# 2021 Masamu CRN Virtual Colloquia Series Abstracts

**Steve Dobson – February 9:**

*Multiple Paternity: How We Used Combinatorials, Bayesian Analyses, and Mixed Models to Study Animal Behavior*

Molecular techniques have revealed widespread multiple paternity of broods among animal species. Date from the laboratory generally report % multiple paternity among broods, mean brood size, and the mean number of sires for broods. Although behavioral ecologists were surprised to find multiple paternity so abundant, just how much multiple paternity should one expect? We used combinatorial formulas to show that if mating males have an equal chance of success at producing offspring, then % multiple paternity depends on the number of offspring and the number of mates. This is a "null model" of multiple paternity, but we only know the number of sires, not the number of mates, from molecular studies. Using assumptions about distributions of brood size and number of sires, we developed a Bayesian MCMC "null model" of multiple paternity. When applied to a large sample of animal species, the null model shows that there is less multiple paternity in nature than one should expect. We contrast species that fit the null model and those that do not, to show that sexual selection favors lower than expected multiple paternity.

A problem with our research is that animal species are not independent, due to shared ancestry of closely related species. With a focus on mammals, we evaluated the extent to which phylogeny can "explain" variation in multiple paternity. This required variance partitioning using mixed models, and another MCMC approach. Levels of % multiple paternity were moderately well explained by phylogeny, indicating that % multiple paternity was similar in closely related groups of mammals. Surprisingly, species that differed most from the null model (viz., those most sexually selected) were not associated with a reduction in the number of mates for females, but rather with a reduction in the number of sires. This suggests that "success matters," such that some males are much more successful than others.

**Sure Mataramvura – February 23:**

*A Hybrid Neural Network GARCH Approach to Forecasting Zimbabwean Inflation Volatility*

Volatility of economic indices like inflation, stocks, exchange rates etc. has generated a lot of interest among researchers especially after the Options Markets' celebrated Black Scholes option pricing formula relied on the assumption of constant underlying asset volatility, itself unobserved on the market. In that sphere (the options market) researchers concluded that, as functions of parameters like strike price, underlying asset volatility is actually not constant and one locus is volatility smile. We posit a priori that Zimbabwe went through periods of economic crisis observed (according to data available) from 1980 and thus modellers face challenges using available data for many technical reasons which to the best of our knowledge has resulted in very little research done in this sphere. This paper provides additional research by modelling and forecasting the inflation volatility present in Zimbabwe, using traditional GARCH models hybridized with Artificial Neural Networks (ANN) and Recurrent Neural Networks (RNN). There are several important conclusions drawn from our results. First, out of the GARCH models, the EGARCH generally performed the best. Second, both the ANN and RNN hybrid models outperformed the

traditional GARCH models by a significant margin. Finally, the hybrid ANN models provided more accurate forecasts during volatile periods when compared with hybrid RNN models.

**Ben Levy – March 9:**

*Modeling the Effect of HIV/AIDS Stigma on HIV Infection Dynamics in Kenya*

In this application we formulate a compartmental system of ordinary differential equations (ODEs) to consider how stigma towards people living with HIV/AIDS has impeded the response to the disease. We take a data-driven approach to embed a time-dependent stigma function within our model for HIV dynamics. We estimate model parameters from published data and explore a range of scenarios to understand the potential impact of different public health interventions on key HIV metrics such as prevalence and disease-related death and to see how close Kenya will get to achieving UN Goals for these HIV and stigma metrics by 2030.

**Samuel Manda – March 23:**

*A Spatial Analysis of COVID-19 in African countries: Evaluating the Effects of Sociodemographic, Geographical, Disease Vulnerability, Preparedness, and COVID-19 Importation Determinants*

The highly contagious coronavirus disease 2019 (COVID-19) pandemic, which started in Wuhan, China, in December 2019, has now become a global public health problem. By 21 March 202, the number of confirmed cases and death in Africa and globally have risen to more than 4.1 million and 110 thousand (Case Fatality Rate (CRF) = 2.68%) and 123.5, 2.7 million (CRF = 2.17%). First cases in the African continents were related to travellers returning from hotspots in Asia, Europe and the United States. The continent's first COVID-19 case was recorded in Egypt on 14 February 2020. Initially confined to the major urban areas, the disease is now widespread in the continent. With the continent’s large populations living in high levels of poverty and crowded informal urban settings, coupled with its fragile health systems, there were global fears that the continent would be particularly devasted by the COVID-19 pandemic. South Africa is the worst-hit African country, with more than 1.5 million cases and 52,000 deaths, followed by countries in North Africa including Morocco (500,000 cases, 8700 deaths); Tunisia (245,000, 8,500); Egypt (194,000, 11,500) and Algeria (116,000, 3057). Burundi (2613, 6); Sao Tome (2145, 34); and Libera (2042, 85) are some of the least-hit in Africa.

Even though Africa has the most vulnerable populations to infectious diseases, there are inter-country variations in the levels of COVD -19 risk. Using publicly available data from the COVID-19 data repository of Our World in Data, we aimed to assess how spatial vulnerability risk factors and neighbourliness between countries could generate potential insights into COVID 19 disparities across Africa by using spatial regression models and hotspot analysis. We found differential effects in risk depending on the Wave of COVID 19. Our findings provide evidence that could guide countries on how to prepare for and respond to the initial and subsequent stages of an infectious disease pandemic like COVID 19. Particularly where a future novel coronavirus is most likely to occur.

**Javier Arsuaga – April 6:**

*Mathematical approaches to the problem of DNA packing in bacterial viruses*

The three-dimensional organization of genomes is a key player in multiple biological processes including the genome packaging and release in viruses. The genome of some viruses, such as bacteriophages or human herpes, is a double stranded DNA (dsDNA) molecule that is stored inside a viral protein capsid at a concentration of 200 mg/ml-800mg/ml and an osmotic pressure of 70 atmospheres. The organization of the viral genome under these extreme physical conditions is believed to be liquid crystalline but remains to be properly understood. A general picture of this organization has been recently given by cryoelectron microscopy (cryoEM) studies that show a series of concentric layers near the surface of the viral capsid followed by a disordered arrangement of DNA fibers near the center of the capsid.

In this talk I will present results from three different mathematical approaches to study the problem of dsDNA packing in bacteriophages. The first approach complements cryoEM observations and uses the formation of knots inside viral capsids as a probe for DNA packing. These results suggest that DNA knots are highly likely upon confinement and that the DNA molecule is chirally organized inside the viral capsid. The second approach aims at identifying the possible sources of the chiral organization of the genome and employs methods from random knotting and brownian dynamics. Our third approach uses continuum mechanics models to describe cryoEM observations as the minima of a liquid crystalline phase. The emergent picture of these approaches suggest that DNA is in a chirally organized liquid crystalline phase in which knots may be the product of liquid crystal defects.

**Jane White – April 20:**

*Crime Modeling*

As the study of crime and criminals, *Criminology* is inherently dynamic, involving individual choices, population behaviours and external triggers. It offers a plethora of theoretical and data-driven research problems that either already do, or could in the future, benefit from insights offered using mathematical modelling tools. In this talk, I hope to inspire you to think about the potential for using these tools in applications of criminology. I will outline the scope of the field and highlight the challenge of linking data, where the focus is on criminal activity, to model structures, where the focus is often on the individual or criminal population. I will do this using a recent modelling project which was parameterised using data from the Western Cape, South Africa. I will draw conclusions from this study and use them as a basis to outline possible future projects where mathematical modelling could significantly support theoretical developments in the field of criminology.

**Farai Mhlanga – May 4:**

*On the pathwise stochastic integration for model-free finance*

This talk is based on a relatively new area in the mathematical finance, where no (possibly wrong) assumptions on the probability measure on the space of price paths (representing possible prices of financial assets on some time interval) are made. The only assumptions considered are the no-arbitrage opportunities widely accepted by the financial mathematics community as essential properties which

prices of financial assets should fulfil. This approach rests on game-theoretic probability and have led to various related notions of outer measures based on the concept of pathwise super-hedging coming from mathematical finance. These outer measures allow to use arbitrage considerations to examine which path properties are satisfied by “typical price paths.” By “typical price paths" in the game-theoretic sense we mean “Investor may get infinitely rich by considering investment in those paths where the considered property fails.” These properties are discussed in greater detail providing a comprehensive outline of various properties and constructions of quadratic variation for model-free càdlàg price paths and integrals driven by such paths. Quadratic variation plays an important role in the analysis of price paths of financial securities and it is sometimes used as the measure of volatility.

**Paul Horn – May 18:**

*Combinatorics and dynamics of a drone network*

In this talk we consider the combinatorics of a model of drone networks, where drones move along trajectories and communicate with each other when in range. When a drone fails, the remaining drones implement a simple ‘switching strategy,’ where they switch trajectories when they do not encounter a drone at an expected communication point. This simple strategy yields both interesting combinatorial and algorithmic problems, and in this talk we will consider several problems concerning the dynamics of the system along with distributed and centralized algorithms for determining information about the drone system. Moreover, we study the robustness of drone networks under randomized failure and show that the switching strategy gives an extremely robust system for coverage.

**Alfred Menezes – June 1:**

*Lattice-Based Cryptography*

Public-key cryptography is used to secure connections between web browsers and web sites, to verify the authenticity of software upgrades, to securely pair two Bluetooth devices, to provide end-to-end security in messaging apps such as WhatsApp and Signal, and so on. However, the public-key cryptosystems in use today, namely RSA and elliptic curve cryptography, are known to be breakable by quantum computers. To address this threat, a lot of work is being done on developing and standardizing public-key cryptosystems that resist attacks even by large-scale quantum computers.

This talk will give an introduction to lattice-based cryptography, the most widely studied kind of quantum-safe public-key cryptography. The lecture will be introductory and will not assume any prior knowledge of cryptography or lattices.

**Michael Washington – June 15:**

*Building Local Capacity to use Simple Models to make Informed Public Health*

Models are a valuable tool to address many policy makers questions and to raise new questions in which they never considered. Unfortunately, modelers are not readily available in low to middle income

Ministries of Health (MOHs), especially in public health. And the ability to attract highly, talented quantitative people to work within the MOHs is difficult due to the demand for their skills in other occupations. In addition, the lack of research to address specific public health problems at the local level are not available to assist in making an informed decision, which increases the need for models. The Data to Policy (D2P) program, developed by the U.S. Centers for Disease Control and Prevention and Vital Strategies as part of the Bloomberg Data for Health Initiative, uses an output-oriented approach to bridge the data to policy gap in public health with economic analyses and simple models. This approach involves team-based training and mentoring of health policy professionals to develop data-driven policy briefs that respond to government health priorities. This presentation will be about the success of the D2P program in using simple models to make public health decisions, and how it helped to build local capacity to perform future analysis.

**Innocent Maposa – July 6:**

*Heritability of vascular health across three generations in the Birth to Twenty-Plus cohort: Implications for vascular disease prevention in sub-Saharan Africa*

Within South Africa, one quarter of all adults are hypertensive and one in five deaths are from cardiovascular disease (CVD)[1]. CVD mortality and morbidity are set to rise with life expectancy now at 64 years (an increase of 10% in the last decade)[2], and increasing levels of overweight and obesity (68% women, 31% men)[3]. Much focus is placed on detecting and treating CVD, but with limited healthcare resources, successful programs are needed to prevent CVD through targeting younger, at-risk individuals. To inform strategies for the prevention of hypertension and cardiovascular disease in South Africa, we sought to determine the heritability of cardiovascular health (blood pressure, arterial stiffness, and cardiac and carotid ultrasound measures) across three generations. We use MRC/Wits Developmental Pathway Unit's data from a cross-sectional study conducted in a longitudinal family cohort within one of the largest tertiary hospital in Africa situated in a historically disadvantaged township. There were (n=66) biological families with all three generations. For exploring associations between pairs of family members on some traits, we use the random family approach which implements restricted-resampling regression methods and for estimating heritability, we use the Linear Mixed Effects Model (aka animal model). In addition to the LMM ReML approach, we use two Bayesian paradigms ie based on Gibbs Samplers and the Hamiltonian Monte-Carlo algorithms. We use the HMC-NUTS sampler as implemented in Stan. We noted that two traits, ie, cIMT and LVMI showed at least one significant pair for possible genetic influence, and these were explored further. On average, from the three estimation methods used, heritability for cIMT is estimated at between 33% and 41% while that of LVMI is between 4% and 28%. Main limitations include small family sample size and missing paternal pedigree information. Low prevalence of hypertension beyond first generation may also have contributed to the uncertainty. Lange et al (2002) in 252 individuals with type 2 diabetes from 122 families estimated a cIMT heritability of 0.32 (se=0.17) but higher among current smokers (h2 = 0.41). Data provide some empirical evidence that subclinical cardiovascular disease may have a significant genetic component and warrants more research to ascertain genes involved. We also show the performance of the algorithms given the high dimensional nature of the data and note that the HMC algorithm is faster compared to the other MCMC algorithm.

**Pete Johnson – July 20:**

*Frobenius Problems in Arbitrary Rings*

Frobenius problems, inspired by a 19th century theorem about the positive integers, are illustrated by the following recreational puzzle, paraphrased and updated from a 1951 Sunday supplement discovered recently in a landfill outside of Cincinnati, Ohio:

In Lower Slobbovia, the slobbinet is a coin worth 5 Slobbovian piasters. Recently the dictator Slobbo III capriciously decreed the coinage of two new coins, the slobbino, worth 7 piasters, and the slob, worth 13 piasters. Slobbo has made sure that he will enjoy an unlimited supply of all 3 coins. What prices will Slobbo be able to pay with exact change, using only coins of these denominations?

The generalization that will be given of such problems to arbitrary rings (in the original, the ring involved is the ring of integers) was stumbled upon and explored by a succession of participants in the Auburn University Research Experience for Undergraduates in Algebra and Discrete Mathematics, in 2008, 2011, 2012, 2015, and 2019. The aim of the talk is to illustrate how mathematics can advance driven by sheer curiosity and aided by sheer luck.

**Suzanne Lenhart – August 24:**

*Optimal control for management of aquatic population models*

Optimal control techniques of ordinary and partial differential equations will be introduced to consider management strategies for aquatic populations. In the first example, managing invasive species in rivers can be assisted by adjustment of flow rates. Control of a flow rate in a partial differential equation model for a population in a river will be used to keep the population from moving upstream. The second example represents a food chain on the Turkish coast of the Black Sea. Using data from the anchovy landings in Turkey, optimal control of the harvesting rate of the anchovy population in a system of three ordinary differential equations (anchovy, jellyfish and zooplankton) will give management strategies. A third example involves partial differential equations for a fishery model.

**Mokhwetha Mabula – September 7:**

*Order-isometries of ordered asymmetrically normed cones*

If $X$ is a real vector space, a function $p : X \to \mathbb{R}$ is called an asymmetric norm on $X$ if is positively homogenous, sub-additive and $p(x) = p(-x) = 0$ if and only if $x = 0$, and $(X, p)$ is called asymmetrically normed space. A subset $K$ of $X$ is called a cone if it is closed under addition, closed under scalar multiplication and the only element of $K$ with additive inverse is zero. A cone $K$ induces a partial ordering on a vector space $X$ by the formula $x \preceq y$ if and only if $y - x \in K$ for all $x, y \in X$. Therefore, the triple $(X, p, \preceq)$ is called an ordered asymmetrically normed space, and $(K, p \preceq)$ is called ordered asymmetrically normed cone.

In this talk, we discuss the notion of order-isometries between two ordered asymmetrically normed cones. The topology induced by asymmetric norm satisfies the separation axiom, $T_0$ but not necessary $T_1$. An application of our definition of order-isometries on cones will be discussed as well. In particular, Banach-Stone theorem.
The work is a joint work with O. Valero and C. Alegre.

**Inger Fabris-Rotelli – September 21:**

*Spatial sampling for a rabies vaccination schedule in rural villages*

Efforts are being made to contain rabies in Tanzania, reported in the southern highland regions, since 1954, and endemic in all districts in Tanzania currently. It has been determined that mass vaccination of at least 70% of a domestic animal population is most effective in reducing trans- mission of rabies. Current vaccination campaigns in Tanzanian villages have many administrative and logistical challenges. Animals roam freely, making a full population vaccination impossible. Spatial sampling of households in villages is proposed, where optimality is measured through the distance traversed by the vaccinator by foot for vaccinating at each sampled household. The walking distance is attained by incorporating a driving network between optimally determined stopping points from which the vaccinator then walks for executing vaccinations, while ensuring the 70% coverage of the animal population. We illustrate the sampling schemes on a real dataset using simulations. A systematic regular spatial sampling is found to be optimal. The vaccination scheme proposed, provides an effective way to manage a vaccination campaign. We further discuss future vaccination models taking into account the distribution of animals, and the growth of villages. Spatial modelling of COVID-19 in South Africa and the effect of spatial coverage of COVID-19 vaccination will also be discussed.

**David Erwin – October 5:**

*Distance domination and generalized eccentricity in graphs with given minimum degree*

Let $G$ be a connected graph and $k \in \mathbb{N}$. The $k$-domination number of $G$ is the smallest cardinality of a set $S$ of vertices such that every vertex of $G$ is within distance $k$ from some vertex of $S$. While for $k = 1$, i.e., for the ordinary domination number, the problem of finding asymptotically sharp upper bounds in terms of order and minimum degree of the graph has been solved, corresponding bounds for $k > 1$ have remained elusive. In this talk we present an asymptotically sharp upper bound on the $k$-domination number of a graph in terms of its order and minimum degree, which significantly improves on bounds in the literature. We also obtain an asymptotically sharp upper bound on the $p$-radius of graphs in terms of order and minimum degree. For $p \in \mathbb{N}$, the $p$-radius of $G$ is defined as the smallest integer $d$ such that there exists a set $S$ of $p$ vertices of $G$ having the property that every vertex of $G$ is within distance $d$ of some vertex in $S$. We also present improved bounds for graphs of given order, minimum degree and maximum degree, for triangle-free graphs and for graphs not containing a 4-cycle as a subgraph.

**Eduard Campillo-Funollet – October 19:**

*Forestry and mathematics: the effect of peri-urban woodland on pollution absorption*

In order to reduce the net carbon emissions to fight climate change, many countries are promoting planting new woodland. Since the number of trees that a country can plant is limited, an optimisation question arises: where is best to plant the trees? Motivated by studies suggesting that there are many benefits to planting new woodland around a city, we derive a reaction-diffusion model to study the effect of a peri-urban forest on the pollution within the city, and we find the optimal radius of the forest in order to minimise the pollutant concentration.

**Bernard Omolo – November 2:**

*A Comparison of Parametric and Semi-parametric Models for Microarray Data*

Microarray technology has revolutionized genomic studies by enabling the study of differential expression of thousands of genes simultaneously. Parametric, nonparametric, and semi-parametric statistical methods have been proposed for gene selection within the last two decades years. In an effort to find the “gold standard", the performance of some common parametric and nonparametric methods have been compared in terms of power to select differentially expressed genes and other desirable properties. However, no such comparisons have been conducted between parametric and semi-parametric models. In this study, we compared a semi-parametric model based on copulas with a parametric model (the quantitative trait analysis or QTA model) in terms of power and the ability to control the Type I error rate. In addition, we proposed a simple algorithm for choosing an optimal

copula. The two approaches were applied to a publicly available melanoma cell lines dataset for validation. Both methods performed well in terms of power, but the copula approach was notably the better. In terms of the Type I error rate control, the two methods were comparable.