The Open University Statistics Group

2018 Report

1. Membership

Miss A. Bromley
Miss C. Calvert
Dr F. Elfadaly
Dr Á.E. Faria
Professor P.H. Garthwaite
Mrs A. Gibb
Dr R.M. Hilliam (to July)
Professor M.C. Jones
Dr C.M. Queen
Dr N.T. Trendafilov
Dr S.K. Vines
Dr H.J. Whitaker (to September)

Long-term associates

Professor C.P. Farrington [Emeritus Professor]
Professor J.C. Gower [Emeritus Professor]
Professor K.J. McConway [Emeritus Professor]
Dr A. Noufaily (Warwick Medical School) [Visiting Research Fellow]
Dr Y.G. Weldeselassie (Warwick Medical School) [Visiting Research Fellow]
Dr H.J. Whitaker (Public Health England) [Visiting Research Fellow]

Full–time research students

Mrs M. Moustafa
Mr M.Z.R. Shabuz (to October)

Part–time research students

Mr T.G. Gebru
Mr R. Harris

2. Introduction

Here is an overview of the substantial activity in Statistics at the Open University in 2018, in both research and teaching.

On the research side, all our usual activities and outputs are described in Section 4 onwards of this report. In the course of the year, members of the Group published a book on the self-controlled case series method for medical statistics and 16 papers in refereed journals, along with a number of other outputs (see Section 6). Pre-publication papers can be found in our technical report series, at

http://www.mathematics.open.ac.uk/research/statistics/technical-reports

On the teaching side, over 2800 students were enrolled on the Statistics modules presented for study from dates in 2018 (see Section 3).

No retirements this year, but two senior and popular members of the group moved on to pastures new. One of these was Heather Whitaker, who first appeared in these reports as a postdoc in 2001, becoming lecturer in 2006 and senior lecturer in 2015. Inter alia, Heather is one of the co-authors of the book mentioned above, and her collaboration with our erstwhile colleague Paddy Farrington was a particularly long and fruitful one. Heather has moved to a new position with Public Health England and we wish her very well indeed in this new endeavour.

Also moving on was Rachel Hilliam, staff tutor since 2011, also at a senior level since 2015, and leading light in many managerial aspects of the Mathematics & Statistics school as a whole. Rachel left for a post at the University of Warwick, but it seems that she is shortly to return …

Congratulations to Carol Calvert on her promotion to Senior Lecturer.

The Statistical Advisory and Consultancy Services continue, supporting research in other disciplines across the Open University. These are run by Álvaro Faria and Paul Garthwaite. Two web pages explain the services:

http://www.mathematics.open.ac.uk/research/research/statistics/advisory-service

http://www.mathematics.open.ac.uk/research/statistics/statistical-consultancy-service

Further information on the Group's research activities is available through our general web pages at http://www.mathematics.open.ac.uk/research/statistics

3. Teaching

Presentation

The Group continued to present six modules of 30 CATS points each with one of them (M140) being presented twice in the year. These, with numbers of students starting the presentations of each module in 2018 given in square brackets, are:

- M140 *Introducing Statistics* (Chairs: Calvert, Bromley) [February 600; October 967]

- M248 *Analysing Data* (Chair: Queen) [434]

- M249 *Practical Modern Statistics* (Chairs: Whitaker, Faria) [182]

- M343 *Applications of Probability* (Chair: Trendafilov) [274]

- M346 *Linear Statistical Modelling* (Chairs: Vines, Garthwaite) [178]

- M347 *Mathematical Statistics* (Chairs: Garthwaite, Elfadaly) [174].

The first level mathematics module, MU123 *Discovering Mathematics*, also contains a little statistical material.

Production

Work on M348 *Applied Statistical Modelling* has begun. It will replace M346 *Linear Statistical Modelling* and is scheduled for first presentation in October 2022. The new module represents a very considerable update compared with the old, prompted partly by its prominent roles in our new undergraduate degree in Data Science (see below) and a new degree in Economics. Karen Vines is production Module Team chair.

Qualifications

Our modules are compulsory components of degrees in Mathematics and Statistics, Mathematics, Computing & IT and Statistics, Mathematics and its Learning and Economics and Mathematical Sciences, as well being specified options in certain other qualifications. The Mathematics and Statistics degree was awarded to 58 students in 2018. A further 33 students were awarded other joint degrees that include a substantial element of statistics.

We also award a Professional Certificate in Practical Statistics, for successful study of M140, M248, M249 and M346, and a Graduate Certificate in Theoretical Statistics and Probability, for M343 and M347. These are new qualifications, as yet awarded to 6 students in total.

We are also in the process of introducing an undergraduate degree in Data Science which will feature all our statistics modules, some compulsory, some optional, as well as others from computing and mathematics. M140, M248 and M348 will also be compulsory modules in a new Economics degree.

4. Research interests

Research in the Statistics Group has been centred for many years around three main themes: *Bayesian Statistics, Medical Statistics* and *Multivariate Statistics*. Clearly, these three general areas cover a huge variety of research projects in Statistics. Research also continues on topics not directly covered by these research themes, of which single case studies, distribution theory, time series forecasting, improved confidence intervals and statistics in the media are but a selection. Ongoing collaborations continue in areas such as medicine, psychology, and earth sciences.

For details, see http://www.mathematics.open.ac.uk/research/statistics and the publications, talks and other items to follow.

5. PhDs awarded

Mohammed Zillur Shabuz for his thesis
Application of transformations for orthogonality
(supervisors: P.H. Garthwaite and N.T. Trendafilov)

Tsegay Gebru for his thesis
Sparse discriminant analysis with more variables than observations
(supervisors: P.H. Garthwaite, N.T. Trendafilov and F. Critchley)

6. Publications

6.1 Book

***Farrington C.P.**, **Whitaker H.** & **Ghebremichael Weldeselassie**, **Y.** **(2018)**
Self-Controlled Case Series Studies: A Modelling Guide with R.* Chapman & Hall/CRC Press.

6.2 Publications in refereed journals

Adachi K. & ***Trendafilov, N.T.*** (2018) Some mathematical properties of the matrix decomposition solution in factor analysis. *Psychometrika*, **83**, 407-424.

Adachi K. & ***Trendafilov, N.T.*** (2018) Sparsest factor analysis for clustering variables: a matrix decomposition approach. *Advances in Data Analysis and Classification*, **12**, 559-585.

Connolly-Andersen, A.M., ***Whitaker, H.J.,*** Klingstrom, J. & Ahlm, C. (2018) Venous thromboembolism associated with hemorrhagic fever with renal syndrome A self-controlled case series study. *Clinical Infectious Diseases*, **66**, 268-273.

Gower, J.C. (2018) Skew symmetry in retrospect. *Advances in Data Analysis and Classification*, **12**, 33-41.

Gower, J.C., Gardner-Lubbe, S. & Le Roux, N.J. (2018) Interaction: Fisher's optimal scores revisited. *Journal of Agricultural, Biological, and Environmental Statistics*, **23**, 92-112.

Hardelid, P., ***Ghebremichael-Weldeselassie, Y., Whitaker, H.,*** Rait, G., Gilbert, R. & Petersen, I. (2018) Effectiveness of live attenuated influenza vaccine in preventing amoxicillin prescribing in preschool children: a self-controlled case series study. *Journal of Antimicrobial Chemotherapy*, **73**, 779–786.

Herman, C. & ***Hilliam, R.*** (2018) The triple whammy: gendered careers of geographically marginalised academic STEM women. *International Journal of Gender, Science and Technology*, **10**, 171-189.

Hilliam, R. & ***Calvert, C.*** (2017) Interactive statistics for a diverse student population. *Open Learning: The Journal of Open, Distance and e-Learning*, DOI: 10.1080/02680513.2017.1386095

Jones, M.C. (2018) Families of complementary distributions. *Statistics and Probability Letters*, **141**, 74-81.

Lakens, D., Adolfi, F.G., Albers, C.J. and 85 others including ***McConway, K.*** (2018). Justify your alpha. *Nature Human Behaviour*, **2**, 168-171.

Liu, E.Y., Smith, L.M., Ellis, A.K., **Whitaker, H.,** Law, B., Kwong, J.C., **Farrington, P.** & Lévesque, L.E. (2018) Quadrivalent human papillomavirus vaccination in girls and the risk of autoimmune disorders: the Ontario Grade 8 HPV Vaccine Cohort Study. *Canadian Medical Association Journal, **190***, E648-E655*.*

Shabuz, Z.R. & **Garthwaite, P.H.** (2018) Bootstrap confidence intervals for the contributions of individual variables to a Mahalanobis distance. *Journal of Statistical Computation and Simulation*, **88**, 2232-2258.

Warren-Gash, C., Blackburn, R., **Whitaker, H.**, McMenamin, J. & Hayward, A.C. (2018) Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. *European Respiratory Journal*, **51**, art. 1701794.

Webb, J., **Hilliam, R.** & Bainbridge, C. (2018) A novel device for preoperative skin preparation to reduce the risk of injury and surgical site infection during lower limb surgery. *Journal of Perioperative Practice*, 28, 109-114.

Whitaker H.J., Ghebremichael-Weldeselassie, Y., Douglas I.J., Smeeth L. & **Farrington C.P**. (2018) Investigating the assumptions of the self-controlled case series method. *Statistics in Medicine*, **37**, 643-658.

Whitaker H.J., Steer, C.D. & **Farrington, C.P**. (2018) Self-controlled case series studies: just how rare does a rare non-recurrent outcome need to be? *Biometrical Journal*, **60**, 1110-1120.

6.3 Other publications

Albers, C.J., **Gower, J.C.** & Kiers. H. (2018) Rank properties for centred three-way arrays. In *Classification, (Big) Data Analysis and Statistical Learning*, Eds: F. Mola, C. Conversano & M. Vichi, Springer, pp. 69-76.

Faria, Á.E. & Santos, A.J. (2018) Dynamic Bayesian smooth transition autoregressive models applied to hourly electricity load in southern Brazil. In: *ITISE 2018– International Conference on Time Series and Forecasting*, pp. 966–981.

Farrington P. & **Whitaker H.** (2018) The self-controlled case series method. Chapter 22 in: *Handbook of Statistical Methods for Case-Control Studies,* Eds: O. Borgan, N. Breslow, N. Chatterjee, M.H. Gail, A. Scott & C.J. Wild, Chapman & Hall/CRC Press.

Hickmann, T., Teixeira Júnior, L.A., **Faria, Á**., Rodrigues, S.B., Côrrea, J.M. & Garcia, E.L. (2018) Temperature field forecast in concrete dam with the use of Arima models and the finite element method. In: *Métodos Quantitativos – Pesquisa Operacional*, Vol. 3, Ed: D.F. Andrade, Editora Poisson, pp. 112–123.

Jones, M.C. (2018) Letter to the Editor concerning "A new method for generating distributions with an application to exponential distribution" and "Alpha power Weibull distribution: Properties and applications". *Communications in Statistics– Theory and Methods*, **47**, 5096.

McConway, K., Joffe, H., Scott, E.M., Spiegelhalter, D. & Walton, H. (2018) Measurement and communication of health risks from pollution. Chapter 9 in *Annual Report of the Chief Medical Officer 2017: Health Impacts of All Pollution – what do we know?* Eds: A.R.H. Dalton, A.R.H. & O. Murphy, Department of Health and Social Care.

7. Seminars at The Open University

Over the year we heard the following external seminar speakers:

Georgi Boshnakov (University of Manchester) *Count data regression based on rich new families of distributions*

Jared Tanner (University of Oxford) *An iterative linear algebra perspective on compressed sensing and matrix completion*

Irini Moustaki (London School of Economics) *Latent variable models and pairwise likelihood framework: old and new developments*

8. External conference papers and posters given

C. Calvert

Horizons in STEM Higher Education Conference: Making Connections and Sharing Pedagogy, Hull *Making student feedback count*

Á.E. Faria

ITISE 2018: International Conference on Times Series and Forecasting, Granada, Spain *Dynamic Bayesian STAR models applied to electricity load in Brazil*

F. Elfadaly

International Society for Bayesian Analysis (ISBA) World Meeting 2018, Edinburgh *An R Shiny tool for Bayesian imputation of missing variables across studies in mixed-effects models* [poster]

P.H. Garthwaite

International Society for Bayesian Analysis (ISBA) World Meeting 2018, Edinburgh. Session discussant for *Learning by iteration: informative priors and their role in updating knowledge*

R. Hilliam

Horizons in STEM Higher Education Conference: Making Connections and Sharing Pedagogy, Hull *Using a dedicated subject website to create a mathematics and statistics community of learners*

CETL-MSOR: Evidencing Excellence in the Mathematical Sciences, Glasgow *Creating an online mathematics and statistics community of learners*

M.C. Jones

26th Annual Meeting of the Royal Statistical Society of Belgium, Ovifat, Belgium *Flexible distributional modelling for parametric survival analysis*

K.J. McConway

Stempra annual training conference, London *Putting statistics in press releases*

Centre for Investigative Journalism Summer School, London *How to read scientific papers and press releases*

N.T. Trendafilov

11th International Conference of the ERCIM Working Group on Computational and Methodological Statistics, Pisa, Italy *Semi-sparse PCA*

9. Other external seminars and talks given

P.H. Garthwaite

RSS Lancashire and East Cumbria Local Group, Preston *Quantifying expert opinion about some sampling models that include covariates*

K.J. McConway

City University, London *How not to be taken in by the numbers*

University of Reading *Don't dumb it down: presenting statistics to the media and the public*

10. Editorial roles

M.C. Jones

- Associate Editor: *Annals of the Institute of Statistical Mathematics* (to July)
- Associate Editor: *Journal of the Korean Statistical Society*
- Associate Editor: *Statistics and Probability Letters*
- Member, Editorial Board: *Statistical Methods in Medical Research* (from April)

K.J. McConway

- Statistical Advisor: *Endocrine-Related Cancer*
- Statistical Advisor: *Journal of Endocrinology*
- Statistical Advisor: *Journal of Molecular Endocrinology*

N.T. Trendafilov

- Associate Editor: *Behaviormetrika*
- Associate Editor: *Computational Statistics*
- Associate Editor: *Japanese Journal of Statistics and Data Science*

H.J. Whitaker

- Associate Editor: *Endocrine Connections*

11. Other activities

C. Calvert

- was promoted to a Senior Lectureship

- won an Open University eSTEeM Scholarship Project of the Year award in the Enhancing the Student Experience category for the project "A flexible start to M140"

- gave a talk at the Open University eSTEeM conference

- lead author of eSTEeM final report on the project "A flexible start to M140"

- obtained internal funding for a flexible start programme for MU123

F. Elfadaly

- is Honorary Lecturer, National Heart and Lung Institute, Imperial College London

Á.E. Faria

- gave a talk on *Quantitative Methods in Research* for Open University Academic Professional Development

C.P. Farrington

- is expert adviser to the Tribunal de Grande Instance de Paris (France) in litigation over the teratogenicity of valproic acid

P.H. Garthwaite

- external examiner for MSc in Applied Statistics with specialisation in Actuarial Science at the University of Mauritius

R. Hilliam

- is an RSS accreditation assessor

- gave a talk at the Open University eSTEeM annual conference

- is a visiting Research Fellow in Medical Statistics at the University of Derby, running statistics and research design training for doctor of professional practice students

M.C. Jones

- examined a PhD at the University of Bath

K.J. McConway

- ran a training event in statistics for journalists at the Science Media Centre, London

- STEM Ambassador (working with schools and young people to promote STEM subjects), including taking part in a STEM careers session at a school in Milton Keynes

- judge for the 'STEM for Britain' parliamentary poster competition for career-young researchers

- member of the Advisory Committee and, later, Trustee, of the Science Media Centre, writing invited comments on statistical aspects of scientific papers, and contributing in other ways to the work of the Centre

- presented video talks on *Judging importance of statistical results* and *Communicating your research*, to form part of *Data Tree*, an online course for researchers and scientists funded by NERC

- committee member, RSS Data Ethics Special Interest Group

C.M. Queen

- external examiner for undergraduate and postgraduate statistics modules at the University of Warwick

N.T. Trendafilov

- member of the program committee of the 4th International Workshop on Machine Learning, Optimization and Big Data (LOD 2018), Volterra, Italy

- elected member of the Board of Directors, European Regional Section of the International Association for Statistical Computing (IASC)

S.K. Vines

- won an Open University eSTEeM Scholarship Project of the Year award in the Innovative/Original Approach to Teaching category for the project "Sonification of depictions of numerical data" (with Chris Hughes)

- co-chair of eSTEeM project "Sonification partial pilot on M140"

- gave a talk at the eSTEeM annual conference

- was a presenter at a seminar arranged by Open University Academic Professional Development in conjunction with the University of Gothenburg, Sweden

- external examiner for undergraduate statistics modules at Kingston University London

H.J. Whitaker

- examined a PhD at the London School of Hygiene and Tropical Medicine