The Open University Statistics Department/Group

2008 Report

1. Membership

Professor P.H. Garthwaite (Head of Group from May)
Professor C.P. Farrington (Head of Group to April)
Dr C.J. Albers
Dr K. Anaya
Professor F. Critchley
Dr A.E. Faria
Professor J.C. Gower
Dr M.N. Hocine
Dr G.E. Iossif
Professor M.C. Jones
Mr I.W. Martin
Dr K.J. McConway
Dr C.M. Queen
Dr N.T. Trendafilov
Dr S.K. Vines
Dr H.J. Whitaker

Long-term visitor

Dr P.C. Taylor (University of Hertfordshire)

Short–term visitors

Dr S. Kato (Keio University, Japan)
Dr P.V. Larsen (university of Southern Denmark)
Professor N. Le Roux (University of Stellenbosch, South Africa)
Professor P.K. Marriott (University of Waterloo, Canada)
Professor P.W. Vos (East Carolina University, U.S.A.)

Full–time research students

Mr T. Collins (Joint with Computing, from October)
Mr S. De
Mr F Elfadaly (from October)
Mr D. Gragn
Ms A. Noufaily
Mr S. Unkel

Part–time research students

Mr N. Calleja
Ms A. Gjini
Mr G. Kafatos
Mr M. Merens
Mr E. Mubwandarikwa
Mr A. Owen
Mr J. Urquhart (joint with Applied Mathematics)

Support staff

Mrs S. Frain, Group Secretary

2. Introduction

2008 was, inevitably, another active year for Statistics at the Open University in both research and teaching. We had an outstanding success in the Research Assessment Exercise. Our overall profile is 5-40-45-5-0. Whichever league table you read this places us within the top 20 Statistics groups in the country. Congratulations to all concerned!

2008 saw no changes in personnel. Our course BM240 *Quantitative Methods in Business* saw its final presentation, while the revision of M346 *Linear Statistical Modelling* was completed. Work started on an exciting new development, *M347 Mathematical Statistics* (see below) and on a revision of M343 *Application of Probability*. The Group's courses attracted a total of over 2300 students in 2008.

In the course of the year, members of the Department published 20 papers in refereed journals (see Section 6), plus contributions to conference proceedings. Pre-publication papers and other research material can be found in our technical report series, at http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm.

Alvaro Faria (assisted by Paul Garthwaite) continued to run a Statistical Advisory Service to support research in other departments at the Open University. A website explains the service: http://www.mathematics.open.ac.uk/advisory.

Further information on the Department is available through our web pages at http://statistics.open.ac.uk/index.html

3. Teaching

Presentation

The Department presented five courses of 30 CATS points each, namely, BM240 *Quantitative Methods in Business*, M248 *Analysing Data*, M249 *Practical Modern Statistics*, M343 *Applications of Probability* and M346 *Linear Statistical Modelling*, and one course of 10 CATS points, SMK184 *Chance, Risk and Health*, which runs four times a year.

We continue to be particularly proud of M249. A standard OU performance indicator is the percentage of respondents to end-of-course surveys that say they definitely or mostly agreed with the statement "Overall I am satisfied with the quality of this course". For M249, the percentage was 99.1% out of 222 respondents. Only 3 of the 136 other courses surveyed got a higher score and 2 of these have only small numbers of responsdents and are special presentations of standard courses that in their main form did worse than M249 on this measure).

Each of the first level mathematics courses, MU120 *Open Mathematics* and MST121 *Using Mathematics*, contains substantial amounts of statistical material (roughly one quarter of each) with which Alvaro Faria and Nickolay Trendafilov were involved.

Kevin McConway contributed to the presentation of SDK125 *Introducing Health Science*.

Production

Karen Vines worked extremely hard on almost single-handedly updating (both computationally and presentationally) the content of M346, the revised version to go out to students in 2009.

The Group has started production of *M347 Mathematical Statistics*. As well as enhancing our third level provision with a new course that gives "the theory behind the methods", this course is notable for leading the way within the Department in terms of electronic presentations M347 is to be presented entirely on-line, and we are investigating ways of making the very best use of modern web-based tools. Catriona Queen and Chris Jones are course team co-chairs.

Gillian Iossif is leading the revision of *M343 Applications of Probability.* This very successful but old, course is in need of a certain amount of improvement and updating.

Alvaro Faria collaborated with Alan Graham (Mathematics Education) on the production of two units on Statistics for MU123 *Discovering Mathematics*, the forthcoming replacement for

Awards

Our courses are core components of degrees in Mathematics and Statistics and in Computing and Statistics (graduates of which are only now starting to emerge) and degrees in Mathematical Sciences and in Economics and Mathematical Sciences, while BM240 is part of the BA in Business Studies.

The undergraduate Diploma in Statistics is now awarded to students who successfully study M248 or BM240, M249, M343 and M346 (or their predecessors). A little under 100 students are awarded the Diploma each year.

4. Ph.D. awarded

Emmanuel Mubwandarikwa for his thesis *Modality Conditions and Prior Weights in the Geometric Combination of Bayesian Forecasting Models* (supervisors: A.E. Faria and P.H. Garthwaite)

5. Research interests

The Department is home to three Research Groups:

Multivariate Statistics (Critchley (chair), Albers, Anaya, De, Faria, Garthwaite, Gower, Gragn, Jones, Queen, Taylor, Trendafilov, Unkel, Vines, Constantine (CSIRO, retired), Cook (University of Minnesota), Marriott (University of Waterloo)).

Bayesian Statistics (Queen (chair), Albers, De, Faria, Farrington, Garthwaite, Jenkinson, McConway, Vines, Whitaker).

Medical Statistics (Whitaker (chair), Farrington, Garthwaite, Hocine, Vines, Jenkinson (University of Aberdeen), Musonda (University of East Anglia).

Clearly, these three general groupings cover a huge variety of research projects in Statistics. Inter alia, they organise informal internal research meetings. Research also continues, of course, on topics not directly covered by these Research Groups, of which distribution theory, kernel smoothing, robust statistics, time series forecasting and influence analysis are but a small selection. Strong ongoing collaborations continue in areas such as medicine, psychology and ecology.

For details, see the publications, talks and other items to follow.

6. Publications

6.1 Publications in refereed academic journals

Albers, C.J. & Schaafsma, W. (2008) Goodness of fit testing using a specific density estimate. *Statistics and Decisions*, 26, 3-23.

Crawford, J.R. & **Garthwaite, P.H.** (2008) On the "optimal" size for normative samples in neuropsychology: Capturing the uncertainty when normative data are used to quantify the standing of a neuropsychological test score. *Child Neuropsychology*, 14, 99-117.

Crawford, J.R., Sutherland, D. & **Garthwaite, P.H.** (2008) On the reliability and standard errors of measurement of contrast measures from the D-KEFS. Journal of the International Neuropsychological Society, 14, 1069-1073.

Critchley, F. & **Jones, M.C.** (2008) Asymmetry and gradient asymmetry functions: density-based skewness and kurtosis. *Scandinavian Journal of Statistics,* 35, 415-437.

Faria, A.E. & **Mubwandarikwa, E**. (2008) Multimodality on the geometric combination of Bayesian forecasting models. *International Journal of Statistics and Management Systems*, 3, 1-25.

Faria, A.E. & **Mubwandarikwa, E.** (2008), The geometric combination of Bayesian forecasting models. *Journal of Forecasting*, 27, 519-535.

Garthwaite, P.H., Chilcott, J.B., Jenkinson, D.J. & Tappenden, P. (2008) Use of expert knowledge in evaluating costs and benefits of alternative service provision: A case study. *International Journal of Technology Assessment in Health Care*, 24, 350-357.

Gower, J.C. (2008) The biological stimulus to multidimensional data analysis. *Electronic Journ@l for History of Probability and Statistics*, 4.

Jones, M.C. (2008) The distribution of the ratio X/Y for all centred elliptically symmetric distributions. *Journal of Multivariate Analysis*, 99, 572-573.

Jones, M.C. (2008) On reciprocal symmetry. *Journal of Statistical Planning and Inference,* 138, 3039-3043.

Jones, M.C. (2008) On a class of distributions with simple exponential tails. *Statistica Sinica*, 18, 1101-1110.

Jones, M.C. (2008) The *t* family and their close and distant relations. *Journal of the Korean Statistical Society*, 37, 293-311. Discussion Paper.

Jones, M.C. & Arnold, B.C. (2008) Distributions that are both log-symmetric and R-symmetric. *Electronic Journal of Statistics*, 2, 1300-1308.

Jones, M.C., Park, H., Shin, K.I., **Vines, S.K.** & Jeong, S.O. (2008) Relative error prediction via kernel regression smoothers. *Journal of Statistical Planning and Inference*, 138, 2887-2898.

Lubbe-Gardner, D.L., le Roux, N,J. & **Gower, J.C.** (2008) Measures of fit in principal component and canonical variate analysis *Journal of Applied Statistics*, 35, 947-965.

Musonda, P., **Hocine, M.N.,** Andrews, N.J., Tubert-Bitter, P. & **Farrington, C.P.** (2008) Monitoring vaccine safety using case series cumulative sum charts. *Vaccine*, 26, 5358-5367.

Musonda, P., **Hocine, M.N., Whitaker, H.J.** & **Farrington, C.P.** (2008) Self-controlled case series method: small sample performance. *Computational Statistics and Data Analysis*, 52, 1942–1957.

Queen, C.M., Wright, B. & **Albers, C.J.** (2008) Forecast covariances in the linear multiregression dynamic model. *Journal of Forecasting*, 27, 175-191.

Twala, B.E.T.H., **Jones, M.C.** & Hand, D.J. (2008) Good methods for coping with missing data in decision trees. *Pattern Recognition Letters*, 29, 950-956.

Whitaker, H.J. (2008) The self-controlled case series method: A way to study the relation between antipsychotics and stroke. *British Medical Journal*, 337, 586-587. Editorial.

6.2 Refereed conference proceedings

Albers, C.J. (2008) Some quadratic optimisation problems in psychometrics. In New Trends in Psychometrics, eds: K. Shigemasu, A. Okada, T. Imaizuma, & T. Hodhina, Universal Academy Press, pp. 1-6.
Gower, J.C. (2008) Asymmetry analysis: The place of models. In New Trends in Psychometrics, eds: K. Shigemasu, A. Okada, T. Imaizuma, & T. Hodhina, Universal Academy Press, pp. 69-78.

7. Seminars and conferences at The Open University

The 18th nearly-annual Open University statistics conference was entitled **Statistics for Public Health Surveillance,** and was organised by Heather Whitaker, Sarah Frain and Paddy Farrington. Speakers at this one-day meeting were:

Marianne Frisén (Göteborg University) *Optimal statistical surveillance of public health*

Stephen Evans and David Prieto (London School of Hygiene and Tropical Medicine) *The use of Bayesian hierarchical models in pharmacovigilance*

Mounia Hocine (Open University) *Sequential case series analysis for surveillance of vaccine safety*

Chris Robertson (University of Strathclyde) *Syndromic surveillance at Health Protection Scotland*

Simon Cauchemez (Imperial College London) *Real-time monitoring of infectious disease outbreaks*

Heather Whitaker (Open University) *Serological surveillance of childhood vaccination programmes*

David Spiegelhalter (University of Cambridge and MRC Biostatistics Unit) *Extreme multiplicity: monitoring large numbers of indicators and areas or institutions*

Leonhard Held (Zurich University) *Multivariate modelling of infectious disease surveillance data*

There were also posters presented by: Nick Andrews (Health Protection Agency); David Conesa (University of Valencia); Patrick Musonda (University of East Anglia); V. Gómez-Rubio (Imperial College London); Baltazar Nunes (Instituto Nacional de Saúde Dr. Ricardo Jorge, Portugal); Michaela Paul (University of Zurich); Anne-Elie Carsin (National Cancer Registry, Ireland); and Theo Kypraios (University of Nottingham).

The group also ran it usual seminar programme with invited speakers. Over the year we heard:

Christian Beckmann (John Radcliffe Hospital, Oxford) *Exploratory FMRI group analysis using tensorial extensions to independent component analysis*

Marc Hallin (Université Libre de Brusselles) *From distribution-freeness to semiparametric efficiency: sixty years of rank-based inference*

David Banks (Duke University, Durham, NC) *Statistical issues in metabolomics*

Inge Koch (University of New South Wales, Sydney) *Dimension selection with independent component analysis and its application to prediction*

Abdel Hannachi (University of Reading) *Optimal interpolation patterns: A new tool in climate data analysis*

Richard Samworth (University of Cambridge) *Maximum likelihood estimation of a multidimensional log-concave density*

Nicky Best (University College London) *Enhancing inference from observational studies by combining multiple data sources: a Bayesian graphical modelling approach*

We ran our fifth Annual Research Students' Research Day, where speakers were Doyo Gragn, Neville Calleja, Marco Merens, John Urquhart, Swarup De, Angela Noufaily, Alun Owen and Steffen Unkel.

And we had two internal research mornings. At the first, in April, Paddy Farrington, Mounia Hocine, Frank Critchley and Alvaro Faria spoke; at the second, in November, speakers were Karim Anaya, Heather Whitaker, Casper Albers and Chris Jones.

8. Conference papers and posters presented

K. Anaya

International Conference on Robust Statistics (ICORS) 2008, Antalya, Turkey *Robustness, cuts and geometry*

C.J. Albers

28th International Symposium of Forecasters, Nice, France *Forecasting traffic flows in road networks: A graphical dynamic model approach*

Royal Statistical Society Annual Conference, Nottingham *Forecasting traffic flows in road networks: A graphical dynamic model approach*

F. Critchley

Royal Statistical Society Annual Conference, Nottingham *Emerging geometries for statistical science*

32nd Annual Conference of the German Classification Society. Hamburg. Germany *Principal axis analysis – with HDLSS bonuses*!

C.P. Farrington

24th International Biometrics Conference, Dublin, Ireland *Individual heterogeneity: effects and estimation for directly transmitted infectious diseases,*

24th International Biometrics Conference, Dublin, Ireland *Case series analysis for dependent recurrent events*

24th International Conference on Pharmacoepidemiology and Therapeutic Risk Management, Copenhagen, Denmark *The self-controlled case series method*

Journées de Statistique, Modélisation et Application, Algiers, Algeria *Transmission des maladies infectieuses: Modélisation et hétérogénéité individuelle.*

J.C. Gower

7th International Conference on Social Science Methodology RC33 – Logic and Methodology in Sociology, Naples, Italy *Assessing and displaying fit in principal component analysis and canonical variate analysis biplots*

24th International Biometric Conference, Dublin, Ireland *Concerning biometrics and psychometrics* [poster]

D. Gragn

31st Research Students' Conference in Probability and Statistics (RSC 2008), Nottingham *Dimensionality reduction for data of high dimension*

M.N. Hocine

Statistics for Public Health Surveillance, The Open University *Sequential case series analysis for surveillance of vaccine safety*

Epidemiology and Biometry: Recent Statistical Methods in Epidemiology, Paris, France *The self-controlled case series method*

29th Annual Conference of the International Society for Clinical Biostatistics, Copenhagen, Denmark *Sequential methodology using the case series method*

Journées de Statistique, Modélisation et Application, Algiers, Algeria *Self-controlled case series models for dependent recurrent events*

M.C.Jones

2008 Cherry Bud Workshop, Keio University, Yokohama, Japan *Scaling for skewness, with spin-offs and insights*

Workshop on Skew Symmetric Probability Distributions, Bertinoro, Italy *Some alternative skew distributions*

Spring Meeting of the Korean Statistical Society, Gwangju, South Korea *The t family and their close and distant relations* [special invited lecture]

C.M. Queen

Royal Statistical Society Conference, Nottingham *Intervention and causality in a dynamic Bayesian network*

N.T. Trendafilov

1st Workshop of the ERCIM Working Group on Computing and Statistics, Neuchâtel, Switzerland *Simple and interpretable discrimination*

S.Unkel

International Meeting of the Psychometric Society (IMPS) 2008, New Hampshire, USA *Simultaneous parameter estimation in exploratory factor analysis by weighted least squares*

17th International Workshop on Matrices and Statistics (IWMS '08) (in honour of Professor T. W. Anderson's 90th birthday), Tomar, Portugal *Simultaneous parameter estimation in exploratory factor analysis*

Royal Statistical Society Annual Conference, Nottingham *Independent factor analysis of climate data*

H.J. Whitaker

Statistics for Public Health Surveillance, The Open University *Serological surveillance of childhood vaccination programmes*

24th International Biometrics Conference, Dublin, Ireland *Measures of disassortativeness and their application to directly transmitted infections*

Journees de Statistiques et Application. Algiers, Algeria *Case series analysis for censored, perturbed or curtailed post-event exposures.*

9. Other seminars and talks given

K. Anaya

University of St. Andrews *Sensitivity analysis, cuts and geometry*

F. Critchley

(with D E Tyler, H Oja and L Dümbgen) Royal Statistical Society Read Paper, London *Invariant coordinate selection*

C.P. Farrington

Presentation to a workshop in Berlin on the case series method and its application: *Case series models for 'interferent' events*

Presentation to a workshop in London on vaccines and childhood survival: *Methodological issues in the design and analysis of cohort studies of non-specific effects of vaccines*

University of Kent *Case series models and extensions*

University of Nottingham *Estimation of contact surfaces for serological survey data*

Drug Safety Research Unit, Southampton *The self-controlled case series method: applications in pharmacoepidemiology*

P.H. Garthwaite

University of New South Wales, Sydney, Australia *Selection of prior weights for weighted model averaging*

M.C. Jones

Université de Neuchâtel, Switzerland *The t family and their close and distant relations*

University of Kent *The t family and their close and distant relations*

University of Liverpool *The t family and their close and distant relations*

10. Editorial roles

F. Critchley

- Associate Editor: *Journal de la Société Francaise de Statistique*
- Editorial Board Member: *Studies in Classification, Data Analysis, and Knowledge Organization* (book series, Springer)

C.P. Farrington

- Editorial Board Member: *Epidemiology and Infection*

J.C. Gower

- Associate Editor: *The Mathematical Scientist*

M.C. Jones

- Associate Editor: *Annals of the Institute of Statistical Mathematics*
- Associate Editor: *Communications in Statistics*
- Associate Editor: *Journal of Statistical Planning and Inference*
- Associate Editor: Journal of the Korean Statistical Society
- Associate Editor: *Pakistan Journal of Statistics*
- Associate Editor: *Statistica Sinica*
- Associate Editor: *Statistical Methodology*

11. Grants Awarded

C.P. Farrington & H.J. Whitaker

A grant of £10,250 from the Wellcome Trust to organise a conference or workshop on statistical methods for database research in epidemiology

12. Other activities

F. Critchley

- organised the "Emerging Geometries for Statistical Science" session at the Royal Statistical Society International Conference, Nottingham

- external referee for KUL (Catholic University of Leuven) Research Professor positions

- international member of Agregacao Committee at Instituto Superior Tecnico, Lisbon

- examined a PhD at University College London

- external thesis adviser, University of Malaya

C.P. Farrington

- member of the RSS Council

- member of the WHO Ad-Hoc Committee on EPI Serology in Relation to Intermittent Preventive Treatment in Infants against Malaria

- member of the Wellcome Trust's Study Design Expert Group

- external examiner for the MSc in Statistical Science at the University of Lancaster

- examined a PhD at the Institute of Child Health, University College London.

P.H. Garthwaite

- visiting academic in the Department of Mathematics and Statistics, University of New South Wales, Sydney, Australia (until April)

- external examiner for undergraduate statistics courses at the University of Mauritius

- member of the MRC Panel of Experts

- examined PhDs for Queensland University of Technology, Brisbane, Australia and for the University of Warwick

J.C. Gower

- member of the Sir Ronald Aylmer Fisher Memorial Committee of Great Britain

- foreign member of the Advisory Board of IOPS (Interuniversitaire Onderzoekschool voor Psychometrie en Sociometrie)

- examined PhD at the University of Leiden, Netherlands

M.C. Jones

- examined PhD for the National College of Business Administration and Economics, Lahore, Pakistan

K.J. McConway

- is Associate Dean (Curriculum and Awards) of the Faculty of Mathematics, Computing and Technology

- continues his media work at an ever increasing pace:
  o academic consultant on two more series of the BBC Radio4/OU series 'More or Less', including a contribution on air to one programme talking about the causation of stomach ulcers UNICEF reports on children and young people;
  o in January, did a piece on the Radio 4 Analysis programme "Jackanory Politics" about narrative in politics, speaking about Kahneman and Tversky's work on heuristics and biases;
  o (with S. Usborne) wrote "Numbers up: The truth about statistics" for The Independent newspaper, 9 April 2008, Extra section, pp. 8-9. Also syndicated in the Belfast Telegraph, 9 April 2008, and the New Zealand Herald (under the title 'Lies, damn lies and 150 coconut deaths'), 12 April 2008.

N.T. Trendafilov

- member of the program committee for a Workshop of the ERCIM Working Group on Computing and Statistics, Neuchâtel, Switzerland