

Australasian Applied Statistics Conference

2024

2 - 5 September,
Rottnest Island,
Western Australia



About AASC and AGUAI

This conference is the most recent in a series of conferences originating with the initial Australasian Genstat conference in Canberra in 1979. At the 2008 conference in Victoria it was voted to broaden the focus of the conference to encompass the wider community of applied statisticians. The conference was therefore renamed as the Australasian Applied Statistics Conference (AASC). The conference attracts biometricians and statisticians working mainly in primary industries and environmental science research in Australia and New Zealand. The first AASC was in 2011 at Palm Cove (Qld), followed by conferences in Queenstown (2012 NZ), Port Lincoln (2014 SA), Bermagui (2016 NSW), Rotorua (2018 NZ) and Inverloch (2022 Vic). The Australasian Genstat Users' Association incorporated (AGUAI) exists to organise the biennial AASC. The AGUAI is an independent, non-commercial, not-for profit, association. Anyone can become a member of the AGUAI by simply attending an AASC and indicating their agreement to membership – there are no membership fees or other requirements. Being a legally incorporated association offers some on-going structure and a degree of legal protection for the committee organising conferences. The AGUAI AGM is scheduled to be held on Tuesday 3 September 2024, where the main item is to select a new AGUAI committee.

Current AGUAI committee organising the AASC 2024 conference:

Officeholders:

Andrew van Burgel (Chairperson, WA)

Bethany Rognoni* (Deputy chairperson, Qld)

Karyn Reeves (Secretary, WA)

Kefei Chen (Treasurer, WA)

Ordinary members:

Tinula Kariyawasam** (WA) Bec O'Leary*** (WA), Vanessa Cave* (NZ), Ruth Butler* (NZ), Angela Anderson* (Qld), Khageswor Giri (Vic), David Baird (NZ), Anna Hepworth (WA), Pauline Ding (NSW).

* Scientific program committee ** Conference website *** Conference booklet

Session guidance for presenters and chairs

We strongly request that everyone keep to the times given:

Invited Talks: 35 minutes with 10 minutes for questions.

Oral Presentation Talks: Sessions of six talks will run with a short break after the first three talks. Up to 10 minutes will be available for questions after all talks have been completed.

Quick-Fire Talks are 10min. Some talks have been allocated 15 minutes.

Posters: The poster session will start with one-minute talks for each poster (possibly with one or two slides).

Session Chairs will be given two 'time left' signs and a 'stop' sign to hold up at the appropriate moment. Presenters should expect to stop talking when the 'stop' sign is shown.

Workshops

Two concurrent workshops are planned for 9 am to 1 pm on Monday 2 September 2024 at the Esplanade Hotel:

Workshop 1: Prof. Dr. Hans-Peter Piepho

Recent developments in design and analysis of experiments in plant breeding and variety testing based on linear mixed models

Workshop 2: Dr. Vanessa Cave & Dr. David Baird, VSN (with input from Dr. Roger Payne).

Reliable and Reproducible Analysis with Genstat

Please see the conference webpage for more details: <https://aasc2024conf.netlify.app/workshops>

Social program

Four options are available for the afternoon of Wednesday 4 September 2024. Please see the conference webpage for more details: <https://aasc2024conf.netlify.app/social>

Invited Speakers

Prof. Dr. Hans-Peter Piepho: Tuesday 9:00 – 9:45am

Presentation: Two-dimensional P-spline smoothing for spatial analysis of field trials



Managing director of Institute of Crop Science, Biostatistics, University of Hohenheim

Hans-Peter Piepho was appointed Professor of Biostatistics at the University of Hohenheim, Stuttgart, Germany in 2001. He has been working as an applied statistician in agricultural research for more than 30 years. His main interests are related to statistical procedures as needed in plant genetics, plant breeding and cultivar testing. Recent interests include envirotype- and marker-enabled breeding, spatial methods for field trials and experimental design for various applications including two-phase experiments and multi-environment trials. Further areas of interest include network meta-analysis and measure of goodness of fit for generalized linear mixed models.

Prof. Lucy Marshall: Tuesday 1:00 – 1:45pm

Presentation: Building Trustworthy Environmental Models



Executive Dean, Faculty of Science and Engineering, Macquarie University

Lucy Marshall is a Professor of Engineering and Executive Dean of the Faculty of Science and Engineering. She is a water resources engineer, with expertise in hydrologic modeling, environmental model optimization, and quantification

of uncertainty in water resources analysis. She has a special interest in understanding how environmental observations can be used to quantify uncertainty in systems undergoing change. Her research has spanned the development of new models in the most heavily instrumented watershed in the United States to making flood predictions in ungauged catchments across Australia.

Lucy completed her BEng (Hons), MEngSc, and PhD at the University of New South Wales (UNSW) in Sydney before moving to Montana State University as an Assistant Professor of Watershed Analysis, where she worked at the interface of engineering and environmental science. She returned to Australia as an Australian Research Council Future Fellow at UNSW, and went on to become the director of the UNSW Water Research Centre. She held multiple leadership positions at UNSW, as the inaugural Associate Dean (Equity and Diversity), Associate Dean (Research), and the academic lead for Athena SWAN. She joined Macquarie as Executive Dean in 2022.

Dr David Baird: Tuesday 3:30 – 4:15pm
Presentation: Genstat - some wider applications



Statistical Consultant and Genstat Developer, VSNi (New Zealand)

David is a biometrician with over 40 years' experience in applying statistics to agriculture, horticulture, biosecurity, entomology, ecology, plant breeding and finance. Through the judicious application of existing and novel statistical methods and tools, David has made many fundamental contributions to addressing critically important problems faced by New Zealand industry, Crown Research Institutes and government agencies. He has been a major contributor to Genstat over the last 30 years, developing the client, spreadsheet facilities, data transfer/manipulation and over 100 statistical or graphical procedures.

Anna Hayes: Wednesday 8:45 – 9:30am

Presentation: ARIMA Modelling To Determine Trigger Thresholds For Changes In Groundwater Levels



Principal Managing Consultant Statistician, Data Analysis Australia

Anna Hayes (Munday) is an applied statistician with over 20 years' experience working as a statistical consultant at Data Analysis Australia. Through managing Data Analysis Australia's consulting team, Anna provides leadership and direction to the company and monitors all projects undertaken by consultants. Anna is an Accredited Statistician of the Statistical Society of Australia, as well as being a past President of the Society's Western Australian Branch. Anna has a wealth of expertise in statistical application, with particular expertise in statistical analysis, forecasting, survey design and analysis, data management, methodology development, statistical reviews, project management and report writing. Anna's experience covers a vast array of industries, including environmental, biological and agricultural sciences, energy, transport, workforce planning, the Courts and legal applications.

Distinguished Prof. Kerrie Mengersen: Wednesday 10:15 – 11:00am

Presentation: The role of priors in Bayesian analysis



Research Chair, Statistics, Queensland University of Technology

Kerrie Mengersen is a Distinguished Professor of Statistics and Director of the Centre for Data Science at QUT, Australia. She is an elected member of the Australian Academy of Sciences and the Australian Academy of Social Sciences, and is a Vice-President of the International Statistical Institute.

Kerrie is passionate about developing methods to break open data and reveal insights that can help address critical challenges in health, environment, society and industry. She can be found perched on a stool at the intersection of statistics, machine learning, AI, and technology, and being constantly amazed and challenged by the current and future-promised traffic.

Prof. Adrian Baddeley: Thursday 8:30 – 9:15am

**Presentation: Spatial statistics and the protection of indigenous
rock art**



John Curtin Distinguished Emeritus Professor, Curtin University

Adrian Baddeley is a retired professor of statistics whose main research interests are spatial statistics and statistical computing. He is a graduate of the Australian National University and Cambridge University, has worked at Cambridge, Bath, Yale, UWA, CSIRO, CWI Amsterdam, and Curtin University, and held honorary professorships at Leiden and Aarhus. Adrian Baddeley is a winner of the Pitman and Hannan medals for statistical research, holds an honorary Doctorate of Science from Aalborg University, is a John Curtin Distinguished Emeritus Professor at Curtin University, and is a Fellow of the Australian Academy of Science.

Program

Monday

9:00am – 1:00pm Workshops at Esplanade Hotel (Fremantle)

1:00 – 2pm LUNCH

3:00pm Travel to Rottne at Fremantle B Shed

6:30-8pm Welcome function at Hotel Rottne

Tuesday

8.45 – 9am Opening

9:00 – 9:45am Invited Talk Hans-Peter Piepho

9:45 – 10:15am MORNING TEA

10:15am – 12pm Abstract Session 1.1

12:00 – 1pm LUNCH

1:00 – 1:45pm Invited Talk Lucy Marshall

1.45 – 3:00pm Abstract Session 1.2

3:00 – 3:30pm AFTERNOON TEA

3:30 – 4:15pm Invited Talk David Baird

4.15 – 5:30pm Abstract Session 1.3

5:30 – 6:00pm AGUAI AGM

7:00 – 9:00pm Poster session with Canapes

Wednesday

8:45 – 9:30am Invited Talk Anna Hayes

9:30 – 9:45 Committee announcements

9:45 – 10:15am MORNING TEA

10:15 – 11:00am Invited Talk Kerrie Mengersen

11:00am – 12:15pm Abstract Session 2.1

12:15 – 1.15pm LUNCH

1:15 – 6:30pm FREE TIME: Social Activities

6:30 – 9:00pm Conference Dinner at Isola

Thursday

8:30 – 9:15am Invited Talk Adrian Baddeley

9:15 – 10:10am Abstract Session 3.1

10:10 – 10:15 Committee announcements

10:15 – 10:45am MORNING TEA

10:45 – 11:30am Abstract Session 3.2

11:30am – 12:15pm Abstract Session 3.3
12:15 – 12:30pm Conference close
12:30 – 1.15pm LUNCH
2pm Ferry departure

AASC 2024: Workshops

Monday 2nd September: 9am-1pm
Esplanade Hotel in Fremantle

Recent developments in design and analysis of experiments in plant breeding and variety testing based on linear mixed models

Prof. Dr. Hans-Peter Piepho, University of Hohenheim

This workshop will consider statistical issues arising in the design and analysis of experiments in plant breeding and variety testing, focusing on recent developments in the field. I will start by considering individual trials and then will move on to the integration of trials across multiple environments, where modelling of genotype-environment interaction is required.

Topics covered for individual trials will include a model-free approach to generating efficient designs, two-phase experimental designs, the use of post-blocking in assessing the merits of alternative designs, recovery of inter-block information, and spatial analysis. In relation to multiple environment trials, I will consider, e.g., sparse testing, the optimal allocation of trials to agro-ecological zones, variance-covariance structures for modelling genotype-environment interaction, the use of environmental covariates to enhance predictions and recommendations, assessment of long-term trends and genetic gain, and two-stage versus single-stage analysis.

Areas of interest:

- Experiment Design
- Spatial Analysis
- Genotype-Environment Interactions

Reliable and Reproducible Analysis with Genstat

Dr. Vanessa Cave & Dr. David Baird (with input from Dr Roger Payne), VSN

This workshop will explore how to move from menus to programs that can be rerun to reproduce the analysis. This includes extracting data from spreadsheet files and databases and producing PDFs containing tables, graphs, and text for clients.

As well as recap highlights of the last ten years of Genstat development, such as

- benefits of HGLMs over GLMMs,
- censored data analysis,
- assessment of bio-equivalence,
- automated REML analyses for spatial data,
- use of the spline facilities for regression and linear mixed models.

The sessions will involve a mixture of examples and practicals, so please bring your laptops (ideally, with Genstat 24 already installed).

Abstracts for Talks: Tuesday

Invited Talk Wed 8:45 - 9:30

Two-dimensional P-spline smoothing for spatial analysis of field trials

Hans-Peter Piepho

University of Hohenheim

Large agricultural field trials may display irregular spatial trends that cannot be fully captured by a purely randomization-based analysis. For this reason, paralleling the development of analysis-of-variance procedures for randomized field trials, there is a long history of spatial modeling for field trials, starting with the early work of Papadakis on nearest neighbor analysis, which can be cast in terms of first or second differences among neighboring plot values. This kind of spatial modeling is amenable to a natural extension using splines, as has been demonstrated in recent publications in the field. Here, we consider the P-spline framework, focusing on model options that are easy to implement in linear mixed model packages. Two examples serve to illustrate and evaluate the methods. A key conclusion is that first differences are rather competitive with second differences. A further key observation is that second differences require special attention regarding the representation of the null space of the smooth terms for spatial interaction, and that an unstructured variance-covariance structure is required to ensure invariance to translation and rotation of eigenvectors associated with that null space. We develop a strategy that permits fitting this model with ease, but the approach is more demanding than that needed for fitting models using first differences. Hence, even though in other areas, second differences are very commonly used in the application of P-splines, our conclusion is that with field trials, first differences have advantages for routine use.

Abstract session 1.1 Tues 10:15 – 12:00

1. Extracting Statistical Elements of Experimental Design with Large Language Models

Emi Tanaka

ANU

The descriptions of experimental designs are often idiosyncratic and verbose, interwoven with details that are geared towards domain experts (e.g. the preparation of the experimental materials). Extracting the statistical elements of the experimental design from these descriptions can be tedious at best, challenging at worst. The emergence of Large Language Models (LLMs) has revolutionized various applications, notably in natural language processing. This talk explores the use of LLMs to streamline the extraction of statistical elements from the descriptions of experiments. This can expedite the distilling of complex experimental design descriptions and aid in formulating an appropriate analysis of experimental data.

2. Estimating infestation rates when naturally infesting products

Carole Wright, Pauline Wyatt, David Mayer and Peter Leach

Department of Agriculture and Fisheries (DAF), Queensland

Phytopsanitary, or quarantine treatments for fresh horticultural produce require a level of treatment efficacy to be determined. This is to provide confidence that no target pest species are moved with the product. The treatment efficacy is often based on the mortality of target pests exposed to the treatment. In naturally infested products, the samples are exposed to mature pests which deposit an unknown number of eggs into each sample. In such cases, the number of pests exposed to the treatment is often unknown and needs to be estimated. The International Plant Protection Convention (IPPC) provides formulae for estimating the infestation rate when naturally infesting. We had concerns over the statistical validity of the formulae which have been an international standard since at least 2014. This talk discusses the deficiencies with the IPPC formulas, our proposed improved method, and the road to acceptance of our concerns and proposed improvement.

3. The need for Interpretable Models – Understanding the impact of confounders and why many attempts to adjust for them are just dodgy accounting

Chris Howden

Sydney Informatics Hub, Sydney University

Statistics competes not only for funding but also relevance with other quantitative fields. Interpretable models drive knowledge creation in science, plus an increasingly diverse range of domains. Unfortunately, there is little understanding that AI or machine learning are rarely interpretable, while statisticians excel in building interpretable models.

We explore the common claim of accounting for confounders by asking “What do researchers mean by this?”, “Are they doing it correctly, or is it just dodgy bookkeeping?”, “How can it be done?”, and “Why is it needed for interpretable models, and why are they so important?”.

Then use this exposition to discuss our strengths and where we need to improve so researchers want us as trusted advisors on the complex issues they face when building interpretable models. Achieving this means a greater share of scarce research dollars and involvement in ground-breaking impactful research. Boosting our profile and ensuring our relevance in a fast-moving quantitative research world.

4. Genome-wide association and genomic selection with multi-environment and multi-trait data in barley

Keifei Chen (Curtin University), Hamid Shirdelmoghanloo (DPIRD), Blakely H. Paynter (DPIRD), Camilla Beate Hill (MU), Sharon Westcott (MU and DPIRD), Tefera Tolera Angessa (MU), Chengdao Li (MU and DPIRD)

The advances in “omics” technologies have enabled unprecedented progress in agricultural and biological sciences. The synergy of high performance computing, high throughput omics approaches, and high dimensional phenotyping has demonstrated the capacity to enhance our understanding of biological mechanisms and provide powerful insights into dissecting the genetic basis of complex traits. Genome-wide association study (GWAS) has become a useful approach to identify mutations that underlie diseases and complex traits. However, it is less suitable for quantitative traits influenced by a large

number of genes with small effects. Genomic selection holds the promises to overcome the limitations by using whole-genome information to predict genetic merits of phenotypes. We present the analytical methods and results of GWAS and genomic selection using multi-environment and multi-trait barley data to identify genomic regions associated with agronomically important traits in barley grown under heat-stressed environments.

5. Single-step genomic selection: accommodating several key issues

Aidan McGarty (UOW), Brian Cullis (UOW), Ahsan Asif (NSW DPI) and Kristy Hobson (NSW DPI)

Genomic selection can be a useful tool in modern plant breeding programs as it allows for the genomic prediction of unphenotyped (unobserved) varieties via the association of a phenotyped trait with high-density genetic marker scores. However, the implementation of models to form this association can be complicated by several issues including, the presence of variety by environment interaction (VEI), complex trial designs and linear dependencies in the matrix of marker scores. The latter can arise in several situations including when genetic clones are present, the matrix of marker scores is centred or there are more varieties than markers. This talk will aim to address several of these issues motivated by an Australian chickpea multi-environment trial (MET) dataset comprising both field trials and controlled environments. The analysis was conducted using DWREml via fitting a single-step factor analytic linear mixed (FALMM) to assess disease resistance.

6. Spatial modelling of plant data with the R-packages ASReml-R and asremlPlus

Chris Brien

University of Adelaide

Candidate two-dimensional spatial models, whose fitting has been implemented in asremlPlus (Brien, 2024) for fitting with ASReml-R, are (i) separable variance models (e.g. assuming ar1 by ar1), (ii) tensor-product smoothing splines, and (iii) tensor-product P-splines for which the degree and the order of differencing the penalty can be specified and for which Piepho et al.'s (2022) P-spline modifications have been incorporated. The models that are fitted and

the three `asremlPlus` functions for fitting and comparing these models will be described. Their application will be illustrated using a two-greenhouse, 1100-pot, high-throughput phenotyping experiment that involved 215 barley lines.

Reference Brien, C. J. (2024) `asremlPlus`: augments `ASReml-R` in fitting mixed models and packages generally in exploring predictions. R package version 4.4.29 <http://cran.at.r-project.org/package=asremlPlus> or <http://chris.brien.name/rpackages>.

Piepho, H.-P., Boer, M. P., & Williams, E. R. (2022). Two-dimensional P-spline smoothing for spatial analysis of plant breeding trials. *Biometrical Journal*, 64(5), 835-857.

Invited Talk Tues 1:00 – 1:45

Building Trustworthy Environmental Models

Lucy Marshall

Macquarie University

Environmental systems are complex, with many interacting variables (biotic and abiotic) across diverse spatial and temporal scales. Historically, modellers have used scientific knowledge of how these systems function to build tractable modelling frameworks. The resultant models are used today in multiple operational settings including natural resource management, early warning systems for hazards such as floods, or ecosystem restoration. This approach to model building is often favoured because it suggests a level of reliability when models are extrapolated to new conditions or scenarios that haven't been observed previously. Despite this, there is an explosion of interest recently in modelling approaches that capitalise on recent advances in machine learning (ML), Artificial Intelligence (AI) and related technologies. These approaches have shown promise in initial applications to environmental problems, but scepticism remains about their usefulness given their disconnect from the knowledgebase of environmental systems. This leads to a fundamental question: How can we effectively combine these innovative approaches with the established understanding of environmental processes? We demonstrate here the integration of traditional process-based environmental models with emerging modelling technologies, aiming to bolster the predictive power and trustworthiness of environmental models. We demonstrate how new hybrid approaches to ML

offer a pathway to refining our understanding of natural processes, enhancing predictions, and addressing pressing environmental challenges with improved accuracy.

Abstract session 1.2 Tues 1:45 – 3:00pm

1. Examining Possible Causal Relationships Between Pornography Use and Mental Wellbeing: An Ecological Momentary Assessment Approach

Nathan I. N. Henry¹ (*Student*), *Mangor Pedersen*¹, *Matt Williams*², *Liesje Donkin*¹

¹*Auckland University of Technology*; ²*Massey University*

To assess causality in behavioral addictions, a longitudinal modelling framework is generally required. Utilizing an Ecological Momentary Assessment (EMA) design, we measured the affective dynamics of mental wellbeing variables before, during and after pornography use among individuals recruited from an online forum. Participants completed a four-week EMA, capturing data on sexual activities and mental wellbeing variables. Bayesian hierarchical mixed-effects modelling was employed to analyze affective dynamics. We found that participants experienced a complex interplay between pornography use and other negative emotional states, which were exacerbated by feelings of guilt and shame. Using temporal markers provided by EMA data, we were able to propose causal relationships between pornography use and its effects on mental wellbeing. The EMA approach produced several unique findings and may be the ideal method for examining the effects of abstinence from pornography use.

2. Utilization of Maternal Healthcare Services Among Women in 33 Low-and Middle -Income Countries: A Pooled Analysis of 2015-2022 Demographic and Health Survey Data

Abdul Baten (Student) (Department of Health Science and Biostatistics, Swinburne University of Technology (SUT), Melbourne, Victoria, Australia) , Raaj Kishore Biswas (Faculty of Medicine and Health, The University of Sydney, New South Wales, Australia), Evie Kendal (SUT), and Jahar Bhowmik (SUT)

Maternal mortality presents a significant global public health challenge in low- and middle- income countries (LMICs). The effective utilization of maternal healthcare services (MHSs) including antenatal care (ANC), skilled birth attendance (SBA), institutional delivery (ID), and postnatal care (PNC), is crucial for achieving improved maternal health outcomes. We investigated the utilization of MHS among women in 33 LMICs around the globe using Demographic and Health Surveys. We fitted complex survey-adjusted logistic regression models for each outcome separately in a combined data set of all 33 surveys. We observe high heterogeneity in the utilization of MHS across countries; for example, Indonesia was 10.12, 8.14, 3.65 and 4.19 times more likely to utilize ANC (95% CI= 9.00,11.39), SBA (7.01, 9.46), ID (3.23, 4.13) and PNC (3.65, 4.81) compared to Bangladesh. Given the heterogeneity in MHS uptakes, country- specific adaptation of successful interventions might be a way forward to achieve relevant Sustainable Development Goals by 2030

3. Comparison of urban-rural inequality in quality antenatal care among women in Bangladesh and Pakistan: a multivariate decomposition analysis

Farjana Misu (Student), Taslima Rahman, Dominic Gasbarro, and Khurshid Alam Murdoch University

Urban-rural inequality in the utilization of quality antenatal care (ANC) is a well-documented challenge in low- and middle-income countries, such as Bangladesh and Pakistan. This study investigates urban-rural inequality in the utilization of quality ANC in Bangladesh and Pakistan. We decomposed inequalities in the utilization of quality ANC among urban and rural women in Bangladesh and Pakistan using the Oaxaca, the Blinder, and related decompositions for nonlinear models. To quantify covariate contributions to the urban-rural inequality, we employed the Blinder-Oaxaca multivariate

decomposition analysis for nonlinear response models. Using data from the latest Demographic Health Surveys (2017-2018), the study reveals significant urban-rural inequality in Bangladesh and Pakistan, which is more pronounced in Pakistan. Wealth difference has the largest contribution percentage among the common significant predictors for both countries. In Pakistan, women's education is the second largest contributor to inequality, while in Bangladesh, it is media exposure. Tailored strategies are required to mitigate these inequalities in ANC.

4. Optimising Multiplicative Noise Data Masking

Pauline Ding, Harrison Rowles, John Brackenbury and Y.X. Lin

National Institute for Applied Statistics Research Australia, University of Wollongong, Australia.

Multiplicative noise masking is a well-known method to perturb data for privacy protection purpose. Twin uniform distribution has been introduced in the literature as a distribution for multiplicative noise, given the simplicity in its mathematical form and the ability to provide good value protection without sacrificing statistical utility. We explore the impact of various distribution parameters on the performances of privacy protection and utility loss when multiplying twin uniform noise for data masking. We proposed an approach to optimise the multiplicative noise scheme with twin uniform noise distribution. We applied the optimisation algorithm to a real accounts payable dataset, and conclude that it yields good results for both privacy and utility.

5. A better way to calculate power for preliminary two sample proportion comparisons

Damian Collins and Paul Hick

NSW DPI

Animal models are essential in pre-clinical development of vaccines. This work is motivated by preliminary studies to determine that a candidate vaccine induces a protective immune response suitable for further evaluation. Such a preliminary study needs adequate statistical power to detect a large “all-or-nothing” difference for the probability of disease (e.g. $P_1 = 0.975$ vs P_2

= 0.3) with small sample sizes (< 10 per group) to minimise the economic and welfare costs of using animals in research. Most “off-the-shelf” power calculators for two proportion comparisons use approximate methods (e.g. `power.prop.test` function in R) despite being inappropriate for small sample sizes. Alternatively, since the sample space of all possible outcomes is small, the true power can readily be calculated for a more “exact” test such as Fisher or Barnard. We propose that this power calculation method be used for preliminary studies in the interests of improving animal welfare.

6. Robust Bayesian experimental design through flexible modelling structures

James McGree

QUT

Bayesian experimental design is a well-established methodology for planning data collections. With such an approach, designs are typically found by maximising the expectation of a utility function with respect to the joint distribution of the parameters and the response, conditional on an assumed statistical model. In practice, specifying such a model can be difficult due to incomplete knowledge of the data generating process. This can be rather problematic as a misspecified model can lead to inefficiencies in data collection and/or conclusions that are misleading. To address this, we present an approach to find Bayesian designs that are robust to the assumed model. To do so, we propose to determine designs based on flexible modelling structures such as those based on spline models. This approach is motivated by real-world sampling problems in environmental monitoring and agriculture where we assess the performance of our methodology against more standard practices.

Invited Talk Tues 3:30 – 4:15

Genstat - some wider applications

David Baird, Roger Payne and Vanessa Cave

VSNi International

Part 1: “Genstat, the birder’s friend”, will describe how Genstat is used to process and check the over 750,000 bird sightings recorded each year in the London area. These need to be imported from a wide range of sources and interpreted, validated and collated so that they can be summarized by the authors of the London Bird Club’s annual reports – see <https://lnhs.org.uk/index.php/publications/london-bird-report>. This involves many different aspects, including text mining, matching and comparison, UTM coordinates and grid references, report writing, tabulation, manipulation and, of course, the spreadsheet. There is still scope for statistical analysis too, and many of the counts suffer upper-censoring.

Part 2: The use of Genstat to predict the allocation of damage to the Christchurch earthquakes will be described. This used a large (186,000 cases) K-nearest-neighbour analysis that was tuned by simulation, and whose standard errors were calculated using bootstrapping. The model was then validated by manually assessing 4000 properties and comparing these to their predictions.

Part 3: Fun & games – finally, we will play some of the classic games available in Genstat: fruit machine, Conway’s Game of life and noughts and crosses. Although primarily there for a bit of light-hearted fun, these games also demonstrate interactive procedures, and illustrate how to create “movies” by the means of an animated series of graphs or PNG images.

Abstract session 1.3 Tues 4.15 – 5:30pm

1. An efficient resampling scheme for outlier detection in linear mixed models

Lu Wang (Student), Alison Smith, Brian Cullis and Carole Birrell

University of Wollongong

Outlier detection is one of the most critical procedures in statistical modelling and should be considered before any formal testing is performed (Cook and Weisberg, 1982). The alternative outlier model (AOM) approach of Cook et al. (1982) was developed in the context of ordinary linear models. The most general approach for outlier detection in a linear mixed model (LMM) framework is presented in Haslett and Hayes (1998) and Haslett and Haslett (2007). Their focus was on the definition and roles of residuals with a general covariance structure.

Gumedze et al. (2010) and Gumedze (2018) considered simple variance component models and proposed likelihood ratio and score test statistics to determine whether individual observations have inflated variance. Under the assumption of independent random effects and residuals, this approach appeared to provide reasonable Type I error rate for multiple testing. The full parametric bootstrap procedure was reported to be computationally demanding.

In this talk, we extend the use of AOM in a LMM framework and derive residual maximum likelihood (REML) score tests applicable for residuals in the model. Our approach also accommodates correlated effects and uses an efficient resampling scheme that does not require re-fitting the null model in each iteration. A simulation study shows that our approach provides accurate Type I error rates.

Reference

Cook, R. D. and Weisberg, S. Residuals and Influence in Regression. New York: Chapman and Hall.1982.

Cook, R.D., Holschuh, N. and Weisberg, S. A note on an alternative outlier model. Journal of the Royal Statistical Society, vol. 44, pp. 370-376.1982.

Haslett, J. and Hayes, K. Residuals for the Linear Model with General Covariance Structure. Journal of the Royal Statistical Society Series B: Statistical Methodology, 60(1), pp. 201-215, 1998.

Haslett, J. and Haslett, S. The Three Basic Types of Residuals for a Linear Model. *Int Statistical Rev*, 75(1), pp. 1–24, 2007.

2. Semi-supervised model-based clustering for ordinal response data

Ying Cui (Student), Louise McMillan and Ivy Liu

Victoria University of Wellington

This paper introduces a semi-supervised learning technique for model-based clustering. Our research focus is on applying it to matrices of ordered categorical response data, such as those obtained from the surveys with Likert scale responses. We use the proportional odds model, which is popular and widely used for analysing such data, as the model structure. Our proposed technique is designed for analysing datasets that contain both labeled and unlabeled observations from multiple clusters. To evaluate the performance of our proposed model, we conducted a simulation study in which we tested the model from six different scenarios, each with varying combinations and proportions of known and unknown cluster memberships. The fitted models accurately estimate the parameters in most of the designed scenarios, indicating that our technique is effective in clustering partially-labeled data with ordered categorical response variables. To illustrate our approach, we use a real-world dataset from aquaculture area.

3. Cokrig-and-Regress for Spatially Misaligned Environmental Data

Zhi Yang Tho, Francis Hui, Alan Welsh, Tao Zou

ANU

Spatially misaligned data, where the response and covariates are observed at different spatial locations, commonly arise in many environmental studies. Motivated by spatially misaligned data collected on air pollution and weather in China, we propose a cokrig-and-regress (CNR) method to estimate spatial regression models involving multiple covariates and potentially non-linear associations. The CNR estimator is constructed by replacing the unobserved covariates (at the response locations) by their cokriging predictor derived from the observed but misaligned covariates under a multivariate Gaussian assumption, where a generalized Kronecker product covariance is used to

account for spatial correlations within and between covariates. Simulation studies demonstrate that CNR outperforms several existing methods for handling spatially misaligned data, such as nearest-neighbour interpolation. Applying CNR to the spatially misaligned air pollution and weather data in China reveals a number of non-linear relationships between $\text{PM}_{2.5}$ concentration and several meteorological covariates.

Reference

Tho, Z. Y., Hui, F. K. C., Welsh, A. H, and Zou T. (2023). Cokrig-and-Regress for Spatially Misaligned Environmental Data. arXiv preprint arXiv:2310.05548.

4. Confidence intervals for proportions estimated by pooled testing based on Firth’s bias correction

*Brad Biggerstaff (CDC, USA), **Graham Hepworth** (University of Melbourne)*

Pooled testing (or group testing) arises when units are pooled together and tested as a group for the presence of an attribute, such as a disease. It originated in blood testing, but has been applied in many fields, including prevalence estimation of mosquito-borne viruses and plant disease assessment – the two fields in which we have encountered the technique.

Confidence intervals for proportions estimated by pooled testing have involved both exact and asymptotic methods. Hepworth and Biggerstaff (2017, 2021) showed Firth’s correction (Firth 1993) to maximum likelihood estimation to reduce bias effectively for pooled testing. Considering the Firth-corrected score as the first derivative function of a penalised likelihood, we develop confidence intervals wholly within this framework, evaluate their performance, and compare them to the existing, recommended asymptotic method.

The methods are illustrated using data on yellow fever virus and West Nile virus infection in mosquitoes.

Reference

Hepworth G. and Biggerstaff, B.J. (2017) Bias correction in estimating proportions by pooled testing. JABES, 22, 602–614.

Hepworth G. and Biggerstaff, B.J. (2021) Bias correction in estimating proportions by imperfect pooled testing. JABES, 26, 90–104.

Firth, D. (1993) Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27–38.

5. Structured Big Data!! How to apply random effects models effectively, and efficiently

Jin Yoon and Justin Boyle

AEHRC CSIRO

The big data era demands new statistical analysis paradigms, specially, structured big data such as multicentre data, there are challenges to analysing complex data including privacy protection, large-scale computation resource requirements, heterogeneity, and correlated observations. Random effects models (REMs) are one of the most popular statistical methods for complexity of the data such as nested structure data, but these methods are often infeasible for big and complicated data due to memory and storage limitations of standard computers. To address this, we use a method “Divide and Combine (D&C or DAC)”, where results are combined over sub-analyses performed in separate data subsets using two approaches, summary statistics DAC (federated DAC) and horizontal DAC (centralised DAC). To assess their efficacy, we apply both approaches to real data and compare them with the gold-standard method of directly fitting REMs to the pooled dataset.

Reference

Chun Wang, et al. (2016). Statistical methods and computing for big data. *Stat Interface*. 9(4): 399-414. Doi:10.4310/SII.2016.v9.n4.a1

[William S. Cleveland and Ryan Hafen. (2014). Commentary Divide and Recombine (D&R): Data science for Large complex data. *Statistical analysis and Data mining*, Vol. 7. Doi:10.1002/sam.11242

Yan Z, et al. (2023). A privacy-preserving and computation efficient federated algorithm for generalized linear mixed models to analyze correlated electronic health records data. *PLoS ONE* 18(1):e0280192. <https://doi.org/10.1371/journal.pone.0280192>

6. Integrating and visualising multi-omics data with the moiraine R package

Olivia Angelin-Bonnet

The New Zealand Institute for Plant and Food Research Limited

The integrated analysis of multi-omics datasets is a challenging problem, owing to their complexity and high dimensionality. Numerous integration tools have been developed, but they are heterogeneous in terms of their underlying statistical methodology, input data requirements, and visual representation of the results generated.

I present the moiraine R package, which provides a framework to consistently integrate and visualise multi-omics datasets. In particular, moiraine enables the construction of insightful and context-rich visualisations that facilitate the interpretation of integration results by domain experts. With moiraine, it is also possible to compare the results obtained with different integration tools, providing confidence into the biological relevance of the results obtained.

Reference

Angelin-Bonnet O (2024). moiraine: Construction of Reproducible Pipelines for Testing and Comparing Multi-omics Integration Tools. R package version 1.0.0, <https://github.com/Plant-Food-Research-Open/moiraine>.

Poster

Reporting randomised factorial trial results: the case of the Topical Analgesia Post - Haemorrhoidectomy trial

James Jin¹, Weisi Xia¹, Runzhe Gao¹, **Alain C. Vandal**², Maree Weston¹, Lincoln Israel¹, Andrew Connolly¹, Primal Singh¹, Darren Svirskis¹, Andrew Hill^{1,2}

¹ Te Whatu Ora Counties Manukau; ² The University of Auckland

A randomised factorial trial is usually designed to detect the smallest clinically meaningful difference for all interventions, absent interaction. When a negative interaction is apparent without a sufficient sample size to test for it, the question arises as to how to best report the results, given the constraints of the statistical analysis plan (SAP). We present the rationale and decisions made in this regard when reporting the results from the Topical Analgesia Post-Haemorrhoidectomy randomised factorial trial, in which a significant difference between a lidocaine-only and a lidocaine-diltiazem formulation injected unexpected variability in the primary outcome, a pain visual analogue scale. While the SAP dictated the reporting of the main effects only, appropriate translation to clinical practice required separate reporting of the four arms. We describe the trial's design, longitudinal data collection process, main analysis, and explain how we dealt with this issue.

Reference

Jin J, Xia W, Gao R*, Vandal AC, Weston M, Israel L, Connolly A, Singh P, Svirskis D & Hill A. (2024). A randomised controlled trial of topical analgesia post-hemorrhoidectomy (TAPH Trial). Diseases of the Colon and Rectum, accepted 2024.05.15.

Using Genstat to compare the responses of sweet corn varieties to fall armyworm, *Spodoptera frugiperda*.

Angela Anderson, Siva Subramaniam

Department of Agriculture and Fisheries (DAF), Queensland

Fall armyworm (FAW) was first detected in 2020 in Queensland. It quickly spread to several locations in Australia, causing up to 80% losses in sweet corn and maize¹. This paper describes the design and analysis of a plot trial conducted in north Queensland which assessed the susceptibility and preference of commercially available sweetcorn varieties to FAW infestation.

Genstat was used to generate the (latinized) doubly resolvable row-column design of four replicates of eight varieties. Variables measured included FAW oviposition rate (egg masses), number of larvae and the intensity of foliage damage. Analyses were conducted in Genstat using hierarchical generalised linear models and with the use of the VAROWCOLDESIGN procedure. This procedure allowed the analysis of the row-and-column design by REML, with the convenience of automatic selection of the best random and spatial covariance model. Significant differences were found between varieties in the number of total alive FAW larvae.

Reference ¹Subramaniam, S et al (2022). Identifying potential parasitoids of the fall armyworm, *Spodoptera frugiperda*, and the risk to Australian horticulture. Final report MT19015. <https://www.horticulture.com.au/globalassets/laserfiche/assets/project-reports/mt19015/mt19015-final-report-complete.pdf>

Genomic selection in the presence of variety by environment interaction: Using single-step factor analytic linear mixed models to define interaction classes

Chris Lisle (InterGrain Pty Ltd), Alison Smith (University of Wollongong), Monique Jordan (University of Wollongong), David Tabah (InterGrain Pty Ltd), Calum Watt (InterGrain Pty Ltd), Jayfred Godoy (InterGrain Pty Ltd), Daniel Mullan (InterGrain Pty Ltd), Brian Cullis (University of Wollongong)

Plant breeding multi-environment trials (METs) are an important component of identifying superior varieties as they allow an assessment of variety by environment interaction (VEI). A key consequence of the presence of VEI

is that it is both statistically invalid and biologically meaningless to base variety selections on measures of overall performance across all environments in the MET. In this poster we address the issue for genomic selection within the framework of a single-step linear mixed model analysis in which a factor analytic model is used for the variety effects in each environment. Our approach for variety selection involves the formation of groups of environments, called interaction classes (iClasses), within which there is minimal crossover VEI. It is then valid to obtain predictions of variety performance across the environments in each iClass. We demonstrate the methods using a commercial Wheat MET dataset.

The definition of no interaction: for detecting synergistic or antagonistic interaction

Damian Collins, Grant Chambers and Nerida Donovan

NSW DPI

Interactions between stressors are defined as synergistic if the combined effect is greater than the *sum* of each effect independently, or antagonistic if less. This implies that the null hypothesis (of no interaction) is that the stressors combine *additively* (since it is the sum). But would a null hypothesis of combining *multiplicatively* be more sensible? For instance, if both stressors reduce growth by 50%, should one expect that the combined effect would be 0% or 25% growth if there was no interaction? The motivation for research comes from investigations into viroid interactions on the growth of citron (*C. medica* L.) plants. The issues surrounding fitting statistical models to studies examining interactions between stressors, such as pathogen interactions, will also be discussed

Quantitative disease traits to describe *Rhizoctonia* root rot severity

Karyn Reeves (DPIRD), **Bethany Rognoni** (USQ), **Clayton Forknall** (DAFQ), **Blake Gontar** (SARDI), and **Daniel Huberli** (DPIRD)

*Department of Primary Industries and Regional Development (DPIRD),
Western Australia*

Rhizoctonia root rot is a fungal soilborne disease that affects cereal and broadleaf crops. It is caused by a long-lived necrotrophic fungus which grows through the top layer of the soil decaying and pruning the plant roots it infects.

In field trials involving *Rhizoctonia* root rot, disease severity has traditionally been assessed using bespoke rating schemes on plants removed from plots with ratings that reflect visual markers of infection, such as lesion colour and the presence of pruned roots, and a grading of how badly a plant is affected. The resulting qualitative ordinal data is then often inappropriately treated as metric data.

We present alternative quantitative traits for assessing disease severity which were used alongside the conventional approach in experiments conducted to assess *Rhizoctonia* root rot incidence and severity. These quantitative traits can be analysed with standard methods, are easier to interpret, and offer an approach which can be consistently applied across trials and through time.

MOTL: enhancing multi-omics matrix factorization with transfer learning

*D. Hirst (Aix Marseille Univ, France), M. T  r  zol (Aix Marseille Univ, France), L. Cantini (Institut Pasteur, France), P. Villoutreix (Aix Marseille Univ, France), **Matthieu Vignes** (Massey University, New Zealand), A. Baudot (Aix Marseille Univ & CNRS, France and Barcelona Supercomputing Center, Spain)*

Joint matrix factorization is a popular method for extracting lower dimensional representations of multi-omics data. It disentangles underlying biological signals, facilitating efficient sample clustering, disease subtyping, or biomarker identification. However, with a limited number of samples, the effectiveness of matrix factorization is reduced. Addressing this limitation, we introduce MOTL (Multi-Omics Transfer Learning), a novel framework for multi-omics matrix factorization with transfer learning based on MOFA (Arguelet et al. 2018). MOTL infers latent factors for a small multi-omics dataset, with

respect to those inferred from a large heterogeneous learning dataset. We designed two evaluation protocols based on simulated and real multi-omics data. We observed that MOTL improves the factorization of multi-omics datasets, comprised of a limited number of samples, when compared to factorization without transfer learning. We showcase the usefulness of MOTL on a small sample glioblastoma dataset, revealing an enhanced delineation of cancer status and subtype thanks to transfer learning.

Reference

Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., Buettner, F., Huber, W., and Stegle, O. (2018). Multi-Omics Factor Analysis - a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology*, 14(6):e8124.

Joint learning of Gaussian graphical models for paired data in high-dimensional and heterogenous population

Dung Ngoc Nguyen (CSIRO Agriculture and Food, Canberra, ACT, Australia)
Alberto Roverato (Department of Statistical Sciences University of Padova, Italy)
Zitong Li (CSIRO Agriculture and Food, Canberra, ACT, Australia)

The Gaussian graphical model (GGM) has become a popular tool for biological network analysis to study conditional dependence between biological variables and predict gene functions, phenotypes and patterns of molecular regulation. We consider the problem of learning a GMM for paired data where observations come from two dependent groups sharing the same variables. Comparing the distribution of a set of variables between two experimental conditions, or groups, is common to many real-world applications such as cancer genomic data or brain networks from fMRI data. To address the problem of high dimensionality and heterogeneity, we apply a mixture of GGMs for paired data and develop a penalized EM algorithm for these models where model estimation and selection can be performed simultaneously. The efficiency of our approach is then demonstrated on constructing gene networks using both real and simulated transcriptome data sets.

Reference

David Edwards. Introduction to Graphical Modelling. Springer, 2nd edition (2000).

Danaher, P., Wang, P., Witten, D.M.: The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical*

Society Series B: Statistical Methodology 76(2), 373–397 (2014).

Lartigue, T., Durrleman, S., Allasonniere, S.: Mixture of conditional Gaussian graphical models for unlabelled heterogeneous populations in the presence of co- factors. SN Computer Science 2(6), 466 (2021).

Roverato, A., Nguyen, D.N.: Exploration of the search space of Gaussian graphical models for paired data. Journal of Machine Learning Research 25(92), 1–41 (2024).

Højsgaard, S. and Lauritzen, S.L.: Graphical Gaussian models with edge and vertex symmetries. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70(5), 1005–1027 (2008).

Lauritzen, S.L.: Graphical Models. Oxford University Press (1996).

Ranciati, S., Roverato, A., Luati, A.: Fused graphical lasso for brain networks with symmetries. Journal of the Royal Statistical Society Series C: Applied Statistics 70(5), 1299–1322 (2021).

Ranciati, S., Roverato, A.: On the application of Gaussian graphical models to paired data problems. arXiv preprint arXiv:2307.14160 (2023).

Quantifying variability in soil organic carbon estimates in Australian agriculture

Rik Rutjens, K. Reinke, S. Georgiou, S. Stylianou and S. Grover

RMIT

As part of the Australian Government’s efforts to reduce greenhouse gas emissions, farmers can gain Carbon Credits by sequestering soil carbon through improved management practices^[1]. However, measuring changes in soil organic carbon (SOC) is difficult, as changes in SOC are typically small, whilst variability in the field (both spatially and temporally) is often substantial and measurement errors can be considerable. Additionally, the sample size is generally small; the Government requires a minimum of 9 samples, whereas plots can be as large as 200 ha. This work therefore aims to identify improved sampling protocols to reliably detect changes in SOC in a practically feasible way. As a first step, a large sampling campaign was undertaken, in which over 1000 cores were extracted from 9 farms in eastern Australia. This poster highlights preliminary results on variability in SOC estimates, and showcases the experimental design for the second round of sampling.

Reference

- [1] Commonwealth of Australia. Carbon Credits (Carbon Farming Initiative – Estimation of Soil Organic Carbon Sequestration using Measurement and Models) Methodology Determination 2021. Canberra, Australia, 2021.
-

A Mutual Influence Model for Two-Mode Network Data

Rui Shen (Student)

ANU

We introduce a mutual influence model for two-mode network that can effectively capture both inward and outward influences within the network. To estimate the unknown parameters, we employ quasi-maximum likelihood estimation and derive the asymptotic properties of our estimators. Furthermore, we propose multiple hypothesis tests to evaluate the inward and outward influence of the network. To assess the model performance, we conduct numerical studies and real data analysis.

We apply the proposed model to a dataset obtained from Yelp, the largest review platform in the US. It is worth noting that our model could similarly be employed on datasets from other platforms, such as eBay and Amazon, which have become increasingly important in the digital economy. We aim to reveal the outward and inward influences of both customers and restaurants. Our dataset contains 171 restaurants and 815 consumers. The mutual influence structure could effectively explain the customer and restaurant features.

Analysis of Stability Classes for Atmospheric Dispersion in the Marine Boundary Layer Using Bayesian Inference

Samuel Davis (Student) (DARE USYD), Matthew Cleary (USYD), Clara Grazian (USYD)

We propose a Bayesian Inference approach to infer the unknown parameters of a Gaussian plume model to predict the behaviour of a marine cloud brightening (MCB) Plume on the Great Barrier Reef. Normalised data from 12 experiments collected from an MCB plume has been used alongside prior knowledge of atmospheric stability to understand the overall time averaged behaviour of the plume. This research employs two different priors for the inferred parameters. The study focuses mostly on the diffusion parameters of the Gaussian plume

model and how they relate to the Pasquill-Gifford Stability classes. It compares the effectiveness of a single and multi-mode prior distribution. While each method has its advantages and disadvantages, we have concluded that the multi-modal Prior stands out as the method which has the best balance between interpretability and performance.

Reference

Latham, John, Kleypas, Joan, Hauser, Rachel, Parkes, Ben, & Gadian, Alan. (2013). Can marine cloud brightening reduce coral bleaching? *Atmospheric Science Letters*, 14(4), 214-219. <https://doi.org/10.1002/asl2.442>

Kahl, Jonathan D.W., & Chapman, Hillary L. (2018). Atmospheric stability characterization using the Pasquill method: A critical evaluation. *Atmospheric Environment*, 187, 196-209. <https://doi.org/10.1016/j.atmosenv.2018.05.058>

Melli, P., & Runca, E. (1979). Gaussian Plume Model Parameters for Ground-Level and Elevated Sources Derived from the Atmospheric Diffusion Equation in a Neutral Case. *Journal of Applied Meteorology and Climatology*, 18(9), 1216-1221. [https://doi.org/10.1175/1520-0450\(1979\)018](https://doi.org/10.1175/1520-0450(1979)018)

Senocak, I., Hengartner, N. W., Short, M. B., & Daniel, W. B. (2008). Stochastic event reconstruction of atmospheric contaminant dispersion using Bayesian inference. *Atmospheric Environment*, 42, 7718-7727. <https://api.semanticscholar.org/CorpusID:43774456>

Harrison, D. P., Harrison, L., Cravigan, L., & Kelaher, B. (2020). Marine cloud brightening proof concept study: boosting coral abundance on the Great Barrier Reef Challenge (Final Report). Sydney Institute of Marine Science, Australia.

Childhood Risk and Resilience Factors for Pasifika Youth Respiratory Health: Accounting for Attrition and Missingness

Siwei Zhai (Student) (The University of Auckland), *Alain C. Vandal* (The University of Auckland, Te Whatu Ora Counties Manukau), *Shabnam Jalili - Moghaddam* (Auckland University of Technology), *Catherine A. Byrnes* (The University of Auckland), *Conroy Wong* (Te Whatu Ora Counties Manukau, The University of Auckland), *Leon Iusitini* (Auckland University of Technology), *El-Shadan Tautolo* (Auckland University of Technology)

Abstract: In New Zealand, 7% of deaths are related to respiratory diseases, with Pacific people at higher risk. Our work investigates the causal effects of early-life risks and resilience factors on early-adulthood lung function amongst Pacific Islands Families Study (PIFS) cohort members ($n = 1,398$). 466 from the cohort participated in the respiratory study. Primary outcome was forced expiratory volume in 1 second (FEV1) z-score at age 18 years. FEV1 and healthy lung function (HLF), defined as the z-score being larger than -1.64, were secondary outcomes. A previous study had evaluated the effects of early-life nutrition factors on the respiratory health of Pacific youth. The results suggested a positive impact of consuming more fruit and vegetables during childhood on respiratory health later in life. The follow-up study will continue to explore the effects of factors from relevant domains based on the PIFS cohort, where a new integrated model will be applied. A simulation will be conducted to determine this model.

Reference

- Riley, L., & Cowan, M. (2014). World Health Organization noncommunicable diseases country profiles. Geneva, Switzerland: WHO Library Cataloguing-in-Publication Data, 156-165.
- Asthma and Respiratory Foundation of New Zealand. (2015). Te Hā Ora (the breath of life): national respiratory strategy.
- Herring, M. J., Putney, L. F., Wyatt, G., Finkbeiner, W. E., & Hyde, D. M. (2014). Growth of alveoli during postnatal development in humans based on stereological estimation. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 307(4), L338-L344.
- Butler, J. P., Loring, S. H., Patz, S., Tsuda, A., Yablonskiy, D. A., & Mentzer, S. J. (2012). Evidence for adult lung growth in humans. *New England Journal of Medicine*, 367(3), 244-247.
- Narayanan, M., Owers-Bradley, J., Beardsmore, C. S., Mada, M., Ball, I.,

Garipov, R., ... & Silverman, M. (2012). Alveolarization continues during childhood and adolescence: new evidence from helium-3 magnetic resonance. *American journal of respiratory and critical care medicine*, 185(2), 186-191.

A nearest-neighbour Gaussian process spatial factor model for censored, multi-depth geochemical data

Tilman M. Davies (University of Otago), Sudipto Banerjee (UCLA)

We investigate the relationships between local environmental variables and the geochemical composition of the Earth in a region spanning over 26,000 km² in the lower South Island of New Zealand. Part of the Southland–South Otago geochemical baseline survey—a pilot study pre-empting roll-out across the country—the data comprise the measurements of 59 chemical trace elements, each at two depth prescriptions, at several hundred spatial sites. We demonstrate construction of a hierarchical spatial factor model that captures inter-depth dependency; handles imputation of left-censored readings in a statistically principled manner; and exploits sparse approximations to Gaussian processes to deliver inference.

Abstracts for Talks: Wednesday

Invited Talk Tues 8:45 – 9:30am

ARIMA Modelling To Determine Trigger Thresholds For Changes In Groundwater Levels

Anna Hayes

Data Analysis Australia

Data Analysis Australia was engaged to design a statistically sound method to determine operational trigger thresholds to determine whether groundwater levels differ beyond what would be considered normal after the building and filling of a series of solar salt evaporation ponds by our client. Due to limited ‘before’ data and high levels of variability in the measured groundwater levels, we developed a method to detect changes using Auto-Regressive Integrated

Moving Average (ARIMA) models that use data from corresponding impact and reference bores to forecast expected groundwater levels at the impact bores and define a trigger to occur when the outcome is not within a certain confidence interval from the forecasts. We determined the appropriate reference bores for each impact bore using a process called Dynamic Time Warping (DTW). I will discuss the development of the statistical methodology from the consulting perspective.

Invited Talk Wed 10:45 – 11:00am

The role of priors in Bayesian analysis

Kerrie Mengersen

Queensland University of Technology

Bayesian methods are now pervasive in applied statistical modelling and analysis. One of the benefits - and challenges - of a Bayesian approach is understanding and formulating priors. In this presentation, I will discuss some of our applied projects in environment and health, and the characterisation of objective, semi-informative and subjective priors. Objective priors will be cast in the context of estimating the probability of presence given a sample comprising only absences and no presences. Semi-informative priors will be addressed through consideration of history matching. Subjective priors will be obtained from citizen science in the context of a Virtual Reef Diver project. The formulation of priors for estimation of mixture models and Hidden Markov models will also be discussed.

This research is joint with a range of collaborators who are listed in the references and will be acknowledged in the presentation.

References

Tuyl F, Gerlach R, Mengersen K. On the Certainty of an Inductive Inference: The Binomial Case. *Statistical Science*, 39 355-356 (2024)

E Santos-Fernandez, J Vercelloni, A Price, et al. Increasing trust in new data sources: crowdsourcing image classification for ecology. *International Statistical Review* 92 (1), 43-61 (2024)

Wang, X., Nott, D. J., Drovandi, C. C., Mengersen, K., & Evans, M. Using History Matching for Prior Choice. *Technometrics*, 60(4), 445–460 (2018).

Abstract session 2.1 Wed 11:00 – 12:15

1. Identifying transient astronomical sources in MeerKAT light curves using Gaussian Processes

Shih Ching Fu (Student)

Curtin University

A key problem in astronomy is the identification of novel celestial objects in the sky. Of particular interest is studying so-called “transient” sources whose changing brightness over time reveals interesting properties about our universe. However, these novel sources must be found among the thousands of other sources that a wide field-of-view astronomical survey might observe. The time trace of a source’s changing brightness over time is known as its light curve, and my work entails the analysis of light curves as observed by MeerKAT, a Square Kilometre Array precursor telescope in South Africa. I have applied Gaussian process (GP) regression to these light curves and found that the distribution of the fitted GP hyperparameters revealed patterns useful for distinguishing between different types of celestial objects. I have compared my results with the variability metrics more commonly used in radio astronomy and found that my approach grants improved discriminatory power and interpretability.

2. Spatial variation in crustacean moult timings

Stephen Bradshaw (Student) (UTAS, DPIRD), Klaas Hartmann (UTAS), Katie Creswell (UTAS), Caleb Gardner (UTAS), Denham Parker (CSIRO)

Understanding crustacean moulting is vital in fisheries, as this phase increases vulnerability, resulting in higher mortality, product loss, and reduced value for certain species. Fisheries often implement closures during moulting, but timing is challenging due to geographic variations or sensitivity to environmental conditions. We introduce a Bayesian model for estimating moult timing using routinely collected datasets in crustacean fisheries, applied to Southern Rock Lobster around Tasmania. Our method, utilizing growth increments, appendage damage, and pleopod regeneration data, offers a robust analysis of spatial variation. It highlights a minor misalignment between the commercial fishing season closure and the moulting period in Tasmania. This flexible approach can be widely applied, integrating diverse data types available in different fisheries, providing a comprehensive understanding of moulting characteristics for effective management.

3. Enhancing Spatial Anomaly Detection for Improved Disease Mapping: A Pragmatic Approach

Farzana Jahan (Murdoch University (MU)), Wafaa Mansoor (MU), Farjana Misu (MU), Susanna Cramb (QUT) and Jessica Cameron (CCQ)

The increase in computational power, coupled with the availability of location data and advanced statistical methods, has fuelled interest in disease mapping and spatial epidemiology [1]. However, the prevalence of geocoding errors and misallocated data introduces significant challenges to the accuracy of these analyses. Yet these anomalies are present in most Australian administrative health datasets, including cancer registry and hospital datasets. These misallocated cases can have a substantial impact on analyses conducted, including misidentifying relationships between covariates and outcomes [2, 3]. This project aims to develop practical and efficient methods to detect spatial anomalies in health datasets. Starting from applying Bayesian spatial models to identify anomalies, followed by utilisation of non-Bayesian smoothing methods such as Gaussian processes, and kriging, and exploring the use of variogram. In the presentation, I will share the findings from the ongoing project followed by future directions in spatial anomaly detection with geospatial datasets in Australia.

4. Bayesian evolutionary optimization for parental selection in Salmon breeding

Jia Liu and Wagdy Mekkawy

CSIRO

Plant and animal breeders face challenges like productivity fluctuations due to climate change and global competition. To sustain long-term genetic gains, we propose optimal design strategies using multi-objective Bayesian optimization. This involves meticulously selecting optimal parental combinations each generation or crop cycle, guiding breeding endeavors.

We propose a novel statistical method using dynamic programming to maximize parental selection in breeding programs. Our approach focuses on maximizing Genomic Estimated Breeding Values (GEBV) and genetic diversity. This involves selecting optimal parental combinations from a population pool,

considering both additive and non-additive genomic random effects, and pedigree information to maintain breeding values.

We compare two acquisition functions—probability of improvement and expected improvement— as design criteria in the optimizing. GEBV are predicted using Genomic Best Linear Unbiased Prediction (GBLUP). Our results show promising long-term genetic gains compared to traditional methods. Additionally, the method is adaptable for plant breeding, incorporating genomic and environmental interactions.

Reference

Recent advances in Bayesian optimization. X Wang, Y Jin, S Schmitt, M Olhofer - ACM Computing Surveys (2023) Modelling and simulation of integrated systems in engineering: issues of methodology, quality, testing and application. Murray-Smith, David J. Elsevier. (2012)

Mating schemes for optimum contribution selection with constrained rates of inbreeding. Sonesson, Anna K and Meuwissen, Theo HE. Genetics Selection Evolution, 32(3), page 231-248 (2000)

5. Joint learning of Gaussian graphical models for paired data in high-dimensional and heterogenous population

Dung Ngoc Nguyen (CSIRO Agriculture and Food, Canberra, ACT, Australia), Alberto Roverato (Department of Statistical Sciences University of Padova, Italy), Zitong Li (CSIRO Agriculture and Food, Canberra, ACT, Australia)

The Gaussian graphical model (GGM) has become a popular tool for biological network analysis to study conditional dependence between biological variables and predict gene functions, phenotypes and patterns of molecular regulation. We consider the problem of learning a GMM for paired data where observations come from two dependent groups sharing the same variables. Comparing the distribution of a set of variables between two experimental conditions, or groups, is common to many real-world applications such as cancer genomic data or brain networks from fMRI data. To address the problem of high dimensionality and heterogeneity, we apply a mixture of GGMs for paired data and develop a penalized EM algorithm for these models where model estimation and selection can be performed simultaneously. The efficiency of our approach is then demonstrated on constructing gene networks using both real and simulated transcriptome data sets.

Reference

David Edwards. Introduction to Graphical Modelling. Springer, 2nd edition (2000).

Danaher, P., Wang, P., Witten, D.M.: The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76(2), 373–397 (2014).

Lartigue, T., Durrleman, S., Allasonniere, S.: Mixture of conditional Gaussian graphical models for unlabelled heterogeneous populations in the presence of co- factors. *SN Computer Science* 2(6), 466 (2021).

Roverato, A., Nguyen, D.N.: Exploration of the search space of Gaussian graphical models for paired data. *Journal of Machine Learning Research* 25(92), 1–41 (2024).

Højsgaard, S. and Lauritzen, S.L.: Graphical Gaussian models with edge and vertex symmetries. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70(5), 1005–1027 (2008).

Lauritzen, S.L.: *Graphical Models*. Oxford University Press (1996).

Ranciati, S., Roverato, A., Luati, A.: Fused graphical lasso for brain networks with symmetries. *Journal of the Royal Statistical Society Series C: Applied Statistics* 70(5), 1299–1322 (2021).

Ranciati, S., Roverato, A.: On the application of Gaussian graphical models to paired data problems. *arXiv preprint arXiv:2307.14160* (2023).

6. Case Studies in Advanced Analysis of Large Strip On-farm Experiments

Zhanglong Cao, Jordan Brown and Suman Rakshit

Curtin University

On-farm experiments (OFE) are gaining attention among farmers and agronomists for testing various research questions on real farms. The Analytics for the Australian Grains Industry (AAGI) has developed several techniques for analysing OFE data. Geographically Weighted Regression (GWR) and the multi-environment trial (MET) technique, which partitions paddocks into pseudo-environments (PEs), have proven effective. Additionally, we have explored the potential of the Generalised Additive Model (GAM) for handling temporal and spatial variability, given its flexibility in accommodating non-linear

variables. In this presentation, we will demonstrate case studies using these techniques to analyse OFE data and compare the outcomes of different approaches.

Abstracts for Talks: Thursday

Invited Talk Thurs 8:30 – 9:15

Spatial statistics and the protection of indigenous rock art

Adrian Baddeley

Curtin University

The peninsula and islands of Murujuga (Dampier Peninsula) in Western Australia contain over one million panels of indigenous rock art, in close proximity to an industrial area. Concern about potential damage to the rock art caused by industrial emissions has led to a research programme which is monitoring the condition of the rock art, studying the rock surfaces, and monitoring air quality across the peninsula. This talk will describe the essential role played by spatial statistical methods in designing the study and in identifying potential evidence of the effects of industrial emissions on the rock art.

Abstract session 3.1 Thurs 9:15 – 10:10am

Should we move on from the CIELAB colour space?

Rebecca O’Leary

Curtin University

The CIELAB colour space index converts colour spectrum into perceptual colour coordinates (L^* , a^* , b^*) defined by CIE international standards. L^* represents lightness, a^* the red- green balance, and b^* the blue-yellow balance. Unlike RGB, CIELAB is perceptually uniform, meaning a given numerical change in coordinates reflects the same perceptible change in colour. This index converts colour spectrum data for wavelengths 380 to 780 nm ($n = 950$) into 3 numbers (L^* , a^* , b^*). Traditionally, the Euclidean distance (ΔE) between L^* , a^* , b^* coordinates was used to detect perceptual colour differences, with $\Delta E > 2$ indicating a difference. The Murujuga Rock Art Monitoring Program examines the impact of industrial air emissions on rock art near Karratha in Western Australia. One component uses spectral colour measurements to monitor long-term trends, but extreme weather has caused instrument faults. These faults are visible in the spectrum plots but disappear when converted to CIELAB, appearing as valid measurements.

2. Spatial kernel smoothing with extreme outliers

Mohomed Abraj, Rebecca O’Leary, Stephanie Hogg and Adrian Baddeley
Curtin University

Spatial kernel smoothing is a technique to visualise spatial point patterns. However, extreme outliers in the data can distort the smoothed surface and mislead the interpretation. These outliers often arise in data collected rocks using portable X-ray fluorescence (pXRF). Geoscientists often classify rocks based on their proportions of elements. For example, the division of igneous rocks into ‘mafic’ and ‘felsic’ rocks is based on the proportion of silica and titanium measured by pXRF, silica typically ranges between 45,000 and 300,000 ppm, but there may be few extreme values around 700,000 ppm or higher. To address these extreme outliers, we present winsorization technique in spatial smoothing, which reduces the impact of spurious outliers while preserving extreme spatial information. This proposed technique is applied to silica and titanium data collected on rocks in Murujuga, Western Australia. The performance is assessed in terms of root mean square error (RMSE) before and after winsorization.

3. Calculating porosity for non-homogenous sections

Stephanie Hogg, Adrian Baddeley, Rebecca O’Leary, and Mohomed Abraj
Curtin University

Methods to establish porosity of homogeneous materials are well established and are used in geology to determine the size of aquifers and hydrocarbon reservoirs. However, these methods are not suited to determining porosity over a changing gradient. To study rock weathering and surface durability, an understanding of how porosity behaves over a vertical section of changing structure is needed. The technique of spatial smoothing can be used to determine a porosity curve that changes with distance from the surface. The technique is based on Scanning Electron Microscope (SEM) images. It uses morphological openings and connected components to define the rock sample surface, and then applies a Nadaraya-Watson smoothing with small bandwidths to estimate of porosity at different surface depths. The technique was applied

to multiple rock samples. Differing bandwidth and opening morphologies were compared.

Abstract session 3.2 Thurs 10:45 – 11:30am

1. Overcoming Challenges of Inconsistent Agronomic Treatments Across Experiments in a Weed Suppression Project

Kerry Bell and Michael Widderick

Department of Agriculture and Fisheries, Queensland

Field crop agronomy projects involve multiple experiments exploring varied agronomic treatment factors, often with inconsistent levels of these treatments investigated across environments. Addressing key questions is challenging because it requires analysing data across a selection of these experiments with consistent treatments and treatment levels. This complexity is exacerbated when needing to address key questions for differing audiences; for example, growers seeking insights into the benefits of changing agronomic practices, and researchers aiming to understand the underlying drivers.

Motivated by a weed suppression project, this presentation explores how experiments and treatments were selected based on the key questions from diverse audiences. The proposed approach maximises the available data, and thus information, by using a tailored definition of “environment” to explore environment x treatment interactions for consistent treatments and/or treatment levels. The approach is performed using linear mixed models, implemented via the Restricted Maximum Likelihood (REML) procedure in Genstat.

Reference

VSN International (2022). Genstat for Windows 22nd Edition. VSN International, Hemel Hempstead, UK. Web page: Genstat.co.uk

2. Dieback Dilemma – a statistical tale

David Reid, Stuart Buck and Nicholas Brazier

Department of Agriculture and Fisheries, Queensland

Following an outbreak of pasture dieback in Queensland in 2015, considerable research effort has been expended in seeking the cause of dieback and developing management strategies. Pasture dieback is a condition that causes death of high yielding tropical and sub-tropical grasses in eastern areas of Queensland and north-eastern New South Wales. Symptoms include leaf discolouration (reddening and/or yellowing) then unthrifty growth, which can be difficult to differentiate from water and temperature stress, nutrient deficiency, fungal infections, and herbicide damage. Symptoms progress until the whole plant dies, becoming grey and brittle. Affected areas spread from roughly circular patches to paddock scale areas of dead pasture. The vague definition of dieback, difficulty in correct diagnosis and the fact that it always results in plant death posed considerable challenges relating to trial design. Come with us on our journey of discovery as we attempted to conquer these obstacles.

3. A stochastic compartmental model for the Merry Island INRAE African Swine Fever outbreak challenge

Matthieu Vignes and Jun Hee Han

Massey University, New Zealand

An African Swine Fever outbreak was simulated among a wild boar (WB) and of pig farm (PF) population on a fictional territory (Picault et al. 2022). Three successive periods were considered after the detection of the initial outbreak, with alternative control measures. Participating teams had to predict epidemiological trajectories. We tailored a stochastic compartmental model to predict spatial densities of newly infected WB and estimate the risk for PF to become infected. It accounted for age, sex and movement behaviour in WB, as well as the biosecurity level of PF. We estimated parameters by evaluating the prediction adequacy to provided epidemiological trajectories at the start of the challenge scenarios. Our final predictions were satisfactory: our model performed similarly to other participating teams and provided a complementary point of view. Decision-makers can assess the gain in

implementing such-and-such measures to contain and eliminate the outbreak. We also discuss our modelling limitations.

Reference S. Picault, T. Vergne, M. Mancini, S. Bareille, P. Ezanno (2022). The African swine fever modelling challenge: Objectives, model description and synthetic data generation. *Epidemics*, 40 (100616).

Abstract session 3.3 Thurs 11:30 – 12:15pm

1. Do Mice Matter? The Impact of Mice on a New Zealand Ecosanctuary

Vanessa Cave

VSNi

Mammal-resistant fences have enabled the eradication of exotic mammals from ecosanctuaries in New Zealand. However, preventing the re-invasion of mice has proven problematical! Indeed, in many fenced ecosanctuaries mice remain present and they can reach high numbers.

Scientists at Manaaki Whenua – Landcare Research have been studying the impacts of mice on biodiversity at Sanctuary Mountain Maungatautari. Two independently fenced sites within the sanctuary were managed to achieve high mouse numbers at the first site and undetectable mouse numbers at the second site. After 2 years, management protocols were switched, with mice eradicated from the first site and their numbers allowed to increase at the second. Data on abundance of invertebrates, seedlings and fungi were collected throughout the duration of the study.

Linear mixed models with smoothing splines were used to model the temporal trends in abundance, leading to the conclusion that mice may be catastrophic in ecosanctuaries that focus on the recovery of invertebrates.

2. How many binary neutron stars are there in our Galaxy - a statistical approach to address observational complexities

Arash Bahramian and Alope Phatak

Curtin University

Neutron stars (NSs) are the densest form matter in the universe, formed from the death of massive stars. A significant fraction of NSs are found in binary systems with stars like our sun, where they “accrete” matter from this companion star. Characterizing the population of NSs is an important component of understanding evolution and death of stars in our Galaxy. From an observational perspective, astronomers only detect these NSs when an episode of accretion “outburst” occurs (lasting days) between long episodes of quiescence (months to decades). These periods of quiescence combined with the comparatively short baseline of observations (few decades), imposes heavy left censoring (absence of data when a quiescence period began), which impacts our ability to infer about the population of NSs in the Galaxy. In this talk, we present the unique complexities of astronomical observations in this context and possible statistical strategies to mitigate these issues.

3. Map the season: summarising climate variability

Anna Hepworth and Meredith Guthrie

Department of Primary Industries and Regional Development (DPIRD)

Annual variations in weather patterns drive many decisions in agriculture. Readily available seasonal summaries, such as those from the Bureau of Meteorology (BOM) are often not focused on the specific details that are of interest to growers. For example, ‘how wet was it today’ and ‘how hot was it’ are relatively easy to determine from weather reports and maps on the BOM website, but these point measures of climate are less important to growers than seasonal summaries, and comparisons with previous seasons. The Climate and Weather team at DPIRD provide maps and graphical summaries of weather relevant to the agricultural industry. These include break of season, available soil water, thermal time and frost duration’s. Maps are also produced quickly in response to notable events, such as heat stress, erosion strength winds, and storms. Maps from the 2024 season will be presented.

Presenter Index

Abraj, Mohomed	42	Lisle, Chris	27
Anderson, Angela	26	Liu, Jia	38
Angelin-Bonnet, Olivia	24		
		Marshall, Lucy	15
Baddeley, Adrian	41	McGarty, Aidan	13
Bahramian, Arash	46	McGree, James	18
Baird, David	19	Mengersen, Kerrie	35
Baten, Abdul	16	Misu, Farjana	17
Bell, Kerry	43		
Bradshaw, Stephen	36	Ngoc Dung, Nguyen	30
Brien, Chris	14		
		O'Leary, Rebecca	41
Cao, Zhanglong	40		
Cave, Vanessa	19, 45	Payne, Roger	19
Chen, Kefei	13	Piepho, Hans-Peter	10
Chin Fu, Shih	36		
Collins, Damian	18, 27	Reeves, Karyn	28
Cui, Ying	21	Reid, David	44
		Rutjens, Rik	31
Davies, Tilman	34		
Davis, Samuel	32	Shen, Rui	31
Ding, Pauline	17		
		Tanaka, Emi	11
Hayes, Anna	35		
Henry, Nathan	15	Vandal, Alain	25
Hepworth, Anna	46	Vignes, Matthieu	29, 45
Hepworth, Graham	23		
Hogg, Stephanie	43	Wang, Lu	21
Howden, Chris	12	Wright, Carole	11
Jahan, Farzana	37	Yang Tho, Zhi	22
		Yoon, Jin	23
Li, Zitong	39	Zhai, Siwei	34

List of Delegates

Alphabetical by last name. Note * attending Tuesday only

First name	Last name	First name	Last name
Angela	Anderson	Zitong	Li
Olivia	Angelin-Bonnet	Chris	Lisle
Robin	Arnason	Jia	Liu
Adrian	Baddeley	Lucy	Marshall
Arash	Bahramian	Aidan	McGarty
David	Baird	James	McGree
Abdul	Baten	Andrew	McLachlan
Kerry	Bell	Kerrie	Mengersen
Stephen	Bradshaw	Farjana	Misu*
Chris	Brien	Abraj	Mohomed
Ruth	Butler	Ngoc Dung	Nguyen
Kym	Butler	Rebecca	O'Leary
Zhanglong	Cao	Hans-Peter	Piepho
Vanessa	Cave	Lincoln	Rayner
Kefei	Chen	Karyn	Reeves
Damian	Collins	David	Reid
Ying	Cui	Renu	Saradadevi
Mario	D'Antuono*	Rik	Rutjens
Tilman	Davies	Rui	Shen
Sam	Davis	Connor	Stewart-Green
Pauline	Ding	Emi	Tanaka
Shih Ching	Fu	Zhi	Yang Tho
Anna	Hayes	Andrew	van Burgel
Nathan	Henry*	Alain	Vandal
Graham	Hepworth	Matthieu	Vignes
Anna	Hepworth	Lu	Wang
Stephanie	Hogg	Carole	Wright
Chris	Howden	Jin	Yoon
Farzana	Jahan	Siwei	Zhai
Nicholas	Lambert		

Australasian Applied Statistics Conference 2024

Rottneest Island, Perth. 2 – 5 September 2024



DAY	Mon 2 Sep	Tues 3 Sep	Wed 4 Sep	Thurs 5 Sep
	9am - 1pm: Workshops @ Esplanade Hotel Fremantle	8.45-9am: Opening 9-9.45am: Prof Dr. Hans-Peter Piepho Chair: David Baird	Chair: Karyn Reeves 9.30-9.45: Committee announcements	8.30-9.15: Prof Adrian Baddeley Chair: Chris Brien 9.15-10.10am: Abstract Session 3.1 1. Bec O'Leary 3. Stephanie Hogg 2. Mohamed Abraj Chair: Karyn Reeves 10.10-10.15am: Committee announcements
Morning Tea		9.45-10.15am	9.45-10.15am 10.15-11am: Prof Kerrie Mengersen Chair: Bec O'Leary	10.15-10.45am 10.45-11.30am: Abstract Session 3.2 1. Kerry Bell 3. Matthieu Vignes 2. David Reid Chair: Emi Tanaka
	9am - 1pm: Workshops @ Esplanade Hotel Fremantle	10.15am-12pm: Abstract Session 1.1 1. Emi Tanaka 4. Kefei Chen 2. Carole Wright 5. Aidan McGarty 3. Chris Howden 6. Chris Brien Chair: Andrew McLachlan	11-12.15pm: Abstract Session 2.1 1. Shih Ching Fu 4. Jia Liu 2. Stephen Bradshaw 5. Zitong Li 3. Farzana Jahan 6. Zhanglong Cao Chair: Graham Hepworth	11.30-12.15pm: Abstract Session 3.3 1. Vanessa Cave 3. Anna Hepworth 2. Arash Bahramian Chair: Emi Tanaka 12.15-12.30: Conference close
Lunch	1-2pm	12-1pm	12.15-1.15pm	12.30-1.15pm

	Travel to Ferry Terminal @ Fremantle B Shed	<p>1-1.45pm: Prof Lucy Marshall</p> <p>Chair: Vanessa Cave</p> <p>1.45-3pm: Abstract session 1.2</p> <p>1. Nathan Henry 4. Pauline Ding 2. Abdul Baten 5. Damian Collins 3. Farjana Misu 6. James McGree</p> <p>Chair: Shih Ching Fu</p>	<p>1.15-6.30pm: Free time Social activities</p>	2pm: Ferry departure
Afternoon Tea		<p>3-3.30pm</p> <p>3.30-4.15pm: Dr David Baird</p> <p>Chair: Ruth Butler</p> <p>4.15-5.30pm: Abstract session 1.3</p> <p>1. Lu Wang 4. Graham Hepworth 2. Ying Cui 5. Jin Yoon 3. Zhi Yang 6. Olivia Tho Angelin-Bonnet</p> <p>Chair: Kerry Bell</p> <p>5.30-6pm AGUAI AGM</p>	<p>1.15-6.30pm: Free time Social activities</p>	
Evening	6.30-8pm: Welcome function @ Hotel Rottnest	<p>7-9pm: Poster Session with Canapes</p>	<p>6.30-9pm: Conference Dinner @ Isola</p>	