



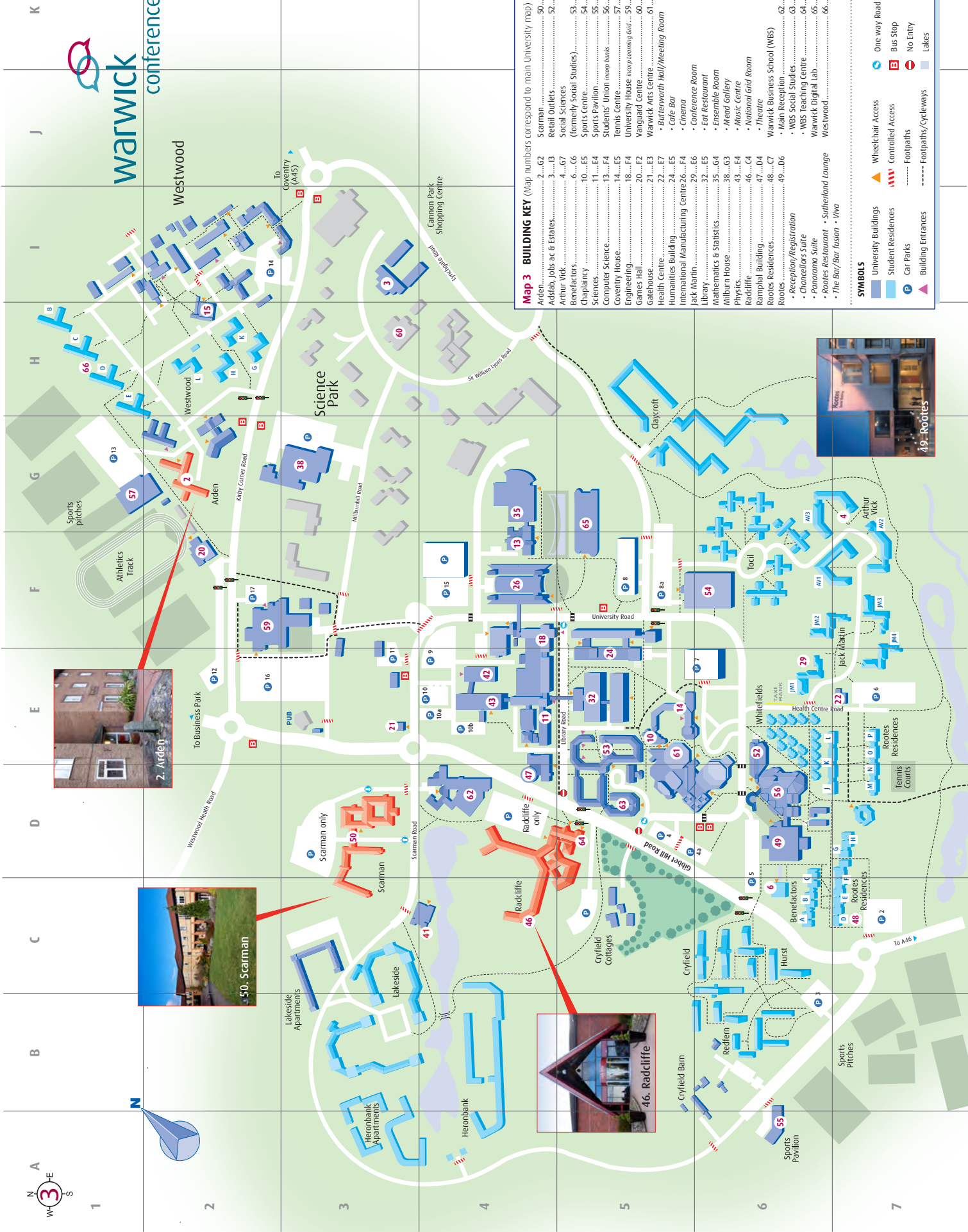
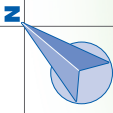
## **Workshops**

May 2010

### **Model Uncertainty**

30 May – 1 June 2010

THE UNIVERSITY OF  
**WARWICK**



**Map 3 BUILDING KEY** (Map numbers correspond to main University map)

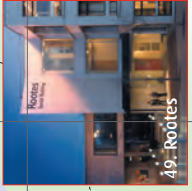
Arden.....	2, 62
Alden, jobs at & Estates.....	3, 15
Arthur Wick.....	4, 67
Benefactors.....	6, 66
Chaplaincy.....	10, 65
Sciences.....	11, 64
Computer Science.....	13, 64
Community House.....	14, 65
Engineering.....	18, 64
University House <i>incap learning aid</i> .....	19, 60
Games Hall.....	20, 62
Gaehouse.....	21, 63
Health Centre.....	22, 67
Humanities Building.....	24, 65
International Manufacturing Centre.....	26, 64
Jack Martin.....	32, 65
Library.....	32, 65
Mathematics & Statistics.....	35, 64
Milburn House.....	38, 63
Physics.....	43, 64
Radcliffe.....	46, 64
Ramphal Building.....	47, 64
Rootes Residences.....	48, 67
Rootes.....	49, 66
Reception/Registration.....	49, 66
Chanakya Suite.....	49, 66
Panorama Suite.....	49, 66
Rootes Restaurant.....	49, 66
The Bar/Bar Vision.....	49, 66
Westwood.....	66

**SCARMAN**

- Scarman
- Retail Outlets
- Social Sciences (formerly Social Studies)
- Sports Pavilion
- Students' Union
- Tennis Centre
- University House
- Vanguard Centre
- Warwick Arts Centre
- Butterworth Hall/Meeting Room
- Cafe Bar
- Cinema
- Conference Room
- Car Restaurant
- Ensemble Room
- Mead Gallery
- Music Centre
- National Grid Room
- Theatre
- Warwick Business School (WBS)
- Main Reception
- WBS Social Studies
- WBS Teaching Centre
- Warwick Digital Lab

**SYMBOLS**

- University Buildings
- Student Residences
- Car Parks
- Building Entrances
- Wheelchair Access
- Controlled Access
- Footpaths
- Footpaths/Cycleways
- One way Road
- Bus Stop
- No Entry
- Lakes



---

## TRAVEL INFORMATION

---

### Venue

The workshops take place in the

Mathematics & Statistics Building,  
University of Warwick,  
Coventry, CV4 7AL,  
UK

The workshop venue is building no. **35** on the campus map and it is also known as “Zeeman Building”.

Check-in for the workshop accommodation is in Scarman House (building no. **50** on the campus map).

### Note:

The University is **not** in the town of Warwick (which is about 8 miles away), and it is **not** the same as Coventry University (which is a different university, located in the centre of Coventry). This is important when telling taxi drivers where you want to be!

### Travelling to the University of Warwick

Information on getting to the University of Warwick from Coventry, as well as from other directions locally and further afield, can be found at <http://www2.warwick.ac.uk/fac/sci/statistics/crism/workshops/model-uncertainty/travel/>.

### Parking

Parking spaces are available at the Scarman House, which is at a walking distance from the Maths and Stats building. Please contact the Scarman reception desk for information about parking. For people not staying at Scarman, Carpark 15 is the closest to the Mathematics and Statistics building.

---

# WELCOME!

---

## Accommodation

Upon arrival, please check in in Scarman House Reception for instructions (building no. 50 on the campus map). This is also the place to ask about car parking, left luggage, etc.

The full address of the Scarman House is:

The Scarman Training and Conference Center,  
University of Warwick,  
Gibbet Hill Road  
Coventry, CV4 7AL  
UK  
Tel: +44 (0) 24 7622 1111  
Fax: +44 (0) 24 7652 0362  
Email: Scarman@warwick.ac.uk

*Note:* Being Bank Holiday, the Scarman House will be open for check-in *only from 5pm of Saturday 29 May 2010*. If you expect to arrive much earlier on Saturday 29, please get in touch in advance with one of the local organizers for information or help.

## Workshop registration

Registration for the workshop will take place at the main atrium of the Mathematics & Statistics Building (no. 35 on the map). The registration time is Sunday 30 May 2010, 8:00-9:00 am.

## Computing facilities

Being bank holiday weekend, we regret to inform you that there will be no computer room available for the participants. However, if you have a WiFi enabled laptop you may access basic internet services<sup>1</sup> from within the Mathematics & Statistics Building by connecting to the “hotspot” wireless network and starting up a web browser. The login details for wireless are:

Username	Password
<i>statsusr</i>	<i>ready2go</i>

## Meals

During the workshops, coffee and tea breaks will be served in the main atrium of the Mathematics & Statistics Building.

Lunch buffets and dinners (for participants who have registered) will be served in Scarman Building (building no. 50).

For participants who have not registered for dinners, dining options include the *Bar Fusion* in Rootes, The *Rootes Restaurant* or the *Eat Restaurant* in the Warwick Arts Centre. The *Varsity Pub* (coordinates E3 in the campus map) is a valid alternative near Scarman. A number of Bars and Cafés are at a walking distance from both Scarman and the Mathematics and Statistics Building. Please visit <http://www2.warwick.ac.uk/services/foodanddrink/> for all on-campus options.

---

<sup>1</sup>but note that sending email via SMTP is not allowed

## Organisers

### CRiSM administrator

- ▷ Paula Matthews ([paula.matthews@stats.warwick.ac.uk](mailto:paula.matthews@stats.warwick.ac.uk))

### Model Uncertainty Workshop

- ▷ Jim Griffin ([J.E.Griffin-28@kent.ac.uk](mailto:J.E.Griffin-28@kent.ac.uk))
- ▷ Gareth Roberts ([G.O.roberts@warwick.ac.uk](mailto:G.O.roberts@warwick.ac.uk))
- ▷ Mark Steel ([M.F.Steel@stats.warwick.ac.uk](mailto:M.F.Steel@stats.warwick.ac.uk))
- ▷ Dario Spanò ([D.Spano@stats.warwick.ac.uk](mailto:D.Spano@stats.warwick.ac.uk))

If it is necessary to contact the organizers at any time during the workshops you can ring Flavio Goncalves 0774 3714979, Dario Spanò at 0796 9797473 or Mark Steel 0753 3012955.

---

## MEDICAL AND EMERGENCY INFORMATION

---

### **Medical assistance**

The University Health Centre is open Monday-Friday 09.00-13.00 and 14.00-17.00. Visitors in need of emergency assistance should contact Security on internal extension 22222.

### **Emergency services and fire procedures**

For help in an emergency dial 22222 from any internal telephone and your call will be directed appropriately. Visitors are asked to familiarise themselves with the University's fire procedures which are displayed in each bedroom.

*On discovering a fire in a building:*

Raise the alarm by breaking the glass in the nearest *Break Glass* point.

*On hearing the continuous ringing of fire bells:*

Stop what you are doing.

Leave by the nearest Fire Exit.

Walk calmly, do not run.

Do not stop to collect personal belongings.

Make your way to the nearest evacuation point, standing well clear of the building.

Do not re-enter the building until told to do so by the Fire Service or University Security Staff.

---

# MODEL UNCERTAINTY

---

	Event	Time	Details
Sunday 30 May	<i>Registration</i>	08.00–08.45	
	Session B.1.1	08.45–09.00	<i>Welcome</i>
		09.00–09.45	Jim Berger
		09.45–10.30	Christian Robert
		10.30–11.00	Dongchu Sun
	<i>Coffee</i>	11.00–11.30	<i>in the main atrium, Maths &amp; Stats</i>
	Session B.1.2	11.30–12.15	David Madigan
		12.15–12.45	Geoff Nicholls
	<i>Lunch</i>	12.45–14.00	<i>in Scarman House (Building 50)</i>
	Session B.1.3	14.00–14.45	Yee Whye Teh
14.45–15.30		Antonio Lijoi	
15.30–16.15		Chris Holmes	
<i>Tea</i>	16.15–16.45	<i>in the main atrium, Maths &amp; Stats</i>	
Session B.1.4	16.45–17.30	Carlos Carvalho	
	17.30–18.00	Kostas Kalogeropoulos	
<i>Dinner buffet</i>	18.30	<i>in Scarman House</i>	
Monday 31 May	Session B.2.1	09.00–09.45	Peter Mueller
		09.45–10.30	Matthew Stephens
		10.30–11.00	Xavier Didelot
	<i>Coffee</i>	11.00–11.30	<i>in the main atrium, Maths &amp; Stats</i>
	Session B.2.2	11.30–12.15	Robert Kohn
		12.15–12.45	Oliver Ratmann
	<i>Lunch</i>	12.45–14.00	<i>in the main atrium, Maths &amp; Stats</i>
	Session B.2.3	14.00–14.45	David Spiegelhalter
		14.45–15.30	Nanny Wermuth
		15.30–16.00	Anthony O’Hagan
<i>Tea</i>	16.00–16.30	<i>in the main atrium, Maths &amp; Stats</i>	
<i>POSTER SESSION</i>	16.30–19.00	<i>in the main atrium, Maths &amp; Stats</i>	
<i>Conference dinner</i>	19.30	<i>in Scarman House</i>	
Tuesday 1 June	Session B.3.1	09.00–09.45	Jon Forster
		09.45–10.30	Ed George
		10.30–11.00	Ioannis Ntzoufras
	<i>Coffee</i>	11.00–11.30	<i>in the main atrium, Maths &amp; Stats</i>
	Session B.3.2	11.30–12.15	Henry Whynn
		12.15–12.45	Aki Vehtari
	<i>Lunch</i>	12.45–14.00	<i>in Scarman House</i>
	Session B.3.3	14.00–14.45	David Dunson
		14.45–15.30	Athanasios Kottas
		15.30–16.00	Elja Arjas
<i>Tea</i>	16.00–16.30	<i>in the main atrium, Maths &amp; Stats</i>	
Session B.3.4	16.30–17.15	Alan Gelfand	
	17.15–17.45	Jim Zidek	
<i>Dinner</i>	18.15	<i>in Scarman House</i>	

---

## Notes

- ▷ Speakers should gauge their presentations so as to leave enough time for discussion within the time allocated. Ideally, about five minutes at the end of each talk should be reserved for questions.
- ▷ The main atrium of the Mathematics & Statistics Building is the open area as you move from the main entrance to the statistics department (where the mural with vertical lines is).
- ▷ All sessions will be held in MS.01. The posters will be displayed in the main atrium.

## Abstracts

### Invited Talks.

#### **PBIC and Effective Sample Size**

by JIM BERGER

A generalization of BIC (called PBIC: Prior-based Bayes Information Criterion) has been available for some time. BIC is often a reasonable approximation to a Bayes factor in i.i.d. situations, but is typically not a good approximation in non-i.i.d. situations. PBIC is not only typically a better approximation - while still being closed form and almost as simple as BIC - but is an exact Bayes factor when the likelihood is normal. The main barrier to implementation of PBIC is that it requires determination of the "effective sample size" corresponding to each parameter in the model. Progress that has been made in determining effective sample size will be discussed.

#### **On the Long Run Volatility of Stocks.**

by CARLOS CARVALHO

(Joint work with Hedibert Lopes and Robert McCulloch)

In "Predictive Systems: Living with Imperfect Predictors", Pastor and Stambaugh (2008) develop a framework for estimating expected returns. In "Are Stocks Really Less Volatile in the Long Run" (2009) they use this framework to assess the conventional wisdom that stocks are less volatile over long horizons than short horizons. They show that this conclusion is only reached by ignoring important parameter uncertainty. They also argue that a key component of prior information concerns the correlation between unanticipated expected return and the unpredictable return. The predictive system framework consists of a vector auto regression in the stock return, the latent expected return for the next period, and a set of variables thought to be able to predict returns. They assume that the innovation covariance is constant over time. This assumption runs counter to much empirical evidence. In this paper we extend the predictive systems framework to account for time varying volatility. We do so in a way that allows us to incorporate complex economically based prior information. In particular, we use prior information about the time series of the correlation between unanticipated expected return and the unpredictable return. In this enriched environment, we examine what kind of prior information is needed to make stock returns less volatile in the long run.

#### **High-dimensional nonparametric Bayes variable selection via nonlinear embeddings.**

by DAVID DUNSON

Joint work with Anirban Bhattacharya. In many application areas, massive dimensional predictors are routinely collected, with some of these predictors being moderately to highly correlated. Standard Bayes and frequentist methods tend to select one of a correlated group of predictors,

and have trouble scaling to very high dimensions (10,000 or more predictors). Sparse latent factor regression models represent a promising approach for dealing with this problem, but current approaches rely on restrictive normality and linearity assumptions. For greater flexibility and to enable a greater degree of dimensionality reduction, we propose a sparse non-linear factor modeling framework. The proposed model relates the high-dimensional measured variables to a non-linear transformation of a low-dimensional vector of latent variables. A carefully structured Gaussian process prior is chosen for the unknown transformation function, with the prior being centered on a sparse linear factor model. The proposed structure allows for the inclusion of infinitely-many factors with the impact of these factors on the response decreasing in the factor index, so that an adaptive truncation can be chosen with all factors above a certain index discarded with negligible impact. Theoretical results are considered and efficient methods are developed for posterior computation via blocked Gibbs sampling combined with grid approximations. The methods are assessed in simulation studies, compared to competitors such as elastic net, and applied to gene expression data.

### **Specification of prior distributions under model uncertainty.**

by JON FORSTER

We consider the specification of prior distributions for Bayesian model comparison, focusing on regression-type models. We propose a particular joint specification of the prior distribution across models so that sensitivity of posterior model probabilities to the dispersion of prior distributions for the parameters of individual models (Lindleys paradox) is diminished. We illustrate the behaviour of inferential and predictive posterior quantities in linear and log-linear regressions under our proposed prior densities with a series of simulated and real data examples.

### **Spatial Modeling of Presence-Only Data over Large Regions**

by ALAN GELFAND

Explaining species distribution using local environmental features is a long standing ecological problem. Often, available data is collected as a set of presence locations only thus precluding the possibility of a presence-absence analysis. We propose that it is natural to view presence-only data for a region as a point pattern over that region and to use local environmental features to explain the intensity driving this point pattern. This suggests hierarchical modeling, treating the presence data as a realization of a spatial point process whose intensity is governed by environmental covariates. Spatial dependence in the intensity surface is modeled with random effects involving a zero mean Gaussian process. Highly variable and typically sparse sampling effort as well as land transformation degrades the point pattern so we augment the model to capture these effects. The Cape Floristic Region (CFR) in South Africa provides a rich class with such species data. The potential, i.e., nondegraded presence surfaces over the entire area are of interest from a conservation and policy perspective. Our model assumes grid cell homogeneity of the intensity process where the region is divided into  $\sim 37,000$  grid cells. To work with a Gaussian process over a very large number of cells we use predictive process approximation. Bias correction by adding a heteroscedastic error component is implemented. The model was run for a number of different species. Model selection was investigated with regard to choice of environmental covariates. Also, comparison is made with the now popular Maxent approach, though the latter is much more limited with regard to inference. In fact, inference such as investigation of species richness immediately follows from our modeling framework.

### **Fully Bayes Model Selection with a Generalized g-Prior**

by ED GEORGE

For the normal linear model variable selection problem, we propose selection criteria based on a fully Bayes formulation with a generalization of Zellner's g-prior. A special case of the prior formulation is seen to yield tractable closed forms for marginal densities and Bayes factors which

reveal new model characteristics of potential interest. Our results allow for  $p \ll n$  and extend to the multivariate regression setup. (This is joint work with Yuzo Maruyama).

### **Bayesian nonparametric clustering of sparse signals**

by CHRIS HOLMES

Bayesian mixture modelling is a widely used method for cluster analysis as it allows for characterisation of uncertainty surrounding principle dependence structures within data. Clustering is often used as an exploratory method to uncover hidden structure or to recover suspected structure. As data becomes ever cheaper to capture there is an increasing need for methods that can adjust towards many recorded variables being irrelevant to the clustering task. Variable selection has a substantial literature in regression or supervised learning settings but is much less studied in unsupervised clustering problems. In this talk we will report on the use of variable selection priors for nonparametric mixture models. Such priors tend to induce sparsity in the posterior model space and help characterise the relative information of those explanatory variables useful for clustering. We demonstrate how the use of hierarchical Dirichlet process priors allows for a principled way to construct these models; providing accurate indication of irrelevant variables whilst being able to quantify the relative relevance of informative variables. We pay particular attention to efficient MCMC sampling schemes for inference allowing for an unknown number of mixture components and an unknown number of relevant variables.

### **Exploring ODE model uncertainty via diffusions, with application to physiological processes.**

by KOSTAS KALOGEROPOULOS

Ordinary differential equations models are extensively used to describe various continuous time phenomena. Although successful in capturing certain aspects of the underlying mechanism, the need for further refinement is becoming increasingly apparent. In many applications, including our motivating example from physiological processes, this may be attributed to the inherent system noise. In our approach the system noise is explicitly modelled and disentangled from other sources of error. This is achieved by incorporating a diffusive component while retaining the mean infinitesimal behavior of the system. In PK/PD applications, the problem of inference on the non linear diffusion parameters is further complicated by the presence of measurement error, individual variability, imbalanced designs and so forth. We present a general inference framework through data augmentation which we illustrate through simulated and real PK/PD data.

### **Adaptive Independent Metropolis-Hastings sampling in challenging models.**

by ROBERT KOHN

Joint work with Paolo Giordani, Michael Pitt and Ralph Silva. We discuss the advantages of Independent Metropolis-Hastings sampling for models where it is very expensive to compute the likelihood or where the likelihood can only be simulated. We also show that as a byproduct of the sampling output we obtain the marginal likelihood, which makes it straightforward to carry out model comparison. The methodology is illustrated by applying it to time series state space models with the simulated likelihood obtained by the particle filter, and dynamic stochastic general equilibrium (DSGE) models.

### **Nonparametric mixture modeling for Poisson processes**

by ATHANASIOS KOTTAS

We present Bayesian nonparametric modeling for non-homogeneous Poisson processes over time or space. The methodology exploits the connection of the Poisson process intensity with a density function. Dirichlet process mixture models for the point process density yield flexible

prior models for the Poisson process. In particular, through appropriate choice of the mixture kernel, the modeling framework enables different shapes and properties for the intensity function depending on the context and application. Simulation-based model fitting provides posterior inference for any functional of the Poisson process that might be of interest. The modeling approach will be illustrated with applications in neuronal data analysis, involving comparison of the firing patterns of a neuron recorded under two distinct experimental conditions, as well as extreme value analysis, involving estimation of intensity of extremes over time. An extension to modeling for marked Poisson processes will also be discussed. Here, the model formulation builds the mark density and intensity function from a joint nonparametric mixture. A key feature of such models is that they yield general specifications for the corresponding conditional mark distribution resulting in flexible inference for different types of marks.

### **Gibbs type priors for Bayesian nonparametric inference on species variety**

by ANTONIO LIJOI

Sampling problems from populations which are made of different species arise in a variety of ecological and biological contexts. Basing on a sample of size  $n$ , one of the main statistical goals is the evaluation of species richness. For example, in the analysis of Expressed Sequence Tags (EST) data, which are generated by sequencing cDNA libraries consisting of millions of genes, one is interested in predicting the number of new gene species that will be observed in an additional sample of size  $m$  or in estimating the so-called sample coverage. In order to deal with these issues, we undertake a Bayesian nonparametric approach based on Gibbs-type priors which include, as special cases, the Dirichlet and the two-parameter Poisson-Dirichlet processes. We show how a full Bayesian analysis can be performed and describe the corresponding computational algorithm.

### **Dynamic Logistic Regression and Dynamic Model Averaging for Binary Classification**

by DAVID MADIGAN

We present a model for dynamic model averaging for binary classification. The model accounts for model uncertainty when parameters are expected to change over time by applying a Markov chain model to the "correct" model and a state space model for the parameters within each model. Parameterization is done in terms of forgetting and approximation of the marginal likelihood is computed via Laplace Approximation. We evaluate the method using simulated data and present an application to laparoscopic surgery in children. (with Tyler McCormick, Adrian Raftery, and Randall Burd)

### **Modeling Dependent Gene Expression**

by PETER MÜLLER

Joint work with Donatello Telesca, Giovanni Parmigiani, Ralph Freedman. We propose a Bayesian modeling approach for inference about dependence of high throughput gene expression. Our goals are to use prior knowledge about pathways to anchor inference about dependence among genes; to account for this dependence while making inferences about differences in gene's behavior across phenotypes; and to explore differences in the dependence itself across phenotypes. Useful features of the proposed approach are a model-based parsimonious representation of expression as an ordinal outcome, a novel and flexible representation of prior information on the nature of dependencies, and the use of a coherent probability model over both the structure and strength of the dependencies of interest. We evaluate our approach through simulations and in the analysis of data on expression of genes in the Complement and Coagulation Cascade pathway in ovarian cancer.

## **On approximating Bayes factors**

by CHRISTIAN ROBERT

Bayesian model choice being based on Bayes factors, efficient approximations of marginal likelihoods are necessary to conduct model comparison in highly complex models. In this talk, we review the existing solutions, including recent ones like nested sampling and particle learning, and show that Chib's (1995) solution can be very stable when it is available. We also reassess the harmonic mean estimator of Gelfand and Dey (1994) in terms of the support of an HPD region derived from the MCMC output. (This is joint work with Nicolas Chopin, Kate Lee, Jean-Michel Marin, and Kerrie Mengersen.)

## **Model Uncertainty: more than a statistical issue?**

by DAVID SPIEGELHALTER

Classical model selection focuses on the ability to fit the data in hand, generally with some form of penalty to prevent over-fitting. Bayesian model uncertainty attempts to identify a plausible model that gave rise to the current observations, which can then be used for predictions. In each case the future is assumed to arise from essentially the same process that gave rise to the past: in particular Bayesian uncertainty statements are conditional on the 'truth' of the assumed model or set of models. But since 'all models are wrong', these statements inevitably understate reasonable uncertainty.

Recent experience with financial models, and disputes about climate change, show that model uncertainty may be more than a technical statistical issue. We shall look at how model uncertainty is dealt with in a range of areas, and in particular examine whether any account is taken of the almost inevitable 'discrepancy' between any formal representation and the actual events that occur.

## **Bayesian testing of multivariate outcomes for association with genetic variants**

by MATTHEW STEPHENS

In many ongoing genome-wide association studies, multiple related phenotypes are available for testing for association with genetic variants. In most cases, however, these related phenotypes are analysed independently from one another. For example, several studies have measured multiple lipid-related phenotypes, such as LDL-cholesterol, HDL-cholesterol, and Triglycerides, but in most cases the primary analysis has been a simple univariate scan for each phenotype. This type of univariate analysis fails to make full use of potentially rich phenotypic data. While this observation is in some sense obvious, much less obvious is the right way to go about examining associations with multiple phenotypes. Common existing approaches include the use of methods such as MANOVA, canonical correlations, or Principal Components Analysis, to identify linear combinations of outcome that are associated with genetic variants. However, if such methods give a significant result, these associations are not always easy to interpret. Indeed the usual approach to explaining observed multivariate associations is to revert to univariate tests, which seems far from ideal. Here we outline a Bayesian approach to this problem based on model selection within Bayesian multivariate regression. We focus particularly on issues of prior specification and model averaging. The approach has close connections with previous work, including Bayesian Seemingly Unrelated Regressions, and Bayesian Graphical models. We illustrate the approach on examples, and illustrate how the method can improve both power and interpretation of association analyses. Although the motivation behind this talk is genetic association studies, the methods and discussion apply quite generally to the more general statistical problem of assessing association between multiple correlated outcomes and a randomised (binary) intervention.

**Bayesian Rose Trees**

by YEE WHYE TEH

Hierarchical structure is ubiquitous in data across many domains. There are many hierarchical clustering methods, frequently used by domain experts, which strive to discover this structure. However, most of these methods limit discoverable hierarchies to those with binary branching structure. This limitation, while computationally convenient, is often undesirable. In this paper we explore a Bayesian hierarchical clustering algorithm that can produce trees with arbitrary branching structure at each node, known as rose trees. We interpret these trees as mixtures over partitions of a data set, and use a computationally efficient, greedy agglomerative algorithm optimizing likelihood ratios to find a rose tree which have high marginal likelihood given the data. Lastly, we perform experiments which demonstrate that rose trees are better models of data than the typical binary trees returned by other hierarchical clustering algorithms.

**How can we deal with insecurity regarding choice of models?**

by NANNY WERMUTH

Sometimes, we hear that mathematical convenience leads to choosing a joint Gaussian distribution as an appropriate framework for analyzing data, sometimes, that only nonparametric approaches are appropriate in the case of model uncertainty. Both are extreme viewpoints, hence some alternatives are contemplated and discussed.

**Bayesian information based learning and majorization.**

by HENRY WYNN

A main theorem in Bayesian learning is that the expected increase in information from an experiment is non-negative. The results appears in various forms and with various levels of generality in the work of Lindley, Renyi, de Groot and others. We also know the class of informational functionals for which the result holds, which includes Shannon information as a special case. The class is closely related to the majorization or rearrangement partial order. The actual posterior information, as opposed the (preposterior) expected value can turn out to be smaller for the posterior than for the prior. This leads to a definition of strong learning as being when the posterior dominates the prior for all the special information functionals and partial learning when it dominates only for some. In some cases it is surprisingly difficult to characterize the ordering in terms of the parameters of the distribution. The theory can also be applied to the area of selective sampling and this leads to a wider discussion of sampling under uncertainty.

**Contributed Talks.****A method for variable-dimensional nonparametric monotonic regression.**

by ELJA ARJAS

In situations where it can be applied, an assumed monotonicity property of the regression function with respect to covariates has a strong stabilizing effect on the estimates. Here we formulate a mathematical construction for multidimensional nonparametric monotonic regression and propose a Bayesian procedure for its estimation. Our explicit construction the regression function postulates it to have a piecewise constant form, each realization being defined by a finite number of random support points at which the function is assigned a random level. This results in a flexible variable- dimensional modeling, where the main interest is in model averaged inference rather than in any single realization of the random regression function. As is well known, the Bayesian paradigm carries an implicit penalty for model complexity, and under vague priors the method usually results in a sparse characterization of the relationships between the considered

variables. The method is illustrated with a concrete example. (Based on joint work with Olli Saarela.)

### **Likelihood-free estimation of model evidence**

by XAVIER DIDELOT

Joint work with Richard Everitt, Adam M. Johansen, Daniel Lawson. Statistical methods of inference typically require the likelihood function to be computable in a reasonable amount of time. The class of likelihood-free methods termed Approximate Bayesian Computation (ABC) is able to eliminate this requirement, by evaluating the  $t$  of a parameter value according to its capacity to generate data similar to the one observed. Likelihood-free methods have gained in efficiency and popularity in the past few years, following their integration with Monte-Carlo Markov Chain (MCMC) and Sequential Monte-Carlo (SMC) in order to better explore the parameter space. They have been applied primarily to the estimation of the parameters of a given model, but can also be used to compare models. Here we present a novel likelihood-free approach to model comparison, which is based on the independent estimation of the evidence of each model under study. Key advantages of this approach over previous techniques are that it can easily be integrated with SMC, and that it does not require a sampler able to mix between models. We illustrate our method on a toy example for which the relative merits of each model are known exactly, and on a realistic example from population genetics.

### **Exploring ODE model uncertainty via diffusions, with application to physiological processes.**

by KOSTAS KALOGEROPOULOS

Ordinary differential equations models are extensively used to describe various continuous time phenomena. Although successful in capturing certain aspects of the underlying mechanism, the need for further refinement is becoming increasingly apparent. In many applications, including our motivating example from physiological processes, this may be attributed to the inherent system noise. In our approach the system noise is explicitly modelled and disentangled from other sources of error. This is achieved by incorporating a diffusive component while retaining the mean infinitesimal behavior of the system. In PK/PD applications, the problem of inference on the non linear diffusion parameters is further complicated by the presence of measurement error, individual variability, imbalanced designs and so forth. We present a general inference framework through data augmentation which we illustrate through simulated and real PK/PD data.

### **A stochastic Dollo-model for phylogenetic inference from binary trait data: model elaboration and checking.**

by GEOFF NICHOLLS

Nicholls and Gray (2008) describe a phylogenetic model for binary trait data. They use their model to estimate branching times on a Indo-European language tree from the lexical data of Ringe et al. (2002), dropping seven languages with missing data. We summarise this work and report three further developments. We give a model for rate heterogeneity in the trait evolution which adds a point process of evolutionary bursts to the background birth-death process. We show how to compute the likelihood when there are trait data missing at random. We fit the model to the full dataset, and check the fit using Bayes factors to measure support for known constraints on the phylogenetic branch lengths and subtree structures. We reject three of thirty historically attested constraints. This is joint work with Robin J. Ryder.

## Population-Based Reversible-Jump MCMC for Bayesian Variable Selection and Evaluation Under Cost- Limit Restrictions

by IOANNIS NTZOUFRAS

In this work we examine a variable-selection strategy for construction of a sickness scale in which predictive accuracy is optimised subject to a bound on cost. Conventional model search algorithms (such as those based on standard reversible-jump Markov chain Monte Carlo (RJMCMC) sampling) in our setting will often fail, because of the existence of multiple modes of the criterion function with movement paths that are forbidden due to the cost restriction. In this paper we develop a populationbased trans-dimensional RJ-MCMC (population RJMCMC) algorithm, in which ideas from the populationbased MCMC and simulated tempering algorithms are combined. Comparing our method with standard RJMCMC, we find that the population-based RJMCMC algorithm moves successfully and more efficiently between distant neighbourhoods of .good. models, achieves convergence faster and has smaller Monte Carlo standard errors for a given amount of CPU time. In a case study of  $n=2,532$  pneumonia patients on whom  $p=83$  sickness indicators were measured, with marginal costs varying from smallest to largest across the predictor variables by a factor of 20, the final model chosen by population RJMCMC, both on the basis of highest posterior probability and specifying the median probability model, was clinically sensible for pneumonia patients and achieved good predictive ability while considering data collection costs.

## Relating Models to Reality.

by ANTHONY O'HAGAN

Statistical models typically assume that observations equal a known function of covariates and parameters (the model function), plus observation errors. There is an implicit assumption that either (a) the model function perfectly captures the underlying reality, or else (b) any model imperfection is subsumed in the observation error. But (a) is generally untenable, and the modelling of the error term rarely justifies (b). The need to represent separately the relationship between model and reality has become central to the field of uncertainty in computer models, because there we effectively have no control over the model function. Lessons can be learnt for statistical modelling generally, particularly in the context of model uncertainty.

## Guiding Bayesian model choice using ABC, with an application to rapidly evolving infectious diseases

by OLIVER RATMANN

Model choice is a cardinal part of data analysis, and has been extensively discussed for scenarios in which the likelihood  $f(x_0|\theta, M)$  of a scientific model  $M$  remains readily evaluable. In many areas of modern science, we now face increasingly often quantities of data for which  $f(x_0|\theta, M)$  cannot be readily evaluated. Approximate Bayesian Computation (ABC) is a developing area of statistics that is motivated by the need to fit, assess and compare models that have a well-defined underlying generative process, but for which classical statistical procedures of estimation and goodness-of-fit encounter formidable computational challenges. ABC circumvents likelihood calculations by comparing model simulations with the data in terms of summary statistics.// To aid Bayesian model choice in this context, methods for model criticism and for model comparison have been introduced. We place these tools within the context of available Bayesian methodology, highlight their limitations and unique opportunities to Bayesian analysis, and propose new methods to guide model choice when it is only feasible to simulate from  $f(\cdot|\theta, M)$ .// Model predictions have long been used to inform model choice; hence, ABC can be naturally embedded in a large array of available Bayesian methods. We explored the application of *sequential Bayesian tools* that comprise a training and testing phase, such as (approximate) posterior or partial predictive densities for model criticism, or posterior, pseudo or intrinsic Bayes' factors for model comparison when the likelihood cannot be evaluated. The extra volatility induced by

simulating data sets typically offsets any gains in precision during the training phase, and we argue for using so-called simultaneous methods [1].// ABC enables us to check several aspects of the data, and thereby opens up multi-dimensional avenues towards more accurate model choice [2]. We discuss in detail the power of multivariate predictive error densities for model criticism, as well as the accuracy of mutually constrained approximate Bayes' factors (aBF's) for model comparison. We also present aBF's for the purpose of multi-dimensional model criticism that simplify to the Savage-Dickey ratio and do not require the computation of the marginal likelihood.// The second part of the talk considers the application of ABC to the study of rapidly evolving infectious diseases such as influenza and HIV. Crucially, ABC makes possible to cross all the relevant biological scales involved (genetics, immunogenetics, ecology) when the summary statistics are suitably chosen. We will demonstrate and discuss the application of the aforementioned tools for model criticism and model choice to compare alternative hypotheses on the nature of immune escape and various forms of acquired immunity in influenza-A dynamics.// We conclude that, based on the above multi-dimensional techniques, we may check and compare complicated models even if aspects thereof remain unidentifiable. Moreover, these methods are particularly appealing in practice as they incur little or no extra computational cost compared to ABC. Joint work with Katia Koelle (Duke, USA), Christophe Andrieu (Bristol, UK), Sylvia Richardson (Imperial College, UK). // References [1] Ratmann et al. Model criticism based on likelihood-free inference, with an application to protein network evolution. PNAS USA, 106(26):10576-10581, 2009. [2] Ratmann et al. Reply to Robert et al.: Model criticism informs model choice and model comparison. PNAS USA, 107(3):E6-7, 2010.

### **Bayes Factor Consistency for ANOVA and Linear Mixed Models.**

by DONGCHU SUN

We first consider the two versions of the one-way ANOVA model that exist in literature, fixed and random effects models. The difference of these two models lies in whether or not the effects of the factor levels are treated as random variables. From the Bayesian point of view, all parameters are considered as random variables, making the distinction between fixed effect model and random effect model rather obscure. The primary goal of this article is to present a unified Bayesian approach to deal with one-way ANOVA models with fixed effects or random effects. We propose a modification of the Zellner-Siow prior, and show that the proposed prior will result good consistency properties in terms of model selection and posterior distributions in the settings of either fixed or random effects models. The situation for a general ANOVA and linear mixed model will be explored.

### **Unifying decision theoretic review of Bayesian predictive model selection.**

by AKI VEHTARI

We provide a unified review of Bayesian model selection methods based on how they approximate the full Bayesian decision theoretic approach of estimating the future predictive performance of the model

$$\mathbb{E}[u] = \iint u(\tilde{y}, p(\tilde{y}|\tilde{x}, D)) p(\tilde{y}|\tilde{x}) p(\tilde{x}) d\tilde{y} d\tilde{x},$$

where  $p(\tilde{y}|\tilde{x}, D)$  is the predictive distribution given the observed data  $D$ ,  $\tilde{y}$  is a future observation and  $\tilde{x}$  a future observation or known fixed quantity. The full approach requires 1) approximating the conditional distribution of  $\tilde{y}$ , 2) approximating the unconditional distribution of  $\tilde{x}$ , 3) integrating over  $\tilde{y}$ , 4) integrating over  $\tilde{x}$ , and 5) integrating over the posterior distribution to obtain the predictive distribution.// We describe prediction scenarios with different forms for  $u$ ,  $p(\tilde{y}|\tilde{x})$ , and  $p(\tilde{x})$  which affect the actual decision problem in model selection. For each predictive model selection method (several methods including word *predictive* in their name and couple others) we describe the explicit and implicit probabilistic and decisional assumptions made, and

how different approximations for 1–5 are combined and how these approximations affect the bias and variance of the predictive performance estimates. Joint work with Janne Ojanen.

**Uncertainty in ensembles of deterministic models with application to probabilistic weather forecasting.**

by JIM ZIDEK

This talk will provide an extension of Bayesian melding . The extension can combine measurements with outputs (simulated data) from an ensemble of deterministic models to determine model bias. Or it can combine the simulated data alone and thereby provide an integrated output with probabilistic estimates of model uncertainty. The model outputs can be on different scales. We apply the Bayesian ensemble melding model to the sea level temperature data at Pacific Northwest area. We combine the measurements with model outputs from an ensemble of five deterministic models for spatial predictions. The predictions of Bayesian ensemble melding model are compared with those of averaging model outputs and Kriging approach. We also show how this purely spatial approach can be turned into a 48 hour ahead probabilistic weather forecaster.

**Poster Session (16.30–19.00, Mon 31 May 2010)**

**An alternative marginal likelihood estimator: a comparative study.**

by SERENA ARIMA

The estimation of normalizing constants of unnormalized probability densities such as marginal likelihoods is a crucial point for Bayesian model selection where the ratio of marginal likelihoods of two competing models defines the well-known Bayes Factor (Kass and Raftery, 1995). Although the problem has received a lot of attention in the literature (see for example Chib (1995), DiCiccio et al. (1997), Gelman and Meng (1998)), the computation of normalizing constants continues to be a formidable challenge in many complex models. Currently, one of the most favorite approach takes advantage of Monte Carlo (MC) and Markov Chain Monte Carlo (MCMC) methods and attempts to estimate this unknown quantity through simulations from the target density or suitably defined auxiliary/importance densities. We consider an alternative marginal likelihood estimator, named In ated Density Ratio estimator (IDR), originally proposed in Petris and Tardella (2007). The IDR estimator is an alternative implementation of the Generalized Harmonic Mean (GHM) (Newton and Raftery, 1994; Raftery et al., 2007). In particular, the IDR method similarly to the original Harmonic Mean Estimator recycles posterior simulation but relies on a different choice of the importance function, defined through a parametrically in ated version of the target density. We use some benchmark simulation plans and some complex models for real data in the field of system biology to explore comparative merits and limits of the IDR approach. In particular we will focus on marginal likelihood estimation for multivariate and multimodal distributions, and posterior simulations for phylogenetic models (Arima and Tardella, 2010). We show some improvements of the original IDR estimator and compare the performance of IDR with the marginal likelihood estimators recently proposed in Ardia et al. (2009) and Calderhead and Girolami (2009). Joint work with Giovanni Petris, Luca

**Comparison of deterministic and stochastic approaches to Bayesian variable selection.**

by ELENİ BAKRA

Reversible jump Markov chain Monte Carlo (RJMCMC) is widely use in the challenging problem of Bayesian model selection. The performance of the RJMCMC algorithm depends to a large extend on a careful tuning of the proposal distribution. We explore, in the context of Bayesian

variable selection, whether deterministic approximations based on the Laplace approximation and on the Bayesian information criterion yield useful as estimates of posterior model probabilities. We compare these to estimates based on versions of RJMCMC where the deterministic estimators are used as proposal distributions. Our conclusion in this context, is generally speaking, that posterior model probability estimates based on Laplace approximations are comparable to RJMCMC estimates, but without the associated additional computational effort.

### **Bayesian compressive sensing for high-dimensional data with model uncertainty.**

by ANJISHNU BANERJEE

Compressive sensing (CS) is a recently proposed methodology in which one randomly projects high-dimensional data to a lower dimensional linear subspace in an online manner. This enables a dramatic reduction in dimensionality, facilitating storage and data processing. In addition, there is a rich theoretical literature providing guarantees of accurate reconstruction of the original signal under sparsity assumptions. CS has been very broadly applied in signal processing, imaging and video applications but essentially all of the literature focuses on using the CS measurements to obtain a point estimate of the reconstructed signal that has good frequentist properties. We instead consider using CS-type random projections to obtain accurate approximations to the full posterior distribution of parameters and functions of interest. We provide bounds on the accuracy of the approximations based on the Kullback-Leibler (KL) divergence between the approximated and true posterior, with the accuracy depending on the number of CS measurements and the true sparsity level. We initially apply the CS approximation to functional data analysis applications in which there is interest in estimating the posterior distribution of a curve allowing for uncertainty in the basis functions having non-zero coefficients. It is shown that the CS method can lead to substantial efficiency gains over the common strategy of focusing on a random subsample of the data. In addition, remarkably accurate inferences can be obtained using a small fraction of the original data under reasonable sparsity assumptions. We also modify the methodology to consider applications to massive-dimensional variable selection, with the motivating application being genetic studies collecting massive numbers of gene expression measurements and/or single nucleotide polymorphisms (SNPs). It is shown that accurate predictive performance can be obtained based on applying CS to the high-dimensional predictors, leading to a dramatic gain in computational efficiency. These results are quite promising in terms of using CS as a general tool for solving large  $p$ , small  $n$  and large  $p$ , large  $n$  problems that are now commonplace in many application areas. Key Words: Basis selection; Correlated predictors; Functional data; Large  $p$ , small  $n$ ; Random projections; Variable selection.

### **Some contributions to the Bayesian nonparametric analysis of species sampling models under Gibbs priors.**

by ANNALISA CERQUETTI

In the species sampling problem a random sample is drawn from an hypothetically infinite population of individuals to make inference on the unknown total number of different species. A full Bayesian nonparametric analysis of this problem has been first proposed in Lijoi et al. (2007, 2008) to derive posterior predictive inference on species richness of an additional sample by treating the multiplicities of the different species observed as a sample from an exchangeable Gibbs partition of type  $\alpha \in [0, 1)$ . In Favaro et al. (2009) completely explicit estimators for conditional species richness under two parameter Poisson-Dirichlet model and its asymptotic distribution have been obtained to deal with the problem of prediction for very large sizes of the additional sample. Here we present two different contributions to this topic. First, moving from Lijoi et al. (2008), but adopting the rephrasing and the groups sequential construction in Cerquetti (2009), we derive a Bayesian nonparametric analysis for conditional species richness under a partition model in Gibbs form of type  $\alpha = -1$  recently introduced in Gnedin (2009) and characterized by a finite but random number of different species. Then, we provide an

alternative approach to the Bayesian finite sample and asymptotic analysis of the two-parameter Poisson-Dirichlet partition model by exploiting a Pitman's characterization by *deletion of classes* property and known results for Beta-Binomial distributions. We obtain simplified and more direct proofs for the estimators of conditional species richness and an alternative scale mixture representation for its asymptotic distribution.

### **Process capability Indices in the case of more than one variance component.**

by DELSON CHIKOBVU

Data arising from multiple sources of variability are very common in practice. Virtually all industrial processes exhibit between-batch, as well as within-batch components of variation. In some cases the between-batch (or between subgroup) component is viewed as part of the common-cause-system for the process. It seems worthwhile to develop a process capability index in even more general settings. In this paper, a three variance component model is considered. A version of the process capability or performance index for the balanced random effects model with three variance components from a Bayesian framework is considered. The index can be used for average of observations for a given time period. In a process to manufacture chronic medication, various properties of the manufactured tablet have to be monitored. Monthly samples of  $J=8$  packages of the tablet are sampled and various physical properties of the tablet are replicated in the laboratory by analysing  $K=5$  tablets per package. The data in Table 1 represents the amount of drug in a tablet (the percentage of the drug per tablet). Packages with tablets sampled for the first  $I=15$  months starting January of a particular year are selected as review data to determine whether the patient gets on average the required dosage of the drug from the batches in a specified time, given that each patient must get an average dosage of at least 20%. A probability matching prior is also derived for the index and a weighted Monte Carlo procedure is used to calculate credibility intervals.

### **A Bayesian model averaging approach with non-informative priors for cost-effectiveness analyses.**

by CATERINA CONIGLIANI

We consider the problem of assessing new and existing technologies for their cost-effectiveness in the case where data on both costs and effects are available from a clinical trial, and we address it by means of the cost-effectiveness acceptability curve. The main difficulty in these analyses is that cost data usually exhibit highly skew and heavy-tailed distributions, so that it can be extremely difficult to produce realistic probabilistic models for the underlying population distribution. Here, in order to integrate the uncertainty about the model into the analysis of cost data and into cost-effectiveness analyses, we consider an approach based on Bayesian model averaging in the particular case of weak prior informations about the unknown parameters of the different models involved in the procedure. The main consequence of this assumption is that the marginal densities required by Bayesian model averaging are undetermined. However, in accordance with the theory of partial Bayes factors and in particular of fractional Bayes factors, we suggest replacing each marginal density with a ratio of integrals that can be efficiently computed via Path Sampling.

### **Searching for models with high posterior probability using constraint integer programming.**

by JAMES CUSSENS

I consider the case when a statistical model is made up of 'components' such that its log posterior probability is a linear combination of scores of these components. This is the case for some graphical models (where the components are the parent sets for each variable) and also some Bayesian approaches to clustering (where the components are the clusters). For problems up

to a certain size, this permits the model selection problem to be recast as a constraint integer programming (CIP) problem, thus allowing state-of-the-art CIP solvers to be applied. Moreover, once the MAP model has been found the solver can be restarted to find other models in order of decreasing probability. The end result is a search-based approach to exploring a posterior distribution which provides an alternative to MCMC.

### **Model selection in dynamic factor models.**

by SHUTONG DING

Sune Karlsson Dynamic factor models have become popular in applied macroeconomics with the increased availability of large data sets. We consider model specification, i.e. the choice of lag length and the number of factors, in the setting of factor augmented VAR models. In addition to the standard Bayesian approach based on Bayes factors and marginal likelihood, we also study model choice based on the predictive likelihood which is particularly appealing in a forecasting context. As a benchmark we compare the performance of the Bayesian procedures with frequentists approaches, such as the factor selection method of Bai and Ng.

### **Bayesian variable selection using cost-adjusted BIC, with application to cost-effective measurement of quality of health care.**

by DIMITRIS FOUSKAKIS

In the field of quality of health care measurement, one approach to assessing patient sickness at admission involves a logistic regression of mortality within 30 days of admission on a fairly large number of sickness indicators (on the order of 100) to construct a sickness scale, employing classical variable selection methods to find an “optimal” subset of indicators. Such “benefit-only” methods ignore the considerable differences among the sickness indicators in cost of data collection, an issue that is crucial when admission sickness is used to drive programs (now implemented or under consideration in several countries, including the U.S. and U.K.) that attempt to identify substandard hospitals by comparing observed and expected mortality rates (given admission sickness). When both data-collection cost and accuracy of prediction of 30-day mortality are considered, a large variable selection problem arises in which costly variables that do not predict well enough should be omitted from the final scale. In this work (a) we develop a method for solving this problem based on posterior model odds, arising from a prior distribution that (1) accounts for the cost of each variable and (2) results in a set of posterior model probabilities that corresponds to a generalized cost-adjusted version of the Bayesian information criterion (BIC), and (b) we compare this method with a decision theoretic cost-benefit approach based on maximizing expected utility. We use reversible-jump Markov chain Monte Carlo (RJMCMC) methods to search the model space, and we check the stability of our findings with two variants of the MCMC model composition (MC3) algorithm. We find substantial agreement between the decision-theoretic and cost-adjusted-BIC methods; the latter provides a principled approach to performing a cost-benefit tradeoff that avoids ambiguities in identification of an appropriate utility structure. Our cost-benefit approach results in a set of models with a noticeable reduction in cost and dimensionality, and only a minor decrease in predictive performance, when compared with models arising from benefit-only analyses.

### **A Bayesian nonparametric approach to the analysis of developmental toxicology data.**

by KASSANDRA FRONCZYK

We develop a Bayesian nonparametric mixture modeling framework for replicated count responses in dose-response settings. We explore this methodology with applications in developmental toxicity studies, in which the primary objective is to determine the relationship between the level of exposure to a toxic chemical and the probability of a physiological or biochemical

response, or death. Data from these experiments typically involve features that can not be captured by standard parametric approaches. To provide flexibility in the functional form of both the response distribution and the probability of positive response, the proposed mixture model is built from a dependent Dirichlet process prior, with the dependence of the mixing distributions governed by the dose level. The methodology is tested with a simulation study, which involves also comparison with semiparametric Bayesian approaches to highlight the practical utility of the dependent Dirichlet process nonparametric mixture model. Further illustration is provided through the analysis of data from two developmental toxicity studies.

### **Latent class based multiple imputation for missing categorical data.**

by MULUGETA GEBREGZIABHER

In this study we propose a latent class multiple imputation (LCMI) approach for analyzing missing categorical covariate data. In the first stage, we impute the missing data using an MCMC based latent class imputer's model and in the second stage, we use likelihood methods to make inference on the association between the outcome and the missing categorical covariate data. We demonstrate its application using two data examples, missing covariate data in an individually matched case-control study of multiple myeloma and a type-2 diabetes study with missing race data. Simulation study is used to investigate the statistical properties of LCMI and comparisons with a frequentist bootstrap based LCMI (Vermunt et al 2008) and other existing methods are made. This is done under all possible missing data mechanisms (including missing completely at random, missing at random and not missing at random). The procedures are compared with respect to bias, asymptotic standard error, type I error, and 95% coverage probabilities of parameter estimates. Simulations show that, under many missingness scenarios, the MCMC based LCMI performs favorably when jointly considering these criteria.

### **Application of Bayesian model uncertainty in RC(M1, M2) association model.**

by S.K., GHOREISHI

In this paper, we briefly discuss the model uncertainty for RC(M1, M2) association models with orthogonal polynomial components. We show how Bayesian model averaging can produce an improvement tool to solve model uncertainty in contingency tables. We evaluate the results of the paper by comparison of association between two contingency tables with the same categorical variables. Key words: Association model, Bayesian model averaging, orthogonal polynomial components.

### **Joint ensembles from separate marginals using Gaussian mixtures: application to irregular time series.**

by JOHN HASLETT

We consider a possibly multivariate stochastic process in continuous time  $Y(t)$ , observed indirectly at irregular points in time  $t_i; i = 1, \dots, n$ . The process is known to be relatively smooth, but with occasional large increments. The observation process involves data  $x = \{x_i; i = 1, \dots, n\}$ , such that we can state posteriors  $\pi(y(t_i)|x_i)$  separately for each  $t_i$ . The ultimate task is to draw joint samples from  $\pi(Y(t)|x)$  given this information. The application is to the palaeoclimate, observed indirectly via proxies in a core of sediment or ice; see for example Haslett et al (2006). Proxies include multivariate pollen counts and/or isotope ratios. The samples are at known depths  $d_i$  — often regular — corresponding to irregular times  $t_i$ . Models calibrated elsewhere permit separate uncertain inference on each  $y(t_i)$  given  $x_i$ . Initial interest focusses on joint inference on  $y = \{y(t_i); i = 1, \dots, n\}$  given  $x$ . Subsequent interest is on regular samples  $\tilde{y} = \{y(t_0 + k\delta); k = 0, \dots, K\}$  for suitable  $K$ . We refer to these latter as ensembles. There is an additional difficulty in that inference on the times  $t_i$ , given the  $d_i$ , is also uncertain. We defer detailed discussion to elsewhere; see Haslett and Parnell (2008) and Parnell and Haslett (2008).

We propose a fast algorithm based on Gaussian mixtures. The  $Y(t)$  process is modelled as an independent increments process with Normal Inverse Gaussian priors. This is a continuous scale mixture of Gaussian random variables. Being long-tailed the model accommodates processes with occasional large increments. Being infinitely divisible, it permits arbitrary interpolations such as are required by  $\tilde{y}$ . The separate posteriors  $\pi(y(t_i)|x_i)$  are approximated by a finite mixture of Gaussian densities. Jointly, these lead to much analytic marginalisation and thus to fast algorithms.

### **Structural uncertainties in health economic decision models.**

by CHRISTOPHER JACKSON

Health economic models compare the cost-effectiveness of medical interventions over the long term. They are used by policy-makers to make decisions about funding, and aim to synthesise all evidence concerning the interventions and the natural history of the patients of interest. They are typically discrete-time Markov models, where each transition probability may be informed by a different dataset. Probabilistic, often fully Bayesian, methods for taking account of parameter uncertainty in these models are well-established. But uncertainty about structural assumptions is generally only described informally, by presenting cost-effectiveness under alternative scenarios. In our work we have addressed statistical model uncertainties, such as the shape of survival curves and the choice of predictors of important events, by model comparison and model averaging. We favour criteria which judge models by their predictive ability, such as DIC and cross-validation, over prior and posterior model probabilities, since the "true" model is likely to be much more complex than any of those considered. Methods which consider the model space as continuous, such as Bayesian semiparametrics, are more flexible than model averaging, though prior choice can be difficult. Expert elicitation is also required to quantify many uncertainties in economic models, such as extrapolations into an unknown future, which have no data to inform them. We illustrate applications to implantable defibrillators, screening for oral cancer and biologic drugs for psoriatic arthritis.

### **AIS(RJ): A trans-dimensional MCMC algorithm.**

by GEORGIOS KARAGIANNIS

One of the main difficulty in variable dimension statistical problems is how to compute the marginal posterior model probabilities. We have developed a within model simulation method called AIS which is based on the Annealing Importance Sampler in the context of the variable dimension problems and allows us to estimate marginal posterior model probability ratios, unbiasedly. Additionally, we present an across model simulation algorithm called AISRJ which is based on the popular reversible jump (RJ) algorithm. AISRJ is a pseudo-marginal MCMC algorithm, a class of algorithms discussed in (Andrieu and Roberts, 2009), which approximates an 'Ideal RJ' algorithm which can be considered as optimal under very weak conditions. We show that AISRJ improves the performance of the standard RJ algorithm in terms of acceptance probability without the need to use very sophisticated moves. The two methods rely on the idea of bridging the models in a specific way in order to ensure good performance. The performance of the proposed algorithms is illustrated on a Bayesian model choice example with generalised linear mixed effect models. Keywords: Variable dimension statistical problems, pseudo-marginal MCMC, reversible jump MCMC, Bayesian statistics.

### **Ecological Model Selection using hybrid Particle and MCMC methods**

by JAMES LAWRENCE

Reversible Jump MCMC and Particle filters are two established model selection (and averaging) techniques. We explore their application to determining the density-dependence structure of natural species, as well as the performance of more advanced hybrid methods, which combine

these as well as more recent particle-based schemes, such as the Particle Learning framework of Polson, Johannes et al.

### **A hierarchical Bayesian framework for constructing sparsity-inducing priors.**

by ANTHONY LEE

Variable selection techniques have become increasingly popular amongst statisticians due to an increased number of regression and classification applications involving high-dimensional data where we expect some predictors to be unimportant. In this context, Bayesian variable selection techniques involving Markov chain Monte Carlo exploration of the posterior distribution over models can be prohibitively computationally expensive and so there has been attention paid to quasi-Bayesian approaches such as maximum a posteriori (MAP) estimation using priors that induce sparsity in such estimates. We focus on this latter approach, expanding on the hierarchies proposed to date to provide a Bayesian interpretation and generalization of state-of-the-art penalized optimization approaches from the signal processing and statistics literature and providing simultaneously a natural way to include prior information about parameters within this framework. We give examples of how to use this hierarchy to compute MAP estimates for linear and logistic regression as well as sparse precision-matrix estimates in Gaussian graphical models.

### **A Bayesian mixture model for detecting unusual time trends in small area estimates: application to COPD mortality in England.**

by GUANGQUAN LI

Space-time modeling of small area data is often used in epidemiology for mapping chronic disease rates and by government statistical agencies for producing local estimates of, for example, unemployment or crime rates. While at any given time point, the spatial variations can be attributable to the differences in the distribution of predictors such as socio-demographic and environmental risk factors, the temporal changes in most local areas tend to resemble each other closely. However, some may exhibit unexpected changes over time, suggesting, e.g., the emergence of localized predictors/risk factor(s) or a change in diagnostic or treatment techniques or the impact of a new policy. Detection of areas with "unusual" temporal patterns is of importance for several reasons, including highlighting areas deserving of further scrutiny, identifying possible predictors/risk factors, or assessing the effectiveness of an intervention or a policy. In this paper, we propose a novel Bayesian mixture model for short time series of small area data that provides estimates of both the common temporal trend and the area-specific temporal trends. For each area, the posterior probability of belonging to the area-specific versus the common trend is used to classify the local time trend as "unusual" or not. To examine the detection ability, we have constructed a comprehensive simulation study, in which the proposed model shows consistently good performance in detecting various departure patterns seen in real world situations. In general, the area under the ROC curve exceeds 0.8 with 1.0 representing a perfect classification whereas 0.5 indicating a worthless test. For real data applications, where the common/unusual status of the areas is unknown, we have developed a simulation-based approach to estimate the false discovery rates (FDR) and thus to determine area-specific cutoff values for the classification. Standard methods (e.g., [1,2]) for FDR estimation are not applicable here as each area has its own "alternative" trend, which violates the identical assumption made in multiple testing. We have applied our model to a retrospective study of chronic obstructive pulmonary disease (COPD) mortality data in England (1990-1997) to assess the impact of government policy to make COPD a "compensatable" disease under the Industrial Injuries Disablement Benefit scheme in 1992, which would be expected to have greatest impact in mining areas. Inspection of the identified unusual trends reveals some interesting "stories" of the corresponding areas, which will be detailed in the paper.

**On Bayesian variable selection using Lasso and related methods.**

by ANASTASIA LYKOU

We propose a Bayesian method that accomplishes both shrinkage and variable selection by using the Lasso (Tibshirani, 1996) and related methods. The proposed method takes advantages of the shrinkage methods, which can reduce efficiently the prediction error avoiding problems appearing when collinearity is present and indirectly implies which variables can be removed from the model. Similar approaches can be found in the literature by Park and Casella (2008) and Hans (2009). Among the challenges is to choose the appropriate shrinkage parameter, which controls the shrinkage applied and the selected variables. Here we explore the sensitivity of the Bayes factor on this choice. We graphically represent this sensitivity on plots similar to the ones used in standard Lasso methods. We facilitate the univariate regressions and their Bayes factors (when compared to the null model) to tune and specify hyperpriors for the shrinkage parameter. Hans, C. (2009). Bayesian Lasso regression. *Biometrika*, 96(4):835-845. Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103(482):681-687. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc. B.*, 58:267-288.

**Bayesian non-parametric signal extraction for Gaussian time series.**

by CHRISTIAN MACARO

We consider the problem of unobserved components in time series from a Bayesian non-parametric perspective. The identification conditions are treated as unknown and analyzed in a probabilistic framework. In particular, informative prior distributions force the spectral decomposition to be in an identifiable region. Then, the likelihood function adapts the prior decompositions to the data. A full Bayesian analysis of unobserved components will be presented for financial high frequency data. Particularly, a three component model (long-term, intra-daily and short-term) will be analyzed to emphasize the importance and the potential of this work when dealing with the Value-at-Risk analysis. A second astronomical application will show how to deal with multiple periodicities.

**A Bayesian Factor model for gene expression detection on oligonucleotide microarrays.**

by VINICIUS DINIZ MAYRINK

A problem of scientific interest is to determine whether a gene is expressed or not in a given probe set of a microarray. Different methods for detection of transcripts have been proposed for the different microarray platforms. A popular detection method for Affymetrix GeneChip oligonucleotide arrays is implemented as part of the preprocessing algorithm MAS 5.0. In short, the one-sided Wilcoxon's signed rank test is applied to obtain p-values which are used to assign detection calls. This presence-absence (P/A) method works relatively well in practice; however, other frameworks improving this solution can be found in the literature. An interesting example is the method named Presence Absence calls with Negative Probe sets (PANP). In brief, the raw data set is first preprocessed using any technique, then selected negative strand matching probe sets are assumed as controls and the empirical c.d.f. of their intensities is used to derive a cutoff value. Its simplicity and improved results with respect to MAS 5.0 P/A procedure make PANP an attractive method; however, defining thresholds based on an empirical c.d.f. seems an arbitrary decision. When several microarrays replicating the same experiment for different samples are available, the analysis of the pattern of expressions across samples is an important source of information for a detection method. No pattern can be observed in a noise probe set, i.e., all probes randomly alternate intensities across samples. On the other hand, a probe set classified as present will exhibit a consistent increasing or decreasing expression sequence across samples for every probe. This result motivates the Bayesian Factor Model (BFM) proposed

in this study. The factor loadings reflect the strength of hybridization between the target of the probe set and each probe, and the factor scores describe the pattern of expression of the target across samples. Within the Bayesian framework, a mixture prior is attributed to the vector of factor loadings; two Normal components are used, and one of them is a distribution centered on zero with a small variance. The detection call is based on the posterior probability that the loadings differ from zero. A simulated data set is used to compare methods, and the results indicate that the BFM outperforms MAS 5.0 P/A and PANP in terms of performance for detecting the true presence or absence of the transcript. Real data sets are also evaluated. The amount of data contained in a microarray represents a challenge in terms of computational cost for MCMC algorithms; therefore, some strategies are considered as alternatives for improving the computational performance.

### **Dynamic logistic regression and dynamic model averaging for binary classification.**

by TYLER HARRIS MCCORMICK

We propose an online binary classification procedure for cases when parameters within a model are expected to change over time either as gradual drift or as sudden spikes. We account for model uncertainty through Bayesian Model Averaging and, for maximum flexibility, allow posterior model probabilities to also change with time. We accomplish this by applying a state-space model to the parameters of each model and a Markov chain model to the correct model, allowing our “correct” model to change over time. Our model accommodates different levels of change in the data-generating mechanism through tuning a “forgetting” factor. A novel tuning algorithm which adjusts the level of forgetting in a completely online fashion using the posterior predictive distribution allows the model to accommodate various levels of change in the data-generating mechanism at different times. We apply our method to data from children with appendicitis who receive either a traditional (open) appendectomy or a laparoscopic procedure. This project is joint work with David Madigan, Adrian Raftery, and Randall Burd.

### **Autoregressive process modeling via the Lasso procedure.**

by YUVAL NARDI

The Lasso is a popular model selection and estimation procedure for linear models that enjoys nice theoretical properties. In this paper, we study the Lasso estimator for fitting autoregressive time series models. We adopt a double asymptotic framework where the maximal lag may increase with the sample size. We derive theoretical results establishing various types of consistency. In particular, we derive conditions under which the Lasso estimator for the autoregressive coefficients is model selection consistent, estimation consistent and prediction consistent. Simulation study results are reported.

### **A construction of vectors of dependent Dirichlet processes.**

by BERNARDO NIPOTI

We present the construction of two dependent Dirichlet processes (DPs) that we obtain by normalizing the increments of two dependent gamma completely random measures (CRMs). These are constructed as linear functionals of dependent Poisson random measures introduced in Griffiths and Milne (1978). If  $p_0$ ,  $p_1$  and  $p_2$  are independent Dirichlet processes with suitable parameters, such a construction gives rise to dependent DPs,  $\tilde{p}_1$  and  $\tilde{p}_2$ , with the property that  $\tilde{p}_1$  is a mixture of  $p_0$  and  $p_1$  while  $\tilde{p}_2$  is a mixture of  $p_0$  and  $p_2$ . The vector of CRMs that we obtain is analytically tractable and the dependence structure between  $\tilde{p}_1$  and  $\tilde{p}_2$  can be described in terms of a parameter in  $[0, 1]$ . Given two exchangeable samples from  $\tilde{p}_1$  and  $\tilde{p}_2$ , respectively, a posterior estimation of such a parameter suggests how distant we are from the two extreme situations of full exchangeability ( $\tilde{p}_1 = \tilde{p}_2 = p_0$ ) and partial exchangeability ( $\tilde{p}_1 = p_1$  and  $\tilde{p}_2 = p_2$ ).

The vector  $(\tilde{p}_1, \tilde{p}_2)$  may be also used to construct dependent mixture models. An extension of the well-known Blackwell-MacQueen sampling scheme allows one to implement a Gibbs sampler and achieve a full Bayesian analysis for clustering and density estimation. This approach also leads one to construct  $k \geq 2$  dependent DPs and more general dependent random probability measures for Bayesian nonparametric inference.

### **Practical implementation of Bayesian predictive model selection in Gaussian process regression.**

by JANNE OJANEN

Bayesian model selection based on the decision-theoretic approach of estimating the future predictive performance of the candidate model requires a utility function and a distribution for the future data. In Bayesian predictive model selection criterion the distribution of the future data is replaced with the Bayesian model average (BMA) predictive distribution describing best our understanding of the uncertainty in the prediction problem. We describe a practical implementation for Bayesian predictive model selection through a series of analytical approximations and simulations in an input variable selection for a robust GP regression model where closed form solutions of the posterior distributions are not available.

### **A Reversible Jump method for generalised linear models with an extension to generalised linear mixed models.**

by ANTONY OVERSTALL

The major difficulty in the efficient implementation of the reversible jump algorithm (Green; 1995, *Biometrika*) for Bayesian model determination is in the joint specification of the mapping function and the proposal distribution. An automatic joint specification of the mapping function and proposal distribution for model determination among generalised linear models is proposed, which is based on orthogonal projections of the current linear predictor. An extension to model determination among generalised linear mixed models is described.

### **A proposal construction for Reversible Jumps MCMC.**

by MICHAEL PAPATHOMAS , PETROS DELLAPORTAS, VASSILIS VASDEKIS

A major difficulty when implementing the reversible jump Markov chain Monte Carlo methodology lies in the choice of good proposals for the parameters of the competing statistical models. We propose a general methodology to construct proposal densities in reversible jump MCMC algorithms so that consistent mappings across competing models are achieved. Our method is based on equating the likelihood under the current model to the expected likelihood under the destination model, the expectation taken with respect to the proposal distribution. Unlike nearly all previous approaches our proposals are not restricted to operate to moves between local models, but they are applicable even to models that do not share any common parameters. We focus on general linear regression models and produce concrete guidelines on proposal choices for moves between any models. These guidelines can be immediately applied to any regression model after applying some standard data transformations to near-normality. We discuss differences with other established methods, and the issue of equating, on average, posterior parameter distributions rather than likelihoods. We illustrate our methodology by providing concrete guidelines for model determination problems in logistic regression and log-linear graphical models. Two real data analyses illustrate how our suggested proposal densities, together with the resulting freedom to propose moves between any models, improve the mixing of the reversible jump Metropolis algorithm.

## **The Bayesian Elastic Net and Related Methods**

by MD. HASINUR RAHAMAN KHAN

The Bayesian Elastic Net is a relatively new method for identifying relevant explanatory variables, particularly in situations with very many correlated covariates as are typically found in gene expression data. We examine the Bayesian Elastic Net and its relationship with other methods.

## **Bayesian approaches to the analysis of spectral metabolic profiling data.**

by MATTIAS RANTALAINEN

Metabolic profiling using Nuclear Magnetic Resonance (NMR) spectroscopy is a well-established methodology for quantitative measurement of metabolite abundances in biological samples. NMR spectroscopy enables concurrent measurement of hundreds of metabolites in a sample, represented as peaks in a spectral profile. In recent years metabolic profiling has gained interest for applications in genomic epidemiology, where changes in the metabolic profile in response to a pathophysiological state are studied in an epidemiological context. Metabolic profiling in genomic epidemiology has the potential to enable novel biomarker discovery as well as providing a tool for disease diagnostics in humans. A valuable property of metabolic profiling is that samples can be collected with minimally invasive procedures, making the technology potentially applicable in a wide range of areas of both basic and clinical research. Here we describe how Bayesian methods can be used for analysis of spectral data originating from NMR spectroscopy. Modelling of spectral NMR-based metabolic profiling data poses numerous statistical challenges including spectral alignment, peak extraction, normalization, handling of structured noise and the presence of outlying observations. Improved inference of the components of variation in metabolic profiling data can improve the information recovery from the spectral profiles, provide means for biomarker discovery and improve the understanding of metabolic responses relating to disease.

## **A graphical diagnostic for identifying influential model choices in Bayesian hierarchical models.**

by IDA SCHEEL

Real-world phenomena are frequently modelled by Bayesian hierarchical models. The building-blocks in such models are the distribution of each variable conditional on parent and/or neighbour variables in the graph. The specifications of centre and spread of these conditional distributions may be well-motivated, while the tail specifications are often left to convenience. However, the posterior distribution of a parameter may depend strongly on such arbitrary tail specifications. This is not easily detected in complex models. In this paper we propose a graphical diagnostic, the Local critique plot, which detects such influential statistical modelling choices at the node level. It identifies properties of the information coming from the parents and neighbours (the local prior) and from the children and co-parents (the lifted likelihood) that are influential on the posterior distribution, and examines local conflict between these distinct information sources. The Local critique plot can be derived for all parameters in a chain graph model.

## **Regularized decompositions for sparse MAP estimation in Factor Analysis.**

by MINGHUI SHI

It has become routine to encounter massive dimensional data sets in a rich variety of applications ranging from genomics to marketing. Large  $p$ , small  $n$  and large  $p$ , large  $n$  statistical problems have become a focus in the statistical community, but there is still a pressing need for new methods that scale to really large settings involving hundreds of thousands to millions of variables. For those kinds of problems, there is clearly a necessity of dimensionality reduction and strong prior information that the high-dimensional data can be sparsely characterized. A promising tool for addressing such problems is sparse Bayesian latent factor regression model-

ing. We proposed a method that scales to much higher dimensions through conducting sparse maximum a posteriori (MAP) estimation in carefully-structured Bayesian factor models. In particular, we choose some priors that automatically induce penalties that lead to many zeros values in the loadings matrix and automatic selection of the number of factors, while also facilitate computation via simple optimization routines. \*It is a joint work with Prof. David B. Dunson.

### **Efficient computation in Bayesian model selection and comparison of generalized linear latent variable models.**

by VITORATOU SILIA, NTZOUFRAS, I. AND MOUSTAKI, I.

Moustaki and Knott (2000) presented a unified framework for latent variable models that uses the nomenclature of the widely used generalized linear models. This general framework was called the Generalized Linear Latent Variable model (GLVM). Due to their generality, these models allow for a unified approach to be implemented in all type of responses (normal or non-normal continuous, categorical, ordinal or count data). Most of the approaches in latent variable models follow the traditional frequentist approach using maximum likelihood estimators and approximate significance tests. Bayesian methods for model comparison or Bayesian model averaging (BMA) can provide an alternative on the topic. In this work we focus on the Bayesian model evaluation and comparison for GLVMs. Our intention is to propose a fully Bayesian method after appropriate specification of the prior distribution. Furthermore, we focus on the efficient estimation of the marginal likelihood for such models. This can be achieved if we exploit the assumption of local independence (given the latent variables, items are independent) and the use of the integrated likelihood which eliminates the latent terms from the actual model likelihood. A thorough comparison of marginal likelihood estimators will be also presented, including the harmonic mean estimator (Kass and Raftery, 1995, Raftery et al, 2007), the bridge sampling estimator (Meng and Wong, 1996), Laplace-Metropolis estimator (Lewis and Raftery, 1997), Chib and Jeliazkov estimator (Chib and Jeliazkov, 2001), Power posterior estimator (Friel and Pettit, 2008) and the Copula based estimator (Nott et al, 2009). Chib, S. and Jeliazkov, I. (2001). Marginal Likelihood from the Metropolis-Hastings Output. *Journal of the American Statistical Association*, 96, 270-281. Friel, N. and Pettit, A.N., (2008). Marginal likelihood estimation via power posteriors. *Journal of Royal Statistical Society*, 70, 589-607. Kass, R.E. and Raftery, A.E. (1995). Bayes Factor. *Journal of the American Statistical Association*, 90, 773-795. Lewis, S.M. and Raftery, A.E. (1997). Estimating Bayes Factors via Posterior Simulation with the Laplace-Metropolis Estimator. *Journal of the American Statistical Association*, 92, 648-655. Meng, X.-L. and Wong, W.H. (1996). Simulating Ratios of Normalizing Constants via a Simple Identity: A Theoretical Exploration, *Statistica Sinica*, 6, 831-860. Moustaki, I. and Knott, M. (2000). Generalised Latent Trait Models. *Psychometrika*, 65, 391-411. Nott, D.J., Kohn, R.J. and Fieldin, M (2009). Approximating the marginal likelihood using copula. Under submission, <http://arxiv.org/abs/0810.5474> Raftery, A., Newton, M., Satagopan, J. and Krivitsky, P. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity.

### **The Shockwave model for time series analysis.**

by RICARDO SILVA

In the typical autoregression model for time-series analysis, a variety of interpretations is given to the error terms. A common interpretation is that the errors represent unspecified events that affect the observed outcomes in the series. For instance, such hidden causes have been described as “shocks” in the econometrics literature. Under this point of view, it is natural to consider causes that directly affect the time series over an extended but finite window of time from the moment they occur. Autoregressive moving average models (ARMA) are a classical family of this type. We present a different perspective that complements ARMA modeling. Error terms, which represent an aggregation over external shocks, become marginally dependent within a

moving time window. In this paper, we describe a model that follows from such assumptions, including a convenient parameterization that exploits recent advances in graphical and copula models, and efficient algorithms for Bayesian inference. The result is a flexible family of non-Markovian, heteroscedastic autoregression models that are learned through Markov chain Monte Carlo with a cost that is linear in the number of time points.

### **Nested sampling and the foundations of computational inference.**

by JOHN SKILLING

Foundations matter. They yield understanding. 0. Probability calculus is solidly based on elementary symmetry. 1. Probability is a measure, whose elements can be arbitrarily re-ordered. Hence dimension and topology aren't fundamental. 2. Prior must support posterior, but not conversely. Hence inference is asymmetrically compressive. 3. Point samples do not have associated volumes. Hence direct Monte Carlo and importance sampling of the prior fail, along with the harmonic mean and Chib approximation which sample the posterior. 4. Because probability is a measure, compression can only be achieved by modulating the prior through a succession of pseudo-likelihood functions. 5. The only pseudo-likelihoods that are impervious to phase transitions are constant, with compression achieved by hard constraint. 6. That uniquely capable algorithm is nested sampling, which is a properly Bayesian method with proven mean-square convergence.

### **Incorporating biological information in Bayesian models for the selection of pathways and genes.**

by FRANCESCO CLAUDIO STINGO

The vast amount of biological knowledge accumulated over the years has allowed researchers to identify various biochemical interactions and define different families of pathways. There is an increased interest in identifying pathways and pathway elements involved in particular biological processes. Drug discovery efforts, for example, are focused on identifying biomarkers as well as pathways related to a disease. We propose a Bayesian model that addresses this question by incorporating information on pathways and gene networks in the analysis of DNA microarray data. These information are used to define pathway summaries, specify prior distributions, and structure the MCMC moves to fit the model. We illustrate the method with an application to gene expression data with censored survival outcomes. In addition to identifying markers that would have been missed otherwise and improving prediction accuracy, the integration of existing biological knowledge into the analysis provides a better understanding of underlying molecular processes.

### **Uncertainty in marginal model specification for dyadic relational data.**

by ANDREW C. THOMAS

Networks of many different classes, including social, biological and technological in nature, are often modelled haphazardly with various binary data schemes that are often incompatible and leave much to be desired for model checking. I will show that a hierarchical modelling approach, with the marginal specification of dyads as the primary unit, allows for a considerable expansion of network-like models not limited to the binary case, as well as the increased flexibility and use of model checking strategies that have not been as applied to this class of data. I demonstrate this methodology on alternate explanations for two approaches popular in the networks literature, the "small-world" and "scale-free" mechanisms, and demonstrate the ability of marginal hierarchical modelling to expand beyond them.

**Model choice for stochastic epidemics in households.**

by CONGHUA WEN

Although methods for parameter estimation for stochastic models of disease transmission are now well-established, the picture is much less clear for model assessment. We consider various approaches for model choice problems in the context of data on disease outbreaks collected at the level of individual households.

**An encompassing prior generalization of the Savage-Dickey density ratio.**

by RUUD WETZELS

Hoijtink, Klugkist, and colleagues proposed an encompassing prior (EP) approach to facilitate Bayesian model selection for nested models with inequality constraints. In this approach, samples are drawn from the prior and posterior distributions of an encompassing model that contains an inequality restricted version as a special case. The Bayes factor in favor of the inequality restriction then simplifies to the ratio of the proportions of posterior and prior samples consistent with the inequality restriction. To date, this formalism has been applied almost exclusively to models with inequality or “about equality” constraints. Here it is shown that the EP approach naturally extends to exact equality constraints by considering the ratio of the heights for the posterior and prior distributions at the point that is subject to test (i.e., the Savage-Dickey density ratio). The EP approach generalizes the Savage-Dickey ratio method, and can accommodate both inequality and exact equality constraints. Unfortunately, the EP approach to exact equality constraints is vulnerable to the Borel-Kolmogorov paradox. We conclude that the general EP approach is an elegant and computationally efficient procedure to calculate Bayes factors for nested models, but that users should be aware of its limitations and underlying assumptions.

**Modelling extracellular recordings with Bayesian hidden Markov models and hierarchical Dirichlet processes.**

by NICOLE WHITE

The reliable detection and classification of spikes in extracellular recordings is fundamental to studies in clinical neuroscience. However, of the techniques currently used in practice, many are highly supervised, are based on “hard” classification and treat the steps of detection and classification as independent processes. Consequently, the quality of classification is in part dependent on the chosen method of detection. To this end, there is a need to develop models that combine these steps into a single, data-driven algorithm. Furthermore, there is a need to develop models that properly manage the uncertainty in detection and classification. We consider Bayesian hidden Markov models (HMMs) as generative models for the analysis of multiple action potentials in extracellular recordings. In particular, we focus on data with low signal-to-noise ratios, a complicating factor in reliable spike detection, and irregular firing activity or “bursting”. The latter feature is explored in two settings; a fixed multiple changepoint and hierarchical Dirichlet process model. A hierarchical Dirichlet process is also used to investigate the problem of modelling an unknown number of neurons. All models are estimated by Gibbs sampling, with posterior predictive checks for identified spikes by neuron. Motivating this work is the analysis of microelectrode recordings of the subthalamic nucleus, a surgical target for Deep Brain Stimulation in the treatment of Parkinson’s disease.

**Sequential learning of model error in dynamic models.**

by RICHARD DAVID WILKINSON

Data assimilation is the process of combining empirical observations with predictions made by a model, to produce estimates that take both the data and the science encoded in the model into account. When performing data assimilation, it is crucial to recognise that discrepancies arise

both because of observational error and model error. Much current work on data assimilation implicitly assumes that the data are observations of the model outputs with independent random errors, but this is fundamentally wrong. When model error is not accounted for, observations can be repeatedly surprising, in the sense that they are repeatedly out in the tails of the prediction (or conversely predictions are too conservative and give too vague a prediction). Dynamic models provide an ideal setting for learning a model's error, as we are often in the situation where predictions are made and outcomes then observed, before further predictions are made. In this talk, we will introduce novel statistical methodology for sequential learning of the model error in dynamic models. We use Gaussian processes to model a state-dependent model error term, and use noisy observations of the system to update our beliefs about the model error. By embedding a sequential Monte Carlo algorithm for state estimation within an iterative algorithm for estimating the Gaussian process, we are able to learn about the error term and converge on a description of model error, distinct from the measurement error on the observations. We will demonstrate the methodology on idealised models where we know the true value of the parameters and the error, and show that our approach is capable of producing predictions of much greater accuracy than standard methods are capable of achieving.

---

# PARTICIPANTS

---

<b>Name</b>	<b>Affiliation</b>	<b>Name</b>	<b>Affiliation</b>
Mouna Akacha	University of Warwick	Tyler McCormick	Columbia University
Lorraine Allchin	University of Oxford	Giorgios Minas	University of Warwick
Elja Arjas	University of Helsinki	Irini Moustaki	London School of Economics
Serena Arima	Università di Roma La Sapienza	Peter Müller	MD Anderson Texas
Khandoker Shuvo Bakar	University of Southampton	Habib Najm	Sandia National Labs
Eleni Bakra	MRC Biostatistics Unit, Cambridge	Yuval Nardi	Israel Inst. of Technology
Anjishnu Banerjee	Duke University	Bent Natvig	University of Oslo
Marilena Barbieri	Università di Roma Tre	Geoff Nicholls	University of Oxford
M.J. (Susie) Bayarri	University of Valencia	Thomas Nicholls	University of Warwick
Jim Berger	Duke University	George Nicholson	University of Oxford
Leonardo Bottolo	Imperial College London	Bernardo Nipoti	University of Pavia
Liseo Brunero	Università di Roma La Sapienza	Ioannis Ntzoufras	Athens U.E.B.
Monica Bugallo	Stony Brook University	Anthony O'Hagan	University of Sheffield
Carlos Carvalho	University of Chicago	Janne Ojanen	Aalto University
Annalisa Cerquetti	Università di Roma La Sapienza	Antony Overstall	University of Southampton
Dalia Chakrabarty	Univ. of Warwick	Michail Papatomas	Coventry University
Delson Chikobvu	University of Free State, South Africa	Antonio Parisi	Un. Roma "Tor Vergata"
Merlise Clyde	Duke University	Chris Penfold	University of Warwick
Caterina Conigliani	Università Roma Tre	Natesh Pillai	University of Warwick
Peter Craig	Durham University	Matti Pirinen	WTCHG, University of Oxford
James Cussens	University of York	Murray Pollock	University of Warwick
Shutong Ding	Örebro University	Mohsen Pourahmadi	Texas A&M Univ
Petar M. Djuric	Stony Brook University	Hazinur Rahaman Kahn	University of Warwick
David Dunson	Duke University	Mattias Rantalainen	University of Oxford
Mohammad Taghi Fathi		Oliver Ratmann	Duke University
Jon Forster	Univ. of Southampton	Fabio Rigat	University of Warwick
Dimitris Fouskakis	National Technical University of Athens	Christian Robert	Paris-Dauphine
Guy Freeman	University of Warwick	Francisco J. Rubio Alvarez	University of Warwick
Kassie Fronczyk	UC Santa Cruz	Richard Savage	University of Warwick
Jørund Gåsemyr	University of Oslo	Ida Scheel	University of Oslo
Mulugeta Gebregziabher	Medical University of South Carolina	Minghui Shi	Duke University
Alan Gelfand	Duke University	Vitoratou Silia	Athens U.E.B.
Ed George	University of Pennsylvania	Ricardo Silva	University College London
Seyed Kamran Ghoreishi	Razi university, I. R. of Iran	John Skilling	Maximum Entropy Data Cons.
Flavio B. Gonçalves	Univ. of Warwick	Jim Smith	University of Warwick
Robert Goudie	University of Warwick	David Spiegelhalter	Cambridge University
Raoul Grasman	University of Amsterdam	Matthew Stephens	University of Chicago
Jim Griffin	University of Kent	Francesco Claudio Stingo	Rice University
John Haslett	Trinity College Dublin	Dongchu Sun	University of Missouri
Chris Holmes	University of Oxford	Yee Whye Teh	University College London
Jane Hutton	University of Warwick	Andrew C. Thomas	Carnegie Mellon University
Christopher Jackson	MRC Biostatistics Unit, Cambridge	Amber Tomas	University of Oxford
Maria Kalli	University of Kent	AJ van der Merwe	University of the Free State
Kostas Kalogeropoulos	London School of Economics	Aki Vehtari	Aalto University
Georgios Karagiannis	University of Bristol	Eric-Jan Wagenmakers	University of Amsterdam
Sune Karlsson	Örebro University	Conghua Wen	University of Nottingham
Soleiman KHAZAEI	University Paris - Dauphine	Nanny Wermuth	University of Gotenburgh
Kei Kobayashi	The Institute of Statistical Mathematics, JP	Ruud Wetzels	University of Amsterdam
Robert Kohn	UNSW	Nicole White	Queensland U. of Techn.
Athanasios Kottas	UC Santa Cruz	David Leslie Wild	University of Warwick
James Lawrence	Cambridge University	Richard David Wilkinson	University of Nottingham
Anthony Lee	University of Oxford	Quin Wills	University of Oxford
Guangquan Li	Imperial College London	Henry Wynn	London School of Economics
Antonio Lijoi	Università di Pavia	Christopher Yau	University of Oxford
Anastasia Lykou	Lancaster University	James Zidek	University of British Columbia
Christian Macaro	Duke University	Piotr Zwiernik	University of Warwick
David Madigan	Columbia University		
Dora Matzke	University of Amsterdam		
Vinicius Diniz Mayrink	Duke University		

# SPONSORS

## Model Uncertainty

