

# 2016 Southwestern Ontario Graduate Mathematics and Statistics Conference

June 13-14, 2016



# 2016 Southwestern Ontario Graduate Mathematics and Statistics Conference

## Schedule of Talks

Monday, June 13th 2016.

- 8:30      **Registration and Coffee**
- 9:00      **Welcome and Introduction**
- 9:10      **Andrew Skelton** University of Guelph.
- 9:25      **Peter Sinclair**, McMaster University.  
*Ultraproducts and their Applications.*
- 9:45      **Melkior Ornik**, University of Toronto.  
*On a Topological Obstruction to Map Extension.*
- 10:05     **Heather Topping**, Queen's University.  
*The Waring Problem for Polynomials.*
- 10:25     **Coffee Break**
- 10:40     **Faizan Mohsin**, University of Toronto.  
*Imputation Methods.*
- 11:00     **Sayantee Jana**, McMaster University.  
*Estimation of the Parameters of a Growth Curve Model under Multivariate Skew Normal Distribution.*
- 11:20     **Demetri Pananos**, University of Waterloo.  
*Vaccine Scars as Critical Transitions.*
- 11:40     **Picture and Lunch (Brass Taps)**
- 1:00      **Jeff Daniel**, University of Guelph.  
*Lasso Regularization of Spatial Poisson Point Processes.*
- 1:20      **Aronica Yang**, University of Guelph.  
*Estimation on Exponential Survival Probability Function on Acute Leukemia Data.*
- 1:40      **Emma Smith**, University of Guelph.  
*Quantifying Health: Measuring the Impact of Fecal Microbiota Transplantation on Quality of Life.*
- 2:00      **Richard Kohar**, Royal Military College.  
*So You Want to Write a Textbook?*
- 2:20      **Coffee Break**

2:40	<b>Robert Henderson</b> , University of Waterloo. <i>Implementing Group Work Exercises in an Online STEM Course.</i>
3:00	<b>Kelvin Shuanjian Zhang</b> , University of Toronto. <i>Optimal Strategy for a Principal Facing Risk-Averse Agent.</i>
3:20	<b>James McNeil</b> , Queen's University. <i>Testing for Structural Change: A Recursive Bootstrap Approach.</i>
3:40	<b>Zheng (Michelle) Song</b> , Queen's University. <i>A Prediction of Successful Bank Telemarketing.</i>
4:00	<b>Break, vote on questions for panel</b>
4:15	<b>Panel discussion</b>
5:15 - 7:00	<b>CMS Social, Grad Student Lounge, 5th Floor University Centre</b>

## Tuesday, June 14th 2016

8:30	<b>Breakfast (coffee, pastries)</b>
9:00	<b>Carolyn Augusta</b> , University of Guelph. <i>Introduction to Natural Language Processing.</i>
9:20	<b>Michal Lisicki</b> , University of Guelph. <i>Optimal Design and Operation of Archimedes Screw Turbines Using Bayesian Optimization.</i>
9:40	<b>Maysum Panju</b> , University of Waterloo. <i>Spectral Learning Demystified.</i>
10:00	<b>Coffee Break</b>
10:20	<b>Taylor Smith</b> , University of Waterloo. <i>Extending the Lyndon-Schutzenberger Theorem.</i>
10:40	<b>Fei Yu</b> , University of Guelph. <i>Analysis and Performance of Cognitive Agents Learning to Cross a Highway Model with Improved Learning Mechanism.</i>
11:00	<b>Abul Sheikh</b> , University of Guelph. <i>Modelling species using Lotka-Volterra equations and constructing food-web networks.</i>
11:20	<b>Lunch (The Bullring)</b>
1:00	<b>Reem Almarashi</b> , Wilfrid Laurier University. <i>Infectious Disease Modelling.</i>
1:20	<b>Zizhen Wang</b> , University of Guelph. <i>R Implementation of a Generalized Linear Model for Effect of Naphthenic Acids on Hatching Probability of Fathead Minnows.</i>

- 1:40 **Stephanie Hughes**, University of Guelph.  
*Syndromic Surveillance Systems and Norovirus Outbreak Reporting in Ontario.*
- 2:00 **Coffee Break**
- 2:20 **Ram Sigdel**, University of Guelph.  
*Properties of Simple Coupled Human Environment System-Forest/Other Land Use.*
- 2:40 **Mona Ghwila**, University of Guelph.  
*The Stability of A Size-Structured Population Model in Terms of its Equivalent Delay System.*
- 3:00 **Kevin Church**, University of Waterloo.  
*Bifurcations in impulsive differential equations.*
- 3:45 **Voting Closes for Best Talk and Best Poster**
- 4:00 **Announcement of results**



# 2016 Southwestern Ontario Graduate Mathematics and Statistics Conference (SOGMSC'16)

## Abstracts



**Peter Sinclair**, University of Waterloo.

*Ultraproducts and their Applications.*

One of the fundamental tools to logicians is the ultraproduct, a method of taking an average of a set of mathematical structures such as fields, topological spaces, or graphs. I will present an informal description of how ultraproducts are constructed, and give examples of how they can be applied outside of logic. In particular, I will discuss the basics of non-standard analysis, including a very short proof of the Bolzano-Weierstrass theorem, and discuss how ultraproducts can be used to transfer results in algebraic geometry from positive characteristic to characteristic zero.

**Melkior Ornik**, University of Toronto.

*On a Topological Obstruction to Map Extension.*

In the study of the reach control problem in control theory, a question of an existence of a nowhere zero map satisfying certain linear conditions on a polytope emerges as a strong necessary condition for solvability of the problem. This question comes in two different settings, dealing with affine and continuous maps. This talk will concentrate on the continuous case, in which case this problem can be expressed as an equivalent question in classical algebraic topology. In the talk, I will briefly present the control theoretical background for the problem, and will set up the equivalent topological question. For reasons of time and clarity, the solution to the problem will then be outlined for the two-dimensional case, which serves to introduce the intuition and machinery necessary in higher dimensions.

**Heather Topping**, Queen's University.

*The Waring Problem for Polynomials.*

The classical Waring Problem asks what the smallest integer  $N(j)$  is so that every integer can be written as the sum of  $N(j)$   $j$ th powers of integers. We may extend this question from integers to linear forms and ask what the smallest integer  $s$  is so that any  $d$ th degree homogeneous form is the sum of at most  $s$   $d$ th powers of linear forms. In this talk, we use a geometric approach to investigate this problem and present a statement of the complete solution to one of the Waring Problems for Polynomials.



**Faizan Mohsin**, University of Toronto.  
*Imputation Methods.*

We apply four imputation methods on a real world data set which is longitudinal in nature with incrementally increasing missing data over time with 60% overall data missing. We compare the following four methods of imputing the missing data: multiple imputation by predictive-mean-matching, a longitudinal data specific multiple imputation method, the single imputation method:  $K$ th Nearest Neighbor and simply replacing the missing values with the mean.

**Demetri Pananos**, University of Waterloo.  
*Vaccine Scares as Critical Transitions.*

There exists strong evidence that vaccines are extremely effective in the prevention of pediatric infectious disease, yet despite this evidence, vaccine refusal is still popular amongst some social circles. Existing mathematical models of vaccinating dynamics are parsimonious with post scare empirical data, yet lack the ability to predict when a vaccine scare will occur. We examine the problem of predicting vaccine scares through the lens of critical transitions theory. We show that a mathematical model of the coupled behavior-disease dynamics of measles exhibits critical slowing down in vaccine sentiment before a critical transition where vaccine coverage drops steeply. Using machine learning, we analyze sentiment in tweets for the United States leading up to the 2014 measles outbreak in Disneyland. We find evidence of critical slowing down in the rising autocorrelation in tweets with negative sentiment.

**Jeff Daniel**, University of Guelph.  
*Lasso Regularization of Spatial Poisson Point Processes.*

The spatial Poisson point process is a statistical model that relates the occurrence of events in space to spatially-indexed covariates, which may be numerical or categorical in nature. Recently, the method of lasso regularization, which allows model fitting and variable selection to be performed simultaneously, was adapted for use with spatial Poisson point processes; however, as currently implemented, the lasso is unable to appropriately handle categorical data. In this talk, I present a generalization of the lasso for spatial Poisson point processes that allows for selection of categorical spatial covariates. I also present an application of this method, modelling the distribution of gorilla nests in an African national park.



**Aronica (Wenjun) Yang, University of Guelph.**

*Estimation on exponential survival probability function on acute leukemia data.*

This talk attempts to estimate exponential survival probability function of acute leukemia patients based on maximum likelihood estimator, method of moments and robust statistics as 1st percentile, median and 3rd percentile. Since Kaplan-Meier survival curve hinged us by exponentiated-form probability function described with related varying parameter per patient (Gupta and Huang et al., 2014), 5 distinct members of exponential family distribution are applied to make statistical inference concerning parameters (Alizadeh, M., et al., 2015). As common in practice, comparisons between proposed probability density functions are performed. Essentially, Weibull distribution characterized by maximum likelihood estimate yields highest goodness of fit on the data globally (Gupta et al., 2004). The illustration on the methods using the acute leukemia data is shown.

**Emma Smith, University of Guelph.**

*Quantifying Health: Measuring the Impact of Fecal Microbiota Transplantation on Quality of Life.*

The RAND 36-Health Survey assesses a patients perceived status in eight dimensions of health including physical functioning, emotional well-being, and bodily pain. This instrument has become extremely popular in recent years as it allows for the calculation of Quality Adjusted Life Years which can be used in subsequent health economic analyses. However, as is the case with most survey instruments, non-response is a huge issue. This is further complicated in a health survey setting as non-response is most likely related to the patients true health status and thus data is not missing at random. This talk will cover the advantages and disadvantages of current methods of handling missing data for these instruments and will end with suggestions of possible improvements.

**Robert Henderson, University of Waterloo.**

*Implementing Group Work Exercises in an Online STEM Course.*

Facilitating group work sessions in tutorial-type course components is a task often faced by teaching assistants. In traditional courses, this is a fairly standard procedure where lecture-style interaction is replaced by the TA acting as a moderator and motivator. Further, group work and peer interaction is known to improve learning. However, implementing group work in an Online course presents some significant challenges. In this talk, I will discuss how we have implemented group work exercises in two successive offerings of an Online first year physics course at the University of Waterloo: the challenges, successes, our observations and some tips for teaching assistants.





**Richard Kohar**, Royal Military College.

*So You Want to Write a Textbook?.*

Many graduate students are thrown in front of undergraduates to teach tutorials or even a class without formal training in teaching. In this talk, I'll discuss my personal experiences about how I learned about educating students, and how I adapted Polya's Problem Solving strategies. Because I was teaching mathematics to students who were Arts majors, I learned to adapt the material to keep their interests: I found logical arguments used in history, applied set theory to social policy making, and demonstrated probability in counterintuitive situations. Doing this reinforced the mathematical concepts in their long-term memory by associating it with their own previous knowledge. In the end, my students shifted from seeing mathematics as rote calculation to problem solving, and all of this generated a lot of new material that eventually landed me a book deal.

**Kelvin Shuangjian Zhang**, University of Toronto.

*Optimal Strategy for a Principal Facing Risk-Averse Agent.*

A monopolist wishes to maximize her profit by finding an optimal price policy. After she publishes a price list of products, each agent  $x$  will choose to buy that product  $y(x)$  which maximizes his own utility, which is strictly decreasing in prices of products. Then the principal will calculate her total profit by summing up the net earnings of each product sold. Note that the distribution of products sold is based on the choices of agents, and thus fundamentally depends on the distribution of agents and also price policy. In this paper, we provide an existence result for this bilevel optimization, by using convex analysis argument. Moreover, we get some uniqueness and robustness results, for certain types of utility. Besides, I also discovered the unique explicit solutions on  $n$ -dimensional symmetric hyperbolic disks, for quasi-linear cases with any  $n \geq 2$ .

**James McNeil**, Queen's University.

*Testing for Structural Change: A Recursive Bootstrap Approach.*

Standard asymptotic tests for structural change are invalid when the regressors are non-stationary. Hansen (2000) presents an alternative fixed regressor bootstrap test which dominates the asymptotic tests but struggles in the presence of heteroskedasticity and when the model is non-stationary in the variance of the regressors. Although this method is valid asymptotically, it has poor finite sample properties because the bootstrap data generating process does not reflect how the data were actually generated. In this paper I examine the properties of structural break tests based on the more conventional residual bootstrap technique along with a wild bootstrap variant for models with heteroskedasticity. While this approach has been suggested before, notably by Jouini (2010), its performance when the model is non-stationary in the variance of the regressors is unexplored. I show that in most circumstances this approach outperforms the fixed regressor bootstrap, although there are still cases where the combination of heteroskedasticity and structural change lead to severely oversized tests.



## **Zheng (Michelle) Song**, University of Guelph.

*A Prediction of Successful Bank Telemarketing.*

Today, bank telemarketing is an important channel for product selling. There are many factors to influence their success. Although it is direct communication between salesperson and potential customers, it may not be the most efficient way to sell products. Sometimes the customers don't want to provide their private information. In this project, we focused on how to reduce the dimension of input data and build the predictive model to help banks increase the rate of bank telemarketing success. The dimension reduction methods considered here were principal components and factor analysis. In addition, the logistic regression would be applied for classification. The findings of this study illustrated that the two dimension reduction methods have different results and the logistic regression analysis is showed that most important predictors focused on customers contact information. The methodology of this study can be applied to more complicated data in the future.

## **Carolyn Augusta**, University of Guelph.

*Introduction to Natural Language Processing.*

The field of machine learning combines mathematics, statistics, engineering, and computer science to yield innovations regarding the speed and accuracy to which we process data in the modern age. Data, though, need not be numeric: how can we classify documents according to common characteristics? When there are millions of documents to classify, as in a database for submitted papers like arXiv, what algorithms can and should we use to ensure the lowest classification error in a reasonable amount of processing time? This tutorial-like talk will introduce some fundamental aspects of natural language processing, and include an example in Python.

## **Michal Lisicki**, University of Guelph.

*Optimal Design and Operation of Archimedes Screw Turbines Using Bayesian Optimization.*

The recent revival of Bayesian optimization has caused widespread use of easily accessible and versatile tools in different areas which involve the search for optimal design or decisions. This method however has not yet been explored in the field of renewable energy systems. This study introduces the main benefits of the procedure to the community through the practical task of optimizing the design and operation of the Archimedes screw turbine (AST) in terms of maximizing the total rate of return for a specific installation. The optimal design is presented as a combination of inputs to a software simulation of a true AST. The contribution of this manuscript is three-fold i) we present the full procedure for sizing an optimal energy system ii) compare various implementations and configurations of the optimization method available under several recent open-source software frameworks and iii) compare the single-objective with the multi-objective approach to optimization within the same scenario. Our experiments demonstrate superior results using Bayesian optimization compared to the standard baseline both in terms of time and number of model evaluations.



**Maysum Panju**, University of Waterloo.

*Spectral Learning Demystified* Note change: Deep Learning Demystified.

Many algorithms in statistical learning are based on inferring maximum likelihood estimates for models, based on iterative algorithms involving some form of Expectation-Maximization. While these algorithms can be very effective when they work, they are subject to the drawbacks of the clunky EM procedure, including the time-consuming iteration scheme and an affinity towards bad local optima. Spectral learning is a relatively recent paradigm shift of machine learning that shows how a moment-matching form of model estimation allows algorithms to break free from these disadvantages, and quickly obtain good models without the heavy iterations. This simple framework is both powerful and mysterious; however, with the correct theoretical approach, the underlying concepts are not too difficult to understand. This talk will present an introduction to how spectral learning works and some situations where it has been effective, in an open and accessible way.

**Taylor Smith**, University of Waterloo.

*Extending the Lyndon-Schutzenberger Theorem.*

The Lyndon-Schutzenberger theorem is a famous result in the field of combinatorics on words. It defines a set of equivalent criteria for two words  $x$  and  $y$  to commute; that is, for  $xy = yx$ . Though much is known about this theorem in its current form, some questions still remain. For instance, can it be generalized to words of higher dimension? In this talk, I will introduce the Lyndon-Schutzenberger theorem, present and prove the equivalence of two additional criteria for the theorem, and extend the theorem to the case of two-dimensional words. I will also offer some open problems and directions for future work. This is a joint work with Jeffrey Shallit.

**Fei Yu**, University of Guelph.

*Analysis and Performance of Cognitive Agents Learning to Cross a Highway Model with Improved Learning Mechanism.*

The presentation introduces a modification of simulation model of cognitive agents learning to cross safely a cellular automaton based highway. Based on the analysis of the simulation results of the initial model its short comings are identified and a modification of learning mechanism is proposed, implemented and performance of the modified simulation model investigated. A large amount of simulation data is generated by the modified simulation model for various combinations of configuration parameters values. An exhaustive comparison analysis of simulation results produced by these two models is conducted. This comparative analysis of the models performance conducted for various combinations of parameters values brings a better understanding of the performance of cognitive agents models as well as their “learning” mechanisms.



**Abul Sheikh**, University of Guelph.

*Modelling species using Lotka-Volterra equations and constructing food-web networks.*

In this project I test a novel visual multi-species ecological communication tool for the public by using actual data to construct and predict a species network using two and three species over 100 years. Here the node size represents the relative species abundance at a given time and an arrow between two nodes indicates the flow of food energy from a prey to a predator, e.g. Prey Predator. To construct the networks, 15 years abundance data for wolf, elk and bison populations was used from the Yellowstone National Park, USA. To construct the models, Lotka-Volterra (L-V) equations were used to create the following prey and predator models: Wolf-Elk, Wolf-Bison and Wolf-Elk-Bison. To calculate the constants for each L-V equation, the prey/predator log-increments was regressed on the predator/prey mean abundance, also known as the log-integral method, then each model was projected and used to construct the three networks. The L-V equations predicted the 15-year wolf data relatively well but the mean square error was high for the elk and bison models. For the Wolf-Elk and Wolf-Bison models, the elk and bison populations steadily decline while the wolf goes extinct after 25 and 10 years, relatively. In contrast, in the three species model, no species go extinct and demonstrate oscillating trends. The practicality of creating a species network for two and three species over time using actual data was demonstrated to be feasible. Hence, constructing larger species networks using this method could offer a powerful way to communicate ecological information to the public.

**Justin Angevaare**, University of Guelph.

*Phylogenetic individual level models: strategies for simulation and inference.*

Phylogenetics is an emergent field that explores the joint dynamics of disease spread and evolution. When epidemic and evolutionary processes occur on similar time scales, phylogenetic models can be used to improve our understanding of disease dynamics. A phylogenetic extension to the individual level models of infectious disease transmission of Deardon et al. (2010) is developed. Computational methods for stochastic simulation and for Bayesian inference are described, and simulation study results are presented. There are opportunities to better inform infectious disease control strategies through the use of these models, when, as is increasingly common, pathogen genetic sequence data are available.



**Reem Almarashi**, Wilfrid Laurier University.

*Infectious Disease Modelling.*

In this work we investigate infectious disease models with immigration of infected individuals. For many such models, there is a unique endemic equilibrium. This means that a surprising phenomenon occurs when the immigration increases from zero. To see this, we first calculate the basic reproduction number  $\mathcal{R}_0$  in the absence of immigration. Without immigration, we have the following common result.

- If  $\mathcal{R}_0 < 1$ , then the only equilibrium is the disease-free equilibrium, on the boundary of  $\mathbb{R}_{\geq 0}^n$ .
- If  $\mathcal{R}_0 > 1$ , then there is the disease-free equilibrium and a unique endemic equilibrium

$$X^* \in \mathbb{R}_{>0}^n.$$

Since non-zero immigration results in a unique equilibrium (which is endemic), for all values of  $\mathcal{R}_0$ , there are implications for what must happen to the disease-free equilibrium as immigration is increases away from zero.

- If  $\mathcal{R}_0 < 1$ , then the disease-free equilibrium moves to the interior  $\mathbb{R}_{>0}^n$
- If  $\mathcal{R}_0 > 1$ , then the disease-free equilibrium moves away from  $\mathbb{R}_{\geq 0}^n$ .

We first study this phenomenon for a specific model (SIV R), and then consider the phenomenon for a general model using matrix theory.

**Zizhen Wang**, University of Guelph.

*R implementation of a generalized linear model for effect of naphthenic acids on hatching probability of fathead minnows.*

As the primary source of acute toxicity of process-affected water, the naphthenic acids (NAs) are a primary problem in the northern Alberta oil sands regions. Recent research has focused on the toxicity of NAs to the highly vulnerable early life-stages of fish like the fathead minnows. Here the fathead minnows have been divided into five fractions with different concentration levels of NAs. We have instigated the effect of different concentration levels of NAs within each fraction on the hatched number of fathead minnow and have examined the Lethal Dose 50 values for each fraction at the overall concentration level. Implementing the differential models in R, the fraction 3 has the most toxicity compared to other fractions. However, resulting from the insufficiency of our data, the ultimate model is not adequate. Also, the transformation on our response variable and explanatory variable is limited and other effective factors should be considered.



## **Stephanie Hughes, University of Guelph.**

*Syndromic Surveillance Systems and Norovirus Outbreak Reporting in Ontario.*

Norovirus is the most common cause of gastroenteritis worldwide, with the total number of cases per year in Ontario second only to the common cold. It is highly infectious and causes outbreaks in closely confined populations, such as long-term care homes and hospitals. Action to reduce the number of cases is required due to its large economic burden. Although laboratory surveillance is practiced in Ontario, it is highly subject to under-reporting and data gaps. Creation of a syndromic surveillance early warning system using TeleHealth Ontario call data will better monitor the spread of disease and detect outbreaks earlier than conventional methods to reduce their scope. This project aims to analyse elements of syndromic surveillance systems and norovirus outbreak reporting in Ontario, and to create a norovirus early warning system to detect outbreaks as early as possible to minimize the burden of disease.

## **Ram Sigdel, University of Guelph.**

*Properties of Simple Coupled Human Environment System-Forest/Other Land Use.*

The present trend of human transgression into any natural system would be very destructive to imbalance most of the ecosystem in future. There are various concerns regarding human role to control the oscillation caused by themselves to the different natural system to guarantee the survival of more species through the co-existence of two corresponding natural system. Either forested land convert to grassland, farmland, urban open land or urban open land, farmland, grassland convert into forested land, certainly human has the major role. Our simple model of human environment system focused on human role to forest and other land use may help to address such concern one step ahead. The qualitative analysis for coupled human-environment system has not often be done before. But through this paper, we have established some global results too. In one hand the absence of injunctive social norm into the model guarantee the global stability of at least one state in which both forested land and other land use co-exist. On the other hand, the presence of injunctive social norms can only guarantee the conditional stability of the interior equilibrium but the system can locally be exist as the mosaic. Under this case, the stability of a state which ignore the importance of rare and endangered species is possible.

## **Mona Ghwila, University of Guelph.**

*The Stability of A Size-Structured Population Model in Terms of its Equivalent Delay System.*

This talk is centered on a stability study of a size-structured population model by considering its equivalent delay system, which consists of a renewal equation for the consumer population birth rate and delay differential equation for the resource concentration. We generalize the delay system by including realistic functional responses and harvesting, also analyze the stability of the equilibrium solution. We discuss how these results affect the dynamics of the population.



**Kevin Church**, University of Waterloo.

*Bifurcations in impulsive differential equations.*

An impulsive differential equations is an ordinary differential equation that exhibits discontinuities in its solutions. For impulsive differential equations with a sufficiently periodic structure, classical bifurcation theory is applicable. However, in the absence of periodicity, one needs a more sophisticated framework to discuss bifurcations of solutions. In this talk, we first introduce the basics of bifurcation theory of non-autonomous dynamical systems and state a non-autonomous analogue of the classical (autonomous) transcritical bifurcation theorem. Next, we provide an extension of this bifurcation theorem to impulsive differential equations, presenting an outline of the proof. We conclude with a simple example motivated by population dynamics.

