Paul David McNicholas

Department of Mathematics & Statistics, McMaster University, Hamilton, Ontario, Canada, L8S 4L8.

Nationality: Irish/Canadian **Telephone (Work):** +1-905-5259140, ext. 23419

E-mail: paulmc@mcmaster.ca **Web:** www.paulmcnicholas.io

Employment

- Director, MacDATA Institute, McMaster University. July 2017–Present.
- Professor, Department of Mathematics and Statistics, McMaster University. July 2014–Present.
- Professor, Department of Mathematics and Statistics, University of Guelph. July 2013–June 2014.
- Bioinformatics Director, University of Guelph. April 2012–May 2014.
- Associate Chair, Department of Mathematics and Statistics, University of Guelph. Sept. 2010–Nov. 2013.
- Associate Professor, Department of Mathematics and Statistics, University of Guelph. July 2010–June 2013.
- Assistant Professor, Department of Mathematics and Statistics, University of Guelph. July 2007–June 2010.

Education

- Ph.D. in Statistics, 2007, *Topics in Unsupervised Learning*, Department of Statistics, Trinity College Dublin.
- M.Sc. in High Performance Computing, 2007, School of Mathematics, Trinity College Dublin.
- B.A. in Mathematics, 2004, First Class Honours with Gold Medal, School of Mathematics, Trinity College Dublin (M.A., 2007).

Awards & Honours

- John L. Synge Award, Royal Society of Canada, 2021.
- Steacie Prize for the Natural Sciences, E.W.R. Steacie Memorial Fund, 2020.
- E.W.R. Steacie Memorial Fellowship, NSERC, 2019.
- Member, College of the Royal Society of Canada, 2017.
- University Scholar, McMaster University, 2017.
- Tier 1 Canada Research Chair in Computational Statistics, NSERC, 2015.
- Chikio Hayashi Award, International Federation of Classification Societies (IFCS), 2013.
- Early Researcher Award, Ontario Ministry of Research & Innovation, 2011.
- Distinguished Professor Teaching Award, University of Guelph Faculty Association, 2011.
- University Research Chair in Computational Statistics, University of Guelph, 2011.
- Barrington Medal, Statistical and Social Inquiry Society of Ireland, 2006.

Named Lectures

- Isobel Loutit Lecture, Statistical Society of Canada, 2020.
- Barrington Lecture, Statistical and Social Inquiry Society of Ireland, 2006.

Monographs

Note: current and former trainees (students and postdoctoral fellows) are typed in bold face in all publications herein.

[1] McNicholas, P.D. and **Tait, P.A.** (2019), *Data Science with Julia*. Boca Raton: Chapman & Hall/CRC Press.

[2] McNicholas, P.D. (2016), *Mixture Model-Based Classification*. Boca Raton: Chapman & Hall/CRC Press.

Peer-Reviewed Articles: Published and Finally Accepted

[3] **Browne, R.P.**, McNicholas, P.D. and Findlay, C.J., 'A partial EM algorithm for model-based clustering with highly diverse missing data patterns', *Stat*. To appear. Accepted, November 2021.

[4] **Gallaugher, M.P.B.**, Tomarchio, S.D., McNicholas, P.D. and Punzo, A., 'Multivariate cluster weighted models using skewed distributions', *Advances in Data Analysis and Classification*. To appear. Accepted, October 2021.

[5] **Gallaugher, M.P.B.**, Tomarchio, S.D., Punzo, A. and McNicholas, P.D., 'Mixtures of contaminated matrix variate normal distributions', *Journal of Computational and Graphical Statistics*. To appear. Accepted, October 2021.

[6] Georgiades, S., **Tait, P.A.**, McNicholas, P.D., Eric Duku, E., Zwaigenbaum, L., Smith, I.M., Bennett, T., Elsabbagh, M., Kerns, C.M., Mirenda, P., Ungar, W.J., Vaillancourt, T., Volden, J., Waddell, C., Zaidman-Zait, A., Gentles, S. and Szatmari, P.M. (2022), 'Trajectories of symptom severity in children with autism: Variability and turning points through the transition to school', *Journal of Autism and Developmental Disorders* **52**(1), 392–401.

[7] Tomarchio, S.D., McNicholas, P.D. and Punzo, A. (2021), 'Matrix normal cluster-weighted models', *Journal of Classification* **38**(3), 556–575.

[8] Vrkljan, B., Beauchamp, M.K., Gardner, P., Fang, Q., Kuspinar, A., McNicholas, P.D., Newbold, K.B., Richardson, J., Scott, D., Zargoush, M. and Gruppuso, V. (2021),'Re-engaging in aging and mobility research in the COVID-19 era: Early lessons from pivoting a large-scale, interdisciplinary study amidst a pandemic', *Canadian Journal on Aging* **40**(4), 669-675.

[9] **Tang, Y.**, Qazi, M.A., Brown, K.R., Mikolajewicz, N., Moffat, J., Singh, S.K. and McNicholas, P.D. (2021), 'Identification of five important genes to predict glioblastoma subtypes', *Neuro-Oncology Advances* **3**(1), vdab144.

[10] McNicholas, S.M., McNicholas, P.D. and Ashlock, D.A. (2021), 'An evolutionary algorithm with crossover and mutation for model-based clustering', *Journal of Classification* **38**(2), 264–279.

[11] **Roick, T.**, Karlis, D. and McNicholas, P.D. (2021), 'Clustering discrete valued time series', *Advances in Data Analysis and Classification* **15**(1), 209–229.

[12] **Tortora, C.**, **Browne, R.P.**, **ElSherbiny, A.**, **Franczak, B.C.** and McNicholas, P.D. (2021), 'Model-based clustering, classification, and discriminant analysis using the generalized hyperbolic distribution: MixGHD R package', *Journal of Statistical Software* **98**:3.

[13] **Subedi, S.** and McNicholas, P.D. (2021), 'A variational approximations-DIC rubric for parameter estimation and mixture model selection within a family setting', *Journal of Classification* **38**(1), 89–108.

[14] Mayhew, A.J., Phillips, S.M., Sohel, N., Thabane, L., McNicholas, P.D., de Souza, R.J., Parise, G. and Raina, P. (2021), 'Methodological issues and the impact of age stratification on the proportion of participants with low appendicular lean mass when adjusting for height and fat mass using linear regression: Results from the Canadian Longitudinal Study on Aging', *The Journal of Frailty and Aging* **10**, 150–155.

[15] Mayhew, A.J., Phillips, S.M., Sohel, N., Thabane, L., McNicholas, P.D., de Souza, R.J., Parise, G. and Raina, P. (2021), 'Do different ascertainment techniques identify the same individuals as sarcopenic in the Canadian Longitudinal Study on Aging?', *Journal of the American Geriatrics Society* **69**(1), 164–172.

[16] Mayhew, A.J., Phillips, S.M., Sohel, N., Thabane, L., McNicholas, P.D., de Souza, R.J., Parise, G. and Raina, P. (2020), 'The impact of different diagnostic criteria on the association of sarcopenia with injurious falls in the CLSA', *Journal of Cachexia, Sarcopenia and Muscle* **11**(6), 1603–1613.

[17] **Paton, F.** and McNicholas, P.D. (2020), 'Detecting British Columbia coastal rainfall patterns by clustering Gaussian processes', *Environmetrics* **31**(8), e2631.

[18] **Murray, P.M.**, **Browne, R.P.** and McNicholas, P.D. (2020), 'Mixtures of hidden truncation hyperbolic factor analyzers', *Journal of Classification* **37**(2), 366–379.

[19] **Gallaugher, M.P.B.** and McNicholas, P.D. (2020), 'Mixtures of skewed matrix variate bilinear factor analyzers', *Advances in Data Analysis and Classification* **14**(2), 415–434.

[20] **Pocuca, N.**, Jevtic, P., McNicholas, P.D. and Miljkovic, T. (2020), 'Modeling frequency and severity of claims with the zero-inflated generalized cluster-weighted models', *Insurance: Mathematics and Economics* **94**, 79–93.

[21] **Wei, Y.**, **Tang, Y.** and McNicholas, P.D. (2020), 'Flexible high-dimensional unsupervised learning with missing data', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**(3), 610–621.

[22] **Tortora, C.**, McNicholas, P.D. and Palumbo, F. (2020), 'A probabilistic distance clustering algorithm using Gaussian and Student-t multivariate density distributions', *SN Computer Science* **1**(2):65.

[23] Punzo, A., **Blostein, M.** and McNicholas, P.D. (2020), 'High-dimensional unsupervised classification via a contaminated mixture', *Pattern Recognition* **98**:107031.

[24] **Gallaugher, M.P.B.** and McNicholas, P.D. (2019), 'On fractionally-supervised classification: Weight selection and extension to the multivariate t-distribution', *Journal of Classification* **36**(2), 232–265.

[25] Turco, C.V., **Pesevski, A.**, McNicholas, P.D., Beaulieu, L.-D. and Nelson, A.J. (2019), 'Reliability of transcranial magnetic stimulation measures of afferent inhibition', *Brain Research* **1723**:146394.

[26] Silva, A., Rothstein, S.J., McNicholas, P.D., and **Subedi, S.** (2019), 'A multivariate Poisson-log normal mixture model for clustering transcriptome sequencing data', *BMC Bioinformatics* **20**:394.

[27] **Tortora, C.**, **Franczak, B.C.**, **Browne, R.P.**, McNicholas, P.D. (2019), 'A mixture of coalesced generalized hyperbolic distributions', *Journal of Classification* **36**(1), 26–57.

[28] **Murray, P.M.**, **Browne, R.P.** and McNicholas, P.D. (2019), Note of Clarification on 'Hidden truncation hyperbolic distributions, finite mixtures thereof, and their application for clustering, by Murray, Browne, and McNicholas, J. Multivariate Analysis 161 (2017) 141-156.', *Journal of Multivariate Analysis* **171**, 475-476.

[29] **Morris, K.**, Punzo, A., McNicholas, P.D. and **Browne, R.P.** (2019), 'Asymmetric clusters and outliers: Contaminated non-elliptical mixtures', *Computational Statistics and Data Analysis* **132**, 145–166.

[30] Mayhew, A.J., Amog, K., Phillips, S., Parise, G., McNicholas, P.D., de Souza, R.J., Thabane, L. and Raina, P. (2019), 'The prevalence of sarcopenia in community dwelling older adults, an exploration of differences between studies and within definitions: A systematic review and meta-analyses', *Age and Aging* **48**(1), 48–56.

[31] **Gallaugher, M.P.B.** and McNicholas, P.D. (2019), 'Three skewed matrix variate distributions', *Statistics and Probability Letters* **145**, 103–109.

[32] **Wei, Y.**, **Tang, Y.** and McNicholas, P.D. (2019), 'Mixtures of generalized hyperbolic distributions and mixtures of skew-t distributions for model-based clustering with incomplete data', *Computational Statistics and Data Analysis* **130**, 18–41.

[33] **Pesevski, A.**, **Franczak, B.C.** and McNicholas, P.D. (2018), 'Subspace clustering with the multivariate-t distribution', *Pattern Recognition Letters* **112**(1), 297–302.

[34] **Shaikh, M.R.**, Antonie, M.L., Murphy, T.B., and McNicholas, P.D. (2018), 'Standardizing interestingness measures for association rules', *Statistical Analysis and Data Mining* **11**(6), 282–295.

[35] Jones, A., Costa, A.P., **Pesevski, A.** and McNicholas, P.D. (2018), 'Predicting hospital and emergency department utilization among community-dwelling older adults: statistical and machine learning approaches', *PLOS ONE* **13**(11): e0206662.

[36] Morton, R.W., Sato, K., **Gallaugher, M.P.B.**, Oikawa, S.Y., McNicholas, P.D., Fujita, S. and Phillips, S.M. (2018), 'Muscle androgen receptor content but not systemic hormones is associated with resistance training-induced skeletal muscle hypertrophy in healthy, young men', *Frontiers in Physiology* **9**, 1373.

[37] Punzo, A., Mazza, A., and McNicholas, P.D. (2018), 'ContaminatedMixt: An R package for fitting parsimonious mixtures of multivariate contaminated normal distributions', *Journal of Statistical Software* **85**:10.

[38] **Gallaugher, M.P.B.** and McNicholas, P.D. (2018), 'Finite mixtures of skewed matrix variate distributions', *Pattern Recognition* **80**, 83–93.

[39] **Tang, Y.**, **Browne, R.P.** and McNicholas, P.D. (2018), 'Flexible clustering of high-dimensional data via mixtures of joint generalized hyperbolic distributions', *Stat* **7**(1), e177.

[40] **Andrews, J.L.**, Wickins, J.R., Boers, N.M. and McNicholas, P.D. (2018), 'teigen: An R package for model-based clustering and classification via the multivariate t distribution', *Journal of Statistical Software* **83**:7.

[41] Skinnider, M.A., Dejong, C.A., **Franczak, B.C.**, McNicholas, P.D. and Magarvey, N.A. (2017), 'Comparative analysis of chemical similarity methods for modular natural products with a hypothetical structure enumeration algorithm', *Journal of Cheminformatics* **9**:46.

[42] **Murray, P.M.**, **Browne, R.P.** and McNicholas, P.D. (2017), 'Hidden truncation hyperbolic distributions, finite mixtures thereof, and their application for clustering', *Journal of Multivariate Analysis* **161**, 141–156.

[43] Punzo, A. and McNicholas, P.D. (2017), 'Robust clustering in regression analysis via the contaminated Gaussian cluster-weighted model', *Journal of Classification* **34**(2), 249–293.

[44] **Murray, P.M.**, **Browne, R.P.** and McNicholas, P.D. (2017), 'A mixture of SDB skew-t factor analyzers', *Econometrics and Statistics* **3**, 160–168.

[45] **Gallaugher, M.P.B.** and McNicholas, P.D. (2017), 'A matrix variate skew-t distribution', *Stat* **6**(1), 160–170.

[46] **Wong, M.H.T.**, Mutch, D.M., and McNicholas, P.D. (2017), 'Two-way learning with one-way supervision for gene expression data', *BMC Bioinformatics* **18**:150.

[47] **Cheam, A.S.M.**, **Marbac, M.**, and McNicholas, P.D. (2017), 'Model-based clustering for spatio-temporal data on air quality monitoring', *Environmetrics* **28**(3), e2437.

[48] **Dang, U.J.**, Punzo, A., McNicholas, P.D., Ingrassia, S. and **Browne, R.P.** (2017), 'Multivariate response and parsimony for Gaussian cluster-weighted models', *Journal of Classification* **34**(1), 4–34.

[49] **Marbac, M.** and McNicholas, P.D. (2016), 'Dimension reduction in clustering', *Wiley StatsRef: Statistics Reference Online*.

[50] **Franczak, B.C.**, **Castura, J.C.**, **Browne, R.P.**, Findlay, C.J. and McNicholas, P.D. (2016), 'Handling missing data in consumer hedonic tests arising from direct scaling: Imputation techniques for consumer hedonic tests', *Journal of Sensory Studies* **31**(6), 514–523.

[51] **Tortora, C.**, McNicholas, P.D. and **Browne, R.P.** (2016), 'A mixture of generalized hyperbolic factor analyzers', *Advances in Data Analysis and Classification* **10**(4), 423–440.

[52] Punzo, A. and McNicholas, P.D. (2016), 'Parsimonious mixtures of multivariate contaminated normal distributions', *Biometrical Journal* **58**(6), 1506–1537.

[53] McNicholas, P.D. (2016), 'Model-based clustering', *Journal of Classification* **33**(3), 331–373.

[54] Punzo, A., **Browne, R.P.** and McNicholas, P.D. (2016), 'Hypothesis testing for mixture model selection', *Journal of Statistical Computation and Simulation* **86**(14), 2797–2818.

[55] Azzalini, A., **Browne, R.P.**, Genton, M.G. and McNicholas, P.D. (2016), 'On nomenclature for, and the relative merits of, two formulations of skew distributions', *Statistics and Probability Letters* **110**, 201–206.

[56] **Morris, K.** and McNicholas, P.D. (2016), 'Clustering, classification, discriminant analysis, and dimension reduction via generalized hyperbolic mixtures', *Computational Statistics and Data Analysis* **97**, 133–150.

[57] **Cheam, A.S.M.** and McNicholas, P.D. (2016), 'Modelling receiver operating characteristic curves using Gaussian mixtures', *Computational Statistics and Data Analysis* **93**, 192–206.

[58] O'Hagan, A., Murphy, T.B., Gormley, I.C., McNicholas, P.D., and Karlis, D. (2016), 'Clustering with the multivariate normal inverse Gaussian distribution', *Computational Statistics and Data Analysis* **93**, 18–30.

[59] **Dang, U.J.**, **Browne, R.P.** and McNicholas, P.D. (2015), 'Mixtures of multivariate power exponential distributions', *Biometrics* **71**(4), 1081–1089.

[60] **Vrbik, I.** and McNicholas, P.D. (2015), 'Fractionally-supervised classification', *Journal of Classification* **32**(3), 359–381.

[61] **Subedi, S.**, Punzo, A., Ingrassia, S. and McNicholas, P.D. (2015), 'Cluster-weighted t-factor analyzers for robust model-based clustering and dimension reduction', *Statistical Methods and Applications* (Journal of the Italian Statistical Society) **24**(4), 623–649.

[62] **Browne, R.P.** and McNicholas, P.D. (2015), 'Multivariate sharp quadratic bounds via Σ-strong convexity and the Fenchel connection', *Electronic Journal of Statistics* **9**(2), 1913–1938.

[63] **Wei, Y.** and McNicholas, P.D. (2015), 'Mixture model averaging for clustering', *Advances in Data Analysis and Classification* **9**(2), 197–217.

[64] **Browne, R.P.** and McNicholas, P.D. (2015), 'A mixture of generalized hyperbolic distributions', *Canadian Journal of Statistics* **43**(2), 176–198.

[65] **Franczak, B.C.**, **Browne, R.P.**, McNicholas, P.D. and Findlay, C.J. (2015), 'Product selection for liking studies: The sensory informed design', *Food Quality and Preference* **44**, 36–43.

[66] **Franczak, B.C.**, **Tortora, C.**, **Browne, R.P.** and McNicholas, P.D. (2015), 'Unsupervised learning via mixtures of skewed distributions with hypercube contours', *Pattern Recognition Letters* **58**(1), 69–76.

[67] **Tang, Y.**, **Browne, R.P.** and McNicholas, P.D. (2015), 'Model-based clustering of high-dimensional binary data', *Computational Statistics and Data Analysis* **87**, 84–101.

[68] Coneva, V., **Simopoulos, C.**, Casaretto, J.A., El-kereamy, A., Guevara, D.R., Cohn, J., Zhu, T., Guo, L., Alexander, D.C., Bi, Y.-M., McNicholas, P.D. and Rothstein, S.J. (2014), 'Metabolic and co-expression network-based analyses associated with nitrate response in rice', *BMC Genomics* **15**:1056.

[69] Ralston, J., Badoud, F., **Cattrysse, B.**, McNicholas, P.D. and Mutch, D.M. (2014), 'Inhibition of stearoyl-CoA desaturase-1 in differentiating 3T3-L1 pre-adipocytes up-regulates Elongase 6 and down-regulates genes affecting triacylglycerol synthesis', *International Journal of Obesity* **38** 1449–1456.

[70] **Misyura, M.**, Guevara, D., **Subedi, S.**, Hudson, D., McNicholas, P.D., Colasanti, J. and Rothstein, S.J. (2014), 'Nitrogen limitation and high density responses in rice suggest a role for ethylene in intraspecific competition', *BMC Genomics* **15**:681.

[71] **Andrews, J.L.** and McNicholas, P.D. (2014), 'Semi-supervised variable selection for clustering and classification', *Journal of Classification* **31**(2), 136–153.

[72] **Murray, P.M.**, **Browne, R.P.** and McNicholas, P.D. (2014), 'Mixtures of skew-t factor analyzers', *Computational Statistics and Data Analysis* **77**, 326–335.

[73] **Franczak, B.C.**, **Browne, R.P.** and McNicholas, P.D. (2014), 'Mixtures of shifted asymmetric Laplace distributions', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36**(6), 1149–1157.

[74] **Browne, R.P.** and McNicholas, P.D. (2014), 'Estimating common principal components in high dimensions', *Advances in Data Analysis and Classification* **8**(2), 217–226.

[75] **Subedi, S.** and McNicholas, P.D. (2014), 'Variational Bayes approximations for clustering via mixtures of normal inverse Gaussian distributions', *Advances in Data Analysis and Classification* **8**(2), 167–193.

[76] **Murray, P.M.**, McNicholas, P.D. and **Browne, R.P.** (2014), 'Mixtures of common skew-t factor analyzers', *Stat* **3**(1), 68–82.

[77] **Bhattacharya, S.** and McNicholas, P.D. (2014), 'A LASSO-penalized BIC for mixture model selection', *Advances in Data Analysis and Classification* **8**(1), 45–61.

[78] Lin, T.-I., McNicholas, P.D. and Ho, H.J. (2014), 'Capturing patterns via parsimonious t mixture models', *Statistics and Probability Letters* **88**, 80–87.

[79] **Browne, R.P.** and McNicholas, P.D. (2014), 'Orthogonal Stiefel manifold optimization for eigen-decomposed covariance parameter estimation in mixture models', *Statistics and Computing* **24**(2), 203–210.

[80] **Xia, Y.** and McNicholas, P.D. (2014). 'A gradient method for the monotone fused LASSO', *Optimization Methods and Software* **29**(3), 463–483.

[81] **Vrbik, I.** and McNicholas, P.D. (2014), 'Parsimonious skew mixture models for model-based clustering and classification', *Computational Statistics and Data Analysis* **71**, 196–210.

[82] **Morris, K.**, McNicholas, P.D. and Scrucca, L. (2013), 'Dimension reduction for model-based clustering via mixtures of multivariate t-distributions', *Advances in Data Analysis and Classification* **7**(3), 321–338.

[83] **Morris, K.** and McNicholas, P.D. (2013), 'Dimension reduction for model-based clustering via mixtures of shifted asymmetric Laplace distributions', *Statistics and Probability Letters* **83**(9), 2088–2093.

[84] **Liseron-Monfils, C.**, Lewis, T., Ashlock, D.A., McNicholas, P.D., Fauteux, F., Stromvik, M. and Raizada, M.N. (2013), 'A pipeline for discovery of co-regulatory motifs in maize and other plant species and its application to the anthocyanin and phlobaphene biosynthetic pathways and the Maize Development Atlas', *BMC Plant Biology* **13**:42.

[85] **Andrews, J.L.** and McNicholas, P.D. (2013), 'Using evolutionary algorithms for model-based clustering', *Pattern Recognition Letters* **34**(9), 987–992.

[86] **Humbert, S.**, **Subedi, S.**, Zeng, B., Bi, Y., Chen, X., Zhu, T., McNicholas, P.D. and Rothstein, S.J. (2013), 'Genome-wide expression profiling of maize in response to individual and combined water and nitrogen stresses', *BMC Genomics* **14**:3.

[87] **Subedi, S.**, Punzo, A., Ingrassia, S. and McNicholas, P.D. (2013), 'Clustering and classification via cluster-weighted factor analyzers', *Advances in Data Analysis and Classification* **7**(1) 5–40.

[88] **Wong, M.H.T.**, Holst, C., Astrup, A., Handjieva-Darlenska, T., Jebb, S.A., Kafatos, A., Kunesova, M., Larsen, T.M., Martinez, D.M., Pfeiffer, A.F.H., van Baak, M.A., Saris, W.H.M., McNicholas, P.D. and Mutch, D.M. (2012), 'Caloric restriction induces changes in insulin and body weight measurements that are inversely associated with subsequent weight regain', *PLOS ONE* **7**(8), e42858.

[89] **Zulyniak, M.A.**, Ralston, J.C., **Tucker, A.J.**, MacKay, K.A., Hillyer, L.M., McNicholas, P.D., Graham, T.E., Robinson, L.E., Duncan, A.M., Ma, D.W.L. and Mutch, D.M. (2012), 'Vaccenic acid in serum triglycerides is associated with markers of insulin resistance in men', *Applied Physiology, Nutrition, and Metabolism* **37**(5), 1003–1007.

[90] **Andrews, J.L.** and McNicholas, P.D. (2012), 'Model-based clustering, classification and discriminant analysis via mixtures of multivariate *t*-distributions', *Statistics and Computing* **22**(5), 1021–1029.

[91] **Browne, R.P.** and McNicholas, P. D. (2012), 'Model-based clustering and classification of data with mixed type', *Journal of Statistical Planning and Inference* **142**(11), 2976–2984.

[92] **Virbik, I.** and McNicholas, P.D. (2012), 'Analytic calculations for the EM algorithm for multivariate skew-t mixture models', *Statistics and Probability Letters* **82**(6), 1169–1174.

[93] McNicholas, P.D. and **Subedi, S.** (2012), 'Clustering gene expression time course data using mixtures of multivariate t-distributions', *Journal of Statistical Planning and Inference* **142**(5), 1114–1127.

[94] Feng, Z.Z., **Yang, X.**, **Subedi, S.** and McNicholas, P.D. (2012), 'The LASSO and sparse least squares regression methods for SNP selection in predicting quantitative traits', *IEEE Transactions on Computational Biology and Bioinformatics* **9**(2), 629–636.

[95] **Browne, R.P.**, McNicholas, P.D. and **Sparling, M.D.** (2012), 'Model-based learning using a mixture of mixtures of Gaussian and uniform distributions', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(4), 814–817.

[96] **Steane, M.A.**, McNicholas, P.D. and Yada, R. (2012), 'Model-based classification via mixtures of multivariate t-factor analyzers', *Communications in Statistics – Simulation and Computation* **41**(4), 510–523.

[97] **Andrews, J.L.** and McNicholas, P.D. (2011), 'Mixtures of modified *t*-factor analyzers for model-based clustering, classification, and discriminant analysis', *Journal of Statistical Planning and Inference* **141**(4), 1479–1486.

[98] McNicholas, P.D. (2011), 'On model-based clustering, classification, and discriminant analysis', *Journal of the Iranian Statistical Society* **10**(2) 181–199.

[99] **Andrews, J.L.** and McNicholas, P.D. (2011), 'Extending mixtures of multivariate t-factor analyzers', *Statistics and Computing* **21**(3), 361–373.

[100] **Xu, R.**, McNicholas, P.D., Desmond, A.F. and Darlington, G.A. (2011), 'A first passage time model for long term survivors with competing risks', *The International Journal of Biostatistics* **7**(1), Article 26.

[101] **Andrews, J.L.**, McNicholas, P.D. and **Subedi, S.** (2011), 'Model-based classification via mixtures of multivariate t-distributions', *Computational Statistics and Data Analysis* **55**(1), 520–529.

[102] Balka, J., Desmond, A.F. and McNicholas, P.D. (2011), 'Bayesian and likelihood inference for cure rates based on defective inverse Gaussian regression models', *Journal of Applied Statistics* **38**(1), 127–144.

[103] McNicholas, P.D. and Murphy, T.B. (2010), 'Model-based clustering of microarray expression data via latent Gaussian mixture models', *Bioinformatics* **26**(21), 2705–2712.

[104] **Shaikh, M.**, McNicholas, P.D. and Desmond, A.F. (2010), 'A pseudo-EM algorithm for clustering incomplete longitudinal data', *The International Journal of Biostatistics* **6**(1), Article 8.

[105] McNicholas, P.D. and Murphy, T.B. (2010), 'Model-based clustering of longitudinal data', *The Canadian Journal of Statistics* **38**(1), 153–168.

[106] McNicholas, P.D. (2010), 'Model-based classification using latent Gaussian mixture models', *Journal of Statistical Planning and Inference* **140**(5), 1175–1181.

[107] McNicholas, P.D., Murphy, T.B., McDaid, A.F. and Frost, D. (2010), 'Serial and parallel implementations of model-based clustering via parsimonious Gaussian mixture models', *Computational Statistics and Data Analysis* **54**(3), 711–723.

[108] Balka, J., Desmond, A.F. and McNicholas, P.D. (2009), 'Review and implementation of cure models based on first hitting times for Wiener processes', *Lifetime Data Analysis* **15**(2), 147–176.

[109] Fu, Y., Kim, L.-T. and McNicholas, P.D. (2009), 'Changes on enological parameters of white wine packaged in bag-in-box during secondary shelf life', *Journal of Food Science* **74**(8), C608–C618.

[110] McNicholas, P.D. and Murphy, T.B. (2008), 'Parsimonious Gaussian mixture models', *Statistics and Computing* **18**(3), 285–296.

[111] McNicholas, P.D., Murphy, T.B. and O'Regan, M. (2008), 'Standardising the lift of an association rule', *Computational Statistics and Data Analysis* **52**(10), 4712–4721.

[112] Ahmad, K., Rogers, S., McNicholas, P.D. and Collins P. (2007), 'Narrowband UVB and PUVA in the treatment of mycosis fungoides: A retrospective study', *Acta Dermato-Venereologica* **87**(5), 413–417.

[113] McNicholas, P.D. (2007), 'Association rule analysis of CAO data (with discussion)', *Journal of the Statistical and Social Inquiry Society of Ireland* **36**, 44–83.

Discussions of Peer-Reviewed Articles

[114] McNicholas, P.D., McNicholas, S.M. and **Tait, P.A.** (2018), Discussion of 'Statistical challenges of administrative and transaction data' by Hand, *Journal of the Royal Statistical Society: Series A* **181**(3), 594–595.

[115] **Gallaugher, M.P.B.** and McNicholas, P.D. (2017), Discussion of 'Random-projection ensemble classification' by Cannings and Samworth, *Journal of the Royal Statistical Society: Series B* **79**(4), 1011–1012

[116] McNicholas, P.D. and **Subedi, S.** (2016), Discussion of 'Perils and potentials of self-selected entry to epidemiological studies and surveys' by Keiding and Louis, *Journal of the Royal Statistical Society: Series A* **179**(2), 362–363.

[117] **Subedi, S.** and McNicholas, P.D. (2015), Discussion of 'Analysis of forensic DNA mixtures with artefacts' by Cowell et al., *Journal of the Royal Statistical Society: Series C* **64**(1), 43–44.

[118] McNicholas, P.D., **Browne, R.P.** and **Murray, P.M.** (2013), Discussion of 'Model-based clustering and classification with non-normal mixture distributions' by Lee and McLachlan, *Statistical Methods and Applications* **22**(4), 467–472.

[119] McNicholas, P.D. and **Browne, R.P.** (2013), Discussion of 'How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification' by Hennig and Liao, *Journal of the Royal Statistical Society: Series C* **62**(3), 352–353.

Peer-Reviewed Proceedings and Book Chapters

[120] **Gallaugher, M.P.B.** and McNicholas, P.D. (2020), 'Parsimonious mixtures of matrix variate bilinear factor analyzers' in T. Imaizumi et al. (eds.), *Advanced Studies in Behaviormetrics and Data Science: Essays in Honor of Akinori Okada*, Springer: Singapore, pp. 177–196.

[121] McNicholas, S.M. , McNicholas, P.D. and **Browne, R.P.** (2017), 'A mixture of variance-gamma factor analyzers'. In S.E. Ahmed (ed.), *Big and Complex Data Analysis: Methodology and Applications*. Cham: Springer International Publishing, pp. 369–385.

[122] **Dang, U.J.** and McNicholas, P.D. (2015), 'Families of parsimonious finite mixtures of regression models'. In Morlini I., Minerva T., and Vichi, M. (eds.), *Advances in Statistical Models for Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization. Switzerland: Springer International Publishing, pp. 73–84.

[123] McNicholas, P.D. (2013), 'Model-based clustering and classification via mixtures of multivariate t-distributions'. In Giudici, P., Ingrassia, S. and Vichi, M. (eds.), *Statistical Models for Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization. Switzerland: Springer International Publishing, pp. 233–240.

[124] **Browne, R.P.** and McNicholas, P.D. (2013), 'Mixture and latent class models in the longitudinal and other settings'. In M. A. Scott, J. S. Simonoff, and B. D. Marx (eds.), *SAGE Handbook of Multilevel Modelling*, SAGE Publications Ltd., pp. 357–370.

[125] Ashlock, D., Schonfeld, J. and McNicholas, P.D. (2011), 'Translation tables: A genetic code in a evolutionary algorithm'. In *IEEE Congress in Evolutionary Computation*, New Orleans, pp. 2685–2692.

[126] McNicholas, P.D. and Zhao, Y. C. (2009), 'Association rules: An overview'. In Y. Zhao, C. Zhang, and L. Cao, eds., *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*, IGI Global, pp. 1–10.

Non-Peer-Reviewed Article (Opinion)

[127] McNicholas P.D. (2016). 'Turning the spit: A perspective on the NSERC Discovery Grant review process'. *Liaison* **30**(4), 45–55.

Non-Peer-Reviewed Proceeding

[128] **Gallaugher, M.P.B.** and McNicholas, P.D. (2018), 'Mixtures of matrix variate bilinear factor analyzers' in *Proceedings of the Joint Statistical Meetings*, American Statistical Association, Alexandria, VA.

Editorials

[129] McNicholas, P.D. (2021), 'Editorial: Journal of Classification Vol. 38-3', *Journal of Classification* **38**(3), 423–424.

[130] McNicholas, P.D. (2021), 'Editorial: Journal of Classification Vol. 38-2', *Journal of Classification* **38**(2), 187.

[131] McNicholas, P.D. (2021), 'Editorial: Journal of Classification Vol. 38-1', *Journal of Classification* **38**(1), 1.

[132] McNicholas, P.D. (2020), 'Editorial: Journal of Classification Vol. 37-3', *Journal of Classification* **37**(3), 549.

[133] McNicholas, P.D. and Steinley, D.L. (2020), 'Editorial: Journal of Classification Vol. 37-2', *Journal of Classification* **37**(2), 275–276.

[134] McNicholas, P.D. (2019), 'Data science', *FACETS* **4**(1), 131–135.

[135] Einbeck, J., Hinde, J., Ingrassia, S., Lin, T.-I. and McNicholas, P.D. (2019), 'Editorial for the 4th Special Issue on advances in mixture models', *Computational Statistics and Data Analysis* **132**, 143–144.

[136] Kestler, H.A., McNicholas, P.D. and Wilhelm, A.F.X. (2018), 'Special issue on "Science of big data: theory, methods and applications"', *Advances in Data Analysis and Classification* **12**(4), 823–825.

[137] Hinde, J., Ingrassia, S., Lin, T.-I. and McNicholas, P. (2016), 'The third special issue on advances in mixture models', *Computational Statistics and Data Analysis* **93**, 2–4.

[138] Böhning, D., Hennig, C., McLachlan, G.J., McNicholas, P.D. (2014), 'The 2nd special issue on advances in mixture models', *Computational Statistics and Data Analysis* **71**, 1–2.

R and Julia Software Packages

Note: R and Julia are free software environment for statistical computing and graphics. All R packages are on CRAN.

i. **Počuča, N.**, **Browne, R.P.**, and McNicholas, P.D. (2021). mixture: Mixture models for clustering and classification. R package version 2.0.4.

ii. **Browne, R.P.**, **Dang, U.J.**, **Gallaugher, M.P.B.** and McNicholas, P.D. (2021), mixSPE: Mixtures of power exponential and skew power exponential distributions for use in model-based clustering and classification. R package version 0.9.1.

iii. **Tortora, C.**, **ElSherbiny A.**, **Browne, R.P.**, **Franczak, B.C.**, McNicholas, P.D. and Amos, D.D. (2020). MixGHD: Model based clustering, classification and discriminant analysis using the mixture of generalized hyperbolic distributions. R package version 2.3.4.

iv. **Tortora, C.**, Vidales, N., Palumbo, F. and McNicholas, P.D. (2020). FPDclustering: PD-clustering and factor PD-clustering. R package version 1.4.1.

v. Punzo, A., Mazza, A. and McNicholas, P.D. (2020). ContaminatedMixt: Model-based clustering and classification with the multivariate contaminated normal distribution. R package version 1.3.4.1.

vi. McNicholas, P.D., **ElSherbiny, A.**, **Jampani, K.R.**, McDaid, A.F., Murphy, T.B. and Banks, L. (2019). pgmm: Parsimonious Gaussian mixture models. R package version 1.2.4.

vii. **Počuča, N.**, **Gallaugher, M.P.B.** and McNicholas, P.D. (2019), MatrixVariate.jl: A complete statistical framework for analyzing matrix variate data. Julia package version 0.2.0. URL: `github.com/nikpocuca`

viii. **Clark, K.M.** and McNicholas, P.D. (2019), oclust: Gaussian model-based clustering with outliers. R package version 0.1.0.

ix. **Gallaugher, M.P.B.** and McNicholas, P.D. (2019), ClickClustCont: Mixtures of continuous time Markov models. R package version 0.1.7.

x. McNicholas, P.D., **Jampani, K.R.**, **Subedi, S.** (2019). longclust: Clustering longitudinal data. R package version 1.2.3.

xi. **Andrews, J.L.**, Wickins, J.R., Boers, N.M. and McNicholas, P.D. (2018). teigen: Model-based clustering and classification with the multivariate t-distribution. R package version 2.2.2.

xii. **Franczak, B.C.**, **Browne, R.P.** and McNicholas, P.D. (2016). sensory: Simultaneous model-based clustering and imputation via a progressive expectation-maximization algorithm. R package version 1.1.

xiii. **Andrews, J.L.** and McNicholas, P.D. (2013). vscc: Variable selection for clustering and classification. R package version 0.2.

xiv. **Athey, T.B.T.** and McNicholas, P.D. (2013). VLF: Frequency matrix approach for assessing very low frequency variants in sequence records. R package version 1.0.

Research Support Held as Principal Investigator

Agency, Program	Total	Year(s)
MITACS, Accelerate	$60,000	2020–21
NSERC, E.W.R. Steacie Memorial Fellowship	$250,000	2019–21
NSERC, Engage	$25,000	2018–19
NSERC, Discovery Grant	$215,000	2017–22
Ontario Research Fund, Small Infrastructure	$149,999	2016
Canada Foundation for Innovation, John R. Evans Leaders Fund	$149,999	2015
Canada Research Chair, Tier 1	$1,400,000	2015–22
NSERC, Collaborative Research & Development Grant	$400,000	2012–16
Industry support (Grant-In-Aid)	$200,000	2012–16
NSERC, Discovery Grant	$125,000	2012–17
NSERC, Research Tools & Instruments	$36,500	2011
NSERC, Discovery Grant	$13,000	2011–12*
Ontario Ministry for Research & Innovation, Early Researcher Award	$150,000	2011–14
University of Guelph, University Research Chair in Computational Statistics	$440,000	2011–14
NSERC, Engage	$25,000	2010
NSERC, Collaborative Research & Development	$220,000	2009–12
Ontario Centres of Excellence, Collaborative Research	$100,000	2009–11
Industry support (Grant-In-Aid)	$150,000	2009–12
Canada Foundation for Innovation, Leaders Opportunity Fund	$118,362	2009
Ontario Ministry for Research & Innovation, Small Infrastructure Fund	$118,362	2009
MITACS, Accelerate	$15,000	2009
NSERC, Discovery Grant	$36,000	2008–11

*Originally awarded for 5 years but shortened on appeal to facilitate re-application.

Research Support Held as Co-Investigator

Agency, Program	Total	Year(s)
CIHR, Project Grant	$1,225,000	2022–2027
Weston Family Foundation, Grant	$12,240,322	2021–2028
SSHRC, Partnership Grant	$2,499,863	2020–26
AGE-WELL Core Research Program, Grant	$600,000	2020
AGE-WELL Core Research Program, Catalyst Grant	$30,000	2020
Labarge Centre for Mobility in Aging, McMaster University	$5,000,000	2019–24
MIRA, McMaster University	$1,000,000	2019–24
Strategic Alignment Fund, McMaster University	$501,357	2019–22

Invited Presentations at Conferences and Meetings

i. *Using Subset Log-Likelihoods to Trim Outliers in Gaussian Mixture Models*, Classification and Data Analysis Group of the Italian Statistical Society, Firenze, Italy, September 2021 (virtual).

ii. *Flexible Clustering of High-Dimensional Data via Mixtures of Joint Generalized Hyperbolic Distributions*, Joint Statistical Meetings, Seattle, WA, August 2021 (virtual).

iii. *Outlier Detection in Model-Based Clustering*, Statistics 2021, Ottawa, July 2021 (virtual).

iv. *Thoughts on Some Problems in Clustering*, Isobel Loutit Lecture (2020), Statistical Society of Canada Annual Meeting, Ottawa, June 2021 (virtual).

v. *Selected Problems in Classification*, Data Science Applied Research and Education Seminar, CANSSI Ontario, Toronto, February 2021 (virtual).

vi. *Data Science, Classification and Three-Way Data*, Colloque des sciences mathématiques du Québec, CRM, Montreal, October 2020 (virtual).

vii. *Mixtures of Skewed Matrix Variate Bilinear Factor Analyzers*, MBC2: An International Workshop on Model-Based Clustering and Classification, Catania, Italy, September 2020 (virtual).

viii. *Using Subset Log-Likelihoods to Predict the Number of Outliers in Gaussian Mixture Models*, AMS Fall Eastern Sectional Meeting, Binghamton, NY, October 2019.

ix. *Contributions to Model Selection and Cluster Analysis*, Special Session in Honour of Prof. Stan Sclove, The Classification Society Annual Meeting, Edmonton, AB, June 2019.

x. *Mixtures of Matrix Variate Bilinear Factor Analyzers*, MBC2: An International Workshop on Model-Based Clustering and Classification, Catania, Italy, September 2018.

xi. *Clustering Longitudinal Data Using Mixture Models*, Joint Statistical Meetings, Vancouver, BC, August 2018.

xii. *Mixtures of Skewed Matrix Variate Distributions*, University of Washington Working Group on Model-Based Clustering, Ann Arbor, MI, July 2018.

xiii. *Selected Problems in Classification* (President's Address), The Classification Society Annual Meeting, Long Island, NY, June 2018.

xiv. *Securing Funding as a New Investigator* (part of a panel discussion), Statistical Society of Canada Annual Meeting, Montreal, QC, June 2018.

xv. *Clustering Longitudinal Data*, ENAR Spring Meeting, Atlanta, GA, March 2018.

xvi. *On Clustering Longitudinal Data*, International Federation of Classification Societies Meeting, Tokyo, Japan, August 2017.

xvii. *Model-Based Clustering of Big Data*, Joint Statistical Meetings, Baltimore, MD, July 2017 (topic contributed).

xviii. *Some Experiences From an Industry Collaboration*, Statistical Society of Canada Annual Meeting, Winnipeg, MB, June 2017.

xix. *Mixtures of Coalesced Generalized Hyperbolic Distributions*, International Conference on Statistical Distributions and Applications, Niagara Falls, ON, October 2016.

xx. *A Cluster, Clustering and Mixture Models*, The Classification Society Annual Meeting, Columbia, MO, June 2016.

xxi. *Clustering Ultra High-Dimensional Data*, Statistical Society of Canada Annual Meeting, St. Catherines, ON, May 2016.

xxii. *Implicit Versus Explicit Variable Selection*, European Research Consortium for Informatics and Mathematics (ERCIM) Working Group on Computing & Statistics, London, England, December 2015.

xxiii. *Averaging and Asymmetry in Cluster Analysis*, International Federation of Classification Societies Meeting, Bologna, Italy, July 2015.

xxiv. *Outlier Detection via Contaminated Mixture Distributions*, Classification and Data Analysis Group of the Italian Statistical Society, Modena, Italy, September 2013.

xxv. *Classification via Mixtures of SAL and Generalized Hyperbolic Distributions*, International Federation of Classification Societies Meeting, Tilburg, The Netherlands, July 2013.

xxvi. *Recent Work on Non-Gaussian Methods*, MBC^2: Workshop on Model Based Clustering and Classification, Catania, Sicily, Italy, September 2012. Keynote speaker.

xxvii. *Mixtures of Generalized Hyperbolic Distributions*, University of Washington Working Group on Model-Based Clustering, ON, Canada, July 2012.

xxviii. *Clustering and Classification of High-dimensional Data via Modified t-factor Analyzers*, ERCIM, London, England, December 2011.

xxix. *Model-based Clustering and Classification via Mixtures of Multivariate t-Distributions*, Classification and Data Analysis Group of the Italian Statistical Society, Pavia, Italy, September 2011.

xxx. *Model-Based Clustering: An Overview*, Sensometrics 10, Rotterdam, The Netherlands, July 2010.

xxxi. *Model-Based Clustering via Mixtures of Multivariate t-Distributions*, University of Washington Working Group on Model-Based Clustering, INRIA, Grenoble, France, July 2010.

xxxii. *Model-Based Clustering of Longitudinal Data*, University of Washington Working Group on Model-Based Clustering, Seattle, WA, July 2008.

xxxiii. *Parsimonious Gaussian Mixture Models*, Royal Statistical Society Conference, Belfast, Northern Ireland, September 2006.

Selected Seminars

i. *Selected Problems in Classification*, Lakehead University, ON, December 2021 (virtual).

ii. *Using Subset Log-Likelihoods to Trim Outliers in Gaussian Mixture Models*, Wirtschaftsuniversität Wien, Vienna University of Economics and Business, Austria, October 2021 (virtual).

iii. *Clustering Higher-Order Data*, Trinity College Dublin, Ireland, October 2019.

iv. *Clustering Higher-Order Data*, Binghamton University, NY, October 2019.

v. *Data Science, Classification and Three-Way Data*, University of Windsor, ON, April 2019.

vi. *Model-Based Clustering of Matrix Variate and Tensor Variate Data*, University of Alberta, AB, March 2019.

vii. *Data Science, Classification and Clustering*, MacEwan University, AB, March 2019.

viii. *Clustering, Classification and Data Science*, University College Dublin, Ireland, October 2018.

ix. *Clustering, Classification and Data Science*, Queens University Belfast, UK, October 2018.

x. *Experiences From an Industry Collaboration*, University of Toronto, ON, September 2018.

xi. *Clustering Three-Way Data Using Mixture Models*, University of Rochester, NY, March 2018.

xii. *Cluster-Weighted Models and Machine Learning for Insurance Data*, Arizona Actuarial Club, AZ, October 2017.

xiii. *Clustering via Mixture Models*, Arizona State University, AZ, October 2017.

xiv. *Clustering via Mixture Models*, Dalhousie University, NS, October 2017.

xv. *Clustering via Mixture Models*, Wharton School, University of Pennsylvania, PA, April 2017.

xvi. *Clustering and Discriminant Analysis via Mixture Models*, Acadia University, NS, March 2016.

xvii. *Model-Based Clustering: Past, Present and Future*, Simon Fraser University, BC, March 2016.

xviii. *Averaging and Asymmetry in Cluster Analysis*, University of Manitoba, MB, September 2015.

xix. *Contaminated Mixtures and Fractionally-Supervised Classification*, Wilfred Laurier University, ON, October 2013.

xx. *Non-Gaussian Model-Based Clustering & Classification*, University of Toronto, ON, November 2011.

xxi. *Non-Gaussian Model-Based Clustering & Classification*, University of Waterloo, ON, October 2011.

xxii. *Model-Based Clustering, Classification and Discriminant Analysis via Mixtures of Multivariate t-Distributions*, University of Ottawa (joint with Carleton University), ON, November 2010.

xxiii. *Recent Work in Model-Based Clustering & Classification*, University of Windsor, ON, March 2010.

xxiv. *Recent Work in Model-Based Clustering & Classification*, University of Waterloo, ON, January 2010.

xxv. *Recent Work in Model-Based Clustering & Classification*, University of Toronto, ON, December 2009.

xxvi. *On Clustering Gene Expression Time Course Data*, University of Western Ontario, ON, October 2009.

xxvii. *Clustering & Data Mining*, Universite de Montreal, QC, January 2009.

xxviii. *Model-Based Clustering: An Overview*, McGill University, Montreal, QC, October 2008.

xxix. *Model-Based Clustering of Longitudinal Data*, Colloque de statistique de Montreal, McGill University, October 2008.

xxx. *Analysis of Gene Expression Time Course Data via Model-Based Clustering*, York University, ON, October 2008.

xxxi. *Model-Based Clustering: An Overview*, National University of Ireland, Galway, April 2008.

xxxii. *Standardizing the Lift of an Association Rule*, University College Dublin, Ireland, November 2007.

xxxiii. *Model-Based Clustering: An Overview*, McMaster University, ON, October 2007.

Training: Quality and Impact

- Former Ph.D. and postdoctoral trainees currently continuing pertinent work as faculty:
  - Jeffrey Andrews (Ph.D., 2012), Associate Professor, University of British Columbia – Okanagan, BC.
  - Ryan Browne (PDF, 2011), Associate Professor, University of Waterloo, ON.
  - Utkarsh Dang (Ph.D., 2014), Assistant Professor, Carleton University, ON.
  - Brian Franczak (Ph.D., 2014), Associate Professor, MacEwan University, AB.
  - Michael Gallaugher (Ph.D., 2020), Assistant Professor, Baylor University, TX.
  - Matthieu Marbac (PDF, 2016), Associate Professor, Ecole nationale de la statistique et de l'analyse de l'information (Campus de Ker Lann), France.
  - Orla Murphy (PDF, 2021), Assistant Professor, Dalhousie University, NS.
  - Nkumbuludzi Ndwapi (Ph.D., 2017), Lecturer, Botswana International University of Science and Technology.
  - Mateen Shaikh (Ph.D., 2013), Assistant Professor, Thompson Rivers University, BC.
  - Sanjeena Subedi (Ph.D., 2012), Assistant Professor and Canada Research Chair, Carleton University, ON.
  - Cristina Tortora (PDF, 2016), Associate Professor, San Jose State University, CA.
  - Irene Vrbik (Ph.D., 2014), Assistant Professor, University of British Columbia – Okanagan, BC.
- Ph.D. students who won the Distinguished Dissertation Award from The Classification Society:
  - Jeffrey Andrews (Ph.D., 2012), winner 2013.
  - Michael Gallaugher (Ph.D., 2020), winner 2021.
  - Paula Murray (Ph.D., 2016), winner 2017.
  - Yang Tang (Ph.D., 2017), winner 2018.
  - Irene Vrbik (Ph.D., 2014), winner 2015.
- Former Ph.D. and postdoctoral trainees who later won the prestigious Chikio Hayashi Award (for excellent researchers ≤ 35 years old), given biennially by the International Federation of Classification Societies:
  - Jeffrey Andrews (Ph.D., 2012), winner 2017.
  - Ryan Browne (PDF, 2011), winner 2015.
  - Brian Franczak (Ph.D., 2014), winner 2019.
  - Sanjeena Subedi (Ph.D., 2012), winner 2019.
  - Cristina Tortora (PDF, 2016), winner 2019.
- Summary of trainees.

	Current		**Previous**	
	Supervised	**Co-Supervised**	**Supervised**	**Co-Supervised**
Postdoctroal	1	0	14	0
Doctoral	7	1	14	5
Master's	1	0	33	20
Undergraduate	1	0	28	4
Total	10	1	89	29

Selected External Service and Leadership

- Editorial activities:
  - Associate Editor, *The Canadian Journal of Statistics*, 2013–2016 and 2022–Present.
  - Editor-in-Chief, *Journal of Classification*, 2020–Present.
  - Member, Editorial Board, *Journal of Classification*, 2016–Present.
  - Associate Editor, *Journal of Multivariate Analysis*, 2019–2020.
  - Data Science Section Editor, *FACETS*, 2018–2020.
  - Associate Editor, *Journal of Classification*, 2018–2020.
  - Associate Editor, *Advances in Data Analysis and Classification*, 2014–2020.
  - Associate Editor, *Environmetrics*, 2014–2019.
  - Guest Editor, *Computational Statistics and Data Analysis* Fourth Special Issue on "Advances in Mixture Models", 2017–2018.
  - Guest Editor, *Econometrics and Statistics* Special Issue on "Advances in Mixture Models", 2015–2017.
  - Guest Editor, *Advances in Data Analysis and Classification*, Special Issue on the "Science of Big Data: Theory, Methods and Applications", 2014–2018.
  - Associate Editor, *Econometrics and Statistics*, 2015–2018.
  - Guest Editor, *Computational Statistics and Data Analysis* Third Special Issue on "Advances in Mixture Models", 2013–2015.
  - Statistical Advisor, BMC-Series Editorial Board, 2013–2016.
  - Associate Editor, *Computational Statistics and Data Analysis*, 2013–2018.
  - Associate Editor, *Statistical Methods and Applications* (Journal of the Italian Statistical Society), 2013–2017.
  - Associate Editor, *Journal of the Royal Statistical Society: Series C*, 2012–2015.
  - Guest Editor, *Computational Statistics and Data Analysis* Second Special Issue on "Advances in Mixture Models", 2011–2013.
- Canadian Statistical Sciences Institute
  - Chair, Data Science Committee, 2018–2019.
  - Associate Director, 2018–2019.
  - Member, Industrial Innovation Committee, 2015–2017.
- Statistical Society of Canada
  - Research Committee: Chair, 2017–2019. Member, 2016–2017.
  - President-Elect/President/Past-President, Business and Industrial Statistics Section, 2013–2016.
  - Local Representative, University of Guelph, 2012–2014.
  - Member, *The Canadian Journal of Statistics* Awards Committee, 2011–2014.
  - Member, Canadian Statistical Institute Development Committee, 2011–2012.
- The Classification Society
  - President, 2018–2019. President-Elect, 2016–2017. Past-President, 2020–2021.
  - Chair, Program Committee, Annual Meeting, Edmonton, Alberta, June 2019.
  - Chair, Program Committee, Annual Meeting, Long Island, New York, June 2018.
  - Member, Program Committee, Annual Meeting, Columbia, Missouri, June 2016.
  - Chair, Organizing Committee (Program and Local Arrangements), Annual Meeting, Hamilton, Ontario, Canada, June 2015.
  - Member, Board of Directors, 2014–Present.
  - Member, Editor-In-Chief Search Committee, *Journal of Classification*, 2014.
- International Federation of Classification Societies (IFCS)
  - Member, Scientific Program Committee, IFCS 2019, Thessaloniki, Greece, August 2019.
  - Member, Scientific Program Committee, IFCS 2017, Tokyo, Japan, August 2017.
  - Member, Scientific Program Committee, IFCS 2015, Bologna, Italy, July 2015.
  - Publication Officer (and Executive Committee Member), 2012–2018.
- Selected Research Service: Provincial, Federal and International
  - Information and Communications Technology, Math and Physics Panel, Ontario Early Researcher Awards.
    - Chair: 2017, 2018, 2020, 2021.
    - Member: 2014, 2015, 2016.
  - NSERC Evaluation Group 1508 (Mathematics and Statistics), 2013/14–2015/16.
    - Statistics Section: Chair, 2014/15, 2015/16. Incoming Chair, 2013/14.

        - Executive Committee: Member, 2013/14–2015/16.
    - Reviewer, Canada Research Chairs program (multiple).
    - NSERC Discovery Grants proposal reviewer (many).
- Member, Highly Qualified Personnel Provincial Advisory Committee, Compute Ontario, 2017–2018.
- Poster Chair, Joint Statistical Meetings 2018 (Vancouver, Canada), 2017–2018.
- External University Tenure and Promotion Reviewing
    - Carleton University, ON.
    - Dalhousie University, NS.
    - Rutgers University, NJ (multiple).
    - South Dakota State University, SD.
    - University of Alabama, AL (multiple).
    - University of Auckland, New Zealand.
    - University of Manitoba, MB.
    - University of Massachusetts Amherst, MA.
    - University of Ottawa, ON.
    - University of Pittsburgh, PA.
    - University of Rochester, NY.
- Habilitation External Examiner
    - University of Lille, France, 2021.
- Ph.D. External Examiner
    - McGill University, QC, 2013, 2021.
    - McMaster University, ON, 2013.
    - University of Milano-Bicocca, Italy, 2018.
    - University of Pavia, Italy, 2021.
    - University of Waterloo, ON, 2013, 2019.
    - University of Western Ontario, ON, 2010, 2020.

Internal (McMaster University) Service Contributions

Semesters	Role, Committee	Level
F21–	Member, Graduate Council	University
F21	Member, Internal Review Panel for CFI Innovation Fund Competition	University
S21–	Member, Data Science in Actuarial and Financial Math Search Committee	Department
S21–	Member, Graduate Committee	Department
S21–F21	Member, RHPCS Director Search Committee	University
F18–S19	Member, Ad Hoc Committee on Data Science	Department
F18–W19	Member, Probability and Statistics Search Committee	Department
F18	Member, Science Research Chairs Committee	Faculty
S17	Member, Canada 150 Research Chairs Committee for Science	Faculty
S17–S18	Member, Research Technology Committee	University
S17–S18	Member, Sherman Centre Advisory Committee	University
F16	Member, School of Graduate Studies Scholarship Committee (NSERC doctoral)	University
S16–S17	Member, Employment Equity Working Committee	University
F15–S16	Member, Senate Executive Committee	University
W16	Member, Internal Review Panel for CFI Innovation Fund Competition	University
W16	Member, Associate Dean Research Search Committee	Faculty
F15	Member, Valedictorian Selection Committee	University
S15–S16	Member, Committee on University Ceremonials and Insignia	University
S15–S16	Member, Senate	University
S15–S18	Member, Awards Committee	Department
S14–S18	Member, Graduate Committee	Department

Internal (University of Guelph) Service Contributions

Semesters	Role, Committee	Level
F12–S14	Member, (Graduate) Admissions and Progress Committee	University
F12–S13	Member, Senate Board of Graduate Studies	University
W12–S14	Chair, Bioinformatics Program Committee & Graduate Coordinator	College
F10, F11, F12	Member, Graduate Awards Committee (NSERC)	University
F10, F11, F13	Member, Promotion and Tenure Committee	Department
F10–S13	Member, University Senate	University
S09–S11	Chair, Graduate Studies Committee & Graduate Coordinator	Department
W08–S10	Member, Appointments Committee	Department

Teaching Summary

- McMaster University (2014–Present):
  - CSE 799: Special Topics (multiple)
  - HTHSCI 4B06: Thesis
  - ISCI 3A12: Integrated Science III (thesis)
  - ISCI 4A12: Integrated Science IV (thesis)
  - MATH 4P06: Senior Research Project (multiple)
  - STATS 4I03/6I03: Inference
  - STATS 4M03/STATS 6M03: Multivariate Analysis
  - STATS 4T06: Senior Thesis (multiple)
  - STATS 4W03: Directed Reading (multiple)
  - STATS 743A: Foundations of Statistics
  - STATS 780/CSE 780: Data Science
  - STATS 794: Directed Reading (multiple)
- University of Guelph (2007–2014):
  - BINF*6970: Statistical Bioinformatics (reading, multiple)
  - MATH*1000: Introduction to Calculus I
  - MATH*1000 DE: Introduction to Calculus I (distance, multiple)
  - STAT*2120: Probability & Statistics for Engineers (multiple)
  - STAT*4050: Generalized Linear Models (reading, multiple)
  - STAT*4340: Statistical Inference (multiple)
  - STAT*4350/6821: Multivariate Statistical Analysis (reading, multiple)
  - STAT*6821: Multivariate Statistical Analysis (reading, multiple)
  - STAT*6841: Statistical Inference (multiple)
  - STAT*6920: Applied Statistical Methods (reading)
  - STAT*6920: Finite Mixture Models (reading, multiple)
  - STAT*6920: Statistics in the 21st Century (reading)