



2013 Joint NZSA + ORSNZ Conference

24 – 27 November 2013

Hamilton, New Zealand



Analytics for a Changing World: From Data to Decisions

Joint Conference of the NZ Statistical Association and Operations Research Society of NZ

ISSN 0114-6602

Conference Welcome

On behalf of the organising committee I would like to welcome you to this joint conference, comprising the 64th annual meeting of the New Zealand Statistical Association and the 47th annual meeting of the Operations Research Society of New Zealand. The University of Waikato is proud to be the host of this illustrious event, particularly in 2013 which has been deemed the International Year of Statistics (<http://www.statistics2013.org>) and celebrates Mathematics of Planet Earth (<http://mpe2013.org/>).

We hope you find the academic programme insightful, illuminating and inspiring. We also hope you enjoy the social programme, kicking off with a reception on the night of Sunday the 24th of November, carrying through with a poster session with canapés followed by the Young Statisticians and Operation Researchers event on the night of Monday the 25th of November, and concluding with a conference dinner on the night of Tuesday the 26th of November. While you are here, we hope you manage to sample some of the other attractions and entertainments that are to be found in Hamilton and the surrounding area.

I would like to extend thanks to all delegates who are attending the conference. To those who are presenting talks or posters, I would like to add additional thanks for their efforts filling the academic programme. In particular, I would like to thank our invited speakers, and the organisers of those invited sessions. We are extremely grateful to our plenary speakers for taking the time and trouble to be with us.

I would also like to thank all of the sponsors that have helped make this conference possible. First mention must go to our platinum sponsor SAS, who not only provided substantial financial support, but in addition, the conference bags, and pens and paper. Thanks are also due to Harmonic Analysis, Catalyst IT and Dragonfly Science, Statistics New Zealand, and the Australian and New Zealand Journal of Statistics for their generous support of prizes and sponsorship. We also thank the University of Waikato for their hosting role, and the Department of Statistics (particularly Lyn Hunt and Tania Robinson) for helping with the administration of the conference. Also thanks to the ORSNZ for hosting the conference website and dealing with financial transactions.

On a more personal level, I would like to extend my thanks to the organising committee. I had not realised how much work is required behind the scenes to get a conference such as this up and running at the start of this project, and without the committee members' generous contributions of time and energy, the conference would have been, at best, a shadow of itself. In particular, I would like to thank Paul Murrell and Cam Walker for untying the Gordian knot that is the academic programme, Andrew Mason for the Herculean jobs of setting up the conference website and managing conference registration, Vanessa Cave and Harold Henderson

for drumming up sponsorship and keeping the organisation project on track, James Curran for his decisiveness and attention to detail, Judi McWhirter for her role as treasurer, Maheswaran Rohan and Ilze Ziedins for their valuable insights, and Maryann Pirie and Kylie Maxwell for their exemplar of how to organise an event with the Young Statisticians and Operations Researchers evening. And of course thanks to all the committee members for the myriad of tasks they will have performed which went completely beyond my ken! Also a deep, heartfelt thanks to the volunteers who will be helping keep the conference running smoothly over the next four days.

And without further ado, on to the 2013 Joint NZSA and ORSNZ Conference!

Steven Miller

Joint Chair of the 2013 Organising Committee

Organisers

ORGANISING COMMITTEE

Vanessa Cave (Joint Chair; AgResearch)

Steven Miller (Joint Chair; University of Waikato)

James Curran (President, NZSA)

Harold Henderson (AgResearch)

Judi McWhirter (University of Waikato)

Paul Murrell (University of Auckland)

Maheswaran Rohan (Department of Conservation)

Andrew Mason (President, ORSNZ)

Cameron Walker (University of Auckland)

Ilze Ziedins (Vice President, ORSNZ)

Andy Philpott (University of Auckland)

Andrea Raith (Treasurer, ORSNZ)

Sponsors

NZSA 2013 gratefully acknowledges the financial support of the platinum sponsor, SAS.



We also gratefully acknowledge the support of



Conference organisers:



Contents

General Information	8
Venues	8
Parking	9
Internet Access	9
Toilet Facilities	10
Other Locations of Interest on Campus	10
Maps	11
Young Statisticians' and Operation Researchers' Event	14
Prizes	15
Conference Programme Overview	16
Conference Timetable	17
Plenary Presentation Abstracts	22
Presentation Abstracts	25

General Information

Venues ***Please refer to the map on page 12**

The main venue for the conference and events is **S-Block** (Map E7).

- **Welcome Reception, Sunday 24 Nov, starting at 5:30pm**
The reception on Sunday night will be held in the ground-floor foyer of S-Block.
- **Conference sessions** from Monday to Wednesday will be held in lecture rooms (S.1.01 – S.1.05) on the first floor of S-Block. The **plenary talks** will be held in S.1.04.
- The **speakers' preparation room** will be S.1.03.
- **Poster Evening, Monday 25 Nov, 4:40pm**
Posters will be displayed in the first-floor foyer of S-Block. **Poster talks** will be in S.1.04, followed by poster viewing with canapés and drinks in the foyer.
- **NZSA AGM and ORSNZ AGM, Tuesday 26 Nov, 4:40pm**
The NZSA AGM will be in S.1.05, and the ORSNZ AGM will be in S.1.01.
- All **lunches** and **tea-breaks** will be served in the first-floor foyer of S-Block.
- **Young Statisticians and Operation Researchers Evening, Monday 25 Nov, 7:30-9:30pm**
The event will be held at the McMeekan Centre on the grounds of the Ruakura Campus. Meet 7pm at the conference venue after the poster session to walk up as a group (a 20 min walk, with van or car pooling if required). A van will be available to drop you back at your accommodation after the event. Halal and vegetarian options will be catered for. Please see the event flyer (**on page 11**) for more information.
- **Conference Dinner, Tuesday 26 Nov, 6:30pm**
It will be held at the Gallagher Academy of the Performing Arts ("the Academy"; Map C6)

Parking *Please refer to the map on page 12

There are more than 1,500 generally available car parking spaces on the University Campus, in areas specifying “General Parking”. The main parking areas are accessed through:

- Gate 1, Knighton Road (Map C9): approximately 640 general parking spaces (please avoid parking in restricted areas or spaces).
- Gate 2B, Knighton Road (Map B5): a few hundred spaces that are all general parking.
- Gate 10, Silverdale Road (Map I6): approximately 800 general parking spaces; a bit of a walk from the conference venue.

There are many disability parks across the campus. The most useful in terms of the conference can be found through:

- Gate 8, Hillcrest Road (Map E7): enter the gate, then head straight, through the turning bay, and park near S-Block.
- Gate 1, Knighton Road (Map C7): enter the gate, turn left at the intersection, drive through the general parking area towards the Academy, and find the parks either near the loading bay for the Academy, or further on near the Student Union buildings.

Internet Access

There is wireless internet coverage over almost the entire campus. Access is available from:

- Eduroam: delegates from participating institutions of the Eduroam network (<http://reannz.co.nz/eduroam>) should be able to connect to the University of Waikato wireless network free of charge. Please see your local host institution’s instructions for setting up your connection.
- Lightwire: delegates who set up a Lightwire account at www.lightwire.co.nz will be able to purchase a prepaid subscription that allows access to the University of Waikato wireless network. See <https://prepay.lightwire.co.nz/plans/> for pricing plans.

At the time of printing, the conference organisers were investigating the possibility of guest accounts for conference delegates on the university computer network. Updated information on this possibility will be available at the conference.

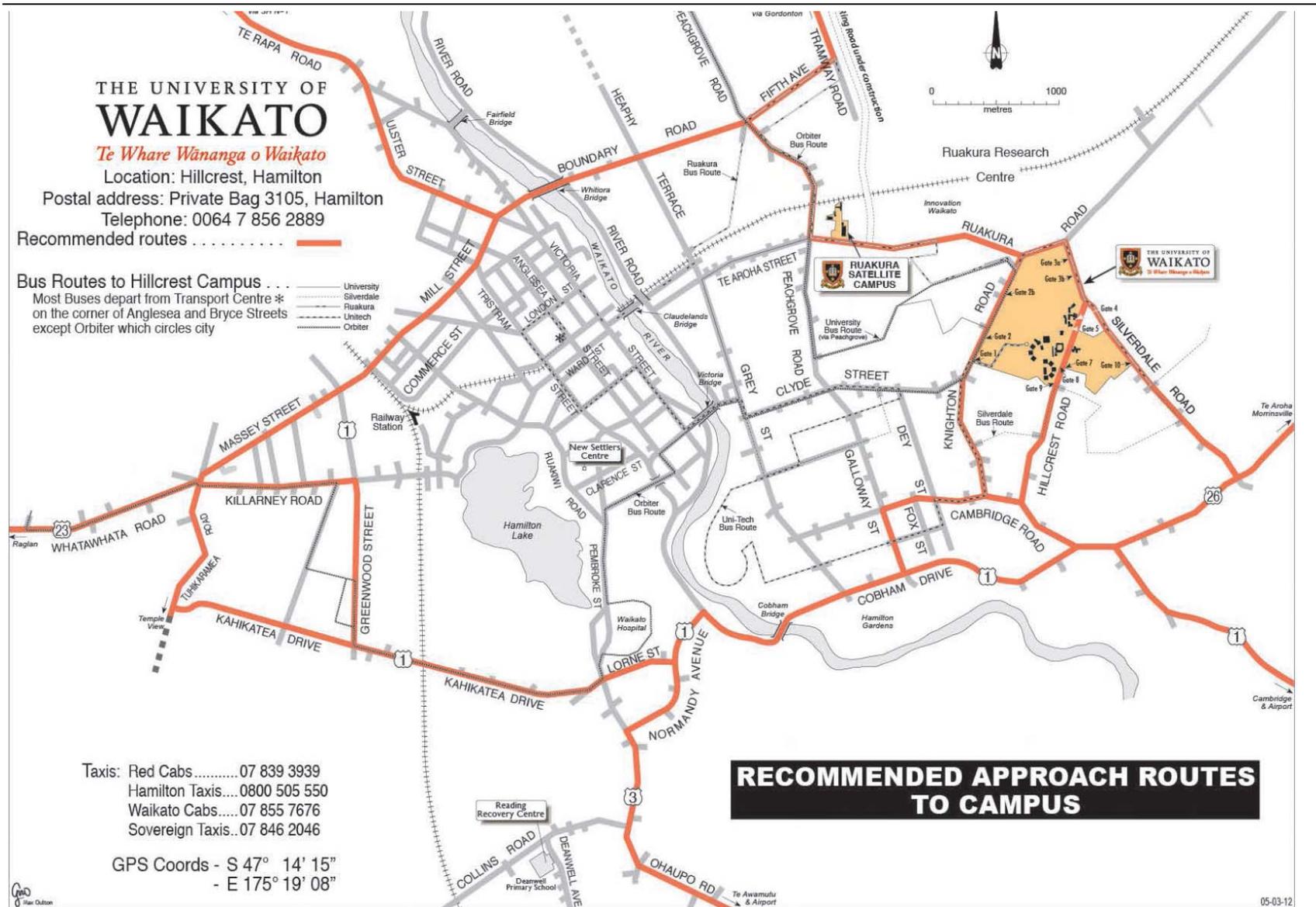
Toilet Facilities

- Toilets are located on the ground floor of S-Block. Go down the main stairs and you will find them on your left. There is also a lift, to the left of the main stairs.
- Alternatively, there are male and female toilets on each floor of G-Block, in the L-Block lecture theatres building, in the Student Centre and at the Academy, and many other places on campus.

Other Locations of Interest on Campus ***Please refer to the map on page 12**

- The Student Centre (Map E7) is the newest completed building on Campus. It incorporates the library, an information centre, Bennetts Bookshop, a café and a few shops.
- A shopping complex is located between S-Block and the Academy (Map D7), in the middle of the campus. There are two cafés, a pharmacy, convenience store and outlets for ASB and ANZ, and ATMs. The campus bus stop for services to the central bus station start and end outside here.
- The Oranga Foodcourt (Map D7) has a range of cafés.
- The Gallagher Academy of Performing Arts (Map C6) has rotating art exhibitions, and the Opus café.
- There are three lakes on campus: Knighton Lake (Map B6-C6), Oranga Lake (Map D6) and Chapel Lake (Map D8). All three have boardwalks that surround them.
- A walk around the Fernery (Map E8) can provide a pleasant respite from the bustle of the campus.

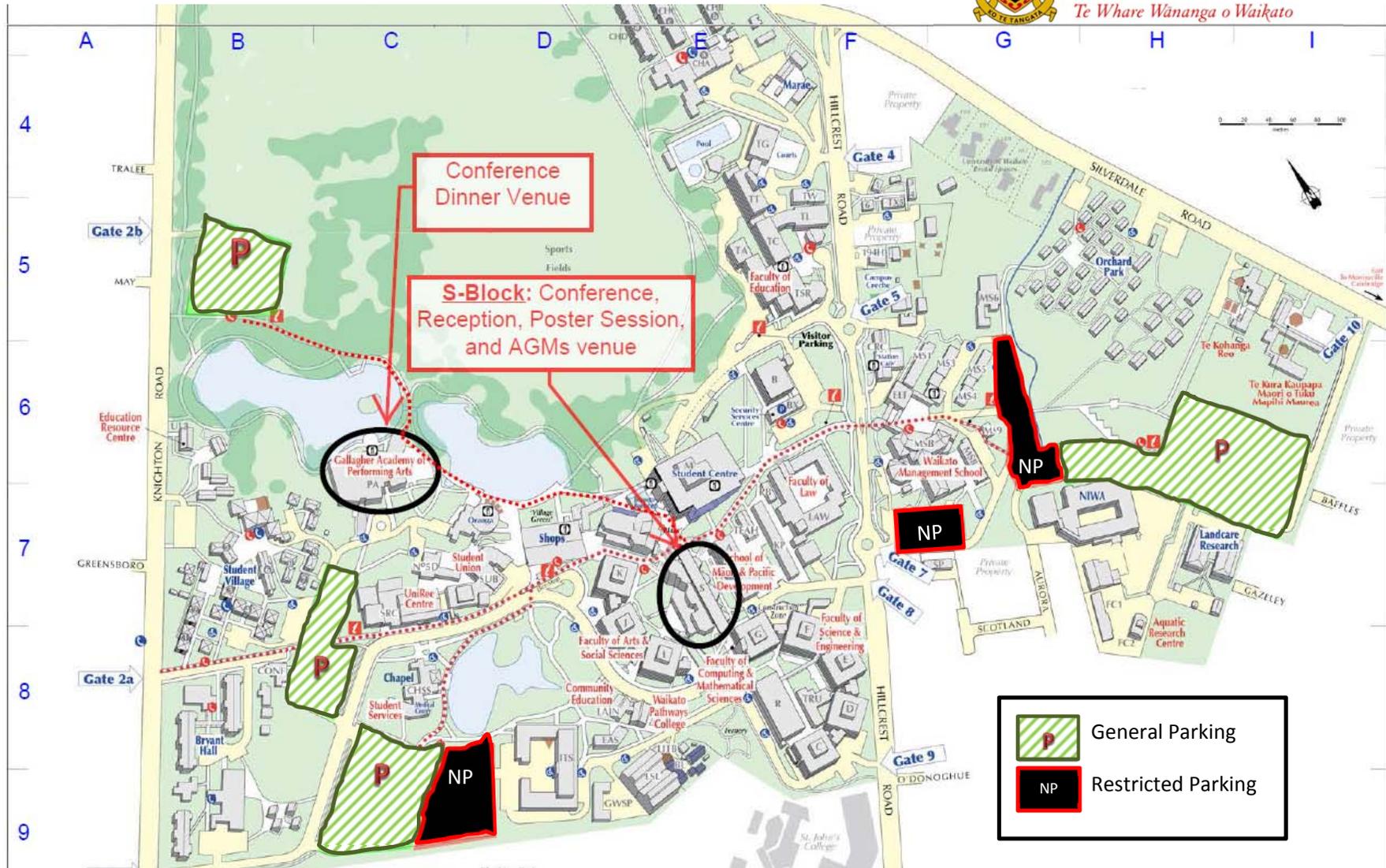
Maps



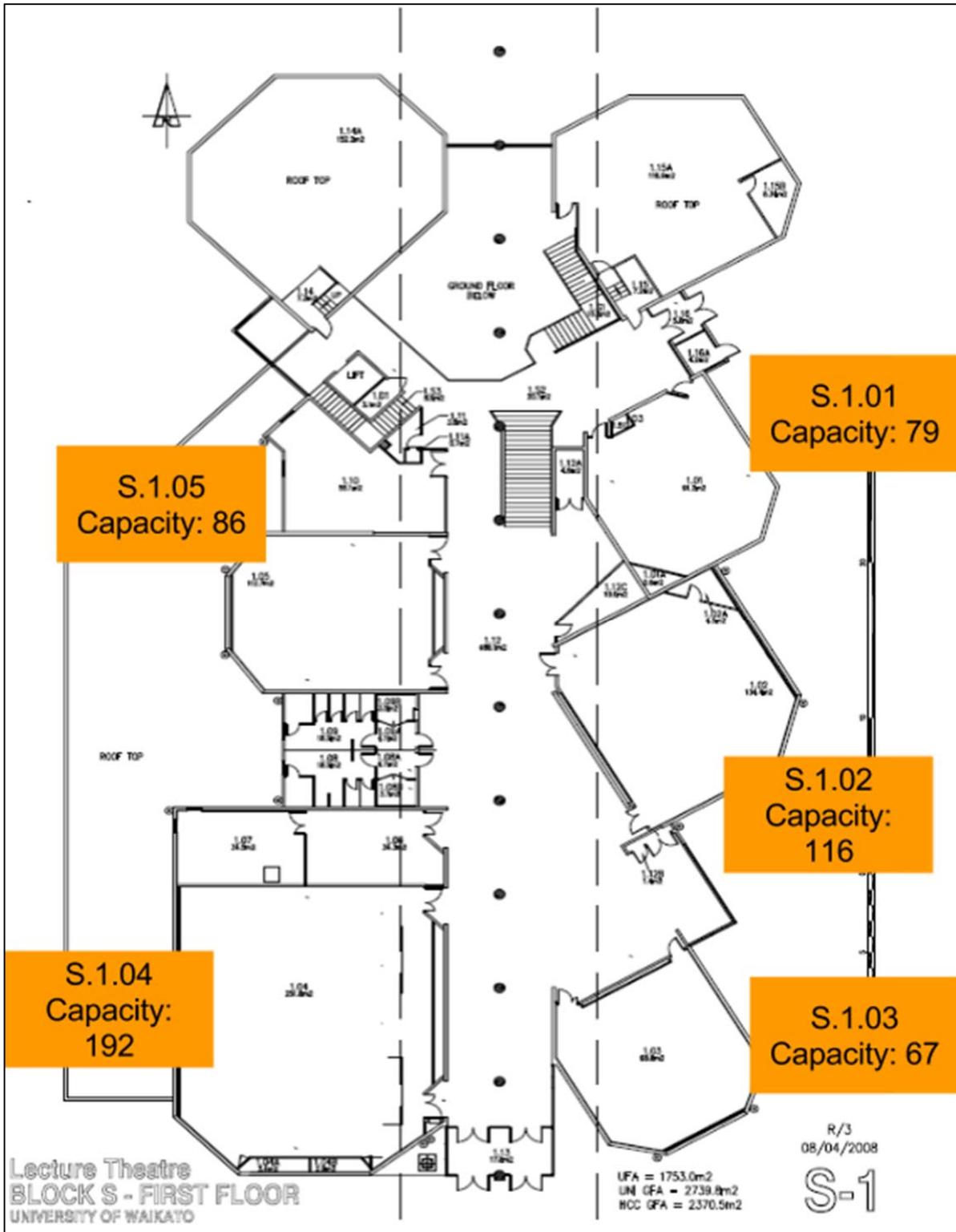
Parking and Locations



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato



Location of the Lecture Rooms



NZSA/ORSNZ conference

Young Statisticians' and Operation Researchers' Event

If you are a young statistician or young operations researcher come along for a BBQ and to network with like-minded individuals.

When: Monday 25th November, 7.30pm – 9.30pm.

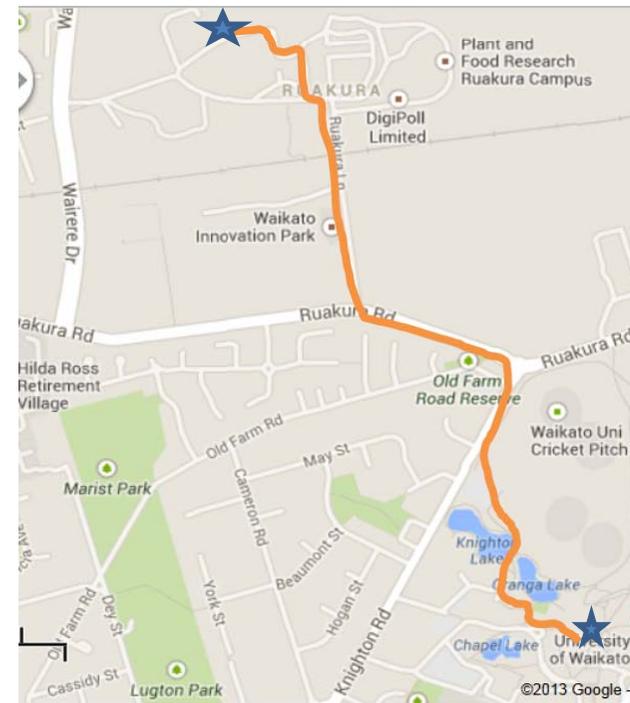
Where: McMeekan Centre, AgResearch Ruakura

Meet 7pm at the conference venue after the poster session to walk up as a group (a 20 min walk, with van or car pooling if required). A van will be available to drop you back at your accommodation after the event.

Halal and vegetarian options will be catered for.

Thank you Statistics NZ and NZSA for providing sponsorship for this event.

Contact Maryann Pirie or Kylie Maxwell if you have any further questions/concerns.



Prizes

Catalyst IT / Dragonfly Open Research Prize, \$500

To be eligible for the prize, each stage of the research process presented must have been carried out with open source tools. For example, data analysis could be performed with R, data visualisation with GGobi, and paper authoring with LaTeX.

The prize may not be awarded; may be awarded to a single person; or be divided at the judges' discretion between up to three entrants Catalyst IT <<http://catalyst.net.nz/>> are specialists in free and open source technologies. Catalyst's mission is to make open source the preferred technology choice of New Zealand. Dragonfly Science <<http://www.dragonfly.co.nz/>> carries out independent statistical modelling and analysis.



Harmonic Analytics Student Prizes

In 2013 Harmonic Analytics are sponsoring prizes for the best student talks at the NZSA conference.



Conference Programme Overview

	Monday 25 Nov	Tuesday 26 Nov	Wednesday 27 Nov
	OPEN		
09:00	Plenary	Plenary	Plenary
09:50	2 talks (4 parallel sessions)	2 talks (4 parallel sessions)	2 talks (4 parallel sessions)
10:30	MORNING TEA (20 minutes)		
10:50	4 talks (4 parallel sessions)	4 talks (4 parallel sessions)	4 talks (4 parallel sessions)
12:10	LUNCH (50 minutes)		
13:00	Plenary	Plenary	Plenary
13:50	Plenary	Plenary	CLOSE
14:20	AFTERNOON TEA (20 minutes)		
14:40	6 talks (4 parallel sessions)	6 talks (4 parallel sessions)	
16:40	Poster Session	NZSA AGM ORSNZ AGM (1 hour)	
17:40			
18:30			
19:30	Young Statisticians and Operation Researchers Evening	Conference Dinner	

Conference Timetable

SUNDAY, 24 November				
1730	Conference Registration Opens/Social Mixer S-Block Ground Floor Foyer			
Note: * denotes student speaker, + denotes the Chair of the session, # denotes entry for Open Research Prize				
MONDAY, 25 November				
0850	OPEN [S.1.04]			
0900	Plenary session [S.1.04] <i>The Need for Speed in the Path of the Deluge</i> Plenary Speaker: Chris Wild			
	[S.1.04]	[S.1.02]	[S1.05]	[S1.01]
0950	(24) <i>Prior Elicitation for Kids: Interactive Spreadsheet Graphics with Sliders Can Be Fun, and Informative</i> Geoff Jones and Wes Johnson	(41) <i>An optimization model for strategic fleet planning in tramp shipping</i> Jørgen Laake and Abraham Zhang	(45*) <i>Predictive risk modeling for hospital readmission</i> Claire Forsythe and Chaitanya Joshi	(38) <i>A K-fold Averaging Cross-validation Procedure</i> Yoosuh Jung and Jianhua Hu
1010	(64+) <i>Developing students understanding of statistical inference: A research project</i> Maxine Pfannkuch and Stephanie Budgett	(42+) <i>Lean and Six Sigma implementation in the logistics industry: a snapshot in Singapore</i> Abraham Zhang, Chia Ting and Xavier Sim	(83+) <i>Environmental factors and stroke incidence: evidence from six large population-based studies in Europe and Australasia</i> Priya Parmar, Valery Feigin, Suzanne Barker-Collo and Rita Krishnamurthi	(124*##+) <i>Do Opinion Polls Predict New Zealand General Election Results?</i> Peter Green
1030	Morning Tea (20 minutes) S-Block First Floor Foyer			
	[S.1.04]	[S.1.02]	[S1.05]	[S1.01]
	<i>In Honour of William Bolstad</i>			
1050	(34) <i>Bayesian semiparametric likelihood approximations for stationery time series</i> Renate Meyer and Claudia Kirch	(127+) <i>The statistician's role in the conduct of randomised trials</i> Alain C. Vandal	(23) <i>A spatiotemporal rainfall model for Auckland: further developments</i> Paul Cowpertwait	(129#) <i>Using the census as a sampling frame: a rare opportunity to understand non-response bias</i> Kylie Maxwell
1110	(16+) <i>Non-MCMC based methods for Computational Bayesian inference</i> Chaitanya Joshi	(126) <i>Data Monitoring Committees a New Zealand Experience</i> Katrina Sharples	(3) <i>Bayesian threshold selection for extremal models using measures of surprise</i> Kate Lee, Yanan Fan and Scott Sisson	(55*) <i>Cluster analysis for longitudinal ordinal data: A likelihood-based approach based on finite mixture models</i> Roy Costilla and Ivy Liu
1130	(69) <i>Randomization and approximation in MCMC</i> Matthew Parry	(53) <i>Network meta-analysis for multi-arm trials</i> Thomas Lumley	(31*) <i>Outliers in Time Series – Using Aggregated Unit Record Data</i> Christopher Bryant	(4) <i>AIC and BIC for survey data</i> Alastair Scott and Thomas Lumley
1150	(37) <i>Statistics and the laws of physics</i> William Bolstad	(52) <i>Extension of Clinical Trials to Population Health Stepped Wedge Design in Randomised Controlled Trials</i> Yannan Jiang	(71*+) <i>Extreme Value Mixture Models – An R Package and Simulation Study</i> Yang Hu and Carl Scarrott	(79+) <i>Longitudinal analysis of child behaviour in the Pacific Islands Families Study</i> Steve Taylor
1210	Lunch (50 minutes) S-Block First Floor Foyer			

1300	Chair: Andy Philpott				Plenary session [S.1.04] <i>Analytics at Fonterra</i> Kevin Ross			
1350	Chair: Andrew Balemi				Plenary session [S.1.04] (125) The Rule of Three in Analytics Annelies Tjetjep			
1420	Afternoon Tea (20 minutes) S-Block First Floor Foyer							
	[S.1.04]		[S.1.02] Derceto YPP (15-min Talks) Chair: Golbon Zakeri		[S1.05]		[S1.01]	
1440	(17) Using a copula-based model of GST data to visualise the New Zealand Richard Vale		(99) Seismic Mitigation Decisions for Building Owners in NZ J. Fei Ying, Suzanne Wilkinson and Jim Corner (100) Multi-objective optimisation in decision making of infrastructure asset management		(39*) Methods for detecting anomalies in routinely recorded animal health data, with particular reference to Foot-and-mouth disease in Vietnam Kate Richards, Mark Stevenson, Martin Hazelton, Cy Lockhart, J Pinto and L Nguyen		(13+) A discussion on Quality and Risk Assessment for an Industrial Application Govindaraju Kondaswamy, Haizhen Wu and Roger Kissling	
1500	(57#+) An R package for optimal pricing of perishable assets Rolf Turner		Lin Chen, Andrea Raith and Theuns Henning (101) Capacity Expansion Optimisation in Transportation Networks		(102*) Improved Models for Stutter Prediction in Forensic DNA Analysis M.A. C. S. Sampath Fernando, James M. Curran, Jo-Anne Bright, John S. Buckleton and Renate Meyer		(29*#) Nonparametric Multivariate Density Estimation Using Mixtures Xuxu Wang and Yong Wang	
1520	(88) National Instantaneous Reserve Prices and Revenue in New Zealand Electricity Market Vladimir Krichtal and Conrad Edwards		Danny Tsai, Andrea Raith and Andy Philpott (103) Modelling risk aversion in generation expansion Corey Kok and Andy Philpott		(27) Extended Data Cloning: Links to Data Visualisation, Smoothing, Confidentiality, and Encryption Stephen Haslett		(51*) Robustifying the geographic relative risk function Khair Jones and Martin Hazelton	
1540	(70) Drawing (piecewise) Smooth Curves Ross Ihaka		(104) Binary Interruptible Load Optimisation Jack Dunn, Golbon Zakeri and Anthony Downward (106) Transmission Tower Painting Optimisation		(48) A View of Data Science Donal Krouse		(30*) Shape-restricted Density Estimation with Applications to Financial Data Yu Liu	
1600	(120) Visualization of a Stock Market Correlation Matrix Alethea Rea and William Rea		Jenny McLean, Anthony Downward, Golbon Zakeri and Andy Philpott (107) Graph partitioning and its Application to the Traffic Assignment Problem		(92+) Challenges in Obtaining and Using Data and Making Decisions in a Post-Emergency Environment John Créquer		(14#) Variables Acceptance Sampling Procedure with Consideration of Model Uncertainty and Measurement Error Haizhen Wu and Govindaraju Kondaswamy	
1620 - 1640	(56*) Decomposing Observed Volatility in Financial Series using Quality Control Theory Nadeeka D. Premarathna, A. Jonathan R. Godfrey and K. Govindaraju		Keith Ruddell		(128) Pulp Mill Power Tool Graeme Everett and Andy Philpott		(111) Symmetric Traveling Salesman Problem and Flows on Hypergraphs- New algorithmic possibilities Tiru Arthanari	
1640 - 1800	Poster Session [Posters displayed in S-Block First Floor Foyer, Talks in S.1.04]							
1930	Young Statisticians and Operation Researchers Evening [McMeekan Centre, AgResearch Ruakura]							

TUESDAY, 26 November				
0900	<p style="text-align: center;"><i>Chair: Alastair Scott</i></p> <p style="text-align: center;">Plenary session [S.1.04] <i>Assessing Change with Longitudinal Data</i> Plenary Speaker: John Neuhaus</p>			
	[S.1.04]	[S.1.02]	[S1.05]	[S1.01]
0950	<p><i>(77) Simulating Cytology with FlexSim HC</i></p> <p>Ploy Pongjetanpong, Michael O'Sullivan and Cameron Walker</p>	<p><i>(18) Supervised learning with immigration data: prediction of unwanted client behaviour</i></p> <p>Robin Willink</p>	<p><i>(33) Visualising orbital effects in ice-core data</i></p> <p>Jean Sanderson</p>	<p><i>(35*) A Fast Algorithm for Estimating a Nonparametric Bathtub-Shaped Hazard Function</i></p> <p>Shabnam Fani</p>
1010	<p><i>(118+) Improving Patient Transit at Auckland City Hospital</i></p> <p>Michael O'Sullivan, Fergus Bassett and Cameron Walker</p>	<p><i>(105+) POSM: optimal dairy farm stock management under uncertainty</i></p> <p>Andy Philpott and Peter Mills</p>	<p><i>(95+) Transforming a public sector team's data production, management, visualisation and analysis</i></p> <p>Peter Ellis</p>	<p><i>(86*#+) New distribution free goodness of fit tests for power-law</i></p> <p>Thuong Nguyen</p>
1030	<p>Morning Tea (20 minutes) First floor foyer of S-Block</p>			
	[S.1.04]	[S.1.02] Derceto YPP (15-min Talks) <i>Chair: Andrea Raith</i>	[S1.05] <i>Chair: Maheswaran Rohan</i>	[S1.01]
1050	<p><i>(44*#) Automatic Conversion of Tables to LongForm Dataframes</i></p> <p>Jimmy Oh</p>	<p><i>(108) 2D shape optimization applied to diamond cutting</i></p> <p>Max Biggs and Tony Downward</p>	<p><i>(82) The efficient computation of the Group Inverse and the Mean First Passage Times in a Markov Chain</i></p> <p>Jeffrey Hunter</p>	<p><i>(9+) Some Observations on the Design of Phase 1 Clinical Trials</i></p> <p>Hans Hockey</p>
1110	<p><i>(2) Reusable Graphics</i></p> <p>Paul Murrell</p>	<p><i>(114*) Creating a Navigational support system for construction projects</i></p> <p>Foad Marzoughi and Tiru Arthanari</p>	<p><i>(72) On fragmentation processes and connections to diversity problems</i></p> <p>Estate Khmaladze</p>	<p><i>(25) Multi-phase trial design for field to laboratory research: soil carbon mineralisation case study</i></p> <p>Esther Meenken, Craig Anderson and Fabiane Vezzani</p>
1130	<p><i>(58#) Video games and MCMC: Parallel Bayesian computation using GPUs</i></p> <p>Chris Jewell</p>	<p><i>(119#) Solving the Classroom Assignment Problem Using Integer Programming</i></p> <p>Antony Phillips, Matthias Ehgott and David Ryan</p>	<p><i>(60) Applying Real Options approach to determine the optimal vaccination timing and threshold</i></p> <p>Elena Moltchanova, Zuzana Chladna and Michael Obersteiner</p>	<p><i>(50*) Modeling for Carryover Effects in Clinical Trials</i></p> <p>S Gwynn Sturdevant and Thomas Lumley</p>
1150	<p><i>(80*#) Software for Reproducible Research Using R</i></p> <p>David Scott</p>		<p><i>(66) Incompletely observed point processes in time: an application to volcanic eruption records</i></p> <p>Ting Wang and Mark Bebbington</p>	<p><i>(121*) Countermatched design as a two-phase sampling design</i></p> <p>Claudia Rivera and Thomas Lumley</p>
1210	<p>Lunch (50 minutes) S-Block First Floor Foyer</p>			
1300	<p><i>Chair: Mark Johnston</i></p> <p style="text-align: center;">Plenary session [S.1.04] <i>It's just common sense, right? So why is it so uncommon?</i> Vicky Mabin</p>			
1350	<p><i>Chair: Harold Henderson</i></p> <p style="text-align: center;">Plenary session [S.1.04] (87) Shayle Searle Memorial Jeffrey Hunter</p>			

1420	Afternoon Tea (20 minutes) S-Block First Floor Foyer			
	[S.1.04]	[S.1.02]	[S1.05]	[S1.01]
1440	(7) <i>Multi-Depot Vehicle Scheduling Problem in Urban Public Transportation with Large Instances</i> Pablo Guedes and Denis Borenstein	(15#) <i>Bayesian Reconstruction of Past Populations and Vital Rates by Age for Developing and Developed Countries</i> Mark Wheldon, Adrian Raftery, Samuel Clark and Patrick Gerland	(1) <i>What's killing our honeybees?</i> Graham Wood	(62*) <i>MANOVA for high-dimensional data</i> Insha Ullah
1500	(94) <i>On Solving The Multi-objective Beam Angle Optimisation Problem Using Single Objective Local Search</i> Guillermo Cabrera G., Matthias Ehrgott, Andrew Mason and Andrea Raith	(20) <i>Bayesian Parallel Computation for Intractable Likelihood</i> Nuttanan Wichitaksorn	(6*#) <i>Likelihood-Based Finite Mixture Models for Ecological Ordinal Data</i> Daniel Fernandez, Richard Arnold and Shirley Pledger	(68) <i>The fixed-effects modified-spluce (FEMS) method: fully quality-adjusted price indexes with no characteristics information</i> Frances Krsinich
1520	(117) <i>Economic Appraisal of Cycling Infrastructure Investments</i> Andrea Raith	(28) <i>Towards the Control of Markov Chains with Constraints: Theory and Applications</i> Boris Miller	(32*) <i>Modelling retinal venous pulsation from video data</i> Brigid Betz-Stablein, Martin Hazelton and Morgan William	(115) <i>Multi-node offer stack optimisation over electricity networks</i> Anthony Downward, Yibo Weng and Golbon Zakeri
1540	(109) <i>Dantzig-Wolfe Decomposition for the Bi-Objective Multi-Commodity Minimum Cost Flow Problem</i> Siamak Moradi, Andrea Raith and Matthias Ehrgott	(40+) <i>EM Algorithm and Gibbs Sampling: Compare and Contrast</i> Murray Jorgensen	(49*) <i>Trait-based ecological community assembly using a hierarchical Bayesian model</i> Paul T. Brown, Chaitanya Joshi and Daniel C. Laughlin	(91) <i>Statistics and counterfeit medicines</i> Sara J Fraser, Austina Clark and Keith Gordon
1600	(113+) <i>Friday Night's A Great Night For Cricket</i> Mark Johnston	(67*) <i>Empirical Bayes estimation for the stochastic block model</i> Shakira Suwan, Dominic Lee and Carey E. Priebe	(89) <i>A Review of the Accuracy of Pith Offset Methods Applied to Kauri Incremental Cores</i> Maryann Pirie	(110+) <i>Do some like it hot? Sex ratios (and time series) in Australia and New Zealand</i> John Haywood
1620 - 1640		(85*) <i>Bayesian age reconciliation for the Auckland Volcanic Field</i> Emily Kawabata, Mark Bebbington, Shane Cronin and Ting Wang	(93+) <i>Determination of the relative contributions of water from soil substrates to stream water</i> Neil Cox, Richard McDowell and Tim Jowett	(112) <i>Shortest Path Algorithms in Traffic Assignment</i> Boshen Chen, Olga Perederieieva and Andrea Raith
1640 - 1740	NZSA AGM [S.1.05]		ORSNZ AGM [S.1.01]	
1830	Conference Dinner [Gallagher Academy of the Performing Arts]			

WEDNESDAY, 27 November				
0900	<p><i>Chair: Thomas Lumley</i></p> <p style="text-align: center;">Plenary session [S.1.04] <i>A Unified Approach to Shrinkage</i> Plenary Speaker: Ken Rice</p>			
	[S.1.04]	[S.1.02]	[S1.05]	[S1.01]
0950	<p><i>(22) Using social media to promote Statistics and Operations Research</i></p> <p>Nicola Ward Petty</p>	<p><i>(19) The shrinking test set: optimal splitting of the dataset in supervised learning</i></p> <p>Robin Willink</p>	<p><i>(43) Modelling New Zealand Building Activity from Building Consents Data</i></p> <p>Sonya McGlone</p>	
1010	<p><i>(90+) A Review of Probability and Statistics Apps for Mobile Devices</i></p> <p>Howard Edwards</p>	<p><i>(81+) PhD project in light of project Management</i></p> <p>Maryam Mirzaei</p>	<p><i>(98+) Examining differences between perceived and actual urban environment measures</i></p> <p>Nick Garrett</p>	<p><i>(130) NeSI: High Performance Computational Platforms and Services for NZ's Research Communities</i></p> <p>John Rugis, Ben Roberts and Mark Cheeseman</p>
1030	<p>Morning Tea (20 minutes) S-Block First Floor Foyer</p>			
	[S.1.04]	[S.1.02]	[S1.05]	[S1.01]
1050	<p><i>(63) Parallel sequencing of barcoded BACs using NGS technology</i></p> <p>Marcus Davy, Roy Storey and Elena Hilario</p>	<p><i>(61#) Jointly optimal bandwidth selection for the planar kernel-smoothed density-ratio</i></p> <p>Tilman Davies</p>	<p><i>(21) Clustering Based Active Learning for Evolving Data Streams</i></p> <p>Bernhard Pfahringer, Dino Ienco, Albert Bifet and Indre Zliobaite</p>	<p><i>(26) A methodology for comparing revisions in time series data</i></p> <p>Conrad MacCormick and Ryan Buchanan</p>
1110	<p><i>(76#) Modelling ecological competition and dispersal in a statistical phylogeographic framework</i></p> <p>Louis Ranjard, David Welch and Stephane Guindon</p>	<p><i>(74) Collecting data and estimating transmissibility and severity in the early stages of an epidemic</i></p> <p>Joshua Ross</p>	<p><i>(8+) Random Projections for Data Mining and Optimization - Theory and Applications</i></p> <p>Bob Durrant</p>	<p><i>(46) The end point of time series: Reporting informatively on trends</i></p> <p>Giles Reid and John Créquer</p>
1130	<p><i>(78+) Identification of synthetic lethal gene pairs for targeted cancer therapies</i></p> <p>Mik Black</p>	<p><i>(65) Determining the source of human campylobacter cases through time</i></p> <p>Jonathan Marshall</p>	<p><i>(122) Mining for differences and changes</i></p> <p>Yong Wang</p>	<p><i>(75) On the determination of variance change points, with application to wavelet analysis</i></p> <p>Leigh Roberts</p>
1150		<p><i>(59+) Bayesian risk prediction for epidemic decision support</i></p> <p>Chris Jewell</p>	<p><i>(54) Exploring Multiple omics Datasets</i></p> <p>Vanessa Cave, Marlon Reis, Paul Maclean and Kuljeet Singh</p>	<p><i>(96+) Fast computation of adjusted forecasts in hierarchical time series</i></p> <p>Rob Hyndman and Alan Lee</p>
1210	<p>Lunch (50 minutes) S-Block First Floor Foyer</p>			
1300	<p><i>Chair: James Curran</i></p> <p style="text-align: center;">Plenary session [S.1.04] <i>Some Recent Advances in Network Tomography</i> Martin Hazelton</p>			
1350	<p>CLOSE [S.1.04]</p>			

Plenary Presentation Abstracts

THE NEED FOR SPEED IN THE PATH OF THE DELUGE

Chris Wild
University of Auckland

[S.1.04] 9:00-9:50 Monday 25 November

There is a rapidly increasing awareness of the so-called "data deluge": the explosion in quantities of data being collected, the explosion of settings in which it is being collected, and expansions in the conceptions and scope of what constitutes data. This is accompanied by advances in ways of visualising data and accessible data-visualisation tools. Consequently, it is imperative to find ways to get students much further, much faster and with better comprehension - a quantum leap in ambition. What can make this possible are some of the same things that gave rise to the deluge, computational power and clever software. We will advance some strategies that envisage maximising awareness and excitement about data and what it can do for you and only later back filling details. We also provide glimpses of two software projects beginning to enable such a future - a fleeting encounter with some data-analysis software followed by a more in-depth look at visualisation approaches to bootstrap and randomisation inference.

ANALYTICS AT FONTERRA

Kevin Ross
Fonterra

[S.1.04] 13:00-13:50 Monday 25 November

Fonterra recognises that advanced analytics capability presents an opportunity to develop and sustain a competitive advantage in the global dairy industry. In order to maximize both the value created from milk and shareholder returns, Fonterra needs to employ a combination of descriptive, predictive and prescriptive analytics. We have developed a range of models to support decisions from milk collection to product mix and asset investment. As part of this initiative, we launched the New Zealand Analytics Forum in 2013. This group of Analytics professionals meets regularly to share best practice principles. We see the potential for this to benefit a wide range of industries, and New Zealand as a whole. I will share some of the insights from both Fonterra's efforts and the Analytics Forum.

ASSESSING CHANGE WITH LONGITUDINAL DATA

John M. Neuhaus (with Charles E. McCulloch)
University of California, San Francisco

[S.1.04] 9:00-9:50 Tuesday 26 November

Investigators often gather longitudinal data to assess changes in responses over time within subjects and to relate these changes to within-subject changes in predictors. Missing data are common in such studies and predictors can be correlated with subject-specific effects. Maximum likelihood methods for generalized linear mixed models provide consistent estimates when the data are “missing at random” (MAR) but can produce inconsistent estimates in settings where the random effects are correlated with one of the predictors. On the other hand, conditional maximum likelihood methods (and closely related maximum likelihood methods that partition covariates into between- and within-cluster components) provide consistent estimation when random effects are correlated with predictors but can produce inconsistent covariate effect estimates when data are MAR. Using theory, simulation studies, and fits to example data this talk shows that decomposition methods using complete covariate information produce consistent estimates. In some practical cases these methods, that ostensibly require complete covariate information, actually only involve the observed covariates. These results offer an easy-to-use approach to simultaneously protect against bias from either cluster-level confounding or MAR missingness in assessments of change.

IT’S JUST COMMON SENSE, RIGHT? SO WHY IS IT SO UNCOMMON?

Vicky Mabin
Victoria University of Wellington

[S.1.04] 13:00-13:50 Tuesday 26 November

Divide and conquer may work in battle, but doesn’t usually work for organisations, where silo thinking and turf wars can lead to poor overall performance. When constraints, complexity and conflicts are managed well, organisations perform significantly better. Drawing on examples from sectors as diverse as manufacturing, health, resource management and education, Professor Vicky Mabin will illustrate how managing these factors using Goldratt’s theory of constraints can improve organisational and personal performance.

A UNIFIED APPROACH TO SHRINKAGE

Ken Rice
University of Washington

[S.1.04] 9:00-9:50 Wednesday 27 November

As data have become "Big", shrinkage estimators of various forms have become standard tools in statistical analysis. Common justifications for them include penalized maximum likelihood, empirical Bayes posterior means, and full Bayes posterior modes. None of these, however, addresses the question of why one might want a shrunken estimate in the first place. In this talk we outline a general approach to shrinkage, as a result of balancing veracity (getting close to the truth) and simplicity (getting close to zero, typically). While yielding "simple" shrunken estimates, the approach does not require any assumption that the truth is actually full of zeros - an assumption that is often unreasonable. Several well-known shrinkage estimates will be derived as special cases, illustrating close connections between them.

SOME RECENT ADVANCES IN NETWORK TOMOGRAPHY

Martin Hazelton
Massey University

[S.1.04] 13:00-13:50 Wednesday 27 November

Volume network tomography is concerned with inference about traffic flow characteristics based on traffic measurements at fixed locations on the network. The quintessential example is estimation of the traffic volume between any pair of origin and destination nodes using traffic counts obtained from a subset of the links of the network. The data provide only indirect information about the target variables, generating a challenging type of statistical linear inverse problem. For much of the past 40 years work on network tomography appeared primarily in the engineering literature, but burgeoning interest by statisticians has seen it dubbed a 'classic' problem in a recent issue of JASA. In this talk I will discuss network tomography for a rather general class of traffic models. I will describe some recent progress on model identifiability, and the development of effective MCMC samplers for simulation-based inference.

Presentation Abstracts

(1) WHAT'S KILLING OUR HONEYBEES?

Graham Wood
Warwick University

[S.1.05] 14:40-16:40 Tuesday 26 November

It is commonly estimated that one third of the food we eat depends on honeybees for pollination. In 2000 the varroa mite, a honeybee parasite, arrived in the North Island and in 2006 it spread to the South Island. Since that time it has caused the death of thousands of NZ honeybee colonies. The nub of the problem appears to be a particular recombinant of two viruses, one carried by varroa (the aptly named varroa destructor virus) and the other a very similar virus carried by the honeybee (the deformed wing virus). The bee defends itself against recombinants by using its interfering RNA mechanism, which results in chopped-up small lengths of the recombinant viruses being available as siRNA, small interfering RNA. Next generation sequencing allows us to sequence vast numbers of these siRNAs, which can then be aligned to a consensus viral genome. This talk describes a method for extracting from such data the recombinant virus mosaics present, and their proportions. It involves setting the problem geometrically, the use of constrained quadratic programming to fit a model, the development of a test for the significance of recombinants and software to do the work (MosaicSolver). The probable culprit will be revealed!

(2) REUSABLE GRAPHICS

Paul Murrell
University of Auckland

[S.1.04] 10:50-12:10 Tuesday 26 November

Statistical graphics (plots) are usually produced as an end point, for example, to serve as a figure in a PDF report. The graphic is designed to be viewed, with attention not usually focused on further reuse of the graphic, for example, inclusion in someone else's PDF report. Even complex interactive displays such as dynamic plots within a web page are designed to be interacted with in predefined ways. Attention is not usually focused on allowing the graphic to be adapted to new uses. This talk will discuss some ways in which statistical graphics can be made more reusable, with examples of reusing graphics and producing reusable graphics in R.

(3) BAYESIAN THRESHOLD SELECTION FOR EXTREMAL MODELS USING MEASURES OF SURPRISE

Kate Lee, Yanan Fan and Scott Sisson
Auckland University of Technology, University of New South Wales, University of New South Wales

[S.1.05] 10:50-12:10 Monday 25 November

Statistical extreme value theory is concerned with the use of asymptotically motivated models to describe the extreme values of a process. A number of commonly used models are valid for observed data that exceed some high threshold. However, in practice a suitable threshold is unknown and must be determined for each analysis. While there are many threshold selection methods for univariate extremes, there are relatively few that can be applied in the multivariate setting. In addition, there are only a few Bayesian-based methods, which are naturally attractive in the modelling of extremes due to data scarcity. In this talk, we demonstrate that the use of Bayesian measures of surprise is suitable in determining thresholds for extreme models. Such measures quantify the level of support for the proposed extremal models without the need to specify any model alternatives.

(4) AIC AND BIC FOR SURVEY DATA

Alastair Scott and Thomas Lumley
University of Auckland

[S.1.01] 10:50-12:10 Monday 25 November

Model selection criteria such as AIC and BIC are widely used in applied statistics. In recent years there has been a huge increase in fitting models to data from large complex surveys, and a resulting demand for versions of AIC and BIC that are valid under complex sampling. We show how to extend both criteria to complex samples. Following the approach of Takeuchi (1976) for possibly-misspecified models, AIC can be extended by replacing the usual penalty term with the trace of the Rao-Scott design effect matrix. BIC can be extended by treating the point estimates under complex sampling as the data available for Bayesian modelling, and then applying the Laplace approximation argument used to construct the standard expression for BIC.

(6) LIKELIHOOD-BASED FINITE MIXTURE MODELS FOR ECOLOGICAL ORDINAL DATA

Daniel Fernandez, Richard Arnold and Shirley Pledger
Victoria University of Wellington

[S.1.05] 14:40-16:40 Tuesday 26 November

Many of the methods to deal with dimension reduction in matrices of data are based on mathematical techniques such as distance-based algorithms or matrix decomposition and eigenvalues. In general, it is not possible to use statistical inference with these techniques because there is no underlying probability model. Recent research has developed a set of likelihood-based finite mixture models for a data matrix of binary or count data (Pledger and Arnold 2013, to appear). My current research extends this earlier work by formulating likelihood-based multivariate methods for ecological ordinal data. My talk will introduce the first results from this research, which applies fuzzy clustering via finite mixtures to the ordered stereotype model. Data from a real ecological example will be shown to illustrate the application of this approach. Finally, I will present the results of a simulation study conducted to determine which information criteria are most appropriate for these particular mixture models when applied to ordinal data.

(7) MULTI-DEPOT VEHICLE SCHEDULING PROBLEM IN URBAN PUBLIC TRANSPORTATION WITH LARGE INSTANCES

Pablo Guedes and Denis Borenstein
Federal University of Rio Grande do Sul

[S.1.04] 14:40-16:40 Tuesday 26 November

The multiple-depot vehicle-scheduling problem (MDVSP) is a classic logistics and transportation problem. The MDVSP is also the basis for solving various related problems, such as the real time vehicle scheduling problem, disruption management; and integrated problems such as the vehicle and crew scheduling problems, and vehicle-crew-rostering problems. Although several mathematical and solution methods have been developed in the literature, large instances, involving thousands of trips and several depots, are still difficult to solve in a reasonable time. The objective of this article is to verify the applicability of the use of the space-time network towards obtaining good solutions for large instances in short time. Time-space network was suggested by Kliwer et al (2006), and it is positioned with respect to two-dimensional axes, one representing time and the other one space or stations. The arcs represent deadheading movements, and waiting periods in the same station. Solution methods for the MDVSP combining time space with integer linear programming solvers and column generation were developed. Extensive testing was carried out using random generated instances, based on demands distribution. Large instances, involving thousands of trips (between 1,000-10,000) and dozen (4-64) of depots, are solved in reasonable times.

(8) RANDOM PROJECTIONS FOR DATA MINING AND OPTIMIZATION - THEORY AND APPLICATIONS

Bob Durrant
University of Waikato

[S.1.05] 10:50-12:10 Wednesday 27 November

Random projections have been used as a dimensionality reduction technique for large data since the appearance of Arriaga and Vempalas seminal FOCS 1999 paper, and they continue to find applications both within the fields of Data Mining and Machine Learning, as well as elsewhere. Starting with some motivating examples from Data Mining and Machine Learning, I will review some key theoretical properties of random projections, and the practical applications this theory has inspired. In particular, I will cover the Johnson-Lindenstrauss lemma which gives conditions under which random projections approximately preserve geometric properties of data and give some related applications, discuss the field of Compressed Sensing from which one can derive guarantees for techniques working with sparse data, and talk about my own recent work using random projections in two very different settings: (i) as a regularizer for classification in the " $N \ll p$ " setting, and (ii) as a divide and conquer technique for large scale continuous optimization using Estimation of Distribution Algorithm (EDA) - a state-of-the-art stochastic local search method.

(9) SOME OBSERVATIONS ON THE DESIGN OF PHASE 1 CLINICAL TRIALS

Hans Hockey
Biometrics Matters

[S.1.01] 10:50-12:10 Tuesday 26 November

In Phase 1 of clinical drug development the pharmacokinetic and pharmacodynamic properties of new compounds need to be explored, as well as safety and tolerability, particularly in relation to dose level and the subsequent determination of the maximum tolerated dose. Hence drug half-life, dose proportionality and other pharmacologic properties of the drug are estimated so as to determine how best to administer the drug. There are also issues of gender and genotype differences, drug-drug interactions, the effect of food, of bioequivalence between different formulations, and so on, all of which are typically studied using healthy volunteers. Determination of efficacy in patients suffering from the targeted indication cannot proceed until there is sufficient knowledge of what effects the bioavailability and other properties of the new drug. Examples of early stage clinical trial designs that I have encountered will be presented and discussed, including cross-over designs. Personal observations, some from an agricultural research perspective, on experimental design and subsequent analyses in early stage clinical trials will be made.

(13) A DISCUSSION ON QUALITY AND RISK ASSESSMENT FOR AN INDUSTRIAL APPLICATION

Govindaraju Kondaswamy, Haizhen Wu and Roger Kissling
Massey University, Massey University, Fonterra Co-operative Group Limited

[S.1.01] 14:40-16:40 Monday 25 November

The adage 'ignorance is not probability' applies to an industrial application that involves tail probability estimation. Model uncertainties can lead to incorrect estimation of the tail probabilities. An industrial example will be presented to show that quantitative risk assessment is more than estimation of a tail probability. Model selection using theoretic information criteria such as AIC, BIC & MDL are popular in the statistical literature. It will be shown that these approaches perform rather poorly for the risk estimation problem but context related actions can be taken to reduce the risks.

(14) VARIABLES ACCEPTANCE SAMPLING PROCEDURE WITH CONSIDERATION OF MODEL UNCERTAINTY AND MEASUREMENT ERROR

Haizhen Wu and Govindaraju Kondaswamy
Massey University

[S.1.01] 14:40-16:40 Monday 25 November

Traditional acceptance sampling procedures rely heavily on the assumption of the underlying distribution being invariant in shape and the measurement errors being negligible. These assumptions are rather idealistic than realistic for short-run processes, because the risks of the acceptance and rejection decisions due to the misspecified assumptions can be considerably large and are even unknown. Allowing flexibility in the shape of the underlying distribution reduces the danger of severely misspecifying the distribution shapes and hence decreases the risks due to misspecification. Shape uncertainty introduces extra difficulty that the operating characteristics (OC) curve of certain sampling plan becomes uncertain, and the (producers and consumers) risks associated with the plan become uncontrollable. Adjustment for the measurement errors also helps to understand more about the true underlying distribution and reduce the risks due to distribution misspecification. However, how to take this measurement error information into account to control the risks is a complex problem. In this talk, we will present a computationally intensive variables sampling procedure that controls the risks of sampling and eliminates the impact of measurement errors, at least in large sample situation.

(15) BAYESIAN RECONSTRUCTION OF PAST POPULATIONS AND VITAL RATES BY AGE FOR DEVELOPING AND DEVELOPED COUNTRIES

Mark Wheldon, Adrian Raftery, Samuel Clark and Patrick Gerland
*Auckland University of Technology, University of Washington, University of Washington, United Nations
Population Division*

[S.1.02] 14:40-16:40 Tuesday 26 November

Bayesian reconstruction, proposed by Wheldon et al. (*J. Am. Stat. Assoc.*, 108, 96--110, 2013), is a method for estimating past populations by age, with fully probabilistic statements of uncertainty. It simultaneously estimates age-specific population counts, fertility rates, mortality rates and net international migration flows from fragmentary data while formally accounting for measurement error. As inputs, Bayesian reconstruction takes initial bias-reduced estimates of age-specific population counts, fertility rates, survival proportions and net international migration. We extend the method to apply to countries without censuses at regular intervals. We also develop a method for using it to assess the consistency between model life tables and available census data, and hence to compare different model life table systems. We show that the method works well in countries with widely varying levels of data quality by applying it to reconstruct the past female populations by age of Laos, a country with little vital registration data where population estimation depends largely on surveys, Sri Lanka, a country with some vital registration data, and New Zealand, a country with a highly developed statistical system and high-quality vital registration data.

(16) NON-MCMC BASED METHODS FOR COMPUTATIONAL BAYESIAN INFERENCE

Chaitanya Joshi
University of Waikato

[S.1.04] 10:50-12:10 Monday 25 November

Thanks to the extremely versatile MCMC methods, it is now possible to solve almost any complex real life problem using the Bayesian approach. Ever since their introduction to the statistical community, Bayesian inference has been almost exclusively implemented using the MCMC methods. However, MCMC methods come with their own drawbacks. As the complexity of the problem and the size of the data increases, the MCMC methods become prohibitively computationally expensive. This results in suboptimal inference. As a result, a lot of research effort in the recent years has been aimed at either improving the efficiency of the MCMC methods or at developing non-MCMC based methods which are computationally efficient yet accurate. One of the promising non-MCMC methods recently developed is the Integrated Nested Laplace Approximation (INLA). INLA provides quick and accurate inference on latent Gaussian models. Although INLA can be used on a wide variety of applications, it has its own limitations too. My research has largely focused on: a) developing INLA-like methods for specific classes of models and b) improving INLA by getting around some of its limitations. In this talk, I'll provide an overview of the INLA method and also talk about my own research in this area.

(17) USING A COPULA-BASED MODEL OF GST DATA TO VISUALISE THE NEW ZEALAND ECONOMY

Richard Vale
Inland Revenue Department

[S.1.04] 14:40-16:40 Monday 25 November

GST (Goods and Services Tax) is one of the data sources used to calculate GDP and other measures of economic performance. Each New Zealand business reports its expenses and sales income when it files a GST return. The joint distribution of expenses and sales changes over time. The purpose of this work is to understand these changes. A five-parameter model of the joint distribution of GST expenses and sales was constructed using a copula. We explain this model and present an animated visualisation written in R and Javascript. This visualisation allows us to track changes in the economy over the last twelve years and identify key events such as the Global Financial Crisis. It also enables us to see a clear business cycle, with the distribution of expenses and sales returning to an earlier state after some time. This analysis shows that we can get very interesting insights by considering the shape of the joint distribution of expenses and sales, rather than just using the aggregate amount of expenses and sales themselves.

(18) SUPERVISED LEARNING WITH IMMIGRATION DATA: PREDICTION OF UNWANTED CLIENT BEHAVIOUR

Robin Willink
Immigration New Zealand

[S.1.02] 9:50-10:30 Tuesday 26 November

Each year Immigration New Zealand (INZ) processes approximately 500,000 visa applications and approves approximately ninety percent. Occasionally a person whose application is approved goes on to exhibit unwanted behaviour of some kind. INZ would like to be able to predict such an outcome at the time of processing. This is one of several projects being explored as INZ looks to develop expertise in analytics. This presentation outlines how this project was formulated as a problem of supervised learning and how it was subsequently conducted. The relevant variables are primarily categorical. It was found that market basket analysis, which looks for local patterns and which some analysts might not regard as modelling, gave better results than a decision tree. Reasons for this are outlined.

(19) THE SHRINKING TEST SET: OPTIMAL SPLITTING OF THE DATASET IN SUPERVISED LEARNING

Robin Willink
Immigration New Zealand

[S.1.02] 9:50-10:30 Wednesday 27 November

Typically, supervised learning involves a set of data on many variables for each of n subjects. One of these variables is an 'outcome' that we would like to be able to estimate from the other variables. Subsequently, the model developed could be used to predict the outcome for a future subject. The model is formed from the data, so its predictive ability can only be assessed honestly using a set of data that is independent. Therefore, it is advisable to split the dataset into a set of size $n-m$ that is used for developing the model and a 'test set' of size m that is used solely for evaluating its performance. But how large should we make m/n ? If the development set is too small then the model will not be optimal, but if the test set is too small then the estimator of performance will have high variance. We discuss choosing m from a development set that starts small but is allowed to grow. So the test set is chosen adaptively and is 'shrinking'. The objective is to maximise a lower confidence limit on predictive ability.

(20) BAYESIAN PARALLEL COMPUTATION FOR INTRACTABLE LIKELIHOOD

Nuttanan Wichitaksorn
University of Canterbury

[S.1.02] 14:40-16:40 Tuesday 26 November

Parallel computation is a fast growing computing environment in many areas including computational Bayesian statistics. However, most of the Bayesian parallel computing have been implemented through the sequential Monte Carlo method where model parameters are updated sequentially and it is suitable for some large-scale problems. This study is the first to revive the use of adaptive griddy Gibbs (AGG) algorithm under the Markov chain Monte Carlo framework and show how to implement the AGG using the parallel computation. The parallel AGG is suitable for (i) small to medium-scale problems where the dimension of model parameter space is not very high, (ii) some or all model parameters are defined on a specific interval, and (iii) model likelihood is intractable. In addition, the parallel AGG is relatively easy to implement and code. Since the AGG is a Gibbs algorithm where each of model parameters is directly drawn from the conditional posterior density, the model marginal likelihood can be conveniently computed and immediately provided after the end of posterior simulation. Three examples including a linear regression model with Student-t error, a nonlinear regression model, and a financial time series model (GARCH), are illustrated to show the applicability of the AGG under the parallel computing environment.

(21) CLUSTERING BASED ACTIVE LEARNING FOR EVOLVING DATA STREAMS

Bernhard Pfahringer, Dino Ienco, Albert Bifet and Indre Zliobaite
Clustering Based Active Learning for Evolving Data Streams

[S.1.05] 10:50-12:10 Wednesday 27 November

Data labeling is an expensive and time-consuming task. Choosing which labels to ask for is increasingly becoming important. In the active learning setting, a classifier is trained by asking for labels for only a small fraction of all instances. While many works exist that deal with this issue in non-streaming scenarios, few works exist in the data stream setting. In this talk we propose a new active learning approach for evolving data streams based on a clustering step, for selecting the most informative instances for labeling. We consider a batch incremental setting: when a new batch arrives, first we cluster the examples, and then, we select the best instances to train the learner. The clustering approach allows to cover the whole data space thus avoiding to oversample examples from only a few areas. We compare our method w.r.t. state of the art active learning strategies over real datasets. The results highlight the improvement in performance of our proposal. Experiments on parameter sensitivity are also reported.

(22) USING SOCIAL MEDIA TO PROMOTE STATISTICS AND OPERATIONS RESEARCH

Nicola Ward Petty
Statistics Learning Centre

[S.1.04] 9:50-10:30 Wednesday 27 November

In the International Year of Statistics, there has been considerable emphasis on getting the message of statistics out to the world. Nicola Ward Petty began a blog in December 2011 on how to learn and teach statistics and operations research. With weekly posts it has now grown to a readership of over 10,000 a month. In addition Nicola's company, Statistics Learning Centre, provides free YouTube videos teaching principles of statistics, to over 30,000 people a month. Twitter helps to raise the profile of the blog and the YouTube site. In this presentation, Nicola will talk about why and how academics and practitioners can be involved in social media such as Twitter, YouTube, Blogs and Facebook. Nicola was awarded a Greenfield Prize by ENBIS in Ankara earlier this year for her contributions to communication of statistical ideas.

(23) A SPATIOTEMPORAL RAINFALL MODEL FOR AUCKLAND: FURTHER DEVELOPMENTS

Paul Cowpertwait
Auckland University of Technology

[S.1.05] 10:50-12:10 Monday 25 November

A preliminary cluster analysis, based on rainfall data from 85 sites in the greater Auckland area, indicates that the Auckland area can be divided into three homogeneous regions. For each month in each of the three regions, a spatiotemporal stochastic rainfall model is fitted to statistical properties calculated from records of ten-minute rainfall time series data. The stochastic model is a revised version of a previously studied model and uses superposed point processes to account for different statistical properties in the different homogeneous regions. One thousand years of ten minute series are simulated at the 85 sites and annual maxima extracted and compared with those in the observed series. This validates the model for a range of applications that include flood studies and the design and analysis of urban drainage systems.

(24) PRIOR ELICITATION FOR KIDS: INTERACTIVE SPREADSHEET GRAPHICS WITH SLIDERS CAN BE FUN, AND INFORMATIVE

Geoff Jones and Wes Johnson
Massey University, University of California, Irvine

[S.1.04] 9:50-10:30 Monday 25 November

Common spreadsheet packages have the facility to insert scroll bars or “sliders” that allow the user to gradually change the values of selected cells. Coupled with the spreadsheet’s graphing facility, this gives a way of interactively exploring the shape of a probability distribution. We have found this to be a useful tool when eliciting priors for Bayesian data analysis. Such a spreadsheet can easily be adapted to different situations so that the “expert” can view scientifically meaningful summaries as opposed to the standard distribution parameters. The sliders can be set up to control either the parameters or particular features of the distribution. Here we present some examples to illustrate the scope and flexibility of the method.

(25) MULTI-PHASE TRIAL DESIGN FOR FIELD TO LABORATORY RESEARCH: SOIL CARBON MINERALISATION CASE STUDY

Esther Meenken, Craig Anderson and Fabiane Vezzani
Plant and Food Research, Plant and Food Research, Federal University of Parana

[S.1.01] 10:50-12:10 Tuesday 26 November

Multi-phase trials consist of more than one stage, for example, samples collected from a field trial that are then subjected to a laboratory based experiment. In this scenario, both phases are subject to independent sources of variation and bias. A multi-phase design helps ensure that final experimental results appropriately account for sources of variation at all phases and are robust and efficient. A four-phase experimental design is described with a non-orthogonal laboratory phase that investigates the effect of different farm management practices on functional diversity of the soil microbiology. This research contributes to environmental issues such as biodiversity, carbon cycling and sequestration. Analysis estimated positive REML variance components at several phases confirming the benefit of this structured design approach to experimental planning.

(26) A METHODOLOGY FOR MEASURING AND MINIMISING REVISIONS IN SEASONALLY ADJUSTED DATA

Conrad MacCormick and Ryan Buchanan
Statistics New Zealand

[S.1.01] 10:50-12:10 Wednesday 27 November

Statistics New Zealand publishes numerous economic and social indicators that contain seasonal patterns. These effects can be induced by cultural biases, tradition or the weather. Seasonal adjustment removes these seasonal influences allowing decision makers to see the underlying time movements in the data. Seasonal adjustment is a time average that will be affected by future releases of data. Each additional period (month or quarter) added to the time series will cause a revision of the full history of the seasonal adjusted data. Thus the current data release will be revised month after month for a period of 3 years. The volatility of these revisions can be such that growths can turn in declines or vice versa. Significant revisions can erode the public's confidence in the economic or social indicator. Thus measuring the size of revisions and minimising revisions becomes a measure of quality of the seasonal model used. In this paper we present a methodology for measuring revisions and choosing between two separate models. The paper presents one application of this methodology where we ask: Which X-12-ARIMA output minimises the revisions.

(27) EXTENDED DATA CLONING: LINKS TO DATA VISUALISATION, SMOOTHING, CONFIDENTIALITY, AND ENCRYPTION

Stephen Haslett
Massey University

[S.1.05] 14:40-16:40 Monday 25 November

There are situations in which it is possible to create new models or data with the same model fit as for another model or dataset. An example, which has a history back to the 1960s is variance-covariance structure changes in linear models where BLUEs (Best Linear Unbiased Estimates) from linear fixed effect models remain unchanged. This has been extended recently to understanding conditions under which BLUEs and/or BLUPs (Best Linear Unbiased Predictions) from linear mixed effect models remain unchanged. It is also possible to change the data, rather than the covariance structure, and still get the same BLUEs. In the multiple linear regression case, this procedure can be repeated many times, each time producing a new dataset with the same multiple regression fit as for the original data. We call these datasets cloned or matching. One advantage of these cloned datasets is that, unlike the more usual model-based alternative, parameter estimates of the original data and the cloned data do not include model error and hence are identical. Such extended data cloning has potential in a wide range of applications from confidentialising or encrypting data to data visualisation and smoothing.

(28) TOWARDS THE CONTROL OF MARKOV CHAINS WITH CONSTRAINTS: THEORY AND APPLICATIONS

Boris Miller
Monash University

[S.1.02] 14:40-16:40 Tuesday 26 November

A problem of stochastic optimal control for Markov chains with a finite number of states in continuous time is considered. The problem statement is assumed a non-stationary finite horizon problem with constraints given as a set of inequalities. It is shown that the optimal control within the class of Markov policies, if it exists, may be found with the aid of maximum principle as a solution of some deterministic optimal control problem. Meanwhile it can be shown that under simple assumptions of convexity, which are common for optimal control problems, this solution is optimal also within more general class of predictive controls. An approach to the numerical solution of the optimal control problem has been suggested and its effectiveness is illustrated by examples from the areas of the INTERNET congestion avoidance and dams' management.

(29) NONPARAMETRIC MULTIVARIATE DENSITY ESTIMATION USING MIXTURES

Xuxu Wang and Yong Wang
University of Auckland

[S.1.01] 14:40-16:40 Monday 25 November

A new method is proposed for nonparametric multivariate density estimation, which extends a general framework that has been recently developed in the univariate case based on nonparametric and semiparametric mixture of normal distributions. The major challenge to a multivariate extension is the dilemma that one can not maximize directly the likelihood function with respect to the whole component covariance matrix, since the likelihood is unbounded for a singular covariance matrix. To select the covariance matrix does not solve the problem of unboundedness and moreover is this would be at least computationally demanding if not infeasible. We consider using a volume parameter h to enforce a minimal restriction on the covariance matrix so that, with h fixed, the likelihood function is bounded and its maximization can be successfully carried out with respect to all the remaining parameters. The role played here by the scalar h is just the same as by the bandwidth in the univariate case and its value can be determined by a model selection criterion, such as the Akaike information criterion. New efficient algorithms are also described for finding the maximum likelihood estimates of these mixtures under various restrictions on the covariance matrix. Empirical studies using simulated and real-world data show that the new multivariate mixture-based density estimator performs remarkably better than two state-of-the-art kernel-based density estimators.

(30) SHAPE-RESTRICTED DENSITY ESTIMATION WITH APPLICATIONS TO FINANCIAL DATA

Yu Liu
University of Auckland

[S.1.01] 14:40-16:40 Monday 25 November

Parametric models are often used for modeling financial data, such as the skewed t-distribution or the stable distribution for stock prices and bond yields, as well as the log-gamma or the log-negative-binomial model for options pricing. Using parametric models can be advantageous in, e.g. computing and interpretation, but they may suffer, sometimes badly, in performance owing to model mis-specification. Here, non-parametric (or semi-parametric) models are used for various types of financial data, in particular models under certain shape restrictions may be safely enforced in practice from a financial perspective. The estimation of a uni-variate density under the log-concavity restriction has been studied. A new algorithm has been developed, which extends the constrained Newton method for fitting a non-parametric mixture model. Compared with other existing algorithm, the new one is quite fast, stable and accurate. Some applications to financial data have been given, also been described a tentative extension that adds smoothness to log-concave density estimate.

(31) OUTLIERS IN TIME SERIES - USING AGGREGATED UNIT RECORD DATA

Christopher Bryant
University of Canterbury

[S.1.05] 10:50-12:10 Monday 25 November

Most work on outliers in time series deals only with aggregate data. However, Statistics New Zealand has access to individual unit records that sum to these aggregates, meaning that they can remove very large individual unit records that lead to aggregate outliers. The method Statistics New Zealand currently uses to identify these very large individual records was developed some years ago. This talk will discuss a possibly more statistically rigorous procedure for approaching extreme unit records. The alternative procedure investigated in this project makes use of extreme value theory in its identification and removal of the large unit values that lead to aggregate outliers in time series. Using extreme value theory allows Statistics New Zealand to have a single outlier cut-off method across multiple datasets. This method will generate cut-offs that are both more stable over time, and easily revisable as economic conditions change.

(32) MODELLING RETINAL VENOUS PULSATION FROM VIDEO DATA

Brigid Betz-Stablein, Martin Hazelton and Morgan William
Massey University, Massey University, Lion's Eye Institute

[S.1.05] 14:40-16:40 Tuesday 26 November

When the foundations of modern statistics were being developing early in the 20th century, datasets were small and simple, consisting of a few variables typically collected from designed experiments. Modern day datasets are often much larger and more complex, and include functional data, genomic datasets, and image data such as satellite images. Video data is another example, and within the field of biostatistics, modelling data extracted from contiguous video clips is in its infancy. We obtained video data of retinal venous pulsation which occurs due to the change in blood volume in vessels throughout the cardiac cycle. We discuss how statistical models can be fitted to video data. Initially RGB intensity data, relating to the volume of blood, must be extracted from the video sequences. The second stage involves modelling these data as a time series, where the trend includes harmonic terms and an additional linear spline component to account for the noise created by participant movement. The model is applied to 654 video sequences from 17 participants. The associations between properties of the fitted trend curve and clinical variables are then investigated, as information from these curves, such as slope and amplitude, may have clinical importance.

(33) VISUALISING ORBITAL EFFECTS IN ICE-CORE DATA

Jean Sanderson
Massey University

[S.1.05] 9:50-10:30 Tuesday 26 November

The majority of standard time series analysis tools require that the signal is composed of observations which are evenly spaced over time. Since paleoclimate series, such as those taken from ice cores, exhibit a naturally irregular time sampling, it is not possible to apply these techniques without modification of the data to a regular grid. However in doing this, it is possible that information at higher frequencies is being lost. This talk focuses on the use of two techniques; empirical mode decomposition (EMD) and wavelet lifting, both of which can be applied directly to the irregularly observed data values. Using EMD the original series can be decomposed into a sum of Intrinsic Mode Functions (IMF) related to different periods of variation within the data. Wavelet lifting is then used to estimate the spectral components of these irregularly spaced series. Applying both methods highlights the strengths and limitations of each, suggesting interesting possible avenues for future work. Results are demonstrated with application to a composite carbon dioxide record dating back nearly 1 million years and the derived orbital forcing signal with the same sampling.

(34) BAYESIAN SEMIPARAMETRIC LIKELIHOOD APPROXIMATIONS FOR STATIONARY TIME SERIES

Renate Meyer and Claudia Kirch
University of Auckland, Karlsruhe Institute of Technology

[S.1.04] 10:50-12:10 Monday 25 November

Time series abound in many fields such as econometrics, medicine, ecology and astrophysics. While Bayesian time series analysis predominantly uses parametric models that are more efficient than nonparametric models if correctly specified but that fail otherwise, robust nonparametric methods prevail in frequentist time series. Here we will propose a new robust semi-parametric approach to Bayesian time series analysis by introducing a nonparametric correction of a parametric likelihood. We use a nonparametric Bernstein polynomial prior on the spectral density with weights induced by a Dirichlet process distribution. We show that Bayesian nonparametric posterior computations can be performed via a MH-within-Gibbs sampler by making use of the Sethuraman representation of the Dirichlet process.

(35) A FAST ALGORITHM FOR ESTIMATING A NONPARAMETRIC BATHTUB-SHAPED HAZARD FUNCTION

Shabnam Fani
University of Auckland

[S.1.01] 9:50-10:30 Tuesday 26 November

The problem of estimating a hazard function under certain shape restrictions can be appealing in practical survival analysis. In this paper, we propose a new algorithm for computing the nonparametric maximum likelihood estimate of a convex hazard function. A new support set expansion technique for the constant part of a convex bathtub-shaped hazard function is presented. The proposed technique successfully deals with the issue of picking out the estimation of a minimum of a convex hazard function. In the case of all exact observations, the proposed algorithm is remarkably faster than the support reduction and bisection algorithm (SRB) due to a single loop instead of double loop and multiple new support points inclusion rather than single inclusion in each iteration. Moreover, this new algorithm can be applied to the situation with interval-censored data and the situation with an interval-censored data mixed with exact observations. We evaluate the performance of our new algorithm via simulation studies and illustrate it on some real data sets. Numerical studies show that the new algorithm works very well in different scenarios and outperforms the SRB algorithm.

(37) STATISTICS AND THE LAWS OF PHYSICS

William Bolstad
University of Waikato

[S.1.04] 10:50-12:10 Monday 25 November

In this talk we look at statistics from the perspective of the laws of physics. We explore these ideas using the linear regression model as an example. The Bayesian approach is seen to be consistent with the laws of physics while the frequentist approach is not.

(38) A K-FOLD AVERAGING CROSS-VALIDATION PROCEDURE

Yoonsuh Jung and Jianhua Hu
University of Waikato, University of Texas MD Anderson Cancer Center

[S.1.01] 9:50-10:30 Monday 25 November

Cross-validation type of methods have been widely used to facilitate model estimation and variable selection. In this work, we suggest a new K-fold cross validation procedure to select a candidate optimal model from each hold-out fold and average the K candidate optimal models to obtain the ultimate model. Due to the averaging effect, the variance of the proposed estimates can be significantly reduced. This new procedure results in more stable and efficient parameter estimation than the classical K-fold cross validation procedure. In addition, we show the asymptotic equivalence between the proposed and classical cross validation procedures in the linear regression setting. We also demonstrate the broad applicability of the proposed procedure via two examples of parameter sparsity regularization and quantile smoothing splines modelling. We illustrate the promise of the proposed method through simulations and a real data example.

(39) METHODS FOR DETECTING ANOMALIES IN ROUTINELY RECORDED ANIMAL HEALTH DATA, WITH PARTICULAR REFERENCE TO FOOT-AND-MOUTH DISEASE IN VIETNAM

Kate Richards, Mark Stevenson, Martin Hazelton, Cy Lockhart, J Pinto and L Nguyen
IFS, Massey University, IVABS, Massey University, IFS, Massey University, Food and Agriculture Organization of the United Nations, Food and Agriculture Organization of the United Nations, Department of Animal Health

[S.1.05] 14:40-16:40 Monday 25 November

The widespread availability of computer hardware and software for recording and storing disease event information means that, in theory, we have the necessary raw materials required to carry out detailed analyses of factors influencing the spatial distribution of disease in animal populations. This said, the reliability of such analyses depends on data quality, with anomalous records having the potential to introduce significant bias, leading to inappropriate decision making. In this paper we present an approach for detecting anomalies in routinely recorded animal health data. Details of outbreaks of foot-and-mouth disease (FMD) in Viet Nam for the period 2006 to 2008 are used as a case study. Details of 2734 individual commune-level outbreaks of FMD that occurred in Viet Nam for the period 1 January 2006 to 31 December 2008 were obtained from the Vietnamese Department of Animal Health. A binomial model was developed where the response was the number of FMD-positive communes per province. A province-level random effect term on the logit scale was used to account for heterogeneity in the data due to unobserved effects. The models were fitted within the Bayesian paradigm using Markov chain Monte Carlo methods. By considering the distribution of the random effect terms we computed an exceedance probability for each province, allowing us to identify provinces where the observed frequency of FMD differed substantially from that which was expected. Our inference here is that these provinces require further investigation and are likely to be those where disease events have been under-reported.

(40) EM ALGORITHM AND GIBBS SAMPLING: COMPARE AND CONTRAST

Murray Jorgensen
Laplace Associates

[S.1.02] 14:40-16:40 Tuesday 26 November

Using one or more case studies I will outline the similarities and differences between the EM Algorithm and Gibbs Sampling for parameter estimation in complex models. Gibbs is used rather than Metropolis-Hastings because the conditional distributions used in both may be the same or related. The details of making the comparison can help with the understanding of both methods. For example the "missing data" in the EM formulation translate to internal nodes in the BUGS graph of a model. A practical motivation for this study comes from the difficulty of fitting complex models with multimodal likelihoods or posteriors. Puzzling results from one of the methods may be set in context by considering the other.

(41) AN OPTIMIZATION MODEL FOR STRATEGIC FLEET PLANNING IN TRAMP SHIPPING

Jørgen Laake and Abraham Zhang
Kristian Gerhard Jebsen Skipsrederi AS, University of Waikato

[S.1.02] 9:50-10:30 Monday 25 November

Maritime transportation is one of the most capital intensive industries. Fleet planning is vital but challenging to shipowners because the industry is extremely volatile. This paper presents a mixed integer programming optimization model for strategic fleet planning in tramp shipping. The model determines the best mix of long term and spot contracts for a given fleet, finds the optimal fleet size and mix for a set of contracts or a mix of both. The model can be used as a basis for a fleet renewal program, helping to decide when to sell and whether to buy old or new ships. It also takes into consideration the time charter market, recommending when to charter in/out vessels. Another area of application is for users that are only engaged in active vessel trading and not in transportation. A numerical example is given to illustrate how to use the model to evaluate different operations strategies.

(42) LEAN AND SIX SIGMA IMPLEMENTATION IN THE LOGISTICS INDUSTRY: A SNAPSHOT IN SINGAPORE

Abraham Zhang, Chia Ting and Xavier Sim
University of Waikato, SIM University, SIM University

[S.1.02] 9:50-10:30 Monday 25 November

Companies must continuously improve their operations to stay competitive. Lean and Six Sigma have been widely adopted in the manufacturing industry to reduce cost and improve speed and quality.

However, limited research has been conducted about their implementation in the logistics industry. This study seeks to understand the current status, and investigates factors affecting Lean and Six Sigma implementation in the logistics industry. In 2012, we conducted an on-line survey and telephone interviews in Singapore, a major logistics hub in Asia. 37.5% of respondents implemented Lean or Lean Six Sigma because they consider cost efficiency and speed as key factors to stay competitive. The remaining respondents did not implement Lean or Six Sigma mainly because they were unsure or did not hear of such methodologies. Among the Lean adopters, most commonly used Lean techniques are 5S, Value Stream Mapping and Kanban. Cost Savings stands as the top benefit gained, followed by Improved Delivery Performance, Reduced Process Cycle Time and Increase in Employee Productivity Rate.

Major implementation challenges were resistance to changes and the failure of employees to understand the rationale in such implementations. The chi-square tests suggest that large firms are more likely to adopt Lean or Lean Six Sigma in comparison with small and medium sized enterprises. There is a significantly higher likelihood for express transport service providers to employ Six Sigma together with Lean, instead of Lean alone, probably due to their high service quality standards.

(43) MODELLING NEW ZEALAND BUILDING ACTIVITY FROM BUILDING CONSENTS DATA

Sonya McGlone
Statistics New Zealand

[S.1.05] 9:50-10:30 Wednesday 27 November

Statistics New Zealand is implementing a programme of work to transform the organisation and achieve the statistical system of the future. An important goal in this transformation is to become an administrative data first organisation. A key strategy is to use more administrative data in outputs. The Value of Building Work Put in Place survey, which measures building activity in New Zealand, is being redesigned from a survey based collection to a collection based primarily on modelled administrative data. Building consents data is received monthly from all territorial authorities throughout New Zealand and will become the primary data source for the modelling approach. The survey currently posts out approximately 15,000 paper forms per year, and the redevelopment aims to substantially reduce this with the use of administrative data. This talk will focus on the proposed modelling methodology, the results of the feasibility investigations, and the subsequent modelling recommendations. I will also discuss the implications of implementing a model based approach, such as the limitations of using administrative data, and the challenges of measuring the quality of the estimates.

(44) AUTOMATIC CONVERSION OF TABLES TO LONGFORM DATAFRAMES

Jimmy Oh
University of Auckland

[S.1.04] 10:50-12:10 Tuesday 26 November

TableToLongForm is an R package that automatically converts hierarchical Tables intended for a human reader into a simple LongForm Dataframe that is machine readable, hence enabling much greater utilisation of the data. It does this by recognising positional cues present in the hierarchical Table (which would normally be interpreted visually by the human brain) to decompose, then reconstruct the data into a LongForm Dataframe. This talk will motivate the benefit of such a conversion with an example Table from Statistics New Zealand, and follows with a short user manual. The talk will then explore the types of Tables the package can convert with a gallery of some of the recognised patterns and structures.

(45) PREDICTIVE RISK MODELING FOR HOSPITAL READMISSION

Claire Forsythe and Chaitanya Joshi
Waikato District Health Board, University of Waikato

[S.1.05] 9:50-10:30 Monday 25 November

I am developing a model for the Waikato District Health Board (Waikato DHB) which will predict patients that are at high risk of readmission after an acute episode in hospital. This will allow the DHB to identify patients upon admission or discharge that have a high probability of returning to hospital acutely within 28 days of discharge. This can potentially reduce hospital costs as instead of a readmission to hospital, high risk patients may be treated in the community at a lower cost than an inpatient bed day. This will also benefit the patient as they do not need to return to hospital. Inpatient hospital data will be sourced from the DHB from July 2010 at an individual patient level. I will use Generalised Linear Models to build relationships between the predictor variables (such as demographics, hospital specialty and hospitalisation history etc) and the response variable, hospital readmission (0 for no readmission, 1 for a readmission). I will identify the significant variables that contribute the most to the prediction of a readmission to get risk probabilities for historic patients. When a patient is admitted to hospital we can determine their risk score and determine whether an intervention needs to occur.

(46) THE END POINT OF TIME SERIES: REPORTING INFORMATIVELY ON TRENDS

Giles Reid and John Créquer
Statistics New Zealand

[S.1.01] 10:50-12:10 Wednesday 27 November

National Statistics Offices (NSOs) face many constraints and difficulties when reporting time series movements, and the guidelines used by different agencies frequently conflict. Trend movements have some advantages in comparison to seasonally adjusted movements or year-on-year comparisons, but the revisions that are made to the end point of a trend series as new data become available can undermine published interpretations and pose a risk to NSOs' reputations as trusted and unbiased sources. The trade-off between the risks posed by revisions and the early detection of shifts and turning points is made particularly difficult at Statistics NZ because the subject experts responsible for the releases are not necessarily technical experts in time series, and our production cycles do not allow much time for extensive modelling or analysis of series movements. Recent developments at Statistics NZ include a new sensitivity analysis tool to help our subject experts to make reliable and informative statements about trend series.

(47) BAYESIAN MODELS FOR THE SOURCE ATTRIBUTION OF ZONOOSES

Poppy Miller
Massey University

[S.1.04] Poster Session: 16:40-18:00 Monday 25 November

Zoonoses figure very prominently in the list of notifiable diseases in New Zealand. In many zoonotic diseases that are important to human health (such as *Campylobacter*), it is difficult to attribute human cases to sources because there is little epidemiological information on the cases. Genotyping systems allow the zoonoses to be categorised, and the relative distribution of the genotypes among the sources (food sources or reservoirs of bacteria) and on the human cases allows inference on the likely source of each genotype. The Hald model is used to estimate parameters for the relative effect of each source and type on the rates of disease. Source effects account for the ability for the source to act as a vehicle of infection; type effects account for a survival or virulence mechanism. Identifiability of the parameters in this model is an issue compounded by data imbalance and low counts for many of the combinations of source and type. We present techniques to overcome these issues within a Bayesian framework, and apply them to *Campylobacter* data from the Manawatu.

(48) A VIEW OF DATA SCIENCE

Donal Krouse
Callaghan Innovation

[S.1.05] 14:40-16:40 Monday 25 November

Data Science is an emerging discipline that can loosely be described as a combination of the best and worst parts of computer science and statistics. Using modern computers, giga, tera and now peta-bytes of data can be generated, filtered and visualised. The challenge for Data Science is to extract information and ultimately value from unprecedented volumes of data - so-called "Big Data". In this talk I present an overview of data science activities and give my own view as a statistician with 30 years experience of 'real data'.

(49) TRAIT-BASED ECOLOGICAL COMMUNITY ASSEMBLY USING A HIERARCHICAL BAYESIAN MODEL

Paul T. Brown, Chaitanya Joshi and Daniel C. Laughlin
*University of Waikato, University of Waikato - Department of Statistics, University of Waikato -
Department of Biological Sciences*

[S.1.05] 14:40-16:40 Tuesday 26 November

The theory of community assembly involves two seemingly opposing processes which influences a plants functional traits in contrasting ways. Environmental filtering suggests that species with similar traits are likely be found in similar environments, leading to the convergence of trait values within communities. However, competition for scarce resources limits the functional similarities of co-occurring species, thus promoting species coexistence. The result of limiting similarity is the divergence of trait values within communities. Assembly theory suggests that individuals are sorted along environmental gradients as their functional traits influence their performance. Recently, Laughlin et al. (2012) have developed Traitspace, a predictive hierarchical Bayesian model that can incorporate both environmental filtering and limiting similarity to predict the relative abundance of tree species in a given ecosystem. The predictive framework is generalizable to any ecosystem, as it can accommodate any species pool and their traits along several environmental gradients. It has been used for the analysis of relative abundance of trees in both Arizona, USA, and Franz Josef, NZ. This talk will discuss the findings of this model to a newly collected dataset from Puketi, NZ, as well as possible extensions to Traitspace.

(50) MODELING FOR CARRYOVER EFFECTS IN CLINICAL TRIALS

S Gwynn Sturdevant and Thomas Lumley
University of Auckland

[S.1.01] 10:50-12:10 Tuesday 26 November

Recently, trials have been used to evaluate pharmaceutical treatments and their ability to delay the onset of diabetes and hypertension. Criticism of one trial focuses on a simulated 80% Type I error rate. Diagnosis of hypertension occurs when a noisy measurement exceeds a threshold, and diagnosis results in treatment, which censors subsequent blood pressure measurements. Approaches to design were studied by simulation using simple comparison of cumulative incidence as the analysis. As an approach to analysis, we studied mixed models treating post-diagnosis measurements as data missing at random. Neither design strategy reliably controls Type I error. The mixed-model analysis approach does control Type I error and give unbiased estimates of treatment effect. Carryover effects on incident hypertension or diabetes require care in design and analysis. We recommend an analysis based on mixed models.

(51) ROBUSTIFYING THE GEOGRAPHIC RELATIVE RISK FUNCTION

Khair Jones and Martin Hazelton
Massey University

[S.1.01] 14:40-16:40 Monday 25 November

In the analysis of the geographic distribution of disease, a popular method to estimate risk for data in the form of spatial coordinates of cases and controls is to use the spatial relative risk function, which is the log ratio of the case and control densities estimated by bivariate kernel smoothing. One issue with this method is that the risk estimate may be very unstable in areas of sparse population density. This can be seen as analogous to the problem of low cell counts in contingency tables. Our proposed solution involves adding pseudo data to the control and case densities, with consideration given to how much pseudo data should be added, and where. We analyse the use of pseudo data through simulations, before looking at some real world applications, including the famous Chorley Ribble cancer dataset.

(52) EXTENSION OF CLINICAL TRIALS TO POPULATION HEALTH STEPPED WEDGE DESIGN IN RANDOMISED CONTROLLED TRIALS

Yannan Jiang
University of Auckland

[S.1.02] 10:50-12:10 Monday 25 November

Randomised controlled trials (RCTs) are considered the gold standard in determining effectiveness of interventions in clinical trials. Traditional individual parallel designs are, however, difficult to implement in evaluation of public health interventions for logistic, financial or ethical reasons. Stepped-wedge design in RCTs involves sequential roll-out of an intervention to individuals in groups over a fixed number of time periods, and is widely applied in cluster randomised trials (CRTs). The order in which a proportion of clusters switch from control to intervention is determined randomly by sequences. All clusters receive the intervention eventually and data are collected repeatedly from individuals at each time period. In this talk, systematic reviews on the applications of stepped wedge design will be summarised and practical considerations for implementing evaluations on population health will be discussed. A case study recently conducted in 14 primary schools in New Zealand (the BISKIT trial) will also be presented.

(53) NETWORK META-ANALYSIS FOR MULTI-ARM TRIALS

Thomas Lumley
University of Auckland

[S.1.02] 10:50-12:10 Monday 25 November

Network meta-analysis allows for results of trials to be combined when the trials do not all compare the same pair of treatments, with estimates of inconsistency between direct and indirect information. The standard approaches to network meta-analysis do not work straightforwardly for multi-arm trials and some ad hoc adjustments have been needed. I will describe an approach using hypergraphs rather than graphs to represent the network and show that it provides a simple mixed-model analysis and avoids the problems that multi-arm trials pose for graph-based network meta-analysis.

(54) EXPLORING MULTIPLE OMICS DATASETS

Vanessa Cave, Marlon Reis, Paul Maclean and Kuljeet Singh
AgResearch

[S.1.05] 10:50-12:10 Wednesday 27 November

Technological advances have resulted in increasing amounts of high-throughput omics data being generated across the entire omics cascade, from DNA to metabolites. Analysis of a single omics dataset can shed light on the genetic factors that help determine the phenotype of an individual. However, increasingly data from multiple omics sources is available on an individual, and a greater understanding of the underlying biological system can often be achieved by conducting a joint analysis. In this talk we apply dimension reduction methods with the aim of extracting potentially useful information from, and revealing interesting relationships between, high-throughput omics datasets. The datasets explored all stem from agricultural research being carried out at AgResearch.

(55) CLUSTER ANALYSIS FOR LONGITUDINAL ORDINAL DATA: A LIKELIHOOD-BASED APPROACH BASED ON FINITE MIXTURE MODELS

Roy Costilla and Ivy Liu
Victoria University of Wellington

[S.1.01] 10:50-12:10 Monday 25 November

Traditional cluster analysis methods are not based on likelihoods and thus the toolbox of statistical inference can't be used for them. For instance, goodness of fit can't be assessed using well-developed criteria such as the AIC. Following the lines of Pledger and Arnold 2013, we develop likelihood-based probability models to cluster ordinal data that arises in longitudinal settings. In this talk, we present the Trends Odds Model (TOM) that includes mixture-based fuzzy clustering to identify similar groups. The estimation procedure is carried out using the EM algorithm. We also illustrate our models using survey data from the Household, Income and Labour Dynamics in Australia (HILDA). In particular, we examine self-reported health status (poor, fair, good, very good and excellent) from the 2001-2010 waves.

(56) DECOMPOSING OBSERVED VOLATILITY IN FINANCIAL SERIES USING QUALITY CONTROL THEORY

Nadeeka D. Premarathna, A. Jonathan R. Godfrey and K. Govindaraju
Institute of Fundamental Sciences, Massey University, Institute of Fundamental Sciences, Massey University, Institute of Fundamental Sciences, Massey University

[S.1.04] 14:40-16:40 Monday 25 November

The quality control technique of rational subgrouping is applied to financial series in order to generate two new measures of volatility. These measures are specifically targeted for separate use by short term traders and longer term investors. Walter Shewhart (1931) proposed that variation in any process can be attributed to either common or special causes by dividing data into rational subgroups. Common cause variation is the underlying variation that is always present in a process, while special cause variation is due to events that are more one-off" in nature. We find, using a wide range of U.S. equity market series, that long term investors would reach quite different conclusions regarding risk if they focused on our measure rather than total variation. We also suggest that simply using lower frequency data does not remove noise from the data as accurately as the technique of rational subgrouping. Further, we find that the periods of time with high volatility are drivers of the non-normality of stock return series.

(57) AN R PACKAGE FOR OPTIMAL PRICING OF PERISHABLE ASSETS

Rolf Turner
AgResearch

[S.1.04] 14:40-16:40 Monday 25 November

I will discuss an R package for determining the optimal price of an asset (item of goods) which is "perishable" in the sense that its value drops to 0 after a deadline has passed. Required inputs include the intensity of customer arrivals at the point of sale and a time-varying price sensitivity function. The latter specifies the probability that a customer will purchase an asset offered at a given price at a given time. I will (very briefly) outline the underlying theory, which leads to a coupled system of differential equations that can be solved numerically. The "deSolve" R package can be called upon to effect the solution. Despite some intricate aspects of the system, the deSolve package can be invoked in an amazingly seamless manner. I will give a few illustrative examples, including one which involves piecewise linear price sensitivity functions. These provide a means by which "informal knowledge" about customer behaviour may be turned into a formally specified function. This makes the proposed technique reasonably simple for users to apply to real problems and thus to produce genuinely useful results in a practical and convenient manner.

(58) VIDEO GAMES AND MCMC: PARALLEL BAYESIAN COMPUTATION USING GPUS

Chris Jewell
Massey University

[S.1.04] 10:50-12:10 Tuesday 26 November

Recent developments in MCMC for Bayesian inference on dynamical models have provided a tantalising vision for decision support in a wide range of applications. In a real-time setting, providing answers in a short time frame is critical to the utility of these methods, though this is often at odds with the computational cost of complex likelihood functions on large datasets. In the post-Moore's Law era, parallel computing has become the standard approach to achieve significant improvements in algorithmic runtime. However, the serial nature of MCMC presents challenges to its implementation on traditional CPU-based architectures. For example, inter-process communication limits scalability for within-chain parallelisation, whereas the length of the burn-in period imposes a lower bound on the runtime of a multiple-chain approach. A promising solution exists in leveraging general-purpose graphics processing unit (GPGPU) technology. GPGPUs may act as co-processors to accelerate the most computationally intensive stages of a MCMC algorithm, combining massive parallelisation with fast communication compared to CPU-based systems. This talk will describe such an approach to real-time Bayesian inference for epidemic models, using Nvidia's CUDA C++ library. Intensive likelihood calculations are delegated to the GPGPU, whilst the CPU directs an overall serial MCMC scheme. Using this technique, a highly scalable parallel algorithm is achieved at a fraction of the hardware cost of the equivalent CPU-based machine.

(59) BAYESIAN RISK PREDICTION FOR EPIDEMIC DECISION SUPPORT

Chris Jewell
Massey University

[S.1.02] 10:50-12:10 Wednesday 27 November

Dynamical models have become a standard tool in epidemiology for assessing the behaviour of an epidemic in response to population demographics and intervention strategies. Agent-based stochastic Poisson process formulations of these models enable the assimilation of a wide range of data sources such as contact networks and the spatial locations of the agents involved. Techniques for simulating epidemic outcomes from such models, conditional on the parameters, is well-characterised. Parameter estimation, however, presents a considerable statistical challenge due to the censoring of individuals' infection times -- cases are only ever detected some time after infection. A Bayesian approach to this problem presents several advantages, chief among which is the power of MCMC data-augmentation techniques, treating the censored observations as latent data. However, weak dependency between the data and process model, and strong dependency between the latent data and parameters, means that standard MCMC algorithms often perform poorly. To address this problem, non-centred (or ancillary) re-parameterisations help by rearranging the dependency structure in the Markov chain state space. Using the example of the UK foot and mouth disease epidemic in 2001, this talk will present some of the issues that arise with respect to efficient MCMC for epidemic models, and show how non-centered parameterisations help to orthogonalise the posterior distribution and improve MCMC efficiency.

(60) APPLYING REAL OPTIONS APPROACH TO DETERMINE THE OPTIMAL VACCINATION TIMING AND THRESHOLD

Elena Moltchanova, Zuzana Chladna and Michael Obersteiner
University of Canterbury, University of Bratislava, IIASA

[S.1.05] 10:50-12:10 Tuesday 26 November

Epidemics are an ever-present danger and although preventive vaccination helps reduce the disease pool considerably, the decision maker is often faced with the choice of buying the vaccine early and perhaps wasting money and losing political credibility or buying late, when more information is available, but when the valuable initiative has been lost. Although guidelines exist, recognising an epidemic and deciding when to take action may still be problematic. We apply a Real Options approach to the standard Susceptible-Infected-Recovered (SIR) epidemic model to develop a dynamic decision-making framework. Unlike the Present Value Optimisation, Real Options approach takes into account the cost of waiting and the knowledge gained by waiting, making it an exciting and valuable tool for real-time strategic optimisation. The value of information is thus explicitly taken into account. We demonstrate some simulation studies and discuss the potential challenges associated with the implementation of the approach. We also demonstrate the sensitivity of the analysis to exogenous parameters and present the possibility of a dynamically learning decision making system.

(61) JOINTLY OPTIMAL BANDWIDTH SELECTION FOR THE PLANAR KERNEL-SMOOTHED DENSITY-RATIO

Tilman Davies
University of Otago

[S.1.02] 10:50-12:10 Wednesday 27 November

The kernel-smoothed density-ratio or 'relative risk' function for planar point data is a useful tool for examining disease rates over a certain geographical region. Instrumental to the quality of the resulting risk surface estimate is the choice of bandwidth for computation of the required numerator and denominator densities. The challenge associated with finding some 'optimal' smoothing parameter for standalone implementation of the kernel estimator given observed data is compounded when we deal with the density-ratio per se. To date, only one method specifically designed for calculation of density-ratio optimal bandwidths has received any notable attention in the applied literature. However, this method exhibits significant variability in the estimated smoothing parameters. In this work, the first practical comparison of this selector with a little-known alternative technique is provided. The possibility of exploiting an asymptotic mean integrated squared error formulation in an effort to control excess variability is also examined, and numerical results seem promising.

(62) MANOVA FOR HIGH-DIMENSIONAL DATA

Insha Ullah
Massey University

[S.1.01] 14:40-16:40 Tuesday 26 November

The traditionally available multivariate analysis of variance (MANOVA) tests such as Wilks lambda and Pillai-Bartlett trace start to suffer from low power as the number of variables approaches the sample size. Moreover, as the number of variables exceeds the number of available observations, these statistics are not available for use. Regularized alternatives have been proposed which not only allow the use of MANOVA in high-dimensional situations but also exhibit more power when the sample size approaches the number of variables. In this paper these alternative approaches have been explored together with a novel approach based on lasso regularization. The performance of different methods has been explored via an extensive simulation study. The MANOVA with lasso regularization maintained good performance in terms of power of the test over the simulation study. The methods are also applied to real data set of soil compaction profiles at various elevation ranges.

(63) PARALLEL SEQUENCING OF BARCODED BACS USING NGS TECHNOLOGY

Marcus Davy, Roy Storey and Elena Hilario
Plant and Food Research, Plant and Food Research, Plant and Food Research

[S.1.04] 10:50-12:10 Wednesday 27 November

Genome assembly connects large scaffold via mate paired ends encompassing an insert size between 3-20 kbp. Larger insert ends can be generated by sequencing BAC or fosmid clones ends by Sanger sequencing, a technique that in comparison to next generation sequencing methods is labour intensive and expensive. Thousands of BAC and fosmid end sequence pairs are required to assist the assembly process. We are sequencing the genome of kiwifruit, *Actinidia chinensis*, by a combination of different NGS insert size libraries, single ends, paired ends and mate paired ends, by Illumina and Roche 454 platforms. We also have a genomic DNA BAC library of kiwifruit, containing 69120 BAC clones ordered in 180 384-well plates. The library is predicted to have 7x genome coverage, with an average insert size of 71 kbp. This talk presents results from a second experiment randomly tagging 30 plates of BAC plasmids with barcoded oligonucleotides at the end of a random hexamer by PCR for 11,500 BACs. In order to aid sequence assembly we are looking for situations where each BAC is encapsulated within two assembled scaffolds in order to join scaffolds, however if information is incomplete, a BAC can span across many scaffolds. This additional information aims to enhance and validate the kiwifruit genome assembly by aligning BAC fragments to kiwifruit scaffolds, with a further goal to reduce the number of scaffolds in the assembly.

(64) DEVELOPING STUDENTS UNDERSTANDING OF STATISTICAL INFERENCE: A RESEARCH PROJECT

Maxine Pfannkuch and Stephanie Budgett
University of Auckland

[S.1.04] 9:50-10:30 Monday 25 November

In this presentation we report on a research project conducted in Year 13, introductory university and workplace classes. The project was a two-year New Zealand wide collaboration among three statisticians, two researchers, sixteen Year 13 teachers, seven university lecturers, one workplace practitioner, and three teacher professional development facilitators. The project team designed innovative computer-based dynamic visualization approaches involving bootstrapping and randomization methods to develop students inferential reasoning. The research sought evidence that these innovations were effective in developing students reasoning and understanding of statistical inference. We discuss the major findings and teaching implications from this project.

(65) DETERMINING THE SOURCE OF HUMAN CAMPYLOBACTER CASES THROUGH TIME

Jonathan Marshall
Massey University

[S.1.02] 10:50-12:10 Wednesday 27 November

Until recently, New Zealand had the highest per-capita incidence rate of campylobacteriosis in the world and, despite recent efforts, notification rates are still significantly higher than in comparable countries. This is not only a significant public health issue, but also an important economic issue, as it is estimated that the 7,000 cases currently notified represent direct and indirect costs to New Zealand of approximately \$32 million dollars per annum. For the majority of these cases there is little epidemiological information, making it difficult to determine the likely source or reservoir of infection. Genotyping techniques, such as Multi-locus sequence typing, however, allow each case to be categorised, allowing assignment of human cases to the most likely source based on the distribution of genotypes on those sources. Campylobacteriosis notifications are highly seasonal, with some evidence that the incidence of specific genotypes is also seasonal. Further, an intervention targetted at the poultry industry in 2007 correlated with a decline in human cases. Thus, it is expected that the proportion of cases attributed to each source changes through time. A Bayesian model for the attribution of Campylobacter cases through time will be presented, with application to cases from the Manawatu.

(66) INCOMPLETELY OBSERVED POINT PROCESSES IN TIME: AN APPLICATION TO VOLCANIC ERUPTION RECORDS

Ting Wang and Mark Bebbington
University of Otago, Massey University

[S.1.05] 10:50-12:10 Tuesday 26 November

Missing data analysis for point processes is much less developed than for regression, mainly due to technical difficulties in dealing with dependent data. Point processes are nowadays frequently used to estimate volcanic hazards, where historical records are often incomplete. We examine the problem of estimating the true (adjusted for missing observations) parameters and hence the hazard in a Weibull/gamma renewal model. Robust regression, robust estimation using repeated medians, and results from the theory of inverse of thinned renewal processes failed to provide consistent estimates from simulated data. Hence we explore a hidden Markov model framework, where the hidden state is a reflection of the number of missing events. This also allows for the completeness level of the record to be estimated, and offers a means of determining where in the record the missing observations are likely to be found. We apply this method to data from the Holocene (last 12,000 years) record of Mt Taranaki, the preliminary estimates of completeness are 86.87% complete (record 5000BC - present) and 78.80% complete beyond that. These figures were independently verified using a model of tephra dispersal. The estimated present hazard is approximately 20% higher than estimated without allowing for missing data.

(67) EMPIRICAL BAYES ESTIMATION FOR THE STOCHASTIC BLOCK MODEL

Shakira Suwan, Dominic Lee and Carey E. Priebe
University of Canterbury, University of Canterbury, Johns Hopkins University

[S.1.02] 14:40-16:40 Tuesday 26 November

Network models are increasingly used to model datasets that involve interacting units, especially random graph models where the vertices represent individual actors and the edges represent the presence or absence of a specified connection between actors. Here we present an empirical Bayes method to estimate block membership of vertices in a random graph generated by a stochastic block model, with particular focus on using a random dot product graph, which is a type of latent position model. An adjacency spectral embedding of a random graph is used to obtain an empirical Bayes prior for the latent positions. Inference with the model is performed using Markov chain Monte Carlo and performance is illustrated using a simulation study.

(68) THE FIXED-EFFECTS MODIFIED-SPLICE (FEMS) METHOD: FULLY QUALITY-ADJUSTED PRICE INDEXES WITH NO CHARACTERISTICS INFORMATION

Frances Krsinich
Statistics New Zealand

[S.1.01] 14:40-16:40 Tuesday 26 November

Statistics New Zealand has been investigating methods for estimating price indexes from both scanner and online data where comprehensive information on characteristics is not available. Matched-model methods can be biased as they do not reflect the implicit price changes associated with the introduction and disappearance of products. With an increasingly dynamic market, in terms of both technological change and price-setting strategies, this potential bias is likely to become more significant over time. This paper shows that the relatively simple fixed-effects, or time-product dummy, index is equivalent to a saturated time dummy hedonic index based on all price-determining characteristics of the products, despite those characteristics not being observed. In production this can be combined with a modified approach to splicing that incorporates the index across the full estimation window to reflect new products with one periods lag without requiring revision.

(69) RANDOMIZATION AND APPROXIMATION IN MCMC

Matthew Parry
University of Otago

[S.1.04] 10:50-12:10 Monday 25 November

Randomization and approximation within an MCMC step can greatly improve targeting efficiency. I give a brief review of current developments in these areas and outline how randomization and approximation can be made to work. Randomizing the proposal step is already a feature of a number of sampling methods, e.g. Hamiltonian MCMC. It is also possible to randomize the acceptance probability, as shown by Nicholls et al. (2012). This gives an explanation as to why round-off error is not fatal to MCMC. Approximation methods, on the other hand, suggest themselves when evaluation of the target is difficult or even intractable. Recent work includes improved exact-approximate methods and methods that use full precision only when proposals based on approximations are accepted. Finally, I will discuss generalized Bayesian updating.

(70) DRAWING (PIECEWISE) SMOOTH CURVES

Ross Ihaka
University of Auckland

[S.1.04] 14:40-16:40 Monday 25 November

When drawing figures, it is often important to include the graph of a piecewise smooth function. It is generally straightforward to do this "by hand" but having a more automated procedure can simplify the process. In this talk I'll look at how such an automatic procedure can be constructed. This will involve short digressions into graphical perception and patent law.

(71) EXTREME VALUE MIXTURE MODELS - AN R PACKAGE AND SIMULATION STUDY

Yang Hu and Carl Scarrott
University of Canterbury

[S.1.05] 10:50-12:10 Monday 25 November

Extreme value theory is used to develop asymptotically motivated models for describing the likelihood of rare events occurring. Such models are typically used to approximate the behavior of the tail(s) of the population distribution. An important challenge in the application of such extreme value models is the choice of a threshold, beyond which point the asymptotically justified extreme value models can provide a good extrapolation to the tail(s). Various approaches have been developed to aid the selection of this threshold, all with their own advantages and disadvantages. Extreme value mixture models are one such approach, combining a suitable model for the bulk of the distribution below the threshold along with an extreme value tail above it. The threshold is usually treated as a parameter to be estimated. A review of such mixture models and other threshold estimation approaches is provided by Scarrott and MacDonald (2012). We have created an R package called *evmix*, available on CRAN, which implements most of the existing mixture models in the literature. This talk will demonstrate usage of the package and will show the results of simulation studies comparing the performance of the generalised mixture models.

(72) ON FRAGMENTATION PROCESSES AND CONNECTIONS TO DIVERSITY PROBLEMS

Estate Khmaladze
Victoria University of Wellington

[S.1.05] 10:50-12:10 Tuesday 26 November

Consider a big block of mass, which is fragmented into some number of smaller blocks in arbitrary proportions. Each smaller block is fragmented similarly, so that after some number of steps we will end up with a very large number, say, N , of fragments or grains, which practically are sand. Surprisingly, we know not only the distribution of sizes (or masses) of these fragments, but also large deviation results for this distribution. Now imagine that each of the fragments emits a Poisson process of some sort of signals. The intensity of emission equals to the mass, which is very small, times N , which is very large. We observe these Poisson processes up to some fixed time. What will this brush of Poisson processes form? How many of fragments will emit no signal? or one signal? or any given k number of signals? Are these numbers just arbitrary random quantities or do they follow some probabilistic law? We show that indeed they follow a law, we show which law do they follow, and that this law is very close to what we observe in many diversity studies.

(74) COLLECTING DATA AND ESTIMATING TRANSMISSIBILITY AND SEVERITY IN THE EARLY STAGES OF AN EPIDEMIC

Joshua Ross
University of Adelaide

[S.1.02] 10:50-12:10 Wednesday 27 November

During the early stages of an outbreak of a novel infectious disease, it is common for governments to collect disease data at the household level in order to assess the transmissibility and severity of the disease. In this talk, I will describe a project in which we are developing novel statistical estimation methods to estimate transmissibility and severity, and are using these methods in combination with a micro-simulation model to evaluate various data collection protocols. I will provide some detail on the estimation methodologies we have developed.

(75) ON THE DETERMINATION OF VARIANCE CHANGE POINTS, WITH APPLICATION TO WAVELET ANALYSIS

Leigh Roberts
Victoria University

[S.1.01] 10:50-12:10 Wednesday 27 November

Simple and intuitive non-parametric methods are provided for estimating variance change points for time series data. Only slight alterations to existing open-source computer code applying CUSUM methods for estimating break points are required to apply our proposed techniques. Our approach, apparently new in this context, is first to define two artificial time series of double the length of the original by reflective continuations of the original: first in advance, then in arrears. We then search for break points forwards and backwards through each of these symmetric extensions to the original time series. A novel feature of this paper is that we are able to identify common break points for multiple time series, even when they collect data at different frequencies. In particular, our methods facilitate the reconciliation of break point outputs from the two standard wavelet filters. Simulation results in this paper indicate that our methods produce accurate results for time series exhibiting both long and short term correlation; and we illustrate by an application to Citigroup stock returns for the last thirty years.

**(76) MODELLING ECOLOGICAL COMPETITION AND DISPERSAL IN A STATISTICAL
PHYLOGEOGRAPHIC FRAMEWORK**

Louis Ranjard, David Welch and Stephane Guindon
University of Auckland

[S.1.04] 10:50-12:10 Wednesday 27 November

Extensive genetic sampling of individuals is becoming common in ecological studies. Phylogenetic reconstructions from such data are used to assess genealogies and estimate evolutionary histories. With greater number of genetic markers and sampled individuals, more information becomes available to understand histories of colonisation of the landscape. Competition between organisms can influence the processes governing the colonisation of new habitats. For example, species or populations arriving first at a suitable location may prevent secondary colonisation. Despite the potential of competition to maintain populations in isolation, past quantitative analyses of evolution have largely ignored that factor because of technical difficulties. A model of colonisation where the probability of migrating depends on the geographical distance and whether the destination location has been colonised in the past will be presented. The model proposes to represent competition as a force preventing secondary colonisation events. Using a Bayesian phylogeographic framework, we aim at estimating the parameters of the model given the localisation of sampled individuals and their genealogy. We simulate migration histories in a two-dimensional finite space and use our approach to estimate the model parameters. Results from these simulations will be presented.

(77) SIMULATING CYTOLOGY WITH FLEXSIM HC

Ploy Pongjetanapong, Michael O'Sullivan and Cameron Walker
University of Auckland

[S.1.04] 9:50-10:30 Tuesday 26 November

We present a simulation of the Department of Cytology in LabPLUS (ADHB), built in FlexSim HC. The tasks carried out by cytologists include emergency callouts, regular (scheduled) offsite clinics, and sample analysis. We present the findings of our analysis, but our main focus is the reporting of our experience modeling with FlexSim HC. We discuss the model building process in this environment, and compare the ease of model development to that of a previous Cytology model built in Stochastic Simulation in Java.

(78) IDENTIFICATION OF SYNTHETIC LETHAL GENE PAIRS FOR TARGETED CANCER THERAPIES

Mik Black
University of Otago

[S.1.04] 10:50-12:10 Wednesday 27 November

In the field of cancer research, The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) projects provide researchers with access to genomic, proteomic, clinical and treatment data from many thousands of individuals. While a number of web-based interfaces have been developed to allow interaction with these data, to date these have had relatively limited functionality, particularly when one is interested in combining data generated on different genomic platforms. This talk will provide an overview of the data available via the TCGA and ICGC, and will demonstrate methods for the retrieval and analysis of these data in R. In addition, use of these data for identifying potential "synthetic lethal" gene partnerships will be highlighted, in the context of developing targeted therapies for cancer.

(79) LONGITUDINAL ANALYSIS OF CHILD BEHAVIOUR IN THE PACIFIC ISLANDS FAMILIES STUDY

Steve Taylor
AUT University

[S.1.01] 10:50-12:10 Monday 25 November

The Pacific Islands Families Study is a longitudinal study focussing on the health and development of a cohort of Pasifika children and their parents. The initial cohort (N=1398) was selected from births in 2000 at Middlemore hospital where at least one parent identified with Pacific Islands origins and was a New Zealand resident. Maternal report of the child's behaviour was measured using the Child Behavior Checklist at interviews as the child reached the ages of two, four and six years. Problem behaviour at a clinically relevant level was indicated by scores above internationally normed cut-off values, separately for two types of behaviour: internalising and externalising. The aim of this research project was to identify risk and protective factors of problem behaviour, including variables with effects that might vary with age. This talk will cover some methodological hurdles encountered during analysis and will present some key results. Logistic multiple regression methods based on generalised estimating equations were used to address matters of correlated observations and missing participants. Many variables suggested by literature or hypothesised were introduced into the models and tested for association, along with their age interaction terms. The model selection process will also be described.

(80) SOFTWARE FOR REPRODUCIBLE RESEARCH USING R

David Scott
University of Auckland

[S.1.04] 10:50-12:10 Tuesday 26 November

There are now many approaches to reproducible research or automatic report generation using R. Besides Sweave which is probably the best known, other software includes odfWeave, SWord, org-babel, knitr, R2HTML and hwriter. I will discuss these examples, some in more detail than others, and then outline some of my own work in enhancing hwriter in the package hwriterPlus.

(81) PHD PROJECT IN LIGHT OF PROJECT MANAGEMENT

Maryam Mirzaei
Victoria University of Wellington

[S.1.02] 9:50-10:30 Wednesday 27 November

PhD is the highest academic qualification for which students produce the longest text of their lives. Such an enormous project undoubtedly needs to be managed. In this paper we examine the application of project management to PhD research project. Both literature survey and case study were used to collect data related to the characteristics of a PhD research as a project. Data have been analysed using project management classification model as well as Theory of constraint thinking processes. It was found that most success criteria of PhD projects are related to scope. Furthermore, many PhD project failures are caused by ill-defined scope. Therefore, in order to get the real benefit of project management in PhD research project, attention should be diverted to tools and techniques that facilitate scope clarification. Applicability of various project management techniques such as Gantt chart and scrum as well as scope analysis methods such as future reality and transition tree have been explored. This paper contributes to better understanding of PhD research as a project and possibilities of increasing its success using modern project management techniques.

**(82) THE EFFICIENT COMPUTATION OF THE GROUP INVERSE AND THE MEAN FIRST PASSAGE TIMES
IN A MARKOV CHAIN**

Jeffrey Hunter
AUT University

[S.1.05] 10:50-12:10 Tuesday 26 November

We explore the derivation of efficient computational procedures for finding the Group Inverse and the mean first passage times associated with a finite ergodic Markov chain. The derivation of the mean first passage times typically involves the solution of a system of linear equations and, for the group inverse, the evaluation of a matrix inverse. By using a perturbation technique, starting from a simple base where no such derivations are formally required, we update a sequence of matrices, formed by linking the solution procedures via generalized matrix inverses, utilising only matrix (and vector) multiplications. The derivation is based upon the ideas outlined in Hunter, J.J., The computation of stationary distributions of Markov chains through perturbations, *Journal of Applied Mathematics and Stochastic Analysis*, 4, 29-46, (1991).

**(83) ENVIRONMENTAL FACTORS AND STROKE INCIDENCE: EVIDENCE FROM SIX LARGE POPULATION-
BASED STUDIES IN EUROPE AND AUSTRALASIA**

Priya Parmar, Valery Feigin, Suzanne Barker-Collo and Rita Krishnamurthi
AUT University, AUT University, University of Auckland, AUT University

[S.1.05] 9:50-10:30 Monday 25 November

There is growing evidence that environmental health relationships pose increasing health risks given future projections of climate change and that the warming trend over recent decades has already contributed to increased morbidity and mortality in many regions of the world. However, robust evidence of associations between particular health disorders and various environmental factors is lacking. Stroke is the second major cause of deaths worldwide and the leading cause of disability in adults in developed countries. We used the stratified Cox proportional hazard model to analyze individual-participant, daily meteorological and geomagnetic activity data from several large population-based stroke incidence studies. Hazard ratio (HR) and corresponding 95% confidence interval (CI) were calculated. Changes in meteorological variables and geomagnetic activity are all associated with increased risk of stroke and should be considered as potentially new risk factors for stroke. Our findings provide a framework to advance stroke prevention through future investigation of the contribution of these environmental factors to the risk of stroke occurrence and pathogenesis.

(84) BLINDNESS IN A VISUAL DISCIPLINE

A. Jonathan R. Godfrey
Institute of Fundamental Sciences, Massey University

[S.1.04] Poster Session: 16:40-18:00 Monday 25 November

No one would argue that statistics is inherently visual, especially at introductory levels where exploratory data analysis techniques are taught. My experiences as a blind person teaching sighted students has given me an insight into the needs of blind students coming through our high school and university statistics courses. My personal experiences have been recognized overseas and led to my attendance to two international workshops where I've shown blind students how easily they can use R as compared to other less useful (in terms of a blind person) statistical software. While I have been at these workshops, I have been able to discuss the latest initiatives in sound and tactile representations of visual material including statistical graphs. I've gained access to some technology not yet available in New Zealand and discussed the problems being faced by students and service providers all over the world. This multimedia presentation will have displays of the technology blind students of today can access that will enhance their educational outcomes in statistics and mathematics courses. To gain an appreciation of the techniques and tools on offer, you will need to listen and feel the solutions, not just read about them.

(85) BAYESIAN AGE RECONCILIATION FOR THE AUCKLAND VOLCANIC FIELD

Emily Kawabata, Mark Bebbington, Shane Cronin and Ting Wang
Massey University, Massey University, Massey University, Otago University

[S.1.02] 14:40-16:40 Tuesday 26 November

The estimation of the hazard from a monogenetic volcanic field, where each eruptive event creates a new volcanic centre, is critically dependent on a likely reconstruction of past events. The Auckland Volcanic Field has produced about 50 volcanic centres during its active phase over the last 250,000 years. The rock fragments and ash ejected by a volcanic eruption are collectively called tephra and partial records, including estimated ages of the eruptions, are available from five tephra deposition locations within the field. However, there is no direct link between the deposit(s) and a particular source volcano. We use an attenuation model, which links estimated eruption volumes with locations of source volcano and tephra deposits, to calculate the likelihood of any combination of volcano and tephra. A Bayesian model, incorporating other age-related data, is suggested to estimate the ages of the volcanic centres and their most likely order.

(86) NEW DISTRIBUTION FREE GOODNESS OF FIT TESTS FOR POWER-LAW

Thuong Nguyen
Victoria University of Wellington

[S.1.01] 9:50-10:30 Tuesday 26 November

In 2013, Khmaladze[1] introduced a new class of asymptotically distribution free goodness of fit tests for discrete distributions, which can be used widely. My talk will concentrate on the pure discrete power-law distribution, where the Khmaladzes method is applied. The fact is that even though power-law has emerged as an extraordinarily frequent model for many physical, biological and social phenomena, up to now we have not had really useful tools for testing it in empirical data since power-law has a very heavy tail. The new method applied to testing power-law give us test statistics which have null distribution independent of the value of the parameter of power-law (and, even, of the fact that it is a power-law.) With power-law as a null hypothesis and Zipfs law, Karlin-Rouault law as alternatives, we show that distribution free version of Kolmogorov-Smirnov and omega-squared statistics, based on modified frequencies, have better statistical power than those based on the traditional vector of chi-squared components in most cases. References: [1] Khmaladze, E. (2013). Note on distribution free testing for discrete distributions. *Annals of Statistics*. (Accepted)

(87) MEMORIAL SESSION TO HONOUR PROF SHAYLE SEARLE

Jeffrey Hunter, Harold Henderson and Shirley Pledger
AUT University, AgResearch, Victoria University of Wellington

[S.1.04] 13:50-14:20 Tuesday 26 November

Professor Emeritus Shayle Searle, Professor of Biological Statistics at Cornell University, died on February 18, 2013. He was born in Wanganui in 1928. Shayle was one of the first statisticians to use matrix algebra in statistical methodology and was an early proponent in the use of applied statistical techniques in animal breeding. Shayle retained strong links with New Zealand and in particular Victoria University of Wellington which conferred on him an Honorary Doctor of Science in 2005. He was the recipient of an Honorary Fellowship of the Royal Society of New Zealand in 1999. The memorial session will highlight his career and his contributions to the field of statistics.

**(88) NATIONAL INSTANTANEOUS RESERVE PRICES AND REVENUE IN NEW ZEALAND
ELECTRICITY MARKET**

Vladimir Krichtal and Conrad Edwards
Transpower NZ

[S.1.04] 14:40-16:40 Monday 25 November

The paper presents a study of the instantaneous reserve price formulation in the proposed Mixed Integer Non-Linear Programming (MINLP) Scheduling Pricing and Dispatch (SPD) model. The model is based on a network that consists of two separate AC systems with the possibility of reserves transfer between two reserves zones via the High Voltage Direct Current (HVDC) link. Different configurations of the HVDC are considered, including round power where two poles operate in different directions. Reserves available in each zone have to cover each risk. Reserve prices and revenue formulae are derived for each risk and compared with existing formulae. Key words: National reserve prices, New Zealand Electricity Market, MINLP, SPD, reserve transfer, HVDC, round power.

(89) A REVIEW OF THE ACCURACY OF PITH OFFSET METHODS APPLIED TO KAURI INCREMENTAL CORES

Maryann Pirie
AgResearch Limited

[S.1.05] 14:40-16:40 Tuesday 26 November

Kauri is a significant resource for dendroclimatology study in the data deficient regions of the Southern Hemisphere. However, the full potential of Kauri as a climate proxy has not been realised because of several issues affecting the validity of these reconstructions. One such issue is the influence of the change in the age and size characteristics of the source material. This can be assessed through comparing the common signal and climate response of near-pith material with material further from the pith. As the pith is rarely present on the samples an estimate of the distance to the pith is required. This talk will review common pith offset estimation methods and assesses their accuracy for estimating pith offsets from kauri incremental cores.

(90) A REVIEW OF PROBABILITY AND STATISTICS APPS FOR MOBILE DEVICES

Howard Edwards
Massey University

[S.1.04] 9:50-10:30 Wednesday 27 November

Mobile devices such as tablets and smartphones are rapidly replacing laptops and notebooks as the primary student e-learning device. This paper reviews some the mobile apps currently available which enable a user to either learn Statistics or to carry out the sorts of summaries and analyses encountered in an undergraduate Statistics course. Implications of these apps for both teaching and learning are discussed.

(91) STATISTICS AND COUNTERFEIT MEDICINES

Sara J Fraser, Austina Clark and Keith Gordon
Dept of Chemistry, University of Otago, Dept of Maths and Stats, University of Otago, Dept of Chemistry, University of Otago

[S.1.01] 14:40-16:40 Tuesday 26 November

Due to the prevalence of counterfeit medicines worldwide and their high-risk contents, a feasible out-of-lab approach to efficiently detect counterfeit products, as well as to qualitatively and quantitatively determine active pharmaceutical ingredient (API) content is desirable. In this study the potential of Raman spectroscopy, combined with various statistical methods, was investigated for this purpose for the first time. Raman spectroscopy was used to classify a group of seized counterfeit medications. Using appropriate data pre-processing, principal components analysis (PCA) and the classification method soft independent modelling on class analogy (SIMCA), it was possible to classify genuine from generic Cialis batches. However SIMCA did not effectively classify samples based on their API. Partial least squares discriminant analysis (PLS-DA), principal components regression (PCR) and support vector machines effectively distinguished between the API of the samples, but were unable to correctly distinguish all samples as genuine or generic/counterfeit. This study highlights the importance of choosing the correct pre-processing procedures and classification method for the data set, and this approach may be used to determine the potential level of harm in counterfeit medicines based on API presence and amount.

(92) CHALLENGES IN OBTAINING AND USING DATA AND MAKING DECISIONS IN A POST-EMERGENCY ENVIRONMENT

John Créquer

Canterbury Earthquake Recovery Authority

[S.1.05] 14:40-16:40 Monday 25 November

In the immediate post-earthquake environment in Canterbury in 2011 data were shared relatively freely to assist with the emergency response. Once the initial response phase was over the needs changed to planning for the short- and medium-term. Decisions evolved over time from immediate emergency responses with short lead times to less well-defined decisions with greater lead times and longer horizons. To inform planning and decisions data were needed to establish the current situation and monitor changes. Typically data were required to be highly geographically detailed and frequent which could not so easily be obtained from established and traditional sources. The challenges included: establishing the need for the data, obtaining data from non-traditional sources, obtaining co-operation of suppliers, assessing data sources and benchmarking new data sources, and combining data with differing definitions. The latter is particularly challenging when using these data in models. The challenges of meeting these data needs in this environment will be discussed. Approaches employed in meeting these challenges and their success will also be traversed.

(93) DETERMINATION OF THE RELATIVE CONTRIBUTIONS OF WATER FROM SOIL SUBSTRATES TO STREAM WATER

Neil Cox, Richard McDowell and Tim Jowett

AgResearch

[S.1.05] 14:40-16:40 Tuesday 26 November

The data consist of concentrations of 31 elements (Ca, Mg) from 4 substrates at each of 13 source sites on a stream and the same elemental concentrations at a single downstream sink, the latter on a number of occasions with a variety of flow rates. The water from these sites flows into the sink. The concentration of the elements in the sink is expected to be proportional to the sum of the upstream loads (load = amount of water * concentration). The aim is to determine (from the element signatures of the 4*13 sources and for any given sink sample) how much is being contributed from each of the 4 substrates. An obvious problem is that we have 52 unknowns (the amount of water from each site and substrate) and 31 observations (the concentrations of the 31 elements at the sink and all the sources). The coefficients represent the water quantities being contributed by each source, so must be non-negative. We will describe our attempts to get estimates and confidence intervals for the percentage contributions from the 4 substrates.

(94) ON SOLVING THE MULTI-OBJECTIVE BEAM ANGLE OPTIMISATION PROBLEM USING SINGLE OBJECTIVE LOCAL SEARCH

Guillermo Cabrera G., Matthias Ehrgott, Andrew Mason and Andrea Raith
Escuela Ingenieria Informatica, Pontificia Universidad Catolica de Valparaiso, Department of Management Science, Lancaster University, Department of Engineering Science, University of Auckland, Department of Engineering Science, University of Auckland

[S.1.04] 14:40-16:40 Tuesday 26 November

In intensity modulated radiation therapy, beam angle and fluence map optimisation problems are usually solved sequentially. Once the number and directions of the radiation beams are determined, intensities for those beam directions are calculated. When these problems have been tackled simultaneously in the literature, only single objective functions of the fluence map have been used to evaluate the quality of a specific beam angle configuration (BAC). In this work, we address both problems simultaneously from a multi-objective point of view using a hybrid strategy combining heuristics and exact methods. Because generating a large set of non-dominated points for each visited BAC is quite expensive in terms of computation time, in this work we use sample non-dominated points to measure the quality of each BAC. Therefore, we do not need to spend time to generate a large set of non-dominated points to decide whether a BAC is considered or not (full-search approach). Using our approach we can either reduce the computational time required to solve the problem significantly or visit a larger number of BACs in the same time than the full-search approach would take without a substantial loss of quality of the final non-dominated set.

(95) TRANSFORMING A PUBLIC SECTOR TEAM'S DATA PRODUCTION, MANAGEMENT, VISUALISATION AND ANALYSIS

Peter Ellis
Ministry of Business, Innovation and Employment New Zealand

[S.1.05] 9:50-10:30 Tuesday 26 November

The Sector Performance team (previously the Tourism Research and Evaluation team) in the New Zealand Ministry of Business, Innovation and Employment began a five year data improvement programme in 2011. Changes implemented by the end of 2013 include a major on-line redevelopment of the International Visitor Survey that has increased estimates of tourism exports by around \$2 billion per year; and world-first use of electronic card transactions combined with survey and national accounts data to estimate regional tourist spend by origin and product. Behind the scenes a complete transformation of the team's tools, workflow, approach to data management and analysis has taken place. Modern applied statistical techniques such as multinomial regression, bootstrapping, correspondence analysis, iterative proportional fitting and generalized additive models are now routinely applied in-house to new data problems; as is extensive use of innovative graphics, statistical overlays on maps, animations, and interactive visualisations. This paper outlines how the success has been possible and lessons learned for organisations seeking to build analytic and statistical capability.

(96) FAST COMPUTATION OF ADJUSTED FORECASTS IN HIERARCHICAL TIME SERIES

Rob Hyndman and Alan Lee
Monash University, University of Auckland

[S.1.01] 10:50-12:10 Wednesday 27 November

In many applications, we may have several times series representing products, groups or geography, as well as separately measured aggregated series. We thus have a hierarchical structure, and may want to produce forecasts of the separate series and the aggregates that respect this hierarchical structure. We could aggregate the separate forecasts, ignoring the forecast of the aggregated series, but this is inefficient. Hyndman et al (2011) proposed a regression based methodology that produces more accurate forecasts. However, if there are many (perhaps several thousand) series being aggregated at different levels of aggregation, we have a regression problem with large number of coefficients to estimate. Standard least squares programs will not provide the best way of fitting these regressions, as they do not exploit the hierarchical structure. In this talk we outline a recursive method for computing these regressions and the adjusted forecasts for an arbitrary hierarchical structure. The method can handle quite large structures, particularly if the number of forecasts being aggregated at the lowest level is large compared to those being aggregated at higher levels.

(98) EXAMINING DIFFERENCES BETWEEN PERCEIVED AND ACTUAL URBAN ENVIRONMENT MEASURES

Nick Garrett
AUT

[S.1.05] 9:50-10:30 Wednesday 27 November

Where we live and how we perceive our local environment has an impact of how physically active we are, and the maintenance of a health and wellbeing throughout our lifespan. A population survey of adult residents in North Shore City was undertaken in 2005, which asked participants questions about their physical activity as well as awareness and use of local places and facilities. In addition, GIS databases were setup that involved a full audit of parks, beach access, sports fields, gyms, halls and any facility that provided places for the public to undertake physical activity. Measures utilising the GIS databases and the road and pathway networks were developed to measure accessibility of the survey participants to all places for physical activity. Logistic regression was utilised to examine the relationship between achieving sufficient physical activity to maintain health, at least 150 minutes of moderate physical activity per week, and perceived or objective measures of accessibility. While the perceived measures demonstrated strong associations, objective measures except beach access showed no association. To further investigate this slightly counterintuitive result, nominal logistic regression was utilised to compare the perceived and objective measures, which then demonstrated the complex relationship between these measures.

(99) SEISMIC MITIGATION DECISIONS FOR BUILDING OWNERS IN NZ

J. Fei Ying, Suzanne Wilkinson and Jim Corner
Auckland University of Technology, University of Auckland, University of Waikato

[S.1.02] 14:40-16:40 Monday 25 November

Seismic retrofit implementation lags behind advances in scientific and engineering advances and has been a major challenge in many active seismic zones. This paper first explores the decision environment and decision process to understand the problems faced by building owners when evaluating the rehabilitation alternatives for substandard buildings. This is accomplished through a series of interviews with 32 government officials, building owners and various other professionals throughout NZ. A standard application of Decision Analysis with Value-Focused-Thinking (VFT) is then applied to these rehabilitation decision problems through two real life case studies to overcome weaknesses of conventional methods used in seismic mitigation decision making. It presents data illustrating the application of this method to a proposed demolition of a hospital building in a low risk zone and an upgrade of a heritage building in a high earthquake risk zone. In both cases, exploitation of Value-Focused-Thinking permits the building owners to shift their attention from the existing alternatives to the fundamental objectives that are important to the organisation, which leads to a better quality decision. The research findings suggest that the method can be adopted in response to similar decision problems faced by building owners with earthquake-prone buildings.

(100) MULTI-OBJECTIVE OPTIMISATION IN DECISION MAKING OF INFRASTRUCTURE ASSET MANAGEMENT

Lin Chen, Andrea Raith and Theuns Henning
University of Auckland

[S.1.02] 14:40-16:40 Monday 25 November

Infrastructure asset management (IAM) is important for modern society. Decision making, as a critical part of IAM, helps in achieving the goals of IAM by selecting strategies. Decision making faces many challenges; therefore, multi-objective optimisation methods are applied to identify efficient solutions in order to assist decision making. Efficient solutions contains supported and non-supported solutions, which can be used as the basis of decision making. The identification of supported solutions is easier; while the identification of non-supported solutions is more challenging. This paper analyses the IAM decision making problem and identifies efficient solutions to assist decision making using a two-phase optimisation approach. More specifically, this paper (1) introduces IAM decision making; (2) formulates practical decision making problems; (3) uses a two-phase optimisation approach that identifies supported solutions at first phase and non-supported solutions at second phase by different methods; and (4) discusses the result. The two-phase optimisation approach is tested with a bi-objective IAM decision making problem. According to the test, practical decision making problems can be solved; and both supported and non-supported solutions are identified. When identifying non-supported solutions, methods perform differently and their application depends on the addressed problem.

(101) CAPACITY EXPANSION OPTIMISATION IN TRANSPORTATION NETWORKS

Danny Tsai, Andrea Raith and Andy Philpott
University of Auckland

[S.1.02] 14:40-16:40 Monday 25 November

The network design problem determines the optimal links in a transportation network to upgrade in order to minimise traffic congestion. Traditional approaches to this problem under an optimisation framework involve solving a bi-level, non-linear, integer program. We explore an alternative approach known as the system optimal network design problem where road users are assumed to act cooperatively. This assumption simplifies the formulation into a single-level non-linear integer program, which we linearise and solve through Kelleys cutting plane method. The effect of increasing the number of network loading conditions considered in the model is explored, and we show the importance of modelling these scenarios in unison to find the optimal solution. The structure of the network design problem with multiple scenarios lends itself to decomposition methods. We compare the solve-time performance between using Benders decomposition and a large mixed-integer program. Results show that the decomposition has greater performance as the network size and number of states increases, but suffers when the number of upgrade decisions increases.

(102) IMPROVED MODELS FOR STUTTER PREDICTION IN FORENSIC DNA ANALYSIS

M.A. C. S. Sampath Fernando, James M. Curran, Jo-Anne Bright, John S. Buckleton and Renate Meyer
University of Auckland, University of Auckland, University of Auckland, Environmental Science and Research Limited (ESR), University of Auckland

[S.1.05] 14:40-16:40 Monday 25 November

Analysing the effect of stutter products is essential in forensic DNA analysis. Therefore, proper selection of statistical models for interpretations of stutter is vital. Bright et al. (2013) studied five statistical models for the behaviour of the stutter ratio (SR). These were two log-normal models and two gamma models on SR and a two-component normal mixture model on $\log(\text{SR})$. Each models performance was assessed by calculating the log-likelihoods and Akaike information criterion (AIC) was used for comparisons. As both the log-normal and gamma distributions are positively skewed, testing the appropriateness of symmetric distributions would be valuable. This study tests normal and Students t models with both common variance and locus specific variance. A two-component normal mixture model, and a two-component t mixture model were also tested on SR. Proposed six models and existing five models were studied with the same IdentifilerTM and NGM SElectTM DNA profiles. In this talk, I will be describing the statistical methodology that we adopted to improve the models developed by Bright and others. I will also compare the performances of proposed models and existing models using AIC and present the interesting findings related to the improvements in the new models.

(103) MODELLING RISK AVERSION IN GENERATION EXPANSION

Corey Kok and Andy Philpott
University of Auckland

[S.1.02] 14:40-16:40 Monday 25 November

The current state of the art model for predicting expansion decisions is the Generation Expansion Model (GEM) which given many constraints, chooses to invest to maximize the welfare of the system. GEM does not model the uncertainty and the resulting risk averse behaviour of the agents who make investment decisions. In this project, a mathematical program was written to model the behaviour of the various agents in the electricity market. Assuming all agents are perfectly competitive and risk neutral, this causes the model to give the same solution as the model which maximizes the welfare of the system. By adding a coherent risk measure, Average Value at Risk (AVaR) to the decision of expansion, we were able to present a more realistic representation of the market. We compared various models using this risk measure. We observed that in our example, without vertical integration between the generation and retail wing causes generation companies to invest less compared with the expected welfare maximising model. Including this vertical integration causes over investment compared with the expected welfare maximising model. We observe that this is the result of differences in profits between the two models across each of the scenarios.

(104) BINARY INTERRUPTIBLE LOAD OPTIMISATION

Jack Dunn, Golbon Zakeri and Anthony Downward
University of Auckland

[S.1.02] 14:40-16:40 Monday 25 November

The electricity market in New Zealand relies on reserves to restore balance between supply and demand following a fault in the network. Interruptible load reserve is one type of reserve that responds either in full, or not at all. Currently these reserves can be partially dispatched, which can lead to an over-response of reserve, referred to as overhang. We investigate two models for reducing this overhang. The Intelligent Selection model constructs a dispatch solution that minimises the overhang without changing the cost of the initial solution. This eliminates the overhang in 15% of solutions. The Constrained On/Off model alters reserve dispatches more significantly to eliminate overhang from the solution entirely. Additional payments are made so that all reserve offers are fairly compensated for these changes. Overhang is eliminated in all solutions for a total cost increase of 0.4%. We consider two methods of determining the price in the new solutions: the price that minimises the sum of additional payments, and the price given by integer programming duality. Both are shown to coincide with the current clearing price, given by the linear programming dual. Finally, we show there are minimal changes to market incentives due to these models.

(105) POSM: OPTIMAL DAIRY FARM STOCK MANAGEMENT UNDER UNCERTAINTY

Andy Philpott and Peter Mills
University of Auckland

[S.1.02] 9:50-10:30 Tuesday 26 November

We describe a model for planning the operation of a dairy farm over a season that is subject to possible drought conditions. The model is solved using a SDDP-type algorithm with four continuous state variables, and a Markov chain determining the climate. The implementation of the model is described, and the policies it generates are compared in simulation with those from some simple decision rules.

(106) TRANSMISSION TOWER PAINTING OPTIMISATION

Jenny McLean, Anthony Downward, Golbon Zakeri and Andy Philpott
University of Auckland

[S.1.02] 14:40-16:40 Monday 25 November

Transpower manages nearly 25,000 galvanised steel lattice towers, and requires a maintenance strategy to prevent tower corrosion. Each tower is rated with a condition assessment: the deterioration of which can be predicted by its corrosion zone, defined by the climate and location. Typically, it takes about one week for a crew to paint a tower. However, when towers are in close proximity, they can be painted concurrently, thus reducing the total time to paint. A modified Bellman-Ford algorithm was developed to generate sets of feasible crew painting schedules. The best schedules were then chosen using two different approaches: a greedy selection method and a Set Partitioning problem. The greedy approach involved sequentially selecting the best schedule generated by the modified Bellman-Ford algorithm and removing these towers from the network. This performed as expected; only the most urgent towers in the network were covered. The Set Partitioning problem delivered a set of viable schedules that covered urgent towers, as well as less urgent towers when in close proximity. In doing so, the schedules were able to make the most of the ability to paint towers concurrently.

(107) GRAPH PARTITIONING AND ITS APPLICATION TO THE TRAFFIC ASSIGNMENT PROBLEM

Keith Ruddell
University of Auckland

[S.1.02] 14:40-16:40 Monday 25 November

Network models are used in numerous applications, including transportation, telecommunications and utilities. In such models one often wishes to suppress detailed local structure in order to emphasise global or long-distance behaviour which is most important to strategic decision-making. One desires a simpler network which preserves the topological and metric relations between regions. This may be achieved by judiciously aggregating nodes - treating several nodes as one. A scheme for aggregating nodes can be described by a partition of the graph. I describe a few of the many methods for partitioning graphs: spectral partitioning, minimum cut algorithms and an adaptation of k-means to a discrete metric space. I then discuss an application of partitioning and aggregation to the traffic assignment problem which resulted in often substantial reductions in computation time for urban transport models.

(108) 2D SHAPE OPTIMIZATION APPLIED TO DIAMOND CUTTING

Max Biggs and Tony Downward
University of Auckland

[S.1.02] 10:50-12:10 Tuesday 26 November

The diamond cutting problem involves maximising the size of a diamond required to lie within a rough stone. The shape of the diamond is fixed, but can be rotated, translated and magnified to find the optimal orientation. We extended existing work on the diamond cutting problem to incorporate non-convex stones and multiple diamonds in two dimensions. These additions are important for realistic modelling, because the majority of rough diamond stones have dents or imperfections which make them non-convex. Furthermore, the combined value of several diamonds cut from a stone is often significantly greater than the value of one large diamond. However the modelling is complex because even for a single diamond within a convex stone, the diamond cutting problem is non-linear and non-convex. We developed three methods to solve the diamond cutting problem with a non-convex stone. The vertex addition method was the fastest and most accurate. This formulation has linear constraints and a non-linear objective. Although, not guaranteed to produce the global optimum solution, we can identify a small range bounding the global optimum, related to the accuracy of a piecewise linear approximation. This causes a trade off between the quality of the solution and the solve time. Two approaches were explored for solving the multiple diamond cutting problem. We developed a greedy sequential method wherein a first diamond was placed optimally and a second diamond was placed in the remaining stone. We also developed a second method which placed diamonds simultaneously in a large integer program but was significantly slower due to the added complexity. Although the difference in solution between the two methods was small for most trials, the high value of diamonds justifies the increase in complexity for the small gain solution quality offered by the simultaneous method.

(109) DANTZIG-WOLFE DECOMPOSITION FOR THE BI-OBJECTIVE MULTI-COMMODITY MINIMUM COST FLOW PROBLEM

Siamak Moradi, Andrea Raith and Matthias Ehrgott

Department of Engineering Science, University of Auckland, Department of Engineering Science, University of Auckland, Department of Management Science, Lancaster University

[S.1.04] 14:40-16:40 Tuesday 26 November

We present a new method for solving the bi-objective multi-commodity minimum cost flow problem. This method is based on the standard bi-objective simplex method and Dantzig-Wolfe decomposition. The method is initialized by optimizing the problem with respect to the first objective, a single objective multi-commodity flow problem, which is solved using standard Dantzig-Wolfe decomposition. Then, similar to the bi-objective simplex method, our method iteratively moves from one non-dominated extreme point to the next by finding entering variables with the maximum ratio of improvement of the second objective over deterioration of the first objective. As we use the Dantzig-Wolfe method in the initial iteration, we do not have a complete set of variables. Our method generates entering variables by finding the optimal solution of single commodity flow sub-problems with the ratio objective function. We find the optimal solution of each sub-problem among the efficient solutions of a bi-objective network flow problem. The solution with the best ratio objective value out of all sub-problems represents the entering variable. The method stops when all the non-dominated extreme points are obtained. The implementation of the method and numerical results are explained.

(110) DO SOME LIKE IT HOT? SEX RATIOS (AND TIME SERIES) IN AUSTRALIA AND NEW ZEALAND

John Haywood

Victoria University of Wellington

[S.1.01] 14:40-16:40 Tuesday 26 November

The global secondary sex ratio (human male: female births) is currently 1.07, giving a proportion of males born (male: total births) of 0.517. Ambient temperature has been positively associated with birth sex ratio within some northern European countries: a higher proportion of males are born in warmer years and in warmer seasons. We examined the hypothesis that ambient temperature predicts fluctuations in the proportion of male births in each of Australia and New Zealand, using over 100 years of time series data. We also performed regional analyses within both countries, including a test of the hypothesis that in areas of Australia with hot temperatures the proportion of male births would decrease when ambient temperatures increased. Further, we examined the effects of NZ stillbirths on the proportions of live births, and we tested whether seasonal ambient NZ temperatures, or deviations from those seasonal patterns, were positively related to the proportion of males born in NZ using monthly data (1980-2009). Results showed, across all hypotheses considered, that fluctuations in ambient temperatures and significant differences in regional temperatures were not related to the proportion of male births. So temperature may influence human sex allocation elsewhere, but such effects of temperature are not universal.

**(111) SYMMETRIC TRAVELING SALESMAN PROBLEM AND FLOWS ON HYPERGRAPHS - NEW
ALGORITHMIC POSSIBILITIES**

Tiru Arthanari
University of Auckland

[S.1.04] 14:40-16:40 Tuesday 26 November

Symmetric traveling Salesman problem (STSP) has many different formulations among them the standard formulation by Dantzig, Fulkerson and Johnson has been studied extensively, both from theoretical and computational perspectives. Multistage insertion formulation is as tight as the standard formulation, but possesses many interesting theoretical properties having only polynomially many constraints. Integer gaps found in computational studies comparing different formulations indicate the superior performance of the MI-formulation. However, these comparisons have used generic LP solvers to solve the MI- formulation of STSP instances. Minimum cost flows on hypergraphs studied by Cambini et al.[3] extends the minimum cost flows in networks / generalized networks and study a hypergraph simplex method for solving such problems. Noticing that the MI problem, is a special case of the minimum cost flow problem in a hypergraph, we propose new algorithms to solve the MI-relaxation problem. We also exploit the Leontief structure of a sub problem to provide additional algorithmic approaches to solve the MI-relaxation problem, using Lagrangean relaxation. Some theoretical bounds on complexity are also derived. Computational experiments planned to compare the new algorithms with the generic commercial LP solver are outlined. From the reported performance of Lagrangean relaxation approaches and that of minimum cost hypergraph simplex method, it is expected that the new proposed algorithms will perform better than solving MI-relaxation using generic LP solvers.

(112) SHORTEST PATH ALGORITHMS IN TRAFFIC ASSIGNMENT

Boshen Chen, Olga Perederieieva and Andrea Raith
University of Auckland

[S.1.02] 10:50-12:10 Tuesday 26 November

Finding shortest paths is a classical problem in optimisation. Many solution algorithms for practical optimisation problems require solving shortest path as sub-tasks. An example of such a problem is Traffic Assignment (TA) which models the behaviour of drivers in a road network assuming that every driver travels on their fastest (or shortest) path. TA algorithms require at every iteration many shortest path calculations in a graph with link costs that slightly change during solution process. This iterative nature together with the specific structure of underlying transport networks raises the question which shortest path methods are most suitable for an iterative TA algorithm. We firstly conduct empirical tests of performance of different shortest path approaches in the context of TA. We also observe that after several initial iterations when TA algorithm is approaching optimum only few shortest paths actually change. We investigate if avoiding some shortest path calculations altogether improves the overall runtime of TA algorithms. We conduct a numerical study of different shortest path algorithms in the context of TA and analyse several shortest path computation avoidance strategies. These ideas were applied to a path equilibration algorithm but can be extended to other TA methods.

(113) FRIDAY NIGHT'S A GREAT NIGHT FOR CRICKET

Mark Johnston
Victoria University of Wellington

[S.1.04] 14:40-16:40 Tuesday 26 November

Each year New Zealand Cricket schedules its domestic competitions across the summer months from October through to April. Six teams (Auckland, Northern Districts, Central Districts, Wellington, Canterbury and Otago) compete in three domestic competitions: Plunket Shield (four day, first-class matches), Ford Trophy (one day, 50 overs per side) and HRV Cup (one evening or afternoon, 20 overs per side Twenty20). For the 2012/13 season, it was proposed that HRV Cup (Twenty20) matches might be able to be scheduled so that there was one match available to be televised every Friday night from October through to December, with the remaining HRV Cup matches scheduled over the traditional Christmas to New Year period. It was also required that several rounds of Plunket Shield matches also be completed prior to Christmas. This paper develops an integer programming model, incorporating a finite state machine, to explore whether or not it is possible to construct such a schedule dovetailing HRV Cup and Plunket Shield matches, taking into consideration the requirements for ground lighting and rest/travel days between matches.

(114) CREATING A NAVIGATIONAL SUPPORT SYSTEM FOR CONSTRUCTION PROJECTS

Foad Marzoughi and Tiru Arthanari
University of Auckland

[S.1.02] 10:50-12:10 Tuesday 26 November

Around the world construction projects fail to meet their objectives. This can be attributed to a whole host of countervailing factors ranging from weather problems to industrial strife. There is strong belief that better measurement and decision-making systems could dramatically improve performance in the construction industry, resulting in economic and structural development in countries around the world. This research aims to develop a framework, particularly useful for projects; starting with projects in the construction industry, with the goal of reducing project failures and increasing project successes. The navigational support system would provide an exploratory map of how the strategy is planned for the construction projects, how the project performance will be navigated towards the desired targets (objectives) with respect to some resources restriction such as time and cost. By making, a navigational support system for construction projects the project managers can understand where the on-going project in terms of project performance is and they can figure out the best way to navigate toward the benchmark targets, which are dynamically changing. The performance benchmarking is a multi-dimensional space and Euclidian distance is not a proper tool because of correlation between critical success factors. In this research Mahalanobis distance will be used to determine the position of performance of projects in the Benchmark space. Also Mahalanobis Taguchi system (MTS) can be used only to monitor and diagnose the performance level of performance with respect to benchmark targets. In the second phase of our research we use Dynamic programming to control the performance to reach the benchmark targets in relation to some restrictions.

(115) MULTI-NODE OFFER STACK OPTIMISATION OVER ELECTRICITY NETWORKS

Anthony Downward, Yibo Weng and Golbon Zakeri
Department of Engineering Science, University of Auckland

[S.1.01] 14:40-16:40 Tuesday 26 November

In this work we examine the problem that electricity generators face when offering power at multiple locations into an electricity market. The amount of power offered at each node can affect the price at the other node, so it is important to optimise both offers simultaneously. Even with perfect information (i.e. known demand, and known offers from competitors) this is a non-convex bi-level optimisation problem. We first show how this can be formulated as an integer program using special ordered sets of type 2 (SOS2) enabling this problem to be solved efficiently. We then extend this work to allow for uncertainty, and hence find the profit maximising offer stacks at each node (as opposed to a single quantity, as in the deterministic case). We demonstrate the intuition that we can gain from this model in a simple two-node example, and finally show the results from the model using real data from the New Zealand electricity market.

(116) PUBLICATION PATTERNS: A CO-CITATION NETWORK FOR BIOMETRICS 1993-2013

Jennifer Wilcock
University of Auckland

[S.1.04] Poster Session: 16:40-18:00 Monday 25 November

The poster presents a network graph of co-citations in papers published in the journal *Biometrics* over the past 20 years. The graph illustrates the areas of study that have dominated in the journal, the extent to which these areas of study are connected and by which groups of cited papers, and which papers have been most highly co-cited. While the analysis is essentially descriptive I hope it will be of intrinsic interest to anyone working in any area of biometrics/statistics and familiar with published literature in our field. The nodes of the graph are papers cited in *Biometrics* since 1993, and the edges of the graph connect two papers cited together, or co-cited, in the same *Biometrics* paper. Edges are weighted by the frequency of co-citation occurring over the 20 year period. To restrict the size of the graph, only papers (nodes) cited at least ten times in *Biometrics* are included, and included co-citations (edges) must have appeared in at least five different *Biometrics* papers. The nodes in the network can be papers from any year and any journal however, since the nodes are based on the articles being cited in *Biometrics* rather than the papers published in *Biometrics* themselves.

(117) ECONOMIC APPRAISAL OF CYCLING INFRASTRUCTURE INVESTMENTS

Andrea Raith
University of Auckland

[S.1.04] 14:40-16:40 Tuesday 26 November

Walking and cycling are gaining increasing popularity as modes of transport around the world as they are active, promote a healthy lifestyle, and have lesser impact on the environment than other modes of transport. Funding for infrastructure investments is typically limited and it is important to prioritise those projects that yield the largest benefits to society. Benefit-to-cost ratios are commonly used as a measure to determine which projects should be funded. Benefits of cycling infrastructure relate to improvements in travel time, reduction in the number of motorised vehicles on the roads, and health and safety benefits. Most of these benefits are derived from an estimate of how many users new infrastructure will have. To estimate use of cycling infrastructure we modify the conventional model (for motorised transport) to allow us to model of cyclist flows. A core assumption in the model is that cyclist route choice may be affected by both perceived safety along a route and travel time. An initial model is developed for cycling infrastructure in Auckland. The new model is compared to current practice, and applied to a recently constructed cyclepath in Auckland which enables the comparison of modelled and observed cyclist flows.

(118) IMPROVING PATIENT TRANSIT AT AUCKLAND CITY HOSPITAL

Michael O'Sullivan, Fergus Bassett and Cameron Walker
University of Auckland

[S.1.04] 9:50-10:30 Tuesday 26 November

Patients arriving on time for treatment is critical for both patient quality of care and efficient delivery of healthcare services. The Patient Transit Service at Auckland City Hospital (ACH) is responsible for picking up patients from wards and delivering them to various services, e.g., Radiology, for treatment. Orderlies and, in some cases, transit nurses are used for patient transits with the aim of delivering patients within the on time window of 15-0 minutes before the patient appointment. This paper describes simulation and optimisation research aimed at improving patient transit at ACH. Historical data on patient transits requests are used to drive a simulation of patient transits and this simulation is calibrated to match the historical data as closely as possible. The calibrated simulation is then used for experimentation with different policies for dispatching orderlies/transit nurses including: fixing the time a request arrives before an appointment; fixing the time an orderly/transit nurse is dispatched to pick up a patient; sending the orderly/transit nurse who will arrive first to each requested pick-up; and optimising the dispatch of orderlies/transit nurses. Results of these experiments will be summarised and the implementation of an optimal dispatch system for orderlies/transit nurses at ACH will be discussed.

(119) SOLVING THE CLASSROOM ASSIGNMENT PROBLEM USING INTEGER PROGRAMMING

Antony Phillips, Matthias Ehrgott and David Ryan
University of Auckland, Lancaster University, University of Auckland

[S.1.02] 10:50-12:10 Tuesday 26 November

In this paper we present an integer programming method for solving the Classroom Assignment Problem in University Course Timetabling. We introduce a novel formulation of the problem which generalises existing models and maintains tractability even for large instances. We also address how the structure of different room assignment problems can affect the relative difficulty of finding and confirming an optimal solution. For situations where a high quality solution is required quickly, we develop a method based on constraint branching which significantly reduces the problem size while attempting to retain the best solutions. Finally we discuss the use of multi-objective optimisation to handle multiple competing solution quality measures. Our model and methods are validated through computational results based on our experiences at the University of Auckland, and on instances from the 2007 International Timetabling Competition.

(120) VISUALIZATION OF A STOCK MARKET CORRELATION MATRIX

Alethea Rea and William Rea
Data Analysis Australia, University of Canterbury

[S.1.04] 14:40-16:40 Monday 25 November

This paper presents a novel application of software developed for constructing a phylogenetic network to the correlation matrix for 48 stocks listed on the New Zealand Stock Exchange. We show that by visualizing the correlation matrix using a Neighbor-Net network some of the problems associated with understanding the large number of correlations between the individual stocks can be overcome. This yields greater insight into how closely individual stocks are related to each other in terms of their correlations and suggests new avenues of research into how to construct small diversified stock portfolios.

(121) COUNTERMATCHED DESIGN AS A TWO-PHASE SAMPLING DESIGN

Claudia Rivera and Thomas Lumley
University of Auckland

[S.1.01] 10:50-12:10 Tuesday 26 November

When studying risk factors for disease incidence in a cohort, some of them could become costly relative to their collection. For example, laboratory analysis of blood and urine samples or records of diet and exercise exposure. Additionally, if the exposure is rare, it would be needless and vain to sample many subjects. Langholz and Borgan (*Biometrika*, 1995) proposed to use countermatching as a technique to sample a group of controls for each of the cases in a cohort study. At each failure time, strata are formed by variables of interest and subsequently a stratified sample is drawn in each risk set. Expressing the countermatched design as a two phase sampling design is of interest when fitting Cox model, either parametrically or semiparametrically. In order to use weighted likelihood or weighted partial likelihood, the inclusion probabilities are found for the final two-phase sample. It leads to an extension of the methods proposed by Samuelsen (*Biometrika*, 1997) for nested-case control designs. The controls from subjects in the risk set are reused for each case so that efficiency of estimators might be improved.

(122) MINING FOR DIFFERENCES AND CHANGES

Yong Wang
University of Auckland

[S.1.05] 10:50-12:10 Wednesday 27 November

In this talk, I will describe a new general methodology for difference and change detection. Following this methodology, two data mining techniques are proposed. One is based on tree-structured models and the other makes use of the bump-hunting technique. Through a case study, it is illustrated how they can help detect changes in the frequencies of event occurrences and uncover unusual clusters of events in a complex environment. The performance of the two methods will be compared and discussed.

(123) A SIMULATION PACKAGE FOR R

Peter Green
Landcare Research

[S.1.04] Poster Session: 16:40-18:00 Monday 25 November

If we are going to invest in the collection of monitoring data, it is important that we know that our design is suitable for making the inferences we need. Does the design have sufficient power to detect what we are trying to find? Are our parameter estimates going to be precise enough for our purposes? When the design is complicated, e.g. in a mixed-effects model, simulation provides a general way of answering these questions. However, setting up such a simulation experiment might be too complicated for many researchers. We are building an R package which provides tools that make it simple to set up and run simulation experiments.

(124) DO OPINION POLLS PREDICT NEW ZEALAND GENERAL ELECTION RESULTS?

Peter Green
Landcare Research

[S.1.01] 9:50-10:30 Monday 25 November

Several research firms regularly publish polls of New Zealand voters party preferences. If changes in underlying party support change slowly, then most of the movement in polling numbers will be due to sampling variation. If journalists over-interpret these random fluctuations there is a risk that the public will be misinformed. When movements in party support are within the margins of error of individual polls, we can increase our power to detect underlying trends by composing polls of polls. Statisticians in the USA have recently had a lot of success with this approach. In New Zealand there is less polling data available, but our electoral system is also much simpler. We will look at polling data and election results for the past three New Zealand general elections.

(125) THE RULE OF THREE IN ANALYTICS

Annelies Tjetjep
SAS

[S.1.04] 13:50-14:20 Monday 25 November

Analytics is a combination of disciplines both theoretically as well as in practice, but there are some fundamentals which are universal for delivering real world insights across research, government and commercial purposes. This is not a presentation on how easy maths and stats is or how technology can miraculously create insights. This is a presentation of some basic trifectas in analytics, experienced and observed, that have led to successful projects, people and organisations.

(126) DATA MONITORING COMMITTEES A NEW ZEALAND EXPERIENCE

Katrina Sharples
Department of Preventive and Social Medicine, University of Otago

[S.1.02] 10:50-12:10 Monday 25 November

In 1996 the Health Research Council (HRC) of New Zealand established a process for providing independent Data Monitoring Committee (DMC) oversight for HRC funded trials. This committee's main responsibility is to constitute independent DMCs where these are required. It also reviews the monitoring plans of HRC funded trials, and provides recommendations to the HRC on the appropriate form of monitoring. The HRC DMC has now completed monitoring of 13 trials, and is currently monitoring a further 9. This talk will provide an overview of the HRC DMC processes, describe some of the issues in monitoring that have arisen, particularly in relation to confidentiality of the trial data, and discuss the results of the recent review of the HRC DMC.

(127) THE STATISTICIAN'S ROLE IN THE CONDUCT OF RANDOMISED TRIALS

Alain C. Vandal.
Auckland University of Technology

[S.1.02] 10:50-12:10 Monday 25 November

The speaker will present some lessons learned from his experience as a statistician on several Phase 2 and Phase 3 randomised trials held in New Zealand in the past 5 years. In particular, the role of the statistician as an adviser on not just trial design and analysis, but also conduct, will be illustrated. A reminder of the main sources of bias in a trial, and how statistical expertise can help to palliate them, will be offered. This talk will provide some context for the other presentations in the Clinical Trials Session.

(128) PULP MILL POWER TOOL

Graeme Everett and Andy Philpott
Norske Skog Tasman Limited, University of Auckland

[S.1.05] 14:40-16:40 Monday 25 November

TBA

(129) USING THE CENSUS AS A SAMPLING FRAME: A RARE OPPORTUNITY TO UNDERSTAND NON-RESPONSE BIAS

Kylie Maxwell
Statistics New Zealand

[S.1.01] 10:50-12:10 Monday 25 November

New Zealand's 2013 Census of Population and Dwellings was used as a sampling frame for two large household surveys: the Mori Social Survey (Te Kupenga) and the Household Disability Survey. One opportunity when using the census as a frame is having demographic information (age, sex, ethnicity, income) about all sample units, including non-respondents. This talk will discuss how we used this information to better understand non-response bias, and how we were able to use this information to adjust for non-response using calibration. While using the census as a sampling frame has some advantages over alternative frames, it also presents some challenges. These include; drawing a sample from a list frame where it is possible to select multiple persons per household, which increases respondent burden, dealing with population groups that move often, and administrative challenges. This talk will discuss these challenges in more detail.

(130) NeSI: HIGH PERFORMANCE COMPUTATIONAL PLATFORMS AND SERVICES FOR NZ's RESEARCH COMMUNITIES

John Rugis, Ben Roberts and Mark Cheeseman
NeSI

[S.1.01] 10:10-10:30 Wednesday 27 November

NeSI (<http://nesi.org.nz>) provides high performance computers and support systems to enable the country's researchers to tackle large problems. NeSI's Computational Science Team is comprised of experts from around the world with scientific domain knowledge and has the mandate to work directly with investigators to improve their research productivity. NeSI is an unincorporated body, with investment from New Zealand universities, Crown Research Institutes and the Crown, through the Ministry of Business, Innovation and Employment. This talk will showcase NeSI's capabilities, introduce its new training programme in scientific programming, and include a Q&A session about what researchers want from their research infrastructure.
