

Frank Dondelinger

Current Work Address

CHICAS, Lancaster Medical School
Lancaster University
Lancaster, LA1 4YW
United Kingdom

Permanent Address

XXX
Lancaster LA1 XXX
United Kingdom

Research Interests

My research focuses on applying statistical and machine learning techniques to data-intensive or high-dimensional problems, particularly in biological and biomedical research. I am interested in problems involving prediction, graphical models and unsupervised learning, using both hierarchical Bayesian and structured penalised regression methods. Much of my research is concerned with situations where a dataset consists of known or unknown heterogeneous subsets, and the goal is to develop a prediction model that leverages the commonalities between the subsets via information sharing. Among other applications, these methods are of use in personalised medicine, for example prediction of drug response in cancer cell lines and patients, and in large-scale population studies, for example prediction of Type-II diabetes incidence in specific populations.

My research interests further include network inference, such as inferring gene regulatory networks from gene expression data. I have applied Bayesian modelling techniques to this problem, including stationary and dynamic Bayesian networks. I have also applied sparse linear regression techniques for network inference (e.g. lasso and graphical lasso). I have developed and tested new priors for information sharing in conjunction with changepoint processes, for situations where the structure of the network can change.

Finally, I have worked on methods for efficient parameter inference in ODE models of biological systems, and combining ODE models with network inference based on linear models for better pathway reconstruction and prediction of pathway dynamics.

Appointments

- October 2015 - Present: *Lancaster University, Lancaster*
Lecturer in Biostatistics in the Lancaster Medical School.
- March 2014 - October 2015: *MRC Biostatistics Unit, Cambridge*
Career Development Fellow in the Statistical Genomics programme.
- August 2012 - February 2014: *The Netherlands Cancer Institute, Amsterdam*
Postdoctoral Fellow in the Division of Biochemistry, Computational Modelling.
- January 2012 - June 2012: *The University of Glasgow, Glasgow*
Research Assistant in the School of Mathematics and Statistics.

Education

- *PhD in Informatics*, Focus on Machine Learning and Systems Biology.
University of Edinburgh, School of Informatics, Edinburgh 2008-2013
Joint appointment at the Institute of Adaptive and Neural Computation (IANC) and at Biomathematics and Statistics Scotland (BioSS).
- *MSc in Artificial Intelligence with Distinction*
University of Edinburgh, School of Informatics, Edinburgh 2007-2008
Major specialism Learning From Data, Minor specialism Bioinformatics.
- *BSc (Hons) in Artificial Intelligence and Computer Science, First Class.*
University of Edinburgh, School of Informatics, Edinburgh 2003-2007

Publications

Journals and Refereed Conferences

- N. Städler, F. Dondelinger, S.M. Hill, R. Akbani, Y. Lu, G.B. Mills, S. Mukherjee, “Molecular heterogeneity at the network level: high-dimensional testing, clustering and a TCGA case study”, [Submitted, Bioinformatics] (2016). *Joint first author.*
- C. J. Oates, F. Dondelinger, N. Bayani, J. Korkola, J.W. Gray, S. Mukherjee, “Causal network inference using biochemical kinetics”, *Bioinformatics* 30, no 17 (2014). *Won Best Paper Award at ECCB 2014.*
- R. Akbani, P.K.S. Ng, H.M.J. Werner, M. Shahmoradgoli, F. Zhang, Z. Ju, W. Liu, J. Yang, K. Yoshihara, J. Li, S. Ling, E.G. Seviour, P.T. Ram, J.D. Minna, L. Diao, P. Tong, J.V. Heymach, S.M. Hill, F. Dondelinger, N. Städler, L.A. Byers, F. Meric-Bernstam, J.N. Weinstein, B.M. Broom, R.G.W. Verhaak, H. Liang, S. Mukherjee, Y. Lu, G.B. Mills: “A pan-cancer proteomic perspective on The Cancer Genome Atlas”, *Nature Communications* 5 (2014).
- F. Dondelinger, M. Filippone, S. Rogers, D. Husmeier, “ODE parameter inference using adaptive gradient matching with Gaussian processes”, Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (2013). *Presented as a talk at AISTATS 2013.*
- F. Dondelinger, S. Lèbre, D. Husmeier, “Non-homogeneous dynamic Bayesian networks with Bayesian regularization for inferring gene regulatory networks with gradually time-varying structure”, *Machine Learning* 90, no. 2 (2012).
- F. Dondelinger, D. Husmeier and S. Lèbre, “Dynamic Bayesian networks in molecular plant science: inferring gene regulatory networks from multiple gene expression time series”, *Euphytica* 183, no. 3 (2011).
- D. Husmeier, F. Dondelinger and S. Lèbre, “Inter-time segment information sharing for non-homogeneous dynamic Bayesian networks”, Advances in Neural Information Processing (2010). *Presented as a poster at NIPS 2010.*

- F. Dondelinger, D. Husmeier and S. Lèbre, “Heterogeneous Continuous Dynamic Bayesian Networks with Flexible Structure and Inter-Time Segment Information Sharing”, Proceedings of the 27th International Conference on Machine Learning (2010). *Presented as a talk at ICML 2010.*
- A. Faisal, F. Dondelinger, C. Beale and D. Husmeier, “Inferring species interaction networks from species abundance data: A comparative evaluation of various statistical and machine learning methods”, *Ecological Informatics* 5, no. 6 (2010). *Joint first author.*

Conferences

- B. MacDonald, F. Dondelinger, D. Husmeier, “Inference in complex biological systems with Gaussian processes and parallel tempering”, Proceedings of the 28th International Workshop on Statistical Modelling (2013).
- F. Dondelinger, S. Rogers, M. Filippone, R. Cretella, T. Palmer, R. Smith, A. Millar, D. Husmeier, “Parameter inference in mechanistic models of cellular regulation and signalling pathways using gradient matching”, Proceedings of the 12th International Workshop on Computational Systems Biology (2012).
- F. Dondelinger, A. Aderhold, S. Lèbre, M. Grzegorzczuk and D. Husmeier, “A Bayesian regression and multiple changepoint model for systems biology”, Proceedings of the 26th International Workshop on Statistical Modelling (2011).
- F. Dondelinger, D. Husmeier, S. Lèbre, “Reconstructing Developmental Gene Networks using Heterogeneous Dynamic Bayesian Networks with Information Sharing”, Proceedings of the Fourth International Workshop on Machine Learning in Systems Biology (2010).

Book Chapters

- S. Lèbre, F. Dondelinger, D. Husmeier, “Nonhomogeneous dynamic Bayesian networks in systems biology”, in J. Wang et al. (eds.): Next Generation Microarray Bioinformatics (2012), Springer, Series Methods in Molecular Biology, Volume 802, page 199.
- K. Lin, D. Husmeier, F. Dondelinger, C.D. Mayer, H. Liu, L. Prichard, G.P.C. Salmond, I.K. Toth and P.R.J. Birch, “Reverse Engineering Gene Regulatory Networks Related to Quorum Sensing in the Plant Pathogen *Pectobacterium atrosepticum*”, in D. Fenyö (ed.): Computational Biology (2010), Springer, Series Methods in Molecular Biology, Volume 673, pages 253-281.

Theses

- F. Dondelinger, “A Machine Learning Approach to Reconstructing Signalling Pathways and Interaction Networks in Biology”, PhD Thesis, University of Edinburgh (2013).
- F. Dondelinger, “Inferring Ecological Networks From Species Abundance Data”, MSc Thesis, University of Edinburgh (2008).

Presentations

Invited

- *High-dimensional statistical approaches for heterogeneous molecular data in cancer medicine*, Royal Statistical Society Annual Conference, Exeter, UK, Autumn 2015.
- *Network inference and dynamic prediction using biochemical kinetics*, International Biometrics Conference, Florence, Italy, Summer 2014.
- *Inferring time-varying gene regulatory networks from gene expression data using information sharing*, Statistical Methods for (post)-Genomics Data, VU University, Amsterdam, Netherlands, Winter 2013.

Other

- *High-dimensional statistical analysis of molecular data.*, Mathematical and Statistical Aspects of Molecular Biology (MASAMB), Helsinki, Finland, Spring 2015.
- *ODE parameter inference using adaptive gradient matching with Gaussian processes*, International Conference on Artificial Intelligence and Statistics (AISTATS), Scottsdale, Arizona, US, Spring 2013.
- *A Bayesian regression and multiple changepoint model for systems biology*, International Workshop on Statistical Modelling (IWSM), Valencia, Spain, Summer 2011.
- *Reconstructing Developmental Gene Networks using Heterogeneous Dynamic Bayesian Networks with Information Sharing*, International Workshop on Machine Learning in Systems Biology, University of Edinburgh, Edinburgh, UK, Winter 2010.
- *Heterogeneous Continuous Dynamic Bayesian Networks with Flexible Structure and Inter-Time Segment Information Sharing*, International Conference on Machine Learning, Haifa, Israel, Summer 2010.
- *Heterogeneous Continuous Dynamic Bayesian Networks with Flexible Structure and Inter-Time Segment Information Sharing*, Conference on Mathematical and Statistical Aspects of Molecular Biology, University of Warwick, Coventry, UK, Spring 2010.
- *Inferring Ecological Networks From Species Abundance Data*, Workshop on Learning in Computational Systems Biology (LICSB), Imperial College, London, UK, Spring 2009.

Teaching and Supervising

Lent Term 2015 : *Supervisions, Statistics IB.*

University of Cambridge, Faculty of Mathematics, Cambridge, UK.

September 2014 - February 2015 : *Supervised MSc project of Christian Groß: Feature Selection for Drug Response Prediction.*

MRC Biostatistics Unit, Cambridge, UK.

June-August 2013 : *Supervised MSc project of James Newling: Drug Response Prediction using l1 penalised regression.*

The Netherlands Cancer Institute, Amsterdam, The Netherlands.

December 2010, December 2011: *Invited Lecture for MSc in Genomics course: Data Analysis and Statistical Modelling.*

University of Edinburgh, Medical School, Division of Pathway Medicine, Edinburgh, UK.

Spring 2009 - Winter 2010: *MSc Tutor, Machine Learning and Pattern Recognition, BSc Tutor, Logic Programming.*

University of Edinburgh, School of Informatics, Edinburgh, UK.

Professional Activities

Referee for: AISTATS, Bioinformatics, BMC Bioinformatics, BMC Systems Biology, Computational Statistics, ECML, Ecological Indicators, JMLR, Machine Learning, Molecular Biosystems, NIPS, PLoS One, Statistical Applications in Genetics and Molecular Biology, Statistics and Computing, Theoretical Biology and Medical Modelling, Transactions on Computational Biology and Bioinformatics, Transactions on Pattern Analysis and Machine Intelligence.

Programme Committee Member for: IJCAI 2015, Machine Learning Track.

Software Packages

EDISON - *Estimation of Directed Interactions from Sequences Of Non-homogeneous gene expression.*

R CRAN package for heterogeneous dynamic Bayesian network reconstruction from time-varying data using information sharing. Developed with Sophie Lèbre.

<http://cran.r-project.org/web/packages/EDISON/index.html>

nethet - High-dimensional exploration of biological network heterogeneity

R Bioconductor package for two-sample testing in graphical Gaussian models and network-based clustering. Developed with Nicolas Städler.

<http://bioconductor.org/packages/devel/bioc/html/nethet.html>

Recent Training Courses

May 2016: *Presentation Skills Workshop*

Lancaster University

March 2016: *Research Grant Costing Workshop*
Lancaster University

February 2016: *Grant Proposal Workshop*
Lancaster University

January 2016: *PhD Supervision Workshop: The Secrets of Highly Successful Research Students*
Lancaster University

December 2015: *Research Ethics and Integrity*
Lancaster University

February 2015: *Successful Grant Writing*
Medical Research Council

January 2015: *High Performance Computing: Programming GPUs using CUDA*
University of Cambridge Computing Service

Programming Skills

- Matlab, R and Java, as well as knowledge of several other languages including PHP, Ruby, Python, C, Prolog and ML.
- Parallel computing on distributed and multi-core architectures.
- Package development experience in R.
- Handling relational databases with MySQL.
- Using Subversion, git and GitHub for version control.

Languages

- Fluent: English, German, French.
- Working Knowledge: Dutch, Spanish.

Interests

Public engagement with science, science writing, podcasting, reading, crocheting, cooking, baking.

References

Dr Sach Mukherjee, Programme Leader Statistical Genomics, MRC Biostatistics Unit, Cambridge, UK.

Professor Dirk Husmeier, School of Mathematics and Statistics, University of Glasgow, Glasgow, UK.

Dr Guido Sanguinetti, Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, Edinburgh, UK.