

BAYES ON THE BEACH

November 10 - 12, 2014

Surfers Paradise, Gold Coast

botb2014.wordpress.com

Bayes on the Beach is an intentionally small international forum for discussing and exploring developments in Bayesian Statistics and its applications. This three day conference will take place in the world-famous Surfers Paradise on the Gold Coast and is designed to have a diverse format including presentations, contributed sessions, workshops, a poster session and tutorials.

The conference is supported by the Australasian Society for Bayesian Analysis and the Bayesian Statistics Section of the Statistical Society of Australia.

Sponsors and participating organisations



Queensland University of Technology
Brisbane Australia



Statistical Society of Australia Inc.



Conference Program: Monday 10th November 2014

		Chair
9:00 – 10:15	Registration	
10:15 – 10:30	Welcome and Conference Opening	
10:30 – 11:30	Keynote Presentation: Professor Louise Ryan, University of Technology, Sydney <i>Title: Covariate Selection with Applications in Epidemiology</i>	Kerrie Mengersen
11:30 – 13:00	Tutorial 1: Professor Peter Green, University of Technology, Sydney and University of Bristol <i>Title: Bayes Nets: Association and Causation among Discrete Variables</i>	
13:00 – 15:15	Lunch	
15:15 – 15:45	Invited Presentation 1: Chris Jewell, Massey University <i>Title: Bayesian Data Assimilation for Vector Borne Disease Response: Theileria Orientalis (Ikeda) in NZ Cattle</i>	
15:45 – 16:15	Invited Presentation 2: Shakira Suwan, University of Canterbury <i>Title: Empirical Bayes Estimation for the Stochastic Blockmodel</i>	Paul Wu
16:15 – 16:30	Afternoon Tea	
16:30 – 17:00	Invited Presentation 3: Wen-Hsi Yang, CSIRO, Dutton Park <i>Title: Hierarchical Nonlinear Multivariate Time-Frequency Functional Data Models</i>	
17:00 – 18:00	Workshop 1	
18:00 – 20:00	Dinner	
20:00 – 22:30	Poster session 1	

Conference Program: Tuesday 11th November 2014

		Chair
9:00 – 10:00	<p>Keynote Presentation: Professor Judith Rousseau, Université Paris Dauphine</p> <p><i>Title: Asymptotic Properties of Empirical Bayes Procedures – In Parametric and Nonparametric Models</i></p>	Caitriona Ryan
10:00 – 10:50	<p>Tutorial 2: Professor Louise Ryan, University of Technology, Sydney</p> <p><i>Title: Bayesian Strategies for Covariate Selection</i></p>	
10:50 – 11:10	Morning Tea	
11.10 – 12.00	<p>Tutorial 2 (continued): Professor Louise Ryan, University of Technology, Sydney</p> <p><i>Title: Bayesian Strategies for Covariate Selection</i></p>	Nicholas Tierney
12:00 – 12:30	<p>Invited Presentation 4: Benoit Lique, University of Queensland</p> <p><i>Title: R2GUESS: A Graphic Processing Unit-Based R Package for Bayesian Variable Selection Regression of Multivariate Responses</i></p>	
12:30 – 13:00	<p>Invited Presentation 5: Clair Alston, Queensland University of Technology</p> <p><i>Title: Modelling Habitat and Planning Surveillance using Landsat Imagery: A Case Study using Imported Red Fireants</i></p>	
13:00 – 15:15	Lunch	
15:15 – 16:15	<p>Keynote Presentation: Professor Peter Donnelly, University of Oxford and Wellcome Trust Centre for Human Genetics</p> <p><i>Title: Bayesian Approaches in Genomics</i></p>	Aleysha Thomas
16:15 – 16:30	Afternoon Tea	
16:30 – 17:00	<p>Invited Presentation 6: Vasileios Giagos, University of New England</p> <p><i>Title: Inference for Reaction Networks using the Linear Noise Approximation</i></p>	
17:00 – 18:00	Workshop 2	
18:00 – 20:00	Dinner	
20:00 – 22:30	Poster session 2	

Conference Program: Wednesday 12th November 2014

		Chair
9:00 – 10:00	<p>Keynote Presentation: Professor Peter Green, University of Technology, Sydney and University of Bristol</p> <p><i>Title: Discovering Structure in Multivariate Data: How Far can a Bayesian Approach Go?</i></p>	Xing Lee
10:00 – 11:00	<p>Tutorial 3: Professor Judith Rousseau, Université Paris Dauphine</p> <p><i>Title: Statistic and Dynamic Mixture Models from Parametric to Nonparametric</i></p>	
11:00 – 11:30	Morning Tea	
11:30 – 12:00	<p>Invited Presentation 7: Renate Meyer, University of Auckland</p> <p><i>Title: Beyond Whittle - Likelihood Approximations for Bayesian Semiparametric Analysis of Stationary Time Series</i></p>	Benjamin Fitzpatrick
12:00 – 12:30	<p>Invited Presentation 8: Mark Wheldon, Auckland University of Technology</p> <p><i>Title: Bayesian Reconstruction of Two-sex Human Populations</i></p>	
12:30 – 13:30	Lunch	
13:30 – 15:00	<p>Tutorial 4: Professor Peter Donnelly, University of Oxford and Wellcome Trust Centre for Human Genetics</p> <p><i>Title: Bayesian Approaches to Detecting Geographic Population Structure with Genomic Data</i></p>	Kerrie Mengersen
15:00 – 15:15	Conference close	

**Conference Program: Poster Session 1, Monday 10th November
2014, 20:00 – 22:30**

Presenter	Title
Jannah Baker , Queensland University of Technology	<i>Spatial Analysis of Type II Diabetic Prevalence and Associated Risk Factors in Queensland Accounting for Missing Data</i>
Marcela Cespedes , Queensland University of Technology	<i>Modelling Dynamics of Brain Atrophy via Bayesian Mixed Effect Regression in Healthy Aging and Alzheimer's Disease Cohorts</i>
Boris Choy , University of Sydney	<i>A Flexible Class of Bivariate Distributions: Estimation and Applications</i>
Amy Cook , Queensland University of Technology	<i>Analysis of Company Data and Creating Recommendation Algorithms</i>
Susanna Cramb , Queensland University of Technology	<i>Late Cancer Diagnosis: A Multivariate Smoothing Approach</i>
Earl Duncan , Queensland University of Technology	<i>Recommendations for the Inclusion of Socioeconomic Status Variables in Spatio-Temporal Bayesian Hierarchical Models</i>
Matt Edwards , University of Auckland	<i>Bayesian Semiparametric Spectral Density Estimation in Gravitational Wave Data Analysis</i>
Benjamin Fitzpatrick , Queensland University of Technology	<i>Addressing Common Challenges in Spatial Modeling of Environments and Ecologies</i>
Su Yun Kang , Queensland University of Technology	<i>Bayesian Adaptive Sampling: Improving the Effectiveness of Reef Monitoring Programs</i>
Xing Lee , Queensland University of Technology	<i>Approximate Bayesian Model Choice for Historical Influenza Outbreak Using ABC</i>
Jim Lewis , Queensland University of Technology	<i>A Systems Approach to Understanding Impact of Interventions on Network Peak Electricity Demand by Residential Consumers</i>
Samantha Low-Choy , Queensland University of Technology	<i>Validating Bayesian Network Models Using Field Data, for Modelling Potential Distribution of Chilean Needle Grass</i>
Ramethaa Pirathiban , Queensland University of Technology	<i>Effect of Visual Stimuli (GIS vs Photo) on Scenario-Based Elicitation of Probabilities of Site Occupancy</i>
David Roberts , University of Queensland	<i>Probabilistic Language Modelling with Hierarchical Bayesian Nonparametrics</i>

**Conference Program: Poster Session 2, Tuesday 11th November
2014, 20:00 – 22:30**

Presenter	Title
Jessie Roberts , Queensland University of Technology	<i>Communication of Bayesian Uncertainty to Non-Experts Decision Makers through Data Visualisation</i>
Guilherme Rodrigues , University of New South Wales	<i>Functional Regression ABC for Non-Parametric Density Estimation</i>
Thais Rodrigues , University of New South Wales	<i>Noncrossing Bayesian Quantile Regression</i>
Jose Romeo , University of Auckland	<i>The Power Variance Function Copula Model in Bivariate Survival Analysis</i>
Aleysha Thomas , Queensland University of Technology	<i>Mixture Modelling as a Tool for Exposomics</i>
Nicholas Tierney , Queensland University of Technology	<i>Missingness, Mixtures, and Multiple {i}Mputation</i>
Zoé van Havre , Queensland University of Technology	<i>Zmix & Zswitch: Methods for Overfitting Mixture Models for Order Estimation</i>
Julie Vercelloni , Queensland University of Technology	<i>Uncertainties in Non-Linear Population Trajectories: a Bayesian Semi-Parametric Hierarchical Approach to Large-Scale Surveys of Coral Cover</i>
Brenda Vo , Queensland University of Technology	<i>An approximate Bayesian Computation Approach for Estimating Parameters Governing the Expansion of Melanoma Cell Colonies in Vitro</i>
Nicole White , Queensland University of Technology	<i>A Spatial Poisson Hurdle Model with Clustered Random Effects to Understand the Provision of Health Services</i>
Richard Woodman , Flinders University	<i>A Comparison of Maximum Likelihood and Bayesian Estimation Approaches to Examine the Effects of Homocysteine and Other Coronary Artery Disease (CAD) Risk Factors on CAD Using Structural Equation Modelling</i>
Paul Wu , Queensland University of Technology	<i>Towards a Bayesian Network Framework for Predicting and Understanding Cumulative Effects in a Complex System</i>
Liang-Cheng Zhang , Griffith University	<i>Estimating Scale and Scope Economies in Australian Universities Under a Bayesian Framework</i>
Xin Zhang , University of New South Wales	<i>Approximation Bayesian Computation Methods Using Composite Likelihood</i>
Wanchuang Zhu , University of New South Wales	<i>A Novel Approach for Markov Random Field with Intractable Normalizing Constant on Large Lattice</i>

Keynote Speaker: Professor Judith Rousseau, Université Paris Dauphine

Research interests:

Bayesian statistics | Parametric and nonparametric | Mixture distributions | MCMC algorithms

Judith Rousseau is currently Professor at University Paris Dauphine. Her work ranges from theoretical aspects of Bayesian procedures, both parametric and nonparametric, to methodological developments, in particular on MCMC or related algorithms or on the elicitation of subjective priors. She is an associate editor of the annals of statistics, Bernoulli and ANZJS and the program secretary of IMS. She is an ISBA fellow.

Keynote presentation

Asymptotic Properties of Empirical Bayes Procedures – In Parametric and Nonparametric Models

Authors: **Judith Rousseau**, Sonia Petrone, C. Scricciolo, V. Rivoirard and S. Donnet

Affiliations: CREST-ENSAE & Université Paris Dauphine & Bocconi University

In this work we investigate frequentist properties of Empirical Bayes procedures. Empirical Bayes procedures are very much used in practice in more or less formalized ways as it is common practice to replace some hyperparameter in the prior by some data dependent quantity. There are typically two ways of constructing these data dependent quantities: using some kind of moment estimator or some quantity whose behaviour is well understood or using a maximum marginal likelihood estimator. In this work we first give some general results on how to determine posterior concentration rates under the former setting, which we apply in particular to two types of Dirichlet process mixtures. We then shall discuss more parametric models in the context of maximum marginal likelihood estimation. We will in particular explain why some pathological behaviour can be expected in this case. We will finally discuss some recent advances on maximum marginal likelihood empirical Bayes approaches in nonparametric setting. The latter is an ongoing work with B. Szabo.

*Keynote Speaker: Professor Louise Ryan, University of
Technology, Sydney*

Research interests:

Applied statistics, especially for applications in medicine and health (biostatistics) | Survival Analysis | Correlated Data Analysis, including clustered data, multiple outcomes and longitudinal | Quantitative Risk Assessment | Meta-Analysis

Professor Louise Ryan has had a sensational career as Henry Pickering Walcott Professor and Chair of the Department of Biostatistics at Harvard, Chief at CSIRO's Division of Mathematics, Informatics and Statistics, and now as Distinguished Professor of Statistics at UTS. Professor Ryan is a member of the Australian Academy of Science and Editor in Chief of Statistics in Medicine. She is well known for her methodological contributions to statistical methods for cancer and environmental health research.

Keynote Speaker: Professor Peter Green, University of Technology, Sydney and University of Bristol

Research interests:

Bayesian inference in complex stochastic systems | Markov chain Monte Carlo methodology | Forensic genetics | Bayesian nonparametrics | Graphical models

Professor Peter Green is a Distinguished Professor at the University of Technology, Sydney, and an Emeritus Professor and Professorial Research Fellow at the University of Bristol; he divides his time between England and Australia. Most of his research has been in computational Bayesian statistics, with interests in inference in complex stochastic systems, asymptotic approximations to posterior distributions, Markov chain Monte Carlo methodology, graphical models, forensic genetics, and Bayesian nonparametrics. Professor Green is also currently Editor of Statistical Science.

Keynote presentation

Discovering Structure in Multivariate Data: How Far can a Bayesian Approach Go?

The structure in a multivariate distribution is largely captured by the conditional independence relationships that hold among the variables, often represented graphically, and inferring these from data is an important step in understanding a complex stochastic system. We would like to make simultaneous inference about the conditional independence graph and parameters of the model; this is known as joint structural and quantitative learning in the machine learning literature. The Bayesian paradigm allows a principled approach to this simultaneous inference task.

There are tremendous computational and interpretational advantages in assuming the conditional independence graph is decomposable, and not too many disadvantages. I will present a new structural Markov property for decomposable graphs, show its consequences for prior modelling, and discuss a new MCMC algorithm for sampling graphs that enables Bayesian structural and quantitative learning on a much bigger scale than previously possible.

This is joint work with Alun Thomas (Utah).

*Keynote Speaker: Professor Peter Donnelly, University of Oxford
and Wellcome Trust Centre for Human Genetics*

Research interests:

Stochastic modelling | Applications of probability and statistics in genetics | Gene mapping, early human evolution, population genetics | Measure-valued diffusions, statistical issues in DNA profiling

Professor Peter Donnelly is the director of the Wellcome Trust Centre for Human Genetics and a Professor of Statistical Science at the University of Oxford. The Australian born, Oxford based mathematician is a Fellow of both the Royal Society and the Academy of Medical Sciences, TED speaker as well as a former Rhodes Scholar. Professor Donnelly is also one of the key members of the International HapMap project. His recent research focuses on genetics of common human diseases and on the development of statistical methods for analysing genetic data.

Bayesian Data Assimilation for Vector Borne Disease Response: *Theileria orientalis* (Ikeda) in NZ Cattle

Presenter: **Chris Jewell**

E-mail: C.P.Jewell@massey.ac.nz

Author: **Chris Jewell**

Affiliation: Massey University

Dynamical models of communicable diseases have become a prominent feature of national-level epidemic response. Developments in Bayesian inference have enabled these models to provide quantitative risk predictions in a real-time setting, learning from spatiotemporal data as it arrives from the field. However, these models rely heavily on accurate covariate data from which to make inference. Incursions of vector borne disease present a particular challenge in this respect, as exemplified by the recent introduction of *Theileria orientalis* (Ikeda), an obligate tick-borne disease of cattle, into New Zealand. Whereas the location of cattle and the animal movement network between farms is well recorded, little is known about the national scale ecology of the tick vector. This talk will present a Bayesian data assimilation approach to this problem, in which vector presence is modelled as a discrete-space latent process with a continuous-time seasonality. A joint likelihood function assimilates the epidemic data and results from a national disease surveillance programme designed for a different disease. A spatiotemporally inhomogeneous Poisson process is used to model the epidemic, with an a priori independent hierarchical binomial surveillance model. This joint model is fitted to observed case detection data using a non-centered trans-dimensional MCMC algorithm, integrating over the marginal posterior of the latent vector surface, censored herd infection times, and the presence of undetected infections. Importantly, the algorithm is implemented using GPGPU technology which accelerates within-chain likelihood calculations to an overnight timeframe. Finally, the predictive distribution is provided as a real time disease forecast for decision support purposes.

Empirical Bayes Estimation for the Stochastic Blockmodel

Presenter: **Shakira Suwan**, University of Canterbury

E-mail: shakira.suwan@pg.canterbury.ac.nz

Authors:

Shakira Suwan, University of Canterbury

Dominic S. Lee, University of Canterbury

Runze Tang, Johns Hopkins University

Daniel L. Sussman, Johns Hopkins University

Minh Tang, Johns Hopkins University

Carey E. Priebe, Johns Hopkins University

Network models are increasingly used to model datasets that involve interacting units, especially random graph models where the vertices represent individual actors and the edges represent the presence or absence of a specified connection between actors. In particular, inference for the stochastic blockmodel is currently of burgeoning interest in the statistical community, as well as in various application domains as diverse as social networks, citation networks, brain connectivity networks (connectomics), etc. Recent theoretical developments have shown that spectral embedding of graphs yields tractable distributional results; in particular, a random dot product latent position graph formulation of the stochastic blockmodel provides a mixture of multivariate Gaussian distributions for the adjacency spectral embedding. Here, we employ this new theory to provide an empirical Bayes methodology for estimating block memberships of vertices in a random graph generated from the stochastic blockmodel, and demonstrate its practical utility. The posterior inference is conducted using a Metropolis-within-Gibbs algorithm. The theory and methods are illustrated through Monte Carlo simulation studies, and experimental results on a Wikipedia data set are presented.

Hierarchical Nonlinear Multivariate Time-Frequency Functional Data Models

Presenter: **Wen-Hsi Yang**, CSIRO, Dutton Park

E-mail: Wen-Hsi.Yang@csiro.au

Authors:

Wen-Hsi Yang, CSIRO, Dutton Park

Christopher K. Wikle, University of Missouri

Scott H. Holan, University of Missouri

Mark L. Wildhaber, US Geological Survey, Columbia, MO

Time-frequency analysis has become a fundamental component of many scientific inquiries. In this domain, crucial aspects of the underlying process of interest often become apparent that would otherwise go undetected. Consequently, time-frequency representations can often serve as powerful predictors in modeling complex processes. In order to facilitate the use of these representations within a statistical modeling framework, we propose a class of flexible Bayesian hierarchical time-frequency functional data models. Importantly, our approach extracts features of the original signal that improve prediction rather than explaining variation. Finally, we illustrate the effectiveness of our approach by applying our model to an example of shovelnose sturgeon spawning success in the Lower Missouri River.

R2GUESS: A Graphic Processing Unit-Based R Package for Bayesian Variable Selection Regression of Multivariate Responses

Presenter: **Benoit Liquet**, University of Queensland

E-mail: b.liquet@uq.edu.au

Authors: **Benoit Liquet**, Gianluca Campanella, Leonardo Bottolo, Sylvia Richardson and Marc Chadeau-Hyam

Recent advances in high throughput ‘omics’ technologies have given rise to a wealth of novel data high dimensional data (ranging from thousands to hundreds of thousand variables) each demonstrating complex correlation structures. These data comprise genetic, epigenetic and transcriptomic profile which have show a great potential in measuring the abundance of biologically relevant molecules over the whole biological system. The analysis of such complex data raises strong statistical challenges relating to the fact that the number of predictors exceeds the number of observations (“large p , small n ” scenario).

Alongside multiple testing correction strategies, variable selection approaches can handle this situation and we propose here a Bayesian implementation of this kind of approach. As such it will seek for the best combination of covariates to predict the (possibly multivariate) outcome. The Bayesian framework it is based on allows the construction of parsimonious regression models adopting prior specifications that translate expected sparsity of the underlying biology and therefore facilitates the interpretation of the results. R2GUESS is an R package interfacing a C++ implementation of a fully Bayesian Variable Selection approach for multivariate linear regression. Using latest computational advancement it is GPU-enabled, and in its current form it enables the analysis of hundreds of thousands of predictors measured in thousands of individuals simultaneously. The efficient exploration of the 2^n dimensional space is enabled by the use of an Evolutionary Monte Carlo sampling scheme comprising a large portfolio of local and global moves. R2GUESS also provides refined numerical and graphical output easing post-processing and subsequent interpretation of the extensive results produced by the GUESS algorithm. Performances of the model and interpretability of its results are illustrated based on examples from several omics platforms.

Modelling Habitat and Planning Surveillance using Landsat Imagery: A Case Study using Imported Red Fireants

Presenter: **Clair L. Alston**

E-mail: c.alston@qut.edu.au

Authors: **Clair L. Alston**¹, Kerrie L. Mengersen¹, Christopher M. Strickland²

Affiliations: ¹Mathematical Sciences School, Queensland University of Technology;

²Australian School of Business, University of New South Wales

The accidental import of fireants into Australia, occurred via the Port of Brisbane in 2001. An eradication program was swiftly implemented, with details of locations of finds being collected methodically over the next decade. To aid in the ongoing surveillance for early detection of these exotic pests, the use of Landsat imagery is an affordable and valuable tool. In this paper we use Classification and Regression Trees (CART) to determine appropriate indices that relate to fireant habitation. Using Classification and Regression Trees (CART) in a resampling environment, we were able to subsample from non-infested areas and determined that visible red, mid infrared, soil brightness index and lower mid infrared were the variables most able to aid in distinguishing between the characteristics of infested and non-infested (or undiscovered) sites. These variables were then used, in various combinations, in a soft clustering technique known as Bayesian mixture models to assess the probability of each pixel as being in habitable terrain either in a multivariate analysis or as a univariate spatial temporal model.

Inference for reaction networks using the Linear Noise Approximation

Presenter: **Vasileios Giagos**

E-mail: vgiagos@une.edu.au

Authors: **Vasileios Giagos**

Affiliation: University of New England

A common research theme across many disciplines, e.g. systems biology, epidemiology, population ecology, is the study of evolving systems of interacting species. These systems are usually observed at discrete times and quite often the observations are partial and subject to measurement error. In this setting, many key research questions can be linked with the kinetic parameters of the system which control how often these interactions occur and it is of great interest to develop appropriate inferential tools.

The dynamical behaviour of a typical system can be expressed, in great detail, as a state-dependent Markov Jump process. Unfortunately, the use of Markov Jump process induces a computational cost which is often unrealistic and approximating models are used instead. The most common approximations are based on Ordinary as well as Stochastic Differential Equations (ODEs, SDEs). The former are very fast to evaluate but ignore the intrinsic stochasticity whereas the latter are harder to implement since the associated likelihood function is generally unknown.

We present a Bayesian methodology [1] based on the Linear Noise Approximation (LNA), which can be viewed as a compromise between ODEs and SDEs approximations: it is stochastic and the evaluation of its likelihood function requires the computational cost of an ODEs-based model. We illustrate our methodology with examples from systems biology, epidemiology and population ecology.

References

[1] Fearnhead, P., Giagos, V. and Sherlock, C. Inference for Reaction Networks Using the Linear Noise Approximation. *Biometrics* 70, no. 2 (1 June 2014): 457–66.

Beyond Whittle - Likelihood Approximations for Bayesian Semiparametric Analysis of Stationary Time Series

Presenter: **Renate Meyer**, University of Auckland

E-mail: renate.meyer@auckland.ac.nz

Authors:

Renate Meyer, University of Auckland, New Zealand

Claudia Kirch, Karlsruhe Institute of Technology, Germany

Standard time series modelling is dominated by *parametric* models like ARMA, GARCH or stochastic volatility models. In particular, *Bayesian* time series analysis (Steel, 2008) is inherently parametric in that a completely specified likelihood function is needed. Even though *nonparametric Bayesian* inference has been a rapidly growing topic over the last decade, only very few nonparametric Bayesian approaches to time series analysis have been developed. Most notably, Carter and Kohn (1997), Gangopadhyay (1998), Choudhuri et al. (2004), and Hermansen (2008) used Whittle's likelihood for Bayesian modeling of the spectral density as the main nonparametric characteristic of stationary time series. On the other hand, frequentist time series analyses are often based on *nonparametric* techniques encompassing a multitude of bootstrap methods (Kreiss and Lahiri, 2011, Kirch and Politis, 2011).

Whittle's likelihood is an approximation of the true likelihood. As shown in Contreras-Cristan et al. (2006), the loss of efficiency of the nonparametric approach using Whittle's likelihood can be substantial. On the other hand, parametric methods are more powerful than nonparametric methods if the observed time series is close to the considered model class but fail if the model is misspecified. Therefore, we suggest a nonparametric correction of a parametric likelihood that takes advantage of the efficiency of parametric models while mitigating sensitivities through a nonparametric amendment. We use a nonparametric Bernstein polynomial prior on the spectral density with weights induced by a Dirichlet process. We show contiguity for Gaussian stationary time series, implement Bayesian posterior computations via a MH-within-Gibbs sampler, and illustrate performance in a simulation study.

References

- C.K. Carter and R. Kohn, Semiparametric Bayesian Inference for Time Series with Mixed Spectra, JRSSB 59, 255-268 (1997).
- N. Choudhuri, S. Ghosal and A. Roy, Bayesian Estimation of the Spectral Density of a Time Series, JASA 99, 1050-1059 (2004).
- A. Contreras-Cristan, E. Gutierrez-Pena and S.G. Walker, A Note on Whittle's Likelihood, Communications in Statistics - Simulation and Computation 35, 857-875 (2006).
- A.K. Gangopadhyay, B.K. Mallick and D.G.T. Dennison, Estimation of Spectral Density of a Stationary Time Series via an Asymptotic Representation of the Periodogram, Journal of Statistical Planning and Inference 75, 281-290 (1998).
- G.H. Hermansen, Bayesian Nonparametric Modelling of Covariance Functions, with Applications to Time Series and Spatial statistics, MSc Thesis, University of Oslo, (2008).
- N.L. Hjort, C. Holmes, P. Muller and S.G. Walker, Bayesian Nonparametrics, Cambridge University Press, Cambridge (2010).
- C. Kirch, D.N. Politis. TFT-Bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. Annals of Statistics 39, 1335-1826 (2011).
- Kreiss, J.-P. and Lahiri, S.N. (2011): Bootstrap Methods for Time Series. In: Handbook of Statistics, Vol 30: Time Series-Methods and Applications, Elsevier-North Holland, 3-26.
- M.F.J. Steel, Bayesian Time Series Analysis, In: The New Palgrave dictionary of economics. Palgrave Macmillan, Basingstoke, Hampshire, New York (2008).

Bayesian Reconstruction of Two-sex Human Populations

Presenter: **Mark C. Wheldon**

E-mail: mwheldon@aut.ac.nz

Authors: **Mark C. Wheldon**, Adrian E. Raftery, Samuel J. Clark, and Patrick Gerland

Affiliations: Auckland University of Technology and Centre for Clinical Research and Effective Practice, Auckland

Bayesian reconstruction is a method for estimating age-specific fertility, mortality, migration and population counts of the recent past with uncertainty. Bayesian reconstruction simultaneously estimates age-specific population counts, fertility rates, mortality rates and net international migration flows from fragmentary data while formally accounting for measurement error. As inputs, Bayesian reconstruction takes initial bias-reduced estimates of age-specific population counts, fertility rates, survival proportions and net international migration. It was developed in collaboration with the United Nations Population Division (UNPD) and is compatible with the processes used to create their biennial World Population Prospects. The original version of Bayesian reconstruction produced estimates for female-only populations. Here we show how two-sex populations can be similarly reconstructed and probabilistic estimates of various sex ratio quantities obtained. We demonstrate the method by reconstructing the populations of India from 1971 to 2001, Thailand from 1960 to 2000, and Laos from 1985 to 2005. We produce probabilistic estimates of sex ratios at birth (SRBs) and sex ratios of mortality and use them to answer some questions about trends in these parameters over time. For example, we found evidence that in India, SRB exceeded its conventional upper limit of 1.06, and increased over the period of study, with posterior probability above 0.9.

Spatial analysis of type II diabetic prevalence and associated risk factors in Queensland accounting for missing data

Presenter: **Jannah Baker**

E-mail: jannah.baker@qut.edu.au

Authors: **Jannah Baker**, Nicole White, Kerrie Mengersen

Affiliation: Queensland University of Technology

Background: Type 2 diabetes mellitus (DM II) is on the rise worldwide, with increasing evidence of geographic inequalities in prevalence. Environmental factors are understood to be an important predictor of DM II but their effects on DM II prevalence geographically are not well understood in many parts of the world.

Methods: We fit Binomial and Poisson generalised linear mixed models (GLMMs) with spatially correlated random effects to data on DM II prevalence across Queensland local government areas, accounting for lifestyle-related risk factors. Lifestyle factors included socioeconomic status, age, overweight/obesity, smoking, insufficient physical activity, fruit and vegetable consumption. Three imputation methods were trialled to account for missing data in covariate information. Sensitivity analyses were conducted to assess robustness of results to choice of priors.

Results: Estimated relative risk of DM II prevalence ranged from 0.48 (Isaac Regional) to 3.073 (Cherbourg Aboriginal Shire) from the spatially smoothed Poisson models. Socioeconomic status was found to be significantly associated with DM II prevalence. Mean imputation was selected as the most accurate method for estimation of missing covariates in this study.

Conclusions: With rises in DM II prevalence worldwide, there is a need to understand geographic inequalities to inform the allocation of related health services. Through spatial modelling to identify high-risk regions of Queensland for DM II prevalence and important geographic and demographic risk factors, we are informed on the allocation of additional resources to reduce the future burden of DM II to our community.

Modelling dynamics of brain atrophy via Bayesian mixed effect regression in healthy aging and Alzheimer's disease cohorts

Presenter: **Marcela Cespedes**

E-mail: marcela.cespedes@student.qut.edu.au

Authors: **Marcela Cespedes**^{1,2}, James McGree¹, Chris Drovandi¹, Kerrie Mengersen¹, James Doecke² and Jurgen Fripp²

Affiliations: ¹Mathematical Sciences School, Queensland University of Technology,
²CSIRO Digital Productivity and Services/Australian E-Health Research Centre

The most common form of dementia in the world is Alzheimer's disease (AD). By age 60 over 5% of the population will be affected by AD and this prevalence doubles every 5-10 years. AD is thought to begin decades in advance prior to symptoms being apparent and the single biggest risk of AD is aging. Numerous neurological studies have captured the rate of deterioration of healthy aging brains and those on the AD pathway, however to date they do so independently.

The aim of this research is to integrate the two and model the trajectory over time of brain atrophy in both healthy and AD patients for the potential use in early detection, in particular, before the onset of dementia symptoms. Such research will therefore be important in intervention strategies to delay the onset of AD.

Data available on potential predictors of brain atrophy include biomarkers such as: neuroimaging, genetic, biochemistry as well as psychological results. The exploration and derivation of new models in the Bayesian framework will incorporate experimental data, as well as priori information from the literature and connectivity networks from diffusion MRI. We hypothesise these models will be more informative and capture several dynamics than current linear mixed models available for longitudinal trajectory of individual brain regions.

A flexible class of bivariate distributions: estimation and applications

Presenter: **Boris Choy**

E-mail: boris.choy@sydney.edu.au

Author: **Boris Choy**

Affiliation: The University of Sydney

A new class of symmetric bivariate distributions is created by mixing random variables. The resulting distributions are heavy-tailed with flexible marginal distributions and have no tail dependence. Other properties will also be explored. With a simple modification, it is extended to a class of asymmetric distributions with tail dependence. Bayesian and non-Bayesian estimation methods are discussed and a real application is presented.

Analysis of Company Data and Creating Recommendation Algorithms

Presenter: **Amy Cook**

E-mail: a20.cook@connect.qut.edu.au

Authors: **Amy Cook**, Paul Wu, Kerrie Mengersen

Affiliation: Queensland University of Technology

An analysis of the purchasing data from a company that sells construction products internationally to other businesses will be used to create an online recommendation system to promote additional purchases. Online recommendation systems have been utilised with great success by website such as Amazon and Netflix and now a recommendation system for business products is in demand. For this project, one million purchase records are available in an excel file format which provides the business insights and validation method for creating the recommendation system. Each row represents a purchase, and lists the item purchased, the product category, net sale amount, location, company ID, data, and the number of employees in the customer company. To gain a basic understanding of the data, an initial round of simple statistical techniques were run. These revealed facts such as: 90% of purchases were within \$952 - \$1290 (with a maximum \$57,200 purchase) and that 50% of purchases were tools as opposed to consumables and accessories. More complex statistical methods such as K means clustering will give further insight into purchasing trends and direct which kind of recommendation system should be employed.

The two most popular and simple recommendation systems are content based and collaborative filtering. Collaborative filtering is the most likely to succeed as it recommends products based on another similar customer. This will pick up the cultural product preferences in building industries in different places around the world. Further considerations that will be built into the system will be appealing to the purchase motivations of a business which are centred around project budgets and any cost, time, or quality advantages a product will provide.

Late cancer diagnosis: a multivariate smoothing approach

Presenter: **Susanna M Cramb**

E-mail: susannacramb@cancerqld.org.au

Authors: **Susanna M Cramb**^{1,2}, Peter D Baade^{1,3,4}, Kerrie L Mengersen²

Affiliations: ¹Cancer Council Queensland, ²School of Mathematical Sciences, Queensland University of Technology, ³School of Public Health and Social Work, Queensland University of Technology, ⁴Griffith Health Institute, Griffith University.

The late diagnosis of cancer increases morbidity and mortality. Cancer screening should enable early detection of cancer, but participation varies between areas. We sought to determine what areas had increased risks of having late cancer diagnosis, and if there was correlation between different types of cancers diagnosed either at early or late stage.

Queensland Cancer Registry data was obtained for females diagnosed between 1 January 1997 and 31 December 2008 with breast cancer, colorectal cancer and/or melanoma. These are the leading cancers diagnosed among females, and have screening procedures available. Multivariate Bayesian hierarchical spatial models were used to jointly model cancers by stage and explore correlation across 478 statistical local areas. A global clustering test (Tango's MEET) was used to determine if the observed variation was significant.

There was significant variation throughout Queensland in the diagnosis of early stage melanoma, breast and colorectal cancers, and also advanced breast cancer. Some remote areas had lower risk of diagnosis for both early and advanced stage cancers in comparison to the Queensland average, but higher proportions of advanced cancers.

Greater access to and participation in screening in remote areas should decrease the high proportion of advanced cancers diagnosed in these areas, particularly for breast cancer. Selected urban areas may also benefit from improving screening participation.

Spatial analyses such as these enable areas with a higher burden of advanced cancer diagnoses to be identified, and could be used to target messages or services in specific areas

Recommendations for the Inclusion of Socioeconomic Status Variables in Spatio-Temporal Bayesian Hierarchical Models

Presenter: **Earl Duncan**

E-mail: earl.duncan@qut.edu.au

Authors: **Earl Duncan**, Nicole White, Kerrie Mengersen

Affiliation: Queensland University of Technology

The inclusion of variables that measure the socioeconomic status of sampling units is often considered to be important. For complex hierarchical models, however, including such variables may not be straightforward and may produce unexpected results. The BaySTDetect model (Li et al. 2012), a Bayesian hierarchical model that aims to detect trends which differ from the common temporal trend, is used to illustrate the difficulties in including a socioeconomic variable – the index of relative socio-economic advantage and disadvantage (IRSAD), one of four socio-economic indexes for areas (SEIFA) developed by the Australian Bureau of Statistics (ABS 2008). This example addresses two key questions: *how* to include IRSAD (i.e. using numerical scores, ordinal ranks, or coding the ranks into quantiles), and *where* to include IRSAD in the model (e.g. in the linear predictor components, through the model indicator parameter, or via some hyper-parameter). Results indicate that the using quantiles may be unnecessary and unhelpful, despite recommendations (ABS 2013) to do so, and demonstrate the need to consider alternative model modifications to include IRSAD in order to achieve convergence and sensible parameter estimates.

References

- Australian Bureau of Statistics (ABS). 2008. *Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 'SEIFA, Statistical Local Areas, Data Cube only, 2006'*, data cube: Excel spreadsheet, cat. no. 2033.0.55.001, www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012006
- Australian Bureau of Statistics (ABS). 2013. *Socio-Economic Indexes for Areas (SEIFA) – Technical Paper, 2011*, cat. no. 2033.0.55.001. Accessed 3 August 2014, www.abs.gov.au/ausstats/abs@.nsf/DetailsPage/2033.0.55.0012011
- Li, G., N. Best, A. Hansell, I. Ahmed and S. Richardson. 2012. BaySTDetect: detecting unusual temporal patterns in small area data via Bayesian model choice. *Biostatistics* **13** (4): 695-710.

Bayesian semiparametric spectral density estimation in gravitational wave data analysis

Presenter: **Matt Edwards**

E-mail: matt.edwards@auckland.ac.nz

Authors: **Matt Edwards**^{1,2}, Renate Meyer¹, Nelson Christensen²

Affiliations: ¹ University of Auckland, Auckland, New Zealand

² Carleton College, Northfield, Minnesota, USA

Astronomy is entering a new and exciting era, with the second generation of ground-based gravitational wave (GW) interferometers (Advanced LIGO, Advanced Virgo, and KAGRA) expected to come on-line in the next few years. Extracting astrophysical information encoded in GW signals is of paramount importance to researchers in the field, and since observations are extremely noisy, accurate predictions rely on the correct characterisation of detector noise. The default model in the GW data analysis literature assumes detector noise is stationary and Gaussian distributed, with a known power spectral density (PSD) that is usually estimated using off-source data. Real GW data often experience departures from such assumptions, and so we propose a Bayesian semiparametric approach to improve this. Using a nonparametric Bernstein polynomial prior on the PSD with weights attained via a Dirichlet process distribution, we update the approximate Whittle likelihood, and sample from the posterior of the PSD using a Metropolis-within-Gibbs sampler. We also simultaneously estimate the reconstruction parameters of a core collapse supernova GW signal that has been embedded in simulated Advanced LIGO noise.

Addressing Common Challenges in Spatial Modelling of Environments and Ecologies

Presenter: **Benjamin R. Fitzpatrick**

E-mail: b1.fitzpatrick@qut.edu.au

Authors: **Benjamin R. Fitzpatrick**, Kerrie Mengersen

Affiliation: Queensland University of Technology

The increasing availability of environmental or ecological survey data with spatial positioning information accompanying each observation facilitates the combination of such data from distinct sampling programs conducted over the same region. This leads to some statistical challenges being present across a broad range of ecological and environmental spatial modelling endeavours. Such challenges include: a) spatial misalignment of (i) observations of different variables and, (ii) observations and the locations (or coverage extents and resolutions) at which inference is desired and b) the availability of large numbers of potentially relevant covariates coupled with the belief that selected model(s) should be sparse. These general challenges are addressed via a pedological modelling case study. The response variable in the case study is the amount of organic carbon present in soil core samples and as such it is a geostatistical variable. Potential covariates examined include: terrain and hydrological metrics calculated from a digital elevation model, vegetation indices calculated from the reflectance of top of canopy active illumination of pasture, foliar projective cover layers derived from satellite imagery and radiometric and electromagnetic imaging layers from an airborne survey. The objective of the case study is to use these covariates to aid in interpolating the soil organic carbon measurements from the soil cores to a full cover predictive map of soil organic carbon levels across the study area.

Bayesian adaptive sampling: Improving the effectiveness of reef monitoring programs

Presenter: **Su Yun Kang**

E-mail: s7.kang@qut.edu.au

Authors: **Su Yun Kang**¹, James McGree¹, Julian Caley² and Kerrie Mengersen¹

Affiliations: ¹Queensland University of Technology, ²Australian Institute of Marine Science

Monitoring programs are essential for understanding patterns, trends and threats in coral reef systems. However, they are costly in terms of dollars, human resources and technology, and they are complex in terms of balancing short and long term requirements. The Australian Institute of Marine Science (AIMS) has recently become interested in developing new statistical methods for designing cost-effective adaptive sampling and monitoring schemes that can better utilise existing information and resources, and which can harness or incorporate available prior information. The aim of this study is to develop potential model-based adaptive monitoring methods for spatio-temporal coral reef data. Different objectives or utilities are used to compare costs and benefits of monitoring using the current approaches as opposed to a Bayesian adaptive sampling approach. Case studies are used to evaluate the efficiency of non-adaptive and adaptive sampling designs with respect to each objective, using different utility or loss functions. The results of the case studies can be used to suggest opportunities to AIMS to help achieve a more informative or cost-effective monitoring scheme.

Approximate Bayesian model choice for historical influenza outbreak using ABC

Presenter: **Xing Lee**

E-mail: xj.lee@qut.edu.au

Authors: **Xing Lee**^{1,2,3}, Chris Drovandi¹ and Tony Pettitt^{1,2}

Affiliation: ¹Mathematical Sciences School, Queensland University of Technology, ²IHBI, Queensland University of Technology, ³Centre of Research Excellence in Reducing Healthcare Associated Infections (CRE-RHAI)

Approximate Bayesian computation (ABC) algorithms were developed to address statistical inferential problems where the likelihood of the model is intractable and simulation from it is straightforward. ABC algorithms accept parameter values which generate simulated data sets similar to the observed data as draws from the desired posterior distribution. The similarity is measured through a discrepancy function, typically the difference between a set of summary statistics.

We have developed an ABC algorithm which addresses model choice problems with more than two competing models by extending the SMC ABC algorithm of Drovandi and Pettitt, 2011, and the regression approach of Fearnhead and Prangle, 2012, where parameter posterior means estimated using stepwise linear regression coefficients, are used as summary statistics in their ABC algorithm.

An additional stepwise multinomial logistic regression is performed on the model indicator variable and the estimated model probabilities are incorporated into the set of summary statistics for model choice purposes. A reversible jump Markov chain Monte Carlo step is included in the SMC ABC move step to increase model diversity and provide a thorough exploration of the model space.

The algorithm was validated through a simple example where it was straightforward to compute the posterior model probabilities without the use of ABC. The application considered here is an historical influenza outbreak on Tristan da Cunha where four common models describing different transmission mechanisms are compared to determine the model which best describes the observed data.

The algorithm described here has shown promise in investigating model choice problems with intractable likelihoods.

References:

1. Drovandi, C. C., & Pettitt, A. N. (2011). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics*, 67(1), 225-233.
2. Fearnhead, P., & Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3), 419-474.

A systems approach to understanding impact of interventions on network peak electricity demand by residential consumers

Presenter: **Jim Lewis**

E-mail: james.lewis@connect.qut.edu.au

Authors: **Jim Lewis**, Kerrie Mengersen, Laurie Buys, Desley Vine, John Bell, Peter Morris and Gerard Ledwich

Affiliation: Queensland University of Technology

Peak demand for electricity is increase while total demand for electricity is either increasing more slowly or is even decreasing. Provision of additional network infrastructure to meet the increase in network peak demand for electricity is increasing costs. Addressing this demand is a major imperative for Australian electricity agencies. There is a need by energy utilities and governments to understand the key factors impacting residential consumers' peak demand for electricity when designing policies and other interventions to help reduce the level of peak demand in relation to average demand.

The complex relationships in the system of demand for electricity by households combine social and technical factors. The wide range of social factors that influence peak demand include household demographics, culture, trust and knowledge. These were combined in a systems model with technical factors such as the customers' location, housing construction and appliances. The systems model that was developed was transformed into a statistical Bayesian network and quantified using diverse types and sources of available information.

The Bayesian network was used to provide insights into the major influential factors and the potential impact of changes to these factors on peak demand. In particular, the model was used to investigate the impact of different market-based and government interventions in various customer locations of interest. The use of this systems approach using Bayesian networks to assist the management of peak demand indifferent modelled locations in Queensland, Australia provided insights about the most important elements in the system and the intervention strategies that could be tailored to the targeted customer segments.

Validating Bayesian Network Models using Field data, for modelling potential distribution of Chilean Needle Grass

Presenter: **Samantha Low-Choy**, Queensland University of Technology

E-mail: s.lowchoy@qut.edu.au

Authors: **Samantha Low-Choy** (QUT), Justine Murray and Rieks van Klinken (CSIRO), Ramethaa Pirathiban (QUT)

Pasture weeds such as Chilean Needle Grass pose substantial biosecurity risk, with implications for control and eradication in some areas should the weed become established and uncontrollable. In consultation with experts, a species distribution model across a region was constructed as a Bayesian Belief Network. It focussed on both susceptibility to establishment of this weed, as well as potential spread. Later, field data became available to help validate the model. Unfortunately due to flooding in the area, the field data did not provide comprehensive coverage of the design space of the model. We discuss how the constrained field data could be used to validate portions of the model. This included an investigation of how part of the BBN could be reformulated as a non-normal regression model with variable selection, using WinBUGS as the computational engine.

Effect of visual stimuli (GIS vs photo) on scenario-based elicitation of probabilities of site occupancy

Presenter: **Ramethaa Pirathiban**

E-mail: r.jeyapalan@qut.edu.au

Authors: **Ramethaa Pirathiban**, Low-Choy. S, Williams. C, and Murray. J

Affiliation: Queensland University of Technology

Heuristics and biases play a major role in designing expert elicitations. The effect they have on the process of elicitations can alter the results. When framing elicitations, we should be equally concerned with not only what we ask experts to assess, but also how we ask it. There is considerable amount of research done in the field of psychology on experimenting with elicitations. Different techniques have been applied and tested to increase the performance of the resulting elicited values. The reason for using between-subjects design in elicitations is to avoid prompting the subjects by juxtaposing the items of interest. However, in reality it confounds the results as different items can provoke different problem interpretations. This study further this idea on between-subject designs and examine its role on elicited values by looking at experimenting with scenario based elicitation for a habitat suitability modelling, accompanying *Elicitor*, a software tool, which assists in quantifying expert knowledge in a form suitable for use as a prior model in Bayesian regression. The aim of this study is to test the effect of the way data is presented to the experts on the elicited values, specifically the use of GIS maps vs photographic images as elicitation “prompts”. Furthermore, this study will emphasise the need for a good framing of elicitation for better results and discuss methods of improving the ability to discern cognitive biases through experimenting with expert elicitation as a cycle of learning from one experiment to the other.

Probabilistic Language Modelling with Hierarchical Bayesian Nonparametrics

Presenter: **David A Roberts**

E-mail: david.roberts@uqconnect.edu.au

Authors: **David A Roberts** and Marcus Gallagher

Affiliation: University of Queensland

Probabilistic language models --- probability distributions over textual data --- are an important element in a number of natural language applications, including text prediction and compression, and machine translation, to name a few. In the past decade, hierarchical Bayesian nonparametric methods have demonstrated strong performance in this area, most notably those based on hierarchical Dirichlet and Pitman-Yor processes (HPYP). This research investigates the efficacy of combining several different approaches undertaken in this area within a single coherent Bayesian model. In particular, a HPYP-based model capable of simultaneously performing word segmentation, part-of-speech tagging, and word prediction in an online manner, is implemented and evaluated experimentally. Preliminary results suggest this model performs well even when provided with only a small amount of training data, and with minimal manual annotation required. As part of this work, an open-source Python/C++ library for performing inference via Gibbs sampling on Chinese restaurant processes will be made available, allowing new and existing HPYP-based models to be implemented with minimal programming effort, but without sacrificing too much computational efficiency.

Communication of Bayesian Uncertainty to non-experts decision makers through Data Visualisation

Presenter: **Jessie Roberts**

E-mail: jessie.roberts@qut.edu.au

Authors: **Jessie Roberts** and Kerrie Mengersen

Affiliation: Queensland University of Technology

As the use of Bayesian inference in applied settings increases so does the need for effective tools that can communicate this uncertainty clearly and efficiently to decision makers, who are often not statisticians or data analysts. Despite this recognised need, there remains a lack of available tools, guidelines, or validated methods for translating this uncertainty to the decision maker.

Data visualisation is recognised as an effective method for efficiently communicating statistical inference to the non-expert, this research aims to extend the current body of research around data visualisation as a communicative tool to incorporate uncertainty communication.

This poster outlines a proposed study that will explore the effectiveness of three different visualisation strategies for communicating uncertainty. The case studies used as the focus of this visualization is a Bayesian analysis of a geospatial health data of cancer incidence and death. The effectiveness of the visualisations will be evaluated through an online survey with respondents from a range of statistical backgrounds as well as 2 focus groups engaging public health policy makers and other relevant decision makers.

Functional regression ABC for non-parametric density estimation

Presenter: **Guilherme Souza Rodrigues**

E-mail: g.rodrigues@unsw.edu.au

Authors: **Guilherme Souza Rodrigues**, David Nott and Scott Sisson

Affiliation: University of New South Wales

In this work we propose a new non-parametric procedure for modelling a set of associated density functions, each of which relates to a different group. This recurrent statistical challenge arises, for example, in the management context when one wants to compare the density of the daily profit of different branches of a given company. As the densities are affected by common factors, and thus might have similar traits, the estimation can be substantially enhanced by sharing strength across the existing groups. In a Bayesian approach, we introduce a hierarchically structured prior, defined over functions, using convenient transformations of Gaussian processes. With this formulation, each observation from the prior corresponds to a set of density functions, rather than a vector of finite dimension. Inference is performed through a combination of Approximate Bayesian Computation (ABC) and functional regression adjustment. The well-known Kernel estimator (computed for each group) plays the role of the summary statistic in the ABC mechanics. This work is understood to be the first to use ABC to estimate infinite dimensional parameters. By avoiding MCMC methods, the proposed technique provides approximate posterior samples at a considerable lower computational cost. We illustrate some interesting properties and give further insight of the method through an instructive example.

Noncrossing Bayesian quantile regression

Presenter: **Thais Carvalho Valadares Rodrigues**

E-mail: t.rodrigues@unsw.edu.au

Authors: **Thais Carvalho Valadares Rodrigues** and Yanan Fan

Affiliation: University of New South Wales

Quantile regression models provide a wider picture of the conditional distribution of the response variable by describing the effect of the covariates at any quantile of interest. However, as they are usually fitted separately, the final estimates may not respect the logical ordering of percentiles, violating basic probabilistic rules and resulting in crossing quantile curves. A fully Bayesian two-stage approach to solve this issue is proposed in this talk. The method is very general, handling, on the first stage, both linear and non-linear standard Bayesian quantile regression based on the asymmetric Laplace distribution. The initial estimates are then adjusted by borrowing strength across nearby quantiles using Gaussian process regression, yielding smoother noncrossing final estimates. Avoiding MCMC on the second stage and controlling the crossing constraint through a single parameter reduces computational cost giving rise to a simple alternative approach. The theoretical aspects of the proposal will be explored and the performance demonstrated on simulated and real examples.

The power variance function copula model in bivariate survival analysis

Presenter: **Jose S. Romeo**

E-mail: j.romeo@auckland.ac.nz

Authors: **Jose S. Romeo**^{1,2}, Renate Meyer¹ and Diego I. Gallardo³

Affiliations: ¹ Department of Statistics, University of Auckland, New Zealand

² Department of Mathematics, University of Santiago, Chile

³ Institute of Mathematics and Statistics, University of Sao Paulo, Brazil

In this work we present the Power Variance Function (PVF) Family copula model as dependence structure for modelling multivariate lifetime data. The PVF copula includes Clayton, Positive Stable (Gumbel) and Inverse Gaussian copulas as special cases. Dependence properties of the copula models are described as well as simulation techniques. Inference is carried out in the Bayesian framework. We perform a simulation study to investigate the properties of the parameters. Finally, we illustrate the usefulness of the methodology in a data set that has been discussed in the literature.

Keywords: Multivariate survival analysis, Archimedean Copulas, Dependence.

Mixture modelling as a tool for Exposomics

Presenter: **Aleysha Thomas**

E-mail: a12.thomas@qut.edu.au

Authors: **Aleysha Thomas**¹, Kerrie Mengersen¹, Nicole White¹, George Mellick³, Fiona Harden², Leisa-Maree Toms²

Affiliations: ¹Mathematical Sciences School, Queensland University of Technology,

²Clinical Sciences School, Queensland University of Technology,

³Clinical Neurosciences, Eskitis Institute for Cell and Molecular Therapies

Exposomics is a relatively new field of research around the 'exposome'. The exposome is described as a variable entity, similar to the genome, containing all exposures in an individual's lifetime. Conceptually, exposomics can be viewed as a 3 node network of a disease, exposure and genes. Methods to analyse the exposome as a whole rather than in parts, is a major challenge for the field. This research focuses on the network of Parkinson's Disease (PD), Persistent Organic Pollutants (POPs) and the associated genes and single nucleotide polymorphisms (SNPs). As a part of exploring methods in exposomics, we investigate the potential of using PD patients' symptom profiles and risk factors, including exposure to pesticides, to identify sub-groups of PD patients. A Bayesian mixture model is used to determine patients' membership into k sub-groups. Multiple models are tested and compared using goodness of fit criteria. These models include those with common and specific variance for each sub-group and a comparison of the use of inequality constraints for each sub-group. We also explore the use of mixture of CART models to identify patient group memberships.

Missingness, Mixtures, and Multiple {i}Mputation

Presenter: **Nicholas Tierney**

E-mail: Nicholas.tierney@gmail.com

Authors: **Nicholas Tierney**¹, Fiona Harden^{2,3}, Maurice Harden³, Kerrie Mengersen¹.

Affiliations: ¹QUT Statistical Science, ²QUT Faculty of Health, ³Hunter Industrial Medicine

This research aims to develop employee health profiles to assist in identifying subgroups at risk of developing chronic disease. We fitted a mixture model to identify these subgroups. A challenge of this research lies in fitting these models to the case study data, due to its features of: i) Large amounts of missing data (>60% missing), ii) Nested structure of the workplace, and iii) Irregular time periods between measurements. In this poster we describe how we account for some of these features.

Zmix & Zswitch: methods for overfitting mixture models for order estimation

Presenter: **Zoe van Havre**

E-mail: zoevanhavre@gmail.com

Author: **Zoe van Havre**

Affiliations: Queensland University of Technology and Paris Dauphine (France)

Finite mixture models represent an intuitive approach to model-based clustering, yet their estimation is complicated by a number of theoretical and computational difficulties, particularly when the number of component is unknown. Recent developments in the Bayesian asymptotic theory of these models now allow us to directly overfit the number of components in a mixture, and encourage any unnecessary groups to have negligible weights in the posterior using prior restrictions. Such a prior choice leads to difficulties in practice however, creating such an extreme posterior surface that MCMC sampling techniques become 'stuck' and unable to mix properly.

This poster presents a set of algorithms which aim to provide an effective and straightforward modelling approach for overfitting mixture models for order estimation. The first method is a Gibbs sampler coupled with Prior Tempering, called "Zmix". Prior Tempering is a new technique based on parallel tempering where the degree of tempering is determined entirely by a hyper-prior on the weights. This deals specifically with the difficulties arising from overfitted mixtures in order to produce well mixed posterior samples of the desired target.

The label switching problem is resolved using a relabelling algorithm (Zswitch) which is able to deal with a large number of empty groups rapidly, does not require comparisons across a large number of labelling permutations, or the development of label invariant loss functions.

We illustrate the application of Zmix and Zswitch on a set of simulated normal univariate mixture models.

Understanding Uncertainties in Non-Linear Population Trajectories: a Bayesian Semi-Parametric Hierarchical Approach to Large-Scale Surveys of Coral Cover

Presenter: **Julie Vercelloni**

E-mail: julie.vercelloni@hdr.qut.edu.au

Authors: **Julie Vercelloni**^{1,2}, M. Julian Caley², Mohsen Kayal³, Samantha Low-Choy¹, Kerrie Mengersen¹

Affiliations: ¹Queensland University of Technology, ²Australian Institute of Marine Science, ³University of California, Santa Barbara.

Recently, attempts to improve decision making in species management have focussed on uncertainties associated with modelling temporal fluctuations in populations. Reducing model uncertainty is challenging; while larger samples improve estimation of species trajectories and reduce statistical errors, they typically amplify variability in observed trajectories. In particular, traditional modelling approaches aimed at estimating population trajectories usually do not account well for nonlinearities and uncertainties associated with multi-scale observations characteristic of large spatio-temporal surveys. We present a Bayesian semiparametric hierarchical model for simultaneously quantifying uncertainties associated with model structure and parameters, and scale-specific variability over time. We estimate uncertainty across a four-tiered spatial hierarchy of coral cover from the Great Barrier Reef. Coral variability is well described; however, our results show that, in the absence of additional model specifications, conclusions regarding coral trajectories become highly uncertain when considering multiple reefs, suggesting that management should focus more at the scale of individual reefs. The approach presented facilitates the description and estimation of population trajectories and associated uncertainties when variability cannot be attributed to specific causes and origins. We argue that our model can unlock value contained in large scale datasets, provide guidance for understanding sources of uncertainty, and support better informed decision making.

An approximate Bayesian Computation approach for estimating parameters governing the expansion of melanoma cell colonies in vitro

Presenter: **Brenda N Vo**

E-mail: n1.vo@qut.edu.au

Authors: **Brenda N Vo**, Christopher C Drovandi, Anthony N Pettitt, and Graeme J Pettet

Affiliation: Queensland University of Technology

Melanoma is the third most commonly diagnosed cancer in Australian men and women (10% of all cancers), with more than 12,500 new cases reported every year (AIHW, 2012). Although much progress has been made in terms of developing and analysing mathematical models of the cell biology processes that drive the expansion of melanoma cell colonies, far less progress has been made in terms of understanding how to estimate model parameters from experimental image-based data, such as time-lapse data and magnetic resonance imaging data. In this talk, we present a novel application of approximate Bayesian computation using a sequential Monte Carlo (SMC-ABC) approach to infer the values of the cell diffusivity, the cell proliferation rate and the cell-to-cell adhesion strength, from a discrete two-dimensional stochastic model of collective cell spreading, and to quantify the uncertainty associated with these estimates. Parameter estimations are based on a detailed experimental data set describing the expanding colonies of human malignant melanoma cells (MM127) in a circular barrier assay, in two separate scenarios with and without Mitomycin-C pretreatment to suppress cell proliferation. The Bayesian sequential learning approach allows uncertainty about parameters to be combined from two different experiments in a principled way. Although the SMC-ABC approach is computationally expensive, compared to other methods that are based on the deterministic models and obtain point estimates, the insights gained from considering the joint posterior distributions of the parameters outweigh the costs.

A Spatial Poisson Hurdle Model with Clustered Random Effects to Understand the Provision of Health

Presenter: **Nicole White**

E-mail: nm.white@qut.edu.au

Authors: **Nicole White**, Kerrie Mengersen

Affiliations: Queensland University of Technology and Cooperative Research Centre for Spatial Information

The management of public health infrastructure is a difficult task, heavily reliant on understanding the demands and capacities of resources in conjunction with how resource utilization varies spatially across the community. In the explanation of demand and capacity, these are often viewed as subjective constructs, and are therefore difficult or impossible to model explicitly. Instead, from a modelling perspective, they may be viewed as latent effects, enabling comparative inferences in the presence of multiple resources. Furthermore, resources may naturally cluster such that, if present, may provide an explanation for how different resources are viewed and accessed by the community.

In response, this talk develops a spatial Poisson hurdle model with fixed and random effects for understanding trends in the utilization of public health infrastructure. Spatial information incorporated into the model by way of covariates and spatially correlated random effects at the areal level. The model is extended to accommodate the possibility of clustering for inferring similarities in latent demand, achieved by the introduction of a Dirichlet Process prior. The development of this model is motivated by the analysis of participation data from a no-fee mammographic screening program, located in Brisbane, Australia.

A comparison of maximum likelihood and Bayesian estimation approaches to examine the effects of homocysteine and other coronary artery disease (CAD) risk factors on CAD using structural equation modelling

Presenter: **Richard J Woodman**

E-mail: Richard.Woodman@flinders.edu.au

Authors: **Richard J Woodman**¹, Xiuqin Hong^{2,3}, Shuiyuan Xiao², Arduino A Mangoni⁴, Shuling Wang³, Jian Chen³, Qian Peng³, Qi Zuo³, Dan Zeng³, Jia Wang³, Na Zhang³, Chenxu Wang³

Affiliations: ¹Flinders Centre for Epidemiology and Biostatistics, School of Medicine, Flinders University, ²School of Public Health, Central South University, Changsha, ³Institute of Gerontology, Hunan Geriatric Hospital, Changsha, ⁴Discipline of Clinical Pharmacology, School of Medicine, Flinders University.

Background: The relationship between coronary artery disease (CAD) and plasma homocysteine is well known, but whether homocysteine has a causal influence on CAD or is simply a risk marker is still unclear. Structural equation modelling (SEM) can assist in understanding complex relationships and causal pathways, however traditional maximum likelihood (ML) approaches to SEM often results in relatively poor model fit due to restrictive assumptions including zero cross-loadings.

Objectives: To examine model fit, effect estimates and the explained variance of homocysteine and CAD using ML-SEM and Bayesian SEM (BSEM) approaches with Mplus software.

Methods: Data was obtained from 242 individuals with CAD (60.3% males, mean (\pm SD) age 76.2 \pm 7.0 and an age-and-sex-matched group of 218 individuals without CAD (58.7% males, age 75.1 \pm 6.6 years). Measures included homocysteine, systolic and diastolic blood pressure, triglycerides, and cholesterol. Body mass index, blood urea nitrogen, C-reactive protein and uric acid were used as markers of insulin sensitivity, renal function, inflammation and oxidative stress respectively. We obtained ML-SEM and BSEM estimates of proposed causal pathways between homocysteine, other risk factors and CAD for males and females separately.

Results: BSEM approaches increased flexibility towards satisfactory model fit. A large proportion of the variance in CAD was explained by both ML-SEM and BSEM approaches. However, effect estimates were generally more conservative and the explained variance of homocysteine lower for models using BSEM.

Conclusions: BSEM allowed greater flexibility in modelling homocysteine and other CAD risk factors than ML-SEM approaches alone. Explained variance and effect estimates were generally more conservative using BSEM.

Towards a Bayesian Network framework for predicting and understanding cumulative effects in a complex system

Presenter: **Paul Wu**

E-mail: p.wu@qut.edu.au

Authors: **Paul Wu**^{1,2}, Kerrie Mengersen¹, Julian Caley²

Affiliations: ¹Queensland University of Technology, ²Australian Institute of Marine Science

Predicting and understanding the behaviour of complex systems is a challenging process due to the evolution of system behaviour over time arising from complex interactions within the system. Known as emergent behaviour or cumulative effects, this is one of the key characteristics of a dynamic complex system made up of many interacting components [1, 2]. Examples of complex systems include ecological systems [2] and critical infrastructure systems, e.g. transportation networks, the electricity grid and telecommunication networks [3].

The Bayesian Network (BN) provides a structured method for the modelling of such complex systems [4]. Based on an underlying acyclic, directed graph, the BN is made up of nodes, which represent factors and/or system components, and the links between them as characterised by conditional probabilities. A BN can be used to analyse causal linkage pathways in a complex system and for predicting cumulative effects [2]. However, due to the acyclic limitation, it is not possible to model feedback loops; this limits the ability to study emergent phenomena that are not known a priori.

One way to overcome this limitation is through a Dynamic Bayesian Network (DBN), which extends the BN by replicating and extending the network structure over discretised points in time, known as time slices [5]. However, interpreting the resultant system trajectory for decision making can be challenging especially for systems that do not converge to a stable equilibrium point [6]. This poster compares the BN and DBN approaches for modelling complex systems using the marine ecosystem as a case study.

References

- [1] A. P. Sage and C. D. Cuppan, "On the systems engineering and management of systems of systems and federations of systems," *Information, Knowledge, and Systems Management*, vol. 2, pp. 325-345, 2001.
- [2] K. R. N. Anthony, J. M. Dambacher, T. Walshe, and R. Beeden, *A framework for understanding cumulative impacts, supporting environmental decisions and informing resilience-based management of the great barrier reef world heritage area*: Australian Institute of Marine Science, Townsville; CSIRO, Hobart; NERP Decisions Hub, University of Melbourne and Great Barrier Reef Marine Park Authority, Townsville, 2013.
- [3] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdependencies," *Control Systems, IEEE*, vol. 21, pp. 11-25, 2001.
- [4] C. A. Pollino, O. Woodberry, A. Nicholson, K. Korb, and B. T. Hart, "Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment," *Environmental Modelling & Software*, vol. 22, pp. 1140-1152, 2007.
- [5] K. P. Murphy, "Dynamic bayesian networks: representation, inference and learning," University of California, Berkeley, 2002.
- [6] C. S. Holling, "Resilience and Stability of Ecological Systems," *Annual Review of Ecology and Systematics*, vol. 4, pp. 1-23, 1973.

Estimating scale and scope economies in Australian universities under a Bayesian framework

Presenter: **Liang-Cheng Zhang**

E-mail: liang-cheng.zhang@griffithuni.edu.au

Authors: **Liang-Cheng Zhang**¹ and Jia-Jia Syu²

Affiliations: ¹Griffith Business School, Griffith University, ²School of Population Health, University of Queensland

The purpose of this study is to estimate scale and scope economies in the Australian university sector. These estimations could give universities and the share-holders the up-to-date cost structure that underpin provision in higher education sector, i.e. the potential for efficiency improvements of HEIs to reflect the true education cost and operate at efficient scale. Estimating the scale and scope economies in higher education, therefore, has been one of the main interests among researchers. Different from past studies using maximum likelihood estimation to estimate scale and scope economies, this study will consider the whole estimation process under a Bayesian framework. Our data comprises a balanced panel data including 37 Australian public universities over the period 2010-12. Bayesian stochastic frontier model is used to account for inefficient production and heterogeneity across universities. This model contains a normalised quadratic cost function with seven outputs and three input factors. Estimates of scale and scope economies will be calculated with the sample data and estimated coefficients from the model. Iterative Monte Carlo Markov Chain (MCMC) simulation is used to derive the unknown parameters (including scale and scope economies) with the prior knowledge. These posterior parameters are estimated from the software, BUGS, running with a burn-in of 20,000 iterations, with 200,000 retained draws. We will present whether universities should jointly or separately produce their outputs and which level of output they should produce to save costs based on our estimates of scale and scope economies.

Keywords: Bayesian framework, stochastic frontier model, scale and scope economies

Approximation Bayesian computation methods using Composite Likelihood

Presenter: **Xin Zhang**

E-mail: xin.zhang@unsw.edu.au

Author: **Xin Zhang**

Affiliation: University of New South Wales

Approximate Bayesian computation (ABC) methods are a family of likelihood-free inference algorithms for Bayesian analysis of intractable models. The application of these algorithms is restricted to models of relatively low-dimensions due to the comparison of high-dimensional summary statistics. To avoid this dimensionality issue, we propose, in this article, to use the composite likelihood (CL), whose components are the densities of subsets of summary statistics, instead of the intractable likelihood function in Bayesian analysis. Further, we utilise the factorisation structure of the CL to design a parallel computation framework for ABC. The performance of our algorithms is first evaluated via a simulated example of multivariate Gaussian distributions with diagonal-block and autoregressive covariance matrices. Then a real example of the foreign exchange rate daily returns data set is analysed through a multivariate g -and- k distribution.

A novel approach for Markov random field with intractable normalizing constant on large lattice

Presenter: **Wanchuang Zhu**

E-mail: z3406438@unsw.edu.au

Author: **Wanchuang Zhu**

Affiliation: University of New South Wales

We introduce a novel method to solve issue of normalizing constant in Markov random field. We take advantage of conditional independence of Markov random field to avoid calculation of normalizing constant. Our algorithm can be applied to Markov random field defined on large lattice, while other methods are problematic in such scenario. Compared to existing methods, our method does not require so many assumptions. Therefore, our method estimates parameters more precisely, especially variance estimation. In addition, our method is used to handle second order neighbourhood structure problem originally. Finally, we study proposed method in both simulated dataset and real dataset.