

The Open University Statistics Department 2005 Report

1. Membership

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

Long-term visitor

Dr P.C. Taylor (University of Hertfordshire)

Short-term visitors

Mr A. Al-Saber (Kuwait University)
Mr K. Anaya (IIMAS-UNAM, Mexico)
Professor P.K. Marriott (University of Waterloo, Canada)

Full-time research students

Mr C.J. Banks (joint with Earth Sciences)
Mr S. De (from October)
Mr D.J. Jenkinson
Mr E. Mubwandarikwa
Mr P. Musonda
Mr A. Sarkar (joint with Computing)

Part-time research students

Ms H. Beaumont
Mr A. Garrett (to March)
Ms A. Gjini
Mr G. Kafatos
Mr A. Owen
Mrs L. Sun

Support staff

Mrs V. Spearman, Departmental Secretary

2. Introduction

2005 was, inevitably, another active year for the Department in both research and teaching.

In September, we welcomed to the Department, Nickolay Trendafilov, Lecturer in Statistics, and Casper Albers and Karim Anaya, Postdoctoral Researchers (Karim to work with Frank Critchley, Casper to work on several projects around the Department).

We were all greatly saddened to hear of the untimely death of our friend and colleague Professor Cliff Lunneborg (University of Washington), a lovely man who was a regular visitor to the Department throughout the 1990s.

Two 'DREaM' events, reviewing developments in statistical methodology both within and between the areas of Diagnostics, Robustness, Exploration and Modelling, were organised as follows: a one day meeting on March 31 and a longer residential workshop, March 30 - April 2. The conference was sponsored by the European Science Foundation. Frank Critchley and Val Spearman worked extremely hard in leading the organisation of these successful events.

In the course of the year, members of the Department published some 26 papers in refereed journals (including three in the *Journal of the American Statistical Association*) plus various other publications (see later). Pre-publication papers and other research material can be found in our technical report series, at

<http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm>

Our teaching continued to flourish. We ran five 30 point courses, three at second level and two at third, as well as contributing to other courses in Faculty and University. The Department's five main courses attracted a total of about 2,000 students.

Alvaro Faria and Paul Garthwaite run, as they have for a few years, a Statistical Advisory Service to support research in other departments at the Open University.

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 (30 CATS points each): [BM240](#) *Quantitative Methods in Business*, [MDST242](#) *Statistics in Society*, [M248](#) *Analysing Data*, [M343](#) *Applications of Probability* and [M346](#) *Linear Statistical Modelling*. BM240 changed from a February start to an October start this year.

Each of the Faculty of Mathematics and Computing's first level mathematics courses, [MU120](#) *Open Mathematics* and [MST121](#) *Using Mathematics*, contains substantial amounts of statistical material (roughly one quarter of each) with which Paul Garthwaite and Catriona Queen were involved.

Kevin McConway contributed to the presentation of U205 *Health and Disease*.

Production

Most of the necessary production work took place this year on a new 10 CATS point Level 1 course for science, mathematics and health students, [SMK184](#) *Chance, Risk and Health*. The course is written around the book 'Dicing with Death' by Stephen Senn. The course is to be

presented as part of the OU's Science Short Course Programme four times a year, starting in February 2006. Chris Jones and Kevin McConway comprised the Course Team.

Work on our new second-level course in Statistics, M249, entitled *Practical Modern Statistics*, continued apace. The course will be presented for the first time in 2007. It will consist of four blocks on the topics of Bayesian statistics, medical statistics, multivariate statistics and time series analysis. On completion, M249 will round off the undergraduate Diploma in Statistics (by replacing the elective element; see below) and will also form a core component of other degrees involving Statistics.

Kevin McConway completed his revision of two units on statistics for MU120 *Open Mathematics* and contributed to SDK125 *Introducing Health Sciences*.

Awards

Our courses are core components of degrees in Mathematical Sciences and Economics and Mathematical Sciences while BM240 is part of the BA in Business Studies. Degrees in Mathematics and Statistics and Computing and Statistics were introduced during 2004 and will have their first graduates in 2007.

The undergraduate [Diploma in Statistics](#) is currently awarded to students who successfully study M248 or BM240, M343 and M346 (or their predecessors) plus one other course from a list of electives. The latest available figures are that 120 students were awarded the Diploma in 2004/05.

4. PhDs awarded

PhDs were awarded to:

Andy Garrett (subject to satisfactory revision) for his thesis *The Role of Subgroups and Sub-Populations in Drug Development and Drug Regulation*. (Supervisors: D.J. Hand, Imperial College, S. Senn, University of Glasgow and C.M. Queen).

Peter Johnston-Wilder for his thesis *Learners' Shifting Perception of Randomness* (joint with Mathematics Education; supervisors included K.J. McConway).

Bheki Twala for his thesis *Effective Techniques for Handling Incomplete Data Using Decision Trees* (Supervisor: M.C. Jones)

Ben Wright (subject to satisfactory revision) for his thesis *A Bayesian Dynamic Approach to Modelling Flow through a Traffic Network* (Supervisor: C. Queen)

5. Research interests

The Department is home to three Research Groups:

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

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

Medical Statistics (Whitaker (chair), Farrington, Garthwaite, Jenkinson, Musonda, Vines).

Clearly, these three general groupings cover a huge variety of research projects in Statistics. 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 sample of topics. Strong ongoing collaborations continue in areas such as medicine, psychology, ecology and archaeology.

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

6. Publications

6.1 Publications in refereed academic journals

Albers, C.J., de Roos, G.T. & Schaafsma, W. (2005) Estimating a frequency unseen: an application to ornithology. *Statistica Neerlandica*, 59, 397-413.

Albers, C.J., Kooi, B.P. & Schaafsma, W. (2005) Trying to resolve the two envelope problem. *Synthese*, 145, 89-109.

Alberts, R., Fu, J., Swertz, M.A., Lubbers, L.A., **Albers, C.J.** & Jansen, R.C. (2005). Combining microarrays and genetic analysis. *Briefings in Bioinformatics*, 6, 135-146.

Blasius, J. & **Gower, J.C.** (2005) Multivariate prediction with nonlinear principal components analysis (application). *Quality and Quantity*, 39, 373-390.

Crawford, J.R. & **Garthwaite, P.H.** (2005) Testing for suspected impairments and dissociations in single-case studies in neuropsychology: evaluation of alternatives using Monte Carlo simulations and revised tests for dissociations. *Neuropsychology*, 19, 318-331.

Crawford, J.R. & **Garthwaite, P.H.** (2005) Evaluation of criteria for classical dissociations in single-case studies by Monte Carlo simulation. *Neuropsychology*, 19, 664-678.

Farrington C.P. & **Whitaker H.J.** (2005) Contact surface models for infectious diseases: estimation from serologic survey data. *Journal of the American Statistical Association*, 100, 370-379.

Garthwaite, P.H., Kadane, J.B. & O'Hagan, A. (2005) Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100, 680-701.

Gower, J.C. & Blasius, J. (2005) Multivariate prediction with nonlinear principal components analysis (methodology). *Quality and Quantity*, 39, 359-372.

Gower, J.C. & Ngouenet, R.F. (2005) Nonlinearity effects in multidimensional scaling. *Journal of Multivariate Analysis*, 94, 344-365.

Hubbard, R., Lewis, S., Smith, C., Godfrey, C., Smeeth, L., **Farrington, C.P.** & Britton, J. (2005) Use of nicotine replacement therapy and the risk of acute myocardial infarction, stroke and death. *Tobacco Control*, 14, 416-421.

Hubbard, R., Lewis, S., West, J., Smith, C., Godfrey, C., Smeeth, L., **Farrington, C.P.** & Britton, J. (2005) Bupropion and the risk of sudden death: a self-controlled case-series analysis using the Health Improvement Network. *Thorax*, 60, 848-850.

Jennrich, R.I. & **Trendafilov, N.T.** (2005) Independent component analysis as a rotation method: A very different solution to Thurstone's box problem. *British Journal of Mathematical and Statistical Psychology*, 58, 199-208.

Jones, M.C. (2005) The uniform mixture of generalized arc-sine distributions. *Journal of the Korean Statistical Society*, 34, 35-38.

Jones, M.C. & Pewsey, A. (2005) A family of symmetric distributions on the circle. *Journal of the American Statistical Association*, 100, 1422-1428.

Kanaan, M. & **Farrington, C.P.** (2005) Matrix models for childhood infections: a Bayesian approach with applications to rubella and mumps. *Epidemiology and Infection*, 133, 1009-1021.

Pewsey, A. & **Jones, M.C.** (2005) Discrimination between the von Mises and wrapped normal distributions: just how big does the sample size have to be? *Statistics*, 39, 81-89.

Prescott, G.J. & **Garthwaite, P.H.** (2005) Bayesian analysis of misclassified binary data from a matched case-control study with a validation sub-study. *Statistics in Medicine*, 24, 379-401.

Prescott, G.J. & **Garthwaite, P.H.** (2005) A Bayesian approach to prospective binary outcome studies with misclassification in a binary risk factor. *Statistics in Medicine*, 24, 3463-3477.

Tata L.J., Fortun P.J., Hubbard R.B., Smeeth L., Hawkey C.J., Smith C.J.P., **Whitaker H.J., Farrington C.P.**, Card T.R. & West J. (2005) Does concurrent prescription of antidepressant and non-steroidal anti-inflammatory drugs substantially increase the risk of upper gastrointestinal bleeding? *Alimentary Pharmacology & Therapeutics*, 22, 175-181.

Tata, L.J., West, J., Smith, C., **Farrington, C.P.**, Card, T., Smeeth, L. & Hubbard, R. (2005) General population based study of the impact of tricyclic and selective serotonin reuptake inhibitor antidepressants on the risk of acute myocardial infarction. *Heart*, 91, 465-471.

Toledano M.B., Nieuwenhuijsen M.J., Best N., **Whitaker H.**, Hambley P., DeHoogh C., Fawell J., Jarup L. & Elliott P. (2005) Relation of trihalomethane concentrations in public water supplies to stillbirth and birth weight in three water regions in England. *Environmental Health Perspectives*, 2005, 113, 225-232.

Trendafilov, N.T. (2005) Fitting the factor analysis model in L1 norm. *British Journal of Mathematical and Statistical Psychology*, 58, 19-32.

van Hijum, S.A.F.T., de Jong, A., Baerends, R.J.S., Karsens, H.A., Kramer, N.E., Larsen, R., den Hengst, C.D., **Albers, C.J.**, Kok, J. & Kuipers, O.P. (2005) A generally applicable validation scheme for the assessment of factors involved in reproducibility and quality of DNA-microarray data. *BMC Genomics*, 6, Article 77.

Warwick, J. & **Jones, M.C.** (2005) Choosing a robustness smoothing parameter. *Journal of Statistical Computation and Simulation*, 75, 581-588.

Whitaker H., Best N., Nieuwenhuijsen M.J., Wakefield J., Fawell J. & Elliott P. (2005) Modelling exposure to disinfection by-products in drinking water for an epidemiological study of adverse birth outcomes. *Journal of Exposure Analysis and Environmental Epidemiology*, 15, 138-146.

6.2 Refereed, invited or commissioned conference contributions, book chapters and research reports

Bartkowiak, A. & **Trendafilov, N.T.** (2005) Feature extraction by the SCoTLASS: An Illustrative example. In *Intelligent Information Processing and Web Mining*, eds: M.A. Klopotek, S.T. Wierzchen & K. Trojanowski, Springer, Berlin, pp. 3-11.

Capilluppi, A., **Faria, A.E.** & Ramil, J. (2005) Exploring the relationship between cumulative change and complexity in an open source system evolution. In *Proceedings of the 9th European Conference on Software Maintenance and Re-engineering (CSMR)*, IEEE Computer Society Press, pp. 21-29.

Davies, C., Anand, P., Artigas, L., Holloway, J., **McConway, K.J.**, Newman, J., Storey, J. & Thompson, G. (2005) Links between governance, incentives and outcomes: a review of the literature. Report for the National Co-ordinating Centre for NHS Service Delivery and Organisation R&D (NCCSDO).

DeRoeck, A., Sarkar, A. & **Garthwaite, P.H.** (2005) Even very frequent function words do not distribute homogeneously. In *The International Conference on Recent Advances in Natural Language Processing (RANLP)*, eds: G. Angelova, K. Bontcheva, R. Mitkov & N Nicolav, INCOMA, Shoumen, Bulgaria, pp. 429-436.

Garthwaite, P.H. (2005) Confidence intervals: Nonparametric. In *Encyclopedia of Statistics in Behavioral Science, Volume 1*, eds: B.S. Everitt & D.C. Howell, Wiley, New York, pp. 375-381.

Gower, J.C. (2005) Biplots. In *Encyclopedia of Statistics in Behavioral Sciences, Volume 1*, eds: B.S. Everitt & D.C. Howell, Wiley, New York, pp. 163-164.

Sarkar, A., **Garthwaite, P.H.** & A DeRoeck, A. (2005) A Bayesian mixture model for term re-occurrence and burstiness. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL)*, eds: I. Dugan & D. Gildea, Association for Computational Linguistics, Ann Arbor, pp. 48-55.

6.3 Letter to Journal Editors

Farrington, C.P. & Miller, E. (2005) Informed choice, balance and the MMR saga. *Lancet Infectious Diseases*, 5, 2-3.

6.4 Research Website

Heather Whitaker is responsible for an extremely impressive and authoritative website about the self-controlled case series methodology: <http://statistics.open.ac.uk/scs/>

7. Seminars and conferences at the Open University

As already mentioned, we ran a one-day conference and three-day residential workshop under the "DREaM" banner, the brainchild of Frank Critchley, chief organiser. Speakers at the one-day meeting were:

Victor Yohai (Buenos Aires and CONICET, Argentina) *An overview of robust statistics*

John Copas (University of Warwick) *Models are usually wrong*

Ursula Gather (Dortmund, Germany) *Outliers*

Marco Riani (Parma, Italy) *Robust diagnostics*

Paul Marriott (Waterloo, Canada) *Local mixture models*

John Gower (Open University) *Exploratory multivariate data analysis and modelling*

The three-day meeting heard from Frank Critchley (Open), Hannu Oja (Tampere), John Kent (Leeds), Guy Brys (Antwerp), Laurie Davies (Duisberg-Essen), Ana Pires (Lisbon), Gentiane Haesbroeck (Liege), G. Willems (Antwerp), Anthony Atkinson (LSE), A. Ruiz-Gazen (Toulouse), K. Joossens (Leuven), Jim Griffin (Warwick) and Chris Jones (Open).

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

Caitlin Buck (University of Sheffield) *Statistical issues in constructing the new internationally agreed radiocarbon calibration curve*

Ahmed Al-Saber (Kuwait University) *Quantifying expert knowledge for modelling air pollution in Kuwait*

Byron Morgan (University of Kent) *100 years of marking*

Daniela DeAngelis (MRC Biostatistics Unit, Cambridge) *Use of surveillance data to estimate the evolution of the HIV epidemic*

Merrilee Hurn (University of Bath) *Loss functions for some awkward estimation problems*

Malcolm Farrow (Newcastle University) *Protein identification using mass spectrometry data and the case of the glass slipper*

Mike Kenward (London School of Hygiene and Tropical Medicine) *Multiple imputation for hierarchical and longitudinal data*

Niel LeRoux (University of Stellenbosch, South Africa) *Some applications of biplot methodology in practice*

John Copas (University of Warwick) *Making the best of the worst case*

We also ran our second Annual Research Students' Research Day, where speakers were Helen Beaumont, Avik Sarkar, Chris Banks, Alun Owen, Patrick Musonda and David Jenkinson.

And at an internal research morning, speakers were Paddy Farrington, Paul Garthwaite and Paul Taylor.

8. Conference papers and posters presented

K. Anaya

Second International Symposium on Information Geometry and Its Applications (IGAI2), Tokyo, Japan *Local mixtures of natural exponential families with quadratic variance function (Poster)*

F. Critchley

ICORS 2005, International Conference On Robust Statistics, Jyvaskyla, Finland *Principal axis analysis*

Second International Symposium on Information Geometry and Its Applications (IGAIA2), Tokyo, Japan *Statistical scientific sensitivity analysis: a geometric progression?*

J.C. Gower

70th Annual Meeting of the Psychometric Society, Tilburg, The Netherlands *Divided by a common language*

DREaM Conference, Open University *Exploratory multivariate data analysis and modelling*

International Biometric Society Multi-Region Conference, Leicester *The visualisation of multiplicative interaction*

International Association for Statistical Computing, 3rd World Conference on Computational Statistics and Data Analysis, Limassol, Cyprus *A matrix minimisation problem*

M.C. Jones

DREaM Conference, Open University *Robustness and modelling via four-parameter distributions*

International Conference on Statistics in Honour of Professor Kai-Tai Fang's 65th Birthday, Hong Kong Baptist University, Hong Kong *The sinh-arcsinh transformation*

H.J. Whitaker

International Biometric Society Multi-Region Conference, Leicester *Infections with varying contact rates: application to varicella*

26th Annual Conference of the International Society for Clinical Biostatistics, Szeged, Hungary. *A semi-parametric self-controlled case series model*

9. Other seminars given

C.P. Farrington

London School of Hygiene and Tropical Medicine *Statistical analysis of cases series*

J.C. Gower

University of Groningen, The Netherlands *Generalised Procrustes analysis with optimal scores*

University of Dortmund, Germany *Exploratory multivariate data analysis and modelling*

VoC, Vereniging voor Ordinatie en Classificatie (Dutch Classification Society), Zeist, The Netherlands *Procrustes problems - an overview*

M.C. Jones

University of Sheffield *Some new families of distributions*

10. Editorships

F. Critchley

- Editorial Board Member: *Studies in Classification, Data Analysis, and Knowledge Organization* (book series, Springer-Verlag).

C.P. Farrington

- Associate Editor: *Journal of the Royal Statistical Society, Series B*
- Editorial Board Member: *Statistics in Medicine* (to September)
- Editorial Board Member: *Lifetime Data Analysis* (to March)
- Editorial Board Member: *Epidemiology and Infection*

J.C. Gower

- Associate Editor: *Journal of Classification*
- Associate Editor: *The Mathematical Scientist*

M.C. Jones

- Associate Editor: *Communications in Statistics*
- Associate Editor: *Journal of Statistical Planning and Inference*
- Associate Editor: *Sankhya*
- Associate Editor: *Statistica Sinica*
- Associate Editor: *Statistical Methodology*

11. Grants

F. Critchley

The European Science Foundation SACD Network provided 13,000 euros to support the DREaM conference at the Open University

European Science Foundation SACD Network (£500) to deliver invited paper at ICORS 2005, International Conference On Robust Statistics, Jyväskylä, Finland.

University of Tokyo and RIKEN Brain Science Institute (£1,000) to deliver invited paper at IGAI2, the Second International Symposium on Information Geometry and Its Applications, Tokyo, Japan.

K.J. McConway

Joint principal investigator (with Jonathan Silvertown and David Gowing of the OU's Biological Sciences Department) for a project on 'Soil moisture gradients and the biodiversity of the Cape flora', to run for two years from September 2005, funded by a grant of £84,192 from the Leverhulme Trust.

12. Other activities

F. Critchley

- research visitor, Instituto Superior Tecnico, Lisbon, Portugal
- examined a PhD at University of Leeds
- is an assessor, Statistician Fast Stream, Civil Service Selection Board

C.P. Farrington

- is a member of the RSS Research Section Committee
- is a member of the WHO Global Advisory Committee on Vaccine Safety
- is a member of the WHO Ad-Hoc Committee on EPI Serology in Relation to Intermittent Preventive Treatment in Infants against Malaria
- is external examiner for the MSc in Modern Epidemiology at Imperial College London
- examined PhDs at the University of Cambridge and at Université de Paris XI, France

P.H. Garthwaite

- is external examiner for undergraduate mathematics courses at the University of Southampton
- is external examiner for the MSc in Statistics at the University of Kent
- is a member of the EPSRC College for peer review of grant applications

J.C. Gower

- was Honorary Secretary of the Sir Ronald Aylmer Fisher Memorial Committee of Great Britain
- is a member of the Council of the International Federation of Classification Societies
- is a Foreign Member of the Advisory Board of IOPS (Interuniversitaire Onderzoekschool voor Psychometrie en Sociometrie)

M.C. Jones

- is listed in ISIHighlyCited.com
- gave the Christmas Lecture at OU Region 4's Staff Development Day in Nottingham

K.J. McConway

- was Director of the OU's Undergraduate Mathematics Programme
- is academic consultant on the BBC Radio4/OU series More or Less, including a contribution on air to one programme

C.M. Queen

- is a member of the EPSRC College for peer review of grant applications

N.T. Trendafilov

- did statistical consulting and data analysis for Shared Solutions Consulting, London, on their Radiography Education Attrition Project.