciplinary research and innovative solutions to practical problems through interactions among statisticians, data scientists, quality professionals, engineers, students, and others from diverse fields. The theme of this year's conference is *Data Science and Statistics for Industrial Innovation*. The technical program focuses on statistical methodology and creative problem solving to address scientific, industrial, and business challenges, drawing upon advances from the fields of statistics, machine learning, and data science.

The Local Organizing Committee and Scientific Program Committee hope that you enjoy the 2024 JRC and your time in Waterloo!

UNIVERSITY OF WATERLOO | Department of Statistics and Actuarial Science
ASA
ims
VIRGINIA TECH

## Sponsors

The 2024 Joint Research Conference acknowledges the generous financial contributions made by Steering Committee Chair Jeff Hooper (Westorross Advisors, LLC) as well as the following organizations.

# Scientific Program Committee

**Nathaniel Stevens**
(Co-Chair)
University of Waterloo

**Yili Hong**
(Co-Chair)
Virginia Tech

**Inez Zwetsloot**
University of Amsterdam

**Alan Vazquez**
Tecnológico de Monterrey

**Anne Driscoll**
Virginia Tech

**James Wilson**
University of San Francisco

**Srijan Sengupta**
North Carolina State University

**Maria Weese**
Miami University

**Ming Li**
Coupang

**Siddhesh Kulkarni**
Bristol Myers Squibb

# Honored Guests

## Conference Honoree and Keynote Speaker: Stefan H. Steiner

**Dr. Stefan H. Steiner** is a Professor in the Department of Statistics and Actuarial Science at the University of Waterloo and he served as chair of this department from 2014-2022. His primary research interests include quality improvement, process monitoring, experimental design and measurement system assessment in the realms of business, industry, and healthcare. He has served on editorial boards for *Journal of Quality Technology*, *Quality Technology and Quantitative Management*, *Annals of Applied Statistics*, *Technometrics*, and *Quality Engineering.* He has also served as President of the Statistical Society of Canada's Business and Industrial Statistics Section as well as Chair of the International Statistical Engineering Association. He is a Fellow of both the American Statistical Association and the American Society for Quality, he was the recipient of the 2021 ASA Donald B. Owen Award, and he received the 2023 ASQ Shewhart Medal.

## Plenary Speaker: Allison Jones-Farmer

**Dr. L. Allison Jones-Farmer** is the Van Andel Professor of Business Analytics at Miami University in Oxford, Ohio. Her research focuses on developing practical methods for analyzing data in industrial and business settings. She is currently the Editor-in-Chief of *Journal of Quality Technology,* is on the editorial board of *Quality Engineering*, and is a former Associate Editor for *Technometrics*. In addition to her research in industrial analytics, Allison enjoys helping organizations improve their analytics capability, developing innovative curricula, and teaching data science. Before joining Miami University, Allison was a Professor of Statistics and Analytics at Auburn University, where she held the C&E Smith chair. She received a B.S. in Mathematics from Birmingham-Southern College, an M.S. in Applied Statistics from the University of Alabama, and a Ph.D. in Applied Statistics from the University of Alabama.

## Plenary Speaker: Dave Campbell

**Dr. Dave Campbell** is a full Professor in the School of Mathematics and Statistics and the School of Computer Science at Carleton University. Academically, he runs a collaborative team researching inferential algorithms at the intersections of statistics with machine learning, computing, and applied mathematics to solve problems inspired by industry and government collaborations. He has co-authored discussion papers in Bayesian Analysis and the Journal of the Royal Statistical Society (series B) and been awarded over $3.5 million in research grants. Dave's career path maintains a theme of Data Science leadership. Most recently he spent two years leading a Data Science team at the Bank of Canada in projects relating to cybersecurity, forecasting banknote demand, understanding drivers of inflation, ensuring data privacy, and more. Before moving to Ottawa in 2019, Dave was a Professor at Simon Fraser University, where he led the creation of their BSc in Data Science. He was the inaugural President of the Data Science and Analytics Section of the Statistical Society of Canada and was a co-organizer of the Vancouver Learn Data Science Meetup (>5000 members).

# Short Course

## Introduction to Large Language Models

*Monday, June 17th (8:30am-5:00pm)*

*Building: Mathematics 3*

*Room: 3103 (for the course) and 3127 (for registrations, breaks, and meals)*

**Description:**

As statisticians, we often work with standard data structures like data frames, which contain numerical and categorical variables. These structures commonly comprise millions of rows and thousands of columns. However, there exists a multitude of other data structures that aren't universally familiar to statisticians, such as text data. Over the last decade, deep learning methods have continuously evolved, particularly in handling text data. Among the most recent advancements are large language models. In this short course, we aim to integrate large language models into our modeling toolkit. Initially, we'll introduce fundamental concepts and then delve into hands-on examples for practical application. Our learning journey will include setting up a computational environment on a cloud system, a process we'll navigate together. To participate, all you'll need is a laptop equipped with an internet browser.

**Instructors:**

**Ming Li** is currently a Staff Data Scientist at Coupang. He was a Director of Data Science in PetSmart, a Science Manager in Amazon, Data Scientist in Walmart and Statistical Leader in GE Global Research Center. He has visited 20 ASA Chapters to teach the ASA traveling course in data science, machine learning and deep learning during 2019, 2020 and 2021. He also organized and presented the 2018 JSM Introductory Overview Lecture: Leading Data Science: Talent, Strategy, and Impact. He was the Chair of Quality & Productivity Section of ASA. With 10+ years' experience in data science and machine learning, he has trained and mentored numerous junior data scientists with diversified backgrounds such as statistician, software developer, database programmer, and business analyst. He was also an instructor of Amazon's internal Machine Learning University and the recipient of Amazon's Best Science Mentor Award. He holds a Ph.D. in Statistics from Iowa State University.

**Xiaoda Liu** is currently a Staff Machine Learning Engineer at Coupang. He was an Applied Scientist at Amazon. With over six years of experience, he has been developing large scale machine learning and deep learning systems for retail and advertising sectors in the companies. His expertise lies in the application of advanced AI techniques to solve complex industry problems. He has mentored numerous data scientists and software engineers. He serves as reviewer for deep learning journals including Neurocomputing, IEEE Transactions on Neural Networks and Learning Systems. He holds a Ph.D. in Petroleum Engineering from Texas A&M University.

# Technical Tour

## Royal City Brewing Co.

*Wednesday, June 19th (3:00pm-7:00pm)*

*Location: Offsite*


The conference programming in the afternoon of Wednesday June 19 consists of a trip offsite to Royal City Brewing Co. in Guelph, Ontario. At 3pm, the group will depart from the conference venue via prearranged buses to the brewery. The event will consist of snacks, a beer tasting, and a tour of the brewery and brewhouse where conference goers will learn about the brewing process. The event will conclude with a reception in the Brew Hall where heavy hors d'oeuvres and one free drink will be provided. Additional drinks may be purchased directly from the brewery by patrons. The buses will then return conference goers back to the University of Waterloo, arriving by 7pm.

# Mary G. and Joseph Natrella Scholarship

The Natrella Scholarship was initiated by a contribution to the ASA Quality and Productivity Section given by Joseph Natrella at the time of Mary Natrella's death to honor her many contributions to the statistical community. The Natrellas always maintained a strong mutual interest in quality applications of statistics. Mary was for many years a staff member of the Statistical Engineering Division of the National Institute of Standards and Technology (NIST). Joe's career was primarily with the Department of Defense and NASA as a mathematician in charge of data processing and computations.

Mary's most important publication, NBS Handbook 91: *Experimental Statistics*, is one of the all-time best-selling publications of NIST. Originally published in 1963 by the Government Printing Office, it was later reprinted by Wiley-InterScience in its Selected Government Publications series, and also has been reprinted in paperback by Dover Publications. The ASA Q&P Section established the scholarship in 2000 to honor Mary's 36 years as an author, teacher and consulting statistician.

Each year, the Quality and Productivity Section awards up to two Natrella Scholarships to students who are currently pursuing a master's or doctoral degree full-time in an accredited college or university. The recipients must have a demonstrated interest in quality applications and excellence in teaching and mentoring, engagement and experience in statistical applications, and service and leadership in the statistics community. The recipients must also demonstrate academic excellence and a strong research program. Winners each year receive a $3500 grant, a $500 stipend toward travel and housing expenses, as well as complimentary registration for the conference and the pre-conference short course. Additionally, the winners give a presentation on their statistical work at that year’s QPRC (or JRC). The scholarship is funded by the Q&P Mary G. and Joseph Natrella Scholarship Fund and the Quality and Productivity Research Conference.

The 2024 winners of the scholarship are shown below. They will be presenting their talks in the Natrella Session on Thursday June 20, 2024 between 9:30-10:30am.

**Luke Hagar, PhD Candidate**
Department of Statistics and Actuarial Science
University of Waterloo

**Shancong Mou, PhD Candidate**
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology

# Schedule

## Monday June 17

| | MONDAY JUNE 17 |
|---|---|
| 8:30-9:00 | **Breakfast & Registration** *(Rm: M3 3127)* |
| 9:00-9:30 | **Short Course** *(Rm: M3 3103)* |
| 9:30-10:00 | |
| 10:00-10:30 | |
| 10:30-11:00 | **Coffee Break** *(Rm: M3 3127)* |
| 11:00-11:30 | **Short Course** *(Rm: M3 3103)* |
| 11:30-12:00 | |
| 12:00-12:30 | |
| 12:30-1:00 | **Lunch** *(Rm: M3 3127)* |
| 1:00-1:30 | |
| 1:30-2:00 | **Short Course** *(Rm: M3 3103)* |
| 2:00-2:30 | |
| 2:30-3:00 | |
| 3:00-3:30 | **Coffee Break** *(Rm: M3 3127)* |
| 3:30-4:00 | **Short Course** *(Rm: M3 3103)* |
| 4:00-4:30 | |
| 4:30-5:00 | |

# Tuesday June 18

| | TUESDAY JUNE 18 | | |
|---|---|---|---|
| 8:00-8:45 | **Breakfast & Registration** *(Rm: Fed Hall Foyer)* | | |
| 8:45-9:00 | **Opening Remarks** *(Rm: Main Hall)* | | |
| 9:00-9:30<br>9:30-10:00 | **Stefan Steiner Keynote** *(Rm: Main Hall)* | | |
| 10:00-10:30 | **SigmaXL Demo** *(Rm: Main Hall)* | | |
| 10:30-11:00 | **Coffee Break** *(Rm: Fed Hall Foyer)* | | |
| 11:00-11:30<br>11:30-12:00<br>12:00-12:30 | ***Technometrics*** *(Rm: Main Hall)* Organizer: Bobby Gramacy Chair: David Steinberg Speakers: Anna Malinovskaya, Rong Pan, Raed Kontar | **Screening Experiments & Applications** *(Rm: Columbia A)* Organizer: Maria Weese Chair: Maria Weese Speakers: David Edwards, Caleb King, Bryce Scott | **Machine Learning & AI** *(Rm: Westmount)* Organizer: Siddhesh Kulkarni Chair: Jock MacKay Speakers: David Banks, Cheryl Flynn Brooks, James Wilson |
| 12:30-1:00<br>1:00-1:30 | **Lunch** *(Rm: Main Hall)* | | |
| 1:30-2:00<br>2:00-2:30<br>2:30-3:00 | ***Journal of Quality Technology*** *(Rm: Main Hall)* Organizer: Allison Jones-Farmer Chair: Allison Jones-Farmer Speakers: Mohsen Ebadi, Wenbo Sun, Xinwei Deng | **Advanced Monitoring Methods** *(Rm: Columbia A)* Organizer: Inez Zwetsloot Chair: Fadel Megahed Speakers: Panagiotis Tsiamyrtzis, Valeria Quevedo, Antonio Lepore | **DOE Data Reduction** *(Rm: Westmount)* Organizer: Alan Vázquez Chair: Rakhi Singh Speakers: HaiYing Wang, Rakhi Singh, Ming-Chung Chang |
| 3:00-3:30 | **Coffee Break** *(Rm: Fed Hall Foyer)* | | |
| 3:30-4:00<br>4:00-4:30<br>4:30-5:00 | ***Quality Engineering*** *(Rm: Main Hall)* Organizer: David Edwards Chair: David Edwards Speakers: Steve Rigdon, Yili Hong, Stephen Walsh | **Computer Experiments** *(Rm: Columbia A)* Organizer: Devon Lin Chair: Xinwei Deng Speakers: Youngdeok Hwang, Minshen Xu, Matthew Pratola | **Statistical Process Control** *(Rm: Westmount)* Organizer: Xiulin Xie Chair: Xiulin Xie Speakers: Marcus Perry, Jonathan Stewart, Yicheng Kang |
| 5:00-5:30<br>5:30-6:00 | **Poster Session** *Rm: (Columbia B)* | | |
| 6:00-6:30<br>6:30-7:00<br>7:00-7:30<br>7:30-8:00<br>8:30-9:00 | **Banquet** *(Rm: Main Hall)* | | |

| | | |
|---|---|---|
| Plenary | Invited | Contributed |
| Software Demo | Meals & Social | Meeting |

# Wednesday June 19

| Time | WEDNESDAY JUNE 19 | | |
|---|---|---|---|
| 8:00-8:30 | **Breakfast & Registration**<br>*(Rm: Fed Hall Foyer)* | | |
| 8:30-9:00<br>9:00-9:30 | **Allison Jones-Farmer Plenary**<br>*(Rm: Main Hall)* | | |
| 9:30-10:00 | **JMP Demo**<br>*(Rm: Main Hall)* | | |
| 10:00-10:30 | **Minitab Demo**<br>*(Rm: Main Hall)* | | |
| 10:30-11:00 | **Coffee Break**<br>*(Rm: Fed Hall Foyer)* | | |
| 11:00-11:30<br>11:30-12:00<br>12:00-12:30 | **Data Science in Industry**<br>*(Rm: Main Hall)*<br>Organizer: Ming Li<br>Chair: Peng Liu<br>Speakers: Tim Hesterberg, Roger Hoerl, Emmanuel Yashchin | **Reliability 1**<br>*(Rm: Columbia A)*<br>Organizer: Ichen Lee<br>Chair: Ichen Lee<br>Speakers: Lance Fiondella, Yueyao Wang, Ming-Yung Lee | **Network Science 1**<br>*(Rm: Westmount)*<br>Organizer: James Wilson<br>Chair: James Wilson<br>Speakers: Kevin Xu, Hui Shen, Benjamin Leinwand |
| 12:30-1:00<br>1:00-1:30 | **Lunch**<br>(Rm: Main Hall) | | |
| 1:30-2:00<br>2:00-2:30<br>2:30-3:00 | **Design and Computer Experiments**<br>*(Rm: Main Hall)*<br>Chair: Luke Hagar<br>Speakers: Steven Barnett, Rituparna Dey, Xiankui Yang, Arun Ravichandran, Kyu Min Shim | **Quality, Reliability, & Industry**<br>*(Rm: Columbia A)*<br>Chair: Yueyao Wang<br>Speakers: Alberto Ferrer, Fangyi Luo, Su-Fen Yang, Yuying Hunag | **Advancements in Statistical Methods**<br>*(Rm: Westmount)*<br>Chair: Ming-Chung Chang<br>Speakers: Trangu Bui, Jared Clark, Benedetta Bruni, Youngjin Cho |
| 3:00-3:30<br>3:30-4:00<br>4:00-4:30<br>4:30-5:00<br>5:00-5:30<br>5:30-6:00<br>6:00-6:30<br>6:30-7:00 | **Brewery Tour & Reception**<br>*(Offsite: Royal City Brewing Co.)* | | |

| | | |
|---|---|---|
| Plenary | Invited | Contributed |
| Software Demo | Meals & Social | Meeting |

# Thursday June 20

| | THURSDAY JUNE 20 | | |
|---|---|---|---|
| 8:00-8:30 | **Breakfast & Registration** *(Rm: Fed Hall Foyer)* | | |
| 8:30-9:00<br>9:00-9:30 | **Dave Campbell Plenary** *(Rm: Main Hall)* | | |
| 9:30-10:00<br>10:00-10:30 | **Network Science 2** *(Rm: Main Hall)* Organizer: Srijan Sengupta Chair: James Wilson Speakers: V. Lyzinski, P. Macdonald | **Measurement System Analysis** *(Rm: Columbia A)* Organizer: Nathaniel Stevens Chair: Mahsa Panahi Speakers: B. Lashkari, Adel Nadi | **Natrella Session** *(Rm: Westmount)* Organizer: Will Guthrie Chair: Peng Liu Speakers: Luke Hagar, Shancong Mou |
| 10:30-11:00 | **Coffee Break** *(Rm: Fed Hall Foyer)* | | |
| 11:00-11:30<br>11:30-12:00<br>12:00-12:30 | **SPM in Industry 4.0** *(Rm: Main Hall)* Organizer: Anne Driscoll Chair: Panagiotis Tsiamyrtzis Speakers: Mostafa Reisi, Fadel Megahed, Xiulin Xie | **Reliability 2** *(Rm: Columbia A)* Organizer: Lu Lu Chair: Nathaniel Stevens Speakers: Rong Pan, Maria Jaenada Malagon, Tony Ng | **Computing and Machine Learning** *(Rm: Westmount)* Chair: Jared Clark Speakers: Junoh Heo, Yen-Chun Liu, Dhari Alenezi, Li-Hsiang Lin, Ismail Shah |
| 12:30-1:00<br>1:00-1:30 | **Lunch & Closing Remarks** *(Rm: Main Hall)* | | |
| 1:30-2:00<br>2:00-2:30<br>2:30-3:00 | **QPRC Management Meeting** *(Rm: Columbia A)* | | |

| | | |
|---|---|---|
| Plenary | Invited | Contributed |
| Software Demo | Meals & Social | Meeting |

# Abstracts (Tuesday June 18)

## Keynote Session: Stefan Steiner

*Tuesday, June 18th (9:00am-10:00am)*

*Room: Main Hall*

### Using Non-random Sample Selection in Variation Reduction Projects

*Stefan Steiner, University of Waterloo*

In this talk we will review the use of non-random sample selection, i.e., selecting parts systematically or selecting extreme parts [called leveraging], in process improvement (variation reduction) projects. The talk is motivated by an industrial case study from my consulting experience concerning vehicle headrest failures. The primary goals of the talk are to show the breadth of applicability and the benefits of systematic and leveraged sample selection. We illustrate the ideas using baseline, measurement system assessment, disassembly/reassembly, component swap, and group comparison investigations.

---

## Software Demo Session: SigmaXL

*Tuesday, June 18th (10:00am-10:30am)*

*Room: Main Hall*

### Statistical Process Control with SigmaXL

*John Noguera, CTO & Co-founder of SigmaXL Inc.*

SigmaXL is a cost-effective and easy to use Excel Add-in for statistical and graphical analysis. This presentation will give an overview of the software and demonstrate some practical and powerful SPC tools. It will consider SPC for autocorrelated data using residuals from automated time series forecasts (e.g., automatic ETS exponential smoothing and seasonal ARIMA with an example using modified Box-Jenkins series G monthly airline passenger data). It will also consider the use of Average Run Length Templates to optimize control charts (e.g., Shewhart charts with tests for special causes, EWMA and CUSUM).

---

## Invited Session: *Technometrics*

*Tuesday, June 18th (11:00am-12:30pm)*

*Room: Main Hall*

*Organizer: Bobby Gramacy, Virginia Tech*

*Chair: David Steinberg, Tel-Aviv University*

### Statistical Process Monitoring of Artificial Neural Networks

*Anna Malinovskaya, Leibniz University Hannover, Institute of Cartography and Geoinformatics*

The rapid progress in artificial intelligence models necessitates the development of innovative real-time monitoring techniques with minimal computational overhead. Particularly in machine learning, where artificial neural networks (ANNs) are commonly trained in a supervised manner, it becomes crucial to ensure that the learned relationship between input and output remains valid during the model's deployment. If this stationarity assumption holds, we can conclude that the ANN provides accurate predictions. Otherwise, the retraining or rebuilding of the model is required. This talk focuses on examining the latent feature representation of data, referred to as "embedding", generated by ANNs to identify the time point when the data stream starts being nonstationary. The proposed monitoring approach employs embeddings and utilizes multivariate control charts based on data depth calculations and normalized ranks. The method's performance is thoroughly compared to benchmark approaches, accounting for various existing ANN architectures and underlying data formats. The goal is to assess its effectiveness in detecting nonstationarity in real time, offering insights into the validity of the model’s output.

## A Class of Hierarchical Multivariate Wiener Processes for Modeling Dependent Degradation Data

*Rong Pan, Arizona State University*

In engineering practice, many products exhibit multiple and dependent degrading performance characteristics (PCs). It is common to observe that these PCs’ initial measurements are nonconstant and sometimes correlated with the subsequent degradation rate, which typically varies from one unit to another. To accommodate the unit-wise heterogeneity, PC-wise dependency, and “initiation-growth” correlation, this article proposes a broad class of multi-dimensional degradation models under a framework of hierarchical multivariate Wiener processes. These models incorporate dual multi-normally distributed random effects concerning the initial values and degradation rates. To infer model parameters, expectation-maximization (EM) algorithms and several tools for model validation and selection are developed. Various simulation studies are carried out to assess the performance of the inference method and to compare different models. Two case studies are conducted to demonstrate the applicability of the proposed methodology.

## Personalized Federated Learning via Domain Adaptation with an Application to Distributed 3D Printing

*Raed Al Kontar, University of Michigan*

Over the years, Internet of Things (IoT) devices have become more powerful. This sets forth a unique opportunity to exploit local compute resources to distribute model learning and circumvent the need to share raw data. The underlying distributed and privacy-preserving data analytics approach is often termed federated learning (FL). A key challenge in FL is the heterogeneity across local datasets. In this paper, we propose a new personalized FL model, PFL-DA, by adopting the philosophy of domain adaptation. PFL-DA tackles two sources of data heterogeneity at the same time: a covariate and concept shift across local devices. We show, both theoretically and empirically, that PFL-DA overcomes intrinsic shortcomings in state of the art FL approaches and is able to borrow strength across devices while allowing them to retain their own personalized model. As a case study, we apply PFL-DA to distributed desktop 3D printing where we obtain more accurate predictions of printing speed, which can help improve the efficiency of the printers.

## Invited Session: Screening Experiments & Applications

*Tuesday, June 18th (11:00am-12:30pm)*

*Room: Columbia A*

*Organizer: Maria Weese, Miami University*

*Chair: Maria Weese, Miami University*

### Row-Constrained Supersaturated Designs for Biological Screening

*David Edwards, Virginia Commonwealth University*

High-throughput screening is widely used across many areas of the biological sciences, perhaps most prominently in drug discovery. In this talk, we propose a statistically principled approach to these screening experiments, using the machinery of supersaturated designs. To accommodate limitations on the number of biological entities that can be applied to a single microplate well, we present a new class of row-constrained supersaturated designs. We develop a computational procedure to construct these designs, provide some initial lower bounds on the average squared off-diagonal values of their main effect information matrix, and study the effect of the constraint on design quality. We also show that the proposed method is statistically superior to existing methods, and demonstrate the use of the new methodology on a real drug-discovery system.

### Designing an Array of Arrays and Other Case Studies in Applying DOE to the Testing and Validation of Statistical Software

*Caleb King, JMP Statistical Discovery, LLC*

Validating statistical software involves a variety of challenges. Of these, the most difficult is the selection of an effective set of test cases, sometimes referred to as the "test case selection problem". To further complicate matters, for many statistical applications, development and validation are done by individuals who often have limited time to validate their application and may not have formal training in software validation techniques. Thus, it is imperative that the adopted validation method be both efficient and effective, as well as easily understood by individuals not trained in software validation techniques. As it turns out, the test case selection problem can be thought of as a design of experiments (DOE) problem. In this talk, we describe examples of how we have applied DOE principles in the testing and validation of statistical software, including how we addressed some of the challenges specific to that type of software. We also briefly discuss some recent research conducted by our group on the problem of "fault localization", where the task is to identify the combinations of inputs that cause observed faults in combinatorial testing.

### How to Optimize 3D Printing for Aerospace Applications

*Bryce Scott, North Carolina State University*

According to Allied Market Research, the 3D printing market was valued at $13.2 billion in 2020 and is expected to reach $94 billion by 2030. Additive manufacturing is a rapidly growing industry with many applications in high tech industries, specifically in the aerospace industry. Given its need for complex parts, the aerospace industry can benefit greatly from the design freedom that 3D printing allows.

How can you use 3D printing if you have no experience? There are myriad settings and variables at play, but it can be difficult to understand exactly how each variable affects things such as print time or quality. Using design of experiments (DOE) and group orthogonal supersaturated designs (GOSSD), we can discover the effects that these variables have on a product.

---

## Invited Session: Machine Learning & AI

*Tuesday, June 18th (11:00am-12:30pm)*

*Room: Westmount*

*Organizer: Siddhesh Kulkarni, Bristol Meyers Squibb*

*Chair: Jock MacKay, University of Waterloo*

### Statistics, AI, and the Knowledge Economy

*David Banks, Duke University*

Statistics came of age when manufacturing was king. But today's industries are focused on information technology. Remarkably, a lot of our expertise transfers directly. This talk will discuss statistics and AI in the context of computational advertising, autonomous vehicles, large language models, and process optimization.

### Responsible Generative AI and Industry Challenges

*Cheryl Flynn Brooks, AT&T*

The emergence of generative AI models is driving new applications and innovations across industries. However, the adoption of these models brings new ethical challenges, and responsible deployment is critical to protect business integrity and brand image. In this talk, we will give an overview of different aspects of responsible generative AI. We will discuss topics including hallucinations in Large Language Models, data leakage and privacy concerns, as well bias and fairness. We will discuss techniques for addressing these issues and delve deeper into open challenges facing industries today.

### Nonparametric feature impact and importance

*James Wilson, University of San Francisco*

Measures of feature importance is used to rank predictors during statistical and machine learning model development in an effort to simplify models and improve generality. Unfortunately, feature importance is often conflated with feature impact, the isolated effect of an explanatory variable on the response variable. This can lead to real-world consequences when importance is inappropriately interpreted as impact, particularly for business or medical insight purposes. The dominant approach for computing importances is through interrogation of a fitted model, which works well for feature selection, but gives distorted measures of feature impact. The same method applied to the same data set can yield different feature importances, depending on the model, leading us to conclude that impact should be computed directly from the data. In this talk, I will give mathematical definitions of feature impact and importance, derived from partial dependence curves, that operate directly on the data. To assess quality, I will show that features ranked by these measures often outperform common existing feature selection techniques using three real data sets.

---

## Invited Session: *Journal of Quality Technology*

*Tuesday, June 18th (1:30pm-3:00pm)*

*Room: Main Hall*

*Organizer: Allison Jones-Farmer, Miami University*

*Chair: Allison Jones-Farmer Miami University*

### Phase I analysis of high-dimensional processes in the presence of outliers

*Mohsen Ebadi, University of Waterloo*

One of the significant challenges in monitoring the quality of products today is the high dimensionality of quality characteristics. We address Phase I analysis of high-dimensional processes with individual observations when the available number of samples collected over time is limited. Using a new charting statistic, we propose a robust procedure for parameter estimation in Phase I. This robust procedure is efficient in parameter estimation in the presence of outliers or contamination in the data. A consistent estimator is proposed for parameter estimation and a finite sample correction coefficient is derived and evaluated through simulation. We assess the statistical performance of the proposed method in Phase I. This assessment is carried out in the absence and presence of outliers. We show that, in both cases, the proposed control chart scheme effectively detects various kinds of shifts in the process mean. Besides, we present two real-world examples to illustrate the applicability of our proposed method.

### A Continual Learning Framework for Adaptive Defect Classification and Inspection

*Wenbo Sun, University of Michigan Transportation Research Institute*

Machine-vision-based defect classification techniques have been widely adopted for automatic quality inspection in manufacturing processes. This article describes a general framework for classifying defects from high volume data batches with efficient inspection of unlabelled samples. The concept is to construct a detector to identify new defect types, send them to the inspection station for labelling, and dynamically update the classifier in an efficient manner that reduces both storage and computational needs imposed by data samples of previously observed batches. Both a simulation study on image classification and a case study on surface defect detection via 3D point clouds are performed to demonstrate the effectiveness of the proposed method.

### Adaptive-Region Sequential Design with Quantitative and Qualitative Factors in Application to HPC Configuration

*Xinwei Deng, Department of Statistics, Virginia Tech*

Motivated by the need to find optimal configurations in the high-performance computing (HPC) system, an adaptive-region sequential design (ARSD) is proposed for the optimization of computer experiments with qualitative and quantitative factors. Experiments with both qualitative and quantitative factors are also encountered in other applications. The proposed ARSD method considers a sequential design criterion under the additive Gaussian process to deal with both qualitative and quantitative factors. Moreover, the adaptiveness of the proposed sequential procedure allows the selection of the next design point from the adaptive design region, achieving a meaningful balance between exploitation and exploration for optimization. Theoretical justification of the adaptive design region is provided. The performance of the proposed method is evaluated by several numerical examples in simulations. The case study of HPC performance optimization further elaborates on the merits of the proposed method.

---

## Invited Session: Advanced Monitoring Methods

*Tuesday, June 18th (1:30pm-3:00pm)*

*Room: Columbia A*

*Organizer: Inez Zwetsloot, University of Amsterdam*

*Chair: Fadel Megahed, Miami University*

### A Bayesian Self-starting Hotelling (BSSH) T2 for Online Multivariate Outlier Detection

*Panagiotis Tsiamyrtzis, Politecnico di Milano & Athens University of Economics and Business*

Online outlier detection in multivariate settings is a topic of high interest in several scientific fields, with the Hotelling's T2 control chart being probably the most widely used method in practice to treat it. The problem becomes challenging though when we lack the ability to perform a proper phase I calibration, like in short runs or in cases where online inference is requested from the start of the process, as in biomedical applications. In this work, we propose a Bayesian self-starting version of the Hotelling's T2 control chart, for multivariate normal data when all parameters are unknown. A conjugate power prior will allow to incorporate different sources of information (when available), providing closed form expressions that are straightforward to be used in practice and most importantly, will allow online inference, breaking free from the phase I calibration stage. From a theoretic perspective, we determine the power of the proposed scheme in detecting a fixed size outlier in the mean vector and we discuss its properties. Apart from monitoring, we deal also with the post-alarm inference aiming to provide the likely source(s) of an alarm, enriching the practitioners root-cause analysis toolbox. A simulation study evaluates the performance of the proposed control chart against its competitors, while topics regarding its robustness are also covered. An application to real data will illustrate its practical use.

### Profile monitoring in aquaculture

*Valeria Quevedo, Universidad de Piura*

There is extensive research in monitoring a process whose characteristics are represented as profiles. The research in this growing literature has made extensive use of both parametric and non-parametric methods in profile monitoring analysis. Most of the parametric and non-parametric techniques for profile monitoring focus on the parameters or features of the model instead of the actual quality characteristic observed. Their focus is on detecting process changes based on the statistical significance of the estimators, which often produce relatively unimportant changes in the true quality characteristic. Additionally, most techniques require all observations from each profile to determine the process state. We study the use of a Shewhart chart based on a Gaussian process model with heteroscedasticity for the online monitoring of profiles, while these are being developed, where the central line is the predictive mean and the control limits are based on the prediction band. The advantage is that we do not have to wait until a profile ends to make process corrections. Our results indicate that our method is effective. We applied our approaches in an aquaculture process to monitor the shrimp weight over 300 ponds.

## An outlier-resistant multivariate profile monitoring framework for a spot welding process in the automotive industry

*Antonio Lepore, University of Naples FEDERICO II*

The quality of a resistance spot welding (RSW) process in the automotive industry is critical to guarantee the structural integrity and solidity of welded assemblies in each vehicle. Among online measurements of RSW process parameters, the dynamic resistance curve (DRC) is recognized as the most informative technological signature of the metallurgical development of a spot weld. The p-dimensional vector of DRCs pertaining to the (p) spot welds of interest on each item can be then used as a multivariate functional quality characteristic of interest for statistical process monitoring (SPM) purposes. Unfortunately, the larger the dimension p, the smaller the fraction of perfectly observed cases. This constitutes an issue in traditional control charting procedures, which are very sensitive to outliers in the Phase I sample, as these may markedly inflate the control interval width and reduce the power to detect process changes in Phase II. To mitigate this issue, traditional robust methods usually down-weight the influence of all p components (casewise), even when outliers occur in one or a few components alone (componentwise). Motivated by the SPM of the aforementioned RSW process, our proposal is a multivariate profile monitoring framework that is resistant to the presence of functional casewise and componentwise outliers in the Phase I sample. The proposed framework is implemented in the R package funcharts, openly available on CRAN, and is compared to alternative monitoring schemes that have already appeared in the literature before.

Acknowledgements

The research activity of A. Lepore was carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1551.11-10-2022, PE00000004). This work reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

---

# Invited Session: DOE Data Reduction

*Tuesday, June 18th (1:30pm-3:00pm)*

*Room: Westmount*

*Organizer: Alan Vázquez, Tecnologico de Monterrey*

*Chair: Rakhi Singh, Binghamton University*

## Subsampling for big rare events data beyond binary responses

*HaiYing Wang, University of Connecticut*

Rare events data occur when certain types of events occur with very small probabilities. Subsampling is very effective to reduce the computational cost of analyzing rare events data with losing significant estimation efficiency. Existing investigations on subsampling with rare events data focus on binary response models. We investigate rare events data beyond binary responses. If sufficient data points for the non-rare observations are sampled, there will be no statistical efficiency loss. In the scenario that there is estimation efficiency loss due to down sampling, we developed optimal sampling design to minimize the information loss.

## A model-free subdata selection method for classification

*Rakhi Singh, Binghamton University*

Subdata selection is a study of methods that select a small representative sample of the big data, the analysis of which is fast and statistically efficient. The existing subdata selection methods assume that the big data can be reasonably modeled using an underlying model, such as a (multinomial) logistic regression for classification problems. These methods work extremely well when the underlying modeling assumption is correct but often yield poor results otherwise. In this talk, I will introduce a model-free subdata selection method for classification problems, called PED. The PED subdata uses decision trees to find a partition of the data, followed by selecting an appropriate sample from each component of the partition. Random forests are used for analyzing the selected subdata. Our method can be employed for a general number of classes in the response and for both categorical and continuous predictors. We study the theoretical properties of PED and show that the PED subdata results in a smaller Gini than a uniform subdata. Further, we demonstrate that the PED subdata has higher classification accuracy than other competing methods through extensive simulated and real datasets.

## Supervised Stratified Subsampling for Regression Problems

*Ming-Chung Chang, Institute of Statistical Science, Academia Sinica*

Predictive analytics involves the use of statistical models to make predictions; however, the power of these techniques is hindered by ever-increasing quantities of data. The richness and sheer volume of big data can have a profound effect on computation time and/or numerical stability. In this presentation, I will introduce a novel approach to subsampling with the aim of overcoming this issue when dealing with regression problems in a supervised learning framework. The proposed method integrates stratified sampling and is model-independent. We assess the theoretical underpinnings of the proposed subsampling scheme, and demonstrate its efficacy in yielding reliable predictions with desirable robustness when applied to different statistical models.

---

# Invited Session: *Quality Engineering*

*Tuesday, June 18th (3:30pm-5:00pm)*

*Room: Main Hall*

*Organizer: David Edwards, Virginia Commonwealth University*

*Chair: David Edwards, Virginia Commonwealth University*

## First to Signal Criterion for Comparing Control Chart Performance

*Steve Rigdon, Saint Louis University*

Control chart performance is often measured using average run length or median run length, which gives the expected or median number of samples to signal. It is often argued that on average, one chart will signal a process change quicker than another, and is therefore a better choice. Average and median run length do not, however, answer the question of which method will be more likely to signal first. We introduce the idea of “first to signal” and compare charts based on this criterion.

## Statistical perspectives on reliability of artificial intelligence systems

*Yili Hong, Virginia Tech*

Artificial intelligence (AI) systems are increasingly popular in many applications. Nevertheless, AI technologies are still developing, and many issues need to be addressed. Among those, the reliability of AI systems needs to be demonstrated so that AI systems can be used with confidence by the general public. In this talk, we provide statistical perspectives on the reliability of AI systems, focusing on the time dimension. That is, the system can perform its designed functionality for the intended period of time. We introduce a so-called "SMART" statistical framework for AI reliability research, which includes five components: Structure of the system, Metrics of reliability, Analysis of failure causes, Reliability assessment, and Test planning. We review traditional methods in reliability data analysis and software reliability, and discuss how those existing methods can be transformed for reliability modeling and assessment of AI systems. Different from traditional reliability studies, the focus of AI reliability is on the software system to include the training data. Thus, we describe recent developments in modeling and analysis of AI reliability for software systems. The talk outlines statistical research challenges in this area, including out-of-distribution detection, the effect of the training set, adversarial attacks, model accuracy, and uncertainty quantification. We discuss how those topics can be related to AI reliability, with illustrative examples. The final element of SMART(test planning), is critical for the demonstration of AI reliability. Therefore we discuss data collection and testing planning, highlighting methods for improving system design in order to achieve higher AI reliability.

## I-optimal or G-optimal: do we have to choose?

*Stephen Walsh, Utah State University*

In optimizing a designed experiment for effective prediction using a second-order response surface model, it's common to consider either G-optimality for optimal worst-case prediction precision or I-optimality for optimal average prediction precision. The I-criterion has been more popular due to its computational simplicity relative to the G-criterion. In this presentation, we'll provide an overview of the development and practical uses of these two design criteria, tracing their historical evolution and impact. Next, we'll explore the concept of multi-objective optimal design, a strategy seeking designs that excel in both criteria and elucidates their trade-offs. Additionally, we'll introduce the Pareto Front, theoretical solution to the multi-objective optimal design problem, which constitutes a set of optimal designs that considers different practitioner preferences on each criterion. We then illustrate how we generate Pareto Fronts on a fully continuous design space via application of particle swarm optimization and discuss our approach for selecting candidate designs from the Pareto Front and assessing their quality for achieving desired balance between experimental goals. Last, we briefly describe the large, database of I/G-optimal Pareto Fronts generated on 17 common response surface design scenarios for two to four factors available on the Quality Engineering website and illustrate the candidate design analysis and selection process on a specific case.

## Invited Session: Computer Experiments

*Tuesday, June 18th (3:30pm-5:00pm)*

*Room: Columbia A*

*Organizer: Devon Lin, Queen's University*

*Chair: Xinwei Deng, Virginia Tech*

### Physics-informed spatio-temporal models

*Youngdeok Hwang, Baruch College, City University of New York*

Statistical models are often derived from physics equations describing dynamical systems. When studying systems heavily influenced by particle transport via wind, it is crucial to consider atmospheric conditions. With the wide availability of numerical weather prediction models (NWPs), it is now possible to utilize atmospheric conditions produced by NWPs. In this talk, we will briefly discuss recent developments in spatio-temporal modeling approaches that incorporate physical knowledge to consider various atmospheric conditions, while leveraging information available from NWPs. We will also discuss ongoing efforts from both planning and modeling perspectives.

### Dimension Reduction for Gaussian Process Models via Convex Combination of Kernels

*Minshen Xu, University of Massachusetts Amherst*

Many computer simulation models in engineering and scientific domains involve a large number of input variables, which can result in high computational cost and low prediction accuracy for the Gaussian process (GP) regression model. However, some simulation models may only be significantly influenced by a small subset of the input variables, referred to as the "active variables". Identifying these active variables can help researchers overcome the two limitations of the GP model and gain a better understanding of the simulated system. To achieve this goal, we propose an approximation of the covariance function of the original GP model involving all the input variables. The approximation is through a convex combination of kernel functions whose input variables are low-dimensional subsets of the complete input variables. To determine the optimal approximation, we develop an iterative algorithm based on the Fedorov-Wynn algorithm from the optimal design literature. We also incorporate the effect heredity principle while selecting the active input variables, which ensures sparsity. Through several examples, we have shown the proposed method outperforms some alternative approaches in correctly identifying the active input variables.

### Bayesian Model Mixing with Applications in Nuclear Physics and Climate

*Matthew Pratola, The Ohio State University*

In many areas of physics and climate modeling, one often encounters the situation where there are multiple competing theoretical models of the physical phenomena being studied, each implemented as a simulator on a computer. In general, it is not known if there is a single global best simulator, or if even any of the simulators are representative of reality. An important question in such a setting is determining the best simulator, or the best combination of simulators, to use for prediction and inference. Bayesian model averaging (BMA) and stacking are two statistical approaches used for learning about such model uncertainties. However, they generally assume a global best model, whereas in our applications simulators are only locally best. To facilitate uncertainty quantification in this setting, Bayesian model mixing (BMM) extends these ideas to capture the localized behavior of each simulator by defining input-dependent weights. This talk outlines a BMM model based on Bayesian Additive Regression Trees (BART) for learning such localized weights in a flexible, non-parametric approach. The proposed methodology is applied to motivating problems in Effective Field Theories (EFTs) for nuclear physics applications, and in ensembles of Global Climate Models (GCMs) used in studying climate change.

---

## Invited Session: Statistical Process Control

*Tuesday, June 18th (3:30pm-5:00pm)*

*Room: Westmount*

*Organizer: Xiulin Xie, Florida State University*

*Chair: Xiulin Xie, Florida State University*

### Joint monitoring of location and scale for modern univariate processes

*Marcus Perry, University of Alabama*

Autocorrelated sequences of individual observations arise in many modern-day statistical process monitoring (SPM) applications. Often times, interest involves jointly monitoring both process location and scale. To jointly monitor autocorrelated individuals data, it is common to first fit a time series model to the in-control process and subsequently use this model to de-correlate the observations so that traditional individuals and moving-range (I-MR) charts can be applied. If the time series model is correctly specified such that the resulting residuals are normal and independently distributed, then applying I-MR control charts to the residual process should work well. However, if the residual process deviates from normality and/or, due to time series model misspecification, contains levels of autocorrela- tion, the false alarm rate of such a strategy can dramatically rise. In this paper we propose improvements to a recently published distribution-free joint monitoring strategy that permits tighter control of the in-control average run length when model misspecification is a concern. We evaluate its average run length performance and conclude that the improved joint monitoring strategy proposed in this work is a very useful tool for today's modern SPM practitioner. The proposed joint monitoring scheme is then applied to a real additive manufacturing process to illustrate its implementation in modern practice.

### Online profile monitoring of random functions using correlation structure

*Jonathan Stewart, Florida State University*

We consider the problem of online profile monitoring of random functions that admit basis expansions possessing Gaussian coefficients for the purpose of out-of-control state detection. In this work, we consider random functions that feature two sources of variation: additive error and random fluctuations through random coefficients in the basis representation of functions. We focus on a two-phase monitoring problem with a first stage consisting of learning the in-control process and the second stage leveraging the learned process for out-of-control state detection. We outline a learning and monitoring methodology for our problem by exploiting the Gaussian process of the basis coefficients to develop a scalable and effective monitoring methodology for the observed processes that makes weak functional assumptions on the underlying process. We will demonstrate the potential of our method through simulation studies that highlight some of the nuances that emerge in profile monitoring problems with random functions. Lastly, we will discuss potential refinement and extensions of our methodology.

## Deep Vision in Smart Manufacturing: MODERN Framework for Intelligent Quality Monitoring and Diagnosis

*Yicheng Kang, Miami University*

Smart manufacturing processes are often installed with a large number of sensors, imaging devices and computers, which not only enable instant communication across various modules of a production system but also aid in intelligent manufacturing management. It is critically important to integrate these enhanced capabilities into the practice of industrial quality control. In this paper, we introduce a deep learning framework for quality monitoring and fault isolation. Using the architecture of an inception residual neural network, we develop a control chart that monitors the likelihood of a product containing defects. We also propose a faulty region estimator that identifies the defective area using transfer learning. To further extend our framework to cases where there is not enough training data, we suggest a transfer monitoring technique that requires a much smaller sample size and a hypothesis testing approach for quantitatively assessing the applicability of our method. Theoretically, we establish the minimax optimal convergence rate for both our defect likelihood estimation and fault diagnosis. Our results lead to a seemingly counterintuitive managerial implication - it may not always be in the manufacturers' best interests to keep upgrading their monitoring equipment regardless of the cost. Empirically, we demonstrate the superior performance of our method in comparison with a state-of-the-art approach using both simulated experiments and real data

# Abstracts (Wednesday June 19)

## Plenary Session: Allison Jones-Farmer

*Wednesday, June 19th (8:30am-9:30am)*

*Room: Main Hall*

### Our Tech-Enabled Future: How Artificial Intelligence is Shaping the Future of Analytics Practice, Research, and Education

*Allison Jones-Farmer, Miami University*

The hallmarks of Industry 4.0 are the advancement of automation and interconnectivity. This tech integration has led to the evolution of Big Data, transforming how we live, communicate, and work. In this talk, we will briefly discuss how Artificial Intelligence (AI) and, more specifically, Generative AI (Gen-AI) emphasize personalization and human-machine collaboration. We highlight how these changes require a paradigm shift in research and education, challenging traditional models and requiring new approaches that align with the demands of a rapidly evolving digital economy. A key part of the discussion will be the abilities and limitations of Gen-AI. We will reflect on the changing roles of researchers, professionals, and educators and encourage the community to collaborate, contribute, and innovate. The goal is to inspire researchers, educators, and practitioners to embrace and contribute to the ongoing technological advances that will inevitably define the future of research, education, and practice in data science fields.

---

## Software Demo Session: JMP

*Wednesday, June 19th (9:30am-10:00am)*

*Room: Main Hall*

### Empowering Industrial Innovation through Statistical Discovery

*Nick Shelton, JMP System Engineer Manager, JMP Statistical Discovery*

Delve into the world of statistical discovery with JMP software. This seminar, tailored for industrial applications, will empower you to conquer real-world challenges in quality control, process optimization, and product innovation with precision and confidence. Witness the power of interactive visualizations and see how these techniques are revolutionizing decision-making and providing actionable insights propelling organizations forward. Don't miss this opportunity to be part of the leading edge of data science and statistics driving industrial innovation.

---

## Software Demo Session: Minitab

*Wednesday, June 19th (10:00am-10:30am)*

*Room: Main Hall*

### Statistical Tools Everyone Can Use

*Scott Kowalski, Senior Manage Global Services, Minitab LLC*

Regardless of analytical background, Minitab Statistical Software (MSS) empowers all parts of an organization to explore data, illustrate insights, and predict future trends. MSS can help visualize, analyze, and harness the power of your data to solve your toughest challenges. Enjoy new levels of productivity and collaboration with Minitab accessible through both your desktop and on the cloud with our web app. Uncover hidden relationships between variables and visualize data interactions using the interactive graph builder. Predictive Analytics tools help to get the most out of your data: discover trends, predict patterns, and identify important factors to answer even the most challenging of questions. In this tutorial, some of the new features of MSS will be presented.

## Invited Session: Data Science in Industry

*Wednesday, June 19th (11:00am-12:30pm)*

*Room: Main Hall*

*Organizer: Ming Li, Coupang*

*Chair: Peng Liu, SAS Institute, Inc.*

### Statistics and Data Science at Instacart and Google

*Tim Hesterberg, Instacart*

I'll share some stories from work at Instacart and Google, focusing on cases where statistics and data science education has room for improvement - where applications fail, or where better understanding of statistical principles and methods would lead to better analysis.

### Data Science Can Learn from the Successes and Shortcomings of the Statistics Discipline

*Roger Hoerl, Union College*

The modern discipline of data science is a few decades old. The statistics discipline, on the other hand, is a couple of centuries old. I argue that like any new discipline, data science has had a successful but somewhat rocky beginning. Further, I argue that this new discipline could learn much from the statistics discipline's centuries of experience, both positive and negative. That is, by imitating the successes, while avoiding the potholes of the shortcomings of the statistics discipline, data science can enhance its own success and long-term prospects significantly, and remain relevant for the next couple of centuries.

## Design and analysis of early warning systems for sensor-enabled equipment

*Emmanuel Yashchin, American University*

In this paper, we delve into a challenge frequently encountered by statistics consultants: developing an initial version of an early warning system (EWS) following a crisis precipitated by repeated failures of a critical piece of equipment. At such a juncture, sensor data becomes invaluable, despite typically being incomplete and possibly compromised by issues such as missing entries, distortions from faulty sensors, and the absence of maintenance records for the equipment. Constructing the EWS requires a heavy reliance on domain knowledge, as many failure modes the system needs to detect may not be reflected in the available data streams. The development of the EWS is a multi-phase process, with a crucial goal being to convince the client of the eventual efficacy of the system in terms of detection speed, false alarm rate and diagnostic accuracy. Throughout the EWS design process, the consultant is likely to navigate not only technical and modeling hurdles, but also ethical dilemmas associated with the use of observational data. This paper discusses a strategy for initiating the EWS and demonstrates its application in a practical setting.

---

# Invited Session: Reliability 1

*Wednesday, June 19th (11:00am-12:30pm)*

*Room: Columbia A*

*Organizer: I-Chen Lee, National Cheng Kung University*

*Chair: I-Chen Lee, National Cheng Kung University*

## Quantitative Assessment of Machine Learning Reliability and Resilience

*Lance Fiondella, University of Massachusetts Dartmouth*

Advances in machine learning (ML) have led to applications in safety-critical domains, including security, defense, and healthcare. These ML models are confronted with dynamically changing and actively hostile conditions characteristic of real-world applications, requiring systems incorporating ML to be reliable and resilient. Many studies propose techniques to improve the robustness of ML algorithms. However, fewer consider quantitative techniques to assess the reliability and resilience of these systems. To address this gap, this talk demonstrates how to collect relevant data during the training and testing of ML suitable for the application of software reliability, with and without covariates, and resilience models and the subsequent interpretation of these analyses. The proposed approach promotes quantitative risk assessment of machine learning technologies, providing the ability to track and predict degradation and improvement in the ML model performance and assisting ML and system engineers with an objective approach to compare the relative effectiveness of alternative training and testing methods. The approach is illustrated in the context of an image recognition model, which is subjected to two generative adversarial attacks and then iteratively retrained to improve the system's performance. Our results indicate that software reliability models incorporating covariates characterized the misclassification discovery process more accurately than models without covariates. Moreover, the resilience model based on multiple linear regression incorporating interactions between covariates tracks and predicts degradation and recovery of performance best. Thus, software reliability and resilience models offer rigorous quantitative assurance methods for ML-enabled systems and processes.

## Variational Inference for Spatial Correlated Failure Time Data under Bayesian Framework

*Yueyao Wang, Zhejiang Gongshang University*

In the modern reliability analysis, the geo-graphically referenced time-to- event data are often collected for analysis. For such reliability data, the spatial dependence among the failure time needs to be properly accounted in the model. In the literature, spatial random effect models, such as the cumulative exposure model is often used for analysis with the Bayesian approach as model inference. However, the inference problem is often high dimensional with respect to the number of spatial locations. Consequently, the conventional Markov Chain Monte Carlo (MCMC) methods for sampling the posterior can be time- consuming when the number of spatial locations is large. In this paper, we investigate the capability of variational inference (VI) for the inference of spatial survival models and aim for a good balance between the estimation accuracy and computational efficiency. Specifically, two divergence metrics, α−divergence and the Kullback Leibler (KL) divergence are used in the VI methods for the spatial cumulative exposure model. In the numerical study, we compare the MCMC and VI methods under spatial GPU lifetime data. The comparison results show that the VI method has a comparable performance as the MCMC approach but with much more efficient computational time.

## Under Time Censored, Use Intermediate Data to Estimate Lifetime Distribution and Model Parameters

*Ming-Yung Lee, Providence University*

Traditionally, degradation models have been explored by measuring degradation data at a fixed time. Due to technological advancement, the measurement methods of accelerated degradation experiments have also changed. Tang and Su (2008) proposed to set several pseudo thresholds before the failure threshold, called nonfailure threshold. And collect the crossing time when these non-failure thresholds are passed for the first time, which is called intermediate data. And use this information to estimate the product's lifetime distribution. However, when the reliability experiment is actually performed, since the experiment time is not long, even if the non-failure threshold is set quite low, the collected intermediate data may still be censored by time. Since traditional lifetime tests only set a failure threshold, only one piece of lifetime data or a censored time can be collected for each test sample. When several non-failure thresholds are set, more sample size of intermediate data can be collected or the censored time are set for each sample. This study will generalize the results of Tang and Su (2008) and explore the lifetime and model parameters estimation under time constraints. In addition to estimating model parameters, mean and median lifespans, and cumulative probabilities, an approximate 100(1-α)% confidence interval or region will be estimated and a contour map will be drawn. Statistical simulations will be performed to examine small sample results. Finally, the high-power LED data proposed by Tan and Singh (2023) will be used as an example analysis.

---

# Invited Session: Network Science 1

*Wednesday, June 19th (11:00am-12:30pm)*

*Room: Westmount*

*Organizer: James Wilson, University of San Francisco*

*Chair: James Wilson, University of San Francisco*

## Latent variable models for continuous-time dynamic networks

*Kevin Xu, Case Western Reserve University*

Networks are ubiquitous in science, serving as a natural representation for many complex physical, biological, and social systems. Probabilistic generative models for networks provide plausible mechanisms by which network data are generated to reveal insights about the underlying complex system. Such complex systems are often time-varying, which has led to the development of dynamic network representations to enable modeling, analysis, and prediction of temporal dynamics. In this talk, I introduce two continuous-time probabilistic generative models for dynamic networks that augment latent variable models for network structure with multivariate Hawkes processes to model temporal dynamics. First is the multivariate community Hawkes (MULCH) model that uses a discrete latent variable representation built on the stochastic block model. Next, I introduce the latent space Hawkes (LSH) model that uses a continuous latent variable representation built on the latent space model. I demonstrate how these models can be used for analysis, prediction, and simulation on several real network data sets, including a network of militarized disputes between countries over time.

## Markovian Change point detection models for evolving networks

*Hui Shen, McGill University*

Networks, namely graphs with context dependent structure, form fundamental building blocks of a host of data analytic pipelines in statistics and machine learning. In many settings (e.g. social networks) such systems undergo temporal evolution leading to the important sub-field of dynamic networks. Increasingly important in this domain are questions of anomaly detection, namely stretches of time where the evolution of the underlying network is “atypical”. Despite significant push from applications, there is much less known for change point detection for networks,  both in terms of benchmark models and performance of standard estimators such as CUSUM-based statistics in such settings. Existing treatments of networks with fixed vertex size often assume statistical independence across time slices. In this work, we propose a general class of Markovian network models for incorporating change points. To detect the location of change point, we consider a general class of CUSUM-based testing statistics, amalgamating both single-layer and multilayer graph structures to detect different types of change point phenomena. The overarching goal is to understand the delicate interplay between the performance of standard estimators, time scales that modulate network dynamics and macroscopic phenomena such as propagation of chaos in high-dimensional networked systems.

## Likelihoods of Weight Loss or: ACRONYM: Augmented degree corrected, Community Reticulately Organized Network Yielding Model

*Benjamin Leinwand, Stevens Institute of Technology*

Modeling networks can serve as a means of summarizing high-dimensional complex systems. Adapting an approach devised for dense, weighted networks, we propose a new method for generating and estimating unweighted networks. This approach can describe a broader class of potential networks than existing models, including those where nodes in different “modules” connect to one another via various attachment mechanisms, inducing flexible and varied community structures. While unweighted edges provide less resolution than continuous weights, restricting to the binary case permits the use of likelihood-based estimation techniques, which can improve estimation of nodal features. The extra flexibility may contribute a different understanding of network generating structures, particularly for networks with heterogeneous densities in different regions.

## Contributed Session: Design and Computer Experiments

*Wednesday, June 19th (1:30pm-3:00pm)*

*Room: Main Hall*

*Chair: Luke Hagar, University of Waterloo*

### Computer Model Calibration for Large-Scale Spatially Distributed Counts

*Steven Barnett, Virginia Tech*

The Interstellar Boundary Explorer satellite (IBEX) was launched by NASA in 2008 and collects data that can help reveal the structure of the heliosphere, the boundary between our solar system and interstellar space. In particular, IBEX detects energetic neutral atoms (ENAs) coming from the heliosphere. Space science theorists have developed a variety of competing models for ENA generation and propagation, and they want to use IBEX data to evaluate the relative strength of evidence for different models. This problem fits naturally within the framework of statistical computer model calibration. However, this problem also comes with some unique challenges, such as non-continuous data, limited but data-intense computer model runs and field measurements, nonstationarity of the response surface, and spatially correlated observations. We present a framework for model calibration that builds upon the canonical Kennedy and O'Hagan approach and incorporates scaled Vecchia-approximated Gaussian processes.

### Causal inference from randomized experiments with ordinal outcomes having a spatio-temporal structure

*Rituparna Dey, Rutgers University, New Brunswick*

When the outcome variable in a two-armed randomized experiment is ordinal, the standard causal parameters like average treatment effect (ATE) are no longer meaningful. Researchers have proposed causal parameters based on the joint probability distribution of the potential outcomes, that are not identifiable from the observed data because only one of the two potential outcomes is observed for each experimental unit. To draw meaningful causal inference, one can estimate sharp bounds of these parameters based on observations from the marginal distributions. Motivated by an industrial experiment, we extend the aforementioned procedure to a situation where the observations have a spatial and temporal structure, the primary goal being to examine the treatment effect over time. We explore methods to sharpen point and interval estimates of the bounds of the causal parameters by utilizing the spatio-temporal association. We also propose bound-based test statistics for the Fisher randomization tests to test the sharp null hypothesis of no treatment effect. The effectiveness of the proposed methodology is demonstrated through simulations and applied to a real-life experiment.

### Quick Non-Uniform Space-Filling Designs

*Xiankui Yang, University of South Florida*

Traditional Space-filling designs have been popular to achieve uniform spread or coverage throughout the design space for broad applications. Non-uniform space-filling (NUSF) designs were recently developed to allow varied densities of design points for emphasizing different regions of the input space. However, due to the use of the point exchange algorithm, the construction of the NUSF designs entails substantial computational costs, particularly in high-dimensional scenarios. To improve computing efficiency, we propose an approach rooted in the core theory of NUSF designs termed Quick Non-Uniform Space-Filling (QNUSF) designs. By leveraging the Hierarchical Clustering Method, QNUSF designs expedite computational processes while retaining the fundamental principles of NUSF designs. Moreover, QNUSF designs can be easily adapted for constructing NUSF designs to achieve the weighted Maximin or Minimax distance characteristics, and hence improve the versatility and applicability for achieving different experimental goals. The computational efficiency and performance of QNUSF designs will be illustrated via several examples with varied shapes and dimensions of the design space for different applications.

## Optimal allocation of sample size for randomization-based inference from $2^K$ factorial designs

*Arun Ravichandran, Rutgers University, New Brunswick*

Optimizing the allocation of units into treatment groups can help researchers improve the precision of causal estimators and decrease costs when running factorial experiments. However, existing optimal allocation results typically assume a super-population model and that the outcome data comes from a known family of distributions. Instead, we focus on randomization-based causal inference for the finite-population setting, which does not require model specifications for the data or sampling assumptions. We propose exact theoretical solutions for optimal allocation in $2^K$ factorial experiments under complete randomization with A-, D- and E-optimality criteria. We then extend this work to factorial designs with block randomization. We also derive results for optimal allocations when using cost-based constraints. To connect our theory to practice, we provide convenient integer-constrained programming solutions using a greedy optimization approach to find integer optimal allocation solutions for both complete and block randomization. The proposed methods are demonstrated using two real-life factorial experiments conducted by social scientists.

## Variance reduction with model-based counterfactual estimation

*Kyu Min Shim, University of Waterloo*

Reducing the variance of the average treatment effect estimator is a critical problem in the context of online controlled experiments. Recent developments in variance reduction utilize pre-experiment data to achieve significant variance reduction under the assumption that pre-experiment and in-experiment data are highly correlated. However, in settings such as e-commerce and social media where trends in data may change rapidly, the validity of such an assumption may be questionable. This work addresses this challenge with a two-stage modeling framework that exploits relationships between covariates and the outcome in both pre-experiment and in-experiment data. Inference is made by estimating the counterfactual outcome of each unit and performing a pairwise comparison. This method of inference is proven to be asymptotically unbiased, with an asymptotic variance that scales with the model’s predictive accuracy though is never larger than that of the naive difference in means estimator.

# Contributed Session: Quality, Reliability, and Industrial Applications

*Wednesday, June 19th (1:30pm-3:00pm)*

*Room: Columbia A*

*Chair: Yueyao Wang, Zhejiang Gongshang University*

## Multivariate Six Sigma for Industry 4.0

*Alberto Ferrer, Universitat Politècnica de València (SPAIN)*

Six Sigma has proven to be a successful methodology for process improvement in the last decades. However, the traditional Six Sigma statistical toolkit, mainly composed of classical statistical techniques (e.g., scatter plots, correlation coefficients, hypothesis testing, and linear regression models from experimental designs), must be revisited to face the new challenges derived from Industry 4.0. In this current context, not only has the amount of registered data increased dramatically, but also there has been a significant change in their nature. These data are mostly collected from daily production (i.e., happenstance data) and often exhibit high correlation, rank deficiency, low signal-to-noise ratio, and missing values. The upgrade of the Six Sigma toolkit with latent variables-based multivariate statistical techniques such as Principal Component Analysis (PCA) and Partial Least Squares (PLS) is essential for addressing the complex data characteristics in Industry 4.0 , giving rise to the so-called Multivariate Six Sigma (Ferrer 2021). In this work, we present a Multivariate Six Sigma case study in the automotive sector arising from the lack of process capability related to the vibration tolerances for a specific part of the car's brake system. We illustrate the integration of latent variables-based multivariate statistical techniques into the five-step DMAIC cycle (i.e., Define, Measure, Analyze, Improve, and Control). These multivariate techniques played a crucial role in extracting valuable information that guided an accessible Design of Experiments (DoE), ultimately leading to the discovery of the root cause. Thus, Multivariate Six Sigma proves to be an effective methodology for process improvement in Industry 4.0 environments.

## Advanced Modeling Techniques for Forecasting and Predictive Stability

*Fangyi Luo, Procter & Gamble*

The objective of forecasting and predictive stability is to predict shelf life and estimate long-term stability failure probability during the early stages of product development. This presentation provides an overview of the advanced modeling techniques used by Procter & Gamble for forecasting and predictive stability. The methods include degradation and reliability modeling, chemometric Multivariate Curve Resolution, Bayesian Analysis of Differential Equations, AI Deep Learning, Bayesian Network modeling, functional data analysis, and hybrid modeling. Application to various consumer products will be shown.

## A New Control Chart to Monitor the Ratio of Proportions

*Su-Fen Yang, National Chengchi University, Taiwan*

This article outlines the construction of a control chart for monitoring the ratio of two binomial proportions within dependent populations. Assuming an in-control process, we presume that the two proportions and the covariance of the populations remain known and constant. The in-control ratio of the two binomial proportions is equivalently transformed into a linear combination of the two proportions. We furnish an unbiased estimator for this linear combination, along with the derivation of its distribution. Subsequently, the ratio of proportions

control chart is developed accordingly. Additionally, we compare the proposed ratio chart with another ratio control chart based on the asymptotic normal distribution of the logarithmic ratio of proportions.
The properties and detection performance of the two proposed ratio control charts are scrutinized and compared through numerical analyses. Demonstrating applicability across varying sample sizes, we establish that the proposed ratio control chart offers accurate process monitoring information and swift out-of-control detection capabilities. To illustrate the practical application of the proposed ratio control chart, we present a real-world example involving the monitoring of SAS-COV-2 infection in COVID-19.
Co-author: Chieh-Hung Lu

### Sequential Bayesian Strategy for Mean Response Surface Modeling of Expensive Stochastic Simulation with Heterogeneous Noise

*Yuying Huang, University of Waterloo*

Most of the recently developed methods of metamodeling for stochastic simulation functions with heterogeneous noise involve measuring at one design point repeatedly to estimate the sample variance in order to obtain sampling noise. This practice may not be the best option when it comes to expensive simulation, where measuring one design point has a nonnegligible cost, resulting in inefficient design when the budget is limited. To address this problem, we propose a novel empirical integrated mean squared error-based sequential design strategy to approximate the mean response surface under a Bayesian framework paired with Gaussian Process modeling. Through the proposed sequential Bayesian strategy, one can develop a globally accurate surrogate model for noisy computer simulations under a small budget. Through different synthetic examples, we show that the proposed strategy has the potential to achieve high predictive accuracy compared to existing state-of-the-art methods. We also demonstrate the performance of our strategy on a real-data simulation of finding the reliable region in podium building seismic design.

---

## Contributed Session: Advancements in Statistical Methods

*Wednesday, June 19th (1:30pm-3:00pm)*

*Room: Westmount*

*Chair: Ming-Chung Chang, Institute of Statistical Science, Academia Sinica*

### Optimal Bayesian Designs for Network A/B Testing

*Trang Bui, University of Waterloo*

We consider the problem of designing an experiment in which experimental units are connected on a network. To find optimal designs for such experiments, the experimental outcomes are assumed to follow a network-correlated model in which units potentially influence one another. Due to network interference, these outcome models are often complex, making the design criteria involve unknown parameters. We mitigate this problem by defining a Bayesian design criterion that incorporates prior distributions of the unknown parameters. Since the defined design criterion usually does not have a closed-form formula, the use of traditional optimal design algorithms is prohibited. We instead adapt meta-heuristic algorithms and Bayesian optimization algorithms to solve the optimal design problem. The performance of these design construction methods is evaluated under different realistic social network data sets and network-correlated outcome models and compared to graphcluster randomization, a popular design-based randomization scheme for network experiments. We deduce the characteristics of good designs for each outcome model and make recommendations for practitioners.

## Modeling Spatially Correlated Survival Data Under Different Distance Functions

*Jared Clark, Virginia Tech*

One common approach to statistical analysis of spatially correlated data relies on defining a correlation structure based solely on the physical distance between the locations of observed values. However, some data have a complex spatial structure that cannot be adequately described with the physical distance alone. In this work, the spatial survival data of focus contains information on GPUs that are linked through a series of logical connections, where it is expected that the failure times of GPUs with few connections between them will be highly correlated. The proposed lifetime regression model includes random effects capturing the dependency due to physical location as well as random effects explaining the dependency due to the number of logical connections between GPUs. The analysis of this GPU dataset serves as an example of models with multiple spatial random effects and the ideas presented can be extended to other applications with complex spatial structures. A Bayesian modeling scheme is recommended for this class of analyses. The examples in this work use the software package, Stan, to produce Markov chain Monte Carlo draws for parameter estimation. This modeling effort is validated through simulation which demonstrates the accuracy of statistical inference. We also apply the developed framework to the large-scale Titan GPU failure time data.

## ESPs: a new cost efficient sampler for expensive posterior distributions

*Benedetta Bruni, Duke University*

Bayesian inverse problem to model complex physical systems require the evaluation of forward simulation models, which can be prohibitively expensive in terms of CPU hours. Therefore it is important to design "cost-efficient" samplers, to achieve a satisfactory representation of the desired posterior under a fixed computational budget. Most of current sampling algorithms (e.g., Hamiltonian Monte Carlo methods) are "sample-efficient", meaning they provide a good representation of the posterior given limited samples, but hare highly cost inefficient, as they require at least one evaluation of the forward model per sample. We present a new sampler, cost-Efficient Stein Points (ESPs). ESPs is an extension of the recent Stein points from Chen et al. (2018, ICML), which achieves sample-efficiency by sequential minimization of the kernel Stein discrepancy with respect to the posterior of interest. The key novelty of ESPs is the use of carefully-constructed Gaussian process surrogate models of the kernel Stein discrepancy, for cost-efficient sequential minimization via Bayesian optimization based on Expected Improvement. We demonstrate the cost-efficiency of ESPs in comparison to state-of-the-art posterior sampling algorithms, via a suite of numerical experiments and a calibration application.

## Revisit Partial Likelihood and Tie Corrections for the Cox Model Using Poisson Binomial Distributions

*Youngjin Cho, Virginia Tech*

In a Cox model, the partial likelihood (PL), as the product of a series of conditional probabilities, is used to estimate the regression coefficients. In practice, those conditional probabilities are approximated by risk score ratios based on a continuous time model, and thus result in parameter estimates from only an approximate PL.

Through a revisit to the original PL idea, we propose an accurate PL computing method for the Cox model, which calculates the exact conditional probability using the Poisson binomial distribution (PBD). We develop new estimating and inference procedures and establish theoretical results for the new procedure. Although ties are common in real studies, current theory for the Cox model does not allow ties. In contrast, our new approach includes the theory for grouped data, which allows ties. Our theory for the new method is also valid for continuous data without ties, thus, providing a unified framework for computing PL for data with or without ties. Our numerical results show that the proposed method is superior to current methods in reducing bias, especially when there are ties or when the variability in risk scores is large. We illustrate the new method using several survival datasets with different characteristics.

# Abstracts (Thursday June 20)

## Plenary Session: Dave Campbell

*Thursday, June 20th (8:30am-9:30am)*

*Room: Main Hall*

### Overcoming Challenges from Multimodal Likelihoods By Incorporating Domain Expertise Or Changing Careers

*Dave Campbell, Carleton University*

Multimodal likelihoods arise when parameters in some parts of the likelihood space at least partially fit the data, but that there exists a direction in which moving parameters degrades the likelihood before eventually rising again. In careers this may lead to a non-linear career path, where career changes can be difficult at first, with the right execution this can lead to new research directions and improved communication across fields. The first part of this talk concerns lessons from spending two years leading a Data Science team at a crown corporation. In differential equation models, multimodal likelihoods may arise when there are several ways in which the model dynamics can fit parts of the data. In mixture models, multimodality might be caused by the models ability to switch model labels, a problem which also impedes inference in text data. The technical portion of this talk focuses on methods for overcoming multi-modality by incorporating model structure into the algorithm.

---

## Invited Session: Network Science 2

*Thursday, June 20th (9:30am-10:30am)*

*Room: Main Hall*

*Organizer: Srijan Sengupta, North Carolina State University*

*Chair: James Wilson, University of San Francisco*

### The effect of shuffled vertex labels on multiple network inference

*Vince Lyzinski, University of Maryland, College Park*

Many multiple network inference methodologies operate under the implicit assumption that the vertex correspondences across networks are a priori known, and this is commonly not the case in practical real data applications. Herein, we explore the effect of misaligned/label-shuffled vertices across networks in the context of multiple inference tasks (testing, joint clustering, classification, etc.), as well as the possibility of ameliorating the label errors via network alignment/graph matching methods. In the process, we reinforce the utility of prudent graph matching as a data preprocessing tool in statistical network inference.

### Mesoscale two-sample testing for networks

*Peter MacDonald, McGill University*

Networks arise naturally in many scientific fields as a representation of pairwise connections. Statistical network analysis has most often considered a single large network, but it is common in a number of applications, for example, neuroimaging, to observe multiple networks on a shared node set. When these networks are grouped by case-control status or another categorical covariate, the classical statistical question of two-sample comparison arises. In this work, we address the problem of testing for statistically significant differences in a prespecified subset of the connections. This general framework allows an analyst to focus on a single node, a specific region of interest, or compare whole networks. In this "mesoscale" setting, we develop statistically sound projection-based tests for two-sample comparison in both weighted and binary edge networks. Our approach can leverage all available network information, and learn informative projections which improve testing power when low-dimensional network structure is present.

---

## Invited Session: Measurement System Analysis

*Thursday, June 20th (9:30am-10:30am)*

*Room: Columbia A*

*Organizer: Nathaniel Stevens, University of Waterloo*

*Chair: Mahsa Panahi, University of Waterloo*

### Measurements system assessment of functional datasets

*Banafsheh Lashkari, University of Waterloo*

A measurement system analysis involves understanding and quantifying the variability in measurement data attributed to the measurement system. A primary goal of such analyses is to assess the measurement system's impact on the overall variability of the measured data, determining its suitability for the intended purpose. While established methods exist for evaluating measurement systems for single variables, their applicability is limited when dealing with complex data types, such as functional data. This study enhances the understanding of measurement system assessment studies, focusing specifically on functional data. We then apply the developed framework to analyze real-world functional datasets.

### Assessing the agreement between multi-operator measurement systems using the probability of agreement

*Adel Nadi, University of Waterloo*

Comparing the performance of different measurement systems and quantifying their level of agreement is an important challenge in medical and industrial contexts. The probability of agreement (PoA) was recently introduced as an intuitive metric that assesses the agreement between an established measurement system and a new one. The PoA is the probability that the difference between single measurements on the same subject by the two systems falls within the range that is deemed to be practically acceptable. Defining the PoA in this way makes its interpretation easy, even for non-statisticians, no matter how complicated the underlying model and estimation procedure might be. The PoA may be visualized with the probability of agreement plot which helpfully summarizes the agreement between systems. The current PoA methodology is based on a one-factor random effects model that represents measurement variation from all sources with a single random effect. However, in quality improvement efforts, a measurement system is often used by multiple operators to collect data. Accounting for their effects is an important part of understanding the measurement variation, and evaluating the operators' effects separately is essential for assessing their contribution to disagreement between systems should it arise. In this presentation, we develop the required methodology for applying the

PoA when the influence of operators is separately accounted for in the measurement error model (either as fixed or random effects). The developed methodology accounts for unbalanced replicate measurements across operators and systems. The methodology is illustrated by comparing the agreement between an old and a new measurement system used for quality inspections.

---

## Invited Session: Natrella Scholarship Session

*Thursday, June 20$^{th}$ (9:30am-10:30am)*

*Room: Westmount*

*Organizer: William Guthrie, NIST*

*Chair: Peng Liu, SAS Institute, Inc.*

### Scalable Design with Posterior-Based Operating Characteristics

*Luke Hagar, University of Waterloo*

To design trustworthy Bayesian studies, criteria for posterior-based operating characteristics – such as power and the type I error rate – are often defined in clinical, industrial, and corporate settings. These posterior-based operating characteristics are typically assessed by exploring sampling distributions of posterior probabilities via simulation. We propose a scalable method to determine optimal sample sizes and decision criteria that maps posterior probabilities to low-dimensional conduits for the data. Our method leverages this mapping and large-sample theory to explore segments of sampling distributions of posterior probabilities in a targeted manner. This targeted exploration approach prompts consistent sample size recommendations with fewer simulation repetitions than standard methods. We repurpose the posterior probabilities computed in that approach to efficiently investigate various sample sizes and decision criteria using contour plots.

### Distributional-assumption-free Robust Learning and Its Applications in Quality Improvements of Complex Engineering Systems

*Shancong Mou, Georgia Institute of Technology*

Image-based anomaly detection is of vital importance for quality improvment across various industrial, environmental, and medical domains. The continuous evolution of sensor technologies has facilitated the collection and storage of high-resolution, multichannel, and multi-view measurement signals, providing detailed product insights. However, annotating defects remains challenging due to the costly labeling process and the infrequent occurrence of defects in modern manufacturing processes, leading to a scarcity of labeled data in a label-rare environment. Moreover, contemporary high-value and safety-critical applications necessitate fine-grained anomaly detection capabilities. Yet, achieving such precision in anomaly detection, in the absence of adequate anomaly supervision signals, presents a big challenge for modern algorithms. This talk focuses on the developed distributional-assumpiton-free robust learning methods for label-efficient monitoring of high-dimensional data. Conventional robust learning methods adhering to low-rank or smooth assumptions, often fall short when dealing with complex signals, like defect detection in products with intricate surface patterns. I will introduce the developed Robust GAN-inversion (RGI) method that generalizes the robust learning method for unsupervised anomaly detection in complex signals in a distributional-assumption-free manner, and discuss its applications in data rich engineering application in advanced manufacturing systems, inclduing anomaly detection and signal resotration. A summary of the current challenges and future research will also be discussed.

## Invited Session: SPM in Industry 4.0

*Thursday, June 20th (11:00am-12:30pm)*

*Room: Main Hall*

*Organizer: Anne Driscoll, Virginia Tech*

*Chair: Panagiotis Tsiamyrtzis, Politecnico di Milano & Athens University of Economics and Business*

### Distributed tensor analysis for structured high-dimensional modeling

*Mostafa Reisi Gahrooei, University of Florida*

Data privacy and security concerns have been exacerbated in recent years and drove the demand to store and analyze data at the edge of networks rather than share it with a centralized server. Federated learning frameworks have been introduced as a solution to these concerns. These frameworks allow local clients to learn local models and collaborate with others to develop a more generalizable aggregated model while handling data privacy issues. At the same time, complex systems are generating more and more structured high-dimensional data (e.g., images and profiles) for which tensor analysis showed promising results by capturing complex correlation structures of data. In this paper, we propose a federated tensor regression framework where multiple local tensor models are learned at the edge, and the low-dimensional embeddings of the parameters are shared with and updated by an aggregator. Experiments on synthetic data sets and case studies show the superiority of our approach over several benchmarks.

### Industrial Statistics and Large Language Models

*Fadel Megahed, Miami University*

Generative AI applications, such as ChatGPT and Claude.ai, have garnered global attention for their ease of use, broad utility, and impressive capabilities. These conversational AI tools are accessible through a simple web browser, significantly lowering the barrier to entry. As generative AI continues to evolve, its economic impact is also growing, with McKinsey & Company estimating its potential annual economic benefit at approximately $7 trillion. Although the industrial statistics and quality monitoring communities may not be directly involved in designing these large language models (LLMs), they possess unique expertise that can significantly enhance these technologies' practical application and effectiveness. This talk will highlight several innovative use cases developed from our research at the intersection of industrial statistics and LLMs:

1. ChatSQC, our chatbot that integrates OpenAI's LLMs with a specialized Statistical Quality Control (SQC) knowledge base.
2. ChatISA, our in-house chatbot designed to support Information Systems and Analytics students at Miami University with coding, project management, exam preparation, and interview readiness, reflecting our team's pioneering work.
3. A reevaluation of how LLM outputs for text classification tasks should be assessed, employing measurement system analysis and psychometrics insights.

We hope to highlight that our community can play a valuable role in the adoption and assessment of LLMs in different organizations. Slides available here: https://fmegahed.github.io/talks/jrc2024/stats_llm.html#1

## Transparent Sequential Learning for Statistical Process Monitoring

*Xiulin Xie, Florida State University*

Statistical process control (SPC) charts provide an important analytic tool for online monitoring of sequential processes. Conventional SPC charts are designed for cases when in-control (IC) process observations are independent and identically distributed at different observation times and the IC process distribution belongs to a parametric (e.g., normal) family. In practice, however, these model assumptions are rarely valid. To address this issue, many new SPC methods have been developed using nonparametric methods, time series analysis and machine learning approaches. However, these approaches have their own limitations to handle certain SPC problems. In this talk, we present our latest research on sequential monitoring processes with complex data structure. In particular, we extend the self-starting process monitoring idea that has been employed widely in modern SPC research to a general learning framework for sequential processes monitoring. Under the new framework, process characteristics to learn are well specified in advance, and process learning is sequential in the sense that the learned process characteristics keep being updated during process monitoring. The learned process characteristics are then incorporated into a control chart for detecting process distributional shift based on all available data by the current observation time. Numerical studies and real data applications show that process monitoring based on the new learning framework is reliable and effective.

---

# Invited Session: Reliability 2

*Thursday, June 20th (11:00am-12:30pm)*

*Room: Columbia A*

*Organizer: Lu Lu, University of South Florida*

*Chair: Nathaniel Stevens, University of Waterloo*

## Joint component-system maintenance planning over entire system task profile

*Rong Pan, Arizona State University*

Engineering systems have become increasingly complex, which demands the study of system-level maintenance strategy to be integrated with a realistic prediction of component-level performance. Most previous research of system maintenance models has focused only on the relationship between component failure rate and maintenance cost. For a system with degrading components, however, it is more intuitive to model the system's health state together with component degradation processes. In this paper, we propose a preventive maintenance strategy based on the simulation of component degradation processes and lifetime predictions. In addition, the proposed strategy utilizes the maintenance opportunities between tasks assigned by the system. Specifically, the unit-time system maintenance cost is minimized with the consideration of interactions between component-level and system-level decision-makings. The approach we propose is more practical due to the simulation of the component degradation processes according to the real task profile of the system and the initial maintenance schedule can be re-adjusted in the event of an unexpected component failure. Therefore, it ensures an overall low maintenance cost and high system reliability over time.

## Robust Estimation for Step-Stress Accelerated Life Tests Under Interval Censoring and Competing Risks

*María Jaenada, Complutense University of Madrid*

Censoring is a key concern in reliability and survival analysis. In reliability tests, interval-censored data arises when the failure times are only known to fall within a specific interval, but the exact failure time cannot be directly observed. Furthermore, when a test unit fails, there are often more than one fatal cause for the failure, known as competing risks. This censoring challenge becomes particularly relevant in experiments involving highly reliable devices with large mean lifetimes, where long experiemental time is needed to conduct accurate inferences under normal operating conditions. An alternative approach involves conducting accelerated life tests (ALTs), where devices undergo increased stress factors, inducing failures and shortening their lifetimes. After suitable inference, the results can be extrapolated to normal conditions. Classical estimation methods based on the likelihood function of the lifetime are quite highly susceptible to data contamination. In this work, we propose a family of robust estimators based on the density power divergence capable of handling interval-censored data in step-stress experiments under competing risks. We examine the theoretical properties of the estimators and assess their performance through numerical simulation.

## Tests for Homogeneity of Component Lifetime Distributions Based on System Lifetime Data

*Hon Keung Tony Ng, Bentley University*

In system reliability engineering, systems are made up of different components, and these systems can be complex. For various purposes, engineers and researchers are often interested in the lifetime distribution of the system as well as the lifetime distribution of the components that make up the system. In many cases, the lifetimes of an n-component coherent system can be observed, but not the lifetimes of the components. In recent years, parametric and nonparametric inference for the lifetime distribution of components based on system lifetime lifetimes has been developed. In this talk, we discuss the problem of testing the homogeneity of component lifetime distributions based on system lifetime data with known system signatures. Existing test procedures for the homogeneity of component lifetime distributions based on complete system lifetime data are reviewed. Then, several nonparametric testing statistics based on the empirical likelihood method are proposed for testing the homogeneity of two or more component lifetime distributions. Both complete and Type-II censored system lifetime data will be considered. The performance of the proposed empirical likelihood ratio tests is compared with other parametric and nonparametric tests in the literature. Finally, some concluding remarks and possible future research directions are provided.

## Contributed Session: Computing and Machine Learning

*Thursday, June 20th (11:00am-12:30pm)*

*Room: Westmount*

*Chair: Jared Clark, Virginia Tech*

### Active Learning for a Recursive Non-Additive Emulator for Multi-Fidelity Computer Experiments

*Junoh Heo, Michigan State University*

Computer simulations have become essential for analyzing complex systems, but high-fidelity simulations often come with significant computational costs. To tackle this challenge, multi-fidelity computer experiments have emerged as a promising approach that leverages both low-fidelity and high-fidelity simulations, enhancing both the accuracy and efficiency of the analysis. In this talk, we introduce a new and flexible statistical model, the Recursive Non-Additive (RNA) emulator, that integrates the data from multi-fidelity computer experiments. Unlike conventional multi-fidelity emulation approaches that rely on an additive auto-regressive structure, the proposed RNA emulator recursively captures the relationships between multi-fidelity data using Gaussian process priors without making the additive assumption, allowing the model to accommodate more complex data patterns. Importantly, we derive the posterior predictive mean and variance of the emulator, which can be efficiently computed in a closed-form manner, leading to significant improvements in computational efficiency. Additionally, based on this emulator, we introduce three active learning strategies that optimize the balance between accuracy and simulation costs to guide the selection of the fidelity level and input locations for the next simulation run. We demonstrate the effectiveness of the proposed approach in a suite of synthetic examples and a real-world problem.

### Experimental design for surrogate modeling of expensive discrete path planning simulators via integer programming

*Yen-Chun Liu, Duke University*

Path planning -- the exploration of feasible paths for navigation -- plays an integral role in broad modern applications, including robotics, surgical planning and assembly planning. For such applications, the feasibility of a selected path is often evaluated via sophisticated virtual simulation models. This introduces two critical challenges: each simulation run can be computationally costly, and the parameter space for path planning is typically quite high-dimensional. A careful design of such simulation experiments is thus critical for timely decision-making. We propose here a novel design approach for path planning experiments, which leverages an underlying Gaussian process surrogate model for response surface exploration. The key novelty of our approach is the use of integer programming (IP) formulations, algorithms and theory for design construction of simulation runs. We present first an IP-based approach for efficient initial design construction with an arbitrary run size $n$. We then outline a flexible IP-based framework for sequential design of subsequent runs, targeting either active learning or black-box optimization of the response surface. Finally, we demonstrate the effectiveness of our design framework over the state-of-the-art, in a suite of path planning numerical experiments and an application to rover trajectory optimization.

## Graph-based Variation Propagation Network for Modeling and Prediction of Hybrid Multi-Stage Manufacturing Systems

*Dhari Alenezi, Georgia Institute of Technology*

Multistage Manufacturing Systems (MMS) are common in industries involving complex processes with multiple stages, each impacting the final product quality. Traditional product quality modeling approaches struggle with the intricate interdependencies and variable structures within these systems, further complicated by the extensive sensor-generated data. In this paper, we introduce Graph-based Variation Propagation Network (GVPNet), an innovative approach for end-to-end learning of product quality representations and their propagation through MMS stages. GVPNet utilizes a heterogeneous Graph Attention Network (hGAT) architecture that uses a graph representation specifically tailored for MMS, facilitating the effective application of graph neural networks (GNN) in this context. The network's ability to aggregate information and learn node embeddings enables it to predict quality variables several stages ahead, thus offering a proactive tool for quality control. GVPNet was applied in two real world case studies and demonstrated its superiority in predictive accuracy and robustness over benchmarks.

## High-Dimensional Multivariate Linear Regression with Weighted Nuclear Norm Regularization

*Li-Hsiang Lin, Georgia State University*

Low-rank matrix estimation plays a crucial role across a broad spectrum of applications. This research is motivated by the need to analyze important electronic absorption spectra of various polycyclic aromatic hydrocarbons (PAHs) from limited data. We introduce an innovative algorithm based on the alternating direction method of multipliers (ADMM) to estimate low-rank matrices within a multivariate linear regression model. This algorithm utilizes a weighted nuclear norm (WNN) penalty—defined as the weighted sum of the singular values of the coefficient matrix—to encourage low rank. The incorporation of weights, typically in non-decreasing order, introduces non-convexity to the objective function. Despite its prevalent use, discussions on the properties of the resulting estimator and the tuning of weight parameters are rare in existing literature. We demonstrate that the sequence of estimators generated by our algorithm converges to a stationary point of the augmented Lagrangian function. Furthermore, we present a theoretical analysis highlighting the critical role of weight tuning in achieving accurate estimation of the true coefficient matrix, alongside examining estimation errors across different scenarios. In orthogonal designs, we detail how weight values affect the precision of singular value estimates. For Gaussian designs, we identify a minimax convergence rate for the estimation error. To facilitate optimal weight parameter tuning, we introduce a generalized cross-validation approach, complemented by an iterative weight adjustment process. Our simulations and PAHs analysis validate the enhanced efficacy and robustness of our method.

## Electricity prices forecasting using functional time series models

*Ismail Shah, Department of Statistical Sciences, University of Padua, Italy*

In the liberalized electricity sector, a widely adopted method for forecasting day-ahead electricity prices is the component estimation-based methodology. This approach typically involves dividing electricity prices into deterministic and stochastic components, treating them independently without considering their direct intra-effects. This research focused mainly on the joint estimation of both components using a relatively new approach based on functional data analysis. To this end, one-day-ahead electricity price forecasts are obtained by jointly estimating both components through functional autoregressive with exogenous variables (FARX) model. The classical approach, where deterministic and stochastic components are estimated and forecasted separately, is also used for comparison. In this case, the deterministic component is forecasted using generalized additive modeling techniques, and the stochastic component is predicted through the functional autoregressive (FAR) and a univariate autoregressive (AR) model. Finally, a functional K-nearest neighbor (FKNN) method is directly applied to the original prices. The performance of these models is tested on real data where one-day-ahead forecasts are obtained for a whole year. Based on various error metrics, the results indicate that the joint estimation of both components using a functional approach leads to better results than estimating them separately.

# Poster Abstracts

## Poster Session

*Tuesday, June 18th (5:00pm – 6:00pm)*

*Room: Columbia B*

### Various Profile Monitoring Techniques for Nonlinear Mixed Models

*Abdel-Salam Abdel-Salam, Qatar University*

Quality control enhances product conformity and efficiency. This study compares monitoring methods, creating robust nonlinear profile model controls, improving quality, and efficiency in industrial applications.

### An Approximate Closed Form Bayesian Propagation Algebraic Algorithm To Accommodate Non-Normal Time Series Data

*Ali Gargoum, UAE University , Department of Statistics and Business Analytics, Al Ain, United Arab Emirates*

In this poster we consider a class of dynamic models with non-normal sampling distributions. An approximate algebraic propagation procedure to accommodate non-normal dynamic processes is developed. This procedure, which is based upon the dynamic generalized linear models, can be applied to complex high-dimensional environments that change dynamically with time. The approximation is very fast and updating is achieved in closed form. An illustrative example of a process for predicting the spread of gaseous waste after a chemical/nuclear accident when the sampling distribution is non-normal is known.

### Multivariate Control Chart with Guaranteed IC performance and Cautious Parameter Learning

*Alvaro Cordero Franco, Universidad Autonoma de Nuevo Leon*

Quality management holds utmost importance in contemporary manufacturing and service industries, with Statistical Process Monitoring (SPC) emerging as a pivotal tool for enhancing process quality. While conventional control charts have been extensively employed for univariate data to detect process shifts and out-of-control points, many real-world scenarios involve multiple interconnected variables influencing process performance. To address this problem, we propose a multivariate SPC technique aimed at effectively monitoring and improving process quality. In this study, we present the development of a multivariate CUSUM control chart ensuring guaranteed in-control (IC) performance and cautious parameter learning. Our developed control chart demonstrates commendable proficiency in detecting moderate to substantial shifts in process location. We achieve this by incorporating guaranteed IC performance, thereby minimizing false alarms during parameter estimation. Additionally, we adopt cautious parameter learning strategies to effectively handle unknown parameter variations. Our approach, validated through simulation and implementation in R software, proves highly effective in maintaining process control and detecting significant changes in process location.

## Selecting A Parent Distribution From A Family Of Distributions To Use In Medical Device Design Verification Testing

*Danni Ma, Hoffman-La Roche Limited (joint collaboration between Global Pharma Technical Operations Quality & Product Development Data Sciences)*

Design verification plays a crucial role in ensuring the safety and efficacy of medical devices. It involves testing various physical phenomena, including forces, to assess the conformance of these devices to specifications. Looking at the design verification testing process, identifying a parent distribution is the first step in following a parametric approach to calculate tolerance intervals. However, there has been a lack of consistency in selecting the appropriate parametric model for analyzing the test data. A common industry practice has been to pre-define a normal distribution and confirm the assumption with a goodness-of-fit test. If there is evidence against the data following a normal distribution, goodness-of-fit tests are used to test and compare multiple parametric models. This approach, however, has led to an inconsistent selection of distributions within datasets for the same product and test method at different time points during shelf-life testing, which may lead to incorrect estimates and decisions (e.g. an unsatisfactory design may pass design verification, or a satisfactory design may fail). Therefore, we propose a simple algorithm for engineers to select and justify the use of a specific predefined probability distribution for describing particular physical phenomena tested in design verification. Specifically, this algorithm involves: identifying suitable distributions from the literature, testing their goodness-of-fit across historical datasets using Anderson-Darling tests, and characterizing the results across the historical datasets to select a distribution. The goodness-of-fit of the proposed distribution can be later assessed at the 0.05 level. The goal is to simplify and standardize the distribution selection process and gain powerful insights of device performance. Through this process we minimize the risk of incorrect conclusions. The practical application of the proposed algorithm will be demonstrated using a case study from a needle-based injection system.

## Semi-Supervised PARAFAC2 Decomposition for Medical Decision Making

*Elif Konyar, University of Florida*

High-dimensional data is common in healthcare applications, particularly from the collection of data during clinical visits of patients. This HD data, commonly referred to as electronic health records (EHR), consists of medical conditions and clinical progress of a large cohort of patients over time. Tensor factorization methods are commonly used to extract meaningful patterns from the EHR data. However, challenges arise in analyzing EHR data because of data irregularity due to varying hospital visits and frequencies, and the presence of missing values, such as incomplete test results or erroneous entries. Existing PARAFAC2 frameworks are unsupervised methods and cannot exploit partially labeled data. In this work, we propose a semi-supervised PARAFAC2 decomposition model to extract meaningful patterns from irregular EHR data by also incorporating label information from a subset of instances. Furthermore, this method can be extended to the cases with partially observed EHR with missing values which is common in real applications. This novel computational phenotyping method has potential to facilitate medical decision making in many healthcare applications.

## Multistate Multilevel Sequential Event Occurrence In Case Of Effective Scholar Promotion

*Hailemichael Menberu Gatahun, University of Science and Technology of China*

Scholars' promotion in higher education is an upward move from one position to a higher position within their organizational hierarchy over time. In an ordered promotion, when the promotion occurs and what happens after the first promotion is an interesting study. This working paper estimates the expected time to be promoted as a full professor and its transition probability from assistant to associate and associate to full professor. It determines motivation factors that have been associated with academic promotion at the two-transition state with a random effect. To achieve these objectives, an efficient novel methodology for multistate sequential promotion transition with a multilevel random effect was developed by extending the semi-parametric survival model. The data is further illustrated by a web-based, published source survey to reach all academic scholars' promotion from the top ten business schools in the United States. The expected empirical finding provides 7.46 years to be promoted from associate to full professor and 4.6 years from assistant to associate professor. The variation between institutes/ universities and between scholars is minimal, and the gender variation highly influences the promotion transition. Males are more likely to be promoted from associate to full professor, even if there is no variation in the transition from assistant to associate. Additionally, some key motivation factors can be used for future scholar promotion studies to move their academic position, and the innovative model is recommended for the researchers to study transition states with a random effect.

## Navigating Missing Data Challenges In Machine Learning With An Application In Tuberculosis Mortality Prediction

*Md Belal Hossain, University of British Columbia*

Large datasets such as linked health administrative data can be used to build models for predicting rare disease outcomes such as long-term mortality in people diagnosed with tuberculosis. However, the challenge is compounded by missing values in important predictors. The complete case data analysis has been widely criticized in the literature, and multiple imputation has been recommended to deal with missing predictor values. Machine learning methods for time-to-event outcomes, such as Cox-LASSO, face limitations as these methods are incompatible for pooling model outputs from multiple imputed data using Rubin's rule, highlighting the need for innovative approaches in handling missing data. To combine outputs from multiple imputed datasets for handling missing data and Cox-LASSO with cross-validation or bootstrapping for internal validation, we explored the performance of three distinct statistical methods: prediction average, performance average, and stacked approaches, paired with two hyperparameter selection techniques: minimum-lambda and 1SE-lambda. We applied these methods to a retrospective cohort of 2,923 participants diagnosed with tuberculosis in British Columbia, Canada. To empirically compare the methods, we also conducted an extensive plasmode simulation to reflect the complex relationships between the risk factors observed in our tuberculosis cohort. In our simulations, the stacked approach provided the most robust predictions in terms of time-dependent c-statistic and calibration slope. This approach with the minimum-lambda also worked uniformly well in predicting tuberculosis mortality, producing a time-dependent c-statistic of 0.93 (standard error [SE]: 0.003), a calibration slope of 1.01 (SE: 0.01), well-calibrated calibration plot, and better net benefits. Our study navigates how to address missing data challenges in machine learning in the context of risk prediction modelling with time-to-event outcomes. Our study also extends its practical utility by offering a web application for tuberculosis mortality prediction and reproducible software codes for future researchers to easily implement these methods in their research.

## Comparison of the analytical and asymptotical variance calculation methods for diagnostic test meta-analysis: a simulation study

*Olana Dabi, University of Waterloo*

Screening procedures in biological science aim to accurately classify individuals with diseases as diseased and individuals without diseases as non-diseased with probability as high as possible. Meta-analysis is important in diagnostic test accuracy (DTA) studies because it allows for the synthesis of findings from various studies. In DTA meta-analysis, the asymptotic (delta method) approach is commonly used to estimate within-study variances for transformed sensitivities and specificities. The asymptotic approach relies on approximations, assumes the transformed random variables are normal, and works better with larger sample sizes. Contrarily, the analytical method employs observed data to directly calculate variances based on transformed random quantities. Thus, choosing the more accurate method is critical and ensures the meta-analysis results are reliable, reflect the actual DTA, and contribute to evidence-based decision-making in clinical practice. Therefore, this study aims to compare the performance of the asymptotical and analytical variance calculation approaches when modeling logit-transformed sensitivities and specificities using the bivariate linear mixed-effects model (LMM) in the context of DTA meta-analyses. We will conduct an extensive simulation study by varying data characteristics and model parameters mimicking real-life scenarios. The methods will be compared in terms of bias, root mean squared error, coverage probability, and width of the 95% confidence interval for the true pooled sensitivity and specificity. We will also demonstrate the methods using real-life meta-analytic data sets. The findings of this study will have implications for researchers and practitioners involved in the assessment and interpretation of DTA studies in the biological sciences, paving the way for better-informed decision-making and patient care.

## Evaluating and validating machine learning methods for diagnosing invasive group A streptococcal (iGAS) disease

*Olena Kislenko, University of Waterloo*

In recent years, the invasive Group A Streptococcus (iGAS) infection has posed a significant and unexpected threat, leading to a notable increase in mortality rates in Canada compared to the 2019 data. The iGAS has become a growing public health concern globally, necessitating innovative approaches for timely and accurate diagnosis, since the usual diagnostic tests based on the throat culture can take up to two days, which may lead to the worsening of the patient’s symptoms and sometimes death as seen in recent trends. This study aims to develop diagnostic tools capable of rapidly and accurately identifying iGAS infection early in the course of the disease. We will approach this problem utilizing publicly available individual participant data meta-analysis (IPDMA) coupled with machine learning techniques to develop a robust diagnostic tool. The machine learning algorithms that will be applied to analyze individual participant data will include logistic regression, XGBoost, random forests, and support vector machines (SVM). Utilizing machine learning approaches may aid in capturing characteristics that are more complex than those initially available in the data. We will also aim to uncover subtle patterns by synthesizing data from multiple sources. This work will delve into the practical implementation of our machine learning models for early iGAS detection while exploring and comparing their characteristics. Our work will demonstrate the potential of machine learning approaches to enhance diagnostic accuracy for iGAS infection and make the diagnosis process considerably easier to perform than the currently available common approaches.

## On Modified Jordan Neural Network structure of Volatility models and Application

*Oyebimpe Adeniji, University of Ibadan*

A recursive neural network structure of GARCH models is developed and incorporated into new variant innovations. An empirical study was carried out on the volatility impact of the Russian-Ukraine Invasion War on crude oil prices. The Impact of the Russian Invasion war on crude prices was observed and reported in this paper. Daily West Texas Intermediate (WTI) crude oil data was used in this study. The visual exploratory analysis indicates non-stationary and high volatility persistence in the series, the descriptive statistics also revealed evidence of stylized facts of volatility. features. Traditional GARCH models and Jordan Neural Network of GARCH models with modified error terms were used to capture the stylized facts. The forecast performance of the proposed models over the existing ones was investigated using the Log-likelihood function, Root Mean Square Error (RMSE), Adjusted Mean Absolute Percentage Error (AMAPE) and Akaike Information Criterion (AIC).

## Statistical Evaluation of Seismic Foundation Models

*Samuel Myren, Virginia Tech, Los Alamos National Laboratory*

Foundation models, which are massive models pretrained on big data through self-supervision that are quickly fine-tuned to accomplish different downstream tasks, have been honed though generative natural language models and are actively being developed for other scientific disciplines such as seismology. However, statistical evaluation of scientific FMs that considers multiple tasks, model interpretability, learning speed, and unbiased train/validation/test splits, presents a significant challenge. This research addresses these challenges for seismic FMs through analysis of deep learning phase picking models. We evaluate and compare four models, one trained using the Stanford Earthquake Dataset (STEAD) (Model 0a) and one trained using a subset of the STEAD data (Model 0b), one trained using the Italian seismic dataset for machine learning (INSTANCE) (Model 1), and one initialized with weights from Model 0b and fine-tuned using INSTANCE data (Model 2). We carefully curate the train/validation/test splits to limit geographical mixing to reduce leakage of signal characteristics between the train and test phases. We also evaluate the effect of the amount of training data used on performance. Beyond this, by grouping sets of the data based on geographic locations and training the model on randomly selected subsets as well as using model snapshots (i.e. ensembling), we obtain uncertainty in predictions and in summary metrics arising from aleatoric and epistemic uncertainty. We utilize multiple seismic test sets to evaluate model performance on in and out of distribution data. To explore interpretability, we use metrics like Brier score and calibration scores. Performing such a wholistic evaluation provides a strong statistical evaluation approach to be applied widely to scientific foundation models.

## Exploring optimization strategies from historical process data

*Sergio García-Carrión, Polytechnic University of Valencia*

The rise of Industry 4.0 has fostered a data-rich environment, wherein most companies gather a massive volume of historical data from daily production, frequently containing some unplanned excitations. However, these data are generally characterized by high collinearity and rank deficiency. Therefore, causal models, which are required for model-based process optimization, cannot be extracted from them by using predictive models such as linear regression (LR) and machine learning (ML) techniques, as independent variations in the input variables are not satisfied (i.e., data are not obtained from a DoE that guarantees this requirement). In this work, we propose two approaches to exploit this type of data (i.e., data from daily production, not coming from a DoE) for optimization purposes. The first one is a retrospective DoE methodology that consists in retrospectively fitting two-level experimental designs by filtering the database for observations that were close to the corresponding design points, and subsequently performing the analysis as in a typical DoE. The possibility of eventual missing treatment conditions was also addressed. Finally, a meta-analysis is conducted via raster plots, which reveal the potential active factors (i.e., factors appearing consistently in the best models) as well as aliasing patterns. The second one is by inverting Partial Least Squares (PLS) models. Since they exhibit uniqueness and causality in the reduced latent space, they are suitable for (restricted) process optimization no matter where the data come from. Both approaches are illustrated and compared through a case study from a real industrial process, highlighting their advantages and disadvantages. Nonetheless, they proved to be useful for screening in the early stages, gaining some insights about the process and providing valuable input towards an efficient subsequent designed experimentation, all with significant cost savings.

## A Bayesian Approach to Response Optimization on Data with Multi-Stratum Structure

*Xiaohua Liu, University of Manitoba*

Multi-stratum design arises naturally in industrial experiments due to the inconvenient and impractical completely randomization. Most research on response optimization in the literature is for completely randomized designs. Accounting for the model uncertainty, we apply the Bayesian model averaging method and predictive approach to investigate the optimization problem for data with multi-stratum structure. With the posterior probabilities of models as weights, we consider the weighted average of the predictive densities of the response over all potential models. The goal of the optimization is to identify the values of the factors that result in a maximum probability of a response in a given range. The method is illustrated with two examples.

## Bayesian Estimation of Negative Binomial Integer-Valued Generalized Autoregressive Conditional Heteroscedasticity (NB-INGARCH) Time Series Models with Minimal Prior Information

*Xiaoyin Wang, Towson University*

This study introduces the NB-INGARCH model, designed for count data with over-dispersion, challenging traditional models relying on conditional Poisson distributions. Parameter estimation of the NB-INGARCH model poses challenges due to a more complex likelihood function compared to the Poisson case. Our paper advocates for Bayesian estimation of the NB-INGARCH process with minimum information prior. Our Bayesian modeling approach, employing various prior models emphasizes minimal prior influence on the joint posterior, thereby establishing a robust framework for parameter inference in data-driven contexts. Numerical results are obtained through Markov Chain Monte Carlo methods. Bayesian estimates demonstrate superior reliability and robustness compared to maximum likelihood estimation (MLE). Estimating the dispersion parameter in the NB-INGARCH model within classical approaches has received limited attention. Typically, a two-stage method is proposed: first estimating the remaining parameters with a fixed dispersion parameter, then optimizing the dispersion parameter using the profile likelihood. In contrast, the Bayesian approach enables joint estimation of all parameters, simplifying the process and addressing parameter uncertainty, leading to a more comprehensive and coherent analysis. The classical approach reveals larger standard errors and negative lower limits in confidence interval of other parameters, which further emphasizing the advantages of the Bayesian approach, especially for parameters like the long-term average. Consistency across different choices of prior distributions is observed for persistence parameters, with remarkable stability in the long-term average across different prior models, suggesting minimal influence from variations in the baseline level parameter. By exploring the alliance of the NB-INGARCH models and Bayesian methodology, our research not only advances the understanding of count time series analysis but also provides practical tools for modeling and inference. This work contributes to the broader landscape of statistical modeling for complex data structures, with potential applications across various domains.

## Data Quality in Crowdsourcing and Spamming Behavior Detection

*Yang Ba, Arizona State University*

As crowdsourcing emerges as an efficient and cost-effective method for obtaining labels for machine learning datasets, it is important to assess the quality of crowd-provided data, so as to improve the analysis performance and reduce the bias of machine learning models. Given the lack of ground truth in most cases of crowdsourcing, we refer to data quality as annotators' consistency and credibility. Unlike the simple scenarios where Kappa coefficient and intraclass correlation coefficient usually can apply, online crowdsourcing requires dealing with more complex situations (e.g. multiple annotations for single tasks and categorical responses). We introduce a systematic method to evaluate data quality and detect spamming threats via variance decomposition and classify spammers into three categories based on their different behavioral patterns. The spammer index is proposed to assess the entire data quality: consistency, and two metrics to measure crowd workers' credibility by utilizing the Markov Chain and Generalized Random Effects models. Furthermore, we showcase the practicality of our techniques and their advantages by correlating them with accuracy in a face verification task via both simulation and real-world data collected from two representative crowdsourcing platforms, as well as providing empirical procedures for assessing data quality and identifying potential spamming behaviors.

## An evaluation of computational methods for aggregate data meta-analyses of diagnostic test accuracy studies

*Yixin Zhao, University of Waterloo*

A Generalized Linear Mixed Model (GLMM) is recommended to meta-analyze diagnostic test accuracy studies (DTAs) based on aggregate or individual participant data. Since a GLMM does not have a closed-form likelihood function or parameter solutions, computational methods are conventionally used to approximate the likelihoods and obtain parameter estimates. The most used computational methods are the Iteratively Reweighted Least Squares (IRLS), the Laplace approximation (LA), and the Adaptive Gauss-Hermite quadrature (AGHQ). However, it has not been clear how these computational methods compare and perform in the context of an aggregate data meta-analysis (ADMA) of DTAs. We compared and evaluated the performance of three commonly used computational methods for GLMM - IRLS, LA, and AGHQ, via a comprehensive simulation study and real-life data examples, in the context of an ADMA of DTAs. By varying several parameters in our simulations, we assessed the performance of the three methods in terms of bias, root mean squared error, confidence interval (CI) width, coverage of the 95% CI, convergence rate, and computational speed. For most scenarios, especially when the meta-analytic data were not sparse (i.e., there were negligible studies with perfect diagnosis), the three computational methods were comparable for the estimation of sensitivity and specificity. However, the LA had the largest bias and root mean squared error for pooled sensitivity and specificity when the meta-analytic data were sparse. Moreover, the AGHQ took a longer computational time to converge relative to the other two methods, although it had the best convergence rate. We recommend practitioners and researchers carefully choose an appropriate computational algorithm when fitting a GLMM to an ADMA of DTAs. We do not recommend the LA for sparse meta-analytic data sets. However, either the AGHQ or the IRLS can be used regardless of the characteristics of the meta-analytic data.

## The value of pure experimentation

*Zahra Jamshidi, University of Calgary, Haskayne School of Business*

Experimentation is a vital tool in innovation, entrepreneurship and business in general. We observe that the value of experimentation often lies in the contrast between the anticipated value of the top result among two or more trials and the expected value of a single attempt. This concept aligns with the statistical theory of the maximum of multiple draws from a probability distribution, known as the last order statistic. The implications of this alignment between experimentation and order statistics have not been fully appreciated or elaborated on in the management literature. We demonstrate that when the probability distribution of potential outcomes is known, the properties of the last-order statistic can provide us with a precise imputation of value to experimentation such that values can be calc
ulated for the advantage of conducting multiple experiments over one. This advantage comes in the form of increase in central tendency (mean or median) and change in dispersion (variance or standard deviation), as well as increased likelihood of outliers. We calculate a number of measures of the advantage of pure experimentation for two common types of probability distribution commonly observed in business outcomes: the Normal Distribution and the Pareto (Power Law) Distribution. We find interesting insights including a "golden rule" of variance reduction that applies to any normal distribution regardless of its parameters.

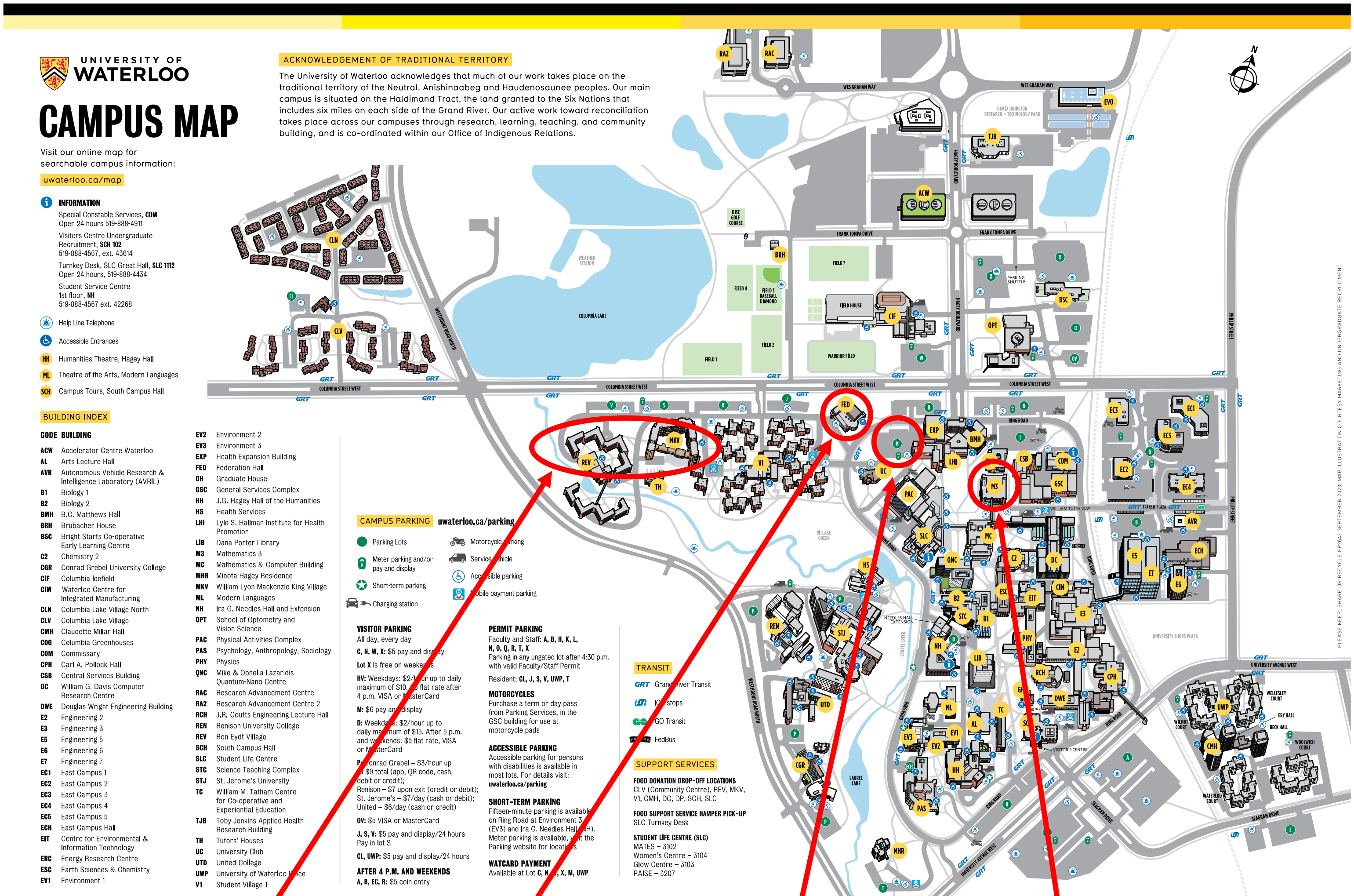
## Testing Conditional Tail Independence

*Zhaowen Wang, Fudan University*

This paper first introduces the definition of conditional tail independence, which considers the extremal dependence structure of bivariate vectors given covariates. Conditional tail quotient correlation coefficient (CTQCC) is proposed to measure the conditional tail dependence. The asymptotic distribution of CTQCC is also established. A new testing procedure based on CTQCC can be used to test the null hypothesis of conditional tail independence. We apply CTQCC to investigate conditional tail dependencies of a large scale problem of daily temperature and humidity in the continental US, given the elevation and. Our results, from the perspective of conditional tail dependence, reveal nonstationarity, spatial clusters, and conditional tail dependence from the extreme temperature and humidity across the continental US.

# Maps

## Campus Map



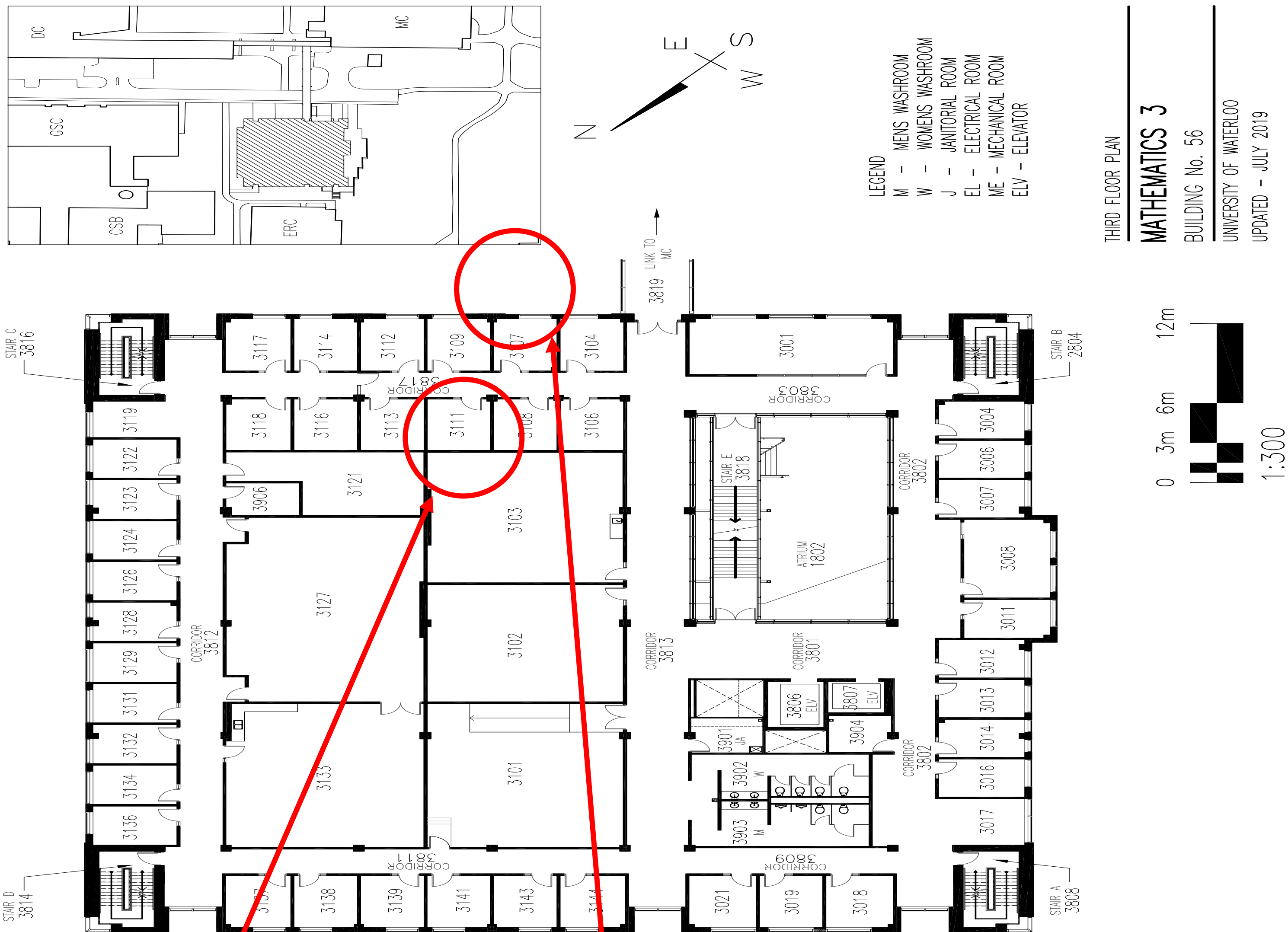

**Residences**
This is the location of the on-campus housing.

**Fed Hall**
This is the main conference location for June 18-20.

**Parking Lot M**
This is a paid parking lot ($14/day) available June 17-20.

**Mathematics 3**
This is the location of the short course on June 17.

## Short Course Map: Mathematics 3




**M3 3103**
This is the location of the short course on June 17.

**M3 3127**
This is the location of meals and breaks during the short course.

## Conference Map: Fed Hall