Australasian Applied Statistics Conference

2026

30th November - 4th December,

Mantra Mooloolaba Beach,

Sunshine Coast,

Queensland



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), Inverloch (2022 Vic) and Rottnest Island (2024 WA). 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 3 December 2026, where the main item is to select a new AGUAi committee.

Current AGUAi committee organising the AASC 2026 conference:
Officeholders:

Bethany Rognoni[1] (Chairperson, QLD)

Colleen Hunt[2] (Treasurer, QLD)

Michael Mumford[3] (Secretary, QLD)

Helena Oakey[2] (Vice-chairperson, SA)

Ordinary members:
Angela Anderson[1] (Qld), David Baird[2] (NZ), Kerry Bell[1] (QLD), Ruth Butler[2] (NZ), Vanessa Cave[2] (NZ), Pauline Ding[2] (NSW), Clayton Forknall[1] (QLD), Ky Mathews[1] (QLD), Rebecca O'Leary[4] (WA), David Reid[1] (QLD), Andrew Van Burgel (WA), Carole Wright (QLD), and Hwan-Jin Yoon[2] (VIC).

Where:

1. Local organising committee
2. Scientific program committee
3. Conference website

4. 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.

Posters: The poster session will start with one-minute talks for each poster with one presentation slide.

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

Three concurrent workshops are planned for Monday 30th November 2026.

Workshop #1 - ASReml-R Workshop 09:00-16:30

Workshop #2 - Genstat Masterclass 09:00-15:30

Workshop #3 - BLUP and Bayesian methods for genomic prediction 09:00-12:30

Please see the conference webpage for more details: `https://aasc2026.netlify.app/workshop`

Social program

Four options are available for the afternoon of Wednesday 2 December 2026. Please see the conference webpage for more details: `https://aasc2026.netlify.app/social`

Sponsorship

AASC 2026 conference is proudly sponsored by VSNi and the Queensland Department of Primary Industries.

Invited Speakers

Prof. Ben Hayes: Day Time
Presentation: Bayesian or machine learning approaches: what works best for genomic prediction?



Professor Hayes has extensive research experience in genetic improvement of livestock, crop, pasture and aquaculture species, with a focus on integration of genomic information into breeding programs, including leading many large scale projects which have successfully implemented genomic technologies in livestock and cropping industries. Author of more than 300 journal papers, including in Nature Genetics, Nature Reviews Genetics, and Science, contributing to statistical methodology for genomic, microbiome and metagenomic profile predictions, quantitative genetics including knowledge of genetic mechanisms underlying complex traits.

Prof. Margarita Moreno Betancur:
Presentation: Causal inference in health research



Professor Margarita Moreno-Betancur is co-Director of the Clinical Epidemiology and Biostatistics Unit at the Melbourne Children's campus. She established an internationally recognised hub of innovation and expertise in statistical methods for observational studies, supported by competitive fellowship and grant funding. Her team conducts methodological research in causal inference and missing data while contributing their expertise to health research studies.

She is Chair of the Victorian Centre for Biostatistics leadership team, Scientific Lead of the LifeCourse platform of cohort studies, and Editor of the journal Epidemiology. She was awarded the Moran Medal by the Australian Academy of Science in 2025.

Associate Prof. Peter Dunn:
Presentation: Modelling continuous data with exact zeros



Peter is an Associate Professor in Biostatistics at the University of the Sunshine Coast, and routinely teaches introductory statistics to over 500 students over two semesters each year.

Peter has over 110 peer-reviewed research publications, has prepared numerous R packages available on CRAN (including the tweedie package, with almost 1 million downloads from CRAN) and co-developed the Dunn-Smyth (or quantile) residuals used in statistical modelling. Peter has been invited to give presentations on the R statistical environment at conferences and workshops across Australia.

Peter won an Australian Office of Learning and Teaching Citation in 2012, a student prize at the 16th International Workshop on Statistical Modelling in Denmark, and the EJ Pitman Prize for student talks at the Australian Statistics Conference in 2002. He has also published two textbooks: Generalized Linear Models With Examples in R (Springer) and Scientific Research and Methodology (CRS Press; also available as a free interactive online textbook).

Dr. Petra Kuhnert:
Presentation: Irresponsible use of AI is a Crime Scene: sifting through the slop to deliver responsible AI solutions



Dr. Petra Kuhnert is a Senior Principal Research Scientist at CSIRO and an expert at the intersection of statistics and machine learning. Her work develops methodology for risk-based decision-making in environmental and agricultural applications where she develops methods that are fit for purpose, relevant to the question at hand, and usable by decision-makers.

Her work, particularly supporting Great Barrier Reef decision-making, has been recognised through awards including 1st Runner Up (APAC Women in AI Innovator of the Year, 2023) and the APAC Women in AI Award for Environment and Biodiversity (2023).

A key contribution to AI-driven decision-making is her award-winning software Vizumap, an R package that helps users understand and communicate uncertainty on maps.

More recently, her focus has included machine-learning emulators to speed up slow-running physical system models, Bayesian methods for decision support, and approaches that use geospatial data (e.g., remote sensing) to improve prediction of terrestrial environmental and agricultural processes. Her previous roles include Associate Science Director for University Engagement and Group Leader of the Statistical Machine Learning Group in CSIRO's Data61.

She has co-authored 100+ journal articles, with 6,000+ citations and a Google Scholar h-index of 30+. Her collaborations span leading government departments, universities and industry.

Dr. Ruth Butler:
Presentation: Consulting in the real world: reflections, comments and data



Dr. Ruth Butler has worked as a biometrician/statistical consultant for more than 35 years, initially in the UK, then from the mid-1990s in New Zealand. She has primarily worked with bio-protection scientists (plant pathology, entomology), but also has significant experience working with other non-medical biological scientists including in soils/agronomy, food research and plant breeding. Ruth has been a Genstat user throughout her career, contributing around 10 Genstat procedures, and has been a beta tester of Genstat for 30 years. Ruth has also been a CycDesigN user since the very first version was released in 1997. Her interests are in good data management practices, well-designed experiments, and in improving communication between statisticians and scientists.

Dr. Salvador Gezan:
Presentation: Fitting Large Genomic Models: Computational Strategies and Approximations within the Linear Mixed Model Framework

Salvador Gezan is a statistician and quantitative geneticist with over 20 years of experience, currently serving as a Statistical Genetics Consultant at VSN International. He specialises in mixed models and their application to biological problems, supporting agricultural, forestry, and aquaculture breeding programmes through the analysis of phenotypic and genotypic data, the development of breeding strategies, and the implementation of genomic selection in operational breeding programmes. He began his career at Rothamsted Research, was an Associate Professor at the University of Florida, and has delivered ASReml workshops globally. Dr. Gezan has over 160 peer-reviewed publications and is a co-author of Statistical Methods in Biology: Design and Analysis of Experiments and Regression.

Program

AASC 2026: Workshops

Monday 30th November: 9am-1pm

Workshop #1 ASReml-R Workshop

Dr. Salvador Gezan (VSNi)

This one-day workshop will focus on modelling complex variance-covariance structures in biological experiments using ASReml-R. ASReml-R is a statistical library in R that fits linear mixed models using REML methodology. This workshop will cover key variance structures for random effects and residuals, including spatial, repeated measures, multivariate, and multi-environment models, with applications drawn from agricultural and medical research contexts.

The workshop is oriented toward practical analysis of real datasets. Participants are expected to have prior familiarity with linear mixed model methodology and R, and should bring their own laptop with R and ASReml-R pre-installed. A training ASReml-R license will be provided.

Morning (4 sessions)

1. **Introduction & Linear Mixed Model Refresher**
   Brief ASReml-R orientation, mixed model theory, fixed vs random effects, and REML estimation.
2. **Variance Structures in ASReml-R**
   The core topic considering diagonal, correlation, unstructured, and factor analytic structures.
3. **Spatial Variance Structures**
   Modelling residual spatial correlation in field trials using AR1 $\times$ AR1 structures.
4. **Repeated Measures & Random Regression**
   Modelling correlations over time, using simple or complex approaches.

Afternoon (3 sessions)

1. **Bivariate & Multivariate Models**
   Random effects and residual correlation structures between responses.

2. **Multi-Environment Trials & Factor Analytic Models**
   G × E interaction, factor analytic variance structures, and extensions to other contexts.

3. **Expanding to Complex Structures**
   Modelling multi-stratum designs with multiple layers of correlation between random effects and residuals.

Workshop #2 Genstat Masterclass

Dr. Vanessa Cave & Dr. David Baird (VSNi)

This full-day masterclass will highlight some new additions and hidden features in Genstat throughout 4 sessions.

1. New Features in Genstat 24
   - New graphics menus: irregular shade and tessellation plots, forest and pirate plots, ROC curves, adding shaded confidence regions to graphs
   - New statistical menus: bimodality and outlier tests, tabulating text, fitting smooth surfaces to irregular points, repeated measures with censoring
   - Menu updates: General ANOVA - transformations and multiple Y variates
   - New spreadsheet features
2. Advanced Data Handling in Genstat
   - Importing and exporting data in programs
   - Working with pointers and substitution
   - Useful functions when programming
   - Merging, appending, subsetting, stacking, and unstacking data in programs
   - Making the most of the spreadsheet facilities
3. Building Your Own Graphs by Programming
   - Better keys with DKEY
   - Labelling points and controlling label positioning

- Non-standard plotting symbols and icons
- Adding text, arrows, error bars, lines and shaded regions to graphs
- Overlaying graphs

4. Extra Features When Fitting Mixed Models
5. Fitting p-splines and other no-standard models
6. Plotting spline fits
7. Exploring fixed and random effects
8. Outlier detection
9. Power analysis for a fixed term or contrast

Workshop #3 BLUP and Bayesian methods for genomic prediction

Professor Ben Hayes (University of Queensland)

Abstracts for Talks: Tuesday

Keynote #1 Tues 9:00 – 9:45

Irresponsible use of AI is a Crime Scene: sifting through the slop to deliver responsible AI solutions

Petra Kuhnert

CSIRO

Artificial Intelligence (AI) is moving at a rapid pace, shaping decisions across sectors and communities. As a result, the number of paper submissions to leading AI venues has surged. For example, in 2025, a little over 71,000 papers were submitted to the main research track of five of the top machine learning conferences, representing a 32% increase since 24 and a 10-fold increase from ten years ago. This rise in volume is becoming increasingly difficult to manage and consume in the academic community. Alongside genuine advances, there is also work that is incremental, overly tuned to curated benchmarks, weakly evidenced or even AI-generated without adequate verification.

If we want AI research and deployed systems that are trusted, robust and transparent, we need a forensic lens on how AI and its underlying analytics are developed and evaluated. For me, Statistics provides that lens, yet it can be sidelined in the rush to publish AI and deploy.

In this talk, I use the metaphor of a crime scene to describe AI crime: errors in predictions, uncertainty and inference that lead to poor decisions and in high-stakes settings can translate into real-world harm. I outline what Responsible, Application-Driven AI (RAD-AI) looks like in practice, and the statistical checks and incentives that help prevent failure modes and AI crime before it reaches the field. I will illustrate these ideas with work we are developing with stakeholders to produce a trusted AI crop growth model.

We would not cut corners at a crime scene, so why would we allow crime to happen with AI!

Abstract contributed session #1 Tues 10:00 – 10:30

1. Genstat Markdown: Reproducible Research with Genstat

James Curran

University of Auckland

Reproducible research ensures that scientific findings can be independently verified by others using the same data and methods. Markdown plays a key role in supporting reproducibility by allowing researchers to combine code, results, and narrative text in a clear, lightweight format. Tools like R Markdown or Jupyter Notebooks use Markdown to embed executable code alongside explanations and outputs, making it easy to document workflows, share analyses, and regenerate results consistently across different environments.

Genstat is a statistical software package designed for data analysis, particularly in biometrical research and applications. Many users know it through its user-friendly interface. However, it also It offers a powerful scripting language.

In this talk I will describe a collaborative project between students and staff at the University of Auckland and VSNi Ltd, that allows Genstat to be used in conjunction with R Markdown to provide flexible reproducible research capability to the Genstat user community.

---

2. Leveraging the core: R Shiny apps which talk to the Genstat Engine

Vanessa Cave

VSNi

Genstat is a statistical software package designed for data analysis, particularly in bioscience research and its applications. While many users are familiar with its user-friendly interface, it also provides a powerful scripting language for advanced statistical computing.

Abstract contributed session #2 Tues 11:00 – 12:30

1. Pizzas, Biometry and the 5Rs

Mario D'Antuono

Retired

In 2008, at the Australasian Genstat Users Association (AGUA) meeting, then known as the Genstat Conference in Marysville, Victoria, I presented a controversial paper on my forthcoming e-book entitled "Biometry: The 5Rs".

The paper was controversial because it presented an argument that in 5 simple lines of code using the R Statistical System (denoted by R) usually in the form of a function call, I could address a task such as the generation of a near-optimal design, or a statistical analysis. These 5 lines can be recursive and hence can call other functions within a function (a nested approach). The approach was stimulated by the work of Karl Broman at the time and his 3Rs (Reproducible Research with R).

In 2019, I retired from the Department of Agriculture, Western Australia. It has taken 7 years for me to find the time and the energy to complete the e-book. The free R software to handle the above has developed to be able to treat many of the common problems in agricultural research.

We illustrate two published examples from the literature, which are case-studies in the e-book; the first involves the construction of a near-optimal p-rep design, and the second, a spatial linear mixed model analysis. We compare the results from the University of Wollongong R packages dwreml-R and odw-R (soon to be released on GitHub) with other software such as CycDesign, speed-R, and DiGGer-R for near-optimal design generation, and asreml-R and Echidna for linear mixed modelling.

2. A modified PQL method for genetic analysis of binary variables

Arthur Gilmour

VSNi

In plant and animal breeding, there is the need to estimate breeding values for binary traits such as lamb survival and presence/absence of disease. For such traits, the average incidence is often outside [0.2, 0.8] and it varies with known fixed effects and covariates. Animal breeders commonly assume the binary trait is based on an underlying normal liability exceeding a threshold. PQL (pseudo quasi-likelihood), an extension of GLM to include random effects in the linear model, is commonly used to analyse such data and predict breeding values, but is known to underestimate the heritability when fitting an animal model. It iterates using a working variable

$$w = \eta + \frac{y - \mu}{\delta},$$

where η is the prediction on the liability scale, $\eta = X\beta + Zu$, y is binary outcome, $\mu = f(\eta)$ is predicted mean on the probability scale, δ is the derivative of the probit link function $f(\eta)$, X is the design matrix for fixed effects (β) and Z is the design matrix for random effects (u). The linear model, $w = X + Zu + e$ is fitted with weights $\delta^2/\text{var}(y)$, $\text{var}(y)$ is $\mu(1-\mu)$. The mPQL model, in the simplest case, calculates $\eta = X$, leading to changes in and δ, and necessitates estimating a dispersion factor when fitting the linear model (which is unchanged). My recommendation when using mPQL is first to fit the model on the probability scale, note the mean (p), calculate the heritability (h_o^2), the dispersion factor ($(1 - h_o^2)$ and predict an approximate heritability for the liability scale using $h_l^2 = h_o^2 p(1-p)/z^2$. Then, fit the mPQL model using the dispersion factor. mPQL is available in Echidna (EchidnaMMS.org)

3. ggincerta: An R package for Uncertainty Visualisation with a Layered Grammar of Graphics

Maggie Ma

Australian National University

A key challenge in spatial visualisation is the simultaneous communication of estimated values and their associated uncertainties, as visual channels are occupied by spatial structure, leaving limited capacity to encode additional information. We developed *ggincerta*, an R package that extends the layered grammar of graphics implemented in *ggplot2*, to facilitate the visualisation of uncertainty in spatial data. Through its native integration with *ggplot2*, *ggincerta* enables users to leverage the full flexibility of the grammar of graphics to create complex and customised visualisations through simpler and more consistent workflows in spatial and more general data settings. In addition to reproducing existing uncertainty visualisation techniques, the package introduces new approaches, including the dual map, a visualisation designed to support selective attention between estimated values and their associated uncertainties. We demonstrate the capabilities of *ggincerta* through several case studies, illustrating how the package can be used to create informative and visually effective maps that communicate both estimates and uncertainty, while also highlighting relevant technical details and visual features. By integrating uncertainty visualisation directly into the *ggplot2* ecosystem and expanding the range of available uncertainty displays, *ggincerta* provides a flexible and powerful framework for researchers and practitioners working with spatial and other information-rich data.

---

4. torch-openreml: a PyTorch-based framework for flexible linear mixed-effects modelling

Patrick Li

Australian National University

We introduce torch-openreml, a PyTorch-based Python library under active early development for fitting linear mixed-effects models using the Average Information Restricted Maximum Likelihood (AI-REML) algorithm. The library represents covariance models as compositional computation graphs, where complex covariance structures are constructed by combining modular components through algebraic operations.

Within this framework, model parameters can be fixed, estimated, or transformed, with analytical gradients provided where possible and PyTorch automatic differentiation used for general cases. The AI-REML solver exploits common covariance structures to improve computational efficiency and benefits from PyTorch support for CPUs, GPUs, and other accelerators.

Although currently in an early alpha stage, torch-openreml supports a wide range of linear mixed-effects modelling applications, such as multi-environment trial analysis, spatial modelling, and genomic prediction. It enables rapid prototyping of novel covariance models without requiring manual gradient derivation, providing an extensible platform for modern mixed-effects modelling.

---

5. One Workflow, Many Questions: A Unified Workflow for Statistical Modelling in the Age of AI

Kathrin Schemann

Sydney Informatics Hub (University of Sydney)

Applied statisticians increasingly work across diverse modelling paradigms, each with distinct assumptions, methods, and reporting requirements. At the same time, the rapid emergence of AI tools creates opportunities to improve efficiency while raising important questions about transparency, reproducibility, and statistical rigor.

This presentation introduces a practical, software-agnostic workflow developed at the Statistical Consulting Unit of the Sydney Informatics Hub, The University of Sydney. The workflow provides a unified framework for addressing descriptive, predictive, and causal modelling objectives while maintaining focus on fundamental statistical principles and research context. The framework follows a shared structure that begins with clearly defining the research question, identifying variables of interest, documenting assumptions, and developing a transparent analysis plan. Key components include study design, bias assessment, variable selection, interaction assessment, model evaluation, interpretation, and communication of results. The presentation also examines how AI can be integrated throughout the analytical process, supporting tasks such as planning, coding, documentation, model implementation, diagnostic review, and reporting. Within the workflow, AI is positioned as a tool to augment—not replace—statistical expertise, subject-matter knowledge, and professional judgment.

By combining established statistical principles with structured opportunities for AI augmentation, the workflow provides a consistent approach to applied statistical practice across modelling objectives, supporting more transparent, reproducible, efficient, and impactful research.

---

6. Nested-factorial treatment models: types, their uses and examples

Chris Brien

Adelaide University

Most commonly factorial experiments involve several crossed treatment factors and the treatment model includes terms for all possible combinations of the treatment factors. However, treatment models that employ nested treatment factors are beneficial in a number of situations and could be used more often. Applicable situations include those in which (i) treatment factors are intrinsically nested, (ii) treatments involve one or more factors plus a control and (iii) treatments have been systematically allocated and/or are pseudoreplicated. Four basic types of nested factors that start with a single nesting factor are (i) one nested factor, (ii) multiple, crossed, nested factors, (iii) multiple, independent, nested factors, and (iv) a hierarchy of multiple, nested factors. Examples involving the different situations and types of nesting will be presented. Constructing the factors using the *R* package *dae* (Brien, 2026, `https://cran.r-project.org/package=dae`) will be described and the properties of the models discussed for the examples via their anatomies produced using *dae*.

Keynote #2 Tues 13:30 - 14:15

Fitting Large Genomic Models: Computational Strategies and Approximations within the Linear Mixed Model Framework

Salvador Green

VSNi International

As genomic datasets grow larger, substantial statistical and computational challenges arise when analysing single-nucleotide polymorphism (SNP) data. Such data are widely used in Genome-wide Association Studies (GWAS), Polygenic Risk Score (PRS) analyses, and genome-wide prediction models. Depending on the field and context, it is common to encounter datasets

containing millions of SNPs and hundreds of thousands of genotyped individuals. Fitting even the simplest linear mixed model at this scale is computationally demanding, requiring the handling of large genomic matrices and the estimation of a large number of marker effects. Statistical and mathematical techniques such as Cholesky decomposition, eigen-transformation, Singular Value Decomposition, and low-rank matrix approximations provide practical solutions for fitting these models and obtaining the required effects and predictions. This talk will present key mathematical details, illustrations, and practical guidelines for addressing large-scale genomic analyses in plant and animal breeding.

Abstract contributed session #3 Tues 14:30 – 15:00

1. Predicting Cotton Fibre Performance using GWAS Summary Statistics

Qian Feng

CSIRO Agriculture and Food Unit

Genomic prediction (GP) refers to the prediction of individual's phenotypes using genome-wide information. In humans, GP has contributed substantially to the development of precision medicine, while in animal and plant breeding it has accelerated genetic gain by enabling earlier and more accurate selection. Traditional GP methods rely on individual-level genotype and phenotype data, which are often subject to privacy, ownership and data sharing restrictions. Recently, summary-statistics-based methods have emerged for GP by leveraging genome-wide association study (GWAS) summary statistics and linkage disequilibrium information from a reference population. These approaches avoid the need of individual-level data and have been widely applied in human genetics. However, their application in plant breeding, including cotton, remains very limited. Here, we investigate the feasibility of summary-statistics-based genomic prediction using a large-scale cotton dataset collected over 40 year-site combinations in central-eastern Australia. Both genotype and phenotype data are available for this dataset, enabling the computation of GWAS summary statistics and a direct comparison between conventional Genomic Best Linear Unbiased Prediction (GBLUP) and a summary-statistics-based BLUP approach. Our empirical results show that the two methods achieve comparable prediction accuracy across multiple fibre yield and quality traits, consistent with theoretical expectations. We further evaluate how the choice of reference population and GWAS summary statistics information influences prediction performance. These results demonstrate the potential of summary-statistics-based GP in

cotton and suggest that this approach could enable the use of larger, more diverse publicly available datasets to support the development of better cotton varieties.

2. Protein Advantage for Mungbean in Australia

Jodine O'Connor

The growing emphasis on plant-based diets and the increasing demand for protein have brought pulse grains into sharp focus. As the second-largest exporter of pulse grains globally, Australia is uniquely positioned to meet this rising demand. To maintain market share and leverage the growing interest in plant-based protein, the Grains Research and Development Corporation (GRDC) and Agriculture Victoria have initiated a national project aimed at exploring innovative strategies to support and expand this vital industry. The project focuses on six pulse crops: mungbean, chickpea, lupin, lentil, faba bean and field pea. To date, historical varieties for each crop have been evaluated across Australia's major pulse growing regions. These trials span 33 sites across five states, conducted in 2024 and 2025. With demand to meeting new markets, a challenge arises for breeding varieties that achieve both high yield and desirable protein levels. Therefore, a statistical methodology is proposed with a protein advantage metric. Using available mungbean data, a multi-environment, across-year bivariate analysis was performed for three sites in QLD and NSW. This analysis examined yield and protein content within a bivariate linear mixed model framework, extended to incorporate multi-environment data. The framework also included spatial modelling for each trial and factor analytic modelling to allow for heterogeneous genetic covariance across trials. The bivariate analysis revealed a strong negative genetic correlation between yield and protein content across all environments. The protein advantage metric was obtained by using genotype deviations from the regression line between protein and yield. When plotted, varieties with a positive protein advantage or a negative advantage are clearly identified. This information provides valuable insights for growers and pulse breeders to select varieties that align with targeted market traits (e.g. protein) and production goals in Australia.

Abstract contributed session #4 Tues 15:30 – 16:30

1. Identifying cellular relationships in finely annotated spatial transcriptomics data

Ellis Patrick

University of Sydney and Sydney Precision Data Science Centre

Recent cell-resolution spatial transcriptomics technologies enable whole-transcriptome profiling of millions of individual cells within intact tissues, creating unprecedented opportunities to study how cellular organisation varies across development, health and disease. However, the same resolution that makes these data biologically powerful also makes them analytically fragile. Although these assays target approximately 20, 000 genes, each cell typically contains only a sparse sample of detected transcripts, with measurements also contaminated by transcript diffusion, imperfect segmentation and cell overlap. This makes expression estimates uncertain and can destabilise fine-grained cell labels. Fine-grained labels can also fragment the data into rare populations for which spatial associations are difficult to estimate precisely. Inference on spatial organisation is further complicated by correlation at multiple scales, arising from cohort design, tissue sampling and the spatial arrangement of cells within structured anatomical regions. Together, measurement error, annotation uncertainty and dependence can produce unstable estimates and reduced power for detecting meaningful spatial relationships.

In this talk, I will present approaches developed by my team to address these challenges in high-dimensional spatial omics data. Using hierarchical modelling, cluster-robust variance estimation, bootstrapping and weighting strategies, we developed more robust and powerful tests for spatial association under annotation uncertainty and multiscale dependence. These methods support inference across cell-cell association testing, proximity-associated gene programs, rare-cell localisation and multicellular tissue architecture, providing a framework for extracting reliable biological signal from sparse, contaminated and correlated spatial omics data.

2. A Location-Scale Linked Autoregressive Moving Average Model for Wind Direction Time Series Data

Piumi Perera

Australian National University

Modelling long-term wind direction time series data poses several challenges, including the circular nature of the observations, temporal processes operating at multiple scales, and the need to jointly capture both mean direction and variability in the presence of serial dependence. Motivated by hourly wind direction observations collected between 2013 and 2016 from a Bureau of Meteorology station in Broome, Western Australia, we propose a location-scale linked autoregressive moving average (LSLARMA) model that prioritises interpretability and data generation. The proposed framework adopts a two-stage approach, comprising a location-scale regression applied to transformed wind direction, followed by a (seasonal) ARMA model fitted to the resulting quantile residuals to capture remaining serial dependence. Application of the LSLARMA model reveals clear annual, monthly, and diurnal patterns in both the mean and dispersion of wind direction. In addition, the model can generate realistic synthetic time series that reproduce key features of the observed wind direction data, while compared with alternative single stage approaches the proposed method delivers improved performance and greater interpretability of the underlying temporal processes.

---

2. Cross trial analysis of rotation by environment interaction in the Western Australian Farming Systems project

Ben Nestor

Department of Primary Industries and Regional Development, WA

The Western Australian Farming Systems (WAFS) project involves ongoing trials comparing four-year standard and diverse cropping rotations under two nitrogen supply levels (low and high). Trials are located across three WA cropping environments: Northampton (northern, medium rainfall), Merredin (eastern, low rainfall), and Lake Grace (southern, medium rainfall). Rotations include combinations of cereals, canola, grain legumes, manured legumes, multi-species, and fallow with different phasings of rotations. A multi-environment trial (MET) analysis was performed to investigate rotation × environment

interactions, with environment defined as the combination of location and nitrogen treatment. The relative performance of each rotation was compared across environments using gross profit margins from three years of rotations as the response variable in a factor analytic model. The best linear unbiased predictions were compared across rotations to identify rotation types that were most profitable in different environments and rotation types that gave more consistent profits across environments. This analysis demonstrates a novel application of MET analysis on cropping rotation trials, providing insights into the effects of rotation and nitrogen treatment across different WA agricultural environments.

---

4. A quantitative framework for dissecting genotype by environment interactions driving crop yield variability

Stephanie Brunner

University of Queensland

Climate change magnifies unpredictable environmental patterns, threatening global crop yields. A wide range of approaches exist to assess varieties across environments through mixed linear models with latent or environmental variables. However, the underlying causes of these genotype by environment interactions (GEIs) remain unexplored. Understanding how genetic and environmental factors interact to drive yield variation is critical for breeding resilient crops. This research develops a framework to investigate the genetic and environmental drivers of GEI to help inform crop breeding decisions. Real world breeding program data was used to develop this framework, ensuring practical applicability. A panel of 965 lines of barley germplasm was evaluated for yield in 12 trials. To dissect the genetic component of GEI a genomic mixed linear model was fitted in which random genetic effects were structured by a genomic relationship matrix leveraging shared genetic information to predict yield across environments. To interpret these predictions biologically, haplotype effects were estimated for marker effects allowing for yield to be explored at the haplotype level. Haplotypes with large GEI for yield were found to have different effects across different environments. The environmental component of GEI was explored using a synthetic environmental covariate approach, which represents major environmental gradients as linear combinations of measured variables (e.g. rainfall). This method attributes GEI variance to interpretable environmental covariates at key growth stages, providing a measure of stability across time. This temporal deconstruction of GEI variance allowed for the analysis of

the impact of individual environmental covariates on haplotype effects for yield bridging the two components of this research. This framework helps to disentangle GEIs and identify environment-responsive haplotypes enabling breeders to target yield stability despite climatic uncertainty.

Poster

A Statistical Framework for Data Streams

Bao Khue Tran

UNSW Sydney

Data streams are potentially unbounded sequences of observations that must be processed incrementally under strict memory and time constraints. Despite their ubiquity, streaming data currently lack a coherent statistical foundation as most existing work is computer science-focused. This raises fundamental questions: How should a data stream be represented as a statistical object? Which classical properties carry over to the streaming setting, and which require reformulation? How can competing methods for the same inferential target be compared?

We propose a rigorous and comprehensive framework for evaluating streaming methods. Using a running example examined at increasing levels of complexity, we show how each layer of our framework addresses gaps in existing approaches, and what is required to perform rigorous statistical inference on data streams.

Quantifying Spatial Variation in Injury-Related Bloodstream Infections across Queensland

Binuri Perera

Queensland University of Technology

Bloodstream infection (BSI) is a serious complication among injured patients. Although injury- related BSI are known to vary across broader remoteness categories in Queensland, the local variations within small areas are yet unknown. This study aimed to examine the disparities in residential location among people with injury-related BSIs in Queensland using population-based linked data.

The linked used data for this study consists of patient admission data from Queensland Hospital Admitted Patient Data Collection (QHAPDC), pathology data and death data from the Death Registry for Queenslanders from 2010 to 2019. Aggregated data on acute injury hospital episodes and estimated residential population were used to calculate incidence rates. Leroux models were fitted to examine the spatial variations using standardised incidence ratios. Markov Chain Monte Carlo simulations were used to fit the models.

Across the entire 10-year period, 1,895 people experienced an injury-related BSI. There was marked spatial variation across the state, with particularly high rates observed in northern, far north and western Queensland. Consistent patterns were observed after accounting for the injury hospitalisations.

Bayesian spatial models enable the identification of high-risk areas with higher incidence based on patients' residential location. The results are beneficial for medical staff and policymakers in identifying disparities based on location and implementing targeted interventions and service improvements.

Efficient feature detection in spiky data using penalised effects in a linear mixed model

Clayton Forknall

Adelaide University

Analysing spiky data, where spiky refers to a response variable that demonstrates severe variation in magnitude in response to one or more explanatory variables, requires care, with challenges arising both in terms of feature detection and model assumptions. We propose a linear mixed model (LMM) based approach for the analysis of spiky data arising from designed experiments. The approach utilises penalties, applied both to the random effects of interest, along with the residual errors, to i) efficiently detect and estimate important features in the response and ii) allow for more flexible t-distributed errors. The approach is motivated by, and demonstrated using, a dataset arising from a mass spectrometry based proteomics experiment, where our aim is to efficiently detect significant peptides and/or proteins that are present at different stages of the barley malting process. Our approach is easily implementable in most LMM software packages, providing a readily applicable solution to a challenging and complex problem.

Improved assessment of rangeland management

Damian Collins

NSW DPIRD

Rangelands suitable for pastoral land use cover 50% of the earth's terrestrial surface. Managing rangelands to reduce or avoid degradation is crucial for food production and ecological sustainability. In recent decades, the use of remote sensing has automated the monitoring of land condition, in particular ground cover, at higher spatial resolution, greater frequency, and lower cost than field surveys. Easier monitoring also allows improved assessment of changes in rangeland management. Assessing these impacts is complicated by climatic drivers of plant growth, especially rainfall and temperature. Methods to discriminate the effects of management on ground cover from climatic effects include regional comparison benchmarking (e.g. Donohue et al. (2022)), reference benchmarking (e.g. Bastin et al. (2012)) or predicted benchmarking approaches such as RESTREND (Evans Geerken (2004)). Regional comparison and reference benchmarking both compare managed areas to reference areas or pixels. Predicted benchmarking approaches seek to standardise against

climatic drivers of growth. This is done independently for each pixel by modelling the time series of ground cover against climatic data time series. To allow for lagged effects of rainfall and temperature, derived climate predictors with various lag and accumulation periods might be derived and the best one(s) selected by comparing R^2 values, for instance. After accounting for climatic effects on ground cover, any remaining trends are interpreted as resulting from land management. We consider if the predicted benchmarking approach can be enhanced with alternative statistical tools. For instance, distributed lag models could predict the effects of climatic drivers with higher accuracy. In addition, pixels could be analysed in a spatio-temporal model instead of independently.

Improvements to macadamia crop yield forecasts by incorporating machine learning in model ensembles

Daniel Powell

Agri-Science Queensland

Crop yield forecasting forms an important component of strategic decision-making for the Australian macadamia industry. With growing pressures from challenging economic conditions, reliable forecasts assist in scheduling, logistics planning, and maintaining market confidence and price stability.

A range of highly variable and often interrelated environmental, economic and biological data sources are considered in developing predictions which create challenges in dealing with high dimensionality and covariance. And although past approaches implementing statistical forecasting models have achieved a high level of success, the intricate nature of influencing factors including climate, management practices, and varietal differences, invites efforts to explore further improvement to the accuracy of the forecasts.

Machine Learning (ML) methods have been shown to outperform other approaches in crop forecasting applications and, having advantages over traditional statistical models in their relaxed assumptions and flexibility to handle many more parameters, appear well suited to address this problem. We present our approach to incorporating ML in predicting crop yields and discuss some of the considerations in adopting these methods in tree crop forecasting. We also highlight the performance of a selection of ML methods used in recent crop estimates for macadamia production in Australia.

To average or not? Analysis of quadrat-level BOTANAL data in grazing experiments.

David Reid

Agri-Science Queensland

This study was motivated by the analysis of data collected from central Queensland properties as part of the Queensland Pasture Resilience Program, a partnership between the Department of Primary Industries, Meat & Livestock Australia (MLA) and the Australian Government through the MLA Donor Company. Pasture yield and composition were assessed using the BOTANAL pasture method (Tothill et al., 1978), which involves visually rating quadrats (typically 30–200 per plot), and calibrating these ratings with an equation derived from a set of ratings and corresponding cut quadrat weights to estimate dry matter yield (kg/ha).

The experiment utilised a split-plot design with three replicates, P fertiliser (+/-) as the main-plot factor, five pasture treatments as sub-plots, and 32 quadrats (samples) per sub-plot. Alternative approaches to analysis were considered, including modelling at the quadrat level or using quadrat-averaged responses at the sub-plot level. Diagnostic residual plots indicated skewness and heteroscedasticity at the quadrat level, while assumptions were adequately met at the sub-plot (treatment) level, which corresponds to the experimental unit level.

These observations raise several practical questions: Is it appropriate to average the quadrat data given potential non-normality in this stratum? Will simply calculating sub-plot means be sufficient to allow a valid analysis? Are such concerns less important given the semi-quantitative nature of BOTANAL data, which rely on subjective ratings calibrated via fitted relationships between scores and observed dry matter?

This paper examines these issues and discusses the implications. In summary, averaging the quadrats is acceptable if skewness is mild/moderate with the quadrat distribution not biasing the mean severely, particularly given the subjectiveness of the data. However, if skewness is severe then apply the appropriate transformation (e.g. log transformation).

Bayesian Variable Selection in Quantile Regression to Assess Joint Impacts of Climate Indices on Monthly Rainfall

Dilani Kaveendri Dambure Hewage

Canterbury Institute of Management, Sydney

Understanding how large-scale climate indices influence rainfall variability is important for assessing drought and flood risk in eastern Australia. This study examines the dependency between the joint impact of the Southern Oscillation Index (SOI), Dipole Mode Index (DMI), and monthly rainfall quantiles in New South Wales (NSW), Australia, with particular emphasis on the tails of the monthly rainfall distribution. To address this, we employ a hierarchical model that incorporates Bayesian variable selection within a spatial quantile regression framework, allowing for the integration of two climate indices and spatial variation in their impacts on different parts of the monthly rainfall distribution. The first level of hierarchy expresses the asymmetric Laplace distribution as a location-scale mixture. The second level expresses the prior of the regression coefficients as a two-component mixture, with components being a point mass at zero and a normal distribution. The third level incorporates spatial dependence by modelling the joint inclusion probabilities of SOI and DMI using multinomial logistic regression, where the argument of the link function varies spatially according to a Gaussian Process prior. The results show that SOI is the dominant driver of monthly rainfall variability across NSW, with strong impacts in winter and spring and a notable impact across all quantiles in November. DMI has an important influence during spring, particularly September and October, across multiple quantiles. The lower and upper quantiles show greater sensitivity to SOI and DMI than the median in selected months and locations. DMI also shows an apparent influence at lower quantiles in January and March, suggesting an effect on drier conditions. These results demonstrate the importance of assessing the joint impact of climate indices across the monthly rainfall distribution, providing more complete inference than mean-based or single-index analyses.

heritable: An *R* package for flexible heritability estimation from complex linear mixed models

Emi Tanaka

Australian National University

Heritability is ubiquitous in quantitative genetics, yet its estimation and interpretation in more complex experimental settings like multi-environment trials remain challenging. We present *heritable*, an *R* package that implements multiple generalised heritability estimators within a unified framework. Our package supports five methods (Cullis, Delta, Oakey, Piepho, Standard) for computing broad-sense and narrow-sense heritabilities directly from model objects from commonly used mixed-modelling software, including *lme*4 and *asreml*. By fitting a single genotype-by-environment model, users can flexibly obtain heritability estimates for the main genotypic effect, the marginal genotypic effect across environments, or stratified genotypic effects within user-defined environmental strata. Our package also provides confidence intervals, visual and tabular summaries to aid interpretation and reporting. Using a simulation study spanning a range of realistic experimental scenarios, we further offer practical recommendations on the choice of heritability measure under varying design complexities. Taken together, this work aims to improve the use, reproducibility, and interpretability of heritability in more biologically realistic settings.

Kernel Propagation of Gradient Boosted Margins for Spatio-temporal Forecasting

Ethan Kane Waters

Queensland's Department of Primary Industries

Gradient boosted trees are widely used due to their flexibility, minimal parametric assumptions and ability to learn complex relationships. However, they assume observational independence, potentially overlooking informative spatio-temporal structures. Existing modelling approaches that can explicitly consider spatio-temporal autocorrelation are often computationally intensive, rely on restrictive parametric assumptions or suffer when predictors are collinear. We propose a post-hoc method that augments gradient boosted tree predictions by using a kernel to propagate raw margins to nearby future observations. Margins are updated as a weighted combination of the original margin and a normalised kernel-weighted average of non-negative neighbouring margins

which is controlled by a mixing parameter. This yields estimates independent of neighbourhood density that are bounded by the original prediction and the local neighbourhood. The method does not modify model training and is compatible with any gradient boosting implementation that exposes raw margins. Empirical results from the motivating biosecurity forecasting task of invasive species presence demonstrate consistent improvements in predictive performance over standard gradient boosting models.

A Bayesian spatio-temporal nearest neighbor Gaussian process model for pooled genetic data

Imke Botha

University of Melbourne

Large scale genetic datasets often aggregate the total allele counts of distinct genetic markers. Inferring haplotype frequencies (i.e. the frequency of multimarker alleles) from these pooled data is a challenge. Previous spatio-temporal modeling in this context has been limited to 3 markers due to the computational cost. In this work, we propose a nearest neighbor Gaussian process (NNGP) model to improve scaling with the number of markers and observations. To infer the parameters of our model, we develop a novel sequential Monte Carlo squared algorithm, which uses particle Gibbs with ancestor sampling to mutate the NNGP function values. The latter has a linear cost in the number of observations and the number of NNGPs, and can be applied to a broad range of NNGP models. As a case study, we analyze genetic data relating to antimalarial drug resistance in Africa, and show our scaling results empirically on a 3 and 6 genetic marker dataset.

Bespoke workshops to empower Agricultural Research Scientists to perform their own routine statistics

Kerry Bell

Queensland's Department of Primary Industries

The demand for statistical collaboration in agricultural research projects in the grains industry conducted within Queensland's Department of Primary Industries (DPI) exceeds the capacity of the biometricians. Empowering researchers to perform routine statistical tasks offers a dual benefit: maintaining the quality of statistical practices in projects while enabling biometricians to focus on advanced methodologies, innovative statistical approaches, and strategic upskilling initiatives. Although the DPI biometry group has implemented various upskilling activities, these have not consistently translated into meaningful practice changes within research projects. To address this, tailored workshops designed around specific project groups have been trialled to bridge the gap between statistical training and practical application. Participants collaborate with biometricians to define learning objectives, explore statistical concepts using their own project data, and develop actionable plans to integrate these learnings into their research workflows. Preliminary feedback from participants indicates that these workshops have improved their confidence in data management and the use of statistics, leading to increased robustness in research outcomes. Additionally, the workshops have facilitated stronger collaboration between researchers and biometricians, creating a foundation for ongoing knowledge exchange and capacity building. This poster will provide an overview of the workshop design process, including the identification of key learning objectives, the integration of project-specific data, and the strategies employed to encourage practice change. The findings highlight the potential of tailored training initiatives to address capacity challenges and ensure the sustainability of high-quality statistical practices in agricultural research.

Estimating volcanic eruption recurrence rates using hierarchical trend renewal processes

Mark Bebbington

Massey University

Long-term (during quiescence) forecasts of volcanic eruptions are vital in estimating risk for land-use planning, but are challenging due to the typically sparse and incomplete data available in geological and/or historical eruption records. The typical solutions is to include in the analysis volcanoes with comparable physical properties and statistical behaviour (whatever these mean) to a target volcano. Analogue volcano recurrence patterns are thus used to estimate model parameters and forecast future eruptions. This approach often fails to consider the specific problem of "missing data", which is common due to the uncertain and possibly unknowable processes in geologic records over thousands to tens of thousands of years. At a finer level, volcanoes can exhibit waxing, waning or cyclic activity in the rate of eruption onsets. This latter variability can be modelled using the trend renewal process using combinations of power, exponential and sine functions. On 14 example volcanoes, cases of constant activity level, increasing trend, wax and wane of activity, and cyclic plus trend behaviours are all observed. We propose using a set of hierarchical trend renewal processes to model analogue volcanoes and account for missing data. From these we create a Bayesian model averaging scheme for forecasting which incorporates model uncertainty by combining the posterior distribution of the forecast times from each of the considered models. We apply this method to forecasting eruptions from Mt Taranaki in New Zealand, which last erupted in 1780 AD and has its entire eruption record preserved only in geological deposits.

Modelling plant growth for biological discovery

Andrew Bowerman

ANU

High-throughput plant phenotyping platforms, such as those in the Australian Plant Phenomics Network, generate dense, repeated measurements of plant growth over time. Extracting biologically meaningful information from these growth trajectories is not straightforward: mechanistic models require an explicit functional form to be specified in advance, which is often laborious to define and fit separately for each genotype or treatment, particularly when growth is irregular or environmentally responsive. We explore statistical approaches, including Gaussian processes and spline-based models, as a flexible alternative for characterising growth trajectories from phenomics data, without requiring a pre-specified functional form. These methods can accommodate nonlinear and irregular growth patterns directly from the data, while still yielding interpretable summaries such as growth rate, timing of key developmental transitions, and genotype- or treatment-level variation. Gaussian processes further provide a natural measure of uncertainty around these estimates, which is valuable when growth trajectories are noisy or unevenly sampled. We present this as a complementary approach to mechanistic modelling, rather than a replacement for it. Statistical characterisation of growth trajectories can help identify patterns that are worth investigating mechanistically, flagging where and when a genotype or treatment departs from expected growth, for example. Because these methods do not require a model to be fitted individually to each genotype or treatment, they are also well suited to the scale of modern phenomics platforms. In this way, statistical modelling offers a practical route to turning large-scale phenotypic data into biological insight.

Fast Approximate MM-Estimation for Outlier Robust Model Selection

Martin Huang

University of Sydney

Stratified robust model selection reduces the impact of large residuals and overrepresented outliers in bootstrap samples but is computationally intensive when fitting iteratively-solved robust estimators across many candidate models. We propose FAMM, a Fast Approximate MM-estimator, implemented as a weighted least squares fit with weights derived from a full-data MM-estimator, to reduce this computational cost. Substituting the MM-estimator with FAMM preserves model selection performance while achieving a substantial speedup in typical dataset settings. We also demonstrate that FAMM satisfies the required conditions for model selection consistency.

A forecasting disaster using sparce, scarce, and sporadic data to forecast a (mainly) unobservable nonlinear system

Melody Whitehead

Forecasts of complex natural systems are required to inform decisions from emergency management to climate response. Forecasting methods rely on interpreting current knowledge and data and extrapolating the result into the future with causal and temporal assumptions.

For volcanic eruption forecasts, where the geological non-linear system operates over thousands to millions of years, the temporal-recency bias of instrumental data forces assumptions like time homogeneity (stationarity). These are significant limitations given that forecasts are intrinsically time-dependent, and system dynamics control future events.

In an attempt to sidestep these assumptions, we take a more holistic approach, using data and knowledge that are not commonly employed for volcano forecasting, thus expanding our available pool of information, but also expanding our catalogue of data and knowledge problems.

Here, we use the exemplar of Ruapehu volcano and the Central Volcanic Plateau in Aotearoa - New Zealand, to demonstrate methods with which we

have tried and failed with, in the hope that this will inspire fruitful discussions on things that might be less ineffectual.

ecreml: An *R*-package to incorporate environmental covariates into a multi-environment trial analysis

Michael Mumford

Queensland Department of Primary Industries

A statistical methodology proposed in Mumford et al. (2023) integrates environmental covariates (ECs) into the multi-environment trial (MET) analysis with the aim of identifying environment drivers of genotype by environment by management practice (G×E×M) interaction effects. The methodology selects important ECs by incorporating them into a linear mixed model (LMM) with complex variance structures using $ASReml - R$. The selection procedure can be used to enhance the prediction of the performance of treatments and/or genotypes under untested environmental conditions. One difficulty arising from the proposed methodology is that the R-code to setup the selection procedure algorithm in conjunction with fitting the LMM is complicated, potentially creating a barrier to applying the methodology to other datasets.

This presentation outlines the recent development of an *R*-package called *ecreml* to compliment the Mumford et al. (2023) methodology. The *ecreml* *R*-package automates the iterative selection procedure which involves identifying important ECs via -fold cross-validation and incorporating them into the LMM using forward selection. Once EC terms are included in the LMM, non-significant EC terms are removed using backwards selection procedures. The R-package can easily be installed via GitHub (`https://github.com/michaelhm-daf/ecreml`) and includes a vignette demonstrating the functionality of the *ecreml*R-package when applied to a motivating MET sorghum dataset.

The *ecreml* *R*-package enables statisticians and agricultural scientists to easily analyse their own MET data using the methodology of Mumford et al. (2023) to identify key ECs contributing to G×E×M interaction effect.

Discrete Copula Lautum Information and Parsimonious Modelling

Mst Bithi Akter

UNSW Sydney

Copula modelling is fundamentally intended to separate marginal behaviour from dependence. Yet, for bivariate discrete distributions, classical copulas do not achieve such a clean separation: they are not uniquely identified, and familiar copula-based measures of association generally retain a dependence on the margins. We work instead with the discrete copula matrix, defined as the reverse Kullback–Leibler (RKL) projection of the joint probability matrix onto the class of matrices with discrete uniform margins. Under suitable support conditions, this construction produces a unique, margin-free representation of the dependence structure.

This paper addresses two questions arising naturally from that construction: how should the overall departure from independence be quantified, and how can the copula matrix be represented parsimoniously? We introduce discrete copula lautum information, defined as the RKL divergence from the independence copula to the discrete copula, and derive the large-sample behaviour of its empirical estimator.

For parsimonious modelling, copula matrices are represented through selected directions in the linear space of matrices with zero row and column sums. Along nested reduced models, the RKL-projection geometry yields an exact decomposition of total copula lautum information into contributions from successive directions and a residual component. This naturally motivates a greedy procedure in which directions are selected according to the additional dependence information they capture. The resulting framework produces interpretable, low-dimensional parametric models while keeping dependence measurement, model selection and estimation within the same information-geometric construction.

Quasi-posteriors for robust adaptive clinical trial design with count outcomes

Navodya Abirami Naotunna Gamacharige

Queensland University of Technology

Modern clinical trials are often designed within a Bayesian framework using trial simulation. This requires specifying assumptions for data generation including a treatment effect and a distribution for the data. Often the performance of the design depends on such assumptions meaning that if they are misspecified, then it could lead to the design performing worse than expected. To address this limitation for clinical trials where count outcomes are observed, we propose to develop designs within a general Bayesian inference framework. This enables replacing the likelihood in standard Bayesian inference with the quasi-likelihood to form a quasi-posterior. This results in inference being largely based on assumptions about the mean and variance model for the data. We propose to leverage such a property to design adaptive trials that are robust across a range of potential data generating models. Such an approach is evaluated through the re-design of a recently conducted clinical trial to assess the performance of citalopram (20 mg and 40 mg) in reducing cocaine use. A range of scenarios are considered including equidispersed, overdispersed, and underdispersed data where we compare our approach to standard practice within trial simulation. The results suggest that a design approach based on the quasi-likelihood instils robustness into the trial design and offers a potential solution to a major limitation in trial design.

Fitting Generalised Linear Mixed Models using Sequential Quadratic Programming

Peter Green

Tadorna Data Science

Finding maximum likelihood estimates in generalised linear mixed models (GLMMs) can be difficult due to the often-intractable integral over the random effects. Ensuring convergence can be tricky, especially in binomial GLMMs, and often multiple optimisers and settings need to be tried to get satisfactory results. This is exacerbated if you want to use parametric bootstrap for your inference. Sequential quadratic programming (SQP) is a method for solving optimisation problems with nonlinear constraints. SQP offers an alternative option for maximising the Laplace approximation to the GLMM likelihood,

bypassing the need for an inner penalised iterative reweighted least squares (PIRLS) loop. This talk discusses the implementation of SQP for GLMMs and compares its performance to other common approaches.

A Bayesian predictive density for Type II censored data for a normal distribution

Takeshi Kurosawa

Tokyo University of Science in Japan

We discuss predictive densities for Type II censored data. Type II censoring can be viewed as a problem of order statistics. For example, a fixed number of batteries is prepared, and some of the batteries run out over time. We aim to predict the time until the next battery runs out, which can be formulated as a Type-II censoring problem. We apply this method to failure-time data from aircraft air-conditioning equipment as an illustrative example. We evaluate predictive densities in terms of the risk based on the Kullback-Leibler divergence. We derive a new Bayesian predictive density for unobserved data under censoring. The derived predictive density has the minimax property within a class of scale-invariant families. In this study, we focus on the Type-II censoring problem for order statistics from normal distributions. We show that the predictive density can be expressed as a truncated t-distribution. As a benchmark, we consider the plug-in predictive density, obtained by substituting the maximum likelihood estimates into the unknown parameters of the true distribution. The plug-in predictive density also belongs to the scale-invariant family under the assumed setting. Therefore, we demonstrate that the proposed predictive density is superior to the plug-in predictive density in terms of risk under the Kullback-Leibler divergence, owing to its minimax property. Furthermore, we conduct simulation studies to demonstrate the validity of the proposed predictive density for predicting unobserved data under censoring.

Analysing longitudinal opioid dispensing patterns among injured patients using sequential mixture hidden Markov models

Tasmiah Sad Sutopa

Queensland University of Technology

Pain management following injury can be complex, and may involve prescription medications, particularly opioids. Prolonged opioid use is known to increase risk of mental and physical harms. Although many studies have explored individual-level factors associated with cross-sectional opioid use, less is known about opioid dispensing states over time. The main aim of this study is to identify and analyse community opioid dispensing patterns among injured patients. The study used a linked dataset comprising hospital admissions, community pharmaceutical opioid dispensing and mortality information for Queensland adults hospitalised due to injury between 2014 to 2015 and dispensed opioids at least once during 2-years of follow-up. We calculated weekly opioid morphine equivalents (OMEs) by aggregating the daily dispensing quantity for seven days. Weekly opioid exposure was categorised into three states: no opioid use, low opioid use (<50 OME), and high opioid use (50+ OME). Sequential Mixture Hidden Markov Models were used to identify groups based on the transition of dispensing states over time. We then applied multinomial logistic regression to explore the association between patient characteristics and group membership. Among 50,152 injured patients, five opioid trajectory groups were identified: Persistent high (3.2%), Persistent low (13.1%), Minimal (39.4%), Intermittent (11.2%), and Persistent no opioid dispensing (33.1%). Compared to patients with no subsequent hospitalisation, those with 5+ encounters had a much higher relative risk (RR: 7.07-13.69) of being in any opioid dispensing group than Persistent no opioid dispensing group. Intensive care unit admission and pre-injury opioid dependence were associated with Persistent high and Intermittent dispensing. Repeated hospitalisation, older age and pre-injury opioid dependence were associated with group membership. The study provides novel evidence on long-term opioid prescription trajectories among injured patients.

A minimal extension of the Bradley-Terry model for paired comparisons with cyclic dominance

Yuki Ohno

Tokyo University of Science

Paired comparison data arise in a wide range of fields, such as sports analysis, social choice, and machine learning. The Bradley-Terry model is a well-established mathematical model for estimating latent rankings of items based on such data. It imposes a probabilistic transitivity constraint under which each pairwise preference is required to be consistent with a global ranking, and this constraint contributes to excellent performance in many applications. However, empirical data sometimes exhibit cyclic dominance, resembling dynamics like "rock-paper-scissors", which may violate the transitivity assumption underlying this classical framework. In such cases, analyses based on this framework may lead to misleading inferences. Instead, we seek to quantify the extent of cyclic dominance in order to assess the degree of departure from transitivity. To this end, we propose a minimal extension of the Bradley-Terry model by introducing only a single nonnegative parameter. When this additional parameter is set to zero, the proposed model reduces to the Bradley-Terry model. It also provides a simple and interpretable measure of departure from the Bradley-Terry model in terms of transitivity. The usefulness of the proposed model is demonstrated through the analysis of real data, where it successfully captures cyclic dominance.

High-throughput phenotyping models predict wheat yield across current and future climates

Zixiong Zhaung

Future climate change includes both elevated CO_2 and warming, both climate factors could significantly influence photosynthesis and global wheat production, one of the major crops for food security1. However, few studies have addressed the combined effects of elevated CO_2 and warming (Ainsworth and Long, 2021). Breeding and selecting wheat that adapt to future climate change requires efficiently evaluating thousands of varieties across both current and future growing conditions. To address challenges at this scale, high-throughput phenotyping technologies offer a novel path to breed better crops for future climates. However, their value for predicting yield under novel climates remains

untested. Over two years I grew 86 diverse Australian wheat lines under current and projected end-of-century conditions (800 ppm CO_2 , +5 °C), phenotyping each with 3D multispectral scanning, digital microscope, hyperspectral leaf-clip reflectance, and a wheat-specific deep-learning model that predicts leaf photosynthetic physiology (Furbank et al. 2021). Over the two years of experiments, future conditions consistently reduced grain yield by 17 ± 2%. This further emphasises the needs to breed wheat varieties better adapted to future climate conditions. From over 140 high-throughput traits 30 features were selected using elastic net and ablation studies and was able to predict wheat yield with an R^2 of 0.62 using recently published in-context learning foundational models4. Additionally, this model could predict the yield of unseen genotypes with no loss of accuracy ($R^2 = 0.60$), which is needed to screen new breeding lines. This result demonstrates that predictive models using scalable, sensor-derived phenotypes mutual across growth environments rather than being environment-specific may be feasible. Additionally, predictive models using phenotyping-based approaches also offer mechanistic insights of adaptation strategies across environments. This further enhances the confidence for breeding and pre-breeding exercises for addressing food security issues driven by climate change.

Fusarium What? A method to identify mungbean lines and parents resistant and tolerant to Fusarium wilt

Lucas Peitton

Queensland Department of Primary Industries

Fusarium wilt in mungbean is caused by fungi within the Fusarium oxysporum and F. solani species complexes and heavily affects grain production in Australia. The need for lines with high yield potential that can either retain yield in the presence of disease or tolerate the effects of the pathogen is imperative. Current selection methods involve the study of disease incidence and grain yield separately, which may result in not using all the available information and thus lead to inefficient selection decisions. A statistical model is presented where geneotypic resistance and tolerance measures are simultaneously derived from a multivariate linear mixed model analysis on grain yield and disease incidence for Mungbean under normal (uninoculated) circumstances and disease pressure (inoculated). The method incorporates genetic relatedness via pedigree information and is implemented in the ASReml-R package. A visual tool will be highlighted that identifies lines with higher yield potential, yield retention under disease pressure and tolerance to disease in the evaluated environment

simultaneously. It has been demonstrated that resistance and tolerance proxies obtained from the genetic effects are independent of grain yield potential and disease incidence, allowing for more efficient selection decisions. As the genetic effects can be partitioned into additive and non-additive components, not only can improved lines be detected, but it is also possible to select potential parents for future crosses. These results are key in developing lines with better performance under disease pressure to ensure yield production for growers with or without the presence of the pathogen.

Abstracts for Talks: Wednesday

Keynote #3 Tues 9:05 – 9:50am

Causal inference in health research

Prof. Margarita Moreno-Betancur

Co-Director of the Clinical Epidemiology and Biostatistics Unit

The ultimate goal of medical and health research is to improve patient outcomes and population health. As a result, an overwhelming majority of clinical and public health research studies ask "causal" questions, concerning the effect of treatments, policies, behaviours and other exposures on health outcomes, even if often unadmitted. In many cases, especially in the current era of data deluge, these studies rely on observational (non-experimental) data to address causal questions. Unfortunately, for a long time the statistics discipline largely shunned the possibility of causal inference beyond randomised trials, and instead focused on the development of tools such as regression models without clarity regarding their usefulness and limitations for addressing the causal questions that substantive areas continued to ask. In recent decades, however, the discipline has seen the rise of a new area focused on determining the settings and approaches that could allow causal inference from observational data. This talk will first provide an overview of some of the fundamental contributions of this statistical area to enable and improve the study of causality in health research and then provide an overview of recent advances in the area seeking further to advance the statistical methods, teaching and practice.

Abstract contributed session #5 Wed 10:00 – 10:30

1. Dynamic risk prediction of ICU discharge alive using longitudinal urea-to-creatinine measurements

Sophie Zaloumis

University of Melbourne

Up to 6 million intensive care unit (ICU) patients worldwide develop ICU acquired weakness (ICU AW) annually. ICU AW is linked to mortality and long term reductions in health related quality of life. Early physical rehabilitation in ICU may improve strength, but identifying patients who could benefit from early rehabilitation and knowing when early rehabilitation should occur are uncertain. Identifying real-time patient biomarkers related to rehabilitation outcomes could help inform these decisions. We assessed the urea to creatinine ratio (UCR) as a real time biomarker of catabolism to guide rehabilitation. A dynamic prediction model used joint modelling: a longitudinal submodel for log2(UCR) (linear mixed effects with natural cubic splines) and a Cox survival submodel for discharge alive, linked by current value and slope. The joint model was fitted with Hamiltonian Monte Carlo via stan_jm (rstanarm R package). Performance was evaluated by calibration, dynamic area under the curve (AUC) and patient level predictions. In 180 patients (167 discharged alive, 13 died), a doubling of UCR (for the same trajectory slope) was associated with a 0.82 fold (95% credible interval (CI) 0.62–1.08) reduction in the hazard of ICU discharge alive. For the same current UCR level, a one unit increase in the current slope of log2(UCR) trajectory was associated with a 0.23 (95% CI 0.06–0.90) reduction in the hazard of ICU discharge alive. The longitudinal model fit well; survival probabilities were sensitive to UCR changes. Higher and rising UCR predicted higher probability of remaining in ICU. Calibration was reasonable in the training and test datasets; dynamic AUCs were unstable in the small test set (n=36). Results suggest UCR may help dynamically predict outcomes to inform rehabilitation decisions.

2. How Are Researchers Using MIMIC? A Methods-Focused Review of Studies in Critical Care

Tayyebeh Kenarangi

VSNi

The Medical Information Mart for Intensive Care (MIMIC) database is a widely used open data resource in critical care research. Its accessibility has led to large numbers of research studies being published; however, there are concerns about the comparability and reproducibility of studies due to variation in design and analysis decisions and how they are reported. Methods: We reviewed studies published between 2015 and 2025 that used the MIMIC database. A random sample of 100 eligible studies was identified through the OpenAlex bibliographic database and assessed using a structured extraction framework. Information was extracted on study aims, cohort definitions, exclusion criteria, outcome specification, handling of missing data, and analytical methods. The review focused on how key design and analytical decisions were reported and justified. Results: Studies reported findings from general ICU and specific clinical populations; the study population could not be determined for 30 of the 100 studies. Predictive modelling was the most common study design ($n = 60$), but only one study included external validation. Core methodological decisions were frequently underreported or inconsistently justified, including length-of-stay thresholds, age criteria, and mortality time horizons. Handling of missing data was not reported in more than half of studies ($n = 56$), despite missing data being common in MIMIC. Conclusions: MIMIC-based studies demonstrate considerable variation in study and analysis that are not always transparently reported. Improved reporting of design and analytical decisions is needed to strengthen the validity, reproducibility, and generalisability of registry-based critical care research.

Abstract contributed session #6 Wed 11:00 – 12:30

1. A geostatistical framework for spatial sampling design and data integration in on-farm experimentation

Zhanglong Cao

Curtin University

On-farm experimentation (OFE) increasingly generates rich spatial datasets, including high-resolution yield maps, proximal sensing surveys, remote-sensing products, and sparse point-source laboratory measurements such as soil cores and tissue tests. However, statistical guidance on how to design cost-effective spatial sampling strategies and integrate these heterogeneous data sources when estimating treatment effects remains limited. This study presents a geostatistical framework addressing three practical questions encountered in large-scale field experiments. First, how densely should point-source measurements be collected? We derive sampling intervals from the variogram range via the kriging equations, and show by Monte Carlo simulation that a multi-scale nested reconnaissance design of roughly one hundred points recovers the spatial structure needed to calibrate the main sampling campaign more reliably than grid or random designs of the same size. Second, how many field cores should be combined within a composite sample? We derive an optimal compositing rule that balances laboratory cost, field cost, and the nugget-to-sill ratio, reducing sampling variance with minimal information loss. Third, how should sparse point samples be integrated with dense spatial covariates? Comparing a kriged-covariate approach against a joint bivariate spatial mixed model, simulation shows that integration can materially improve treatment-effect estimation, with benefits governed by point-sample density and, for the joint model, whether the cross-layer spatial covariance is estimable. We characterise when each approach is preferable, and when neither improves on the dense data alone.

2. ST-MICE: A Trajectory-Aware Multiple Imputation Framework for Spatial Coordinates and Associated Variables

Amuchechukwu Ibenegbu

UNSW Sydney

Spatiotemporal trajectory datasets are central to environmental and scientific applications such as tropical cyclone monitoring, ocean circulation analysis and movement ecology. They contain ordered geographic coordinates and associated physical or environmental variables, but are often incomplete due to sensor failure, communication gaps, irregular sampling and quality-control filtering. A key challenge is that the coordinates defining where an observation belongs may be missing, while variables measured or associated with that location may also be absent. Standard imputation methods are not fully suited to this setting because they often ignore nonlinear trajectory geometry or assume that coordinates are fully observed. We propose ST-MICE, a trajectory-aware multiple imputation framework for imputing missing coordinates and associated spatiotemporal variables. ST-MICE represents each trajectory using a principal curve and derives a normalised arc-length parameter that captures geometric progression along the path. This representation guides coordinate interpolation and serves as a geometry-aware predictor within a chained-equation framework for refining associated variables. By combining coordinate reconstruction with multivariate refinement, ST-MICE preserves trajectory structure while exploiting dependencies among coordinates and variables. The method is evaluated on publicly available environmental trajectory datasets, including tropical cyclone tracks and ocean surface drifter paths, under missing completely at random (MCAR) and missing at random (MAR) mechanisms, and is compared with several baselines. Results show that ST-MICE achieves competitive and often superior performance, especially for nonlinear trajectories and for simultaneous missingness in coordinates and associated variables.

3. Tools for Causal Interpretation

Chris Howden

Statistical Consulting Unit at Sydney University

Causality is becoming the new buzz word. Everywhere you look there are causal workshops and talks, some journals are even encouraging authors to use Directed Acyclic Graphs (DAG's). Researchers across disciplines are asking how they can move beyond reporting associations to making credible causal claims. In this talk I will discuss when we should be aiming for causality instead of correlation and suggest a practical workflow for causal inference applicable across most domains. Some of the more common tools for causal interpretation will be discussed such as randomisation & Target Trials along with how we might classify them into different types of tools – each with different advantages, limitations, assumptions and tricks for implementation. An important distinction will be made between "objective" causal tools that work largely irrelevant to the users experience e.g. randomisation, versus "subjective" tools that require experience and skill to use effectively e.g. DAG's and Target Trials. Understanding these pros, cons, assumptions and differences will help researchers select the appropriate tools for making stronger, more defensible causal claims.

---

4. Parsimonious modelling of unstructured and structured random-effect covariance in glmmTMB

Maeve McGillycuddy

UNSW Sydney

Generalised linear mixed models provide a framework for modelling dependent data. Many applications require multiple covariance structures to account for different sources of correlation; however, finding a single open-source modelling tool capable of fitting these structures can be challenging. We present two new covariance structures in the R package glmmTMB that broaden the range of models that can be fitted while retaining the package's familiar syntax. The first method introduces reduced-rank covariance structures for high-dimensional random effects. This approach parsimoniously approximates a covariance matrix using latent variables. The second method introduces a proportional covariance structure, allowing users to specify random effects

whose covariance matrix is proportional to a known covariance matrix. This has a range of applications, including phylogenetically and pedigree structured random effects, and spline smoothers defined through basis functions. Together, these methods allow users to combine latent variables, structured random effects, smoothers, and other mixed-model terms within a single modelling framework. We illustrate their applications using ecological and plant breeding examples, demonstrating models that were previously difficult or impossible to fit using existing software.

---

5. Feature selection in censored-covariate regression models

Rajan Shankar

University of Sydney

Censored-covariate regression models arise when a response variable is modelled using a covariate that is only partially observed due to censoring. Existing methods for fitting these models focus on parameter estimation, and do not directly address the problem of selecting features that may affect the response variable, the censored covariate, or both. We propose a penalised likelihood method that jointly models the response variable and a single right-censored covariate, and applies separate L_1 penalties to the feature coefficients in the two model components. Parameters are estimated using an Expectation-Minimisation (EM) algorithm, where the minimisation step separates into two optimisation problems corresponding to the response-variable and censored-covariate model components. A simulation study shows that the method can recover distinct sets of active features for the two model components under adequate signal. We apply our method to Huntington disease (HD) data from the PREDICT-HD study, using cognitive ability measured by the Symbol Digit Modalities Test score as the response variable and time-to-diagnosis of HD as the right-censored covariate, and using brain region volumes as features. The selected features suggest partly distinct brain-region profiles for cognitive ability and time-to-diagnosis, including basal-ganglia regions in the time-to-diagnosis model component. A stability analysis shows that the HD data contains weak signal, clarifying the extent to which the selected features should be trusted.

---

6. Model-based Assessment of Functional and Phylogenetic Diversity

Shaoqian Huang

UNSW Sydney

Biodiversity is a quantity of fundamental interest, yet its measurement remains challenging. Two key aspects of biodiversity are functional diversity (FD) and phylogenetic diversity (PD).

Commonly used distance-based diversity measures are typically affected by sampling intensity. Moreover, there is currently no established statistical framework for analyzing how diversity varies along environmental gradients.

Here, we propose a model-based approach to assessing PD and FD that is robust to variation in sampling intensity and enables the study of diversity changes along environmental gradients.

By reparameterizing abundance models, we decompose variation in abundance into components associated with total abundance (alpha-diversity) and relative abundance (beta-diversity). Simulation results further indicate that changes in PD and FD arise when the beta-diversity coefficients are aligned with the principal axes of the phylogenetic or functional similarity matrix, which can be interpreted as linear contrasts of the beta-diversity coefficients. In addition, we develop a biplot by projecting the compositional effects of each observation, providing a visualization of how diversity differentiates across gradient levels in a percentage cover dataset.

Overall, this framework integrates the measurement of PD and FD changes along environmental gradients into a model-based setting, offering improved interpretability, flexibility, and computational efficiency.

Abstracts for Talks: Thursday

Keynote #4 Thurs 9:05 – 9:50

Bayesian or machine learning approaches: what works best for genomic prediction?

Ben Hayes

The University of Queensland

The foundational paper for genomic prediction, predicting complex traits from genome wide DNA markers, was published in 2001. Since then, genomic prediction has been widely adopted in crop and livestock production, and in human genetics for prediction of disease risk. In this overview, statistical methodologies for genomic prediction are compared, and we then speculate on the future of genomic prediction methodologies.

Abstract contributed session #7 Thur 10:00 – 10:30

1. AI-Integrated Genomic Prediction Framework for Parent Selection in Australian Macadamia Germplasm

Lin Ni-Jung

University of Queensland

Macadamia, an Australian native subtropical perennial tree crop, presents critical breeding challenges due to inherent heterogeneity, long juvenility, single plant basis progeny evaluation, and the complex genotype by environment interactions. Aiming to develop and validate an efficient parent selection framework by integrating advanced statistical methods, genomic tools and artificial intelligence (AI). Phenotypic yield data of 10,156 individuals across ten trials in Queensland and New South Wales since the 1990s. Genomic comprehensive analysis included 2,010 accessions across four species (*Macadamia integrifolia*, *M. tetraphylla*, *M. ternifolia*, and *M. jansenii*) that were genotyped by 12,655 DArT markers, revealing 45.45% missing marker rate. Informative markers were filtered by call rate at 90% and minor allele frequency at 1%, remaining 1,931 markers. Three imputation algorithms: (1) mode, (2) k-nearest neighbors (KNNi), and (3) random forest (missForest), implemented in R. Imputation accuracy was assessed by randomly masking 5% of the data, 78.60%, 87.14%, and 84.73%, respectively. Population structure identified 2 to 4

genetic clusters explaining 17.45% to 20.05% of the total variation, with clear species differentiation and admixture in breeding germplasm. Estimate prediction through family variance for half-sib and full-sib accounted for 31.69% and 15.84% of the variation, respectively. Pedigree-based best linear unbiased prediction (ABLUP) estimated narrow-sense (h2) and broad-sense (H2) heritability at 25.29% and 39.20%, matching plot means (A*BLUP) estimates. An AI-integrated framework combining phenotypic analysis of multi-environment trials, linear genomic models, and non-linear approaches will be developed, including random forest, support vector machine, multilayer perceptron, and stacking ensemble learning. This framework has strong potential for shortening breeding cycles and accelerating genetic gain in macadamia breeding programs.

---

2. Parameter Tuning in Statistics and Machine Learning Models for Plant Genetic Improvement Using Bayesian Optimization

Zitong Li

CSIRO

Genomic prediction has become a central tool in modern plant and animal breeding, enabling early selection of superior candidates using genome-wide marker data. A wide range of high dimensional statistics and machine learning models have been applied in this context, including penalized regression, Bayesian regression, ensemble methods such as random forests, and deep learning. However, their predictive performance is highly sensitive to hyperparameter choices, and efficient tuning remains challenging due to high-dimensional, heterogeneous search spaces and limited computational budgets. Bayesian optimization provides a principled and sample-efficient framework for hyperparameter tuning in various machine learning models. Nevertheless, its effectiveness can vary across model classes and problem settings, and practical guidelines for its application in genomic prediction remain limited. In addition, hyperparameter tuning is typically formulated as a single-objective problem focused on predictive accuracy, despite the importance of model simplicity and interpretability in breeding applications. In this work, we present a systematic study of Bayesian optimization for genomic prediction across a diverse set of model classes. We evaluate its performance under different hyperparameter tuning scenarios and analyse how optimization behaviour varies across models and search spaces. We further investigate the impact of objective selection in hyperparameter tuning by comparing predictive performance-based criteria

with model selection criteria, and explore a multi-objective formulation to balance accuracy and model parsimony. Our results provide practical insights into effective hyperparameter tuning strategies for genomic prediction, demonstrated on both simulated data and real cotton datasets.

Abstract contributed session #8 Thur 11:00 – 12:15

1. Resource optimised designs: Determining variety replication status

David Hughes

University of Wollongong

Early evaluation stages in plant breeding programs involve large numbers of test lines, grown in experiments in order to select the best. Widely used partially replicated designs of Cullis et al. (2006) have led to significant improvements in selection accuracy. In their design construction, Cullis et al. (2006) did not use information on the genetic relatedness of lines, due to the lack of suitable design software at the time. Cullis et al. (2020) used the ODW package to show that selection accuracy can be substantially improved with designs that incorporate genetic relatedness in the allocation of plots to test lines. This talk will present an additional step that uses genetic relatedness to determine an optimal subset of test lines to be replicated, subject to seed availability. Advantages of this approach are demonstrated to resource allocation using an in silico study based on an early stage selection experiment in chickpeas.

---

2. Estimating cell-type-specific gene co-expression and spatial autocorrelation for spot-level spatial transcriptomics data

Max Woollard

Accurately estimating gene-gene co-expression is key to understanding biological processes. However, it is important to recognise the context of these gene-gene co-expressions. First, these co-expression structures can vary with different cell types, meaning we should expect the corresponding covariance matrices for different cell types to be different(1). Furthermore, gene expression activity varies across space with spatial autocorrelation, which must be taken into account to accurately estimate the covariance matrices(2). Spatial transcriptomics data addresses these challenges by measuring transcriptomic activity at many spots in a tissue, and these spots often contain a mixture of different cell types(3).

To this end, we propose CSSpatialNet, a statistical model to describe both cell-type-specific gene covariances and and cell-type-specific gene expressions' spatial autocorrelations from spot-level spatial transcriptomics data. The parameters of the model are then estimated by maximising the likelihood of the observed spot-level data, which is computationally efficient using the standard gradient descent method.

Our simulation studies show CSSpatialNet provides more accurate estimation of the cell-type-specific covariance matrices than competing methods that either ignore cell-type or discard spatial information Additionally, CSSpatialNet estimates the spatial autocorrelation of each gene within each cell type. We are also evaluating our method on real spatial transcriptomics data at a range of resolutions, such as Visium Mouse Embryo data. By accounting for the varying levels of spatial autocorrelation among genes and cell types, we expect to see more accurate cell-type-specific gene co-expression estimates.

3. From Pixels to Phenotypes: Building Statistical Pipelines for Drone-Based Field Experiments

Colleen Hunt

Queensland Department of Primary Industries

Early biomass measurement in sorghum is essential for assessing plant health at early developmental stages and identifying factors such as soil deficiencies or disease. Conventionally, biomass estimation relies on destructive sampling, which limits repeated measurements on the same plants and hinders longitudinal analysis of biomass accumulation. This limitation has historically impeded efforts to explain growth trajectories. In this study, destructive biomass measurements from a sorghum field trial were used as ground truth to develop predictive models estimating plant biomass from drone imagery. Calibration analysis was conducted to ensure precise measurement of early plant biomass via image analysis. The developed framework facilitated multiple biomass assessments throughout the growing season. The 2023 heterosis trial included 150 inbred lines and 457 hybrid crosses. Early biomass was predicted at two time points using both destructive weight measurements and estimates obtained from drone imagery. A multi-trait dynamic linear mixed model was applied to evaluate the accuracy of the image-based predictions. Results demonstrated a high correlation between destructive measurements and imaging estimates, indicating that biomass imaging for early sorghum biomass assessment is both reliable and effective.

---

4. Multi-trait multi-environment models enhances prediction and interpretation of grain yield in wheat field trials

Nick Fradgley

University of Sydney

Linear mixed models are widely used for analysis of genotype by environment interactions (GEI) in multi environment field trials (METs) for crop breeding. Well established methods employ factor analytic approximation of heterogeneous genetic covariance structures across environments to group environments into interaction classes (iClasses) that minimise cross-over interactions for single traits of interest (Smith et al., 2021). While grain yield is the main trait of interest for cereal crops such as wheat, multiple secondary traits, such as

phenology, plant architecture and resistance to often unpredictably occurring diseases, are known to drive complex GEI effects for grain yield. Enhanced prediction of secondary adaptive traits may help improve prediction of grain yield GEI and enhance selection of well adapted genotypes.

Here, we extend the factor analytic linear mixed model framework for joint modelling of multiple secondary traits—including days to heading, plant height, and disease scores for Yellow (Stripe) rust and Fusarium crown rot—in addition to grain yield, to a multi-year MET as part of the CAIGE program (Trethowan et al., 2024). Accurate model fitting required partitioning genetic covariance among traits and environments, while separately accounting for correlated residual spatial trends among traits within environments. Secondary traits were generally more heritable than grain yield, exhibited stronger genetic correlations across environments, and showed higher proportions of additive genetic variance. Importantly, genetic correlations between secondary traits and grain yield varied markedly in both magnitude and direction across yield iClasses, reflecting their contrasting adaptive processes. This approach could be particularly useful when highly heritable secondary traits are assessed in earlier stages of a breeding program before multi-environment yield trials, thus allowing more accurate selection from a larger set of candidate genotypes.

---

5. Validity and interpretability of estimated effects from mixed model based meta-analysis

Khageswor Giri

Agriculture Victoria Research

Mixed models are a commonly used method to undertake meta-analysis for synthesising evidence across different experiments/studies. Such analyses rely on a key implicit assumption that all contributing studies estimate the same underlying quantity on a comparable measurement scale. In this presentation, we use meta-analysis of data from individual dairy cow pasture dry matter intake (DMI) from a series of experiments to illustrate how violation of this assumption can affect the validity and interpretability of estimated effects.

Individual cow pasture DMI, estimated using the alkane method, was analysed across multiple controlled experiments using linear mixed models with experiment, cow and cow within experiment as random effects. Initial diagnostics of the fitted model revealed a pronounced temporal pattern in the estimated experiment effects: experiments conducted before 2021 tended to have negative

effects, whereas those conducted after 2021 tended to have positive effects. The shift corresponded to an average difference of approximately 5 kg DMI per cow per day between earlier and later experiments, suggesting a systematic change that could not be readily attributed to biological factors alone.

Further investigation indicated that a refinement to the laboratory alkane estimation procedure occurred around the same period, providing a plausible explanation for the observed positive shift in estimated pasture DMI. We discuss diagnostic approaches for detecting such scale shifts, practical mitigation strategies for mixed-model based meta-analysis, and the broader implications for statisticians working with large, multi-study datasets where measurement/ estimation protocols may evolve over time.

AGUAi annual general meeting (AGM)

12:45-13:30

Keynote #5 Thurs 13:30 – 14:15

Modelling continuous data with exact zeros

Associate Prof. Peter Dunn

University of the Sunshine Coast

Statistical models for continuous data (e.g., linear regression) and discrete data (e.g., logistic regression) are commonly used. Continuous data with exact zeros, while less common, appear in many applications. In this talk, different models are discussed for modelling this type of data, with a focus on using Poisson-gamma models, which are Poisson sums of gamma distributions. Poisson-gamma distributions are part of the EDM family, so can be used for modelling within the generalised linear models (GLMs) framework. This means that the fitting, model selection and residual analysis tools familiar from GLMs also apply to GLMs fitted to continuous data with exact zeros.

Abstract contributed session #9 Thur 14:30 – 15:00

1. Harnessing aerial imagery to overcome spatial complexity in yield trials

Zachary Aldiss

University of Illinois Urbana-Champaign

Early-stage wheat breeding trials often suffer from low heritability because environmental noise, soil heterogeneity, and water gradients obscure true genetic potential. This study evaluates the integration of Unmanned Aerial Systems (UAS)-derived spectral data into spatial Linear Mixed Model (LMM) frameworks to account for residual variation and enhance genetic gain. Using data from 176 US wheat trials via a collaborative, NIFA-funded public breeding initiative, we fit univariate spatial LMMs and evaluated multitrait frameworks pairing grain yield with high-throughput phenotyping (HTP) latent variables and agronomic trait test weight used as a benchmark. Molecular data from 146 trials was further integrated to assess additive genetic variance partitioning. Our analysis affirms that HTP integration is a highly scalable strategy for reliability recovery, particularly across diverse breeding programs operating in spatially complex environments where traditional ANOVA-based methods struggle. By applying a standardized secondary trait from UAS imagery, we effectively recovered trial reliability in noisy environments where spatial trends otherwise dominate. Furthermore, the multitrait frameworks allowed for the extraction of biological signals from canopy spectral data, which successfully partitioned residual variation and improved the precision of variety effect estimation. However, our findings demonstrate that the value of HTP is strictly conditional on the baseline reliability of the trial, with diminishing returns observed in environments that already feature high univariate heritability.

2. Healing the History: A Robust Estimation Framework for Addressing Misidentification in POPAN Capture-Recapture Models

Adam Wong-Toi

Following heavy 20th-century commercial whaling, Southern Hemisphere humpback whales have experienced rapid demographic recovery. Historically, monitoring relied on shore-based point censuses, which did not track individual life histories over time (Noad et al., 2020). Modern platforms, like Happywhale, collate fluke images to provide individual photo-identification. While these multi-decade datasets capture expansive longitudinal histories, they raise key analytical challenges. This study evaluates the vulnerability of open-population capture-recapture models (POPAN) to capture heterogeneity and misidentification errors. By applying both standard POPAN and finite-mixture 'Capmix' models to a 41-year dataset comprising 15,792 individual capture histories, we estimated contemporary population dynamics. Through simulation, the impact of individual misidentification is robustly quantified. To address this bias, an iterative stochastic reallocation algorithm was developed that uses the joint likelihood to reallocate misidentified "ghost" individuals to statistically plausible capture histories. In simulation, with a 1.5% per-image misidentification rate, the bias in the superpopulation size, N*, was approximately 7.2%, while survival and recruitment metrics remained stable. The iterative reallocation algorithm successfully reduced this abundance bias to 0.05%. Applying the framework to the Happywhale eastern Australian humpback whale data yielded a 2024 population estimate of 63,584 individuals, and an annual population growth rate of 10%, consistent with literature from land-based studies (Noad et al., 2020). This framework successfully corrects for misidentification bias when supplied with an estimated error rate from external matching validation exercises. It provides the necessary correction to citizen science datasets, securing a precise and structurally sound framework for long-term population modelling by healing the history.

Abstract contributed session #10 Thur 15:30 – 16:15

1. Ensemble modelling and visualising uncertainty in fisheries stock assessments

Alise Fox

Fisheries Queensland

Fisheries stock assessments play a central role in decision-making under uncertainty, informing management of complex ecological systems with significant social and economic consequences. A key challenge is not only quantifying uncertainty, but representing it in ways that are both statistically robust and accessible to stakeholders.

This presentation uses a recent stock assessment as a case study to examine how ensemble modelling and visualisation can improve the communication of uncertainty. In this assessment, spawning biomass was estimated from an ensemble of 12 plausible model scenarios rather than a single "best" model.

The ensemble approach allowed structural uncertainty—arising from assumptions about biology, data weighting and other model components—to be explicitly incorporated into final outputs. However, presenting the results of multiple models introduces challenges for interpretation, particularly when communicating to stakeholders with varying levels of statistical literacy.

I will present a range of strategies used to summarise ensemble outputs, including probability distributions of biomass and simplified summaries of scenario variability. These approaches aim to balance transparency with interpretability, ensuring that uncertainty is neither obscured nor overwhelming.

A key focus of the talk is how ensemble outputs reshape the role of visualisation and interpretation in applied statistics. Rather than supporting a single model narrative, visualisations must synthesise information across competing models, highlight robustness of conclusions, and reveal where uncertainty meaningfully affects decisions.

This work demonstrates that combining ensemble modelling with thoughtful visual design provides a powerful framework for communicating uncertainty in high-stakes applied contexts.

2. Keeping the eastern king prawn stock assessment in line

Joanne Wortmann

Fisheries Queensland.

Commercial trawl vessels have changed considerably over time through the adoption of new and improved technologies and fishing gear. As a result, the relative fishing efficiency (or fishing power) of trawl vessels has been explicitly quantified in eastern king prawn stock assessments. A recently incorporated factor in estimating fishing power is line trawling behaviour. Line trawling occurs when a vessel trawls back and forth along the same path during a night, targeting specific areas where prawns are likely to aggregate. This behaviour can improve catchability by repeatedly targeting high-density patches, increasing the likelihood that prawns disturbed or missed in earlier passes are captured in subsequent trawls. This focused, iterative fishing strategy also exploits aggregating behaviour, effectively concentrating effort where catch rates are highest and reducing time spent searching. A linear mixed-effects model indicates that line trawling increases fishing efficiency, resulting in average catch increases of 11–19%. Catch and effort data used to derive the index of abundance in the eastern king prawn stock assessment are obtained from commercial logbooks (i.e. fishery–dependent data). It is therefore important to adjust for – that is, remove – the effects of changes in fishing power over time, so that trends in catch rates reflect changes in stock abundance rather than improvements in fishing efficiency. More generally, this work provides a case study in accounting for changes over time in how the data are collected, demonstrating how shifts in technology and behaviour can introduce bias into trends if not explicitly modelled.

3. Integrated Species Distribution Modeling for Fisheries Abundance Data: A Single Index Approach

Quan Vu

Fisheries Queensland.

In fisheries ecology, species abundance data are often collected by multiple surveys, each with unique characteristics. This article focuses on Atlantic sea scallop abundance data along the northeast coast of the United States, collected from two bottom trawl surveys which cover a larger spatial domain but have low catch efficiency, and a dredge survey which is more efficient but limited to domains where the species are believed to be present. To model such data, integrated species distribution models (ISDMs) have been proposed to incorporate information from multiple surveys, by including common environmental effects along with correlated survey-specific spatial fields. However, while flexible, these ISDMs can be susceptible to overfitting, which can complicate interpretability of the shared environmental effects and potentially lead to poor predictive performance. To overcome these drawbacks, we introduce a novel single index ISDM, built from a single index (with spatial random effects) that represents a latent measure of the true species distribution, and survey-specific catch efficiency functions which map the single index to the survey-specific expected catch. Our results show that the single index ISDM offers more meaningful interpretations of the environmental effects and survey catch efficiency differences, while potentially achieving better predictive performance than existing ISDMs. This is joint work with Francis Hui, Alan Welsh, Samuel Muller, Eva Cantoni, and Chris Haak.

Abstracts for Talks: Friday

Abstract contributed session #11 Fri 09:05 – 10:35

1. Advanced Statistical Modelling and Sensor Technologies for Enhanced Animal Welfare Assessment

Chanatda Somchit

New Zealand Institute for Bioeconomy Science

Advanced statistical modelling plays a critical role in transforming high-volume sensor data into meaningful and actionable indicators of animal welfare. Sensor-based technologies are increasingly used to provide continuous, objective measures of animal health, behaviour, and welfare. Identifying behaviour and environmental thresholds beyond which animals begin to show signs of compromised health including heat stress and lameness is fundamental to effective monitoring and management. Equally important is the validation of accelerometer-based systems to ensure accurate behaviour classification. Statistical methodologies such as generalized additive mixed models (GAMMs) including those developed for large-scale and giga datasets by Wood et al. (2015, 2017) are well suited to analysing complex, non-linear relationships among behaviour, environmental conditions, and health outcomes. These methods support the identification of risk thresholds for the early detection of welfare challenges and timely intervention. Complementary machine-learning techniques including decision tree algorithms provide robust and interpretable approaches for behaviour classification and sensor validation. Together, these analytical frameworks enable the extraction of reliable insights from complex datasets and support evidence-based decision-making. This presentation will highlight analytical frameworks that integrate advanced statistical and machine-learning methods with high-frequency sensor data to support animal welfare assessment. Case studies will demonstrate the application of GAMMs to identify environmental thresholds associated with heat stress and lameness in cattle, and the development of decision tree algorithms for behaviour classification in multi-species grazing ruminants using accelerometer data.

2. Teaching Clients to Catch Their Own Statistical Fish: Reflections on Building Statistical Capability Through Consulting

Alexandra Green

Sydney Informatics Hub

At the Sydney Informatics Hub, I have recently experimented with a consulting approach inspired by the idea of teaching clients to catch their own statistical fish. Rather than repeatedly providing analyses, the goal is to help researchers move from "how do I do this?" to "I can do this myself". This talk reflects on a series of one-to-one consultations with public health researchers who were primarily SPSS point-and-click users. The approach began with me leading an initial analysis and providing a reproducible R workflow. Rather than treating each new project as a separate engagement, that workflow became a template that could be adapted and extended as new research questions emerged. Existing analyses were modified to accommodate different outcome types, allowing clients to learn new statistical methods within a familiar analytical framework. Over time, clients developed both computational and statistical confidence. One client progressed from asking "How do I interpret this graph?" to "How can I fit a count model and assess whether zero inflation is required?". Researchers increasingly ran analyses independently, adapted code for new projects, assessed model assumptions, and selected appropriate modelling approaches without me conducting the analysis. Shared, reproducible code also encouraged peer learning, with clients troubleshooting and building on one another's work. I reflect on the trade-off between short-term efficiency and long-term capability building, and how reusable code, project-based learning, and progressive transfer of responsibility can support statistical literacy and independence. Finally, I consider whether a consulting engagement's success might be measured not only by analyses completed, but by clients' ability to complete the next one themselves.

3. Version-Controlled Protocols and Equivalence Testing for Reproducible Drone Phenomics

Arden Burrell

University of Sydney

Across the Australian Plant Phenomics Network, drone phenotyping data are collected by different operators (network nodes), across multiple sites and years. These data are comparable only if every team collects, calibrates, and processes them the same way. The Airborne Phenomics Excellence project found that when operators repeat a drone flight or fly a shared target, results disagree, and the dominant, fixable cause is process, not sensors. The variance lived in field and processing procedure, and in whether the agreed method was written down, current, and followed. This presentation describes our response: managing scientific protocols the way software teams manage code, and then testing statistically whether they work.

Procedures live in a shared version-controlled repository, are revised through reviewed pull requests, locked as a published read-only wiki that teams follow in the field, and issued as tagged, citable releases, giving every dataset a permanent record of the method it was collected under.

But writing protocols down does not prove they deliver the comparability the network depends on. Conventional significance testing answers this poorly: failing to reject a difference is not evidence of equivalence, and with large pixel samples trivial differences become "significant". In a controlled validation campaign, we frame cross-sensor calibration as a formal equivalence problem, deriving an equivalence bound and applying two one-sided tests (TOST) for mean reflectance, then assessing distributional equivalence (quantile TOST, Wasserstein distance, variance ratios) and a multivariate, classifier-based effect size across the visible-near-infrared and shortwave-infrared regions.

The result is a per-wavelength verdict (equivalent, non-equivalent, or inconclusive) showing where a procedure works and where recalibration is needed, a practical way to control variance and bias between operators, sites, and years that recur in multi-year, multi-site monitoring.

4. Integrating Spatial Data and On-Farm Experimentation to Analyse Wheat Variety Performance in Western Australia

Sandra Tanz

Curtin University

On-farm experimentation (OFE) provides an opportunity to evaluate crop performance under commercial farming conditions while capturing the spatial variability inherent in production systems. However, this variability also presents statistical challenges for experimental design and analysis.

This pilot study investigates the integration of spatial data with large-scale OFE to better understand wheat variety performance across five grower-managed sites in Western Australia. Each trial was established within commercial paddocks using strip trial designs aligned with historical production zones. Two core wheat varieties, Scepter (a widely grown commercial cultivar) and Dale (a near-release InterGrain line) were included at all sites, with additional varieties incorporated where locally relevant.

The study focuses on developing and evaluating statistical methods that account for spatial variation by defining pseudo-environments using both equally spaced zones and environmental covariates derived from soil and topographic information. Trial analyses incorporate additional spatial datasets, including elevation and NDVI time series from drone and satellite imagery, to better characterise environmental heterogeneity and genotype-by-environment interactions.

The project aims to develop robust, spatially aware workflows for OFE trial design and analysis, encompassing data cleaning, spatial gridding, pseudo-environment assignment, and linear mixed modelling to account for spatial variation across commercial paddocks. By integrating spatial data, environmental covariates, and modern statistical methods, this work demonstrates the potential of OFE to complement conventional small-plot variety trials, improve variety evaluation, support agronomic decision-making, and contribute to predictive breeding in commercial farming systems.

5. Bivariate analysis of ratio component traits in horticultural breeding programs

Ky Mathews

Queensland Department of Primary Industries

Ratio traits are commonly used in agriculture as selection indices for complex characteristics. For example, flavour characteristics in strawberry are measured as a ratio of total soluble sugars to acidity, the so-called "brix-to-acid ratio". In macadamia, fertility is measured as a ratio of the number of self-pollinated to open-pollinated florets on a raceme. The distributional properties of these ratio traits are often non-normal and heteroscedastic. For breeding decisions, selection on the component traits is also of interest and each has its own genetic variance that is ignored when incorporated into a ratio trait. In this study, we demonstrate a bivariate mixed-model approach for repeated measures, multi-year datasets is possible. Further, ancestral information is included to enable selection for variety release and for parents in future breeding cycles. Compared with direct univariate analysis of the derived ratio, the bivariate approach retains information on the component traits, quantifies their genetic relationship, and provides more interpretable outputs for selection of candidate varieties and future parents.

---

6. Modelling stability of irrigated cotton farming system soil organic carbon content and function

Pauline Ding

University of Wollongong

Long-term changes in stocks and stability within irrigated cotton farming systems in semi-arid Australia is of great public interest. Soil organic carbon underpins nutrient cycling, soil structural stability, and ecosystem resilience, yet soil organic carbon responses to crop rotation and diversification in irrigated Vertosols remain uncertain.

In this project, archived soil samples were collected between 1995 and 2020 from a long-term farming systems experiment at the Australian Cotton Research Institute, Narrabri, NSW. Treatments included continuous cotton and cotton rotated with wheat, vetch, or faba bean under minimum tillage management. The experiment commenced in 1995 in the West Field with five cropping

treatments replicated four times. The following year, 1996, three additional treatments added in adjacent East Field, where the cotton phase occurred 12 months latter than the West Field. Soil samples were collected at 30cm, 60cm and 90 cm depth and analysed using dry combustion and near-infrared spectroscopy. In this study, we use linear mixed model to fit multivariate long-term soil collections and evaluate long-term soil carbon dynamics in intensive agricultural systems. This is a joint work with CISRO Agriculture and Food.

Keynote #6 Fri 11:00 – 11:45

Consulting in the real world: reflections, comments and data

Dr. Ruth Butler

StatsWork 2022 Ltd

I have worked as a statistical consultant for a long time now, working primarily in Agriculture and Horticulture science institutes. Along the way, I have developed some strong opinions about statistics and how statisticians work. I will discuss some of these. I will then present a couple of relatively recent pieces of work, each with a statistical challenge.

Presenter Index

Akter, Mst Bithi 35
Aldiss, Zachary 56

Bebbington, Mark 31
Bell, Kerry 30
Botha, Imke 29
Bowerman, Andrew 32
Brien, Chris 16
Brunner, Stephanie 22
Burrell, Arden 63
Butler, Ruth 66

Cao, Zhanglong 44
Cave, Vanessa 11
Collins, Damian 25
Curran, James 11

Ding, Pauline 66
Dunn, Peter 55
D'Antuono, Mario 12

Feng, Qian 18
Forknall, Clayton 24
Fox, Alise 58
Fradgley, Nick 54

Gamacharige, Navodya Abirami Naotunna 36
Gezan, Salvador 17
Gilmour, Arthur 13
Giri, Khageswor 55
Green, Alexandra 62
Green, Peter 37

Hayes, Ben 49
Hewage, Dilani Kaveendri Dambure 27
Howden, Chris 46
Huang, Martin 33
Huang, Shaoqian 48
Hughes, David 51
Hunt, Colleen 53

Ibenegbu, Amuchechukwu 45

Kenarangi, Tayyebeh 43
Kuhnert, Petra 10
Kurosawa, Takeshi 37

Li, Patrick 15
Li, Zitong 51

Ma, Maggie 14
Mathews, Ky 65
McGillycuddy, Maeve 47
Moreno-Betancur, Margarita 41
Mumford, Michael 34

Nestor, Ben 21
Ni-Jung, Lin 50

Ohno, Yuki 39
O'Connor, Jodine 18

Patrick, Ellis 19
Peitton, Lucas 41
Perera, Binuri 23
Perera, Piumi 20
Powell, Daniel 25

Reid, David 26

Sad Sutopa, Tasmiah 38
Schemann, Kathrin 16
Shankar, Rajan 47
Somchit, Chanatda 61

Tanaka, Emi 28
Tanz, Sandra 64
Tran, Bao Khue 22

Vu, Quan 60

Waters, Ethan Kane 29
Whitehead, Melody 34
Wong-Toi, Adam 57
Woollard, Max 52
Wortmann, Joanne 59

Zaloumis, Sophie 42
Zhaung, Zixiong 40

Time	Monday	Tuesday	Wednesday	Thursday	Friday
8:30					
8:45		Welcome & Housekeeping			
9:00	Workshops	Plenary	Plenary	Plenary	Contributed session 6x15 minutes
9:15					
9:30					
9:45		short break	short break	short break	
10:00		Contributed session 3x10 minutes	Contributed session 3x10 minutes	Contributed session 3x10 minutes	
10:15					
10:30		Morning tea	Morning tea	Morning tea	Morning tea
10:45					
11:00		Contributed session 6x15 minutes	Contributed session 6x15 minutes	Contributed session 6x15 minutes	Plenary and closing
11:15					
11:30					
11:45					Closing ceremony and prizes
12:00					
12:15					
12:30		Lunch	Lunch	Lunch	Lunch
12:45					
13:00				AGUAi AGM	
13:15					
13:30		Plenary	Social Events	Plenary	Conference End
13:45					
14:00					
14:15		short break		short break	
14:30		Contributed session 2x15 minutes		Contributed session 2x15 minutes	
14:45					
15:00		Afternoon Tea		Afternoon Tea	
15:15					
15:30		Contributed session 4x15 minutes		Contributed session 4x15 minutes	
15:45					
16:00					
16:15					
16:30		Poster Session			
16:45					
17:00					
17:15					
17:30	Welcome Reception			Conference Dinner	