Abstracts Book for the 2018 Joint NZSA/ORSNZ Conference

Edited by A. Jonathan R. Godfrey and Katharina Parry

Last updated: 11:40, 26 Nov 2018

Book of Abstracts [-]

This volume contains abstracts for all presentations made during the 69th Annual Conference of the NZ Statistical Association and the 52nd Annual Conference of the Operations Research Society of NZ, held jointly at the Manawatu campus of Massey University during November 2018.

Our sponsors

The NZSA Student Prizes have been sponsored by Harmonic Analytics, and the ORSNZ Young Practitioners prizes by Suez Smart Solutions.

We appreciate the support of our sponsors.

Summary of talks

We have six keynote addresses, leaving 70 contributed talks. This includes a handful of lightning talks.

Tuesday 27 November

We welcome conference attendees to the Manawatu campus of Massey University today. We do hope you enjoy your stay in Palmerston North.

Wi-fi on campus

You should look for the visitor wi-fi service. Massey people should use the Massey networks and everyone else will use the "MUGuests" (unsecured) network.

Find out more about this service on the frequently asked questions page

This is a three device 600MB capacity per person service.

Alternative services that can be reached include Eduroam and Inspire, the latter is a local ISP that provides fairly good service across Palmerston North.

Overview

Time		What's on
12:00		Registration desk opens
2:00		Opening
2:10		Plenary address: Prof. Jean Yang Chair Katharina Parry
3:10		Afternoon tea break
3:40		Contributed talk session 1
5:20		End of formal sessions
5:30		Informal event for students and early career statisticians

Maths & Stats Teachers' Symposium

There will be an event for secondary school teachers from the lower North Island and anyone interested in education at the secondary school level in mathematics and statistics subjects on this day. It starts at 9.00a.m. and will be colocated with our conference. We have invited symposium attendees to join us for our opening keynote speaker and the afternoon's contributed sessions.

Contributed Session 1

Time	AH1	AH2	AH3	AH5
Theme	Biostatistics	Optimisation 1	Lolly Scramble 1a	Lolly Scramble 1b
3:40	Miller	J Xiao	Tularam	Cosic
4:05	Bilton	Jagtenberg	Morris	Hedderley
4:30	Lamont	Krishnan	Alsaedi	Govindaraju
4:55	Mohammad	Foster	Penny	

Chair	Volunteer needed	Volunteer needed Penny	Govindaraju

Wednesday 28 November

Overview

Time		What's on
9:00		Opening
9:05		Plenary address: Assoc. Prof. Ilze Ziedins Chair: Mike O'Sullivan
10:00		Morning tea break
10:30		Contributed talk session 2
12:10		Lunch break and ORSNZ Annual General Meeting in AH2
1:10		Contributed talk session 3
2:00		Time to allow movement between rooms
2:05		Plenary address: Prof. Geoff Jones Chair: Steve Haslett
3:00		Afternoon tea break
3:30		Contributed talk session 4
5:10		End of formal sessions
5:15		Meetings: NZSA Annual General Meeting in AH3 and ORSNZ Council Meeting in AH2
6:30		ORSNZ dinner, including presentation of awards, at Wharerata (on campus)

Contributed Session 2

Time	AH1	AH2	AH3
Theme	Health	Natural Hazards	Optimisation 2
10:30	KM Lin	Huang	Mason
10:55	Adams	Frigerio-Porta	Downward
11:20	Cleland	Shahzadi	Gosse
11:45	C Lin	Bebbington	Raith
Chair	volunteer wanted	Bebbington	Mason/Raith

Contributed Session 3

Time	AH1	AH2	AH3
Theme	Forensics	Transport 1	Business
1:10	Gupta	Tidswell	S Marshall
1:35	Wen	Prendergast	Thomas

Chair	Curran	volunteer wanted	volunteer wanted

Contributed Session 4

Time	AH1	AH2	AH3
Theme	Nature & Energy	Dimension Reduction	Theory and methods
3:30	Stewart	Chen	Turner
3:55	Adiga	Xu	Raubenheimer
4:20	Shanker	Tran	Yee
4:45	Smith	L Li	X Xiao
Chair	volunteer wanted	volunteer wanted	volunteer wanted

Thursday 29 November

Overview

Time		What's on
9:00		Opening
9:05		Plenary address: Prof. Stefan Nickel Chair: Andrea Raith
10:00		Morning tea break
10:30		Contributed talk session 5
12:10		Lunch break
1:10		Contributed talk session 6
2:00		Time to allow movement between rooms
2:05		Plenary address: Prof. Alan Welsh Chair:
3:00		Afternoon tea break
3:30		Analytics Forum
5:10		End of formal sessions
6:00		Pre-dinner meet-up at the Coachman
7:00		NZSA dinner, including presentation of awards

Contributed Session 5

Time	AH1	AH2	AH3
Theme	Lolly Scramble	Health	Energy
10:30	Mahmoodjanlou	Maindonald	Read
10:55	F Li	Davies	Philpott
11:20	Angelin-Bonnet	Cabrera-Guerrero	Noble
11:45	Tabassum	O'Sullivan	McDonald

Chair	J.Marshall	O'Sullivan	McDonald

Contributed Session 6

Time	AH1	AH2	AH3
1:10	Lightning talks	Krsinich	Sibanda
1:35	Thomasen	Pearson	Hazelton
Chair	volunteer wanted	volunteer wanted	volunteer wanted

Lightning talks

Speakers for the five lightning talks will be given at most four minutes to speak and then 30 seconds to field one question before the next speaker will start. The five speakers are Jonathan Godfrey, Rina Parry, Jonathan Marshall, Sarah Pirikahu, and Ellie Johnson. N.B. The order of the speakers has not yet been finalised.

Analytics Forum

We are pleased that with the support of Fonterra,

we are able to host an Analytics Forum with the theme, "Analytics in the Primary Sector". The forum will be held in AH1.

Speakers are Vanessa Cave, Kim Frew, and Mark Piper.

Friday 30 November

Time		What's on
10:00		Opening
10:05		Inaugural ANZ Journal of Statistics Discussion Paper
11:15		Comfort break
11:20		Panel discussion
12:30		Wind up and thanks
1:00		Safe travels

Inaugural ANZ Discussion Paper

Dr Francis Hui will present the paper: Hui, F.K.C., Muller, S. and Welsh, A.H. (2019) Testing Random Effects in Linear Mixed Models: Another Look at the F-test (with discussion).

To be followed by discussants, Chris Triggs and Geoff Jones, and an opportunity for the presenter to respond. Questions from the audience will close this session.

Chair: Prof. Martin Hazelton, Theory and Methods Editor, ANZJS.

Panel discussion - What is the future for statistics

Our panel (John Maindonald, Marie Luckman, Ian Westbrooke, and Vanessa Cave) will each have ten minutes to put their views before the panel as a whole responds to questions from the audience.

Chair: Jonathan Godfrey

Adams, Thomas

A Simulation Model of Auckland City Hospital's Emergency Department

Thomas Adams, Michael O'Sullivan, and Cameron Walker

Department of Engineering Science, University of Auckland

The Ministry of Health agrees on several health targets with the district health boards in New Zealand. One of the targets is for 95% of patients to be admitted to hospital, discharged, or transferred from an Emergency Department within 6 hours. The management of the Emergency Department at Auckland City Hospital are considering restructuring the staff in an effort to reduce waiting times and ensure this target is met. In particular they are contemplating moving to a zone or `pod' model with staff dedicated to each part of the Emergency Department. Additionally in the near future more staff will be available to work in the Emergency Department. This presentation presents a simulation model designed to assist the management in deciding whether and how to restructure their staff, and the effect of trialling several staffing configurations within the simulation on patient's waiting times and length of stay.

This presentation is eligible for the ORSNZ Young Practitioners Prize.

Adiga, Rishi

Stochastic Mixed Integer Programming for Geothermal Well Decision Optimization

Rishi Adiga, Andy Philpott, and John O'Sullivan

Department of Engineering Science, University of Auckland

Drilling geothermal wells has a very high capital cost, and as such, it is imperative to maximize value from wells by selecting them optimally. An important technology used when making well placement and scheduling decisions is computer simulation of production. This is usually done manually, with experts creating reservoir models, and simulating and comparing different production scenarios. These reservoir models have the problem of non-uniqueness; different models can fit available data equally well but give differing predictions.

This paper uses Stochastic Mixed Integer Programming models to address this problem. An economic model was created to calculate Net Present Values for a set of candidate wells at different start times, and the interactions between them, using AUTOUGH2 simulation results of an example geothermal system. Binary decision variables were used to select the combination of wells and start times that would maximize total NPV. Stochastic, multistage optimization models were formulated to account for uncertainty in the numerical simulation data, reflecting a realistic geothermal project, and to make hedging decisions based on this uncertainty.

This presentation is eligible for the ORSNZ Young Practitioners Prize.

Alsaedi, Yasir

Solar Price Forecasting Based on Various Univariate Time-Series Models: A Case Study from Queensland, Australia

Yasir Alsaedi[12], Gurudeo Anand Tularam[2], and Victor Wong[3]

1. Department of Mathematics, Umm Al-Qura University, Makkah, Saudi Arabia

2. Environmental Futures Research Institute , Griffith University, Brisbane

3. Department of Accounting, Finance and Economics, Griffith University, Nathan, Queensland

In the coming years, it is anticipated that the generation of power through solar means will play a key role in the electricity markets. Therefore, energy analysts and government organisations alike require guidelines to help them choose the most appropriate

forecasting techniques for achieving accurate predictions of solar pricing trends.In this study, three types of univariate models are considered, namely the simple exponential smoothing (SES), Holt-Winters exponential smoothing (HWES) and autoregressive integrated moving average (ARIMA) models. In order to determine the most appropriate model, four different strategies are applied as selection criteria in order to quantify the accuracy of the model predictions, namely the mean squared error (MSE), root-mean-square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) strategies. The three models are compared by applying them to the time series of the average solar prices for Queensland, Australia. Data covering August 2012 to March 2018 are extracted from the database of Solar Choice, Australia's solar power installation brokering and solar quote comparison service. The comparison indicates that the SES model performs better than the HWES model in terms of its predictive power, with a confidence interval of 95%. However, the ARIMA (1, 1, 1) model yields the best results, which leads us to conclude that this sophisticated and robust model outperforms the other simple, albeit flexible, models in relation to the solar market. This study aims to help policy makers and industry marketing strategists select the best forecasting method for the solar market.

Angelin-Bonnet, Olivia

How `R` and `Julia` hang out together to simulate complex biochemical interaction networks

Institute of Fundamental Sciences, Massey University

Simulating the expression of genes and the interactions between their products using computational models is a challenging task. For a simulation to be representative of the biological complexity, one must build complex mathematical models to represent the system behaviour. Under a stochastic framework, simulating the system state over time amounts to computing the sequence of biochemical reactions occurring at each time step. This is achieved using a stochastic simulation algorithm, and results in long computational running time. In order to fasten up the computation, we decided to use the programming language Julia to run such simulations. During this talk I will be presenting the concept of expression data simulation algorithm, and explain how we used the R package `XRJulia` to call Julia functions from within R.

This presentation is eligible for the NZSA Student Prize.

Bebbington, Mark

Intraeruption forecasting using semi-Markov models

Mark Bebbington[1] and Susanna Jenkins[2]

1. Volcanic Risk Solutions, Massey University, Palmerston North

2. Earth Observatory of Singapore, Nanyang Technological University, Singapore

Forecasting eruption onsets has received much attention, in both the short- and long-term. However, unlike an earthquake, an eruption is not easily reduced to an instant in time. Any usable definition of an eruption has to allow for activity over scales ranging from minutes to decades, and can do so only by allowing for multiple eruptive phases. These phases can be defined by having different styles (e.g., effusive and/or explosive) of activity and/or quiescent periods between them. We have coded a database of multiple-phase eruptions into 8 major styles. The result contains c. 700 multi-phase eruptions, with eruptions having up to 33 non-quiescent phases. The resulting record of transitions between states is relatively dense, and so a probability tree is infeasible to model the possible phase sequences. Instead we will turn to Markov chain models. Markov chains describe the state path under the assumption that only the present state determines the probability of the next state, but the definition of 'state' can be extended in this case to include preceding or following quiescence. This enables us to calculate likelihoods for the next step of the eruption, conditional on (e.g.) the elapsed duration of the current phase, and the duration of the quiescence preceding it. The results are illustrated on recent eruptions, and the effect of partitioning the data base to isolate volcano types or compositions matching those of the target volcano examined.

Bilton, Timothy

Estimating genetic relatedness in polyploid species from sequencing data

Timothy P. Bilton[1,2], Matthew R. Schofield[1], Ken G. Dodds[2] and Michael A. Black[3]

1. Department of Mathematics and Statistics, University of Otago, Dunedin

2. Invermay Agricultural Centre, AgResearch, Mosgiel

3. Department of Biochemistry, University of Otago, Dunedin

Genetic relatedness refers to the proportion of genetic information between two individuals that is derived from a common ancestor. Computing estimates of genetic relatedness is important in many genetic applications, such as genomic selection, parentage analysis and inferring genetic population structure. Estimators for genetic relatedness have been developed and extensively used in diploid species that have two sets of chromosomes

(e.g., humans and most animal species). However, some species (known as polyploids) have a more complex genetic architecture in that they have more than two sets of chromosomes. For these species, relatedness estimation is more complicated, as there are additional genetic phenomena at play, and to date there has been little research in computing relatedness in polyploids. Additionally, estimating relatedness with data generated from the latest sequencing technology is complicated by the presence of errors in the form of missing parental information and incorrectly sequenced bases. These errors, if not taken into account, lead to underestimation of relatedness. We derive theoretical results extending relatedness estimation to polyploids and propose a new estimator that accounts for the errors associated with sequencing data. Using simulated and real data, we examine the performance of these estimators under various scenarios. Lastly, we discuss some limitations of the estimators and scope for future research.

 This presentation is eligible for the NZSA Student Prize.

Cabrera-Guerrero, Guillermo

Solving the Direct Aperture Optimisation Problem using Local Search Strategies

Leslie Pérez Cáceres, Ignacio Araya, Denisse Soto, and Guillermo Cabrera-Guerrero

Escuela de Ingeniería Informática, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

In this work, we aim to solve the direct aperture optimisation problem (DAO) in radiation therapy for cancer treatment by means of two novel heuristic local search strategies. In the DAO problem, the goal is to find a set of deliverable apertures shapes and intensities so we can irradiate the tumor according to a medical prescription without producing any harm to the surrounding healthy tissues. Traditional inverse-planning approach starts by selecting a set of beam angles radiation will be delivered from (Beam Angle Optimisation problem). Then, the intensities for such angles are computed by optimising some “meaningful” objective function (Fluence Map Optimisation problem). Finally, the apertures shapes that will allow us to deliver the radiation computed during the previous step are generated (Multi-Leaf Collimator Sequencing Problem). Clearly, this sequential approach provokes that the number of apertures shapes and the time patients will be exposed to radiation depends, to a large extent, on both the beams selected when solving the BAO problem and the intensities computed when solving the FMO problem.

Unlike this sequential approach, in the DAO problem, constraints associated to the number of deliverable aperture shapes as well as physical constraints are taken into account during the fluence map optimisation process. Thus, we do not longer need any leaves sequencing

procedure after solving the DAO problem. This integrated approach allows us to find an intensity map that not only optimise the radiation that is delivered to the patient but also minimises the number of deliverable apertures shapes.

To solve the DAO problem, we try two novel heuristic local search strategies on a prostate case and compare the obtained treatment plan to the one obtained using the traditional sequential approach. Results show that our algorithms are able to find treatment plans that are very competitive when considering the number of deliverable aperture shapes.

Cave, Vanessa

Digital Agriculture at AgResearch

AgResearch Ltd, Hamilton

Agriculture is one of New Zealand's largest export income earners. However, digital technologies and analytics are transforming the agricultural industry. AgResearch, a Crown Research Institute that exists to provide science and technology for the benefit of the New Zealand's agricultural sector, is involved in a number of scientific research projects aimed at preparing New Zealand for the era of "Digital agriculture". Digital agriculture refers to the use of new and advanced technologies, integrated into one system, to help make farming (more) profitable and sustainable. It offers the potential to provide solutions for the challenges facing New Zealand agriculture. In this presentation, we will describe a suite of recent research projects that can be loosely grouped under the Digital Agriculture umbrella.

This talk is part of the Analytics Forum to be held on Thursday 29 November in AH1.

Chen, Tong

Approximations to the distribution of a large quadratic form in normal variables

Tong Chen and Thomas Lumley

Department of Statistics, University of Auckland

Quadratic forms of Gaussian variables occur in a wide range of applications in statistics. They can be expressed as a linear combination of chi-squareds. The coefficients in the linear combination are the eigenvalues $\lambda_1, \ldots, \lambda_n$ of ΣA, where A is the matrix representing the quadratic form and Σ is the covariance matrix of the Gaussians. There is quite a bit of literature on this problem, but it mostly deals with approximations on small quadratic forms ($n < 10$) and moderate p-values ($p > 10^{-2}$). Motivated by genetic applications, we look at large quadratic forms ($n > 1000$) and very small p-values ($p < 10^{-4}$). We compare existing methods under genetic settings and show that a leading-eigenvalue approximation,

which only takes the largest k eigenvalues, has computational advantage without any loss in accuracy. For time complexity, a leading-eigenvalue approximation reduces the computational complexity from $O(n^3)$ to $O(n^2k)$ on extracting eigenvalues and avoids speed problems with approximating the sum of n terms. For accuracy, under large quadratic forms, moment methods are inaccurate for very small p-values and Farebrother's method is not useable if the minimum eigenvalue is small, but the saddlepoint approximation has bounded relative error even in the extreme right tail.

This presentation is eligible for the NZSA Student Prize.

Cleland, Isaac

Improved Staff Rostering Algorithms

Isaac D. Cleland, Andrew Mason, and Michael O'Sullivan

Department of Engineering Science, University of Auckland

The Staff Rostering Problem involves optimising the assignment of staff to shifts, while fulfilling various rules and preferences associated with these assignments. These problems are difficult to solve automatically; they are NP-Hard, regularly have a large number of variables, have a non-linear or multi-objective cost function, and are tightly constrained. We have developed a suite of algorithms to solve these problems using both Column Generation (Branch and Price) and Column Generation based Matheuristics which we will outline in this presentation.

The International Nurse Rostering Competition (INRC) was a global competition held in 2010 to benchmark the best Staff Rostering solvers against each other on a large number of difficult- to-solve Nurse Rostering Problems. We will present some preliminary results from our suite of algorithms in solving the INRC problems together with some real world case studies.

This presentation is eligible for the ORSNZ Young Practitioners Prize.

Cosic, Jelena

Bayesian Network as a Modelling Tool for Increasing Knowledge on the Factors Influencing Vineyard Longevity and Sustainability

Jelena Cosic[1], Steffen Klaere[1], Matthew Goddard[2] and Bruno Fedrizzi[3]

1. Department of Statistics and School of Biological Sciences, University of Auckland

2. School of Biological Sciences, University of Auckland

3. School of Chemical Sciences, University of Auckland

The long-term project "Resilient and Profitable NZ wine industry" has the objective to study the impact of different vineyard management techniques on the vineyard longevity and profitability, and to increase the knowledge of the factors influencing longevity and profitability. To find meaningful answers appropriate quantifiable outcomes need to be obtained. Profitability of a vineyard can be quantified by its yield and quality of the end product, while health will be studied in a more holistic way by developing a vineyard ecosystems model incorporating the data obtained from different areas of interest. The empirical nature of data collection makes a computational ecosystem modelling approach the most suitable. Such approaches are quite common and popular in ecology, and are promising for this project. Of particular interest are Bayesian Networks (BNs) which have received increased attention throughout several research fields for their ability to incorporate prior knowledge and to handle incomplete data. BN have also been shown to efficiently avoid overfitting the data, and avoiding the observation of "chimeric" effects. We will use BN to model vineyard ecosystems incorporating microbial, fungal and eukaryotic molecular data, chemical profiles, meteorological information, and other markers at different points in the life cycle of vineyards, and discover the differences vineyard managements make with respect to resilience and profit. Some of the challenges that we see are: variables that have been measured on different time scales, a large amount of microbial data and uncertainty of the interactions of components included in our ecosystem.

Davies, Tilman

Modelling multilevel spatial behaviour in binary-mark muscle fibre configurations

Tilman M. Davies[1], Matthew R. Schofield[1], Jon Cornwall[2], and Philip W. Sheard[3]

1. Department of Mathematics and Statistics, University of Otago, Dunedin

2. School of Medicine, University of Otago, Dunedin

3. Department of Physiology, University of Otago, Dunedin

The functional properties of skeletal muscles depend on the spatial arrangements of fast and slow muscle fibre types. Qualitative assessment of muscle configurations suggest that muscle disease and normal ageing are associated with visible changes in the spatial pattern, though a lack of statistical modelling hinders our ability to formally assess such trends. We design a nested Gaussian CAR model to quantify spatial features of dichotomously-marked muscle fibre networks, and implement it within a Bayesian framework. Our model is applied to data from a human skeletal muscle, and results reveal spatial variation at multiple levels across the muscle. The model provides the foundation for future research in describing the extent of change to normal muscle fibre type parameters under experimental or pathological conditions.

Downward, Anthony

Multistage stochastic programming with stagewise-dependent objective coefficient uncertainty

Anthony Downward, Oscar Dowson, and Regan Baucke

Department of Engineering Science, University of Auckland

We present a new algorithm for solving linear multistage stochastic programming problems with objective function coefficients modelled as a stochastic process. This algorithm overcomes the difficulties of existing stochastic dual dynamic programming methods, which require discretisation of the objective coefficient states. Using an argument based on the finiteness of the set of possible cuts, we prove that the algorithm converges almost surely. Finally, we demonstrate the practical application of the algorithm on a hydro-bidding example with the spot-price modelled as an auto-regressive process.

Foster, Shasa

Infinite Horizon in Stochastic Dual Dynamic Programming

Shasa Foster, Ben Fulton, Tony Downward, and Andy Philpott

Department of Engineering Science, University of Auckland

This report discusses the extensions made to stochastic dual dynamic programming to produce an infinite horizon stochastic dual dynamic programming methodology useful for solving large multistage convex stochastic optimization problems. The infinite horizon methodology was implemented by extending SDDP.jl, an existing package (in the programming language Julia) using stochastic dual dynamic programming. Parallel processing and cut selection heuristics present in SDDP.jl were integrated into the infinite horizon extensions of SDDP.jl. The infinite horizon SDDP algorithm was then applied to a hydro-thermal scheduling model of the NZEM (the DOASA model) for various configurations of the NZEM.

This presentation is eligible for the ORSNZ Young Practitioners Prize.

Frew, Kim

Analytics Use Cases for Value Creation at Fonterra

Fonterra, Auckland

I will speak to the use cases Fonterra is exploring to create new value, how we ensure people are empowered at the centre of analytics design and delivery and the potential for analytics in our ecosystem for innovation.

This talk is part of the Analytics Forum to be held on Thursday 29 November in AH1.

Frigerio Porta, Gabriele

A statistical model for earthquake and/or rainfall triggered landslides

Gabriele Frigerio Porta, Mark Bebbington, Geoff Jones, and Xun Xiao

Institute of Fundamental Sciences, Massey University, Palmerston North

Coseismic and rainfall-triggered landslides are a common hazard in many terrains, and the risk associated with them can be quantified, usually by probabilistic modelling. These events are well-documented as a special case of a cascading hazard chain, and the

assessment is commonly done via spatial modelling of susceptibility (suppressing temporal dependence) or tailoring models to specific areas and events.

The interaction between Earthquakes and rainfall is not usually implemented in a model, as it is considered coincidental. However, because landslides have multiple triggering factors, there is a need for a statistical model that incorporates both features, in a manner such that the separate and joint effects can be estimated. This helps with understanding the interactions between primary events in the triggering of a single secondary hazard type that is crucial for generally applicable multi-hazard methodologies.

The presented work aims at the apportioning of the relative and combined effect on landslide triggering by earthquakes and rainfall using a discrete approximation to a multivariate hierarchical point process. Doing so provides a building block in a general framework with the potential to be extended to other chains of events. A case study on the Italian region of Emilia Romagna is included, using one of the longest and most complete landslide data sets known. The resulting model uses a Zero-Inflated Poisson to handle the 98+% of space-time with no landslide occurrence. Interpretations of the triggering influence between the multiple factors in different configurations are derived.

Time permitting, a rainfall simulation 'plug-in' will be presented.

 This presentation is eligible for the NZSA Student Prize.

Godfrey, Jonathan

Efficient production of a conference book of abstracts

A. Jonathan R. Godfrey and Katharina Parry

Institute of Fundamental Sciences, Massey University, Palmerston North

Creation of an online book of abstracts has proven challeging for other conference organisers. We chose to use the bookdown approach generally used for technical manuals which have dozens to hundereds of printed pages as the basis for this conference's book of abstracts for a variety of reasons:

- it can be viewed online using a variety of devices;
- it was easy to update;
- it created new opportunities for inclusion of R commands and output;
- it looks good;
- and if anyone wants the printed version, they can do it themselves using a pdf file we provide.

 This lightning talk will show you how easy it was to prepare this key element of organising a conference; we may have learned some lessons worth sharing along the way.

Gosse, Michelle

Creation of a realistic population for a microsimulation

Michelle A. Gosse, Mark S. Bebbington, and Jonathan C. Marshall

Institute of Fundamental Sciences, Massey University, Palmerston North

Population-based microsimulations using agents require external validity so that the simulation results approximate reality. The construction of a realistic population is particularly crucial for agent-based simulations. Agent assignment to households is a detail that tends to be glossed over in publications. This presentation will present an overview of the construction of a simulated New Zealand town to model disease transmission and spread from a food-borne illness. Four external datasets were used to construct the population, of which three were random-rounded for confidentiality reasons. As well as households, agents have been allocated to schools (children) and employers (employed agents). Most of the work has been implemented in R, using probabilistic methods disaggregate random-rounded data. Some useful tips for implementation in R will be covered.

Govindaraju, Kondaswamy

Guardbanding for decisions

Institute of Fundamental Sciences, Massey University, Palmerston North

The *guardbanding* principle is used industry in order to manage the risk of noncompliance to specifications which are often externally imposed. In a risk management perspective, guardbanding is an *epistemic* solution to *aleatory* problems. Most statistical inference methods for decision problems do not explicitly incorporate epistemic uncertainties as a matter of guardbanding.

We trialled the guardbanding principle for an industrial problem successfully, and some of the successes achieved will be presented. The principle of guardbanding is also extended to legal metrology applications, and other conformity testing problems where limiting values take precedence over specifications. The talk will be followed by a general discussion on the scope and limitations of guardbanding for general statistical decision problems.

Gupta, Anjali

Comparison of Raman Spectra of Inks using Functional Data Analysis

Anjali Gupta[1], James Curran[1], Patrick Buzzini[2], and Christopher Triggs[1]

1. Department of Statistics, University of Auckland

2. Department of Forensic Science, Sam Houston State University

Forensic scientists commonly use Raman spectroscopy to generate spectral data for ink samples from inkjet printers. The purpose is to examine whether the ink samples in question have originated from the same printer or not. In traditional sense, this can be thought of as an application of Hotelling's T^2 test but for very high-dimensional data. However, commonly used statistical methods fail to analyse spectral data due to its high-dimensionality and insufficient sample observations in forensic laboratories. This leads to singular variance-covariance estimators and unstable comparisons amongst the samples. In order to compare spectral data, we propose to take advantage of the work done in the field of functional data analysis and use the Hotelling's T^2 test statistic derived by Pini et al. (2018). We conduct the comparisons between cyan, magenta and yellow inks from several printers using this method. I will discuss the benefits of using functional data analysis and the results of these comparisons in this talk.

This presentation is eligible for the NZSA Student Prize.

Hazelton, Martin

Approximate samplers for linear inverse problem

Martin Hazelton, Mike McVeagh, and Bruce van Brunt

Institute of Fundamental Sciences, Massey University, Palmerston North

Statistical inverse problems occur when we wish to learn about some random process that is observed only indirectly. Inference in such situations typically involves sampling possible values for the latent variables of interest conditional on the indirect observations. This talk is concerned with linear inverse problems for count data, for which the latent variables are constrained to lie on the integer lattice within a convex polytope (a bounded multidimensional polyhedron). An illustrative example arises in transport engineering where we observe vehicle counts entering or leaving each zone of the network, then want to sample possible interzonal patterns of traffic flow consistent with those entry/exit counts. Other areas of application include inference for contingency tables, and capture-recapture modelling in ecology.

In principle such sampling can be conducted using Markov chain Monte Carlo methods, through a random walk on the lattice polytope. However, it is challenging to design

algorithms for exact sampling that are both scalable and have guaranteed theoretical properties. In this talk I will describe some current work on developing approximate samplers that are nonetheless guaranteed to connect all lattice points carrying appreciable posterior probability.

Hedderley, Duncan

A comparison of partial least squares and random forests for metabolomics analysis

Duncan Hedderley[1], Tony McGhie[1], Sarah Cordiner[1], Mary Ann Lila[2], and Carol L Cheatham[3]

1. New Zealand Institute of Plant and Food Research, Palmerston North

2. Plants for Human Health Institute, North Carolina State University, North Carolina

3. Nutrition Research Institute, University of North Carolina at Chapel Hill, North Carolina

Metabolomics is the untargeted measurement of metabolites in samples order to discover unanticipated relationships between treatment groups and their metabolites. A standard approach has been to use partial least squares (PLS) to discriminate treatment groups. The R mixOmics package includes 'sparse' procedures which identify the most important variables contributing to the linear combinations which are PLS dimensions. Alternatively, random forests (Breiman 2001) use a classification tree approach to find variables which best discriminate between groups. We compare the information these approaches provide using data from a study on the effect of supplementing older people's diets with blueberries.

Huang, Willie

Modelling non-volcanic tremor observations using autoregressive hidden Markov models

Chia-Chuan (Willie) Huang and Ting Wang

Department of Mathematics and Statistics, University of Otago, Dunedin

Non-volcanic tremor is a series of low frequency seismic vibrations detected in the tectonic areas. Such tremor events are associated with slow slip events which are commonly observed within the same source region of megathrust earthquakes. Understanding the connections between these three types of seismic activities may therefore aid forecasts of large destructive earthquakes. Insights into such mechanism can be extracted by studying the spatiotemporal migration pattern of tremors.

A hidden Markov model was proposed recently to study the spatiotemporal variation of tremor activities in Southwest Japan. However, the dependence structure of that hidden Markov model was unable to capture the correlation between tremor clusters. We developed a 1D autoregressive hidden Markov model to analyse the spatiotemporal behaviour of the tremor activity in the Tokai region, Southwest Japan. To facilitate the analysis, we reduced the dimension of the data by projecting the 2D locations onto a line northeast of the region. This new model captured major correlation arose in the tremor activity. It classified the tremor systems into three types: active, transit and background. By comparison, the forecasting ability of the new model outperforms the hidden Markov model developed in earlier research. Future studies will relax the model assumptions further to provide a more robust model.

 This presentation is eligible for the NZSA Student Prize.

Hui, Francis

Testing Random Effects in Linear Mixed Models: Another Look at the F-test (with discussion)

Francis K. C. Hui[1], Samuel Muller[2], and Alan H. Welsh[1]

1. Mathematical Sciences Institute, Australian National University, Canberra

2. School of Mathematics and Statistics, University of Sydney

In many applications of mixed models, an important step is assessing the significance of all or a subset of the random effects included in the model. There has been extensive research conducted into testing the significance of random effects (or variance components) in linear mixed models, from correcting the asymptotic null distribution to simulation based methods and variations thereof.

This talk re-examines one of the earliest and simplest methods of random effects testing in linear mixed models, namely the F-test based on linear combinations of the responses, or FLC test. For current statistical practice, we argue that the FLC test is underused and should be given more consideration especially as an initial or "gateway" test for linear mixed models. We present a very general derivation of the FLC test that is applicable to a broad class of models where the random effects and/or normally distributed errors can be correlated. We discuss three advantages of the FLC test often overlooked in modern applications of linear mixed models: computation speed, generality, and its exactness as a test. Empirical studies provide new insight into the finite sample performance of the FLC test, identifying cases where it is competitive or even outperforms modern methods in terms of power, as well as settings where it performs worse than simulation based methods for testing random effects, all the while being faster to compute.

While straightforward to understand and implement, the FLC test stimulates deeper thinking into the notion of treating random effects as fixed effects, and its implications on

estimation and inference in mixed models more generally. In the latter portion of this talk, we make connections between the principle behind the FLC test and estimation approaches such as penalized quasi-likelihood and variational approximations, as well as model selection techniques such as information criteria, penalized likelihood methods, and the concept of degrees of freedom. Ultimately, we hope to motivate future research into methods that, in some way, regard random effects as fixed, what happens to these methods asymptotically, and towards more empirical comparisons with standard approaches that treat the random effects as they are.

Biography

Francis Hui is a lecturer in statistics at the Mathematical Sciences Institute, Australian National University (ANU). Francis completed his PhD in ecological statistics in 2014 at the University of New South Wales, and afterwards undertook a postdoctoral fellowship at the ANU supervised by Alan Welsh (ANU) and Samuel Mueller (University of Sydney). He has been unable to leave the glorified country town that is Canberra since. Francis' research spans a mixture of methodological, computational, and applied statistics, including but not limited to dimension reduction and variable selection, longitudinal and correlated data analysis, approximate likelihood estimation and inference, and semiparametric regression. Much of his application is driven by multi-species distribution modeling in community ecology, but more recently he has dabbled in longitudinal modeling for mental health and income. All of his research is complemented by copious amounts of tea drinking and unhealthy amounts of anime watching.

Session details

This plenary address will be delivered in AH1 on Friday 30 November at 10:00.

Jagtenberg, Caroline

Organizing an International Optimization Challenge

Caroline Jagtenberg[1] and Joaquim Gromicho[2]

1. Department of Engineering Science, University of Auckland

2. ORTEC, Zoetermeer, The Netherlands

Optimization challenges are important because they stimulate research in new topics and provide the means to compare algorithms objectively. The organization of such a challenge involves many aspects, and the process is typically not visible to anyone outside the organizing committee. This article, however, describes an international optimization challenge - the third edition of the VeRoLog Solver Challenge - from the perspective of the organizers. The challenge was hosted by ORTEC, one of the world's leaders in optimization software and analytics solutions. The problem central to the challenge was based on a rich Vehicle Routing Problem (VRP) that occurs in practice at one of ORTEC's customers. In

total, 28 teams world wide participated in the challenge that ran for eight months. This article describes the background of the challenge and reflects on the organization. Hopefully, this will provide useful insights for anyone who considers hosting an optimization challenge in the future, as well as inspire researchers to participate in the next VeRoLog Solver Challenge.

 This presentation is eligible for the ORSNZ Young Practitioners Prize.

Johnson, Ellie

Identifying outbreaks of campylobacteriosis: A case study in Hawke's Bay

Ellie Johnson and Jonathan Marshall

Institute of Fundamental Sciences, Massey University, Palmerston North

In August 2016, a large water-borne outbreak of campylobacteriosis occurred in Havelock North in the Hawke's Bay, resulting in an estimated 5,000 cases of disease from a population of 15,000. In such large outbreaks, identifying the common exposure is possible as a wealth of epidemiological information are available from the large number of cases and, with knowledge of the exposure, future risk may be reduced or eliminated through targetted mitigation. However, most cases of food and water-borne campylobacteriosis in New Zealand and world wide are due to isolated events where only a small number of people get sick. In such situations, identifying a common exposure to link cases can be difficult due to sparse epidemiological information. One technique is to model the underlying spatial U_i and temporal R_t trends in notification rate λ_{it}, with a spatio-temporal field of outbreak indicators $X_{R[i]t}$ across regions $R[i]$ over and above these trends:

$$\begin{aligned} Y_{it} &\sim \text{Poisson}(n_{it}\lambda_{it}) \\ \lambda_{it} &= R_t + U_i + \beta_{R[i]}X_{R[i]t} \end{aligned}$$

We apply this model to Hawke's Bay campylobacteriosis notifications from 1997 through 2016, and show that many of the potential outbreaks prior to the large outbreak in August 2016 also occurred in the town of Havelock North.

 This is a lightning talk.

Jones, Geoff

Diagnostic Test Evaluation or: How I Learned to Stop Worrying and Love Bayesian Statistics

Inaugural Professorial Address by Geoff Jones

Institute of Fundamental Sciences, Massey University, Palmerston North

In some inferential situations, adopting a Bayesian framework makes the analysis much simpler. There are other situations where, because of identifiability issues, Bayesian analysis becomes the only viable possibility.

In this talk I will trace my gradual adoption of Bayesian methods through a number of examples, and focus eventually on the use of Bayesian latent class analysis for data from diagnostic testing when a "gold-standard" test is not available. Issues to be discussed include the use of multiple tests, the assumptions of conditional independence and homogeneity of test performance, sampling from under-identified models, and modelling between-group heterogeneity in prevalence.

Biography

Geoff Jones was recently made Professor of Biostatistics at Massey University Palmerston North, where he has been working (or at least has been employed) since 1997. He completed his PhD at the University of California, Davis in 1996, after many years teaching in secondary schools in the UK and Malaysia. He has collaborated widely with scientists in a range of disciplines, but particularly with the Veterinary Epidemiologists at Massey's Epicentre, most recently in the area of diagnostic testing. He is currently a member of the OiE (World Organization for Animal Health) Collaborating Centre for Diagnostic Test Validation Science. He is particularly proud of his work with Professor Steve Haslett for the UN World Food Programme, on small area estimation of deprivation indices to guide the allocation of food aid in Third World countries.

Session details

This plenary address will be delivered in AH1 on Wednesday 28 November at 2:00.

Krishnan, Harish

A Mathematical Model to Solve Google's Ad Allocation Problem

Harish Krishnan and Andrew Mason

Department of Engineering Science, University of Auckland

Sponsored search auctions are ubiquitous these days and search engine marketing companies such as Google make billions of dollars in revenue. These auctions are usually second price auctions. Advertisers spend significant portion of their marketing budget on sponsored search and this is only expected to increase in the future. Search Engines are likely to solve a complex optimization problem of allocating ads to different advertisers to achieve one or all of the following objectives.

1. Maximize Ad Revenue
2. Improve User Experience
3. Maximize Return on Investment for Advertisers.

In this paper, the auction process is explained formally. Empirical data collected from experiments to understand the auction process is analyzed. Slates based optimization model - a model from the existing literature is explained and critiqued using the empirical data and a new slots based optimization model that is consistent with the data is proposed.

Simulated results indicate that the slots based optimization model outperforms the random allocation of ads by 22% and a Greedy Approach based on Google's documented rules by 12%.

This presentation is eligible for the ORSNZ Young Practitioners Prize.

Krsinich, Frances

Rental indexes and decomposed house price indexes from big(ger) data

Ministry of Business, Innovation and Employment, Wellington

MBIE and Stats NZ have recently collaborated to develop an improved rental index for the NZ Consumers Price Index directly from tenancy bond data. This requires new methods, which are gaining traction internationally for measuring inflation from big data such as scanner and online data.

MBIE are also using these methods to develop a house price index from CoreLogic sales and valuation data, which can be decomposed into land and structure components. In addition to an improved house price measure for New Zealand, this will enable land price measures

and price-to-cost measures for the entire stock of residential property, able to be disaggregated down to the territorial authority level.

Lamont, Alastair

A missing value approach for breeding value estimation

Alastair Lamont[1], Matthew Schofield[1], and Richard Barker[2]

1. Department of Mathematics and Statistics, University of Otago, Dunedin

2. Division of Sciences, University of Otago, Dunedin

For a particular trait, an individual's breeding value is the genetic value it has for its progeny. Accurate breeding value estimation is a critical component of selective breeding, necessary to identify which animals will have the best offspring. As technology has improved, genetic data is often available, and can be utilised for improved breeding value estimation. While it is cost efficient to genotype some animals, it is unfeasible to genotype every individual in most populations of interest, due to either cost or logistical issues. This missing data creates challenges in the estimation of breeding values. Modern approaches adapt least-squares estimation to allow for unobserved genetic data. To do so requires particular assumptions and approximations which may not be well-suited to typical livestock data. Breeding values can also be estimated using Bayesian methods, but existing approaches do not accommodate missing genetic data. We specify a model for genetic inheritance, an approach used in human applications. This allows missing genotypes to be estimated alongside other parameters using Bayesian methods.

This presentation is eligible for the NZSA Student Prize.

Li, Fangyao

Air Pollution Prediction by Using the Matching Pursuit Algorithm

Fangyao Li[1], Christopher M. Triggs[1], Bogdan Dumitrescu[2], and Ciprian Doru Giurcăneanu[1]

1. Department of Statistics, University of Auckland

2. Department of Automatic Control and Computers, University Politehnica of Bucharest, Romania

Particulate matter (PM) is one of the most dangerous air pollutants. PM is measured as $PM_{2.5}$ and PM_{10}, where 2.5 and 10 are the maximum diameters of the particles. $PM_{2.5}$ is

much more expensive to measure compared with PM_{10}. Hence, it is important to estimate the $PM_{2.5}$ concentration for the areas where the measurements do not exist. Statistical methods are vital in prediction air pollution data. We will briefly introduce the PM data collected at four sites of Auckland by the National Institute of Water and Atmospheric Research (NIWA), then illustrate the prediction of $PM_{2.5}$ concentration by using the matching pursuit algorithm (MPA). Novel stopping rules for MPA, based on information theoretic criteria, will also be discussed (Li et al. 2018).

This presentation is eligible for the NZSA Student Prize.

Li, Lingyu

Fuzzy clustering using adjacent-categories logit model via finite mixture model

School of Mathematics and Statistics, Victoria University of Wellington

Traditional analysis of ordinal data treats the outcome either as nominal or continuous variables. The nominal approach ignores the ordinal property whereas the continuous approach introduces assumptions about the ordinal level spacing - thus these traditional approaches can lead to a loss of statistical power, or can introduce bias.

This talk presents cluster analysis of ordinal data utilising the natural order information of ordinal data. Three models usually used in ordinal modelling are discussed: the proportional odds model, the adjacent-categories model and the ordered stereotype model.

In our research, the data take the form of a matrix where the rows are subjects, and the columns are a set of ordinal responses by those subjects to, say, the questions in a questionnaire. We implement model-based fuzzy clustering via a finite mixture model, in which the subjects (the rows of the matrix) and/or the questions (the columns of the matrix) are grouped into a finite number of clusters. We will explain how to use EM (Expectation Maximisation) algorithm to estimate the model parameters. Specifically, we illustrate the details of using Adjacent-Categories logit model to perform row/column and bi-clustering. This clustering method differs from other typical clustering methods such as K-means or hierarchical clustering, because it is a likelihood-based model, and thus statistical inference is possible.

This presentation is eligible for the NZSA Student Prize.

Lin, Christina

Healthcare Pathway Discovery, Conformance, and Enrichment

Michael O'Sullivan[1], Andreas W. Kempa-Liehr[1], Christina Lin[1], Delwyn Armstrong[2], and Randall Britten[3]

1. Department of Engineering Science, University of Auckland

2. Waitemata District Health Board, Auckland

3. Orion Health, Auckland

Healthcare pathways are critical for reducing clinical variability and maximizing health outcomes. This project will investigate the utilization of Business Process Modelling (BPM) to provide a scaffold for healthcare pathway discovery, conformance analysis, and enrichment.

The efficacy of the BPM approach is demonstrated via a case study that applies the process mining pipeline to discover appendicitis and cholecystitis healthcare pathways from hospital records. Two years' worth of data from 2015 to 2017 on appendicitis and cholecystitis pathways are used. The resulting pathways have been reviewed by clinical experts and this review confirmed that the two discovered healthcare pathways comply with their knowledge of real clinical cases.

The healthcare pathways are subsequently enriched with demographic data and machine learning methods are utilised to explore factors that influence patient recovery time. A partial least squares regression model which estimates patient recovery time based on the available information at the end of surgery is developed. The machine learning models presented here have the potential to be very useful for hospital scheduling purposes.

The designed process mining pipeline is effective for analysis of simple healthcare pathways. Future work would apply the same pipeline to complex healthcare pathways.

This presentation is eligible for the ORSNZ Young Practitioners Prize.

Lin, Kuan-Min

A random forest model to classify head and neck radiotherapy treatment plans

Kuan-Min Lin and Andrea Raith

Department of Engineering Science, University of Auckland

Radiotherapy treatment planning involves making conflicting treatment trade-offs between irradiating cancerous tissues and sparing surrounding critical organs. However, due to geometrical structural variations, assessing the quality of a plan is difficult. Plan tuning is conducted in a trial-and-error manner without knowing the improvement potential. This planning process leads to variable plan quality as well as inefficiency in producing treatment plans. This study proposes a planning tool based on random forest and data envelopment analysis (DEA) to support the planning process. DEA performs plan evaluation by comparing a plan to an ideal defined by the dataset and has been originally proposed for assessing the quality of prostate radiotherapy plans. For treatment sites with a more complicated structural relationship such as head and neck, DEA losses discrimination power due to the increased number of clinical criteria. We propose to use a random forest model to assess the clinical criteria and classify if a plan is satisfactory. Only plans deemed satisfactory are subsequently assessed for improvement potential by DEA.

This presentation is eligible for the ORSNZ Young Practitioners Prize.

Mahmoodjanlou, Ahmad

Measuring Stability of Stochastic Model of Traffic Flow

Ahmad Mahmoodjanlou, Martin Hazelton, and Katharina Parry

Institute of Fundamental Sciences, Massey University, Palmerston North

In this talk, we look at scenarios where some unforeseen occurrence happens to a traffic network on a particular day. For example, temporary road closures or an accident disrupts the normal flow of traffic. We would observe that on such days the cost of using affected routes increases, given that the incident would mean it takes a long time to pass through that route. Using a measure called the coefficient of reactivity, we investigate the how fast a system returns to normal after various types of such one-off events. We then show how this measure can be used to determine in which systems a stochastic model of traffic flow in can be well approximated by its deterministic counterpart.

This presentation is eligible for the NZSA Student Prize.

Maindonald, John

Use of the Beta-binomial for Modeling Dose-Mortality Data

Statistics Research Associates, Wellington

In the exposure-mortality data that is of interest, the exposure measure may be time in coolstorage, or exposure to varying levels of a fumigant for a fixed time period. Each replicate provides information, for each of a number of exposure levels, on mortality. There are then several replicates for each combination of one or more species and/or lifestage and/or temperature or other conditions. The aim is to compare tolerance between species/lifestage combinations, and to predict, if possible, a level of exposure that is likely to lead to close to 100% mortality.

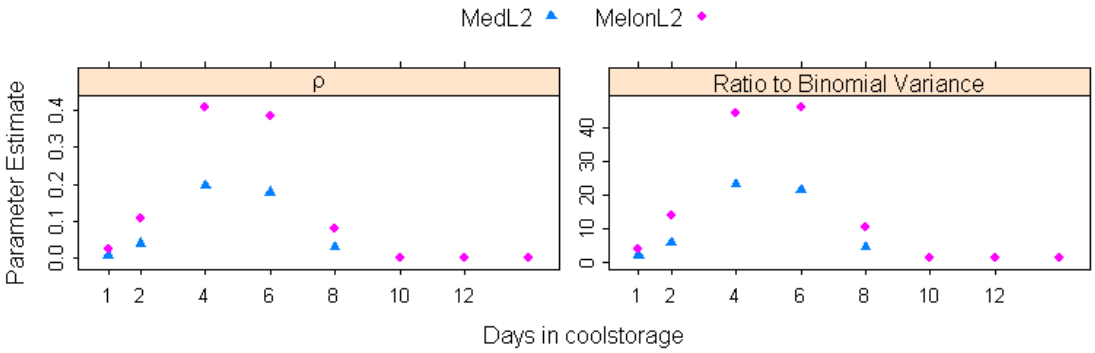
Abilities have recently been added to the glmmTMB package for R that facilitate the use of models that assume a beta-binomial error, with the scale parameter modeled as a function of explanatory variables. Lines (or, in principle, curves), fitted assuming a suitable link function, can be modeled either as fixed or as random effects. Possibilities are a random intercept that is drawn from a normal distribution, or a random bivariate normal intercept and slope.

Interesting and important implications, not previously considered in the analysis of organic produce quarantine data, or in the design of the relevant experiments, flow from the modeling of the scale parameter. It is easiest to explain how these effects operate when expressed in terms of the intra-class correlation ρ, which is a simple function of glmmTMB's choice of scale parameter. The graph shows the estimated patterns of change of ρ with days in coolstorage, for the larval 2 lifestages of two different fruitfly species, in one dataset examined. Predicted mortalities range from 1 day:0.08; 4: 0.44; 8:0.999 for MedL2, and 1:0.14; 4:0.3; 8:0.66 14:0.997 for MelonL2.

An intra-class correlation of ρ implies that, given a probability π, the variance cannot be reduced below $\pi(1-\pi)\rho$, no matter how large the sample size n. Thus, if $\rho = 0.1$, the variance cannot be reduced below $0.1\pi(1-\pi)$. A sample size of 90 reduces it to 10% above this minimum. Points to note are:

- The assumption that the binomial variance is multiplied by an amount that is the same at all points on the scale, as commonly made when analyzing using quasibinomial errors, gives much too much weight to points at mid-range mortalities, and too little to points at high mortalities.
- A further consequence is that slope estimates will be overly influenced by statistical variation in mid-range mortalities. There are strong implications for the use of insect material.

The beta-binomial may be too extreme in downplaying the benefits of increasing sample size. This calls for further investigation.



Marshall, Jonathan

markeR: a shiny app for marking assignments

Institute of Fundamental Sciences, Massey University, Palmerston North

Marking assignments is a lot of work. In ye olden times with pen and paper, it was relatively quick to provide concise, useful comments to each student on their work highlighting where they did particularly well or went wrong. With paperless assignments, this is more time consuming, as PDF annotation typically takes longer and is more cumbersome than hand writing comments on paper. Further, the manual tasks of collating and entering marks is still subject to arithmetic and data entry error with manual checking being required. Where multiple markers are involved in marking the same question, there is additionally consistency issues to be ironed out which is hard to check manually.

markeR is a shiny app designed for marking assignment work. It allows mark entry and a selection of comments to be applied to each question, storing the results per question in a database and producing a summary report for each student with both marks and comments on each question. While the main goal is to make the marking process more efficient, it has the added benefit of providing additional data on which comments are used most often, and how the marks are distributed per sub-question that may be useful for adapting teaching or assessment practice.

This is a lightning talk.

Marshall, Sarah

Modelling a renewing free repair warranty using an alternating geometric process

Richard Arnold[1], Stefanka Chukova[1], Yu Hayakawa[2], and Sarah Marshall[3]

1. School of Mathematics and Statistics, Victoria University of Wellington

2. School of International Liberal Studies, Waseda University, Japan

3. Department of Mathematical Sciences, Auckland University of Technology

The cost of rectifying warranty claims can be significant and therefore accurate estimation of the warranty cost is important. We model the product life cycle using an alternating geometric process (AGP), in which there is an alternating sequence of operational and repair times. To accommodate the ageing of the product and repair equipment, we use a decreasing geometric process to model the consecutive operational times and an increasing geometric process to model the consecutive repair times. We model the warranty claims process and evaluate the warranty servicing costs under non-renewing, renewing, and restricted renewing free repair warranties. In this talk, theoretical results relating to the renewing and restricted renewing free repair warranties will be presented. Properties of the model will be demonstrated using a simulation study.

Mason, Andrew

Applying optimisation and analytics to help find missing Māori shareholders.

Andrew J. Mason, Sam Gilmour, Andy Philpott, and Jon Symons

Department of Engineering Science, University of Auckland

For historical reasons, Māori assets such as land or companies are often owned by a large number of individuals. Parininihi ki Waitotara Incorporated (PKW) is one such organisation that was formed in the 1970's to take ownership of a number of Māori land blocks in Taranaki. PKW has over 5000 so-called 'missing' shareholders who are owed almost $4.5 million in unpaid dividends. PKW face a significant challenge in trying to trace these shareholders, a challenge that keeps increasing as shareholders pass away and are succeeded by family members. We are developing analytics software systems to help PKW locate these missing shareholders. These systems need to solve a number of analytics problems ranging from optical character recognition of historical Māori Land Court documents through to the probabilistic optimised matching of a network of people and

associated relationships. This presentation discusses these problems in more detail, presents the algorithms we are developing, and discusses our results to date.

McDonald, Barry

Modelling correlated wind farms and the risk of extreme power fluctuations

Barry W. McDonald and Sheng Gong

Institute of Natural and Mathematical Sciences, Massey University, Auckland

Wind farms produce maximum power when the wind speeds are within a certain interval. Outside that interval the power output from a wind farm can change very dramatically with wind speed. If major changes occur at different farms, within a short time period, this can have serious short-term implications for the national power supply. We discuss some models for wind changes and power output changes in the light of the fact wind speeds at nearby farms are positively correlated.

Miller, Poppy

Identifying hotspots of rat activity and quantifying their effect on the risk of leptospirosis in urban slums

Poppy Miller[1,2], Chris Jewell[1], Peter Diggle[1], and Kate Hacker[3]

1. Lancaster University, United Kingdom

2. AgResearch, Palmerston North

3. Yale University, New Haven, Conneticutt

Leptospirosis is much more prevalent in urban slums than in most other parts of the world. It is suspected that one of the key predictors of leptospirosis risk is exposure to rats. However, rodent control has so far been largely ineffective at reducing the burden of leptospirosis in urban slum environments where Norway rats (Rattus norvegicus) are the primary reservoir hosts.

Our study aims to quantify the risk attributable to rat exposure compared to other known risk factors in an urban slum. As part of this, we estimate a spatio-temporal rat activity surface over the study area which can be used to target rodent control more effectively.

The study design was a spatially continuous constrained random sample (340 points) at locations throughout a Brazilian urban slum community. An additional 100 random points were added at close range, to distinguish between short range spatial variation and

underlying noise, allowing identification of hotspots at very small ranges. Tracking boards at each study location detect if a rat crosses them. All residents of the study area were asked to participate in the study, and their leptospirosis infection status was measured every 6 months.

At all points, we performed environmental surveys and used satellite imagery to derive spatially relevant covariates. We analyzed the rat activity data using a generalised linear spatial model to evaluate the association between rat activity and nearby environmental characteristics, and to create high-resolution predictive maps of relative rat activity/abundance. This allowed identification of defined environmental features of slum communities, which predict rat abundance.

We then predicted rat abundance near each study participants homes, and used these to quantify of the risk of leptospirosis attributable to nearby rat abundance compared to other risk factors.

This presentation is eligible for the NZSA Student Prize.

Mohammad, Khandoker Akib

Efficient Estimation For The Cox Proportional Hazard (PH) Cure Model

School of Mathematics and Statistics, Victoria University of Wellington

While analysing time-to-event data, it is usually assumed that a certain fraction of subjects will never experience the event of interest and they are said to be cured. When this important feature of survival models is taken into account then the models are commonly referred to as cure models. In the presence of covariates, the conditional survival function of the population can be modelled by using cure model which depends on the probability of being uncured (incidence) and the conditional survival function of the uncured subjects (latency). Usually in the Cox PH cure model, a combination of logistic regression and Cox PH regression is used to model the incidence probability and latency respectively. Here profile likelihood approach has been used to estimate the cumulative hazard and the regression parameters. However, the estimator of the baseline hazard is an implicit function that is why it is very difficult to find an estimate of variance of profile likelihood estimator in the Cox PH cure model. We can solve of the problem of implicit function by considering the 'statistical generalized derivative' which is used to calculate the score function and do not require differentiability of the score function to show the asymptotic normality of the profile likelihood estimator. We applied the theory of 'statistical generalized derivative' in the Cox PH cure model to calculate the efficient score function and establish the asymptotic normality of the estimator without assuming the derivative of the score function in the model.

This presentation is eligible for the NZSA Student Prize.

Morris, Lindsay

Spatio-temporal modelling for point referenced data

School of Mathematics and Statistics, Victoria University of Wellington

Point referenced spatial data (often referred to as geostatistical data) describes measurements that have been observed at a particular location, and finds application in climatology, ecology, environmental health, real estate marketing and many more. Gaussian processes (GPs) are the most common method for modelling spatio-temporal processes that produces point referenced data. They are ideal as they only require specification of the mean and covariance function. However, restrictive assumptions about the process are usually made, with the most constraining being the notion of stationarity. While this assumption makes inference and prediction simpler, most spatial processes have covariance structures that change over the study domain. Hence it is inappropriate to model under an assumption of stationarity.

This presentation describes the method of partitioning spatial point referenced data into smaller subdomains in order to reduce the impact of non-stationarity. Several conditional autoregressive (CAR) models are fit (in a Bayesian framework) to New Zealand particulate matter data that has been partitioned using K-means clustering. The posterior distributions of the covariance parameters are compared to those from the full data, and data partitioned based on the main island the observation is from.

This presentation is eligible for the NZSA Student Prize.

Nickel, Stefan

Health Care Logistics

Institute of Operations Research, Karlsruhe Institute of Technology, Germany

Healthcare logistics addresses the efficient planning, realization and control of patient-, material- and information-flow within the healthcare sector. Therefore, the use of Operations Research (OR) methods plays a crucial role. It is important to not only put emphasis upon the economic efficiency but also to take the quality of care and patient satisfaction into account. On the other hand, healthcare logistics should not get involved in (core) medical decisions.

Healthcare logistics addresses healthcare facilities and service providers at all levels, for example general practitioners (GP) providing primary care or emergency departments (ED) treating patients with pressing health issues. Care pathways containing several different providers as well as the interaction of providers and services, e.g. when patients are transferred to a hospital by their GP, are also targeted. Usually, processes in healthcare grew historically ("We have always done it this way."). Consequently, processes have not been analyzed critically until reforms of the health system have put increasing pressure on

the providers. Nowadays, especially hospitals are looking for possibilities to improve their processes. The success of logistics concepts in healthcare lies in resource conservation for non-value-adding activities (not directly relevant for the healing process, e.g., administrative work) and high resource utilization for value-adding activities (e.g., surgery) while the personnel shall not be over-utilized (i.e., no overtime). Moreover, the interaction of appropriate logistics concepts with modern OR models allow a patient centered treatment, by respecting the needs of a patient and allowing a smoother process. The digitalization of the healthcare sector offers additional opportunities.

In this talk, we give an overview on how OR methods can be used in order to support process optimization in healthcare organizations. For a subset of healthcare logistics applications arising in different healthcare sectors OR models and numerical results mainly from real world projects will be presented. Examples include: location planning for ambulances and GPs, appointment planning, emergency department simulation, layout planning for hospitals and many more. We will also give some information on how healthcare logistics research is organized in Karlsruhe.

Biography

Stefan Nickel is a full professor at the Karlsruhe Institute of Technology - KIT (Germany) and one of the directors of the Institute of Operations Research. He obtained his PhD in mathematics at the Technical University of Kaiserslautern (Germany) in 1995. From 1995 to 2003 he was assistant and associate professor in mathematics at the Technical University of Kaiserslautern. After a full professor position at the Saarland University (Chair of Operations Research and Logistics) from 2003 to 2009, he joined the Karlsruhe Institute of Technology as the Chair in Discrete Optimization and Logistics in April 2009. Since 2014 he is the dean of the Department of Economics and Management at KIT. Stefan Nickel was also member of the scientific advisory board as well as of the management board of the Fraunhofer Institute for Applied Mathematics (ITWM) in Kaiserslautern from 2004-2016. Since 2011 he additionally holds the positions of one of the directors of the Karlsruhe Service Research Institute (KSRI) and of the Research Center for Computer Science (FZI). From 2006-2015 he was editor-in-chief of Computers & Operations Research and is still consulting editor. Moreover, he is editor-in-chief of Operations Research for Health Care. He has coordinated the Health Care working group within the German OR society (GOR) and has been the president of GOR from 2013-2014.

Stefan has authored or co-authored 5 books as well as more than 100 scientific articles in his research areas Locational Analysis, Supply Chain Management, Health Care Logistics, and Online Optimization. He has been awarded the EURO prize for the best EJOR review paper (2012) and the Elsevier prize for the EJOR top cited article 2007-2011. In addition he conducted several industry projects with well-known companies such as BASF, Lufthansa, Miele, and SAP.

Session details

This plenary address will be delivered in AH1 on Thursday 29 November at 9:00.

Noble, Alasdair

A new approach to calculating flux measurements from chamber based soil greenhouse gas data.

Alasdair Noble[1], Rod Venterea[2], Asger Pedersen[3], Soren Petersen[4], Tim Parkin[5], and Cecile de Klein[6]

1. Agresearch, Lincoln

2. USDA - ARS, St. Paul, USA

3. Department of Clinical Medicine , Aarhus University, Denmark

4. Department of Agroecology - Soil Fertility, Aarhus University, Denmark

5. USDA - ARS, Ames, USA

6. Agresearch, Invermay

Measuring the flux of a greenhouse gas being emitted from soil is important for assessing National totals for Climate Agreement protocols. In a "good" experiment a chamber is placed on the soil and four measurements are taken over time. Diffusion theory suggests that the concentration of the gas of interest (Nitrous Oxide in this case) will increase non-linearly over time. The variable of interest is the instantaneous gas emission at time 0, ie the slope of the line at that time. Various linear and non-linear models are fitted and the best fitting model is generally accepted to predict the slope. With so few data points R squared values of over 99% are common for all of the models so the decision process is not clear cut. We propose a model averaging approach which improves the stability of the estimates.

O'Sullivan, Mike

Case Studies in Healthcare Simulation

Mike O'Sullivan and Cameron Walker

Department of Engineering Science, University of Auckland

Simulation models enable healthcare service providers to evaluate performance of their current system, examine how this performance changes with changes to the demand for their service (e.g., future population growth), as well as experiment with changes to that system before implementing any changes. This presentation will include case studies of simulation projects with real-world healthcare services and learnings from those projects.

Parry, Katharina

What's in a full stop?

Institute of Fundamental Sciences, Massey University, Palmerston North

CohMetrix is a programme that allows you to enter a text, containing for example a snippet out of a book, and in return you get a breakdown of how "easy" or "cohesive" the language used is. The text needs to be entered with punctuation complete, however, with song texts, is the way we punctuate necessarily as consistent as with other texts? And do potential differences in where we place a full stop have a significant impact on the output we get from CohMetrix?

This lightning talk will give some insights on these questions.

Pearson, Janet

Analysis of a national gambling and violence study using a block-wise selection process

Janet E. Pearson[1], Maria Bellringer[2], Katie Palmer du Preez[2], Denise Wilson[3], Jane Koziol-McLain[4], Nick Garrett[1], and Max Abbott[2]

1. Department of Biostatistics and Epidemiology, Auckland University of Technology

2. Gambling and Addictions Research Centre, Auckland University of Technology

3. Taupua Waiora Centre for Maaori Health Research, Auckland University of Technology

4. Interdisciplinary Trauma Research Centre, Auckland University of Technology

We have data on a convenience sample of 164 adult (18+) gamblers recruited from New Zealand gambling treatment services, as part of a larger study to investigate the occurrence of family violence. Here, we are interested in the effect of gambler gender on the relationship between having dependent children at home, and the gambler perpetrating or being a victim of family violence. Further, we wish to adjust for the many inter-related potential confounder variables collected, and to do so in a way that illuminates the effect of adjusting for variables in the macro social domain, and then the more micro psychological and gambling domains. The identified process to manage this is described as follows. We sequentially add blocks of conceptually related and statistically correlated variables, to each of our two models (one predicting perpetration, and one predicting victimisation). Thus, in each model, we can see the effect on the primary predictor (the interaction of gambler gender and having dependent children) of adjusting for each block in turn. This enables us to investigate the effect of adjusting for later blocks, once earlier blocks have been taken into account. Before inclusion in the final multivariable model, only the most

predictive variables are kept in each block (controlling for the primary predictor), enabling a more parsimonious final model, reduction in the number of redundant variables, reduction in multicollinearity, and better accuracy in prediction. With the primary predictor in place, we then adjust in sequence and cumulatively, for significant predictors from each block: socio-demographic, then psycho-social, and finally gambling factors, in the process removing variables that are no longer significant in the final multivariable logistic regression model. We show how this block-wise selection process uncovers interesting insights about the data.

Penny, Richard

Producing Consistent Time Series in a Changing Data Environment

StatsNZ, Christchurch

StatsNZ is responsible for producing many of the key time series used a range of users for different purposes. Throughout StatsNZ's history it has had to evolve and adapt to a changing data environment, while maintaining consistent time series. With the increasingly dynamic data environment it is using more tools to inform its decisions on balancing the current relevancy of time series measurements with long-term consistency. Several examples will be covered.

Philpott, Andy

100% renewable electricity generation

Andy Philpott[1], Anthony Downward[1], and Michael Ferris^2

1. Department of Engineering Science, University of Auckland

2. Department of Computer Science and Wisconsin Institute of Discovery, University of Madison, Wisconsin

Using data from New Zealand, we discuss models of the national electricity system that are aimed at delivering 100% renewable electricity. Wind and solar energy are intermittent, and run-of-river hydroelectricity plants convert uncertain inflows into uncertain levels of energy, so the planning problem must account for these uncertainties. To compensate, hydroelectric reservoirs provide a means of storing energy, and geothermal power provides a predictable base-load energy source. Our model maximizes risk adjusted social welfare. This is compared with a competitive equilibrium with risk-averse investors.

Piper, Mark

Analytics at Fonterra for Research and Development

Research & Development Group, Fonterra, Palmerston North

I will talk about my role inside Fonterra as the head of our Research and Development function, how Analytics is used inside Fonterra to assist in the development of new products, improvement of existing formulations and to improve the food quality standards too.

This talk is part of the Analytics Forum to be held on Thursday 29 November in AH1.

Pirikahu, Sarah

Bayesian networks as a real-time prediction tool for swimmability

Sarah Pirikahu[1], Beverley Horn[1], and Roger Hodson[2]

1. Institute of Environmental Science and Research, Christchurch

2. Environment Southland, Invercargill

New Zealand government have set a national target of making 90% of New Zealand's large rivers and lakes swimmable by 2040. In order to achieve this goal, it is important to investigate the interaction of environmental and microbiological factors over the entirety of a river catchment. Bayesian networks can act as a tool for analysing the interactions between variables in a complex system, whilst providing a framework that can be easily understood by a non-specialist audience. In this talk we share our experiences of developing a Bayesian network to describe the swimmability of the Aparima River at Thornbury, Southland, and discuss the potential for Bayesian networks as a predictive tool for swimmability.

 This is a lightning talk.

Prendergast, Jesse

Simulating First- and Last- Mile Transport in Auckland

Jesse Prendergast, Andrea Raith, and Andrew Mason

Department of Engineering Science, University of Auckland

Using public transport is an excellent way for commuters to avoid congestion, keep costs down and reduce their carbon footprint. However, public transport can often be

underutilised when bus stops and train stations are too far away for commuters to comfortably walk, or there is no parking nearby. Auckland is no exception to this. Ferries in Devonport and Half Moon Bay, and the northern busway (a dedicated bus lane running from the North Shore into the central city) are all struggling from lack of parking or inaccessibility. Ride hailing companies such as Uber, Lyft and Zoomy compound this problem by taking commuters off buses as they catch door-to-door transport, which places more cars on the road. Even when the main portion of commuters' journeys are served by public transport, commuters have to get from their origin to the start of the public transport journey, and from its end to their destination. When the trip between the start and end of the public transport leg of the journey and origin or destination of the trip is completed by a shuttle-like service, this is referred to as First- and Last- Mile Transport. In this project we are building a simulation framework to test different First- and Last- Mile strategies in Auckland to investigate how a system might be deployed. Transport requests with ride time constraints are randomly generated within the simulated region, and capacitated vehicles are dispatched to pick up and deliver these requests. Vehicle allocation and routing is performed using online algorithms. This simulation framework can be used for testing the impacts of demand, pricing and scheduling strategies on the robustness and efficiency of the system. Commuters waiting time, comfort levels, and ride time can be determined for strategic decision making.

This presentation is eligible for the ORSNZ Young Practitioners Prize.

Raith, Andrea

Understanding the impact of incentives for electric vehicles with an integrated mode choice and traffic assignment model

Bob Grün and Andrea Raith

Department of Engineering Science, University of Auckland

The drastic increase in the emissions of carbon dioxide is one of the major contributors to global warming. A large amount of the overall carbon dioxide emissions is caused by the transport sector. Depending on their speed, petrol-based vehicles produce more than 50 times more carbon dioxide than electric vehicles. The slower the speed is, the larger is this difference. Since daily commuting within larger cities often means travelling in congested road networks, an increase of electric vehicles could contribute to a reduction of carbon dioxide. For this reason, we consider different approaches to increase the general uptake of electric vehicles.

We present an integrated mode choice and traffic assignment approach for petrol based and electric vehicles and comment on existence and uniqueness of its solution. In two case studies, one for a motorway section and one for central Auckland (New Zealand), we

consider different incentives for electric vehicle uptake and compare the resulting traffic flow as well as overall travel time and emissions. Since the resulting number of electric vehicles depends on the cost of the car and the users' value of time, we conclude the case studies with a sensitivity analysis on the input data.

Raubenheimer, Lizanne

BFF estimation methods for the difference between two binomial proportions

School of Mathematical and Statistical Sciences, North-West University, South Africa

Bayesian, frequentist and fiducial (BFF) methods for estimating the difference between two binomial proportions will be considered. Three non-informative priors will be used, the Jeffreys prior, a divergence prior and the probability matching prior. A probability matching prior is a prior distribution under which the posterior probabilities of certain regions coincide with their coverage probabilities. Fiducial inference can be viewed as a procedure that obtains a measure on a parameter space while assuming less than what Bayesian inference does, i.e. no prior. Fisher introduced the idea of fiducial probability and fiducial inference. In some cases the fiducial distribution is equivalent to the Jeffreys posterior. The performance of the Jeffreys prior, divergence prior and the probability matching prior will be compared to a fiducial method and other classical methods for constructing confidence intervals for the difference between two independent binomial parameters.

Read, Grant

Optimisation Applications at EPECentre

EGR Consulting Ltd, Christchurch, and the Electric Power Engineering Centre, University of Canterbury

EPECentre (the Electric Power Engineering Centre) was founded in 2002 with the aim of being New Zealand's centre of excellence for electric power engineering. Since then it had developed a number of successful research programmes, including the "Green Grid" initiative which has involved integrating modelling of future trends in renewable electricity generation and household demand, together with in depth knowledge of electricity networks and power management, to ensure that New Zealanders have access to reliable, safe, and affordable renewable energy. But modelling the performance of electricity sector systems involves applying optimisation techniques at various levels. This presentation summarises some past and present applications at EPECentre, and suggests some for the future.

Shahzadi, Amina

Estimating record completeness using a 2-state hidden Markov model

Amina Shahzadi[1], Ting Wang[1], Matthew Parry[1], and Mark Bebbington[2]

1. Department of Mathematics and Statistics, University of Otago, Dunedin

2. Institute of Fundamental Sciences, Massey University, Palmerston North

Natural phenomena such as earthquakes and volcanic eruptions can cause catastrophic damage. Such phenomena can be modelled using point processes. A major problem related to point process modelling of the earthquake and volcanic eruption data, however, is an underestimation of hazard caused by missing data. In particular, the number of missing observations between each pair of consecutively observed events in such type of records is unknown. Modelling volcanic eruptions as a renewal process, we develop a hidden Markov model (HMM) to tackle the problem with missing data in volcanic eruption records. In this model, the states of the hidden process correspond to whether or not there is one or more missing event(s) between each pair of consecutively observed events. We, therefore, propose a two-state model, where state 1 represents no missing observation, and state 2 represents that there is a variable, unknown, number of missing observations between a pair of consecutively observed events in the observed record. We model the number of missing observations in state 2 as a Poisson variable. The inter-arrival times of the observed process are assumed to have a gamma distribution in state 1, and a compound Poisson-gamma distribution in state 2. We apply this model to a global volcanic eruption record and demonstrate how we estimate the completeness of the record and the future hazard rate.

This presentation is eligible for the NZSA Student Prize.

Shanker, Liam

The Impact of Climate Change on the New Zealand Electricity Market

Liam Shanker[1], Golbon Zakeri[1], Geoffrey Pritchard[1], and Hannah Andrews[1,2]

Department of Engineering Science, University of Auckland

Meridian Energy Ltd., Auckland

The New Zealand electricity sector, dominated by hydroelectric generation, is arguably highly vulnerable to climate change. While the current generation resources seems adequate for maintaining security of supply, in the face of climate change, the inflow patterns may change drastically and we need to reassess the generation resource adequacy.

With current use of historically observed inflows, it is difficult to model how this system will behave under a varying climate. This research establishes a novel process for modelling historical inflows using quantile regression models. Our models can be utilized to generate sample paths of inflows, adjusted to represent various climate scenarios. Meridian's hydrothermal power system model, LPCon, is employed to present forward-looking commentary about the electricity sector under different climate scenarios.

We find that by 2035, at a national scale, the average price and generation should not be much affected, but the seasonality within these parameters will likely change. Within this near future horizon, we report that a reduction in annual thermal generation may be offset by an average increase in hydroelectric generation along the Waikato river hydro scheme. However, by 2100, we anticipate a degree of stress in the market due to an expected decrease in water availability. In the worst case, we see that thermal generation will increase to ensure the security of supply, and higher prices may be observed.

 This presentation is eligible for the ORSNZ Young Practitioners Prize.

Sibanda, Nokuthaba

Randomisation method for cluster randomised trials with unbalanced cluster sizes

Nokuthaba Sibanda[1], Francesca Storey[2], Fiona Cram[3], Stacie Geller[4], Bev Lawton[2]

1. School of Mathematics and Statistics, Victoria University of Wellington

2. Centre for Women's Health Research, Victoria University of Wellington

3. Katoa Ltd, Auckland

4. Department of Medicine, University of Illinois

Cluster-randomized clinical trials (CRT) are trials in which the unit of randomization is not a participant but a group or cluster (e.g. a hospital or a GP practice). They are suitable when the intervention applies naturally to the cluster (e.g. practice change); when lack of independence among participants may occur or when it is most ethical to apply an intervention to all within a group (e.g. school-level immunization). Because participants in the same cluster receive the same intervention, CRT may approximate clinical practice, and may produce generalizable findings. However, when not properly designed or interpreted, CRT may induce biased results.

CRT designs have features that add complexity to statistical estimation and inference. Key among these is the cluster-level correlation in response measurements induced by the randomization. An important consideration is the experimental unit of inference; often it is desirable to consider intervention effects at the level of the individual rather than the

cluster. Finally, given that the number of clusters available may be limited and these clusters may differ in size, simple forms of randomization may not achieve balance between intervention and control arms at either the cluster- or participant-level. Failure to account for imbalance may induce bias and reduce validity. This presentation focuses on the complexities of randomization in the design of CRTs, such as the imbalances in size and covariate factors across clusters.

Smith, Haddon

Determining effective placement strategies for denitrifying bioreactors in New Zealand river systems

Haddon Smith[1], Thomas Graham-Murdoch[1], Richard Clarke[1], Naresh Singhal[2], Alys Clarke[3], Vinod Suresh[3], Andrew Mason[1], and Stephen Waite [1]

1. Department of Engineering Science, University of Auckland

2. Department of Civil and Environmental Engineering, University of Auckland

3. Auckland Bioengineering Institute, University of Auckland

The state of New Zealand's waterways is a subject of growing concern, due to decreasing trends in a range of water quality measures. In particular, the nitrogen species nitrate shows increasing concentrations at monitored sites within agricultural regions. This has prompted research into how nitrate can be removed from waterways, with the use of denitrifying bioreactors being one promising method. Aside from the design of such bioreactors, the problem of determining the most effective spatial placement within river networks is one research avenue that can help maximize the impact of bioreactor installations. The aim of this project was to use optimization techniques alongside water quality modelling theory to find effective placements solutions. A number of methods were developed so the advantages and disadvantage of different approaches could be evaluated. A mixed integer program (MIP), solved using commercial solver Gurobi, was shown to be effective. A greedy sequential heuristic algorithm was also implemented with successful results. A dynamic programming method was used, allowing the use of non-linear objective functions, however it was shown to be limited in its ability to scale to large networks. The placement strategies resulting from optimisation are shown to give additional benefit to water quality compared to random or intuitively generated solutions, and give a number of insights into how bioreactors should be placed. Firstly, placements in high concentration rivers are more effective that placements in lower concentrations areas. Secondly, placing units high up in river networks is shown to be advantageous to overall river health. Thirdly, the performance of the greedy sequential heuristic indicates that there is a low level of dependence between the effectiveness placement locations. Fourthly, a small number of large installations is shown to be more effective than a large number of small installations in some cases. Following on from this, the robustness of solutions to seasonal variation and variations in bioreactor extraction rates was studied. Furthermore, a number

of insights were gained into what characteristics designers of bioreactors should try to achieve, such as the extraction rates necessary for different levels of impact on water quality. Lastly, the value of using optimization in choosing the placement of treatment facilities in rivers systems is demonstrated, through the use of the MIP model to find placement strategies designed to meet water quality improvement aims published by the Waikato River Authority.

This presentation is eligible for the ORSNZ Young Practitioners Prize.

Stewart, Russell

Optimising makespan and energy consumption in task scheduling for parallel systems

Russell Stewart[1], Andrea Raith[1], and Oliver Sinnen[2]

1. Department of Engineering Science, University of Auckland

2. Department of Electrical and Computer Engineering, University of Auckland

In parallel computing multiple processors are employed to execute a program faster than it could be executed by a single processor. To achieve high efficiency, the scheduling of the tasks of an application onto the processors of the parallel system is crucial. Such a task schedule determines both the allocation of tasks to one of the processors, and the order in which they are executed. This task scheduling problem is a challenging optimisation problem (strongly NP-hard), even for the case where there is only one objective, e.g. to minimise the total execution time, also known as makespan. Today task scheduling is often a constrained optimisation problem, where the makespan needs to be minimised, while keeping energy consumption below a threshold. Here, we consider a multi-objective version of this scheduling problem that aims to minimise both makespan and energy consumption. Based on a model of processor energy consumption, a multi-objective integer linear programming problem is formulated. Different approaches to modelling energy consumption of processor idle times before, during, and after task execution, are described. The task scheduling problems can be solved using multi-objective optimisation methods based on weighted sum or ε-constraint methods. Computational performance and characteristics of Pareto sets of solutions of this multi-objective problem are discussed.

This presentation is eligible for the ORSNZ Young Practitioners Prize.

Tabassum, Tanjila

Causal Relationship of Economic Factors with GDP Growth of Two Emerging Markets of South Asia: Bangladesh and India

School of Economics and Finance, Victoria University of Wellington

Substantial change in world economy in last few decades has made emerging markets more influential than ever before. This study tries to investigate causal relationship of three economic factors (GCF, FDI, and trade) with GDP growth of two emerging markets of South Asia: Bangladesh and India. External factors like globalization (KOF index) and financial crisis (VIX index) have also been incorporated as and where deemed necessary. Time series data (1977-2015) of both countries taken from sources like World Development Indicator have been analysed. Johansen Co-integration method indicates that there exist long run equilibrium relationship among stationary form of the concerned variables for both countries. The results of Granger Causality test suggest that (1) Trade and GDP growth granger cause each other, (2) GCF and KOF granger cause GDP growth and (3) FDI for India has bidirectional relationship with GDP growth.

 This presentation is eligible for the NZSA Student Prize.

Thomas, Glenn

Leveraging the statistical problem to solve the business problem: A case study from Watercare.

Harmonic Analytics, Auckland

Oftentimes, the motivation for a statistical problem stems from a business problem. However the statistical problem of modelling the data is often much more interesting. As a data scientist, the challenge is to bring the results back around to deliver business value. In this session, I give some examples of bridging the gap in the context of understanding wastewater overflows and smart meter data for Watercare, the water supplier for Auckland. Some questions that the data scientist faces are:

- What data and analyses should I be using?
- How do I manage expectations of the business?
- How do I communicate the findings so that they will have an impact on the business?

We discuss these and more, and some of the challenges and learnings of thinking statistically in a business environment.

Thomasen, Lisa

The Bedside Manner of a Statistician

Fonterra, Palmerston North

"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of." - Ronald Fisher (1938)

At med school, doctors learn the importance of an empathetic bedside manner. Bedside manner is an overlooked skill in a statistician's toolbox which leaves us to learn the hard way how to tell a scientist that their experiment is dead.

I will outline some of the soft skills of statistical consulting including "statistical first aid" and the "stats defibrillator" which can be used to revive sickly datasets.

As in health care, prevention is always better than a cure. I will share some of my preventative measures and "stats herbal remedies" which can be employed to avoid that awkward moment of pronouncing the dataset dead.

Tidswell, James

Accounting for excess cold emissions in traffic assignment models with emission objectives

James Tidswell, Andrea Raith, and Anthony Downward

Department of Engineering Science, University of Auckland

There has been increased awareness of greenhouse gases and associated climate change, where the transport sector makes a significant contribution to greenhouse gases via vehicle emissions. This has provoked interest in estimating vehicle emissions through the use of traffic models, such as in the Vehicle Routing Problem and the Traffic Assignment problem. Emissions models used within these traffic models generally consider the 'hot' emissions, which are the emissions produced while the vehicle engine is at operating temperature, and are a function of average speed. In this presentation we consider accounting for the initial portion of a vehicle user's trip where the vehicle engine starts at below operating temperature, producing excess 'cold' emissions, which are generally a function of trip length. We examine two methods of modelling cold emissions and incorporate them into a Traffic Assignment model, addressing the path-dependent shortest path problem that arises with path-length-dependent emissions costs on arcs, and propose an algorithm to identify the shortest path in a network with path-dependent arc costs. Incorporating this algorithm into the traffic assignment model allows the evaluation of solutions to the traffic assignment problem with path-dependent objectives.

This presentation is eligible for the ORSNZ Young Practitioners Prize.

Tran, Kien

Fuzzy Clustering of Nominal Data using Finite Mixtures

School of Mathematics and Statistics, Victoria University of Wellington

Clustering techniques are often performed to reduce the dimension of very large datasets, whose direct analysis using techniques such as regression can be computationally infeasible.

The clustering of non-independent nominal variables in particular poses distinct difficulties due to its lack of a distance metric and a measure of correlation. This prevents the use of many techniques such as k-means (Lloyd 1993), hierarchial clustering (see, eg. Hastie et al., 2009), Copula-based methods ((Nelsen, 1999)) or some variations of Principal Component Analysis (see, eg. Chavent et al., 2011).

This paper proposes a clustering approach for this data type based on finite mixtures (McLachlan and Peel, 2000) , pairwise composite likelihood (review by Varin et al., 2011) and transformation of nominal levels. If applicable, this would provide a parsimonious, likelihood-based fuzzy clustering model suitable for statistical inference; as well as the potential for extension to more general mixed type data.

This presentation is eligible for the NZSA Student Prize.

Tularam, Gurudeo Anand

Traditional vs Non-traditional Teaching and Learning Strategies - the case of E-learning

Environmental Futures Research Institute, Griffith University, Brisbane and Palmerston North Boys High School

Traditional teaching approaches are generally teacher-directed, where students are taught in a manner that is conducive to sitting and listening. It is often argued that these approaches may not provide students with valuable learning skills, and rather that non-traditional approaches to teaching and learning might better endow students with such skills. The teaching of mathematics that is usually referred to as “non-traditional” uses constructivist philosophies as its bases; by implicating strategies in which the individual develops, constructs or rediscovers knowledge in an attempt to make sense of his or her universe, or by employing social constructivist ideals which suggest that group work, language and discourse are vital for learning. Also, it is often argued that non-traditional teaching is done using a problem solving approach, where the learner is the problem

solver. Non-traditional teaching and learning (NTTL) in mathematics and engineering needs to be well understood so that appropriate and meaningful comparisons with traditional techniques can be made. The computer based teaching technology known as "e-learning" is now often used in tertiary level mathematics and engineering teaching. The e-learning methodology is considered to be more in-line with the non-traditional than the traditional teaching approaches. This paper critically reviews the literature that makes comparisons of the traditional and non-traditional teaching approaches in mathematics and engineering education, and specifically examines the advantages and disadvantages of these approaches along with the manner in which they influence the academic performance of students in mathematics and engineering courses.

Turner, Rolf

Hidden Markov models, probability tables and the Levenberg-Marquardt algorithm.

Department of Statistics, University of Auckland

The underlying problem is one of fitting hidden Markov models to sequences of discrete valued observations for which there is no appropriate parametric distribution. The distributions have to be specified non-parametrically via tables (one table for each state of the underlying Markov chain). I have tried fitting the models in the "usual" way via the EM algorithm, and also tried a "brute force" method whereby the likelihood is maximised using the nlm() function. Both approaches proved unsatisfactory, so I implemented a Levenberg-Marquardt algorithm in this context. I had previously used the Levenberg-Marquardt algorithm to fit hidden Markov models in settings in which the observations are Poisson distributed. Conceptually it is straightforward to adapt the procedure to the non-parametric context. In this talk I will describe some of the challenges involved in working out the details. Once these details were dealt with, the LM algorithm turned out to be fast, effective and reliable. I will briefly present some of the results of a data analysis that makes use of this algorithm.

Welsh, Alan

Using the Bootstrap in Generalized Regression Estimation

James G. Booth[1] and Alan H Welsh[2]

1. Department of Statistical Science, Cornell University, city/state?

2. Mathematical Sciences Institute, The Australian National University, Canberra

Regression adjustments are widely used for adjusting the sample mean of a variable of interest to account for deviations of the means of related auxiliary variables from their

known population values. The adjustments produce estimators with variances smaller than that of the original sample mean. The method has a long history in the survey literature, and is closely related to covariance analysis in designed experiments and the control variates method used for variance reduction in Monte Carlo studies.

We discuss a generalized regression estimation procedure that can lead to much improved estimators of general population characteristics, such as quantiles, variances, and coefficients of variation. The key idea behind the approach is that the regression of an estimator of a target population parameter on the estimator of a (possibly vector-valued) auxiliary parameter is computed directly using the bootstrap. The method is quite general and requires minimal assumptions, the main ones being that the asymptotic joint distribution of the target and auxiliary parameter estimators is multivariate normal, and that the population values of the auxiliary parameters are known. The assumption on the asymptotic joint distribution implies that the relationship between the estimated target and the estimated auxiliary parameters is approximately linear with coefficients determined by their asymptotic covariance matrix. Use of the bootstrap to estimate these coefficients avoids the need for parametric distributional assumptions. First-order correct conditional confidence intervals based on asymptotic normality can be improved upon using quantiles of a conditional double bootstrap approximation to the distribution of the studentized target parameter estimate.

Biography

Alan Welsh holds the E.J. Hannan Professorship in Statistics at the Australian national University. He has contributed to statistical methodology and theory, including robustness and model selection in linear models, robustness and bootstrapping in linear mixed models, semiparametric estimation, modelling zero-inflated data, modelling compositional data and the analysis of data from sample surveys. He has written a major monograph on Aspects of Statistical Inference (Welsh 1996) and jointly authored another on Maximum Likelihood Estimation for Sample Survey Data (Chambers et al 2012). He was the Editor-in-Chief of the Australian and New Zealand Journal of Statistics from 2012-2015. Alan was elected a Fellow of the Australian Academy of Science in 2007, Fellow of the Institute of Mathematical Statistics in 1990, and Fellow of the American Statistical Association in 2002. He was awarded the Moran Medal of the Academy of Science for outstanding research in statistics by a person under the age of 40 in 1990, the Pitman Medal of the Statistical Society of Australia in 2012 and the E.A. (Alf) Cornish Award of the Australasian Region of the International Biometric Society in 2017.

Session details


This plenary address will be delivered in AH1 on Thursday 29 November at 2:00.

Wen, Zhijian

Bayesian modelling for glass refractive index

Wen Zhijian[1], James Curran[1], Sallyann Harbison[2], and Douglas Elliot[2]

1. Department of Statistics, The University of Auckland

2. Institute of Environmental Science and Research Ltd, Auckland

Glass is a common type of evidence in forensic science. Broken glass recovered from a particular suspect may have simliar physical common characteristics with glass collected at a crime scene and therefore can be used as evidence. Statistical treatment of this evidence involves computing a measure of the weight of evidence. This may be done in Bayesian framework that incorporates information from the circumstances of the crime. One of the most crucial quantities in this calculation is the assessment of the relative rarity the characteristics of the glass - essentially the probability distribution used to model the physical characteristics of recovered glass. Typical characteristics used in case work are the chemical profile of glass and the refractive index measurement. There is a considerable body of scientific literature devoted to the modelling of this information. Aitken and Lucy (2004) described a multivariate kernel density estimation to analyse the chemical profile of glass. Lucy and Zadora (2011) used kernel density estimation to model the change of the refractive index before and after annealing procedures. Our aim is to construct a Bayesian semi-parametric model, the Dirichlet Process Mixture Model, to model the glass refractive index measurement. The results of the analysis can be used to compute the density of such quantity given certain refractive index and extra information on the source glass.

This presentation is eligible for the NZSA Student Prize.

Xiao, Jade

Towards Faster Multi-Objective Shortest Path Queries

Jiantao Shen, Jade Xiao, and Andrea Raith

Department of Engineering Science, University of Auckland

Optimisation problems frequently have multiple conflicting objectives and no single solution. In the context of shortest paths, one might wish to simultaneously minimise travel time, fuel consumption, and the number of transits on a transportation network. Existing algorithms for the multi-objective shortest path problem are not fast enough, or else sacrifice correctness for speed. This motivates extending powerful speed-up techniques for the standard shortest path problem into the multi-objective domain. We present our adaptations of the algorithms based on landmarks, reach, arc flags, and contraction hierarchies, followed by the results of computational experiments. This research

contributes to our understanding of the effectiveness of existing speed-up techniques in the multi-objective context.

 This presentation is eligible for the ORSNZ Young Practitioners Prize.

Xiao, Xun

A New Closed Form Estimator for Beta Distribution

Institute of Fundamental Sciences, Massey University, Palmerston North

In this study, we propose a new closed form estimator for two-parameter beta distribution. It is derived from the score equation of generalized beta distribution. The asymptotical properties of this estimator is investigated. It surprisingly resembles MLE in terms of mean square error and we further explore the linkage between the new estimator and MLE. With its closed form, it is very easy to develop some improved estimators such as one step estimator and jackknife estimator.

Xu, Danli (Lois)

Area-proportional Visualisation for Circular Data

Danli (Lois) Xu

Department of Statistics, University of Auckland.

Data visualisation is important for statistical analysis, as it helps to communicate information efficiently and understand the significance of data in a visual context. It is particularly helpful to display circular data in a 2-dimensional space due to its non-linear support space and special characteristics, as they are able to reveal the underlying circular structure which is otherwise not obvious in 1-dimension.

In this talk, we will first formally categorise circular plots into two types, height- and area-proportional ones, and then describe a new general methodology to visualise circular data area-proportionally in a 2-dimensional space. Formulae are given that are fairly simple yet effective to produce desired circular histograms, rose diagrams and dot plots. More importantly, a density curve can now be produced and superimposed to match either the height- or area-proportional display of the raw data. This allows a visual assessment of the goodness of fit of selected models and enables a quick comparison among various models. In addition, plots are also developed for circular data with multiple classes.

 This presentation is eligible for the NZSA Student Prize.

Yang, Jean

Extracting the most out of your biomedical data

Charles Perkins Centre, The University of Sydney

With the advancement of many high-throughput biotechnologies, a research interest of many scientists has been to utilize multiple data sources to gain further insights into biology and deeper understanding of complex diseases. Data integration not only enables scientists to address and ask very specific questions, it also allows them to explore and understand the complex relationships among different phenotypes. In this talk, I will discuss how recent statistical visualization and machine learning approaches are used to address different questions in bioinformatics research. For example, the modeling of heterogeneity in multi-omics data together with extracting different types of features can help to improve the prognosis of disease outcome. Finally, I will discuss the impact of such research on our statistics and data science curriculum.

Biography

Professor Jean Yang is an applied statistician with expertise in statistical bioinformatics. She was awarded the 2015 Moran Medal in statistics from the Australian Academy of Science in recognition of her work on developing methods for molecular data arising in cutting edge biomedical research. Her research stands at the interface between medicine and methodology development and has centered on the development of methods and the application of statistics to problems in -omics and biomedical research. In particular, her focus is on developing methods for integrating omics and clinical data to answer a variety of scientific questions. As a statistician who works in the bioinformatics area, she enjoys research in a collaborative environment, working closely with scientific investigators from diverse backgrounds.

Session details

This plenary address will be delivered in AH1 on Tuesday 27 November at 2:00.

Yee, Thomas

Generally Altered, Inflated and Truncated Poisson and Binomial Distributions

Thomas Yee[1] and Chenchen Ma[2]

1. Department of Statistics, University of Auckland

2. Department of Mathematics, Zhejiang University, China

The zero-altered (hurdle), zero-inflated and zero-truncated distributions are very well-known to statistical practitioners for handling surplus and absences of 0s in responses. This talk sketches early work on generalizing these to other values of the support so that 0 is not necessarily the only special value. We focus on the Poisson and binomial models, however the work is easily extended to other count distributions. Software implementing some of the methods will also be described; they include several new family functions in the VGAM R package, as well as d- and p- and q-type functions. Some applications will also be described. This is also work with Theodora Jin.

Ziedins, Ilze

Improving health care delivery: from theory to practice

Department of Statistics, University of Auckland

New Zealand has an excellent public health system, operating within a very constrained funding environment.
Providing timely health care is crucial, yet patients often experience delays in receiving treatment. Are there ways in which these delays could be reduced? Could the existing capacity be used in different ways? I will discuss some of the challenges in answering these questions, and give examples of both theoretical and simulation based approaches. These problems lie at the nexus of statistics, applied probability, and operations research.

Biography

Ilze Ziedins works on modelling and analysis of stochastic networks, with applications to health care, transportation, and communications networks. She was an undergraduate at Waikato, then a PhD student and junior research fellow at Cambridge, and a lecturer at Heriot-Watt University in Edinburgh, before joining the University of Auckland. Her recent work has concentrated on models of networks with selfish routing, and modelling and optimization of health care delivery.

Session details

This plenary address will be delivered in AH1 on Wednesday 28 November at 9:00.

Aitken, Colin GG, and David Lucy. 2004. "Evaluation of Trace Evidence in the Form of Multivariate Data." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 53 (1). Wiley Online Library:109–22.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1):5–32. https://doi.org/10.1023/A:1010933404324.

Li, F., C.M. Triggs, B. Dumitrescu, and C.D. Giurcăneanu. 2018. "The Matching Pursuit Algorithm Revisited: A Variant for Big Data and New Stopping Rules." *Signal Processing* (in press).

Lucy, David, and Grzegorz Zadora. 2011. "Mixed Effects Modelling for Glass Category Estimation from Glass Refractive Indices." *Forensic Science International* 212 (1-3). Elsevier:189–97.