

CONTACT AND  
GENERAL INFO

2500 boulevard de l'Université  
Sherbrooke, QC, J1K-2R1  
toby.dylan.hocking@usherbrooke.ca  
+1(819)821-8000x65565

Birth: 17 March 1984 in Newport Beach, California  
Citizenship: USA, Permanent residence: Canada.  
Languages: English (native), French (fluent since 2009).  
Web: <http://tdhock.github.io>, Erdős number = 3.

RESEARCH  
INTERESTS

Fast, accurate, and interpretable machine learning algorithms. Free/open-source statistical software for machine learning, data visualization, and optimization. Applications in genomics, neuroscience, medicine, microbiome, cybersecurity, robotics, satellite/sonar imagery, climate/carbon modeling.

PROFESSIONAL  
EXPERIENCE

**Université de Sherbrooke**, Canada (2024–present).

Tenured Associate Professor, Département d'informatique.  
IVADO researcher, MILA associate academic member.  
“Learning Algorithms, Statistical Software, Optimization (LASSO).”

**Northern Arizona University**, Flagstaff, Arizona, USA (2018–2024).

Tenure-Track Assistant Professor, School of Informatics, Computing, and Cyber Systems.  
“Optimization algorithms for machine learning and interactive data analysis.”

**McGill University**, Montreal, Canada (2014–2018).

Postdoc with Guillaume Bourque, Department of Human Genetics.  
“Changepoint detection and regression models for peak detection in genomic data.”

**Tokyo Institute of Technology**, Tokyo, Japan (2013).

Postdoc with Masashi Sugiyama, Department of Computer Science.  
“Support vector machines for ranking and comparing.”

**Sangamo BioSciences**, Richmond, CA, USA (2006–2008).

Research Assistant with Jeff Miller in the Zinc Finger Technology group.  
“A web app for visualization and statistical analysis of experimental data.”

## EDUCATION

**École Normale Supérieure**, Cachan, France (2009–2012).

Math Ph.D. with Francis Bach, Département d'Informatique; Jean-Philippe Vert, Institut Curie.  
“Learning algorithms and statistical software, with applications to bioinformatics.”

**Université Paris 6**, Paris, France (2008–2009).

Master of Statistics, internship at INRA with Mathieu Gautier and Jean-Louis Foulley.  
“A Bayesian Outlier Criterion to Detect SNPs under Selection in Large Data Sets.”

**University of California, Berkeley**, CA, USA (2002–2006).

Double B.A. in Statistics, Molecular and Cell Biology; thesis in Statistics with Terry Speed.  
“Chromosomal copy number analysis using SNP microarrays and a binomial test statistic.”

HONORS AND  
AWARDS  
(SELECTED)

PI on NSERC Discovery grant, 210,000\$CAD, 2026–2031. “Optimization algorithms for complex objective functions in supervised machine learning, with applications to classification and change-point detection.”

Consultant on NIH grant (Autism Data Science Initiative) Sep 2025–Aug 2028, USD\$4,249,998 total, USD\$22,896 my share. “Advancing Success & Developmental Outcomes in Autism Spectrum Disorder through Analysis of Secondary Data (ASD3 Outcomes Project).”

PI on grant from Quebec Education Minister (FabriqueREL—free online educational resources in French), 15,000\$CAD, 2025–2026. “French translation for user manual of animint2 interactive data visualization R package.”

Project co-lead on grant from UK Medical Research Council (Better Methods, Better Research program), £50,000, 2025–2028 (my share: £3,694). “UKRI1573: Maintaining and expanding the data.table R package for fast and efficient data manipulation.” Lead Heather Turner.

PI on travel grant from DATAIA (Artificial Intelligence Institute of Université Paris-Saclay), €25,500, Jan–June 2025. “Efficient algorithms and software for change-point detection.”

PI on National Science Foundation grant 2303612 from program Pathways to Enable Open-Source Ecosystems (POSE), US\$731,881, Sep 2023–Aug 2025. “POSE: Phase II: Expanding the data.table ecosystem for efficient big data manipulation in R.” Co-PIs Igor Steinmacher and Marco Gerosa.

Co-PI on National Institute of Mental Health grant 3R01MH134177-02S1, US\$455,660, Aug 2023 to June 2027. “Addressing Structural Disparities in Autism Spectrum Disorder through Analysis of Secondary Data (ASD3).” Role: advise PHD student and summer salary for machine learning analysis of health records related to childhood autism. US\$40K under my control = US\$10K in each of 4 academic years.

Senior Personnel on Department of Energy grant, US\$3,600,000, Sept 2022 to Sept 2025. “Friends and Foes: microbial interactions and soil biogeochemistry after 23 years of experimental warming.” Role: summer salary for machine learning analysis of microbial interaction data. US\$30K under my control = US\$10K in each of 3 academic years.

Co-PI on National Science Foundation grant 2125088 from program URoL-Understanding the Rules of Life, US\$3,000,000, Sept 2021 to Aug 2026. “MIM: Discovering in reverse – using isotopic translation of omics to reveal ecological interactions in microbiomes.” Role: supervise PHD student working on machine learning analysis of new metabolic flux data, in order to infer new types of interactions between microbes. US\$200K under my control = US\$40K in each of 5 academic years.

Senior personnel on NAU US\$819K sub-contract of Department of Energy grant, US\$9,000,000, 2022-2025, Lawrence Livermore National Laboratory Science Focus Area program entitled “Microbes Persist: Systems Biology of the Soil Microbiome” led by PI Jennifer Pett-Ridge. Role: machine learning analysis of microbiome interaction networks, identifying taxa and traits that are associated with soil carbon cycling processes. US\$15K under my control = US\$5K in each of 3 academic years.

Senior personnel on Missouri Department of Education grant, US\$1,509,570, July 2021–June 2023, contract entitled “MMD-DCI Research, Development, & Leadership” led by PI Ronda Jenson. Role: summer salary and mentoring graduate students on interpretable machine learning algorithms for Predictive Modeling Framework for District Continuous Improvement. US\$50K total under my control = US\$25K in each of two academic years.

Air Force Research Laboratory, Summer Faculty Fellowship, US\$20,000, May–July 2021, “Machine learning algorithms for understanding physically unclonable functions based on resistive memory devices.”

PI on R Consortium Grant, US\$34,000, Jan–Dec 2020, “RcppDeepState: an easy way to fuzz test compiled code in R packages.”

“Mobilité entrant” travel award, research about dynamic programming algorithms for constrained change-point detection, with Guillem Rigail in Université Evry, France, 2016.

International userR conference, Best Student Poster Award, “Adding direct labels to plots,” 2011.

INRIA/INRA (French computer science and agricultural research institutes), Ph.D. scholarship, 2009 (declined).

UC Berkeley, Department of Statistics VIGRE research scholarship, 2005.

PAPERS IN  
PROGRESS AND  
UNDER REVIEW

Rust K, **Hocking TD**. A Log-linear Gradient Descent Algorithm for Unbalanced Binary Classification using the All Pairs Squared Hinge Loss. Under review in *Journal of Machine Learning Research*, arXiv:2302.11062.

Amoakohene D, Chetia A, Steinmacher I, Hocking TD. Asymptotic benchmarking using the atime package. *R Journal* (in press).

PEER-REVIEWED  
JOURNAL PAPERS

Hocking TD, Thibault G, Bodine CS, Arellano PN, Shenkin AF, Lindly O. SOAK: Same/Other/All K-fold cross-validation for estimating similarity of patterns in data subsets. *Statistical Analysis and Data Mining* 19(1), 2026.

Agyapong D, Propster JR, Marks J, and **Hocking TD**. Cross-validation for training and testing co-occurrence network inference algorithms. *BMC Bioinformatics*, 26(74), 2025.

Nguyen TL and **Hocking TD**. Penalty Learning for Optimal Partitioning using Multilayer Perceptron. *Statistics and Computing*, 35, 2025.

Gurney KR, Aslam B, Dass P, Gawuc L, **Hocking TD**, Barber JJ, and Kato A. Assessment of the climate trace global powerplant CO2 emissions. *Environmental Research Letters*, 19(11):114062, 2024.

Kaufman J, Stenberg A, **Hocking TD**. Functional Labeled Optimal Partitioning. *Journal of Computational and Graphical Statistics*, DOI:10.1080/10618600.2023.2293216. (2024)

Bodine CS, **Hocking TD**, Buscombe D. Automated river substrate mapping from sonar imagery with machine learning. *Journal of Geophysical Research — Machine Learning and Computation* 1(3). (2024)

Tao F, Houlton BZ, Frey SD, Lehmann J, Manzoni S, Huang Y, Jiang L, Mishra U, Hungate BA, Schmidt MWI, Reichstein M, Carvalhais N, Ciais P, Wang Y-P, Ahrens B, Hugelius G, **Hocking TD**, Lu X, Shi Z, Viatkin K, Vargas R, Yigini Y, Omuto C, Malik AA, Peralta G, Cuevas-Corona R, Di Paolo LE, Luotto I, Liao C, Liang Y-S, Saynes VS, Huang X, Luo Y. Reply to: Model uncertainty obscures major driver of soil carbon. *Nature* (2024), volume 627, pages E4-E6.

Tao F, Huang Y, Hungate BA, Manzoni S, Frey SD, Schmidt MWI, Reichstein M, Carvalhais N, Ciais P, Jiang L, Lehmann J, Mishra U, Hugelius G, **Hocking TD**, Lu X, Shi Z, Viatkin K, Vargas R, Yigini Y, Omuto C, Malik AA, Peralta G, Cuevas-Corona R, Di Paolo LE, Luotto I, Liao C, Liang YS, Saynes VS, Huang X, Luo Y. Microbial carbon use efficiency promoting global soil carbon storage. *Nature* (2023). DOI:10.1038/s41586-023-06042-3.

Hillman J, **Hocking TD**. Optimizing ROC Curves with a Sort-Based Surrogate Loss Function for Binary Classification and Change-point Detection. *Journal of Machine Learning Research* 24(70):1-24, 2023.

Runge V, **Hocking TD**, Romano G, Afghah F, Fearnhead P, Rigai G. gfpop: an R Package for Univariate Graph-Constrained Change-point Detection. *Journal of Statistical Software* (2023), Volume 106, Issue 6.

Harshe K, Williams J, **Hocking TD**, Lerner Z. Predicting Neuromuscular Engagement to Improve Gait Training with a Robotic Ankle Exoskeleton. *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 5055-5060, Aug. 2023, doi: 10.1109/LRA.2023.3291919.

- Hocking TD**, Srivastava A. Labeled Optimal Partitioning. *Computational Statistics* DOI: 10.1007/s00180-022-01238-z (2022).
- Mihaljevic JR, Borkovec S, Ratnavale S, **Hocking TD**, Banister KE, Eppinger JE, Hepp CM, Doerry E. SPARSEMODr: Rapidly simulate spatially explicit and stochastic models of COVID-19 and other infectious diseases. *Biology Methods & Protocols*, Volume 7, Issue 1, 2022.
- Barnwal A, Cho H, **Hocking TD**. Survival regression with accelerated failure time model in XGBoost. *Journal of Computational and Graphical Statistics* (2022).
- Hocking TD**, Rigai G, Fearnhead P, Bourque G. Generalized Functional Pruning Optimal Partitioning (GFPOP) for Constrained Changepoint Detection in Genomic Data. *Journal of Statistical Software*, Vol. 101, Issue 10 (2022).
- Chaves AP, Egbert J, **Hocking TD**, Doerry E, Gerosa MA. Chatbots language design: the influence of language use on user experience. *ACM Transactions on Computer-Human Interaction* 29, 2, Article 13 (2022).
- Hocking TD**, Vargovich J. Linear Time Dynamic Programming for Computing Breakpoints in the Regularization Path of Models Selected From a Finite Set. *Journal of Computational and Graphical Statistics* (2021), doi:10.1080/10618600.2021.2000422.
- Hocking TD**. Wide-to-tall data reshaping using regular expressions and the nc package. *R Journal* (2021), doi:10.32614/RJ-2021-029.
- Liehrmann A, Rigai G, **Hocking TD**. Increased peak detection accuracy in over-dispersed ChIP-seq data with supervised segmentation models. *BMC Bioinformatics* 22, Article number: 323 (2021).
- Fotoohinasab A, **Hocking TD**, Afghah F. A Greedy Graph Search Algorithm Based on Changepoint Analysis for Automatic QRS-Complex Detection. *Computers in Biology and Medicine* 130 (2021).
- Abraham A, Prys-Jones T, De Cuyper A, Ridenour C, Hempson G, **Hocking TD**, Clauss M, Doughty C. Improved estimation of gut passage time considerably affects trait-based dispersal models. *Functional Ecology* (2021); 35: 860-869.
- Hocking TD**, Rigai G, Fearnhead P, Bourque G. Constrained dynamic programming and supervised penalty learning algorithms for peak detection in genomic data. *Journal of Machine Learning Research* 21(87):1-40, (2020).
- Hocking TD**. Comparing namedCapture with other R packages for regular expressions. *R Journal* (2019). doi:10.32614/RJ-2019-050
- Jewell S, **Hocking TD**, Fearnhead P, Witten D. Fast Nonconvex Deconvolution of Calcium Imaging Data. *Biostatistics* (2019), doi: 10.1093/biostatistics/kxy083.
- Depuydt P, Koster J, Boeva V, **Hocking TD**, Speleman F, Schleiermacher G, De Preter K. Meta-mining of copy number profiles of high-risk neuroblastoma tumors. *Scientific Data* (2018).
- Alirezaie N, Kernohan KD, Hartley T, Majewski J, **Hocking TD**. ClinPred: Prediction Tool to Identify Disease-Relevant Nonsynonymous Single-Nucleotide Variants. *American Journal of Human Genetics* (2018). doi:10.1016/j.ajhg.2018.08.005
- Sievert C, Cai J, VanderPlas S, Khan F, Ferris K, **Hocking TD**. Extending ggplot2 for linked and dynamic web graphics. *Journal of Computational and Graphical Statistics* (2018).
- Depuydt P, Boeva V, **Hocking TD**, *et al.* Genomic Amplifications and Distal 6q Loss: Novel

Markers for Poor Survival in High-risk Neuroblastoma Patients. *Journal of the National Cancer Institute* (2018). DOI:10.1093/jnci/djy022.

**Hocking TD**, Goerner-Potvin P, Morin A, Shao X, Pastinen T, Bourque G. Optimizing ChIP-seq peak detectors using visual labels and supervised machine learning. *Bioinformatics* (2017) 33 (4): 491-499.

Shimada K, Shimada S, Sugimoto K, Nakatochi M, Suguro M, Hirakawa A, **Hocking TD**, Takeuchi I, Tokunaga T, Takagi Y, Sakamoto A, Aoki T, Naoe T, Nakamura S, Hayakawa F, Seto M, Tomita A, Kiyoi H. Development and analysis of patient-derived xenograft mouse models in intravascular large B-cell lymphoma. *Leukemia* (2016).

Chicard M, Boyault S, Colmet-Daage L, Richer W, Gentien D, Pierron G, Lapouble E, Bellini A, Clement N, Iacono I, Bréjon S, Carrere M, Reyes C, **Hocking TD**, Bernard V, Peuchmaur M, Corradini N, Faure-Contier C, Coze C, Plantaz D, Defachelles A-S, Thebaud E, Gambart M, Millot F, Valteau-Couanet D, Michon J, Puisieux A, Delattre O, Combaret V, Schleiermacher G. Genomic copy number profiling using circulating free tumor DNA highlights heterogeneity in neuroblastoma. *Clinical Cancer Research* (2016).

Maidstone R, **Hocking TD**, Rigai G, Fearnhead P. On optimal multiple changepoint algorithms for large data. *Statistics and Computing* (2016). doi:10.1007/s11222-016-9636-3

Suguro M, Yoshida N, Umino A, Kato H, Tagawa H, Nakagawa M, Fukuhara N, Karnan S, Takeuchi I, **Hocking TD**, Arita K, Karube K, Tsuzuki S, Nakamura S, Kinoshita T, Seto M. Clonal heterogeneity of lymphoid malignancies correlates with poor prognosis. *Cancer Sci.* (2014) Jul;105(7):897-904.

**Hocking TD**, Boeva V, Rigai G, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Richer W, Bourdeaut F, Suguro M, Seto M, Bach F, Vert J-P. SegAnnDB: interactive Web-based genomic segmentation. *Bioinformatics* (2014) 30 (11): 1539-1546. DOI:10.1093/bioinformatics/btu072

**Hocking TD**, Wutzler T, Ponting K and Grosjean P. Sustainable, extensible documentation generation using inlinedocs. *Journal of Statistical Software* (2013), 54(6), 1-20. DOI:10.18637/jss.v054.i06

**Hocking TD**, Schleiermacher G, Janoueix-Lerosey I, Boeva V, Cappo J, Delattre O, Bach F, Vert J-P. Learning smoothing models of copy number profiles using breakpoint annotations. *BMC Bioinfo.* (2013), 14:164. DOI:10.1186/1471-2105-14-164

Gautier M, **Hocking TD**, Foulley JL. A Bayesian outlier criterion to detect SNPs under selection in large data sets. *PloS ONE* 5 (8), e11913 (2010).

Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C, Katibah GE, Amora R, **Hocking TD**, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Amacher SL. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nature biotechnology* 26 (6), 702-70 (2008).

PEER-REVIEWED  
CONFERENCE  
PAPERS

In addition to peer-reviewed journals, I publish papers at highly competitive computer science conferences like *ICML* and *NeurIPS*, with double-blind peer reviews, and  $\approx 20\%$  acceptance rates.

Oliveira P, Amoakohene D, **Hocking TD**, Gerosa M, Steinmacher I. Governance Matters: Lessons from Restructuring the data.table OSS Project. *IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 2025.

Barr J, **Hocking TD**, Morton G, Thatcher T, Shaw P. Classifying Imbalanced Data with AUM Loss. 2022 Fourth International Conference on Transdisciplinary AI (TransAI).

**Hocking TD**, Barr J, Thatcher T. Interpretable linear models for predicting security vulnerabilities

in source code. 2022 Fourth International Conference on Transdisciplinary AI (TransAI).

Barr J, Shaw P, Abu-Khzam FN, Thatcher T, **Hocking TD**. Graph Embedding: A Methodological Survey. 2022 Fourth International Conference on Transdisciplinary AI (TransAI).

Kolla AC, Groce A, **Hocking TD**. Fuzz Testing the Compiled Code in R Packages. IEEE 32nd International Symposium on Software Reliability Engineering (ISSRE 2021), pp. 300-308, doi: 10.1109/ISSRE52982.2021.00040.

Fotoohinasab A, **Hocking TD**, Afghah F. A Graph-Constrained Change-point Learning Approach for Automatic QRS-Complex Detection. *Asilomar Conference on Signals, Systems, and Computers* (2020).

Fotoohinasab A, **Hocking TD**, Afghah F. A Graph-constrained Change-point Detection Approach for ECG Segmentation. *42th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2020)*.

**Hocking TD**, Bourque G. Machine Learning Algorithms for Simultaneous Supervised Detection of Peaks in Multiple Samples and Cell Types. *Pacific Symposium on Biocomputing* 25:367-378 (2020).

Drouin A, **Hocking TD**, Laviolette F. Maximum margin interval trees. *Neural Information Processing Systems (NeurIPS)*, 2017.

**Hocking TD**, Rigai G, Bourque G. PeakSeg: constrained optimal segmentation and supervised penalty learning for peak detection in count data. *International Conference on Machine Learning (ICML)*, 2015.

**Hocking TD**, Rigai G, Bach F, Vert J-P. Learning sparse penalties for change-point detection using max-margin interval regression. *International Conference on Machine Learning (ICML)*, 2013.

**Hocking TD**, Joulin A, Bach F, Vert J-P. Clusterpath: an Algorithm for Clustering using Convex Fusion Penalties. *International Conference on Machine Learning (ICML)*, 2011.

BOOKS, CHAPTERS,  
MANUALS **Hocking TD**. Introduction to Machine Learning and Neural Networks. Chapter 36 (pages 309-317) in textbook *Land Carbon Cycle Modeling: Matrix Approach, Data Assimilation, and Ecological Forecasting*, edited by Yiqi Luo and Benjamin Smith, published in 2022 by Taylor and Francis.

**Hocking TD** (2018). Animated interactive data visualization using animint2 in R. Free online textbook, 20 web pages/chapters with interactive graphics and exercises.

**Hocking TD** (2026). Visualisation interactive de données dans R avec animint2. Université de Sherbrooke, Département d'informatique. fabriqueREL. Licence CC BY-SA. Free online textbook, 20 web pages/chapters with interactive graphics and exercises. DOI:10.71892/11143/1448

CONFERENCE  
TUTORIALS **Hocking TD**, Killick R. Introduction to optimal change-point detection algorithms, *useR* 2017.

**Hocking TD**, Ekstrøm CT. Understanding and creating interactive graphics, *useR* 2016.

INVITED TALKS  
(SELECTED) Semi-plenary talk at JDS Marseille, Institut Henri Poincaré, round table panelist for Impact IA colloquium at Paris Saclay, AgroParisTech, Centre Borelli group meeting at ENS Paris Saclay, Madrid useR group, Julie Josse lab meeting at INRIA Montpellier, Institut de Maths de Toulouse, Toulouse R User Group, Toulouse Data Viz meetup, INRAE Toulouse, ETH Zurich, LMU Munich, Rencontres R Mons Belgium (2025); ASU Data-oriented Mathematical and Statistical Sciences seminar, Joint Statistical Meetings Toronto, Université Laval, Université Sherbrooke (2023); Institute of Mathematical Statistics London, IEEE conference in Prescott Arizona, University of Arizona Math/Stats seminar (2022); ASU West ML Day, TRIPODS University of Arizona, IEEE NJACS

(2021); NAU Math Department Colloquium (2018); University of Waterloo, Université de Montréal, Sainte-Justine Children’s Hospital, University of Québec à Montréal, Polytechnique Montréal (2017); Université Laval (2016); McGill Barbados epigenomics workshop (2015); Sapporo Japan Workshop on Machine Learning and Applications to Biology (2013); Google Research New York, Université Rennes, Université Angers, INRIA Lille (2012); Institut de Biologie de Lille (2011).

CONSULTING  
(SELECTED)

Acronis SCS, cybersecurity company in Phoenix (2022). Interpretable and non-linear machine learning algorithms which use source code analysis to predict software vulnerabilities.

Plotly, data visualization startup in Montreal (2015). Original author of ggplot functionality in plotly R package.

TEACHING

All of my course materials are freely available online, <https://tdhock.github.io/teaching/>

Summer 2026, Université de Sherbrooke, ROP543/631, Optimisation non linéaire (nonlinear optimization).

Winter 2026, Université de Sherbrooke, IFT704, Apprentissage automatique à grande échelle (large-scale machine learning).

Fall 2024, Université de Sherbrooke, IFT603/712, Techniques d’apprentissage (machine learning).

Fall 2024, Université de Sherbrooke, IFT187, Éléments de bases de données (databases).

Fall 2023, Northern Arizona University, CS470, Artificial Intelligence.

Fall 2023, Northern Arizona University, CS599, Deep Learning.

Fall 2023, Northern Arizona University, CS599, Unsupervised Learning.

Spring 2023, 2 hour lecture “Introduction to Deep Learning in R” for graduate student training “Research Bazaar Arizona.”

Spring 2023, Northern Arizona University, CS470, Artificial Intelligence.

Spring 2023, Northern Arizona University, CS105/205/305, Computing Tools.

Fall 2022, Northern Arizona University, CS499, Deep Learning.

Fall 2022, Northern Arizona University, CS105/205/305, Computing Tools.

Spring 2022, Northern Arizona University, CS570, Deep Learning.

Fall 2021, Northern Arizona University, CS499/599, Unsupervised Learning.

Summer 2021, 1 hour lecture “Introduction to Machine Learning and Neural Networks” for summer school on “New Advances in Land Carbon Cycle Modeling.”

Spring 2021, Northern Arizona University, CS470, Artificial Intelligence.

Fall 2020, Northern Arizona University, CS499/599, Unsupervised Learning.

Summer 2020, 90 minute lecture “Introduction to Machine Learning and Neural Networks” for summer school on “New Advances in Land Carbon Cycle Modeling.”

Spring 2020, Northern Arizona University, CS499, Deep Learning.

Fall 2019, Northern Arizona University, CS/EE599, Reproducible Machine Learning Research.

Spring 2019, Northern Arizona University, CS499, Optimization algorithms for machine learning.

PROFESSIONAL  
SERVICE

2012–present, co-administrator and mentor for R project in Google Summer of Code (teaching free/open-source software development, how to create and improve R packages).

2018–present, editor for Journal of Statistical Software.

2021–present, machine learning editor for rOpenSci Statistical Software.

2021–present, co-author of R Development Guide and member of R Contribution Working Group (resources for making it easy/accessible to contribute improvements to base R).

2023–present, co-author of Omics CRAN Task View.

2023, co-organized WNAR session on changepoint detection, with Ning Hao and Selena Niu.

2022, organized two topic contributed sessions about changepoint detection for Institute of Mathematical Statistics meeting in London.

2022, Member of steering committee for R project in Google Season of Docs.

2017–2018, president of organizing committee for “R in Montreal 2018” conference.

2010–present: peer reviewer for Technometrics, International Conference on Machine Learning (ICML), Advances in Neural Information Processing Systems (NeurIPS), Journal of Machine Learning Research (JMLR), Artificial Intelligence Review, Journal of Computational and Graphical Statistics (JCGS), R Journal, Bioinformatics, PLOS Computational Biology, BMC Bioinformatics, IEEE Transactions on Pattern Analysis and Machine Intelligence, Information and Inference, Journal of Statistical Computation and Simulation, Computo, Genome Biology.

SOFTWARE ONLINE  
(SELECTED)

Numerous original/novel free/open-source software contributions using R, C, C++, Python, and JavaScript. Available from standard repositories (CRAN for R, PyPI for Python) as well as GitHub: <https://github.com/tdhock>, links to my software contributions <https://tdhock.github.io/software/>

**R:** contributions to regular expressions in base R (2011–2023). Co-author since 2024 of `data.table` package (I wrote governance, `measure` function for data reshaping). Maintainer of numerous R packages (20 on CRAN as of April 2026) for machine learning (cross-validation benchmarks, changepoint detection, classification, regression, ranking, etc), `directlabels` for labeled figures, `animint2` for animated interactive figures, `inlinedocs` for documentation generation, `atime` for asymptotic benchmarks.

**Python:** contributions to `pandas` module for data manipulation (`str.extractall` regex functionality). Author of GUI/web server software for labeling and changepoint detection in genomic data (`annotate_regions`, `SegAnnDB`, `PeakLearner`).

REFERENCES

**Bruce Hungate**, collaborator on research papers and grants related to machine learning for ecological data analysis.

Regents’ Professor, Biological Sciences, Northern Arizona University  
E-mail: [Bruce.Hungate@nau.edu](mailto:Bruce.Hungate@nau.edu). Telephone: 928-523-0925

**Yiqi Luo**, collaborator on research and teaching related to machine learning for climate science (summer school, textbook).

Professor, School of Integrative Plant Science, Cornell University  
E-mail: [y12735@cornell.edu](mailto:y12735@cornell.edu)

**Olivia Lindly**, collaborator on research papers and grants related to interpretable machine learning algorithms for understanding autism.

Associate Professor, Department of Health Sciences, Northern Arizona University

Phone: +1 (928) 523-5155, E-mail: Olivia.Lindly@nau.edu

**Jarrett “Jay” Barber**, colleague familiar with my teaching related to statistical machine learning.

Associate Professor, School of Informatics, Computing and Cyber Systems,

Northern Arizona University, Phone: +1 (928) 523-6869, E-mail: Jarrett.Barber@nau.edu

**Alex Groce**, collaborator on research papers and grants.

Associate Professor, School of Informatics, Computing and Cyber Systems,

Northern Arizona University, Phone: +1 (928) 523-8263, E-mail: Alex.Groce@nau.edu

**Fatemeh Afghah**, collaborator on research papers and grants.

Associate Professor, Clemson University, Phone: +1 (864) 656-0100, E-mail: fafghah@clemson.edu,

Web: <https://fafghah.people.clemson.edu/>

**Guillem Rigail**, collaborator on research papers and grants.

Researcher at INRAE (French Agronomy Research Institute)

Phone: +33 (0) 1 64 85 35 44, E-mail: guillem.rigail@inrae.fr

Web: <http://www.math-evry.cnrs.fr/members/Grigail/welcome>

**Rebecca Killick**, collaborator on a conference tutorial.

Professor of Statistics at Lancaster University

Phone: +44 (0)1524 593780, E-mail: r.killick@lancaster.ac.uk

Web: <http://www.lancs.ac.uk/~killick/>