Mathew W. McLean

Location: Sydney, Australia
Mobile: +61 4 5238 9372
Email: mathew.w.mclean [at] gmail.com
Citizenship: Canadian
Australian Work Rights: Subclass 186 Visa
Website: http://mwmclean.github.io/

Education

Ph.D. in Operations Research, Cornell University, 2013

Concentration: Applied Probability and Statistics

Advisors: David Ruppert and Giles Hooker

Dissertation: On Generalized Additive Models for Regression with Functional Data

M.S. in Operations Research, Cornell University, 2013

B.Sc. (Hons.) in Statistics, University of Manitoba, 2008

Professional Experience

Data Scientist, Implementation Team, Displayr, 2021-2024/Apr

Developed and tested features (mostly written in R and JavaScript) for our market research software products, Q and Displayr, to simplify and automate data cleaning, analysis, visualization, and reporting for our users. Since joining, the company has grown from 20 to 80+ employees and nearly $14 million in ARR.

Setup CircleCI pipelines and an orb for continuous integration and deploying of R packages to Azure blog storage, achieving 5x reduction in build times.

Collaborate with Platform team to help maintain Ubuntu VMs for running R calculation requests sent from our web application, Displayr.

Developed novel algorithm for classifying open-ended text survey responses involving Google's Universal Sentence Encoder, clustering, and random forests. Later explored prompt engineering using ChatGPT for this task and compared with our existing algorithm.

Team Lead, Data Science Team, Displayr, 2019-2021

Maintained/co-authored a suite of 40+ R packages for statistical modeling, visualization, data cleaning and integrating with various web APIs.

Lead a small team of data scientists to write efficient, intuitive tools for performing text analysis, driver analysis, regression, and choice modeling using Scrum.

Our team wrote and maintained 1,000s of integration and unit tests for all data science features in our products.

Computational Statistician, Data Science Team, Displayr, 2017-2019

Designed user interface as well as interactive visualization and HTML widget output for data science and machine learning features in our apps, Displayr and Q.

Developed models implemented in stan and R for Hierarchical Bayes choice modeling, simulation, and experimental design.

Wrote blog posts detailing usage of new features and replied to customer questions.

Postdoctoral Research Fellow, University of Technology Sydney, 2015-2017

Supervisor: Matt P. Wand

Developed novel, fast, approximate algorithms for fitting Bayesian semiparametric and nonparametric models; applied methods to streaming data from robotics/computer vision.

Maintainer of http://realtime-semiparametric-regression.net which demonstrates the online fitting of semiparametric regression models using data scraped in real-time from various websites.

Research Assistant Professor, Texas A&M University, 2013-2015

Supervisor: Raymond J. Carroll

Developed new semiparametric methods to account for measurement error when analyzing dietary intake data; estimation performed using Markov Chain Monte Carlo implemented in R and C.

Used generalized linear models to analyze count data from a large RNA-Seq experiment to study the effects of diet on colon cancer.

Invited talks on Bayesian modelling for longitudinal data, hypothesis testing with mixed models, reproducible research, and authoring `R` software packages.

Four submitted or published peer-reviewed journal articles.

Sole author and maintainer for `RefManageR`, an R package available on CRAN. The package allows for importing bibliographic references via web APIs, PDFs, and `.bib` files into `R` and provides tools for easily manipulating, printing, and searching the stored references.

Produced reproducible research manuscripts and presentations in `Markdown`, LaTeX, and `HTML5` formats with `knitr`/`Sweave` and interactive web applications using RStudio's Shiny framework.

Reviewed original applied statistics research submitted for publication for ten statistics journals.

Research Assistant and Teaching Assistant, Cornell University, 2008-2013

Research in functional data analysis, semiparametric regression, variable selection for high dimensional models, Bayesian hierarchical modelling, mixed linear models, and time series analysis.

Three papers published in statistics journals with code contributed to `R` package `refund`.

Five invited talks and poster presentations on regression models for functional data.

Lectured to classes of approximately 30 students on introductory statistics for engineers, stochastic processes, and data mining; as well as ran labs demonstrating use of `Excel` and `R` for statistics.

Publications, talks, and software are available at `https://mwmclean.github.io/`.

Publications

McLean, M. W. and M. P. Wand (June 2019). "Variational Message Passing for Elaborate Response Regression Models". In: *Bayesian Analysis* 14.2, pp. 371–391. DOI: 10.1214/18-BA1098.

Triff, K., McLean, M. W., E. Callaway, J. Goldsby, I. Ivanov, and R. S. Chapkin (May 2018). "Dietary fat and fiber interact to uniquely modify global histone post-translational epigenetic programming in a rat colon cancer progression model". In: *International Journal of Cancer* 143.6, pp. 1402–1415. DOI: 10.1002/ijc.31525.

McLean, M. W. (2017). "RefManageR: Import and Manage BibTeX and BibLaTeX References in R". In: *The Journal of Open Source Software*. R package also accepted to ROpenSci. DOI: 10.21105/joss.00338.

Triff, K., McLean, M. W., K. Kranti, J. Pang, E. Callaway, B. Zhou, I. Ivanov, and R. S. Chapkin (June 2017). "Assessment of Histone Tail Modifications and Transcriptional Profiling During Colon Cancer Progression Reveals a Global Decrease in H3K4me3 Activity". In: *BBA - Molecular Basis of Disease* 1863.6, pp. 1392–1402. DOI: 10.1016/j.bbadis.2017.03.009.

McLean, M. W., G. Hooker, and D. Ruppert (2015). "Restricted Likelihood Ratio Tests for Linearity in Scalar-on-Function Regression". In: *Statistics and Computing* 25.5, pp. 997–1008. DOI: 10.1007/s11222-014-9473-1. arXiv: 1310.5811 [stat.ME].

McLean, M. W., G. Hooker, A. M. Staicu, F. Scheipl, and D. Ruppert (2013). "Functional Generalized Additive Models". In: *Journal of Computational and Graphical Statistics* 23.1, pp. 249–269. DOI: 10.1080/10618600.2012.729985.

Matteson, D. S., McLean, M. W., D. B. Woodard, and S. G. Henderson (2011). "Forecasting Emergency Medical Service Call Arrival Rates". In: *Annals of Applied Statistics* 5.2B, pp. 1379–1406. DOI: 10.1214/10-AOAS442. arXiv: 1107.4919 [stat.AP].

Manuscripts

McLean, M. W., C. J. Oates, and M. P. Wand (2017). *Real-Time Semiparametric Regression via Sequential Monte Carlo*. Submitted.

McLean, M. W., F. Scheipl, G. Hooker, S. Greven, and D. Ruppert (2017). *Bayesian Functional Generalized Additive Models with Sparsely Observed Covariates*. arXiv: 1305.3585 [stat.ME]. Submitted.

McLean, M. W. (2014). *Straightforward Bibliography Management in R Using the RefManageR Package*. arXiv: 1403.2036 [cs.DL].

Programming Skills

R (including tidyverse, plotly, rstan, and htmlwidgets), Linux, Emacs/ESS, Git, CircleCI, GitHub Actions, RStudio, python, docker, HTML, JIRA, Scrum, SQL

Past: SAS, C, Java, MS Office, MATLAB

Certifications

SAFe Practice Consultant (SPC) (2024)

SAFe Practice Consultant (RTE) (2024)

SAFe Advanced Scrum Master (SASM) (2024)

Professional Agile Leadership - Evidence Based Management (PAL-EBM) (2024)

Certified Agile Leader (CAL 1) (2024)

Professional Scrum Master II, 2024

Professional Scrum Product Owner I, 2024

Professional Scrum Facilitation Skills I, 2024

Introduction to Relational Databases (IBM/Coursera), 2024

Python Project for Data Engineering (IBM/Coursera), 2024

Python for Data Science, AI & Development (IBM/Coursera), 2023

Professional Scrum Master I, 2022

Honors, Awards, & Fellowships

NSERC PGS-D, 2011–2013

NSERC PGS-M (CGS-M Declined), 2008–2009

NSERC Undergraduate Student Research Award, University of Manitoba, 2007 and 2008

Dr. Cyril H. Goulden Memorial Scholarship, 2008

U. of M. Students Union Scholarship, University of Manitoba, 2005, 2006, and 2007

Queen Elizabeth II Entrance Scholarship, 2004-2005

Open-Source Software

`RefManageR` an R package available on CRAN, for which I am sole author and maintainer. The package allows for importing, printing, and working with BibTeX and BibLaTeX bibliographic references in R.

`refund` an R package available on CRAN for regression with functional data, maintained by Lei Huang (2012). I have contributed the following components that allow for fitting, prediction, and visualization of functional generalized additive models: `fgam()`, `af()`, `lf()`, `predict.fgam()`, `vis.fgam()`. I assisted with adding this functionality to the function `pfr()` in the same package.

Online Appendix to "Functional Generalized Additive Models", containing R code used in simulations.

Contributions to the R package `curvHDR` to make the package suitable for CRAN, such as adding a package namespace and including compiled code.

Talks and Poster Presentations

Bayesian Hierarchical Models for Sparse Functional Data

ERCIM 2014, Pisa, Italy, December, 2014 (invited)

Restricted Likelihood Ratio Tests for Regression with Functional Data

2014 WNAR/IMS Conference, Honolulu, HI, June, 2014 (invited)

ENAR 2014 Spring Meeting, Baltimore, MD, March, 2014 (contributed)

Reproducible Research and Dynamic Documents with R

IAMCS Machine Learning and Applied Statistics Workshop Series, College Station, TX, 2014 (invited)

Authoring R packages

IAMCS Machine Learning and Applied Statistics Workshop Series, College Station, TX, 2014 (invited)

Functional Generalized Additive Models

Department of Statistics, University of Auckland, Auckland, New Zealand, August, 2017 (invited)

Annual Conference of the German and Austrian Statistical Association, Vienna, Austria, September, 2012 (invited)

ORIE PhD Student Colloquium, September, 2012 (talk)

Imaging, Communications and Finance: Stochastic Modeling of Real-world Problems, Columbia University, June, 2011 (poster)

SAMSI Closing Workshop on the Analysis of Object Data, SAMSI, Research Triangle Park, June, 2011 (poster)

Statistical Methods for Very Large Datasets Conference, Johns Hopkins University, June, 2011 (poster)

Joint Statistical Meetings, Vancouver, Canada, August, 2010 (talk)

SCoOp: Statistical Computation of Option Prices (poster)

University of Manitoba

Professional Activities

Referee for Biometrics, Biometrika, Journal of the American Statistical Association, Environmetrics, Journal of Computational and Graphical Statistics, Journal of the Royal Statistical Society: Series C, Computational Statistics and Data Analysis, Journal of Multivariate Analysis, Statistica Sinica, Journal of Statistical Software, Statistical Modelling

Cornell Operations Research Graduate Student Association President, 2011-2012