Conference Proceedings

25th Annual Research Students Conference in Probability and Statistics

University of Warwick

18th-21st March 2002

Contents

Welcome to RSC2002

Introduction to RSC2002

The University of Warwick would like to welcome you to the 25th Annual Research Students' Conference in Probability and Statistics. This 4-day event has been organised by postgraduates, for postgraduates, and will hopefully provide a friendly environment for you to meet other students with similar interests. The RSC has become rather popular in recent years and in this, its 25th year, we are pleased to welcome over 140 delegates.

For many of you, this may be your first experience of giving a talk or displaying a poster. We hope this will provide an appropriate forum for you to present your research. If you are attending for the first time you may have chosen not to make a presentation but we are sure you will still benefit from attending the talks and meeting people who are working in a similar field.

In the same way as last year at Newcastle, students have the chance to present for up to 30 minutes. Talks will run in parallel sessions, so the timetable must be strictly adhered to. This will be the responsibility of the chairperson, who should read the instructions given in this booklet.

Posters will be situated in the engineering concourse and delegates will get the opportunity to peruse them over tea and coffee. A prize will be awarded for the best poster. You will have a chance to vote for the winner until 1pm on Wednesday 21st March. The winner will be announced at the conference dinner. If you are presenting a poster, please read the instructions given.

We would like to take this opportunity to thank all within the Department of Statistics here at Warwick, and also the sponsors, who have helped to make this event possible.

We hope you will make the most of your stay at Warwick, and that RSC2002 will be both an enjoyable and valuable experience.

Simon Bond   Ben Cowling   Sandeep Shah

An Introduction to Warwick University

You are staying on the campus of the University of Warwick – a lively modern campus about 3 miles South of the centre of Coventry, where the city meets the County of Warwickshire. The University is the place of study for 15,115 undergraduate and postgraduate students.

Campus shops include a post office, supermarket, pharmacy and bookshop; there are branches of the major banks, a Health Centre, a Chaplaincy, a launderette and even a travel agency. There are bars, cafés and restaurants, an art gallery, cinema, two theatres and a concert hall. For sports enthusiasts, there is a Sports Centre with a swimming pool, squash courts, fitness room, and a climbing room. You can walk through Tocil Woods, jog around the campus lakes, or follow the outdoor sculpture trail.

The nearest supermarket off-campus is Tescos, situated near the Claycroft halls of residence. The nearest pub is the Varsity, just a few minutes around the corner from the gatehouse.

To get into Coventry city centre, you can take the number 12 bus from outside the Arts Centre. Alternatively, a taxi ride will cost approximately £5.

Coventry City Centre

A modern, multi-cultural city with an ancient past, Coventry is just three miles from campus. It boasts excellent shopping, ranging from the modern West Orchards Shopping Centre and the recently redeveloped Lower Precinct to the small shops of medieval Spon Street. Coventry offers a lively nightlife with many pubs and clubs. We can recommend Whitefriars in Gosford Street for a traditional pub atmosphere, or Brown's, in Earl Street, for a modern up-beat night. A new city-centre leisure complex, the Skydome, houses two nightclubs as well as a multi-screen cinema, ice rink, restaurants and bars.

Coventry is a busy manufacturing city (Jaguar, Massey Ferguson, the Peugeot Motor Company) with a population of 300,000. Substantially rebuilt after the devastation of the Second World War, it retains some medieval buildings and has a striking modern cathedral, designed by Sir Basil Spence.

Earlsdon, a suburb on the South side of the city, is a hub of student life. It is close to the University of Warwick and has good local shops, pubs and restaurants.

Other Local Towns

KENILWORTH, just a short bus ride from the University, the small town of Kenilworth is well worth a visit. Its magnificent red sandstone castle, built in 1120, is one of the region's major tourist attractions. It also has many interesting restaurants and pubs.

ROYAL LEAMINGTON SPA, an elegant Regency town and home to many Warwick students, is 8 miles (13 kilometres) from campus. It has excellent shopping, and is popular with students for its nightlife. There are some nice restaurants here too.

WARWICK, the historic and attractive county town, is dominated by its castle. Standing on a site first fortified one thousand years ago, the famous Warwick Castle is the finest medieval fortress in the country set in 24 hectares of grounds and gardens. Visitors can enjoy re-enactments of battles and gain a flavour of life from centuries ago with the resident band of fully costumed actors.

STRATFORD-UPON-AVON, famous the world over as the birthplace of Shakespeare, lies only 15 miles (24 kilometres) from campus. Here you can walk the streets where Shakespeare spent his childhood, or see a play at one of the three Royal Shakespeare Company theatres.

Useful Contact Numbers

N.B. All of the telephones on campus require you to press (*) before dialling an external number.

Internal Telephone numbers:

Daliwal Pharmacy	72860
Health Centre	24888
Security	22222
Sports Centre	23011

External Telephone numbers:

Emergency		999
Emergency Organiser Contact		07765 255085
Local Police (non-emergency)		024 7653 9000
NHS Direct		0845 46 47
Takeaway Food:	Cheung Sing (Chinese)	024 7667 7702
	Shah Bagh (Indian)	024 7667 2467
Travel – Bus:	National Express	0121 622 4373
Travel – Taxi:	Godiva Taxis	024 7622 2227
Travel – Train:	National Rail Enquiries	08457 484950

Timetable of Events

Monday 18th March

14.00 – 16.00	Delegates Arrive Conference Registration (*Rootes Reception*)
16.00 – 18.45	Plenary Session (*Humanities H0.52*) 16.00 Welcome 16.05 Prof. Sir David Cox 16.55 Alex Glaser *EPSRC/RSS Graduate Training Program* 17.15 Prof. Peter Green 18.05 Prof. Jim Smith
18.45 – 19.30	Dinner (*Rootes Restaurant*)
20.00 – 23.00	Pub Quiz (*Zippy's, Student Union*)

Tuesday 19th March

07.30 – 09.00	Breakfast (*Rootes Restaurant*)
09.15 – 11.00	Session 1 (*Engineering Block*)
11.00 – 11.30	Break for tea and coffee (*Engineering Concourse*)
11.30 – 12.45	Session 2 (*Engineering Block*)
12.45 – 14.00	Lunch (*Rootes Restaurant*)
14.00 – 15.20	Session 3 (*Engineering Block*)
15.20 – 16.20	Poster Session (*Engineering Concourse*) tea and coffee will also be available
16.25 – 17.45	Session 4 (*Engineering Block*)
18.00 – 19.00	Dinner (*Rootes Restaurant*)
20.00 – 23.00	Barn Dance (*Cryfield Pavilion*)

Wednesday 20th March

07.30 – 09.00	Breakfast (*Rootes Restaurant*)
09.15 – 11.10	Session 5 (*Engineering Block*)
11.10 – 11.30	Break for tea and coffee (*Engineering Concourse*)
11.30 – 12.45	Session 6 (*Engineering Block*)
12.45 – 14.00	Lunch (*Rootes Restaurant*)
14.00 – 16.00	Session 7 (*Engineering Block*)
16.30 – 18.30	Sponsors' Wine Reception (*Engineering F107*)
19.00 – 19.15	Depart for Conference Dinner (meet in *Rootes Reception*)
19.30 – 23.00	Conference Dinner (*Royal Britannia Hotel*)

Thursday 21st March

07.30 – 09.00	Breakfast (*Rootes Restaurant*)
09.30	Delegates depart

What about the Evenings?

To make your time here more enjoyable, we have organised some evening entertainment.

Monday 18th March

After your long journey to Warwick, we know you will probably just want to sit down and relax, so we've organised a pub quiz in the student union starting at 8.00pm. We've booked out an upstairs room, hired a DJ and prepped a quizmaster, and it should be a great evening.

We hope this relaxed atmosphere will give you a chance to mingle, and make new friends.

Tuesday 19th March

We've lined up a superb ceilidh for our Tuesday night entertainment. The band are called Glorishears, and are one of the Midlands leading ceilidh bands. It will be held in the newly-built Cryfield pavilion, just a short walk from Rootes, and ideal for a barn dance. The band are starting at 8.00pm, and they'll go on playing until 11.00pm.

You can visit the Glorishears website at:
http://www.folkfax.freeserve.co.uk/gshears/

Wednesday 20th March

On our final evening, we'll first have a wine reception where you will have a chance to meet our sponsors. Then we will travel into Coventry for our conference banquet in the ballroom at the Royal Britannia Hotel in Coventry. Coaches will depart at 7.15pm, and return afterwards at 10.30pm.

We hope that you'll enjoy your time at RSC2002.

Timetable of Talks

Plenary Session. Monday 18th March, 16.00 – 18.45.

16.00	Welcome
16.05	Prof. Sir David Cox *Do we need formal theories of inference?*
16.55	Alex Glaser *EPSRC/RSS Joint Graduate Training Program*
17.15	Prof. Peter Green
18.05	Prof. Jim Smith *Causality*

Session 1. Tuesday 19th March, 09.15 – 11.00

Time	Room	Presentation
09.20	F107	Fractional Bayes Factors applied to a medical cost data set – *Peter Gregory*
09.20	F110	Using Boosting in Classification – *Kassim Mwitondi*
09.20	F111	Dynamic lattice Markov spatio-temporal models for environmental data – *Linda Garside*
09.45	F107	Analysis of Sickle Anaemia Data: a Bayesian Semiparametric Approach – *Samsom Babatunde Adebayo*
09.45	F110	Local versus Global Models for Classification Problems – *Veronica Vinciotti*
09.45	F111	Spatial Clustering and the Addition of Covariates – *Susan Gooding*
10.10	F107	The use of statistical modelling to identify divergent performance in health care – *David Ohlseen*
10.10	F110	Dynamic principal component methods for studying shape variability – *Osho Ajayi*
10.10	F111	Bayesian methods for clustering – *Ruth Fuentes-Garcia*
10.35	F107	Weighted naive Bayes modelling for data mining – *Jose Ferreira*
10.35	F110	Estimating Labour Force Gross Flows in the Presence of Misclassification and Validation Information – *Nikolaos Tzavidis*
10.35	F111	Inference Techniques for Spatio-Temporal Pollution Data – *Sammy Rashid*

Session 2. Tuesday 19th March, 11.30 – 12.45

Time	Room	Presentation
11.30	F107	The Analysis of High-Density Genotyping Information – *Michael Heeneman*
11.30	F110	The use of index policies to solve a machine maintenance problem – *Helen Mitchell*
11.30	F111	Duck a lá reversible jump – *Lara Jamieson*
11.55	F107	Chromosomal Evolution and Phylogenetic Inference – *George Savva*
11.55	F110	A Whittle index policy for a two-class queueing system with quadratic holding costs – *Marilyn O'Keefe*
11.55	F111	Conditioning an additive functional of a finite state space Markov chain – *Zorana Najdanovic*
12.20	F107	Robust Hierarchical Models for Oligonucleotide Gene Expression Data – *Wiesner Vos*
12.20	F110	Experiences Modelling a Traffic Network – *Ben Wright*
12.20	F111	Limiting Behaviour in an Urn Model – *Christina Goldschmidt*

Session 3. Tuesday 19th March, 14.00 – 15.20

Time	Room	Presentation
14.00	F107	Fitting Generalized Linear Mixed Models to Strawberry Inflorescence Data – *Diana Cole*
14.00	F110	Can noise be useful in statistics and how? – *Kamila Zychaluk*
14.00	F111	Use of Generalized Procrustes Analysis in Assessing Behavioural Expression in Dogs – *Matthew Burnell*
14.25	F107	Longitudinal data analysis of anticoagulant data – *Catherine Fullwood*
14.25	F110	Using Gaussian Kernels To Estimate The Area Under The ROC – *Carlos Cuevas-Covarrubias*
14.25	F111	Simultaneous Confidence Intervals in Linear Modelling – *Jonathan Donnelly*
14.50	F107	Least Squares Fitting of Compact Set-Valued Data – *Talal Maatouk*
14.50	F110	If You Carry On Like That You'll Go Blind… - *Stuart Gardiner*
14.50	F111	New developments in the lilypond – *Codina Cotar*

Poster Session. Tuesday 19th March 15.20 – 16.20

Flare in Systemic Lupus Erythmatosus – *Elizabeth Allen*
Bootstrap Confidence Regions in Shape Analysis – *Getulio Amaral*
Generalized Linear Models for Wind Speeds over Northern Europe – *Steven Bate*
Long Term Survival after the A-bomb – *Judith Anzures Cabrera*
Bound on an Approximation for the Distribution of the Extreme Fluctuations of Exchange Rates – *Nathanaël Benjamin*
Changes in alcohol consumption in the UK – *Laura Gross*
Simulation, MCMC and Bayesian Statistics – *Elizabeth Heron*
Stepuniform coupling for Perfect Simulation – *Giovanni Montana*
Parallel exact sampling of Gaussian Markov Random Fields – *Ingelin Steinsland*
Simulating Multivariate Extreme Value Distributions – *Alec Stephenson*
An Investigation of Floods Along the River Thames – *Elizabeth Traiger*

Session 4. Tuesday 19th March, 16.25 – 17.45

Time	Room	Presentation
16.25	F107	Image analysis and wavelets – *Katherine Hunt*
16.25	F110	Investigating Spatial Variations of Disease in Epidemiology – *Louise Choo*
16.25	F111	Uncertainty in Financial Models of Large and Complex Government Projects – *Kevin McNally*
16.50	F107	Analysing Protein Structure Using a Wavelet Lifting Transform – *Nikki Carlton*
16.50	F110	Regression of underdispersed count data – *Anton Altmann*
16.50	F111	Fitting Mixtures by Sequential Simplification – *Yanzhong Wang*
17.15	F107	Differentiating Noisy Time Series: Wavelet Methods for Radiocommunications data – *Paul Baxter*
17.15	F110	Variance Estimation in the Presence of Imputed Values – *Gabriele Beissel*
17.15	F111	Estimating the Tangency Portfolio on the Markowitz Efficient Frontier – *Constantinos Chappas*

Session 5. Wednesday 20th March, 09.15 – 11.10

Time	Room	Presentation
09.20	F107	Modelling Cancer Mortality in Europe – *Carolyn Davies*
09.20	F110	Inference for Exceedances over High Thresholds – *Lee Fawcett*
09.20	F111	Bayesian Graphical Gaussian Model Selection – David O'Donnell
09.45	F107	Deterministic Estimates of Disease Progression Parameters from Retrospective Data – *David Harrison*
09.45	F110	Quantifying the Impact of Climate Change upon Extreme Sea Levels – *Adam Butler*
09.45	F111	Exploratory Analysis using Graphical Models – *Constantinos Kallis*
10.10	F107	Exploring Childhood Undernutrition and Mortality in Africa –*Ngianga-Bakwin Kandala*
10.10	F110	Estimating Characteristics of Extreme Events – *Chris Ferro*
10.10	F111	The Observed Association Structure from Graphical Log-Linear Models with a Binary Latent Variable – *Maria Fatima Salgueiro*
10.40	F107	Fisher, Combination of Observations (Meta-analysis) and Ancillary Statistics – *Keith O'Rourke*
10.40	F110	Asymptotically Independent Joint Tail Modelling in Practice – *Alexandra Ramos*
10.40	F111	Multicausal Prior Families for Patterns of DAGs – *Ali Daneshkhah*

Session 6. Wednesday 20th March, 11.30 – 12.45

Time	Room	Presentation
11.30	F107	Some properties of the Logistic Proportional Hazard Model – *Yasin Al-Tawarah*
11.30	F110	Meta Analysis - what it is, what it isn't – *Dan Jackson*
11.30	F111	Choice of Statistical Methods for Comparing Heuristic Performance – *Zahid Hussain*
11.55	F107	Tha gamma frailty model - some issues – *Peter Barker*
11.55	F110	Group Screening for Experiments with Large Numbers of Factors – *Anna-Jane Vine*
11.55	F111	New skew normal distributions and their applications – *High Seng Chai*
12.20	F107	Randomised Enrollment in Multi-Centre trials – *Matthew Jones*
12.20	F110	Use of prognostic modelling to optimise clinical trial design in acute ischaemic stroke – *Fiona Young*
12.20	F111	How confident are you of your normality? – *Richard Samworth*

Session 6. Wednesday 20th March, 11.30 – 12.45.

Session 7. Wednesday 20th March, 14.00 – 16.00

Time	Room	Presentation
14.00	F110	Presentation of Statsoft Software– *Matthew Coates*
14.00	F111	Estimation of the association parameter for bivariate Archimedean Copulas – *Dimitris Nicoloutsopoulos*
14.30	F110	Omitted and Supernumerary Values in Time Series – *David Cairns*
14.30	F111	Methodology at the Office for National Statistics, and Academic Research – *Gareth James*
14.55	F110	Modelling Water Quality in the Clyde Estuary – *Andrew McMullan*
14.55	F111	Learning by playing: results in multi-agent reinforcement learning – *David Leslie*
15.20	F110	Forecasting non-stationary time series by wavelet process modelling – *Piotr Fryzlewicz*
15.20	F111	Modelling mate choice & sexual selection – *Miguel Marques dos Santos*

Instructions for Poster Displays

Thank you for volunteering to display a poster. Please read the following information and contact an organiser if you have any other queries.

- The allocated size for your poster is 1m x 1m. Please try to display your poster within this area.

- The boards should be available for you to display your poster from Monday evening. Please make sure your poster is up by 9.30am on Tuesday 19th March.

- If you require any drawing pins etc., please ask one of the organisers.

- Posters will be displayed for the full duration of the conference, and there will be a formal poster session on Tuesday afternoon. Here you should make yourself available to answer any questions from delegates. Hopefully this will provide an opportunity for you to meet people who are working in similar areas.

There will be a prize for the best poster, and delegates will be able to vote until 1.00pm on Wednesday 20th March. The winner will be announced at the conference dinner. **We wish you the best of luck**.

Instructions for Chairs

Thank you for volunteering to chair a session. In order for things to run smoothly, there are a few things you need to remember:

- Look for the chair person's pack which will be left in the room. This will contain timing notices, prompting signs, and OHP notices.
- Introduce yourself to the speakers before the session begins.
- Ensure the OHP displays the relevant slide for your session, and is in focus.
- Ensure the session begins on time. There should be a clock in the room.
- Introduce the first speaker, giving their name and talk title.
- Show the relevant cue cards when there are 5/2/1 minutes remaining.
- Do not let the speaker run over their allocated time. Politely intervene if necessary.
- Thank the speaker and invite questions, ensuring this takes no longer than 5 minutes (there will be ample opportunity for further discussion over coffee).
- Encourage applause.
- Do not start the next talk until the allocated time, to allow people to move between sessions (talks are timetabled simultaneously in all three seminar rooms).
- When all talks are finished, thank all the speakers again and display the OHP slide for the next session.
- On the second OHP, display the slide which advertises RSC2003.

Abstracts for Talks

The abstracts for all the talks are given on the following pages, in alphabetical order of surname.

Samson B. Adebayo, Ludwig Maximilians University of Munich	Osho O. Ajayi, Glasgow University
Analysis of Sickle Anaemia Data: a Bayesian Semiparametric Approach	Dynamic principal component methods for studying shape variability
Sickle cell anaemia is a genetical problem. It is an hereditary disorder of the red cells in the blood which makes the sufferer experience various kinds of ill-health and discomfort from pains due to inability of free-flow of blood into all parts of the body. Three categories of sickle cell genotype (SS, SC and CC) were considered in this paper. A multinomial probit analysis was adopted for modelling influence of some haematological parameters on genotype of known sickle cell patients at the University of Ilorin Teaching Hospital, Nigeria. Because of the weakness of the conventional parametric regression models which are not flexible enough to reveal the possible nonlinear effects of the metrical covariates, a semiparametric analysis within a Bayesian framework that uses recent Markov chain Monte Carlo techniques was employed.	In the statistical analysis of shape data, the variability structure of sample configurations can be studied using principal component methods usually in the tangent space to a specified icon. A useful icon is the estimated transformation invariant mean. The estimation process of this icon implicitly assumes that all the sampling units are of equal importance. This may not always be true. In this work, we show that a locally estimated mean shape may be a preferred icon in certain situations. A weighting mechanism is introduced which, depending on the distance from the point of estimation only allows a sample unit to proportionally contribute into the estimation of the icon. This produces a smoothly changing mean shape over a predefined covariate. The resulting principal components from the tangent space to the icon are also smoothly changing. We explore this result for use in in a dynamic variability study for shape data.

Yasin Al-Tawarah, Keele University

Some properties of the Logistic Proportional Hazard Model

MacKenzie (1996) introduced the 3-parameter Generalized Time-Dependent Logistic family of survival models. In this paper we focus on the Logistic PH (LPH) Model in which the baseline hazard function, $\lambda_o(t \mid \alpha, \gamma)$ is a 2-parameter logistic function of time.

An advantage of the model is that when $\beta = 0$, the model reduces to a testable parametric form, namely, the time-dependent logistic model (MacKenzie, 1996). The lack of such a property in Cox's (1972) PH model is sometimes referred to as the 'absent intercept' problem. In particular, $\beta = 0$ may either imply that the model is PH and the covariates have no influence or it may simply imply that the covariates have no influence because the model is mis-specified. The score test of $\beta = 0$, per se, provides no specific information about which situation obtains. An additional advantage is that the LPH model is wholly parametric and we use the method of maximum likelihood to estimate the parameters. In addition, this model has a closed survival function, which can be used in the analysis of interval censored data.

We discuss the technical properties of the LPH model and use it to describe the survival pattern of 855 incident cases of lung cancer studies in Northern Ireland. The factors age and sex at diagnosis are used in the analysis and the results are compared with those obtained by using the proportional hazards model of Cox. (1972).

Anton Altmann, City University, London

Regression of Underdispersed Count Data

When one wishes to model a dependent variable against a set of explanatory variables, it is important to consider the distribution of the dependent variable, since it is needed to specify the likelihood function to be maximised. If the dependent variable is the frequency of an event e.g. visits to the doctor, then the Poisson distribution is the most obvious choice. However, in practice, the mean = variance assumption of this is rarely satisfied. In the case where the variance is greater than the mean (overdispersion) the negative binomial distribution is an adequate replacement. However, none of the well-known distributions can accomodate situations where the variance is lower than the mean (underdispersion). This talk will outline a method of obtaining maximum likelihood estimates for underdipersed data, with application to American Football touch-downs/field goals.

Peter Barker, University of Lancaster	Paul Baxter, University of Essex
Tha gamma frailty model- some issues	Differentiating Noisy Time Series: Wavelet Methods for Radiocommunications Data
The standard regression model for survival data is the Cox proportional hazards model; covariate effects are assumed to be proportional with time and a baseline hazard function describes the time dependence. The gamma frailty model is a natural extension to the Cox proportional hazards model which allows covariate effects to converge with time without placing a parametric form on the baseline hazard. We propose an extension to the gamma frailty model to allow for both converging and proportional effects with time. We outline the estimation procedure and fit the model to a suitable data set.	This is joint work with G. J. G. Upton Radiocommunications signals pose particular problems in the context of statistical signal processing. This is because short-term fluctuations (noise) are a consequence of atmospheric effects whose characteristics vary in both the short and longer term. These problems are further confounded when we try to recover the rate of change in the signal with time, a quantity of great interest to the Electrical Engineer. We discuss the use of wavelet methods to denoise the data, the problems encountered in calculating the derivative, and possible solutions.

Gabriele Beissel, University of Southampton

Variance Estimation in the Presence of Imputed Values

Imputation methods are widely used for handling missing data in surveys. Providing valid variance estimation techniques for point estimators that are based on imputed values is of great interest since standard variance formulae are not valid when imputation has been used. Applying the naive variance formula could lead to considerable underestimation of the true variance. We present a valid variance formula for an 'improper' multiple imputation method based on a random hot-deck procedure within imputation classes. These classes are constructed using a regression model.

We use data from the Labour Force Survey, a large survey of households, which includes information of hourly earnings of employees in the U.K. The aim is to estimate the proportion P of employees earning below the National Minimum Wage in Great Britain. However, the variable for hourly earnings is subject to nonresponse. The aim is to impute the missing values taking into account information on other covariates, and to estimate the proportion P via a point estimator based on observed and imputed values in the sample. Under the assumption of ignorable nonresponse, an imputation method using a random hot deck procedure within imputation classes based on a regression model, was carried out and compared to more established methods such as predictive mean matching, as investigated in Skinner and Beissel (2001). The imputation is applied multiple times.

Based on this imputation method we derive a variance formula of the point estimator as well as an unbiased estimator of this formula taking into account imputation, response and sampling variability and the complex weighting scheme of the LFS, using a design-based approach (Rao and Sitter, 1995). The response mechanism is assumed to be uniform within imputation classes. A comprehensive simulation study shows good results for point and variance estimators, using different assumptions about the nonresponse mechanism.

In addition, we consider variance estimation using Rubin's multiple imputation formula and compare the results with the variance estimation formulae derived earlier. This formula is designed for proper multiple imputation, however, and we find that this approach underestimates the variance for the improper imputation procedure.

Matthew Burnell, University of Glasgow

Use of Generalized Procrustes Analysis in Assessing Behavioural Expression in Dogs

As part of a study based at the University of Glasgow’s Veterinary School, assessing the level of clinical pain in dogs, a Composite Pain Measurement Scale (CMPS) has been developed. The CMPS contains 7 categories composed of a series of behavioural descriptors, which contribute differing weights towards an overall pain score. One of the categories is headed as ‘Demeanour’ and ‘response to people’ and includes terms such as ‘depressed’ or ‘anxious’. However comments made by observers testing the CMPS in the small animal hospital suggested they found this category more difficult to assess than others.

Therefore to address this concern, volunteers (veterinary surgeons and nurses) were asked to observe the behaviour in a suitable environment and produce their own descriptors for each dog (some post-surgical) for general demeanour and response to people (the handler). For each observer, a matrix described their own responses for each dog in binary format, and then these individual configurations were matched up in a geometric sense using a multivariate technique known as Generalized Procrustes Analysis (Gower, 1975). From this a consensus configuration could be reached and measures of distance of each observer to the consensus and to each other could be calculated. Further, correlating the original matrices to the principal axes of the consensus configuration provided ‘Word Charts’ that enabled assessment of the semantic convergence of the observers’ word choice. That is, it can be seen whether observers apply their own set of terminologies in a coherent manner to describe the dog behaviour (with regard to demeanour and response to people) and whether these assessments are based on commonly perceived and systematically applied criteria. Results from this experiment will be used to enrich the descriptors in this category in the CMPS.

Adam Butler, University of Lancaster	David Cairns, University of Sheffield
Quantifying the impact of climate change upon extreme sea levels	Omitted and Supernumerary Values in Time Series: A Dendrochronologists Dilemma
Data generated using a deterministic hydrodynamical model for sea-levels in the North Sea, run for current and future climate scenarios, provide a basis for assessing the impact of climate change upon the risk of coastal flooding. We propose methods for assessing changes in extreme sea-levels using spatial extreme value techniques, and report the results of preliminary simulation studies.	Dendrochronology is defined as the study of climate changes and past events by comparing the successive annual growth rings of trees or old timber. It has applications in areas as diverse as archaeology and climate study. One area of dendrochronology of great importance is the construction of master chronologies for certain species of trees in certain geographical areas. These master chronologies are constructed by averaging the ring widths from a number of dated samples in order to obtain a chronology which contains all the general characteristics of that species' growth over time. However, in order to produce a master chronology, a dendrochronologist needs to be able to date individual samples accurately. This becomes extremely difficult in the presence of omitted and supernumerary rings. Certain species of tree in extreme climates fail to produce one and one ring only each and every year. These problems can be overcome if they occur rarely by attempting to resynchronise the series by examining it graphically, with reference to another sample that has some presumed overlap with it. These problems are much more difficult to overcome if there is no sample to compare it with or if there are a number of omitted and supernumerary rings. As the successive ring widths form a time series, this is a problem which can formulated and examined statistically. The background which motivated the study will be presented as well as a description of the initial avenues of study which have been pursued, in particular the statistical formulation of the problem and areas of robust methods and outlier study which are related.

Nikki Carlton, University of Bristol	High Seng Chai, University of Southampton
Analysing Protein Structure Using a Wavelet Lifting Transform	New skew normal distributions and their applications
The analysis of protein structure is of great interest because the function of proteins is intricately related to their three dimensional structure. We are able to gain a better understanding of most life processes by studying molecular interactions involving proteins. It has recently been suggested that all protein structures might be classified using a limited number of protein folds. Methods of classifying protein structures already exist, most based on clustering or nearest neighbour techniques. However, many of these techniques are time consuming or require supervision from molecular biology experts. We aim to produce an analysis of protein structure which is relatively quick and easy to implement. We intend to do this by using a wavelet lifting transform to produce a multiscale representation of the protein structure. It is possible to represent a protein using a set of points in three-dimensional space by representing the location of amino acid residues by C_{alpha} atoms. We can perform a multiscale decomposition of these points using an extension of a two-dimensional wavelet lifting scheme for irregularly spaced data into three dimensions. I intend to present the basics of our three-dimensional lifting scheme and outline some of the potential for its application.	The most commonly adopted approach to account for non-normality is the use of suitable monotonic transformation of variables so that the transformed variables are approximately normal. However, there are two major difficulties in association with this method. Firstly, in multivariate setting, transformations are usually on each individual component separately, and achievement of multivariate normality is only hoped for. The other weakness of the transformation method is its difficulty in interpretation, especially when each variable is transformed by using different functions. We provide an alternative method to handle non-normal observations by introducing a new class of multivariate distributions, namely the skew-normal distributions. This family of distributions allows for a great deal of flexibility in degree of skewness and tail behavior. Using such a rich class of distributions, modelling can be performed on the original scale of the data. Thus, the new class of continuous multidimensional distributions gives great flexibility and ease of interpretation in real data fitting that an applied statistician would typically require.

Constantinos Chappas, University College London	Louise Choo, University of Bath
Estimating the Tangency Portfolio on the Markowitz Efficient Frontier	Investigating Spatial Variations of Disease in Epidemiology
The tangency (or market) portfolio on the Markowitz Efficient Frontier is notoriously unstable when based solely on the asset returns' sample moments. A number of estimators have been proposed in order to rectify this problem, most notably shrinkage estimators for the optimisation inputs. In this talk we will be proposing three estimators which attempt to stabilise the market portfolio weights. For the first estimator, we use bootstrap replications to approximate the theoretical distribution of the sample moments estimator and produce a more robust alternative. In the second case, a more general two-parameter family of estimators is proposed, and a method of choosing the optimal parameters. Finally, the last estimator is based on a modification of the utility function to allow for structure among the assets and account for the possibility of transaction costs.	In areas of low population, sparse numbers of disease cases often give rise to data exhibiting extra-Poisson heterogeneity or overdispersion. (That is, within each area of interest the observed disease counts fluctuate about the mean more than would be expected due to variations in Poisson sampling alone.) These variations in estimate precision must be taken into account otherwise differentiating between spurious geographical variation in the disease rates and the true spatial pattern will be difficult. This need to address extra-Poisson heterogeneity has led to the use of Bayesian techniques for smoothing estimates of the relative risk (RR). Current state-of-the-art models for disease mapping include the conditional autoregressive model (CAR) and the simultaneous autoregressive model (SAR). In this talk, I give a motivation for a different model and consider merits to be discussed.

Diana Cole, University of Kent at Canterbury	Codina Cotar, University of Bristol
Fitting Generalized Linear Mixed Models to Strawberry Inflorescence Data	New developments in the lilypond
Many methods now exist for finding the approximate maximum likelihood estimates of the parameters of generalised linear mixed models. The methods include: penalised likelihood, using the EM algorithm, Bayesian methods, and simulated maximum likelihood. In this talk, the main methods are reviewed and compared for fitting binomial logistic generalised linear mixed models to strawberry inflorescence data. These data are sparse, and because of the way they are structured, require the specification of a large number of random effects. Simulations matched to the original data are used to show that a modified EM method due to Steele (Biometrics, 1996) is the most accurate for data of this kind.	We investigate the lilypond model. First we consider the 1-dimesional model studied by Daley, Mallows and Shepp:- Suppose points are distributed on the real line according to the Poisson distribution Po(1). At time t=0, circles start growing at a common constant rate. Whenever two circles touch, both those circles stop growing. New results are obtained for the case when we consider only the first k circles on the positive half-line and ignore all the other circles. We find the probability that at the time the k circles have stopped growing, 0 is not covered by a circle. We then generalize this model by considering instead of a Poisson process, an ordinary renewal process on the real line. We find the probability that 0 is not covered both when all the circles are taken into account and in the k circles case.

Carlos Cuevas-Covarrubias, University of Warwick	Ali Daneshkhah, University of Warwick
Using Gaussian Kernels To Estimate The Area Under The ROC	Multicausal Prior Families For Patterns of DAGs
The ROC curve is a standard way of summarizing the statistical properties of a marker or screening variable in the two-group classification problem. The area under the curve provides a good overall measure of performance. Given samples of cases from the two populations, the ROC curve and its area are usually estimated from the empirical distribution functions, and from the Mann-Whitney statistic, respectively. We show that more accurate estimation of the area is possible by smoothing the data using Gaussian kernels. Our approach is to find the optimum bandwidth for estimating the area, and then to find the best smoothed ROC curve consistent with that overall amount of smoothing. In this way, good smoothed estimates of ROC are obtained which use less smoothing than has been generally recommended in the smoothing literature. Two examples are discussed.	The multicausal essential graph is defined on the equivalence class of DAGs in which local and global parameters independence is assumed for every DAG in this class. The multicausal graph asserts a causal directionality. We characterise family of distributions consistent with the DAGs which exhibits this causality. In a special case when the essential graph is undirected, this family of prior distributions is the Hyper-Dirichlet family. We explore the interventioned causality implicit in these local and global independence.

Carolyn Davies, University of Glasgow

Modelling cancer mortality in Europe

Previous research in cancer mortality throughout Europe has revealed large spatial variations within and between countries. Research also suggests that there are many factors associated with the aetiology of cancers, such as dietary factors, socio-economic status, smoking. Without detailed data on individual mortality and exposure to risk factors it is still possible to explore the geographical variations in cancer mortality and their relationships with population characteristics.

Cancer mortality data and population data are available for 188 regions in 12 EU countries (WHO, 1997). The 1991 data have been standardised for age and sex. The natural hierarchical structure of the data suggest multilevel modelling methods are appropriate to model the distribution of cancer mortality across Europe. Other data were obtained at the region or country level allowing some of the potential causal factors to be taken into account when modelling. Fitting the explanatory variables GDP, consumption of animal fats, fresh fruit and vegetables, tobacco and alcohol, shows what effect these factors have on cancer mortality and explains some of the variation in the fitted model.

A variance components model was fitted allowing the preliminary exploration of hierarchical effects, within area effects and relationships with explanatory variables. However, this method doesn’t take into account the geographic structure present in the data. Fitting a spatial model with multilevel structure takes into account the fact that areas close to each other in geographical space may share common environmental factors which influence cancer mortality. Mapping mortality rates and residuals from these fitted models allows clear visual exploration of patterns existing in Europe. It is also of interest to compare these models, investigating the benefits of fitting the more complex spatial model eg more variation explained or allowing better interpretation of results.

Jonathan Donnelly, University of Southampton

Simultaneous Confidence Intervals in Linear Modelling

The problem considered is the construction of simultaneous confidence intervals for the mean of the response variable when the explanatory variables are constrained to lie within a particular interval. The multiple linear regression models considered here are all assumed to have a Normal error, and the confidence intervals are of the hyperbolic type, i.e. their width is proportional to the standard error of the mean. To date exact confidence intervals of this type are only available in the specific case of simple linear regression, i.e. only one explanatory variable, and so currently any confidence intervals for the mean response are conservative when there are more than one predictor variable. This talk will outline a method to construct exact confidence bands when there are more than one explanatory variable.

Lee Fawcett, University of Newcastle	Jose Ferreira, Imperial College, London
Inference for Exceedances over High Thresholds	Weighted Naive Bayes Modelling for Data Mining
As a procedure for statistical inference, the technique of modelling annual maxima with a generalised extreme value distribution is highly inefficient. Data may have been collected daily, or even hourly, but only one observation per year is used in the analysis. A more modern approach to modelling extreme values of a population or random process is based on the generalised Pareto distribution (GPD). This idea was first suggested by Pickands (1975)*, and involves modelling all observations which exceed some specified threshold level. The asymptotic results surrounding the implementation of the GPD require the underlying sequences to be independently and identically distributed; however, most environmental time series exhibit some form of departure from this ideal. The effect of local temporal dependence and the `clustering' of extreme values on maximum likelihood estimates of the GPD parameters is considered, and the effects of `declustering' the series prior to analysis are examined through simulation studies. An alternative method, which explicitly models the dependence structure of such processes through Markov chains, is also presented. Pickands, J. (1975) Statistical inference using extreme order statistics. Ann. Statist., 3, 119-131.	Naive Bayes modelling has proved particularly popular for data mining applications. Despite the simple, and perhaps unrealistic nature of such models in most problems, they have been shown to predict well and are computationally efficient at handling large data sets. In this talk we propose a new model to improve on naive Bayes by allowing partition weights rather than the more standard variable weights. The information given by each predictor variable is separately extracted by means of a recursive partition structure. This information is then combined across predictors using a weighted product model form, an extension of the naive Bayes model. Empirical results are presented comparing this new method with other methods in the Statistics and Machine Learning literature, for several data sets. Two typical data mining applications, a chromosome identification problem and a forest cover type identification problem are used to illustrate the ideas. The new approach is fast and surprisingly accurate.

Chris Ferro, University of Lancaster	Piotr Fryzlewicz, University of Bristol
Estimating Characteristics of Extreme Events	Forecasting non-stationary time series by wavelet process modelling
It is often important to be able to describe characteristics of extreme events: the duration of storms for example. There are now a number of different approaches to this problem in the extreme-value tool-box. We demonstrate some of them with an analysis of some environmental time-series data.	This is joint work with Sebastien Van Bellegem and Rainer von Sachs, Universite catholique de Louvain, Belgium We propose a new method for forecasting non-stationary time series using wavelets. Using the Locally Stationary Wavelet model (Nason et al. (2000), JRSSB), we formally extend the theory of linear prediction to a class of processes whose second order structure changes over time. We introduce an automatic "adaptive forecasting" algorithm for choosing the parameters of prediction. Applied to a meteorological time series, our method significantly outperforms its classical competitors.

Ruth Fuentes Garcia, University of Bath	Catherine Fullwood, University of Lancaster
Bayesian methods for clustering	Longitudinal data analysis using regression trees
The problem we are considering is a model-based approach for cluster analysis using a mixture of normal distributions. In a Bayesian context, a finite mixture model where the number of components is variable essentially represents a way to deal with the one-dimensional model-based cluster analysis. Reversible Jump Markov Chain Monte Carlo methods (Green, 1995) have successfully achieved this (Richardson and Green, 1997). The aim of the talk is to present an attempt to extend this methodology to higher dimensions. Some problems encountered will be described and a possible alternative method to implement a MCMC model-based cluster analysis will be outlined.	Classification and Regression trees (CART) Breiman et al (1984) are a form of exploratory analysis based on extracting meaningful subgroups, strata with common covariate values, within univariate data. We discuss application of this method to both balanced and unbalanced longitudinal data. The methods are applied to data from patients receiving anticoagulant treatment. The response variable percent time in range (PTR) gives us an indication of how controlled the patient's treatment is. We wish to see whether it is possible to group patients with similar characteristics at the start of treatment to predict how their PTR might behave throughout.

Stuart Gardiner, Nottingham Trent University

If You Carry On Like That You'll Go Blind...

When dealing with glaucoma patients, clinicians routinely measure the sensitivity of the patient's visual field to light. This results in a map of the sensitivity at 54 fixed points in the eye. However, these measurements are notoriously noisy; partly due to actual day-to-day variability in the clarity of the patient's vision, and partly due to human error caused by relying on the patient pressing a button on seeing the visual stimulus. The work aims to develop a spatial filter to reduce this noise, to improve detection of the presence and extent of disease.

The relationships between sensitivities at different points are dictated by the shape of the retinal nerve fibre layer. Defects result from damage to this layer; and follow its structure, rather than occurring in straight lines. This makes a simple Gaussian filter unsuitable. Therefore, the new filter must be based on the actual relationships between sensitivities at each pair of points.

In this talk, the derivation of a new filter will be explained. This will involve a database of 98000 fields, correlations, multiple regressions, and lots of pretty pictures.

Linda Garside, Newcastle University

Dynamic lattice Markov spatio-temporal models for environmental data

The dynamic lattice model has interactions occurring in two dimensional space evolving over time creating a 2+1D model. The latent process and observations in this model are expressed using a dynamic linear model. The dependency through time is expressed as a linear combination of the point at the previous time and the associated second order neighbours.

Consideration is given to the system along the edges. Here, less information is available and the system needs to be adjusted accordingly to retain the stationarity which is required. An alternative, but computationally inefficient method is to embed the process within a larger spatial area.

High correlation between parameters and latent process leads to a single block Metropolis Hastings MCMC scheme being used for parameter estimates. This method is then applied to Atlantic sea temperatures within the region 20 to 30 degrees latitude and -80 to -20 degrees longitude.

Christina Goldschmidt, University of Cambridge	Susan Gooding, University of Lancaster
Limiting Behaviour in an Urn Model	Spatial Clustering and the Addition of Covariates
Balls in urns provide a simple and useful way of describing probabilistic problems. We consider a particular model with coloured balls and a scheme for deletion and replacement, which could describe something like the spread of a disease. We investigate the limiting behaviour of the model as the number of colours goes to infinity. In order to do this, we introduce a Markov process and demonstrate a large deviations-style fluid limit result (the equivalent for a stochastic process of the law of large numbers). Finally, we show that there is a phase transition between two very different kinds of behaviour.	The detection of spatial clustering is an important tool in developing the aetiology of various diseases. I am particularly concerned with developing methods of detecting spatial clustering where the underlying population density is heterogeneous. Therefore, for example, one could be testing whether an observed pattern of leukaemia cases could have risen due to chance alone, given human settlement patterns, or whether it is due to some underlying clustering mechanism. Existing tests of spatial clustering cannot incorporate covariates into their analysis. It would be beneficial for covariates to be included in tests of spatial clustering since ignoring them may spuriously lead to significant results. For example, it is well known that social deprivation is spatially clustered and that some diseases are more prevalent in areas of high deprivation. For this situation the existing tests may detect spatial clustering when in fact it is not present. This talk will introduce a new cases-control test of spatial clustering which incorporates relevant covariate information and in turn overcomes the above problem.

Peter Gregory, University of Sheffield	David Harrison, MRC Biostatistics Unit, Cambridge
Fractional Bayes Factors applied to a medical cost data set	Deterministic estimates of disease progression parameters from retrospective data
When undertaking Bayesian analysis of a data set we select an appropriate model for the data and then incorporate prior knowledge about the parameters of the selected model to calculate the joint posterior distribution for the model parameters. If, however, it is not possible to find a realistic model for the likelihood then a transformation of the data may provide a reasonable solution. We will formulate a Bayesian analysis after using a Power Transformation on the data. To compare possible models, the posterior odds may be expressed as the prior odds multiplied by the Bayes factor. We will show how weak prior information may be formulated using partial Bayes factors. The Fractional Bayes Factor is a resolution of the undefined training sample. We will give examples of model comparisons for a medical cost data set.	When working with case-control data, the estimation of parameters from complex temporal models involving progression in multiple dimensions may not be practicable using a full stochastic model. We demonstrate one such situation and explore how estimates can be obtained using a deterministic epidemiological modelling approach. In a recent study of breast cancer screening, the relative risks associated with dense mammographic parenchymal patterns were found to be higher for grade 3 tumours than for the less aggressive grade 1 and 2 tumours. This observation could be accounted for by (1) a greater biological predisposition of dense tissue to grade 3 cancers; (2) dedifferentiation, the progression of grade over time, with tumours in dense breasts being missed at screening; or (3) both of these phenomena simultaneously. We consider three deterministic epidemiological models corresponding to these three possibilities and estimate the model parameters from size- and grade-specific relative risks. It is concluded that dedifferentiation alone could not account for the patterns in the data, but there is evidence that both processes may be taking place.

Mike Heeneman, University College London	Katherine Hunt, University of Bristol
Analysis of High Density Genotyping Information	Image Analysis and Wavelets
What can you do with a set of markers when you are given no family history. How do you get the most information for the breeders from genetic markers under these circumstances? Simple regressions will offer some hints as to which markers to use, though multivariate methods will explain more of the underlying covariance structure. Sample sizes usually demand the use of dimensional reducing methods such as PCR or PLS.	Our aim is to construct a statistical model to relate a response image to an explanatory image. We may wish to relate two images in order to, for example, restore missing pixels in the response image or to calibrate images (e.g. to remove time dependent effects). We look briefly at existing regression methods before introducing a new model based on the two-dimensional non-decimated wavelet transform (2D NWT). Using data from a LANDSAT-type sensor, we assess the effectiveness of our model in comparison with a simple linear regression model. We conclude by suggesting extensions to our model and by considering alternatives methods for image analysis.

Zahid Hussain, University of Huddersfield	Daniel Jackson, University of Warwick
Choice of Statistical Methods for Comparing Heuristic Performance	Meta Analysis - what it is, what it isn't. Its strengths and weaknesses
Heuristic techniques are basically used to solve NP-hard problems (also known as combinatorial problems) by exploring a large space of possible solutions to find a point at which an objective or cost function achieves its minimum value. All of these techniques approach the solution of NP-hard problem by iterative improvement on a starting set of random solution(s) and deliver non-deterministic approximate solution. Because these heuristic methods are of stochastic nature, knowing the quality of their solutions and studying their statistical properties in order to improve their performance for large size combinatorial problems have been an important issues in optimisation study. Heuristic researchers make their claims about the performance of a particular metaheuristic on the basis of their empirical analysis. Often their claims do not provide enough statistical evidence and possibly claims of heuristics may not be statistical well founded. There is also a need to provide experimental information publicly for reproducibility purposes. This paper focuses on these issues.	Meta Analysis is a fairly new branch of statistics. Many know little about it and this talk should rectify this! I will discuss and explain the various issues. It will be low in technical detail and high in ideas. Concepts will be explored and formulae avoided. In particular, the notions of publication bias and in between study variation will be addressed.

Gareth James, Office for National Statistics	Lara Jamieson, University of Cambridge
Methodology at the Office for National Statistics, and Academic Research	Duck a lá reversible jump
The methodology group of the Office for National Statistics has a number of roles, which include the continuous improvement of statistical methods, and quality assurance of National Statistics. Methodologists conduct research and provide advice on a wide range of statistical topics. This presentation will give a flavour of the work I have done since joining ONS in October 2001. This includes project work on sample re-design, regression, ANCOVA and other statistical analyses. Prior to joining ONS, I was a research student at Cardiff University, and comparisons between work in the methodology group of ONS and academic research, will be drawn. A brief overview of the work I presented in my thesis, which includes errors-in-variables regression models, transformations and normality will also be provided.	National and International authorities are becoming increasingly concerned with the management and maintenance of key wildlife species and their natural habitats. Central to any management programme is the construction of realistic models of the underlying population dynamics and their interaction with the local environment. We focus upon the analysis of a series of data on North American ducks and construct a state space model to investigate the importance of population size and other key covariates upon the underlying population dynamics. The analysis is formulated as a model selection problem and we use reversible jump MCMC techniques to distinguish between competing models. In doing so, we demonstrate how these powerful new techniques can be used to discriminate between competing biological hypotheses and learn how best to manage and maintain these particular populations.

Matthew Jones, University of Cardiff	Constantinos Kallis, University of Warwick
Randomised Enrolment in Multi-Centre trials	Exploratory analysis using graphical models
In multi-centre trials, the most appropriate model to describe the probability of arrival of a given number of patients, at a given centre, in a given period of time is the Poisson process with parameter lambda_i, i=1...,N, where N is the number of centres in the study. Three different estimators of the combined response to treatment (CRT) are studied. These are the least squares estimators of the CRT under fixed effects models of increasing complexity. A simulation study is performed to analyse the behaviour of the estimators under different conditions. Comparisons are made based on the behaviour of the estimators mean squared error (MSE). Theoretical approximations are made for the mean of the MSE for each of the estimators. To obtain these approximations estimates of the negative moments of the Poisson distribution are derived, and are found to be superior to existing estimates.	Aortic aneurysm is a clinical condition that kills older men without warning. The mortality rate for aortic rupture lies between 80% and 90% within a small period of time. Screening is arguably the most effective way of finding cases of abdominal aortic aneurysm (AAA). Several easily measured variables are related to the presence and progress of AAA. However, reliance on pairwise marginal associations may be very misleading, as shown for example by Simpson's paradox. We use graphical modelling on a database of 3000 individuals from the Birmingham Community Aneurysm Screening Project (CASP) to learn about the interrelationships between AAA and factors such as diastolic blood pressure, smoking and age. The statistical package MIM was used for the analysis and presentation of results.

Ngianga-Bakwin Kandala, University of Munich, Germany	David Leslie, University of Bristol
Exploring causes of Child Mortality in Africa	Learning by playing: results in multi-agent reinforcement learning
In this work the causes of childhood mortality among children under five in Africa shall be explored, with the major objective to assess possible temporal and spatial variations in the effects of underlying factors. In practice, the impact of some factors on survival might change with the child's age, for instance, some factors rather increase the risk of mortality within the first months of life, while other factors affect late survival. Child survival might also be affected by some environmental factors such as diseases prevalence or lack of health facilities specific to some locations. The primary focus of this work is to flexibly model non-linear effects of metrical and spatial covariates arising in hierarchically structured survey data sets reflecting recent developments in statistics. The analysis is based on data from the 1992 Demographic and Health Surveys from Malawi, Tanzania and Zambia. Bayesian dynamic probit model for time discrete survival data is used and Markov chain Monte Carlo techniques are employed to explore the spatio-temporal effects as well as non-linear effects of underlying factors within the Generalized linear models framework. Inference is fully Bayesian and uses recent Markov chain Monte Carlo techniques.	Many models of adaptive learning in normal form games require that each player uses an estimate of the expected reward received when each of their actions is played against the current mixed strategies of their opponents. This is generally achieved in one of two ways: players can maintain a model of opponent play based upon observation and use the (known) game structure to calculate the expected rewards (e.g. fictitious play), or the average reward is observed directly in the limit of infinite population size for evolutionary models (e.g. the replicator dynamics). Using two-timescales stochastic approximation, we show how to obtain model-free asymptotically-accurate estimates of the expected action values, and how this can result in a stochastic process whose limiting behaviour is given by that of a standard deterministic dynamical system of adaptive game theory.

Talal Maatouk, University of Glasgow	Andrew McMullan, University of Glasgow
Least Squares Fitting of Compact Set-Valued Data	Modelling Water Quality in the Clyde Estuary
In some practical studies the data consist of intervals. This is usually called set-valued data. I have extended the traditional techniques of single-valued quantities into the context of set-valued data. I propose methods of least squares fitting of such data for constructing multi-function models which are a best fit to a family of set-valued observations, for the case where each datum is a nonempty compact interval of the real line. Two models will be considered: (1) Interval-valued output observed from interval-valued input. (2) Interval-valued output resulting from single-valued input. The solution will be given in two approaches, Parametric and Nonparametric.	The Scottish Environmental Protection Agency (SEPA) is the public body responsible for environmental protection in Scotland. As part of its remit SEPA has been monitoring the water quality in Scotland's rivers and sea lochs, including the River Clyde. Data have been collected on the River Clyde for many years, but statistical analysis has been limited. The data are collected on surveys, which are carried out approximately 16 times per year, but at irregular dates. The variables collected include water temperature, salinity and dissolved oxygen. The main variable of interest in this study will be the analysis of dissolved oxygen, which is most closely linked to water quality. Measurements from the surface level at one particular station on the river are used. A flexible starting point is a model including a mean trend term and seasonal variation as follows: $y_i = \alpha_i(x_i) + \beta_i(x_i)\cos(2\pi x_i - \theta_i(x_i))$ Where y denotes dissolved oxygen, x denotes year in decimal form to include survey time within the year and n denotes sample size. Examination of the data shows that the seasonal term is adequately modelled by a cosine curve with phase parameter theta. The model above maximises flexibility by allowing all parameters to vary smoothly over time. It is the aim of this talk to show how this model, and sub-models where particular parameters are held fixed, can be fitted to observed data. These models are applied to the water quality example and informal conclusions on the most suitable description are drawn from the fitted models. It will then be informative to test the models in a more formal manner, to deduce which of the models best describes the data. This will be done using results from the approximate F test suggested by Hastie and Tibshirani (1990).

Kevin McNally, University of Sheffield	Helen Mitchell, University of Newcastle
Uncertainty in Financial Models of Large & Complex Government Projects	The use of index policies to solve a machine maintenance problem
A non-technical presentation that will focus mainly on the background to the problem of estimating the cost of providing a government service – why such a cost estimate is required & the problems in providing an accurate estimate. The current methodology used by the National Audit Office will be discussed, with reference to a specific project – the redevelopment of the Ministry of Defence Main Building. The presentation will finish with a discussion of the problems faced with making such an estimate, the problems of the existing methodology that the NAO use, and the interests of the National Audit Office. The current direction of my research concludes the talk.	Consider a repairman who is responsible for the upkeep of a number of machines, each of which is in some state of disrepair. At any one time, the repairman can attend to one machine, which incurs some maintenance cost. Whilst his attention is away from the other machines, there is a chance that they will break completely, and have to be replaced at some cost much higher than it would have cost to repair. If a machine does not break, it's state deteriorates slightly. Our aim is to find some policy for the repairman to use, as a method for choosing between machines. We do this be modelling the scenario as a system of Restless Bandits and seek calibrating indices for the machines. Unlike the Generalised Bandit Problem, an index policy is not known to be optimal and indeed in many cases no index exists. We can show that in this case an index (in the Whittle sense) does exist, and we can assess its effectiveness in our scenario.

Kassim Mwitondi, University of Leeds	Zorana Najdanovic, University of Warwick
Using Boosting in Classification	Conditioning an Additive Functional of a Finite State Space Markov Chain
The objective of pattern recognition is to find an allocation rule for multivariate observations, which assigns an observation to one of a small number of known classes. The rule is typically produced by a ``learning algorithm'' which uses a set of training data with known class labels. The rule can then be applied to new observations for which the classes are unknown. The performance of the algorithm is often measured by its ability to predict classes on those new observations. One recently developed enhancement for pattern recognition is called boosting, a technique that generally improves the performance of a learning algorithm. Its strength derives from its capability to transform weakly performing classifiers, such as tree stumps, into strongly performing classifiers, with substantially reduced errors. The technique has been widely praised, in particular, for being resistant to over-fitting. In this talk, we briefly present the mechanics of the technique and some results based on a simulated two-class data set in two dimensions with a sine-wave boundary. For this simple example, classifiers based on a single variable, such as tree stumps, perform poorly, but boosting is much better. We also explore the underlying link between boosting and Bayesian discriminant analysis.	We consider a finite state space continuous time Markov chain and its additive functional, and want to condition the Markov chain on the event that its additive functional stays non-negative. We discuss the cases when the additive functional is with a positive and with a negative drift. In the case of a positive drift there is only one way of conditioning and we, in addition, study the relation through a time reversal between the new process obtained by this conditioning and some other processes which are obtained also from the original Markov chain. In the case of a negative drift, we show one of possible ways of conditioning and the connection of the resulting process with another process obtained from the original Markov chain by double conditioning.

Dimitris Nicoloutsopoulos, University College London	David O'Donnell, University of Southampton
Estimation of the association parameter for bivariate Archimedean Copulas	Bayesian Graphical Gaussian Model Selection
A bivariate distribution H with margins F, G can always be expressed as H(x,y) = C(F(x),G(y)), where C is a copula. When b(C(u,v)) = b(u) + b(v) for some convex decreasing function b defined on (0,1]: b(1) = 0 the copula C is said to belong to the Archimedean class. This representation is a convenient hub to statistical investigations since the generator b characterises the copula family. We examine the problem of estimating the association parameter when the margins are unknown. Our approach is an extension of an existing method.	Graphical Gaussian models (GGM's) are used to model conditional independence structures between continuous variables. Each model can be represented by a graph (with vertices and edges) and comprises a family of (Gaussian) distributions for the variables satisfying a particular set of conditional independence constraints. We describe methodologies for performing Bayesian inference for these models, including both parameter and model selection, primarily through the use of MCMC methods.

David Ohlseen, MRC Biostatistics Unit, Cambridge	Marilyn O'Keeffe, University of Newcastle
The use of Bayesian methods to identify Divergent Performance in Health Care	A Whittle index policy for a two-class queueing system with quadratic holding costs
There has been a rapid increase in the use of statistical modelling to assess the performance of health care providers. Perhaps the most common aim of an analysis is to identify units that appear to be divergent. The variety of modelling techniques have typically employed different decision rules. For example classical fixed effect models have been combined with tests of homogeneity whereas models developed within the Bayesian framework have examined Posterior p-values and credibility intervals for unit ranks. Until now there has been no common metric on which to compare methods, this has lead to the use of an informal sensitivity analysis by comparing a variety of methods. A recent example of this was presented by Spiegelhalter et al (2002), who considered fixed effect models using confidence intervals, random effect models using cross-validation posterior p-values and institutional ranking with credibility intervals. This talk will provide an overview of previous methods and show how these models can be incorporated into a single framework.The idea is based on conflicts of evidence, which has been used in the past to identify troublesome features of multilevel models (O'Hagan, 2001). Spiegelhalter, D.J., Aylin, P., Best, N.G. and Murray, G.D. (2002), Commissioned analysis of surgical performance by using routine data: lessons from the Bristol inquiry ,Journal of the Royal Statistics Society Series 165, 2 O'Hagan, A. (2000) HSSS Model Criticism In Highly Structured Stochastic Systems (ed. P.J. Green, N.L. Hjort and S.T. Richardson), OUP, Oxford	Most of the work to date in modelling multi-class queueing systems has assumed that costs for delaying customers are linear with the number of customers in the queue. Such assumptions are often claimed to be unrealistic. Following work by Whittle (1988), we develop an index policy for a single-class system, which we would expect to perform well in minimising non-linear holding costs. We consider a two-class M/M/1 system with quadratic delay costs and form an index for the average cost problem. Early results are presented. Whittle, P. (1988). Restless bandits: activity allocation in a changing world. J.A.P., A25,287-298.

Keith O'Rourke, University of Oxford	Alexandra Ramos, University of Surrey
Fisher, Combination of Observations (Meta-analysis) and Ancillary Statistics	Asymptotically independent joint tail modelling in practice
Essentially identical studies are often conducted in various fields of inquiry. In order that all the evidence can be quantified, statistical methods are needed to both assess the similarity of estimates from these essentially identical studies (was there replication?) and appropriately combine the estimates in hand (what is the total evidence?). This is often referred to as meta-analysis or combination of estimates (pooling). Formally - as each study could consist of a single observation - any single study could be viewed as a combination of observations. This talk will trace out RA Fisher's possible use of a "combination of observations" conceptualisation in the development of his early concept of ancillary statistics. It will also present a "combination of single observation likelihoods" plot to visually assess JA Nelder's claim that the Stack-loss data set contains no outliers (non-replicating observations). Nelder J.A. There are no outliers in the Stack-loss data. Student 2000, 211-216	A fundamental issue in applied multivariate extreme values (MEV) analysis is modelling dependence within joint tail regions. We have developed a pseudo-polar framework for modelling extremal dependence that extends the existing classical results and provides a constructional procedure for obtaining parametric joint tail dependence models. The practical application of such a model is the focus of our study. Our presentation concentrates on the bivariate case for simplicity and covers applications to simulated and environmental data and details joint estimation of dependence and marginal parameters via likelihood methodology.

Sammy Rashid, University of Sheffield

Inference Techniques for Spatio-Temporal Pollution Data

Atmospheric pollutants, such as Ozone and Nitric Oxide can cause damage to crops, animals and human health. Accurate information is needed to understand the science of their generation and dispersion, and perhaps to aid formulation of policies to reduce their effects.

A difficulty, however, is that detailed measurements of atmospheric pollutants have been made at only a small number of locations around the country, though at these locations hourly time series over a number of years are available.

The aim of the project is to investigate inference techniques for data where spatial coverage is sparse but temporal coverage is dense, using the available atmospheric pollution data.

The initial focus of my project is to study the problem of assessing ozone exposure of forest trees. Scientists have suggested that cumulative exposure to concentrations of ozone over 40 parts per billion (ppb), during the growing season, reduce the growth (yield). This cumulative excess measure is known as AOT40 (Annual Accumulated Excess Ozone Concentration Over A Threshold Of 40 ppb).

The talk will outline the development of a Bayesian method used to estimate predictive distributions for AOT40, using data from all the UK rural monitoring stations for the years 1987 to 2000.

Maria Fatima Salgueiro, University of Southampton

The Observed Association Structure from Graphical Log-Linear Models with a Binary Latent Variable

We investigate the observed association structure between two or three binary manifest variables arising from a latent class model with a single binary latent variable. The latent class model is represented as a graphical log-linear model. These models are a family of parametric statistical models used for modelling discrete data cross-classified in contingency tables (see Whittaker, 1990).

We prove that marginalizing over the latent variable will give the saturated model for the manifest variables. However, when performing data analysis, type II errors (i.e. false acceptances of the null hypotheses) can occur, and therefore the observed association structure will often not correspond to this saturated model. A simulation study is used to estimate the statistical power of the selection procedure and, hence, the probability of a latent class model being chosen. In this context we refer to power of the model selection procedure as the probability of selecting the true model given specified true model parameters (see Smith, 1992).

Results show that when performing model selection to detect a latent class model, for certain combinations of observed odds ratios and marginal cell probabilities, the data analyst must still consider an observed association structure that is not necessarily the one induced by the saturated model of the manifest variables. The importance of the total sample size is also discussed.

Smith, P.W.F. (1992). Assessing the power of model selection procedures used when graphical modelling. In Proceedings of COMPSTAT 1992, Vol.1, 275-280, Heidelberg: Physica-Verlag.

Whittaker, J. (1990). Graphical Models in Applied Multivariate Statistics. Chichester: John Wiley & Sons.

Richard Samworth, University of Cambridge	Miguel Marques dos Santos, University of Bristol
How confident are you of your normality?	Modelling Mate Choice & Sexual Selection
Suppose X has a multivariate normal distribution with mean vector theta and covariance matrix I. Charles Stein's dramatic discovery in 1956 that there exists a better estimator of theta than X has sparked a long search for confidence sets for theta which improve on the one consisting of a sphere centred at X. I will propose that a sphere centred at the Stein estimator should be used, and will describe the asymptotic expansions from which we compute the radius.	It has been claimed that genetic models are the only valid basis for understanding sexual selection and mate choice. We demonstrate that this view is incorrect by showing that there is a general method of translating genetic models into game-theoretic models. We illustrate the method by showing that the Fisher process can work in the context of a culturally transmitted male trait. This model also establishes that the "sexy son" argument is correct and derives the value to the female of mating with a particular type of male.

George Savva, John Innes Centre, Norwich

Chromosomal Evolution and Phylogenetic Inference

All of an organism's genetic information is contained within it's genomes. These genomes consist of long strands of DNA called chromosomes, along which are found the genes.

There are genes which are found in almost every living organism, and it has long been known that by comparing the DNA sequence of a particular gene across different species, it is possible to estimate the evolutionary relationships between them, using models of how DNA sequences are expected to evolve in time.

Recently, a new method has been developed, based on comparisons of the entire genomes of different organisms rather than small pieces of DNA. These work by comparing the gene content (ie which genes are found in which genomes) and the gene order (the order of the genes on the chromosomes). We can then use a model of how entire genomes evolve to estimate relationships between species.

Currently, there are only a few whole genome sequences available for study, but that number is increasing rapidly, and some preliminary analyses on those that are available have produced some very interesting results.

Nikolaos Tzavidis, University of Southampton

Estimating Labour Force Gross Flows in the Presence of Misclassification and Validation Information

Survey data, collected using panel surveys, are widely used for economic research. An example is the Labour Force Survey (LFS), which collects labour market related data. There are several analytical benefits of panel data compared to cross-sectional data. In the case of the LFS many economists are interested in the labour force gross flows. Using these flows economists can identify the labour market condition. Despite the benefits, there are many problems encountered in the estimation of the labour force gross flows. Sampling attrition, response error and rotation group bias are some of the most common non-sampling errors. In a longitudinal framework these errors can introduce severe biases in the estimation of the labour force gross flows. We believe that the conclusions from an analysis that fails to take these problems into account can be quite misleading.

We focus our interest on the response error problem and on double sampling methods (Bross 1954 and Tenebein 1972), which aim at correcting for measurement error when the variables of interest are discrete and validation information is available. We describe alternative double sampling schemes and we study their impact on the gross flows estimators attempting to correct for response errors. A standard assumption used to define such estimators is that of Independent Classification Errors (ICE) (Singh and Rao 1995). Existing literature indicates that the adjustments under ICE can be regarded as an upper bound. In this context, we propose two alternative estimators that seek to reduce the undesirable effects of the ICE assumption. Furthermore, we develop variance estimators for the adjusted gross flows under the double sampling approach. The performance of these estimators is assessed using a simulation study. Our variance approximations work well, while the adjusted gross flows suggest that if we ignore response errors we will tend to estimate a more dynamic labour market.

Bross, I. (1954). Misclassification in 2x2 Tables. Biometrics 10, 478-486.

Tenebein, A. (1972). A Double Sampling Scheme for Estimating from Misclassified Multinomial Data. Technometrics, Vol. 14, 187-202.

Singh, A.C. and Rao, J.N.K. (1995). On the Adjustment of Gross Flows Estimates for Classification Error with Application to the Data from the Canadian Labour Force Survey. Journal of the American Statistical Association, Vol. 90, 478-488.

Veronica Vinciotti, Imperial College, London	Anna-Jane Vine, University of Southampton
Global versus local models for classification	Group screening for experiments with large numbers of factors
In a classification task, an object of known measurement vector x is to be assigned to a class c, given the information provided by a database of objects for which x and c are both known. For simplicity, we will consider the binary class case, with c = 0,1. The statistical approach to this problem estimates the probability p(1|x), directly or via Bayes theorem, and compares it with a threshold t to predict the class of object x. The classification of x depends only on the side of the decision surface (p(1|x) = t) the point x falls in. A good estimation of this surface should then be the main focus of a classification procedure. Unless the overall model is properly specified, it follows that a local model, focusing on the region around p(1|x) = t, will do better than a global model which fits the entire data space. This talk will consider the special case of logistic regression and describe a way to implement a local version of the standard maximum likelihood procedure. We compare the local and standard methods on a set of real databases.	The response in an experiment can depend on the levels of a number of different factors. Factorial experiments assess the different factors simultaneously, providing valuable information on possible interactions between the factors. As the number of factors to be investigated increases, the number of observations needed can rapidly become economically infeasible. An important approach in trying to achieve a practical number of runs in an experiment, is to group factors together and define new grouped factors to represent each group. Traditionally, classical group screening estimates only grouped main effects at a first stage of experimentation whereas interaction group screening also estimates grouped interactions at the first stage. In a second stage of experimentation the individual factors within the important groups are investigated. Limitations of the current theory that are of practical importance include the requirement of equal group sizes and equal probabilities of grouped factorial effects being declared active. The purpose of this talk is to outline extensions to this theory and to present some examples.

Wiesner Vos, University of Oxford	Yanzhong Wang, University of Glasgow
Robust Hierarchical Models for Oligonucleotide Gene Expression Data	Fitting Mixtures by Sequential Simplification
Array-based gene expression methods involving Affymetrix oligonucleotide gene expression arrays produce large data sets with several levels of variation. The purpose of my research is to develop statistical methods and strategies for the analysis of Affymetrix gene expression data, in order to derive an understanding of the relevant biological processes in gene expression studies. This talk deals with the application of robust hierarchical models to these data sets. The talk will illustrate the usefulness of comparing the robust variance component estimates obtained from these models with traditional variance component estimates for the purpose of identifying genes with misbehaving probes and probe sets, as well as for identifying misbehaving arrays. The talk will also propose how robust models can be used for normalization across replicate arrays and to determine differentially expressed genes across various experimental conditions.	Mixture models, because of its flexibility and computational tractability, are widely used to model complex densities. Among techniques for learning mixtures, the most popular appears to be EM algorithm. But as a local search algorithm, EM has a number of limitations: (i) It is slow to converge, (ii) The true number of mixing components is assumed given, (iii) It is sensitive to initialisation, and (iv) It may get stuck in one of many local maxima of the likelihood function. To overcome these limitations, an Iterative Pairwise Replacement Algorithm (IPRA) was introduced based on sequential simplification. It starts at kernel density estimates, combines pairwise similar components by L2E, and only refits locally. We extend this method into higher dimension problems, by using minimum spanning tree to limit searches and updating the MST after each iteration. The components weight ratio is also bounded to avoid possible 'spikes'. Simulation result shows the algorithm (without fine-tuning) achieves similar solution to EM in terms of L2E value, but much faster and provides better location for the estimated component centres.

Ben Wright, Open University	Fiona Young, University of Glasgow
Experiences Modelling a Traffic Network	Use of Prognostic Modelling io Optimise Clinical Trial Design in Acute Ischaemic Stroke
The talk is about my experiences modelling traffic flows through the M26/A2/A282 motorway junction. The flows are measured by the number of vehicles passing counting sites in each hour. This means we have a multivariate time series and we wish to develop a Bayesian forecasting and monitoring system. The problems such a model must overcome are the complexity of the problem, its multivariate nature and the requirement for the system to work quickly. This talk looks at one way of modelling the network, but with emphasis on the problems and issues encountered while applying the model.	Stroke is the 3rd most common cause of mortality in the UK. However, few clinical trials of stroke treatments have shown statistically significant benefits of the intervention being tested, even though many stroke physicians believe that the trials may have missed important therapeutic effects on the patients. There are a number of reasons why the trials may have been unsuccessful: 1. The drugs do in fact have no beneficial effect 2. The patient population is too heterogeneous and hence outcomes vary widely 3. The clinical measurements scales used to measure outcome lack statistical power This talk will focus on clinical measurement scales. The analysis of the outcome measurements usually involves dichotomising these scales into 'Good' or 'Bad' outcome. This can cause a loss of sensitivity in the identification of a treatment effect. Simulations, based on a proportional odds logistic regression model, have investigated several methods of choosing the cut off point for this dichotomization. Initial results suggest that a trial would have increased power to detect a beneficial treatment if patient-specific trial endpoints were set using information from the initial stroke severity.

Kamila Zychaluk, University of Birmingham

Can noise be useful in statistics and how?

That is, apart from giving statisticians jobs in denoising...

It turns out that additional noise is not always an intruder but it can also be helpful. We propose a method which is opposite to denoising, that is instead of removing noise we deliberately add some to the original sample. This may appear strange, but such `controlled' random noise can help in finding solutions, or alternative solutions, to some statistical problems in curve estimation.

We briefly outline how our method can be employed in the estimation of curve derivatives and how the arising estimator can be used in a bias reduction method for a density estimate. The methodology also leads to an interesting corollary, which we discuss. Finally, we look at how extra noise can help in dealing with tied observations in cross validation in the density estimation setting.

Abstracts for Posters

The following pages list the abstracts for posters, in alphabetical order.

Elizabeth Allen, University College London

Flare in Systemic Lupus Erythmatosus

Systemic Lupus Erythmatosus is a syndrome of multifactorial aetiology, characterized by widespread inflamation, most commonly affecting women during the child-bearing years.

In studies of many chronic medical conditions such as SLE the health status of a patient may be characterised using a finite number of disease states. The statistical models used for such a disease must reflect the variation in the disease over relatively long periods of time.

The BILAG index provides a comprehensive approach to measuring disease activity in SLE, assigning one of five scores to disease activity in eight organ systems. Given this disease activity in SLE can be regarded as a multivariate multistate process where the five BILAG scores correspond to states within eight correlated univariate processes.

A number of possible approaches to modelling disease activity in SLE are considered and the results of the final analysis are presented.

Getulio Amaral, University of Nottingham

Bootstrap Confidence Regions in Shape Analysis

The bootstrap method for constructing confidence regions with directional data of Fisher, Hall, Jing and Wood (vol 91, pgs 1062:71,JASA) can be adapted to the statistical shape analysis context. This poster presents a bootstrap approach for building a confidence region for the mean shape. The method is applied to a real data set and some simulation results are presented in order to compare the bootstrap method to other procedures. The simulation results show that the bootstrap method is better in some important situations.

*This research is funded by CNPQ, an agency of the Brazilian government.

Steven Bate, University College London	Judith Anzures Cabrera, University of Warwick
Generalized Linear Models for Wind Speeds over Northern Europe	Long Term Survival after the A-bomb
Within this poster we show how Generalized Linear Models (GLMs) can be used to model wind speeds across Northern Europe. We attempt to model a very large data set, which consists of daily observations taken over a network of 120 sites across Northern Europe for a period of 41 years (1958-1998). When fitting GLMs to this data set, we assume either a gamma or Weibull distribution for the wind speeds, while all predictor variables can be viewed as belonging to one of four broad sub-groups, these being, geographic effects, seasonal effects, autocorrelation effects and effects relating to teleconnection patterns such as the North Atlantic oscillation. We show how the fitted GLMs can provide us with a detailed understanding of the relationship between wind speeds and the various predictor variables, as well as enabling us to identify possible trends in wind speeds over time. We can also use the fitted GLMs to generate realistic future scenarios for wind speeds.	A Weibull model was fitted to explain the long term survival times of people who were registered in a Medical Centre in Hiroshima in 1968, including people who were and were not exposed to the radiation of the atomic bomb. Age, gender, logarithm of radiation and the interaction of age with gender and logarithm of radiation were found to be the covariates that better explain the survival times. Despite the fact that the Kaplan-Meier estimation of the survivor function showed that the late-entries had a higher probability of surviving than the registered population. A conditional-survival model using the same covariates of the model for the registered population was found to be suitable to model the survival times of people that came later to the study.

Nathanaël Benjamin, University of Oxford	Laura Gross, University of Hull
Bound on an Approximation for the Distribution of the Extreme Fluctuations of Exchange Rates	Changes in alcohol consumption in the UK
Large fluctuations of Euro and Sterling/US dollar exchange rates are investigated with respect to compound Poisson approximations as described in Embrechts, Klueppelberg and Mikosch (1997). In particular, recent theoretical bounds on compound Poisson approximations derived by Barbour, Novak and Xia (2000) are compared to data, using ARIMA-GARCH models.	Longitudinal studies, or repeated measures investigations as they are sometimes known, are comprised of repeated observations of an outcome variable and a set of covariates for each of many subjects. Since the repeated observations are made on the same individuals, correlation among a subject's measurements must be taken into account. It is this fact that distinguishes longitudinal studies from cross-sectional ones where data is recorded at one time point only. This project examines longitudinal alcohol data gathered from the Health and Lifestyle Surveys (HALS) of 1984/5 (HALS1) and 1991/2 (HALS2). The first main aim is to describe the marginal expectation of the outcome, for example, the probability of being a drinker at any given time, as a function of several predictor variables. Population averaged marginal models are investigated, in particular, the generalized estimating equations (GEE's) of Liang and Zeger. Since missing data is commonplace in the HALS data, the GEE's must be adapted to take account of this. Hence, the idea of weighted general generalized estimating equations (WGEE) arises. The second main aim is to model the changes in drinking patterns over time for various groups of the population. Based on transitional models, multinomial logit models are used which allow an ordinal rather than just a binary response variable to be used, These are used to model the probabilities of changing states over time, for example, becoming a drinker at HALS2 having been a non-drinker at HALS1. This project is a study of a large longitudinal data set that suffers from the problem of a significant quantity of missing data.

Elizabeth Heron, Trinity College, Dublin	Giovanni Montana, University of Warwick
Simulation, MCMC and Bayesian Statistics	Stepuniform coupling for Perfect Simulation
The use of MCMC in the world of Bayesian inference is commonplace. The methods are seen by some as a "magic solution" when traditional avenues of analysis have failed. Many research projects involve the use of these methods, often without a proper appreciation of the finer points underlying the algorithms. In our experience, good random number generation is tricky, when combined with the subtleties of MCMC it becomes hard. It is good to be sure that the basics are correct before tackling a complex problem. In this poster presentation we examine the performance of the Gibbs sampler and the more general Metropolis Hastings algorithm for a simple problem. Some considerations to be aware of when generating gamma deviates are highlighted. A complex situation where such methods have been applied is in the analysis of fatigue crack propagation in steel. For this model we have implemented a more complicated version of the above method and we present some results on this current research.	We describe a coupling method based on slicing density functions and its use within the "coupling from the past" protocol for perfect simulation with Markov chains. An application of the coupler is given to sampling from the equilibrium distribution of some non-linear autoregressive models.

Ingelin Steinsland, Trinity College, Dublin	Alec Stephenson, Lancaster University
Parallel exact sampling of Gaussian Markov Random Fields.	Simulating Multivariate Extreme Value Distributions
Exact sampling from Gaussian Markov Random Fields (GMRFs) are in theory straight forward. But it involves inverting a n*n matrix to sample a n-dimensional field. We can take advantage of the spare structure of the precision matrix (the inverse of the covariance matrix) the Markov property gives. This has been done for speeding-up sampling (H. Rue (2001) Fast sampling of Gaussian Markov random fields. Journal of the Royal Statistical Society, Series B, 63(2):325-338). We have now extended this to parallel sampling of GMRF. There are two main reasons for parallelization: a) To gain further speed-up and b) to be able to sample from fields that else would be too large. Methods and software are from numerical analysis. We have adjusted the software, tested sampling on a parallel computer, and given the methods a statistical interpretation.	Methods are given for simulating from symmetric and asymmetric versions of the multivariate logistic distribution, and from other multivariate extreme value distributions based on the well known logistic model. We consider two general approaches. The first approach uses transformations to derive random variables with a joint distribution function from which it is easy to simulate. The second approach derives from a specification of conditionally independent marginal components, conditioning on positive stable random variables. This specification extends to models of nested or hierarchical type and leads to an efficient way of incorporating marginal censoring. All algorithms used are available on request from the author. They are also included in the R package EVD, which is available from http://www.maths.lancs.ac.uk/~stephena/.

Elizabeth Traiger, University of Oxford

An Investigation of Floods Along the River Thames

This is joint work with P.J. Northrop, D.R. Cox

In 1999, the Flood Estimation Handbook was published as a guide for investigating the nature of floods in the United Kingdom. It includes statistical methods as well as data of yearly maximal flows for gauged sites across Britain and Northern Ireland. We use results from Extreme Value theory to analyze data from the Thames River. A graphical method is used to assess the independence between annual maxima. The possibility of climate change is addressed by the fitting of trends to the data. Spatial dependence between the flood series at sites along the Thames in investigated using correlation analysis.

Alphabetical List of Presentations

Forename	Surname	Session	Title
Samson	Adebayo	1a	Analysis of Sickle Anaemia Data: a Bayesian Semiparametric Approach
Osho	Ajayi	1b	Dynamic principal component methods for studying shape variability
Elizabeth	Allen	Poster	Flare in Systemic Lupus Erythmatosus
Yasin	Al-Tawarah	6a	Some properties of the Logistic Proportional Hazard Model
Anton	Altmann	4b	Regression of underdispersed count data
Getulio	Amaral	Poster	Bootstrap Confidence Regions in Shape Analysis
Judith	Anzures-Cabrera	Poster	Long Term Survival after the A-bomb
Peter	Barker	6a	Tha gamma frailty model - some issues
Steven	Bate	Poster	Generalized Linear Models for Wind Speeds over Northern Europe
Paul	Baxter	4a	Differentiating Noisy Time Series: Wavelet Methods for Radiocommunications data
Gabriele	Beissel	4b	Variance Estimation in the Presence of Imputed Values
Nathanaël	Benjamin	Poster	Bound on an Approximation for the Distribution of the Extreme Fluctuations of Exchange Rates
Matthew	Burnell	3c	Use of Generalized Procrustes Analysis in Assessing Behavioural Expression in Dogs
Adam	Butler	5b	Quantifying the impact of climate change upon extreme sea levels
David	Cairns	7b	Omitted and Supernumerary Values in Time Series
Nikki	Carlton	4a	Analysing Protein Structure Using a Wavelet Lifting Transform
High Seng	Chai	6c	New skew normal distributions and their applications
Constantinos	Chappas	4c	Estimating the Tangency Portfolio on the Markowitz Efficient Frontier
Louise	Choo	4b	Investigating Spatial Variations of Disease in Epidemiology
Matthew	Coates	7b	Statsoft Presentation
Diana	Cole	3a	Fitting Generalized Linear Mixed Models to Strawberry Inflorescence Data
Codina	Cotar	3c	New developments in the lilypond
Carlos	Cuevas-Covarrubias	3b	Using Gaussian Kernels To Estimate The Area Under The ROC
Ali	Daneshkhah	5c	Multicausal Prior Families For Patterns of DAGs
Carolyn	Davies	5a	Modelling Cancer Mortality in Europe
Jonathan	Donnelly	3c	Simultaneous Confidence Intervals in Linear Modelling
Lee	Fawcett	5b	Inference for Exceedances over High Thresholds
Jose	Ferreira	1a	Weighted naive Bayes modelling for data mining
Chris	Ferro	5b	Estimating Characteristics of Extreme Events
Piotr	Fryzlewicz	7b	Forecasting non-stationary time series by wavelet process modelling
Ruth	Fuentes-Garcia	1c	Bayesian methods for clustering
Catherine	Fullwood	3a	Longitudinal data analysis of anticoagulant data
Stuart	Gardiner	3b	If You Carry On Like That You'll Go Blind…
Linda	Garside	1c	Dynamic lattice Markov spatio-temporal models for environmental data
Christina	Goldschmidt	2c	Limiting Behaviour in an Urn Model
Susan	Gooding	1c	Spatial Clustering and the Addition of Covariates
Peter	Gregory	1a	Fractional Bayes Factors applied to a medical cost data set
Laura	Gross	Poster	Changes in alcohol consumption in the UK
David	Harrison	5a	Deterministic estimates of disease progression parameters from retrospective data

Michael	Heeneman	2a	The Analysis of High Density Genotyping Information
Elizabeth	Heron	Poster	Simulation, MCMC and Bayesian Statistics
Katherine	Hunt	4a	Image analysis and wavelets
Zahid	Hussain	6c	Choice of Statistical Methods for Comparing Heuristic Performance
Dan	Jackson	6b	Meta Analysis - what it is, what it isn't.
Gareth	James	7c	Methodology at the Office for National Statistics, and Academic Research
Lara	Jamieson	2c	Duck a lá reversible jump
Matthew	Jones	6a	Randomised Enrollment in Multi-Centre trials
Constantinos	Kallis	5c	Exploratory analysis using graphical models
Ngianga-Bakwin	Kandala	5a	Exploring childhood undernutrition and mortality in Africa
David	Leslie	7c	Learning by playing: results in multi-agent reinforcement learning
Talal	Maatouk	3a	Least Squares Fitting of Compact Set-Valued Data
Miguel	Marques dos Santos	7c	Modelling mate choice & sexual selection
Andrew	McMullan	7b	Modelling Water Quality in the Clyde Estuary
Kevin	McNally	4c	Uncertainty in Financial Models of Large and Complex Government Projects
Helen	Mitchell	2b	The use of index policies to solve a machine maintenance problem
Giovanni	Montana	Poster	Stepuniform coupling for Perfect Simulation
Kassim	Mwitondi	1b	Using Boosting in Classification
Zorana	Najdanovic	2c	Conditioning an additive functional of a finite state space Markov chain
Dimitris	Nicoloutsopoulos	7c	Estimation of the association parameter for bivariate Archimedean Copulas
David	O'Donnell	5c	Bayesian Graphical Gaussian Model Selection
David	Ohlseen	1a	The use of statistical modelling to identify divergent performance in health care
Marilyn	O'Keeffe	2b	A Whittle index policy for a two-class queueing system with quadratic holding costs
Keith	O'Rourke	5a	Fisher, Combination of Observations (Meta-analysis) and Ancillary Statistics
Alexandra	Ramos	5b	Asymptotically independent joint tail modelling in practice
Sammy	Rashid	1c	Inference Techniques for Spatio-Temporal Pollution Data
Maria Fatima	Salgueiro	5c	The Observed Association Structure from Graphical Log-LinearModels with a Binary Latent Variable
Richard	Samworth	6c	How confident are you of your normality?
George	Savva	2a	Chromosomal Evolution and Phylogenetic Inference
Ingelin	Steinsland	Poster	Parallel exact sampling of Gaussian Markov Random Fields
Alec	Stephenson	Poster	Simulating Multivariate Extreme Value Distributions
Elizabeth	Traiger	Poster	An Investigation of Floods Along the River Thames
Nikolaos	Tzavidis	1b	Estimating Labour Force Gross Flows in the Presence of Misclassification and Validation Information
Veronica	Vinciotti	1b	Local versus Global Models for Classification Problems
Anna-Jane	Vine	6b	Group Screening for Experiments with Large Numbers of Factors
Wiesner	Vos	2a	Robust hierarchical models for oligonucleotide gene expression data.
Yanzhong	Wang	4c	Fitting Mixtures by Sequential Simplification
Ben	Wright	2b	Experiences Modelling a Traffic Network
Fiona	Young	6b	Use of prognostic modelling to optimise clinical trial design in acute ischaemic stroke
Kamila	Zychaluk	3b	Can noise be useful in statistics and how?

Alphabetical List of Chairs

Forename	Surname	Session	Forename	Surname	Session
Ali	Alshabani	5c	Matthew	Gander	7c
Getulio	Amaral	3c	Sara	Geneletti	5b
Peter	Barker	4b	Alex	Glaser	2c
Paul	Baxter	7b	Michael	Heeneman	6a
Gabriele	Beissel	6c	Ngianga-Bakwin	Kandala	1b
Denise	Brown	2a	David	Leslie	2b
High Seng	Chai	4c	David	Ohlseen	5a
Louise	Choo	3a	Sofia	Olhede	4a
Alexander	Cox	1c	Keith	O'Rourke	6b
Jonathan	Donnelly	1a	Nikolaos	Tzavidis	3b

Key to Sessions

Tuesday 19th March 2002

1	09.15 – 11.00
2	11.30 – 12.45
3	14.00 – 15.20
Poster	15.20 – 16.20
4	16.25 – 17.45

Wednesday 20th March 2002

5	09.15 – 11.10
6	11.30 – 12.45
7	14.00 – 16.00

All talks take place in the School of Engineering.
Those sessions listed with an "a" take place in F107.
Those sessions listed with a "b" take place in F110.
Those sessions listed with a "c" take place in F111.
Posters will be displayed in the Engineering Concourse

Alphabetical List of Delegates

Title	Forename	Surname	University	Email
Mr	Anton	Altmann	City University	anton321@hotmail.com
Mr	John	Harrison	City University	johnharrison@hotmail.com
Miss	Hannah	Aldgate	Imperial College London	hannah.aldgate@ic.ac.uk
Mr	Jose	Ferreira	Imperial College London	jt.ferreira@ic.ac.uk
Mr	Matthew	Gander	Imperial College London	matthew.gander@ic.ac.uk
Mr	Adrien	Jamain	Imperial College London	adrien.jamain@ic.ac.uk
Miss	Bing bing	Liu	Imperial College London	bing.liu@ic.ac.uk
Mr	Ross	McDonald	Imperial College London	Ross.McDonald@ic.ac.uk
Ms	Sofia	Olhede	Imperial College London	s.olhede@ic.ac.uk
Mr	Alexandre	Pintore	Imperial College London	alexandre.pintore@ic.ac.uk
Miss	Veronica	Vinciotti	Imperial College London	v.vinciotti@ic.ac.uk
Mr	George	Savva	John Innes Centre	george.savva@bbsrc.ac.uk
Mr	Yasin	Al-Tawarah	Keele University	mad26@keele.ac.uk
Mr	David	Harrison	MRC Biostatistics Unit, Cambridge	david.harrison@mrc-bsu.cam.ac.uk
Mr	David	Ohlseen	MRC Biostatistics Unit, Cambridge	david.ohlssen@mrc-bsu.cam.ac.uk
Mr	Stuart	Gardiner	Nottingham Trent University	Stuart.Gardiner@ntu.ac.uk
Ms	Linjuan	Sun	Open University	L.Sun@open.ac.uk
Mr	Ben	Wright	Open University	b.j.wright@open.ac.uk
Mr	Sourabh	Bhattacharya	Trinity College Dublin	bhattacs@tcd.ie
Ms	Eleisa	Heron	Trinity College Dublin	eheron@tcd.ie
Mr	Georgios	Stefanou	Trinity College Dublin	stefanog@tcd.ie
Ms	Ingelin	Steinsland	Trinity College Dublin	steinsli@tcd.ie
Ms	Liz	Allen	University College London	liz@stats.ucl.ac.uk
Mrs	Anne	Andrew	University College London	anne@stats.ucl.ac.uk
Mr	Steven	Bate	University College London	steven@stats.ucl.ac.uk
Mr	Constantinos	Chappas	University College London	c.chappas@ucl.ac.uk
Miss	Sara	Geneletti	University College London	sara@stats.ucl.ac.uk
Mr	Michael	Heeneman	University College London	michael_heeneman@hotmail.com
Mr	Dimitris	Nicoloutsopoulos	University College London	dimitris@stats.ucl.ac.uk
Mr	John	Aquilina	University of Bath	mapja@maths.bath.ac.uk
Miss	Louise	Choo	University of Bath	mapllc@bath.ac.uk
Mr	Alexander	Cox	University of Bath	mapamgc@maths.bath.ac.uk
Miss	Ruth	Fuentes-Garcia	University of Bath	R.S.Fuentes@bath.ac.uk
Mr	Chris	Page	University of Bath	c.j.page@bath.ac.uk
Mr	Richard	Sharp	University of Bath	maprms@bath.ac.uk
Mr	Surbjeet	Singh	University of Bath	mapss@bath.ac.uk
Ms	Kamila	Zychaluk	University of Birmingham	zychaluk@maths.bham.ac.uk
Miss	Nikki	Carlton	University of Bristol	nikki.carlton@bris.ac.uk
Miss	Codina	Cotar	University of Bristol	Cody.Cotar@bris.ac.uk
Mr	Piotr	Fryzlewicz	University of Bristol	p.z.fryzlewicz@bris.ac.uk
Mr	David	Hastie	University of Bristol	david.hastie@bristol.ac.uk
Mr	Tim	Heaton	University of Bristol	Tim.Heaton@bristol.ac.uk
Miss	Katherine	Hunt	University of Bristol	katherine.hunt@bristol.ac.uk
Mr	David	Leslie	University of Bristol	d.s.leslie@bris.ac.uk
Mr	Miguel	Marques dos Santos	University of Bristol	miguel.m.santos@bristol.ac.uk
Miss	Juliet	Chapman	University of Cambridge	juliet.chapman@cimr.cam.ac.uk
Ms	Christina	Goldschmidt	University of Cambridge	cag27@cam.ac.uk
Miss	Olivia	Grigg	University of Cambridge	olivia.grigg@mrc-bsu.cam.ac.uk
Miss	Lara	Jamieson	University of Cambridge	lej25@statslab.cam.ac.uk
Mr	Richard	Samworth	University of Cambridge	rjs57@cam.ac.uk
Mr	Mark	Simmonds	University of Cambridge	mark.simmonds@mrc-bsu.cam.ac.uk
Mr	Matthew	Jones	University of Cardiff	JonesMC1@cf.ac.uk

Mr	Matthew	Killeya	University of Durham	m.r.killeya@durham.ac.uk
Mr	Benjamin	Wyllie	University of Durham	b.j.wyllie@durham.ac.uk
Mr	Paul	Baxter	University of Essex	pdbaxt@essex.ac.uk
Miss	Julie	Badcock	University of Exeter	j.badcock@ex.ac.uk
Mrs	Heather	Turner	University of Exeter	Heather.L.Turner@ex.ac.uk
Mr	Osho	Ajayi	University of Glasgow	shola@stats.gla.ac.uk
Miss	Denise	Brown	University of Glasgow	denise@stats.gla.ac.uk
Mr	Matthew	Burnell	University of Glasgow	matthew@stats.gla.ac.uk
Miss	Carolyn	Davies	University of Glasgow	carolyn@msoc.mrc.gla.ac.uk
Miss	Michela	Gnaldi	University of Glasgow	michela@stats.gla.ac.uk
Mr	Mohsen	Hosseini	University of Glasgow	Mohsen@stats.gla.ac.uk
Miss	Katsiaryna	Kaval	University of Glasgow	katya@stats.gla.ac.uk
Mr	Talal	Maatouk	University of Glasgow	talal@stats.gla.ac.uk
Mr	Andrew	McMullan	University of Glasgow	a.mcmullan@stats.gla.ac.uk
Mr	Luigi	Palla	University of Glasgow	luigi@stats.gla.ac.uk
Mr	Yanzhong	Wang	University of Glasgow	wang@stats.gla.ac.uk
Miss	Fiona Beveridge	Young	University of Glasgow	fby1w@clinmed.gla.ac.uk
Mr	Zahid	Hussain	University of Huddersfield	z.hussain@hud.ac.uk
Miss	Laura	Gross	University of Hull	laura@bayesian-bay.freeserve.co.uk
Mrs	Diana	Cole	University of Kent at Canterbury	djp5@ukc.ac.uk
Miss	Zarina	Mohd Khalid	University of Kent at Canterbury	zm6@ukc.ac.uk
Mr	David	Morris	University of Kent at Canterbury	dlm4@ukc.ac.uk
Mrs	Dilhari	Attygalle	University of Lancaster	d.attygalle@lancaster.ac.uk
Mr	Peter	Barker	University of Lancaster	p.n.barker@lancaster.ac.uk
Mr	Adam	Butler	University of Lancaster	a.butler@lancaster.ac.uk
Mrs	Deborah	Costain	University of Lancaster	d.costain@lancs.ac.uk
Ms	Audrienne	Cutajar Bezzina	University of Lancaster	a.bezzina@lancaster.ac.uk
Mr	Chris	Ferro	University of Lancaster	c.ferro@lancaster.ac.uk
Miss	Rosemeire	Fiaccone	University of Lancaster	r.fiaccone@lancaster.ac.uk
Miss	Catherine	Fullwood	University of Lancaster	c.fullwood@lancaster.ac.uk
Miss	Suely	Giolo	University of Lancaster	s.giolo@lancaster.ac.uk
Miss	Susan	Gooding	University of Lancaster	s.gooding@lancaster.ac.uk
Mr	Iraj	Kazemi	University of Lancaster	i.kazemi@lancaster.ac.uk
Mr	Md Azman	Shahadan	University of Lancaster	a.shahadan@lancaster.ac.uk
Mr	Alec	Stephenson	University of Lancaster	a.stephenson@lancaster.ac.uk
Ms	Ting-Li	Su	University of Lancaster	ting-li.su@lancaster.ac.uk
Mr	Kassim	Mwitondi	University of Leeds	mwitondi@amsta.leeds.ac.uk
Mr	Nathan	Green	University of Liverpool	nat1@liverpool.ac.uk
Mr	Samson Babatunde	Adebayo	University of Munich, Germany	samson@stat.uni-muenchen.de
Mr	Ngianga-Bakwin	Kandala	University of Munich, Germany	kandala@stat.uni-muenchen.de
Mr	Mohammed	Al-Sadi	University of Newcastle	alsadi@hotmail.com
Mr	Ryan	Dunn	University of Newcastle	R.T.Dunn@ncl.ac.uk
Mr	Lee	Fawcett	University of Newcastle	lee.fawcett@ncl.ac.uk
Miss	Linda	Garside	University of Newcastle	l.m.garside@ncl.ac.uk
Mr	Philip	Giles	University of Newcastle	P.R.Giles@ncl.ac.uk
Mr	Richard	Lumley	University of Newcastle	Richard@lumley1.screaming.net
Miss	Helen	Mitchell	University of Newcastle	h.m.mitchell@ncl.ac.uk
Mrs	Marilyn	O'Keeffe	University of Newcastle	Marilyn.O'Keeffe@ncl.ac.uk
Mr	Peter	Philipson	University of Newcastle	p.m.philipson@ncl.ac.uk
Miss	Elizabeth	Smith	University of Newcastle	elizabeth.smith2@ncl.ac.uk
Mr	Ali K.	Alshabani	University of Nottingham	pmxaksa1@nottingham.ac.uk
Mr	Getulio	Amaral	University of Nottingham	getulio.amaral@maths.nottingham.ac.uk
Ms	Xue	Wang	University of Nottingham	pmxxw@maths.nottingham.ac.uk
Mr	Nathanaël	Benjamin	University of Oxford	nathanael.benjamin@centraliens.net
Mr	Keith	O'Rourke	University of Oxford	orourke@stats.ox.ac.uk

Mr	Alastair	Pickett	University of Oxford	pickett@stats.ox.ac.uk
Miss	Elizabeth	Traiger	University of Oxford	traiger@stats.ox.ac.uk
Mr	Wiesner	Vos	University of Oxford	vos@stats.ox.ac.uk
Miss	Carla	Ysusi Mendoza	University of Oxford	carlamym@yahoo.com
Mr	Ailton	Andrade	University of Sheffield	j.a.andrade@sheffield.ac.uk
Mr	David	Cairns	University of Sheffield	stp01dac@sheffield.ac.uk
Mr	Peter	Gregory	University of Sheffield	
Mr	Kevin	McNally	University of Sheffield	K.McNally@shef.ac.uk
Mr	Sammy	Rashid	University of Sheffield	stp00snr@sheffield.ac.uk
Miss	Elsa	Valdes-Marquez	University of Sheffield	stp01ev@sheffield.ac.uk
Ms	Gabriele	Beissel	University of Southampton	gbeissel@socsci.soton.ac.uk
Mr	High Seng	Chai	University of Southampton	hsc264@soton.ac.uk
Mr	Wee_Swan Leslie	Chiang	University of Southampton	wc801@soton.ac.uk
Mr	James	Dartnall	University of Southampton	jed@maths.soton.ac.uk
Mr	Marcel	De Toledo Vieira	University of Southampton	marcel@socsci.soton.ac.uk
Mr	Jonathan	Donnelly	University of Southampton	jhmd@maths.soton.ac.uk
Miss	Kathryn	Hoad	University of Southampton	kah@maths.soton.ac.uk
Mr	Ebrahim	Khodaie Biramy	University of Southampton	ekb@soton.ac.uk
Mr	David	O'Donnell	University of Southampton	dod@maths.soton.ac.uk
Miss	Veronica	Rapley	University of Southampton	ver@maths.soton.ac.uk
Miss	Maria Fatima	Salgueiro	University of Southampton	mffs@socsci.soton.ac.uk
Mr	Rob	Stapleton	University of Southampton	rds@maths.soton.ac.uk
Mr	Nikolaos	Tzavidis	University of Southampton	ntzav1@socsci.soton.ac.uk
Miss	Anna-Jane	Vine	University of Southampton	aev@maths.soton.ac.uk
Miss	Emily	Webb	University of Southampton	elw@maths.soton.ac.uk
Mr	Alex	Glaser	University of Surrey	a.glaser@eim.surrey.ac.uk
Mrs	Alexandra	Ramos	University of Surrey	a.ramos@eim.surrey.ac.uk
Miss	Judith	Anzures-Cabrera	University of Warwick	J.Anzures-Cabrera@warwick.ac.uk
Mr	Simon	Bond	University of Warwick	bond@stats.warwick.ac.uk
Mr	Ben	Cowling	University of Warwick	bjc@stats.warwick.ac.uk
Mr	Carlos	Cuevas-Covarrubias	University of Warwick	strbt@csv.warwick.ac.uk
Mr	Ali	Daneshkhah	University of Warwick	A.Daneshkhah@warwick.ac.uk
Mr	Dan	Jackson	University of Warwick	Daniel.Jackson@warwick.ac.uk
Mr	Constantinos	Kallis	University of Warwick	C.Kallis@warwick.ac.uk
Mr	Giovanni	Montana	University of Warwick	montana@stats.warwick.ac.uk
Miss	Zorana	Najdanovic	University of Warwick	zn18@hotmail.com
Mr	Sandeep	Shah	University of Warwick	strci@csv.warwick.ac.uk

RSC2003

University of Surrey, Guildford

7th – 10th April 2003

for more details, and to keep updated with RSC2003 information, visit www.maths.surrey.ac.uk/rsc2003/

Meet Our Sponsors

Thanks to all our sponsors, who have helped to make RSC2002 possible:

The Royal Statistical Society
http://www.rss.org.uk/

Springer Verlag
http://www.springer.co.uk/

Biomathematics and Statistics Scotland
http://www.bioss.ac.uk/

Eurandom
http://www.eurandom.nl/

Biometrics (GlaxoSmithKline Statistics)
http://www.biometrics.com/

Insightful
http://www.insightful.com/

Office for National Statistics
http://www.statistics.gov.uk/

Pfizer
http://www.pfizer.co.uk/

MRC Biostatistics Unit, Cambridge
http://www.mrc-bsu.cam.ac.uk/

Statsoft Ltd.
http://www.statsoft.co.uk/

ROYAL STATISTICAL SOCIETY FOUNDED 1834

Springer Verlag

BioSS

EURANDOM

gsk GlaxoSmithKline

Insightful™ intelligence from data

national STATISTICS

Pfizer

MRC Biostatistics Unit Cambridge

STATISTICA
High performance single-user and enterprise-wide statistical software
data analysis data mining quality control