ISSN 1932-6157 (print)
ISSN 1941-7330 (online)

THE ANNALS *of* APPLIED STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

continued

Vol. 14, No. 3—September 2020

ISSN 1932-6157 (print)
ISSN 1941-7330 (online)

THE ANNALS
of
APPLIED STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles—*Continued from front cover*

Vol. 14, No. 3—September 2020

THE ANNALS OF APPLIED STATISTICS Vol. 14, No. 3, pp. 1063–1580 September 2020

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

The Annals of Applied Statistics **[ISSN 1932-6157 (print); ISSN 1941-7330 (online)],** Volume 14, Number 3, September 2020. Published quarterly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, Ohio 44122, USA. Periodicals postage pending at Cleveland, Ohio, and at additional mailing offices.

POSTMASTER: Send address changes to *The Annals of Applied Statistics*, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1063–1087
https://doi.org/10.1214/20-AOAS1336

STATISTICAL METHODS FOR REPLICABILITY ASSESSMENT

BY KENNETH HUNG[1] AND WILLIAM FITHIAN[2]

[1]*Department of Mathematics, University of California, Berkeley, kenhung@berkeley.edu*
[2]*Department of Statistics, University of California, Berkeley, wfithian@berkeley.edu*

Large-scale replication studies like the Reproducibility Project: Psychology (RP:P) provide invaluable systematic data on scientific replicability, but most analyses and interpretations of the data fail to agree on the definition of "replicability" and disentangle the inexorable consequences of known selection bias from competing explanations. We discuss three concrete definitions of replicability based on: (1) whether published findings about the signs of effects are mostly correct, (2) how effective replication studies are in reproducing whatever true effect size was present in the original experiment and (3) whether true effect sizes tend to diminish in replication. We apply techniques from multiple testing and postselection inference to develop new methods that answer these questions while explicitly accounting for selection bias. Our analyses suggest that the RP:P dataset is largely consistent with publication bias due to selection of significant effects. The methods in this paper make no distributional assumptions about the true effect sizes.

REFERENCES

ACHENBACH, J. (2015). Many scientific studies can't be replicated. That's a problem. *The Washington Post*.

AMRHEIN, V., KORNER-NIEVERGELT, F. and ROTH, T. (2017). The Earth is flat ($p > 0.05$): Significance thresholds and the crisis of unreplicable research. *PeerJ* **5** e3544. https://doi.org/10.7717/peerj.3544

ANDERSON, C. J., BAHNÍK, Š., BARNETT-COWAN, M., BOSCO, F. A., CHANDLER, J., CHARTIER, C. R., CHEUNG, F., CHRISTOPHERSON, C. D., CORDES, A. et al. (2016). Response to comment on "Estimating the reproducibility of psychological science". *Science* **351** 1037c.

ANDREWS, I. and KASY, M. (2018). Identification of and correction for publication bias. GitHub 1–85.

BAKER, M. (2015). Over half of psychology studies fail reproducibility test.

BARRETT, L. F. (2015). Psychology is not in crisis. *The New York Times* A23.

BENJAMIN, D. J., BERGER, J. O., JOHANNESSON, M., NOSEK, B. A., WAGENMAKERS, E.-J., BERK, R. A., BOLLEN, K. A., BREMBS, B., BROWN, L. et al. (2018). Redefine statistical significance. *Nat. Hum. Behav.* **2** 6–10.

BENJAMINI, Y. and HELLER, R. (2008). Screening for partial conjunction hypotheses. *Biometrics* **64** 1215–1222. MR2522270 https://doi.org/10.1111/j.1541-0420.2007.00984.x

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

BENJAMINI, Y. and HOCHBERG, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Stat.* **25** 60–83.

BENJAMINI, Y. and YEKUTIELI, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *J. Amer. Statist. Assoc.* **100** 71–93. With comments and a rejoinder by the authors. MR2156820 https://doi.org/10.1198/016214504000001907

CAMERER, C. F., DREBER, A., FORSELL, E., HO, T.-H., HUBER, J., JOHANNESSON, M., KIRCHLER, M., ALMENBERG, J., ALTMEJD, A. et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science* **351** 1433–1436.

CAMERER, C. F., DREBER, A., HOLZMEISTER, F., HO, T.-H., HUBER, J., JOHANNESSON, M., KIRCHLER, M., NAVE, G., NOSEK, B. A. et al. (2018). Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nat. Hum. Behav.* **343** 229–268.

CAREY, B. (2015). Many psychology findings not as strong as claimed, study says. *The New York Times* A1.

DODSON, C. S., DARRAGH, J. and WILLIAMS, A. (2008). Stereotypes and retrieval-provoked illusory source recollections. *J. Exper. Psychol., Learn., Mem., Cogn.* **34** 460–477.

Key words and phrases. Replicability, multiple testing, postselection inference, publication bias, meta-analysis.

DUVAL, S. and TWEEDIE, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56** 455–463.

ETZ, A. and VANDEKERCKHOVE, J. (2016). A Bayesian perspective on the reproducibility project: Psychology. *PLoS ONE* **11** e0149794. https://doi.org/10.1371/journal.pone.0149794

FARRIS, C., TREAT, T. A., VIKEN, R. J. and MCFALL, R. M. (2008). Perceptual mechanisms that characterize gender differences in decoding women's sexual intent. *Psychol. Sci.* **19** 348–354.

FISHER, R. A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron* **1** 3–32.

FISHER, R. A. (1924). The distribution of the partial correlation coefficient. *Metron* **3** 329–332.

FITHIAN, W., SUN, D. L. and TAYLOR, J. E. (2014). Optimal inference after model selection. arXiv:1410.2597v2.

GELMAN, A. and CARLIN, J. (2014). Beyond power calculations. *Perspectives on Psychological Science* **9** 641–651.

GELMAN, A. and O'ROURKE, K. (2013). Discussion: Difficulties in making inferences about scientific truth from distributions of published p-values. *Biostatistics* **15** 18–23.

GELMAN, A. and TUERLINCKX, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Comput. Statist.* **15** 373–390.

GILBERT, D. T., KING, G., PETTIGREW, S. and WILSON, T. D. (2016a). Comment on "Estimating the reproducibility of psychological science." *Science* **351** 1037a.

GILBERT, D. T., KING, G., PETTIGREW, S. and WILSON, T. D. (2016b). A response to the reply to our technical comment on "Estimating the reproducibility of psychological science."

GILBERT, D. T., KING, G., PETTIGREW, S. and WILSON, T. D. (2016c). More on "Estimating the reproducibility of psychological science".

GOODMAN, S. N. (2013). Discussion: An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics* **15** 23–27.

GOODMAN, S. N., FANELLI, D. and IOANNIDIS, J. P. A. (2016). What does research reproducibility mean? *Sci. Transl. Med.* **8** 341ps12. https://doi.org/10.1126/scitranslmed.aaf5027

HEDGES, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statist. Sci.* **7** 246–255.

HELLER, R., GOLLAND, Y., MALACH, R. and BENJAMINI, Y. (2007). Conjunction group analysis: An alternative to mixed/random effect analysis. *NeuroImage* **37** 1178–1185.

HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6** 65–70. MR0538597

HUNG, K. and FITHIAN, W. (2020). Supplement to "Statistical methods for replicability assessment." https://doi.org/10.1214/20-AOAS1336SUPPA, https://doi.org/10.1214/20-AOAS1336SUPPB

IOANNIDIS, J. P. A. (2013). Discussion: Why "An estimate of the science-wise false discovery rate and application to the top medical literature" is false. *Biostatistics* **15** 28–36.

JAGER, L. R. and LEEK, J. T. (2014). An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics* **15** 1–12. https://doi.org/10.1093/biostatistics/kxt007

JOHNSON, V. E., PAYNE, R. D., WANG, T., ASHER, A. and MANDAL, S. (2017). On the reproducibility of psychological science. *J. Amer. Statist. Assoc.* **112** 1–10. MR3646548 https://doi.org/10.1080/01621459.2016.1240079

KLEIN, R. A., VIANELLO, M., HASSELMAN, F., ADAMS, B. G., ADAMS, R. B., ALPER, S., AVEYARD, M., AXT, J. R., BAHNÍK, Š. et al. (2018). Many Labs 2: Investigating variation in replicability across sample and setting.

LARSEN, J. T. and MCKIBBAN, A. R. (2008). Is happiness having what you want, wanting what you have, or both? *Psychol. Sci.* **19** 371–377. https://doi.org/10.1111/j.1467-9280.2008.02095.x

LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. MR3485948 https://doi.org/10.1214/15-AOS1371

MOREY, R. D. and LAKENS, D. (2017). Why most of psychology is statistically unfalsifiable. https://doi.org/doi:10.1021/es060015i

NOSEK, B. A. and ERRINGTON, T. M. (2017). Making sense of replications. *eLife* **6**. https://doi.org/10.7554/eLife.23383

NOSEK, B. A. and GILBERT, E. (2016). Let's not mischaracterize replication studies: Authors.

OPEN SCIENCE COLLABORATION (2015). Estimating the reproducibility of psychological science. *Science* **349** 943.

PURDIE-VAUGHNS, V., STEELE, C. M., DAVIES, P. G., DITLMANN, R. and CROSBY, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *J. Pers. Soc. Psychol.* **94** 615–630. https://doi.org/10.1037/0022-3514.94.4.615

SAMPSON, A. R. and SILL, M. W. (2005). Drop-the-losers design: Normal case. *Biom. J.* **47** 257–268. MR2145117 https://doi.org/10.1002/bimj.200410119

SIMONSOHN, U., NELSON, L. D. and SIMMONS, J. P. (2014a). *p*-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science* **9** 666–681.

SIMONSOHN, U., NELSON, L. D. and SIMMONS, J. P. (2014b). *P*-curve: A key to the file-drawer. *J. Exp. Psychol. Gen.* **143** 534–547.

SRIVASTAVA, S. (2015). Moderator interpretations of the Reproducibility Project.

SRIVASTAVA, S. (2016). Evaluating a new critique of the Reproducibility Project.

STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 479–498. MR1924302 https://doi.org/10.1111/1467-9868.00346

STROEBE, W. (2016). Are most published social psychological findings false?. *J. Exp. Soc. Psychol.* **66** 134–144.

THE ECONOMIST (2016). The scientific method. *The Economist*.

VALENTINE, J. C., BIGLAN, A., BORUCH, R. F., CASTRO, F. G., COLLINS, L. M., FLAY, B. R., KELLAM, S., MOŚCICKI, E. K. and SCHINKE, S. P. (2011). Replication in prevention science. *Prev. Sci.* **12** 103–117.

VAN AERT, R. C. M. and VAN ASSEN, M. A. L. M. (2017). Bayesian evaluation of effect size after replicating an original study. *PLoS ONE* **12** e0175302–23. https://doi.org/10.1371/journal.pone.0175302

VAN AERT, R. C. M. and VAN ASSEN, M. A. L. M. (2018). Examining reproducibility in psychology: A hybrid method for combining a statistically significant original study and a replication. *Behav. Res. Methods* **50** 1515–1539. https://doi.org/10.3758/s13428-017-0967-6

VAN DIJK, E., VAN KLEEF, G. A., STEINEL, W. and VAN BEEST, I. (2008). A social functional approach to emotions in bargaining: When communicating anger pays and when it backfires. *J. Pers. Soc. Psychol.* **94** 600–614.

WEINSTEIN, A., FITHIAN, W. and BENJAMINI, Y. (2013). Selection adjusted confidence intervals with more power to determine the sign. *J. Amer. Statist. Assoc.* **108** 165–176. MR3174610 https://doi.org/10.1080/01621459.2012.737740

YEKUTIELI, D. (2012). Adjusted Bayesian inference for selected parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 515–541. MR2925372 https://doi.org/10.1111/j.1467-9868.2011.01016.x

ZÖLLNER, S. and PRITCHARD, J. K. (2007). Overcoming the winner's curse: Estimating penetrance parameters from case-control data. *Am. J. Hum. Genet.* **80** 605–615.

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1088–1121
https://doi.org/10.1214/20-AOAS1350

EFFICIENCY IN LUNG TRANSPLANT ALLOCATION STRATEGIES

BY JINGJING ZOU[1], DAVID J. LEDERER[2] AND DANIEL RABINOWITZ[3]

[1]*Division of Biostatistics and Bioinformatics, Department of Family Medicine and Public Health, University of California, San Diego, j2zou@health.ucsd.edu*

[2]*Regeneron Pharmaceuticals, Inc. and Columbia University Irving Medical Center, dl427@cumc.columbia.edu*

[3]*Department of Statistics, Columbia University, dan@stat.columbia.edu*

Currently in the United States, lung transplantations are allocated to candidates according to each candidate's lung allocation score (LAS). The LAS is an ad hoc ranking system for patients' priorities of transplantation. The goal of this study is to develop a framework for improving patients' life expectancies over the LAS based on a comprehensive modeling of the lung transplantation waiting list. Patients and organs are modeled as arriving according to Poisson processes, patients' health status evolving a waiting time inhomogeneous Markov process until death or transplantation, with organ recipient's expected post-transplant residual life depending on waiting time and health status at transplantation. Under allocation rules satisfying minimal fairness requirements, the long-term average expected life converges, and its limit is a natural standard for comparing allocation strategies. Via the Hamilton–Jacobi–Bellman equations, upper bounds for the limiting average expected life are derived as a function of organ availability. Corresponding to each upper bound is an allocable set of (state, time) pairs at which patients would be optimally transplanted. The allocable set expands monotonically as organ availability increases which motivates the development of an allocation strategy that leads to long-term expected life close to the upper bound. Simulation studies are conducted with model parameters estimated from national lung transplantation data. Results suggest that, compared to the LAS, the proposed allocation strategy could provide a 7.7% increase in average total life. We further extend the results to the allocation and matching of multiple organ types.

REFERENCES

AHN, J. H. and HORNBERGER, J. C. (1996). Involving patients in the cadaveric kidney transplant allocation process: A decision-theoretic perspective. *Manage. Sci.* **42** 629–641.

AKAN, M., ALAGOZ, O., ATA, B., ERENAY, F. S. and SAID, A. (2012). A broader view of designing the liver allocation system. *Oper. Res.* **60** 757–770. MR2979420 https://doi.org/10.1287/opre.1120.1064

ALAGOZ, O., MAILLART, L. M., SCHAEFER, A. J. and ROBERTS, M. S. (2004). The optimal timing of living-donor liver transplantation. *Manage. Sci.* **50** 1420–1430.

ALAGOZ, O., MAILLART, L. M., SCHAEFER, A. J. and ROBERTS, M. S. (2007a). Choosing among living-donor and cadaveric livers. *Manage. Sci.* **53** 1702–1715.

ALAGOZ, O., MAILLART, L. M., SCHAEFER, A. J. and ROBERTS, M. S. (2007b). Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Oper. Res.* **55** 24–36. MR2290886 https://doi.org/10.1287/opre.1060.0329

ALBRIGHT, C. and DERMAN, C. (1972). Asymptotic optimal policies for the stochastic sequential assignment problem. *Manage. Sci.* **19** 46–51. MR0414063 https://doi.org/10.1287/mnsc.19.1.46

ASHLAGI, I., GILCHRIST, D. S., ROTH, A. E. and REES, M. A. (2011). Nonsimultaneous chains and dominos in kidney- paired donation-revisited. *Am. J. Transplant.* **11** 984–994.

ASHLAGI, I., GAMARNIK, D., REES, M. A. and ROTH, A. E. (2012). The need for (long) chains in kidney exchange.

ATA, B., DING, Y. and ZENIOS, S. A. (2018). An achievable-region-based approach for kidney allocation policy design with endogenous patient choice. *Manuf. Serv. Oper. Manag.*

Key words and phrases. Lung transplantation, lung allocation score (LAS), allocation rules, Markov decision models, Hamilton–Jacobi–Bellman equations, simulation studies.

BERTSIMAS, D., FARIAS, V. F. and TRICHAKIS, N. (2013). Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. *Oper. Res.* **61** 73–87. MR3042742 https://doi.org/10.1287/opre.1120.1138

BERTSIMAS, D. and NIÑO-MORA, J. (1996). Conservation laws, extended polymatroids and multiarmed bandit problems; a polyhedral approach to indexable systems. *Math. Oper. Res.* **21** 257–306. MR1397215 https://doi.org/10.1287/moor.21.2.257

BERTSIMAS, D. and NIÑO-MORA, J. (2000). Restless bandits, linear programming relaxations, and a primal-dual index heuristic. *Oper. Res.* **48** 80–90. MR1753226 https://doi.org/10.1287/opre.48.1.80.12444

BONSER, R. S., TAYLOR, R., COLLETT, D., THOMAS, H. L., DARK, J. H. and NEUBERGER, J. (2012). Effect of donor smoking on survival after lung transplantation: A cohort study of a prospective registry. *Lancet* **380** 747–755.

CECHLÁROVÁ, K. and LACKO, V. (2012). The kidney exchange problem: How hard is it to find a donor? *Ann. Oper. Res.* **193** 255–271. MR2874765 https://doi.org/10.1007/s10479-010-0691-4

DERMAN, C., LIEBERMAN, G. J. and ROSS, S. M. (1972). A sequential stochastic assignment problem. *Manage. Sci.* **18** 349–355. MR0299189 https://doi.org/10.1287/mnsc.18.7.349

DERMAN, C., LIEBERMAN, G. J. and ROSS, S. M. (1975). A stochastic sequential allocation model. *Oper. Res.* **23** 1120–1130. MR0439031 https://doi.org/10.1287/opre.23.6.1120

FLEMING, W. H. and SONER, H. M. (2006). *Controlled Markov Processes and Viscosity Solutions*, 2nd ed. *Stochastic Modelling and Applied Probability* **25**. Springer, New York. MR2179357

GILL, R. D. and JOHANSEN, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18** 1501–1555. MR1074422 https://doi.org/10.1214/aos/1176347865

GITTINS, J. C. (1989). *Multi-Armed Bandit Allocation Indices*. *Wiley-Interscience Series in Systems and Optimization*. Wiley, Chichester. MR0996417

HALMOS, P. (1974). *Measure Theory*. Springer.

HASANKHANI, F. and KHADEMI, A. (2017). Efficient and fair heart allocation policies for transplantation. *MDM Policy Pract* **2** 2381468317709475. https://doi.org/10.1177/2381468317709475

HORNBERGER, J. and AHN, J. H. (1997). Deciding eligibility for transplantation when a donor kidney becomes available. *Med. Decis. Mak.* **17** 160–170.

KOPP, B. T., GRONER, J., TOBIAS, J. D., WHITSON, B. A., KIRKBY, S. and HAYES, D. (2015). Cigarette smoking effect on survival after lung transplant in cystic fibrosis. *Experimental and Clinical Transplantation : Official Journal of the Middle East Society for Organ Transplantation* **13** 529–534.

KOTLOFF, R. M. and THABUT, G. (2012). Lung transplantation. *Am. J. Respir. Crit. Care Med.* **184**.

LIU, V., ZAMORA, M. R., DHILLON, G. S. and WEILL, D. (2010). Increasing lung allocation scores predict worsened survival among lung transplant recipients. *Am. J. Transplant.* **10** 915–920.

MAXWELL, B. G., LEVITT, J. E., GOLDSTEIN, B. A., MOONEY, J. J., NICOLLS, M. R., ZAMORA, M., VALENTINE, V., WEILL, D. and DHILLON, G. S. (2014). Impact of the lung allocation score on survival beyond 1 year. *Am. J. Transplant.* **14** 2288–2294.

MERLO, C. A., WEISS, E. S., ORENS, J. B., BORJA, M. C., DIENER-WEST, M., CONTE, J. V. and SHAH, A. S. (2009). Impact of U.S. lung allocation score on survival after lung transplantation. *J. Heart Lung Transplant.* **28** 769–775.

NIÑO-MORA, J. (2001). Restless bandits, partial conservation laws and indexability. *Adv. in Appl. Probab.* **33** 76–98. MR1825317 https://doi.org/10.1239/aap/999187898

OTO, T., GRIFFITHS, A. P., LEVVEY, B., PILCHER, D. V., WHITFORD, H., KOTSIMBOS, T. C., RABINOV, M., ESMORE, D. S., WILLIAMS, T. J. et al. (2004). A donor history of smoking affects early but not late outcome in lung transplantation. *Transplantation* **78** 599–606.

RIGHTER, R. (1989). A resource allocation problem in a random environment. *Oper. Res.* **37** 329–338. MR0990612 https://doi.org/10.1287/opre.37.2.329

ROTH, A. E., SÖNMEZ, T. and UTKU ÜNVER, M. (2004). Kidney exchange. *Q. J. Econ.* **119** 457–488.

ROTH, A. E., SÖNMEZ, T. and UTKU ÜNVER, M. (2005). Pairwise kidney exchange. *J. Econom. Theory* **125** 151–188. MR2186970 https://doi.org/10.1016/j.jet.2005.04.004

RUSSO, M. J., IRIBARNE, A., HONG, K. N., DAVIES, R. R., XYDAS, S., TAKAYAMA, H., IBRAHIMIYE, A., GELIJNS, A. C., BACCHETTA, M. D. et al. (2010). High lung allocation score is associated with increased morbidity and mortality following transplantation. *CHEST Journal* **137** 651–657.

RUSSO, M. J., WORKU, B., IRIBARNE, A., HONG, K. N., YANG, J. A., VIGNESWARAN, W. and SONETT, J. R. (2011). Does lung allocation score maximize survival benefit from lung transplantation?. *J. Thorac. Cardiovasc. Surg.* **141** 1270–1277.

RUTH, R. J., WYSZEWIANSKI, L. and HERLINE, G. (1985). Kidney transplantation: A simulation model for examining demand and supply. *Management Science. Journal of the Institute of Management Science. Application and Theory Series* **31** 515–526.

SHIGEMURA, N., TOYODA, Y., BHAMA, J. K., GRIES, C. J., CRESPO, M., JOHNSON, B., ZALDONIS, D., PILEWSKI, J. M. and BERMUDEZ, C. A. (2013). Donor smoking history and age in lung transplantation: A revisit. *Transplantation* **95** 513–518.

SU, X. and ZENIOS, S. A. (2004). Patient choice in kidney allocation: The role of the queueing discipline. *Manuf. Serv. Oper. Manag*. **6** 280–301.

SU, X. and ZENIOS, S. A. (2005). Patient choice in kidney allocation: A sequential stochastic assignment model. *Oper. Res*. **53** 443–455. MR2148501 https://doi.org/10.1287/opre.1040.0180

SU, X. and ZENIOS, S. A. (2006). Recipient choice can address the efficiency-equity trade-off in kidney transplantation: A mechanism design model. *Management Science*. *Journal of the Institute of Management Science*. *Application and Theory Series* **52** 1647–1660.

UNOS (2011). OPTN SRTR 2011 Annual Data Report: Lung. Technical Report.

UNOS (2017). Organ Procurement and Transplantation Network (OPTN) Policies. 1–223.

VALAPOUR, M., SKEANS, M. A., SMITH, J. M., EDWARDS, L. B., CHERIKH, W. S., UCCELLINI, K., ISRANI, A. K., SNYDER, J. J. and KASISKE, B. L. (2017). OPTN/SRTR 2015 annual data report: Lung. *Am. J. Transplant*. **17** 357–424. https://doi.org/10.1111/ajt.14129

VOCK, D. M., TSIATIS, A. A., DAVIDIAN, M., LABER, E. B., TSUANG, W. M., FINLEN COPELAND, C. A. and PALMER, S. M. (2013). Assessing the causal effect of organ transplantation on the distribution of residual lifetime. *Biometrics* **69** 820–829. MR3146778 https://doi.org/10.1111/biom.12084

WEBER, R. (1992). On the Gittins index for multiarmed bandits. *Ann. Appl. Probab.* **2** 1024–1033. MR1189430

WHITTLE, P. (1980). Multi-armed bandits and the Gittins index. *J. Roy. Statist. Soc. Ser. B* **42** 143–149. MR0583348

WHITTLE, P. (1988). Restless bandits: Activity allocation in a changing world. *J. Appl. Probab.* **25A** 287–298. A celebration of applied probability. MR0974588 https://doi.org/10.1017/s0021900200040420

ZENIOS, S. A. (1999). Modeling the transplant waiting list: A queueing model with reneging. *Queueing Syst. Theory Appl*. **31** 239–251. MR1706048 https://doi.org/10.1023/A:1019162331525

ZENIOS, S. A., CHERTOW, G. M. and WEIN, L. M. (2000). Dynamic allocation of kidneys to candidates on the transplant waiting list. *Oper. Res*. **48** 549–569.

ZENIOS, S. A., WEIN, L. M. and CHERTOW, G. M. (1999). Evidence-based organ allocation. *Am. J. Med*. **107** 52–61. https://doi.org/10.1016/s0002-9343(99)00166-7

ZOU, J. (2015). Efficiency in Lung Transplant Allocation Strategies. Ph.D. dissertation, Columbia Univ. MR3358231

ZOU, J., LEDERER, D. J. and RABINOWITZ, D. (2020). Supplement to “Efficiency in lung transplant allocation strategies.” https://doi.org/10.1214/20-AOAS1350SUPP

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1122–1145
https://doi.org/10.1214/20-AOAS1348

MARKOV DECISION PROCESSES WITH DYNAMIC TRANSITION PROBABILITIES: AN ANALYSIS OF SHOOTING STRATEGIES IN BASKETBALL

BY NATHAN SANDHOLTZ* AND LUKE BORNN†

Department of Statistics and Actuarial Science, Simon Fraser University, [*]*nsandhol@sfu.ca;* [†]*lbornn@sfu.ca*

In this paper we model basketball plays as episodes from team-specific nonstationary Markov decision processes (MDPs) with shot clock dependent transition probabilities. Bayesian hierarchical models are employed in the modeling and parametrization of the transition probabilities to borrow strength across players and through time. To enable computational feasibility, we combine lineup-specific MDPs into team-average MDPs using a novel transition weighting scheme. Specifically, we derive the dynamics of the team-average process such that the expected transition count for an arbitrary state-pair is equal to the weighted sum of the expected counts of the separate lineup-specific MDPs.

We then utilize these nonstationary MDPs in the creation of a basketball play simulator with uncertainty propagated via posterior samples of the model components. After calibration, we simulate seasons both on-policy and under altered policies and explore the net changes in efficiency and production under the alternate policies. Additionally, we discuss the game-theoretic ramifications of testing alternative decision policies.

REFERENCES

CARPENTER, B., GELMAN, A., HOFFMAN, M. D., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P. et al. (2017). Stan: A probabilistic programming language. *J. Stat. Softw.* **76** Issue 1.

CERVONE, D., D'AMOUR, A., BORNN, L. and GOLDSBERRY, K. (2016). A multiresolution stochastic process model for predicting basketball possession outcomes. *J. Amer. Statist. Assoc.* **111** 585–599. MR3538688 https://doi.org/10.1080/01621459.2016.1141685

FRANKS, A., MILLER, A., BORNN, L. and GOLDSBERRY, K. (2015). Characterizing the spatial structure of defensive skill in professional basketball. *Ann. Appl. Stat.* **9** 94–121. MR3341109 https://doi.org/10.1214/14-AOAS799

GOLDMAN, M. and RAO, J. (2014a). Optimal stopping in the NBA: An empirical model of the Miami Heat. Available at SSRN 2363709.

GOLDMAN, M. and RAO, J. M. (2014b). Misperception of risk and incentives by experienced agents. Available at SSRN 2435551.

GOLDNER, K. (2012). A Markov model of football: Using stochastic processes to model a football drive. *J. Quant. Anal. Sports* **8**.

GRIMSTEAD, C. M. and SNELL, J. L. (1997). *Introduction to Probability*, 2nd revised ed. Amer. Math. Soc., Providence, RI.

GUDMUNDSSON, J. and HORTON, M. (2017). Spatio-temporal analysis of team sports. *ACM Comput. Surv.* **50** Art. ID 22.

HIROTSU, N. and WRIGHT, M. (2002). Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. *J. Oper. Res. Soc.* **53** 88–96.

LI, W. and NG, M. K. (2014). On the limiting probability distribution of a transition probability tensor. *Linear Multilinear Algebra* **62** 362–385. MR3175419 https://doi.org/10.1080/03081087.2013.777436

NEIMAN, T. and LOEWENSTEIN, Y. (2011). Reinforcement learning in professional basketball players. *Nat. Commun.* **2** Art. ID 569. https://doi.org/10.1038/ncomms1580

Key words and phrases. Basketball, Bayesian hierarchical model, Markov decision process, optical tracking data, simulation.

OLIVER, D. (2004). *Basketball on Paper*: *Rules and Tools for Performance Analysis*. Potomac Books, Inc., Washington, DC.

POLSON, N. G. and SCOTT, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Anal*. **7** 887–902. MR3000018 https://doi.org/10.1214/12-BA730

PUTERMAN, M. L. (2014). *Markov Decision Processes*: *Discrete Stochastic Dynamic Programming*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR1270015

ROUTLEY, K. D. (2015). A Markov game model for valuing player actions in ice hockey. In *Conference on Uncertainty in Artificial Intelligence* (*UAI*) 782–791.

SANDHOLTZ, N. and BORNN, L. (2020). Supplement to "Markov decision processes with dynamic transition probabilities: An analysis of shooting strategies in basketball." https://doi.org/10.1214/20-AOAS1348SUPP

SKINNER, B. (2012). The problem of shot selection in basketball. *PLoS ONE* **7** Art. ID e30776. https://doi.org/10.1371/journal.pone.0030776

SKINNER, B. and GOLDMAN, M. (2017). Optimal strategy in basketball. In *Handbook of Statistical Methods and Analyses in Sports*. *Chapman & Hall/CRC Handb*. *Mod*. *Stat*. *Methods* 229–244. CRC Press, Boca Raton, FL. MR3837238

ŠTRUMBELJ, E. and VRAČAR, P. (2012). Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. *Int*. *J*. *Forecast*. **28** 532–542.

SUTTON, R. S. and BARTO, A. G. (2018). *Reinforcement Learning*: *An Introduction*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3889951

THOMAS, A. C., VENTURA, S. L., JENSEN, S. T. and MA, S. (2013). Competing process hazard function models for player ratings in ice hockey. *Ann*. *Appl*. *Stat*. **7** 1497–1524. MR3127956 https://doi.org/10.1214/13-AOAS646

VRAČAR, P., ŠTRUMBELJ, E. and KONONENKO, I. (2016). Modeling basketball play-by-play data. *Expert Syst*. *Appl*. **44** 58–66.

WARD, J. H. JR. (1963). Hierarchical grouping to optimize an objective function. *J*. *Amer*. *Statist*. *Assoc*. **58** 236–244. MR0148188

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1146–1163
https://doi.org/10.1214/20-AOAS1337

STATISTICAL METHODS FOR ANALYSIS OF COMBINED CATEGORICAL BIOMARKER DATA FROM MULTIPLE STUDIES

BY CHAO CHENG[1,2,*] AND MOLIN WANG[1,3,4,†],

[1]*Department of Epidemiology, Harvard T.H. Chan School of Public Health,* [*]*c.cheng@yale.edu;* [†]*stmow@channing.harvard.edu*

[2]*Department of Mathematical Sciences, Tsinghua University*

[3]*Department of Biostatistics, Harvard T.H. Chan School of Public Health*

[4]*Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School*

In the analysis of pooled data from multiple studies involving a biomarker exposure, the biomarker measurements can vary across laboratories and usually require calibration to a reference assay prior to pooling. Previous researches consider the measurements from a reference laboratory as the gold standard, even though measurements in the reference laboratory are not necessarily closer to the underlying truth in reality. In this paper we do not treat any laboratory measurements as the gold standard, and we develop two statistical methods, the exact calibration and cut-off calibration methods, for the analysis of aggregated categorical biomarker data. We compare the performance of both methods for estimating the biomarker-disease relationship under a random sample or controls-only calibration design. Our findings include: (1) the exact calibration method provides significantly less biased estimates and more accurate confidence intervals than the other method; (2) the cut-off calibration method could yield estimates with minimal bias and valid confidence intervals under small measurement errors and/or small exposure effects; (3) controls-only calibration design can result in additional bias, but the bias is minimal if the exposure effects and/or disease prevalences are small. Finally, we illustrate the methods in an application evaluating the relationship between circulating vitamin D levels and colorectal cancer risk in a pooling project.

REFERENCES

CHENG, C. and WANG, M. (2020). Supplement to "Statistical methods for analysis of combined categorical biomarker data from multiple studies." https://doi.org/10.1214/20-AOAS1337SUPP

CHOI, H. K., ATKINSON, K., KARLSON, E. W. and CURHAN, G. (2005). Obesity, weight change, hypertension, diuretic use, and risk of gout in men: The health professionals follow-up study. *Arch. Intern. Med.* **165** 742–748.

COLDITZ, G. A., MANSON, J. E. and HANKINSON, S. E. (1997). The Nurses' Health Study: 20-year contribution to the understanding of health among women. *J. Women's Health* **6** 49–62. https://doi.org/10.1089/jwh.1997.6.49

CROWE, F. L., APPLEBY, P. N., TRAVIS, R. C., BARNETT, M., BRASKY, T. M., BUENO-DE-MESQUITA, H. B., CHAJES, V., CHAVARRO, J. E., CHIRLAQUE, M.-D. et al. (2014). Circulating fatty acids and prostate cancer risk: Individual participant meta-analysis of prospective studies. *J. Natl. Cancer Inst.* **106** Art. ID dju240.

FERNÁNDEZ, C. and STEEL, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *J. Amer. Statist. Assoc.* **93** 359–371. MR1614601 https://doi.org/10.2307/2669632

FLETCHER, D. J. and UNDERWOOD, A. J. (2002). How to cope with negative estimates of components of variance in ecological field studies. *J. Exp. Mar. Biol. Ecol.* **273** 89–95.

GAIL, M. H., WU, J., WANG, M. et al. (2016). Calibration and seasonal adjustment for matched case-control studies of vitamin D and cancer. *Stat. Med.* **35** 2133–2148. MR3513504 https://doi.org/10.1002/sim.6856

GALLICCHIO, L., HELZLSOUER, K. J., CHOW, W.-H., FREEDMAN, D. M., HANKINSON, S. E., HARTGE, P., HARTMULLER, V., HARVEY, C., HAYES, R. B. et al. (2010). Circulating 25-hydroxyvitamin D and the risk

Key words and phrases. Between-study variability, calibration, multiple studies, categorical biomarker data, measurement error.

of rarer cancers: Design and methods of the Cohort Consortium Vitamin D Pooling Project of Rarer Cancers. *Am. J. Epidemiol.* **172** 10–20.

GANGULI, B., STAUDENMAYER, J. and WAND, M. P. (2005). Additive models with predictors subject to measurement error. *Aust. N. Z. J. Stat.* **47** 193–202. MR2155119 https://doi.org/10.1111/j.1467-842X.2005.00383.x

GODAMBE, V. P. (1991). *Estimating Functions*. Oxford University Press, New York.

HU, Y. and RIDDER, G. (2012). Estimation of nonlinear models with mismeasured regressors using marginal information. *J. Appl. Econometrics* **27** 347–385. MR2905032 https://doi.org/10.1002/jae.1202

KEY, T. J., APPLEBY, P. N., ALLEN, N. E. and REEVES, G. K. (2010). Pooling biomarker data from different studies of disease risk, with a focus on endogenous hormones. *Cancer Epidemiol. Biomark. Prev.* **19** 960–965.

KEY, T. J., APPLEBY, P. N., TRAVIS, R. C., ALBANES, D., ALBERG, A. J., BARRICARTE, A., BLACK, A., BOEING, H., BUENO-DE-MESQUITA, H. B. et al. (2015). Carotenoids, retinol, tocopherols, and prostate cancer risk: Pooled analysis of 15 studies. *Am. J. Clin. Nutr.* **102** 1142–1157.

LAI, J. K., LUCAS, R. M., BANKS, E., PONSONBY, A.-L. and GROUP, A. I. (2012). Variability in vitamin D assays impairs clinical assessment of vitamin D status. *Intern. Med. J.* **42** 43–50.

LANZKRON, P. J., ROSE, D. J. and SZYLD, D. B. (1991). Convergence of nested classical iterative methods for linear systems. *Numer. Math.* **58** 685–702. MR1090255 https://doi.org/10.1007/BF01385649

LEITHY, H. A. E., WAHED, Z. A. A. and ABDALLAH, M. S. (2016). On non-negative estimation of variance components in mixed linear models. *J. Adv. Res.* **7** 59–68. https://doi.org/10.1016/j.jare.2015.02.001

LIANG, H., THURSTON, S. W., RUPPERT, D., APANASOVICH, T. and HAUSER, R. (2008). Additive partial linear models with measurement errors. *Biometrika* **95** 667–678. MR2443182 https://doi.org/10.1093/biomet/asn024

MCCULLOUGH, M. L., ZOLTICK, E. S., WEINSTEIN, S. J., FEDIRKO, V., WANG, M., COOK, N. R., ELIASSEN, A. H., ZELENIUCH-JACQUOTTE, A., AGNOLI, C. et al. (2018). Circulating vitamin D and colorectal cancer risk: An international pooling project of 17 cohorts. *J. Natl. Cancer Inst.* **111** 158–169.

MITCHELL, E. M., LYLES, R. H., MANATUNGA, A. K., DANAHER, M., PERKINS, N. J. and SCHISTERMAN, E. F. (2014). Regression for skewed biomarker outcomes subject to pooling. *Biometrics* **70** 202–211. MR3251681 https://doi.org/10.1111/biom.12134

RAO, C. R. (1972). Estimation of variance and covariance components in linear models. *J. Amer. Statist. Assoc.* **67** 112–115. MR0314185

ROBINSON, L. D. and JEWELL, N. P. (1991). Some surprising results about covariate adjustment in logistic regression models. *Int. Stat. Rev.* **58** 227–240.

SLOAN, A., SONG, Y., GAIL, M. H., BETENSKY, R., ROSNER, B., ZIEGLER, R. G., SMITH-WARNER, S. A. and WANG, M. (2019). Design and analysis considerations for combining data from multiple biomarker studies. *Stat. Med.* **38** 1303–1320. MR3920616 https://doi.org/10.1002/sim.8052

SNELLMAN, G., MELHUS, H., GEDEBORG, R., BYBERG, L., BERGLUND, L., WERNROTH, L. and MICHAËLSSON, K. (2010). Determining vitamin D status: A comparison between commercially available assays. *PLoS ONE* **5** Art. ID e11555. https://doi.org/10.1371/journal.pone.0011555

THOMPSON, W. A. JR. (1962). The problem of negative estimates of variance components. *Ann. Math. Stat.* **33** 273–289. MR0133912 https://doi.org/10.1214/aoms/1177704731

TSILIDIS, K. K., TRAVIS, R. C., APPLEBY, P. N., ALLEN, N. E., LINDSTRÖM, S., ALBANES, D., ZIEGLER, R. G., MCCULLOUGH, M. L., SIDDIQ, A. et al. (2013). Insulin-like growth factor pathway genes and blood concentrations, dietary protein and risk of prostate cancer in the NCI Breast and Prostate Cancer Cohort Consortium (BPC3). *Int. J. Cancer* **133** 495–504.

TWOROGER, S. S. and HANKINSON, S. E. (2006). Use of biomarkers in epidemiologic studies: Minimizing the influence of measurement error in the study design and analysis. *Cancer Causes Control* **17** 889–899.

WANG, N., CARROLL, R. J. and LIANG, K.-Y. (1996). Quasilikelihood estimation in measurement error models with correlated replicates. *Biometrics* **52** 401–411.

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1164–1181
https://doi.org/10.1214/20-AOAS1324

OPTIMAL EMG PLACEMENT FOR A ROBOTIC PROSTHESIS CONTROLLER WITH SEQUENTIAL, ADAPTIVE FUNCTIONAL ESTIMATION (SAFE)

BY JONATHAN STALLRICH[1,*], MD NAZMUL ISLAM[1,†], ANA-MARIA STAICU[1,‡], DUSTIN CROUCH[2], LIZHI PAN[3] AND HE HUANG[4]

[1]*Department of Statistics, North Carolina State University,* [*]*jwstalli@ncsu.edu;* [†]*mnislam@ncsu.edu;* [‡]*astaicu@ncsu.edu*
[2]*Department of Mechanical, Aerospace, and Biomedical Engineering, University of Tennessee, Knoxville, dcrouch3@utk.edu*
[3]*School of Mechanical Engineering, Tianjin University, melzpan@tju.edu.cn*
[4]*Joint Department of Biomedical Engineering, University of North Carolina and North Carolina State University, helen-huang@unc.edu*

Robotic hand prostheses require a controller to decode muscle contraction information, such as electromyogram (EMG) signals, into the user's desired hand movement. State-of-the-art decoders demand extensive training, require data from a large number of EMG sensors and are prone to poor predictions. Biomechanical models of a single movement degree-of-freedom tell us that relatively few muscles, and, hence, fewer EMG sensors are needed to predict movement. We propose a novel decoder based on a dynamic, functional linear model with velocity or acceleration as its response and the recent past EMG signals as functional covariates. The effect of each EMG signal varies with the recent position to account for biomechanical features of hand movement, increasing the predictive capability of a single EMG signal compared to existing decoders. The effects are estimated with a multistage, adaptive estimation procedure that we call Sequential Adaptive Functional Estimation (SAFE). Starting with 16 potential EMG sensors, our method correctly identifies the few EMG signals that are known to be important for an able-bodied subject. Furthermore, the estimated effects are interpretable and can significantly improve understanding and development of robotic hand prostheses.

REFERENCES

BUTTERWORTH, S. (1930). On the theory of filter amplifiers. *Wireless Engineer* **7** 536–541.

CARDOT, H., FERRATY, F. and SARDA, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica* **13** 571–591. MR1997162

CARDOT, H. and SARDA, P. (2008). Varying-coefficient functional linear regression models. *Comm. Statist. Theory Methods* **37** 3186–3203. MR2467760 https://doi.org/10.1080/03610920802105176

CIUPERCA, G. (2019). Adaptive group LASSO selection in quantile models. *Statist. Papers* **60** 173–197. MR3905445 https://doi.org/10.1007/s00362-016-0832-1

CROUCH, D. L. and HUANG, H. (2016). Lumped-parameter electromyogram-driven musculoskeletal hand model: A potential platform for real-time prosthesis control. *Journal of Biomechanics* **49** 3901–3907.

CROUCH, D. L. and HUANG, H. H. (2017). Musculoskeletal model-based control interface mimics physiologic hand dynamics during path tracing task. *J. Neural. Eng.* **14** 036008. https://doi.org/10.1088/1741-2552/aa61bc

DAVENPORT, C. A. (2013). Semiparametric regression models for interacting covariates. Ph.D. thesis, North Carolina State University, Raleigh, NC. MR3218062

DELP, S. L., ANDERSON, F. C., ARNOLD, A. S., LOAN, P., HABIB, A., JOHN, C. T., GUENDELMAN, E. and THELEN, D. G. (2007). OpenSim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Trans. Biomed. Eng.* **54** 1940–1950.

EILERS, P. H. and MARX, B. D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemom. Intell. Lab. Syst.* **66** 159–174.

Key words and phrases. Electromyography signal, varying functional regression, functional variable selection, adaptive group LASSO, correlated functional predictors, sequential adaptive functional estimation.

FAN, Y., JAMES, G. M. and RADCHENKO, P. (2015). Functional additive regression. *Ann. Statist.* **43** 2296–2325. MR3396986 https://doi.org/10.1214/15-AOS1346

FERRATY, F., GONZÁLEZ-MANTEIGA, W., MARTÍNEZ-CALVO, A. and VIEU, P. (2012). Presmoothing in functional linear regression. *Statist. Sinica* **22** 69–94. MR2933168 https://doi.org/10.5705/ss.2010.085

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.

GERTHEISS, J., MAITY, A. and STAICU, A.-M. (2013). Variable selection in generalized functional linear models. *Stat.* **2** 86–101. MR4027303 https://doi.org/10.1002/sta4.20

GOLDSMITH, J., CRAINICEANU, C. M., CAFFO, B. S. and REICH, D. S. (2011). Penalized functional regression analysis of white-matter tract profiles in multiple sclerosis. *NeuroImage* **57** 431–439.

GUO, P., ZENG, F., HU, X., ZHANG, D., ZHU, S., DENG, Y. and HAO, Y. (2015). Improved variable selection algorithm using a LASSO-type penalty, with an application to assessing hepatitis B infection relevant factors in community residents. *PLoS ONE* **10** e0134151. https://doi.org/10.1371/journal.pone.0134151

HAHNE, J. M., BIESSMANN, F., JIANG, N., REHBAUM, H., FARINA, D., MEINECKE, K.-R. and PARRA, L. C. (2014). Linear and nonlinear regression techniques for simultaneous and proportional myoelectric control. *IEEE Trans. Neural Syst. Rehabil. Eng.* **22** 269–279.

HARGROVE, L. J., ENGLEHART, K. B. and HUDGINS, B. (2007). A comparison of surface and intramuscular myoelectric signal classification. *IEEE Trans. Biomed. Eng.* **54** 847–853.

HOLZBAUR, K. R., MURRAY, W. M. and DELP, S. L. (2005). A model of the upper extremity for simulating musculoskeletal surgery and analyzing neuromuscular control. *Ann. Biomed. Eng.* **33** 829–840.

HUANG, H. H., ZHANG, F., SUN, Y. L. and HE, H. (2010). Design of a robust EMG sensing interface for pattern classification. *Journal of Neural Engineering* **7**.

IVANOFF, S., PICARD, F. and RIVOIRARD, V. (2016). Adaptive Lasso and group-Lasso for functional Poisson regression. *J. Mach. Learn. Res.* **17** 1-46. MR3504615

JIANG, N., ENGLEHART, K. B. and PARKER, P. A. (2009). Extracting simultaneous and proportional neural control information for multiple-DOF prostheses from the surface electromyographic signal. *IEEE Trans. Biomed. Eng.* **56** 1070–1080.

JIANG, N., REHBAUM, H., VUJAKLIJA, I., GRAIMANN, B. and FARINA, D. (2014). Intuitive, online, simultaneous, and proportional myoelectric control over two degrees-of-freedom in upper limb amputees. *IEEE Trans. Neural Syst. Rehabil. Eng.* **22** 501–509.

KRSTAJIC, D., BUTUROVIC, L. J., LEAHY, D. E. and THOMAS, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics* **6**. https://doi.org/10.1186/1758--2946--6--10

LEEB, H., PÖTSCHER, B. M. and EWALD, K. (2015). On various confidence intervals post-model-selection. *Statist. Sci.* **30** 216–227. MR3353104 https://doi.org/10.1214/14-STS507

MATSUI, H. and KONISHI, S. (2011). Variable selection for functional regression models via the L_1 regularization. *Comput. Statist. Data Anal.* **55** 3304–3310. MR2825412 https://doi.org/10.1016/j.csda.2011.06.016

MCLEAN, M. W., HOOKER, G., STAICU, A.-M., SCHEIPL, F. and RUPPERT, D. (2014). Functional generalized additive models. *J. Comput. Graph. Statist.* **23** 249–269. MR3173770 https://doi.org/10.1080/10618600.2012.729985

MEIER, L. (2009). grplasso: Fitting user specified models with group lasso penalty. *R Foundation for Statistical Computing*.

MEINSHAUSEN, N. (2007). Relaxed Lasso. *Comput. Statist. Data Anal.* **52** 374–393. MR2409990 https://doi.org/10.1016/j.csda.2006.12.019

MERCIER, C., REILLY, K. T., VARGAS, C. D., ABALLEA, A. and SIRIGU, A. (2006). Mapping phantom movement representations in the motor cortex of amputees. *Brain* **129** 2202–2210.

O'RAHILLY, R. and MÜLLER, F. (1983). *Basic Human Anatomy: A Regional Study of Human Structure*. Saunders, Philadelphia.

PANNU, J. and BILLOR, N. (2017). Robust group-LASSO for functional regression model. *Comm. Statist. Simulation Comput.* **46** 3356–3374. MR3656106 https://doi.org/10.1080/03610918.2015.1096375

PAUL, D., BAIR, E., HASTIE, T. and TIBSHIRANI, R. (2008). "Preconditioning" for feature selection and regression in high-dimensional problems. *Ann. Statist.* **36** 1595–1618. MR2435449 https://doi.org/10.1214/009053607000000578

R CORE TEAM (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

RAMSAY, J. and HOOKER, G. (2017). *Dynamic Data Analysis: Modeling Data With Differential Equations*. *Springer Series in Statistics*. Springer, New York. MR3645102 https://doi.org/10.1007/978-1-4939-7190-9

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

RAMSEY, J. O., WICKHAM, H., GRAVES, S. and HOOKER, G. (2014). Package fda: Functional data analysis. *R Foundation for Statistical Computing*.

RESNIK, L. (2011). Development and testing of new upper-limb prosthetic devices: Research designs for usability testing. *Journal of Rehabilitation Research and Development* **48** 697–706.

RESNIK, L., ETTER, K., KLINGER, S. L. and KAMBE, C. (2011). Using virtual reality environment to facilitate training with advanced upper-limb prosthesis. *Journal of Rehabilitation Research and Development* **48** 707–718.

RESNIK, L., HUANG, H. H., WINSLOW, A., CROUCH, D. L., ZHANG, F. and WOLK, N. (2018). Evaluation of EMG pattern recognitiion for upper limb prosthesis control: A case study in comparison with direct myoelectric control. *Journal of NeuroEngineering and Rehabilitation* **15**.

ROBERTS, D. R., BAHN, V., CIUTI, S., BOYCE, M. S., ELITH, J., GUILLERA-ARROITA, G., HAUENSTEIN, S., LAHOZ-MONFORT, J. J., SCHRÖDER, B. et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40** 913–929.

SCHEME, E. and ENGLEHART, K. (2011). Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use. *Journal of Rehabilitation Research and Development* **48** 643–659.

SHERWOOD, B. and MAIDMAN, A. (2017). Package 'rqPen'. *R Foundation for Statistical Computing*.

STALLRICH, J., ISLAM, M. N., STAICU, A.-M., CROUCH, D. L., PAN, L. and HUANG, H. H. (2020). Supplement to "Optimal EMG placement for a robotic prosthesis controller with sequential, adaptive functional estimation (SAFE)." https://doi.org/10.1214/20-AOAS1324SUPPA, https://doi.org/10.1214/20-AOAS1324SUPPB.

WEI, F. and HUANG, J. (2010). Consistent group selection in high-dimensional linear regression. *Bernoulli* **16** 1369–1384. MR2759183 https://doi.org/10.3150/10-BEJ252

WOOD, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics* **62** 1025–1036. MR2297673 https://doi.org/10.1111/j.1541-0420.2006.00574.x

WU, Y., FAN, J. and MÜLLER, H.-G. (2010). Varying-coefficient functional linear regression. *Bernoulli* **16** 730–758. MR2730646 https://doi.org/10.3150/09-BEJ231

XIAO, L., ZIPUNNIKOV, V., RUPPERT, D. and CRAINICEANU, C. (2016). Fast covariance estimation for high-dimensional functional data. *Stat. Comput.* **26** 409–421. MR3439382 https://doi.org/10.1007/s11222-014-9485-x

YANG, Y. and ZOU, H. (2013). gglasso: Group lasso penalized learning using a unified BMD algorithm. *R Foundation for Statistical Computing*.

YANG, Y. and ZOU, H. (2015). A fast unified algorithm for solving group-lasso penalize learning problems. *Stat. Comput.* **25** 1129–1141. MR3401877 https://doi.org/10.1007/s11222-014-9498-5

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561 https://doi.org/10.1198/016214504000001745

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574 https://doi.org/10.1111/j.1467-9868.2005.00532.x

ZHAO, S., SHOJAIE, A. and WITTEN, D. (2017). In defense of the indefensible: A very naive approach to high-dimensional inference. Preprint. Available at arXiv:1705.05543.

ZHOU, P., LOWERY, M. M., ENGLEHART, K. B., HUANG, H., LI, G., HARGROVE, L., DEWALD, J. P. A. and KUIKEN, T. A. (2007). Decoding a new neural-machine interface for control of artificial limbs. *Journal of Neuralphysiology* **98** 2974–2982.

ZIEGLER-GRAHAM, K., MACKENZIE, E. J., EPHRAIM, P. L., TRAVISON, T. G. and BROOKMEYER, R. (2008). Estimating the prevalence of limb loss in the United States: 2005 to 2050. *Archives of Physical Medicine and Rehabilitation* **89** 422–429.

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1182–1206
https://doi.org/10.1214/20-AOAS1322

ACTIVE MATRIX FACTORIZATION FOR SURVEYS

BY CHELSEA ZHANG[1,*], SEAN J. TAYLOR[2], CURTISS COBB[3] AND JASJEET SEKHON[1,†]

[1]*Department of Statistics, University of California, Berkeley,* [*]*cyz@berkeley.edu;* [†]*sekhon@berkeley.edu*
[2]*Lyft, Inc., seanjtaylor@gmail.com*
[3]*Facebook, Inc., ccobb@fb.com*

Amid historically low response rates, survey researchers seek ways to reduce respondent burden while measuring desired concepts with precision. We propose to ask fewer questions of respondents and impute missing responses via probabilistic matrix factorization. A variance-minimizing active learning criterion chooses the most informative questions per respondent. In simulations of our matrix sampling procedure on real-world surveys as well as a Facebook survey experiment, we find active question selection achieves efficiency gains over baselines. The reduction in imputation error is heterogeneous across questions and depends on the latent concepts they capture. Modeling responses with the ordered logit likelihood improves imputations and yields an adaptive question order. We find for the Facebook survey that potential biases from order effects are likely to be small. With our method, survey researchers obtain principled suggestions of questions to retain and, if desired, can automate the design of shorter instruments.

REFERENCES

ABADIE, A., DIAMOND, A. and HAINMUELLER, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J. Amer. Statist. Assoc.* **105** 493–505. MR2759929 https://doi.org/10.1198/jasa.2009.ap08746

ADAMS, R. P., DAHL, G. E. and MURRAY, I. (2010). Incorporating side information in probabilistic matrix factorization with Gaussian processes. Preprint. Available at arXiv:1003.4944.

ANSOLABEHERE, S. and JONES, P. E. (2010). Constituents' responses to congressional roll-call voting. *Amer. J. Polit. Sci.* **54** 583–597.

ANSOLABEHERE, S. and SCHAFFNER, B. (2010). CCES common content, 2012. V3 [Version]. Available at http://hdl.handle.net/1902.1/17705 (accessed June 2, 2014).

ATHEY, S., BAYATI, M., DOUDCHENKO, N., IMBENS, G. and KHOSRAVI, K. (2018). Matrix completion methods for causal panel data models Technical report, National Bureau of Economic Research, Cambridge, MA.

ATTENBERG, J. and PROVOST, F. (2011). Inactive learning?: Difficulties employing active learning in practice. *ACM SIGKDD Explor. Newsl.* **12** 36–41.

BEN-MICHAEL, E., FELLER, A. and ROTHSTEIN, J. (2018). The augmented synthetic control method. Preprint. Available at arXiv:1811.04170.

BROOCKMAN, D. E., KALLA, J. L. and SEKHON, J. S. (2017). The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. *Polit. Anal.* **25** 435–464.

CANDÈS, E. J. and RECHT, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. MR2565240 https://doi.org/10.1007/s10208-009-9045-5

CAO, Y. and XIE, Y. (2015). Categorical matrix completion. In 2015 *IEEE* 6*th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* (*CAMSAP*) 369–372. IEEE, Piscataway, NJ.

CHAKRABORTY, S., ZHOU, J., BALASUBRAMANIAN, V., PANCHANATHAN, S., DAVIDSON, I. and YE, J. (2013). Active matrix completion. In 2013 *IEEE* 13*th International Conference on Data Mining* 81–90. IEEE, Piscataway, NJ.

CHALONER, K. and VERDINELLI, I. (1995). Bayesian experimental design: A review. *Statist. Sci.* **10** 273–304. MR1390519

CHEN, H., FELT, M.-H. and HUYNH, K. P. (2017). Retail payment innovations and cash usage: Accounting for attrition by using refreshment samples. *J. Roy. Statist. Soc. Ser. A* **180** 503–530. MR3601781 https://doi.org/10.1111/rssa.12208

Key words and phrases. Active learning, adaptive surveys, matrix factorization, multidimensional adaptive testing, optimal design, survey imputation.

CHIANG, K.-Y., HSIEH, C.-J. and DHILLON, I. S. (2015). Matrix completion with noisy side information. In *Advances in Neural Information Processing Systems* 3447–3455.

PRODUCTIVITY COMMISSION (2018). Superannuation: Assessing efficiency and competitiveness. Productivity Commission draft report overview. Published April.

DAVENPORT, M. A., PLAN, Y., VAN DEN BERG, E. and WOOTTERS, M. (2014). 1-bit matrix completion. *Inf. Inference* **3** 189–223. MR3311452 https://doi.org/10.1093/imaiai/iau006

DILLMAN, D. A., SINCLAIR, M. D. and CLARK, J. R. (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opin. Q.* **57** 289–304.

EARLY, K., MANKOFF, J. and FIENBERG, S. E. (2017). Dynamic question ordering in online surveys. *J. Off. Stat.* **33** 625–657.

EDWARDS, P., ROBERTS, I., CLARKE, M., DIGUISEPPI, C., PRATAP, S., WENTZ, R. and KWAN, I. (2002). Increasing response rates to postal questionnaires: Systematic review. *BMJ* **324** Art. ID 1183.

ELAHI, M., RICCI, F. and RUBENS, N. (2016). A survey of active learning in collaborative filtering recommender systems. *Comput. Sci. Rev.* **20** 29–50. MR3519039 https://doi.org/10.1016/j.cosrev.2016.05.002

FAN, W. and YAN, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Comput. Hum. Behav.* **26** 132–139.

FITHIAN, W. and MAZUMDER, R. (2013). Flexible low-rank statistical modeling with side information. Preprint. Available at arXiv:1308.4211.

FULLERTON, A. S. and XU, J. (2012). The proportional odds with partial proportionality constraints model for ordinal response variables. *Soc. Sci. Res.* **41** 182–198.

GABRY, J. and GOODRICH, B. (2016). rstanarm: Bayesian applied regression modeling via Stan. R Package Version 2.

GARNETT, R., KRISHNAMURTHY, Y., XIONG, X., SCHNEIDER, J. and MANN, R. (2012). Bayesian optimal active search and surveying. Preprint. Available at arXiv:1206.6406.

GOLBANDI, N., KOREN, Y. and LEMPEL, R. (2010). On bootstrapping recommender systems. In *Proceedings of the* 19*th ACM International Conference on Information and Knowledge Management* 1805–1808. ACM, New York.

GONZALEZ, J. M. and ELTINGE, J. L. (2008). Adaptive matrix sampling for the consumer expenditure quarterly interview survey. In *Proceedings of the Section on Survey Research Methods* 2081–2088. American Statistical Association, Alexandria, VA.

GROVES, R. M. (2011). Three eras of survey research. *Public Opin. Q.* **75** 861–871.

HASTIE, T., MAZUMDER, R., LEE, J. D. and ZADEH, R. (2015). Matrix completion and low-rank SVD via fast alternating least squares. *J. Mach. Learn. Res.* **16** 3367–3402. MR3450542

HEBERLEIN, T. A. and BAUMGARTNER, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *Am. Sociol. Rev.* 447–462.

HERZOG, A. R. and BACHMAN, J. G. (1981). Effects of questionnaire length on response quality. *Public Opin. Q.* **45** 549–559.

JOSSE, J., HUSSON, F. et al. (2016). missMDA: A package for handling missing values in multivariate data analysis. *J. Stat. Softw.* **70** 1–31.

KARIMI, R., FREUDENTHALER, C., NANOPOULOS, A. and SCHMIDT-THIEME, L. (2011b). Towards optimal active learning for matrix factorization in recommender systems. In 2011 *IEEE* 23*rd International Conference on Tools with Artificial Intelligence* 1069–1076. IEEE, Piscataway, NJ.

KARIMI, R., FREUDENTHALER, C., NANOPOULOS, A. and SCHMIDT-THIEME, L. (2011a). Non-myopic active learning for recommender systems based on matrix factorization. In 2011 *IEEE International Conference on Information Reuse & Integration* 299–303. IEEE, Piscataway, NJ.

KLOPP, O., LAFOND, J., MOULINES, É. and SALMON, J. (2015). Adaptive multinomial matrix completion. *Electron. J. Stat.* **9** 2950–2975. MR3439190 https://doi.org/10.1214/15-EJS1093

KREUTER, F. (2013). Facing the nonresponse challenge. *Ann. Am. Acad. Polit. Soc. Sci.* **645** 23–35.

KROSNICK, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl. Cogn. Psychol.* **5** 213–236.

KÜNZEL, S. R., SEKHON, J. S., BICKEL, P. J. and YU, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl. Acad. Sci.* **116** 4156–4165.

LAWRENCE, N. D. and URTASUN, R. (2009). Non-linear matrix factorization with Gaussian processes. In *Proceedings of the* 26*th Annual International Conference on Machine Learning* 601–608. ACM, New York.

LIM, Y. J. and TEH, Y. W. (2007). Variational Bayesian approach to movie rating prediction. In *Proceedings of KDD Cup and Workshop* **7** 15–21.

LOWELL, D., LIPTON, Z. C. and WALLACE, B. C. (2019). Practical obstacles to deploying active learning. In *Proceedings of the* 2019 *Conference on Empirical Methods in Natural Language Processing and the* 9*th International Joint Conference on Natural Language Processing* (*EMNLP–IJCNLP*) 21–30.

MARCUS, B., BOSNJAK, M., LINDNER, S., PILISCHENKO, S. and SCHÜTZ, A. (2007). Compensating for low topic interest and long surveys: A field experiment on nonresponse in web surveys. *Soc. Sci. Comput. Rev.* **25** 372–383.

MARLIN, B. M. and ZEMEL, R. S. (2009). Collaborative prediction and ranking with non-random missing data. In *Proceedings of the Third ACM Conference on Recommender Systems* 5–12. ACM, New York.

MAZUMDER, R., HASTIE, T. and TIBSHIRANI, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* **11** 2287–2322. MR2719857

MONTGOMERY, J. M. and CUTLER, J. (2013). Computerized adaptive testing for public opinion surveys. *Polit. Anal.* **21** 172–192.

MULDER, J. and VAN DER LINDEN, W. J. (2009). Multidimensional adaptive testing with optimal design criteria for item selection. *Psychometrika* **74** 273–296. MR2507381 https://doi.org/10.1007/s11336-008-9097-5

MUNGER, G. F. and LOYD, B. H. (1988). The use of multiple matrix sampling for survey research. *J. Exp. Educ.* **56** 187–191.

NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *J. Mach. Learn. Res.* **13** 1665–1697. MR2930649

PORTEOUS, I., ASUNCION, A. and WELLING, M. (2010). Bayesian matrix factorization with side information and Dirichlet process mixtures. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*.

RECHT, B. (2011). A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** 3413–3430. MR2877360

REITER, J. P. and RAGHUNATHAN, T. E. (2007). The multiple adaptations of multiple imputation. *J. Amer. Statist. Assoc.* **102** 1462–1471. MR2372542 https://doi.org/10.1198/016214507000000932

RENNIE, J. D. and SREBRO, N. (2005). Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the* 22*nd International Conference on Machine Learning* 713–719. ACM, New York.

ROBERTS, C., GILBERT, E., ALLUM, N. and EISNER, L. (2019). Research synthesis: Satisficing in surveys: A systematic review of the literature. *Public Opin. Q.* **83** 598–626.

RUBENS, N. and SUGIYAMA, M. (2007). Influence-based collaborative active learning. In *Proceedings of the* 2007 *ACM Conference on Recommender Systems* 145–148. ACM, New York.

RUBIN, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys. Wiley Classics Library*. Wiley Interscience, Hoboken, NJ. MR2117498

RUBINSTEYN, A. and FELDMAN, S. (2016). fancyimpute: Multivariate imputation and matrix completion algorithms implemented in Python.

SALAKHUTDINOV, R. and MNIH, A. (2008). Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *Proceedings of the* 25*th International Conference on Machine Learning* 880–887. ACM, New York.

SCHNABEL, T., SWAMINATHAN, A., SINGH, A., CHANDAK, N. and JOACHIMS, T. (2016). Recommendations as treatments: Debiasing learning and evaluation. Preprint. Available at arXiv:1602.05352.

SEEGER, M. and BOUCHARD, G. (2012). Fast variational Bayesian inference for non-conjugate matrix factorization models. In *Artificial Intelligence and Statistics* 1012–1018.

SEGALL, D. O. (2009). Principles of multidimensional adaptive testing. In *Elements of Adaptive Testing* 57–75. Springer, Berlin.

SENGUPTA, N., SREBRO, N. and EVANS, J. (2018). Simple surveys: Response retrieval inspired by recommendation systems. Preprint. Available at arXiv:1901.09659.

SETTLES, B. (2009). Active learning literature survey Technical report, Dept. Computer Sciences, Univ. Wisconsin-Madison.

SHEEHAN, K. B. (2001). E-mail survey response rates: A review. *J. Comput.-Mediat. Commun.* **6** JCMC621.

SILVA, J. and CARIN, L. (2012). Active learning for online Bayesian matrix factorization. In *Proceedings of the* 18*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 325–333. ACM, New York.

SREBRO, N., RENNIE, J. and JAAKKOLA, T. S. (2005). Maximum-margin matrix factorization. In *Advances in Neural Information Processing Systems* 1329–1336.

SREBRO, N. and SALAKHUTDINOV, R. R. (2010). Collaborative filtering in a non-uniform world: Learning with the weighted trace norm. In *Advances in Neural Information Processing Systems* 2056–2064.

SUTHERLAND, D. J., PÓCZOS, B. and SCHNEIDER, J. (2013). Active learning and search on low-rank matrices. In *Proceedings of the* 19*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 212–220. ACM, New York.

THOMAS, N., RAGHUNATHAN, T. E., SCHENKER, N., KATZOFF, M. J. and JOHNSON, C. L. (2006). An evaluation of matrix sampling methods using data from the National Health and Nutrition Examination Survey. *Surv. Methodol.* **32** 217–231.

TOURANGEAU, R., KREUTER, F. and ECKMAN, S. (2015). Motivated misreporting: Shaping answers to reduce survey burden. In *Survey Measurements. Techniques, Data Quality and Sources of Error* 24–41. Campus Verlag, Frankfurt.

TRAN, D., KUCUKELBIR, A., DIENG, A. B., RUDOLPH, M., LIANG, D. and BLEI, D. M. (2016). Edward: A library for probabilistic modeling, inference, and criticism. Preprint. Available at arXiv:1610.09787.

UDELL, M. and TOWNSEND, A. (2019). Why are big data matrices approximately low rank? *SIAM J. Math. Data Sci.* **1** 144–160. MR3949704 https://doi.org/10.1137/18M1183480

WAGNER, J. (2010). The fraction of missing information as a tool for monitoring the quality of survey data. *Public Opin. Q.* **74** 223–243.

XU, M., JIN, R. and ZHOU, Z.-H. (2013). Speedup matrix completion with side information: Application to multi-label learning. In *Advances in Neural Information Processing Systems* 2301–2309.

YAMMARINO, F. J., SKINNER, S. J. and CHILDERS, T. L. (1991). Understanding mail survey response behavior a meta-analysis. *Public Opin. Q.* **55** 613–639.

ZHANG, C., TAYLOR, S. J., COBB, C. and SEKHON, J. (2020). Supplement to "Active matrix factorization for surveys." https://doi.org/10.1214/20-AOAS1322SUPPA, https://doi.org/10.1214/20-AOAS1322SUPPB

ZHOU, T., SHAN, H., BANERJEE, A. and SAPIRO, G. (2012). Kernelized probabilistic matrix factorization: Exploiting graphs and side information. In *Proceedings of the* 2012 *SIAM International Conference on Data Mining* 403–414. SIAM, Philadelphia, PA.

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1207–1241
https://doi.org/10.1214/20-AOAS1327

SIZE ESTIMATION OF KEY POPULATIONS IN THE HIV EPIDEMIC IN ESWATINI USING INCOMPLETE AND MISALIGNED CAPTURE-RECAPTURE DATA

BY ABHIRUP DATTA[1,*], ANDREW PITA[1,†], AMRITA RAO[2,‡], BHEKIE SITHOLE[3], ZANDILE MNISI[4] AND STEFAN BARAL[2,§]

[1]*Department of Biostatistics, Johns Hopkins University,* [*]*abhidatta@jhu.edu;* [†]*apita3@jhmi.edu*
[2]*Department of Epidemiology, Johns Hopkins University,* [‡]*arao24@jhu.edu;* [§]*sbaral@jhu.edu*
[3]*FHI360 LINKAGES Program, bhekie.sithole@gmail.com*
[4]*Research Department, Ministry of Health, zandimnisi@gmail.com*

In 2020, our understanding of the distributions of HIV risks in the most burdened settings, including eSwatini, remains limited. In part, this is driven by the limited availability of the size and burden of the populations at the greatest risk for HIV. Given pervasive social and healthcare stigmas, the size estimations of these populations often rely on the multiplier method—a variant of the capture-recapture approach where the first survey is replaced by an enumeration of population members who used some service or attended an event. To characterize the distributions of marginalized communities in eSwatini, multiple data sources are available at each region for the multiplier method. Current practices in such circumstances produce multiple population size estimates at each region ignoring the correlation among these estimates. We recast the multiple multiplier method as a special case of capture-recapture problem with incomplete data and propose a fully model based approach for size estimation using multiple capture-recapture data with arbitrary pattern of incompleteness. We use a data augmentation scheme that allows us to model the correlations in the data and produce a unified estimate of population size per region. A hierarchical model ties together the models for multiple regions, allowing us to borrow strength across the regions and enabling extrapolation to areas without data. In eSwatini we also encounter data misalignment where counts from some of the data sources are not available for each region but as an aggregate over few regions. We propose a solution to the general misalignment problem which considers data-source-specific patterns of misalignment. We use simulation studies to demonstrate the accurate inferential capabilities of our Bayesian multiplier method. This approach is then used to produce uncertainty-quantified population size estimates of key populations in eSwatini. Lastly, we propose a Bayesian nonparametric extension for incomplete capture-recapture that allows nonindependent data sources.

REFERENCES

ARNOLD, R., HAYAKAWA, Y. and YIP, P. (2010). Capture-recapture estimation using finite mixtures of arbitrary dimension. *Biometrics* **66** 644–655. MR2758846 https://doi.org/10.1111/j.1541-0420.2009.01289.x

BAO, L., RAFTERY, A. E. and REDDY, A. (2015). Estimating the sizes of populations at risk of HIV infection from multiple data sources using a Bayesian hierarchical model. *Stat. Interface* **8** 125–136. MR3322160 https://doi.org/10.4310/SII.2015.v8.n2.a1

BARAL, S. D., KETENDE, S., MNISI, Z., MABUZA, X., GROSSO, A., SITHOLE, B., MAZIYA, S., KERRIGAN, D. L., GREEN, J. L. et al. (2013). A cross-sectional assessment of the burden of HIV and associated individual- and structural-level characteristics among men who have sex with men in Swaziland. *J. Int. AIDS Soc.* **16** 18768.

BARAL, S., KETENDE, S., GREEN, J. L., CHEN, P. A., GROSSO, A., SITHOLE, B., NTSHANGASE, C., YAM, E., KERRIGAN, D. et al. (2014). Reconceptualizing the HIV epidemiology and prevention needs of female sex workers (FSW) in Swaziland. *PLoS ONE* **9** e115465.

Key words and phrases. Bayesian, capture-recapture, epidemiology, HIV, misalignment, multiplier method.

BASU, S. and EBRAHIMI, N. (2001). Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence. *Biometrika* **88** 269–279. MR1841274 https://doi.org/10.1093/biomet/88.1.269

BERGER, B. O., GROSSO, A., ADAMS, D., KETENDE, S., SITHOLE, B., MABUZA, X. S., MAVIMBELA, M. J. and BARAL, S. (2018). The prevalence and correlates of physical and sexual violence affecting female sex workers in Swaziland. *J. Interpers. Violence* **33** 2745–2766. https://doi.org/10.1177/0886260516629385

BROWN, C. A., GROSSO, A. L., ADAMS, D., SITHOLE, B., KETENDE, S., GREENE, J., MAZIYA, S., MNISI, Z., MABUZA, X. et al. (2016). Characterizing the individual, social, and structural determinants of condom use among men who have sex with men in Swaziland. *AIDS Res. Hum. Retrovir.* **32** 539–546. https://doi.org/10.1089/AID.2015.0240

CASTLEDINE, B. J. (1981). A Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika* **68** 197–210. MR0614956 https://doi.org/10.1093/biomet/68.1.197

CHEN, Y., DIACONIS, P., HOLMES, S. P. and LIU, J. S. (2005). Sequential Monte Carlo methods for statistical analysis of tables. *J. Amer. Statist. Assoc.* **100** 109–120. MR2156822 https://doi.org/10.1198/016214504000001303

DARROCH, J. N. (1958). The multiple-recapture census. I. Estimation of a closed population. *Biometrika* **45** 343–359. MR0119360 https://doi.org/10.2307/2333183

DARROCH, J. N. (1959). The multiple-recapture census. II. Estimation when there is immigration or death. *Biometrika* **46** 336–351. MR0119361 https://doi.org/10.2307/2333531

DATTA, A., LIN, W., RAO, A., DIOUF, D., KOUAME, A., EDWARDS, J. K., BAO, L., LOUIS, T. A. and BARAL, S. (2019). Bayesian estimation of MSM population in Côte d'Ivoire. *Stat. Public Policy* **6** 1–13. https://doi.org/10.1080/2330443X.2018.1546634.

DATTA, A., BANERJEE, S., HODGES, J. S. and GAO, L. (2019). Spatial disease mapping using directed acyclic graph auto-regressive (DAGAR) models. *Bayesian Anal.* **14** 1221–1244. MR4044851 https://doi.org/10.1214/19-BA1177

DATTA, A., PITA, A., RAO, A., SITHOLE, B., MNISI, Z. and BARAL, S. (2020). Supplement to "Size estimation of key populations in the HIV epidemic in eSwatini using incomplete and misaligned capture-recapture data." https://doi.org/10.1214/20-AOAS1327SUPP.

DIACONIS, P. and STURMFELS, B. (1998). Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.* **26** 363–397. MR1608156 https://doi.org/10.1214/aos/1030563990

DOBRA, A. (2012). Dynamic Markov bases. *J. Comput. Graph. Statist.* **21** 496–517. MR2945478 https://doi.org/10.1080/10618600.2012.663285

DOBRA, A., TEBALDI, C. and WEST, M. (2006). Data augmentation in multi-way contingency tables with fixed marginal totals. *J. Statist. Plann. Inference* **136** 355–372. MR2211344 https://doi.org/10.1016/j.jspi.2004.07.002

DUNSON, D. B. and XING, C. (2009). Nonparametric Bayes modeling of multivariate categorical data. *J. Amer. Statist. Assoc.* **104** 1042–1051. MR2562004 https://doi.org/10.1198/jasa.2009.tm08439

EDWARDS, J. K., HILEMAN, S., DONASTORG, Y., ZADROZNY, S., BARAL, S., HARGREAVES, J. R., FEARON, E., ZHAO, J., DATTA, A. and WEIR, S. S. (2009). Estimating sizes of key populations at the national level: Considerations for study design and analysis. *Epidemiology* **29** 795–803. https://doi.org/doi:10.1097/EDE.0000000000000906

FEARON, E., CHABATA, S. T., THOMPSON, J. A., COWAN, F. M. and HARGREAVES, J. R. (2017). Sample size calculations for population size estimation studies using multiplier methods with respondent-driven sampling surveys. *JMIR Public Health Surveill.* **3** e59. https://doi.org/10.2196/publichealth.7909

FIELDING-MILLER, R., MNISI, Z., ADAMS, D., BARAL, S. and KENNEDY, C. (2014). "There is hunger in my community": A qualitative study of food security as a cyclical force in sex work in Swaziland. *BMC Public Health* **14** 79.

FOG, A. (2015). BiasedUrn: Biased Urn Model Distributions. R package version 1.07.

FONNER, V. A., KERRIGAN, D., MNISI, Z., KETENDE, S., KENNEDY, C. E. and BARAL, S. (2014). Social cohesion, social participation, and HIV related risk among female sex workers in Swaziland. *PLoS ONE* **9** e87527. https://doi.org/10.1371/journal.pone.0087527

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2013). *Bayesian Data Analysis*. CRC Press/CRC, Boca Raton.

GEORGE, E. I. and ROBERT, C. P. (1992). Capture-recapture estimation via Gibbs sampling. *Biometrika* **79** 677–683. MR1209469 https://doi.org/10.2307/2337223

GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. MR1380810 https://doi.org/10.1093/biomet/82.4.711

GROVER, E., GROSSO, A., KETENDE, S., KENNEDY, C., FONNER, V., ADAMS, D., SITHOLE, B., MNISI, Z., MAZIYA, S. L. et al. (2016). Social cohesion, social participation and HIV testing among men who have sex with men in Swaziland. *AIDS Care* **28** 795–804. https://doi.org/10.1080/09540121.2015.1131971

HECKATHORN, D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Soc. Probl.* **44** 174–199.

HOLLAND, C. E., PAPWORTH, E., BILLONG, S. C., TAMOUFE, U., LEBRETON, M., KAMLA, A., ELOUNDOU, J., MANAGA, F., YOMB, Y. et al. (2015). Antiretroviral treatment coverage for men who have sex with men and female sex workers living with HIV in Cameroon. *J. Acquir. Immune Defic. Syndr.* **68** S232–S240.

HOLLAND, C. E., KOUANDA, S., LOUGUÉ, M., PITCHE, V. P., SCHWARTZ, S., ANATO, S., OUEDRAOGO, H. G., TCHALLA, J., YAH, C. S. et al. (2016). Using population-size estimation and cross-sectional survey methods to evaluate HIV service coverage among key populations in Burkina Faso and Togo. *Public Health Rep.* **131** 773–782. https://doi.org/10.1177/0033354916677237

KENNEDY, C. E., BARAL, S. D., FIELDING-MILLER, R., ADAMS, D., DLUDLU, P., SITHOLE, B., FONNER, V. A., MNISI, Z. and KERRIGAN, D. (2013). "They are human beings, they are Swazi": Intersecting stigmas and the positive health, dignity and prevention needs of HIV-positive men who have sex with men in Swaziland. *J. Int. AIDS Soc.* **16** 18749.

LINCOLN, F. C. et al. (1930). Calculating waterfowl abundance on the basis of banding returns.

LOGIE, C. H., DIAS, L. V., JENKINSON, J., NEWMAN, P. A., MACKENZIE, R. K., MOTHOPENG, T., MADAU, V., RANOTSI, A., NHLENGETHWA, W. et al. (2018a). Exploring the potential of participatory theatre to reduce stigma and promote health equity for lesbian, gay, bisexual, and transgender (LGBT) people in Swaziland and Lesotho. *Health Educ. Behav.* **46** 146–156.

LOGIE, C. H., PEREZ-BRUMER, A., JENKINSON, J., MADAU, V., NHLENGETHWA, W. and BARAL, S. (2018b). Marginalization and social change processes among lesbian, gay, bisexual and transgender persons in Swaziland: Implications for HIV prevention. *AIDS Care* **30** 33–40.

MANRIQUE-VALLIER, D. (2016). Bayesian population size estimation using Dirichlet process mixtures. *Biometrics* **72** 1246–1254. MR3591609 https://doi.org/10.1111/biom.12502

MANRIQUE-VALLIER, D. and FIENBERG, S. E. (2008). Population size estimation using individual level mixture models. *Biom. J.* **50** 1051–1063. MR2649394 https://doi.org/10.1002/bimj.200810448

MANRIQUE-VALLIER, D. and REITER, J. P. (2014). Bayesian estimation of discrete multivariate latent structure models with structural zeros. *J. Comput. Graph. Statist.* **23** 1061–1079. MR3270711 https://doi.org/10.1080/10618600.2013.844700

NORRIS, J. L. and POLLOCK, K. H. (1996). Nonparametric mle under two closed capture-recapture models with heterogeneity. *Biometrics* 639–649.

PETERSEN, C. G. J. (1896). The yearly immigration of young plaice in the Limfjord from the German Sea. *Rep. Dan. Biol. Stn.* **6** 1–48.

RAO, A., STAHLMAN, S., HARGREAVES, J., WEIR, S., EDWARDS, J., RICE, B., KOCHELANI, D., MAVIMBELA, M. and BARAL, S. (2017). Sampling key populations for HIV surveillance: Results from eight cross-sectional studies using respondent-driven sampling and venue-based snowball sampling. *JMIR Public Health Surveill.* **3** e72. https://doi.org/10.2196/publichealth.8116

RISHER, K., ADAMS, D., SITHOLE, B., KETENDE, S., KENNEDY, C., MNISI, Z., MABUSA, X. and BARAL, S. D. (2013). Sexual stigma and discrimination as barriers to seeking appropriate healthcare among men who have sex with men in Swaziland. *J. Int. AIDS Soc.* **16** 18715. https://doi.org/10.7448/IAS.16.3.18715

SABIN, K., ZHAO, J., CALLEJA, J. M. G., SHENG, Y., GARCIA, S. A., REINISCH, A. and KOMATSU, R. (2016). Availability and quality of size estimations of female sex workers, men who have sex with men, people who inject drugs and transgender women in low- and middle-income countries. *PLoS ONE* **11** e0155150. https://doi.org/10.1371/journal.pone.0155150

SEKAR, C. C. and DEMING, W. E. (1949). On a method of estimating birth and death rates and the extent of registration. *J. Amer. Statist. Assoc.* **44** 101–115.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433

SMITH, P. W., FORSTER, J. J. and MCDONALD, J. W. (1996). Monte Carlo exact tests for square contingency tables. *J. Roy. Statist. Soc. Ser. A* 309–321.

THE PHIA PROJECT, I. (2017). Swaziland HIV incidence measurement survey 2: A population-based HIV impact assessment. Report.

VERMUNT, J. K., VAN GINKEL, J. R., VAN DER ARK, L. A. and SIJTSMA, K. (2008). Multiple imputation of incomplete categorical data using latent class analysis. *Sociol. Method.* **38** 369–397.

WEIR, S., TATE, J., HILEMAN, S., KHAN, M. and JACKSON, E. (2005). Place: priorities for local AIDS control efforts: A manual for implementing the place method.

WEIR, S. S., BARAL, S. D., EDWARDS, J. K., ZADROZNY, S., HARGREAVES, J., ZHAO, J. and SABIN, K. (2018). Opportunities for enhanced strategic use of surveys, medical records, and program data for HIV surveillance of key populations: Scoping review. *JMIR Public Health Surveill.* **4** e28. https://doi.org/10.2196/publichealth.8042

WHO (2010). Guidelines on estimating the size of populations most at risk to HIV. Report, UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance.

YAM, E. A., MNISI, Z., SITHOLE, B., KENNEDY, C., KERRIGAN, D. L., TSUI, A. O. and BARAL, S. (2013). Association between condom use and use of other contraceptive methods among female sex workers in Swaziland: A relationship-level analysis of condom and contraceptive use. *Sex. Transm. Dis.* **40** 406–412. https://doi.org/10.1097/OLQ.0b013e318283c16d

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1242–1257
https://doi.org/10.1214/20-AOAS1341

A SEMIPARAMETRIC MIXTURE METHOD FOR LOCAL FALSE DISCOVERY RATE ESTIMATION FROM MULTIPLE STUDIES

BY SEOK-OH JEONG[1], DONGSEOK CHOI[2] AND WONCHEOL JANG[3]

[1]*Department of Statistics. Hankuk University of Foreign Studies, seokohj@hufs.ac.kr*
[2]*OHSU-PSU School of Public Health, Oregon Health & Science University, choid@ohsu.edu*
[3]*Department of Statistics, Seoul National University, wcjang@snu.ac.kr*

Antineutrophil cytoplasmic antibody associated vasculitis (AAV) is extremely heterogeneous in clinical presentation and involves multiple organ systems. While the clinical presentation of AAV is diverse, we hypothesized that all AAV share common pathways and tested the hypothesis based on three different microarray studies of peripheral leukocytes, sinus and orbital inflammation disease. For the hypothesis testing we developed a two-component semiparametric mixture model to estimate the local false discovery rates from the p-values of three studies. The two pillars of the proposed approach are Efron's empirical null principle and log-concave density estimation for the alternative distribution. Our method outperforms other existing methods, in particular when the proportion of null is not that high. It is robust against the misspecification of alternative distribution. A unique feature of our method is that it can be extended to compute the local false discovery rates by combining multiple lists of p-values.

REFERENCES

ALCORTA, D. A., BARNES, D. A., DOOLEY, M. A., SULLIVAN, P., JONAS, B., LIU, Y., LIONAKI, S., REDDY, C. B., CHIN, H. et al. (2007). Leukocyte gene expression signatures in antineutrophil cytoplasmic autoantibody and lupus glomerulonephritis. *Kidney Int*. **72** 853–64.

BAGNOLI, M. and BERGSTROM, T. (2005). Log-concave probability and its applications. *Econom. Theory* **26** 445–469. MR2213177 https://doi.org/10.1007/s00199-004-0514-4

BALABDAOUI, F. and DOSS, C. R. (2018). Inference for a two-component mixture of symmetric distributions under log-concavity. *Bernoulli* **24** 1053–1071. MR3706787 https://doi.org/10.3150/16-BEJ864

BORDES, L., DELMAS, C. and VANDEKERKHOVE, P. (2006). Semiparametric estimation of a two-component mixture model where one component is known. *Scand. J. Stat.* **33** 733–752. MR2300913 https://doi.org/10.1111/j.1467-9469.2006.00515.x

CHANG, G. T. and WALTHER, G. (2007). Clustering with mixtures of log-concave distributions. *Comput. Statist. Data Anal.* **51** 6242–6251. MR2408591 https://doi.org/10.1016/j.csda.2007.01.008

COORDINATORS, N. R. (2017). Database resources of the national center for biotechnology information. *Nucleic Acids Res*. **45** (Database issue) D12–D17.

CULE, M., GRAMACY, R. and SAMWORTH, R. (2009). LogConcDEAD: An R package for maximum likelihood estimation of a multivariate log-concave density. *J. Stat. Softw.* **29**.

CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 545–607. MR2758237 https://doi.org/10.1111/j.1467-9868.2010.00753.x

DÜMBGEN, L. and RUFIBACH, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15** 40–68. MR2546798 https://doi.org/10.3150/08-BEJ141

DÜMBGEN, L. and RUFIBACH, K. (2011). logcondens: Computations related to univariate log-concave density estimation. *J. Stat. Softw.* **39** 1–28.

EDGAR, R., DOMRACHEV, M. and LASH, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. **30** 207–210.

Key words and phrases. False discovery rate, log concave, microarray, mixture model, next generation sequencing data.

EFRON, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.* **23** 1–22. MR2431866 https://doi.org/10.1214/07-STS236

EFRON, B. (2009). Empirical Bayes estimates for large-scale prediction problems. *J. Amer. Statist. Assoc.* **104** 1015–1028. MR2562003 https://doi.org/10.1198/jasa.2009.tm08523

EFRON, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Institute of Mathematical Statistics (IMS) Monographs* **1**. Cambridge Univ. Press, Cambridge. MR2724758 https://doi.org/10.1017/CBO9780511761362

EFRON, B., TURNBULL, B. and NARASIMHAN, B. (2015). locfdr: Computes local false discovery rates. R package version 1.1-8. https://CRAN.R-project.org/package=locfdr.

FABREGAT, A., SIDIROPOULOS, K., GARAPATI, P., GILLESPIE, M., HAUSMANN, K., HAW, R., JASSAL, B., JUPE, S., KORNINGER, F. et al. (2016). The reactome pathway knowledgebase. *Nucleic Acids Res.* **44** (Database issue) D481–D487.

FABRIZIO, M., NONINO, M., BONO, G., FERRARO, I., FRANÇOIS, P., IANNICOLA, G., MONELLI, M., THÉVENIN, F., STETSON, P. B. et al. (2011). The Carina Project. IV. Radial velocity distribution. *Publ. Astron. Soc. Pac.* **123** 384–401.

FRIEDMAN, M. A., CHOI, D., PLANCK, S. R., ROSENBAUM, J. T. and SIBLEY, C. (2019). Gene expression pathways across multiple tissues in anti-neutrophil cytoplasmic antibody-associated vasculitis reveal core pathways of disease pathology. *J. Rheumatol.* https://doi.org/10.3899/jrheum.180455

GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035–1061. MR2065197 https://doi.org/10.1214/009053604000000283

GRAYSON, P. C., STEILING, K., PLATT, M., BERMAN, J. S., ZHANG, X., XIAO, J., ALEKSEYEV, Y. O., LIU, G., MONACH, P. A. et al. (2015). Defining the nasal transcriptome in granulomatosis with polyangiitis (Wegener's). *Arthritis Rheumatol.* **67** 2233–2239.

HELD, L. (2019). On the Bayesian interpretation of the harmonic mean p-value. *Proc. Natl. Acad. Sci. USA* **116** 5855–5856. MR3939777 https://doi.org/10.1073/pnas.1900671116

HU, H., WU, Y. and YAO, W. (2016). Maximum likelihood estimation of the mixture of log-concave densities. *Comput. Statist. Data Anal.* **101** 137–147. MR3504841 https://doi.org/10.1016/j.csda.2016.03.002

HUNG, H. M. J., O'NEILL, R. T., BAUER, P. and KÖHNE, K. (1997). The behavior of the P-value when the alternative hypothesis is true. *Biometrics* **53** 11–22. MR1450180 https://doi.org/10.2307/2533093

HUNTER, D. R., WANG, S. and HETTMANSPERGER, T. P. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist.* **35** 224–251. MR2332275 https://doi.org/10.1214/009053606000001118

JEONG, S.-O., CHOI, D. and JANG, W. (2020). Supplement to "A semiparametric mixture method for local false discovery rate estimation from multiple studies." https://doi.org/10.1214/20-AOAS1341SUPP

JOHNSTONE, I. M. and SILVERMAN, B. W. (2005). Empirical Bayes selection of wavelet thresholds. *Ann. Statist.* **33** 1700–1752. MR2166560 https://doi.org/10.1214/009053605000000345

KALLENBERG, C. G. M. (2014). Key advances in the clinical approach to ANCA-associated vasculitis. *Nat. Rev. Rheumatol.* **10** 484–493.

KUMAR PATRA, R. and SEN, B. (2016). Estimation of a two-component mixture model with applications to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 869–893. MR3534354 https://doi.org/10.1111/rssb.12148

MACFARLANE, D. G., BOURNE, J. T., DIEPPE, P. A. and EASTY, D. L. (1983). Indolent Wegener's granulomatosis. *Ann. Rheum. Dis.* **42** 398–407.

PLONER, A., CALZA, S., GUSNANTO, A. and PAWITAN, Y. (2006). Multidimensional local false discovery rate for microarray studies. *Bioinformatics* **22** 556–565.

RITCHIE, M. E., PHIPSON, B., WU, D., HU, Y., LAW, C. W., SHI, W. and SMYTH, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43** e47. https://doi.org/10.1093/nar/gkv007

ROSENBAUM, J. T., CHOI, D., WILSON, D. J., GROSSNIKLAUS, H. E., HARRINGTON, C. A., SIBLEY, C. H. et al. (2015). Orbital pseudotumor can be a localized form of granulomatosis with polyangiitis as revealed by gene expression profiling. *Exp. Mol. Pathol.* **99** 271–278.

SELLKE, T., BAYARRI, M. J. and BERGER, J. O. (2001). Calibration of p values for testing precise null hypotheses. *Amer. Statist.* **55** 62–71. MR1818723 https://doi.org/10.1198/000313001300339950

SODERBERG, D. and SEGELMARK, M. (2016). Neutrophil extracellular traps in ANCA-associated vasculitis. *Front. Immunol.* **7** 256.

WALTHER, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.* **97** 508–513. MR1941467 https://doi.org/10.1198/016214502760047032

WALTHER, G. (2009). Inference and modeling with log-concave distributions. *Statist. Sci.* **24** 319–327. MR2757433 https://doi.org/10.1214/09-STS303

WILSON, D. J. (2019). The harmonic mean p-value for combining dependent tests. *Proc. Natl. Acad. Sci. USA* **116** 1195–1200. MR3904688 https://doi.org/10.1073/pnas.1814092116

YAN, J. (2007). Enjoy the joy of copulas: With a package copula. *J. Stat. Softw.* **21**.

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1258–1284
https://doi.org/10.1214/20-AOAS1345

A NOVEL CHANGE-POINT APPROACH FOR THE DETECTION OF GAS EMISSION SOURCES USING REMOTELY CONTAINED CONCENTRATION DATA

BY IDRIS ECKLEY[1], CLAUDIA KIRCH[2] AND SILKE WEBER[3]

[1]*Mathematics and Statistics, Lancaster University, i.eckley@lancs.ac.uk*

[2]*Department of Mathematics, Institute of Mathematical Stochastics, Center for Behavioral Brain Sciences (CBBS), Otto-von-Guericke University Magdeburg, claudia.kirch@ovgu.de*

[3]*Institute of Stochastics, Karlsruhe Institute of Technology (KIT), silke.weber@kit.edu*

Motivated by an example from remote sensing of gas emission sources, we derive two novel change-point procedures for multivariate time series where, in contrast to classical change-point literature, the changes are not required to be aligned in the different components of the time series. Instead, the change points are described by a functional relationship where the precise shape depends on unknown parameters of interest such as the source of the gas emission in the above example. Two different types of tests and the corresponding estimators for the unknown parameters describing the change locations are proposed. We derive the null asymptotics for both tests under weak assumptions on the error time series and show asymptotic consistency under alternatives. Furthermore, we prove consistency for the corresponding estimators of the parameters of interest. The small-sample behavior of the methodology is assessed by means of a simulation study, and the above remote sensing example analyzed in detail.

REFERENCES

ASTON, J. A. D. and KIRCH, C. (2012). Evaluating stationarity via change-point alternatives with applications to fMRI data. *Ann. Appl. Stat.* **6** 1906–1948. MR3058688 https://doi.org/10.1214/12-AOAS565

ASTON, J. A. D. and KIRCH, C. (2018). High dimensional efficiency with applications to change point tests. *Electron. J. Stat.* **12** 1901–1947. MR3815301 https://doi.org/10.1214/18-EJS1442

AUE, A., HÖRMANN, S., HORVÁTH, L. and REIMHERR, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.* **37** 4046–4087. MR2572452 https://doi.org/10.1214/09-AOS707

AUE, A., HÖRMANN, S., HORVÁTH, L., HUŠKOVÁ, M. and STEINEBACH, J. G. (2012). Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory* **28** 804–837. MR2959126 https://doi.org/10.1017/S0266466611000673

BARDWELL, L., FEARNHEAD, P., ECKLEY, I. A., SMITH, S. and SPOTT, M. (2019). Most recent change-point detection in panel data. *Technometrics* **61** 88–98. MR3933661 https://doi.org/10.1080/00401706.2018.1438926

BHATTACHARJEE, Y. (2008). New efforts to detect explosives require advances on many fronts. *Science* **320** 1416–1417.

BICKEL, P. J. and LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969 https://doi.org/10.1214/009053607000000758

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

CHEN, Y.-H. and PRINN, R. G. (2006). Estimation of atmospheric methane emissions between 1996 and 2001 using a three-dimensional global chemical transport model. *J. Geophys. Res., Atmos.* **111**.

CHO, H. and FRYZLEWICZ, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 475–507. MR3310536 https://doi.org/10.1111/rssb.12079

FRIED, R. and IMHOFF, M. (2004). On the online detection of monotonic trends in time series. *Biom. J.* **46** 90–102. MR2055450 https://doi.org/10.1002/bimj.200410007

Key words and phrases. Nonaligned change points, epidemic model, projection methods, dependent errors, multivariate change points.

HERRNDORF, N. (1984a). A functional central limit theorem for ρ-mixing sequences. *J. Multivariate Anal.* **15** 141–146. MR0755822 https://doi.org/10.1016/0047-259X(84)90073-3

HERRNDORF, N. (1984b). A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.* **12** 141–153. MR0723735

HIGHAM, N. J. (2002). *Accuracy and Stability of Numerical Algorithms*, 2nd ed. SIAM, Philadelphia, PA. MR1927606 https://doi.org/10.1137/1.9780898718027

HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57** 1–17. MR0273727 https://doi.org/10.1093/biomet/57.1.1

HIRST, B., JONATHAN, P., DEL CUETO, F. G., RANDELL, D. and KOSUT, O. (2013). Locating and quantifying gas emission sources using remotely obtained concentration data. *Atmos. Environ.* **74** 141–158.

HIRST, B., RANDELL, D., JONES, M., JONATHAN, P., KING, B. and DEAN, M. (2017). A new technique for monitoring the atmosphere above onshore carbon storage projects that can estimate the locations and mass emission rates of detected sources. *Energy Proc.* **114** 3716–3728.

HLÁVKA, Z., HUŠKOVÁ, M., KIRCH, C. and MEINTANIS, S. G. (2017). Fourier-type tests involving martingale difference processes. *Econometric Rev.* **36** 468–492. MR3608385 https://doi.org/10.1080/07474938.2014.977074

HORVÁTH, L. and HUŠKOVÁ, M. (2012). Change-point detection in panel data. *J. Time Series Anal.* **33** 631–648. MR2944843 https://doi.org/10.1111/j.1467-9892.2012.00796.x

HORVÁTH, L., KOKOSZKA, P. and STEINEBACH, J. (1999). Testing for changes in multivariate dependent observations with an application to temperature changes. *J. Multivariate Anal.* **68** 96–119. MR1668911 https://doi.org/10.1006/jmva.1998.1780

HORVÁTH, L. and RICE, G. (2014). Extensions of some classical methods in change point analysis. *TEST* **23** 219–255. MR3210268 https://doi.org/10.1007/s11749-014-0368-4

HUŠKOVÁ, M. and STEINEBACH, J. (2000). Limit theorems for a class of tests of gradual changes. *J. Statist. Plann. Inference* **89** 57–77. MR1794412 https://doi.org/10.1016/S0378-3758(00)00094-X

HUŠKOVÁ, M. and STEINEBACH, J. (2002). Asymptotic tests for gradual changes. *Statist. Decisions* **20** 137–151. MR1924608

JIRAK, M. (2015). Uniform change point tests in high dimension. *Ann. Statist.* **43** 2451–2483. MR3405600 https://doi.org/10.1214/15-AOS1347

KILLICK, R., NAM, C. F. H., ASTON, J. A. D. and ECKLEY, I. A. (2012). changepoint.info: The changepoint repository. Available at http://changepoint.info.

KIRCH, C., MUHSAL, B. and OMBAO, H. (2015). Detection of changes in multivariate time series with application to EEG data. *J. Amer. Statist. Assoc.* **110** 1197–1216. MR3420695 https://doi.org/10.1080/01621459.2014.957545

MATTESON, D. S. and JAMES, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *J. Amer. Statist. Assoc.* **109** 334–345. MR3180567 https://doi.org/10.1080/01621459.2013.849605

MEI, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika* **97** 419–433. MR2650748 https://doi.org/10.1093/biomet/asq010

NAM, C. F. H., ASTON, J. A. D., ECKLEY, I. A. and KILLICK, R. (2015). The uncertainty of storm season changes: Quantifying the uncertainty of autocovariance changepoints. *Technometrics* **57** 194–206. MR3369676 https://doi.org/10.1080/00401706.2014.902776

PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR0088850 https://doi.org/10.1093/biomet/41.1-2.100

POLITIS, D. N. (2003). Adaptive bandwidth choice. *J. Nonparametr. Stat.* **15** 517–533. MR2017485 https://doi.org/10.1080/10485250310001604659

PREUSS, P., PUCHSTEIN, R. and DETTE, H. (2015). Detection of multiple structural breaks in multivariate time series. *J. Amer. Statist. Assoc.* **110** 654–668. MR3367255 https://doi.org/10.1080/01621459.2014.920613

RIGAILL, G. J., CADOT, S., KLUIN, R. J. C., XUE, Z., BERNARDS, R., MAJEWSKI, I. J. and WESSELS, L. F. A. (2012). A regression model for estimating DNA copy number applied to capture sequencing data. *Bioinformatics* **28** 2357–2365.

SIEGMUND, D., YAKIR, B. and ZHANG, N. R. (2011). Detecting simultaneous variant intervals in aligned sequences. *Ann. Appl. Stat.* **5** 645–668. MR2840169 https://doi.org/10.1214/10-AOAS400

TRUOUNG, C., OUDRE, L. and VAYATIS, N. (2018). A review of change point detection methods.

WANG, Y., MEI, Y. and PAYNABAR, K. (2018). Thresholded multivariate principal component analysis for phase I multichannel profile monitoring. *Technometrics* **60** 360–372. MR3847172 https://doi.org/10.1080/00401706.2017.1375993

WANG, T. and SAMWORTH, R. J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 57–83. MR3744712 https://doi.org/10.1111/rssb.12243

WEBER, S. M. (2017). Change-point procedures for multivariate dependent data. Ph.D. thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe. https://doi.org/10.5445/IR/1000068981

ZHANG, N. R., SIEGMUND, D. O., JI, H. and LI, J. Z. (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika* **97** 631–645. MR2672488 https://doi.org/10.1093/biomet/asq025

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1285–1303
https://doi.org/10.1214/20-AOAS1338

DOES TERRORISM TRIGGER ONLINE HATE SPEECH? ON THE ASSOCIATION OF EVENTS AND TIME SERIES

BY ERIK SCHARWÄCHTER[*] AND EMMANUEL MÜLLER[†]

Data Science and Data Engineering, University of Bonn, [*]*scharwaechter@bit.uni-bonn.de;* [†]*mueller@bit.uni-bonn.de*

Hate speech is ubiquitous on the Web. Recently, the offline causes that contribute to online hate speech have received increasing attention. A recurring question is whether the occurrence of extreme events offline systematically triggers bursts of hate speech online, indicated by peaks in the volume of hateful social media posts. Formally, this question translates into measuring the association between a sparse event series and a time series. We propose a novel statistical methodology to measure, test and visualize the systematic association between rare events and peaks in a time series. In contrast to previous methods for causal inference or independence tests on time series, our approach focuses only on the *timing* of events and peaks and no other distributional characteristics. We follow the framework of event coincidence analysis (ECA) that was originally developed to correlate point processes. We formulate a discrete-time variant of ECA and derive all required distributions to enable analyses of peaks in time series with a special focus on serial dependencies and peaks over multiple thresholds. The analysis gives rise to a novel visualization of the association via quantile-trigger rate plots. We demonstrate the utility of our approach by analyzing whether Islamist terrorist attacks in Western Europe and North America systematically trigger bursts of hate speech and counter-hate speech on Twitter.

REFERENCES

ALVI, H. (2014). The diffusion of intra-Islamic violence and terrorism: The impact of the proliferation of Salafi/Wahhabi ideologies. *Middle East Rev. Int. Aff.* **18** 38–50.

BARTLETT, J. and MILLER, C. (2012). The edge of violence: Towards telling the difference between violent and non-violent radicalization. *Terrorism Polit. Violence* **24** 1–21.

BENNETT, W. L. (2005). Social movements beyond borders: Organization, communication, and political capacity in two eras of transnational activism. In *Transnational Protest and Global Activism* (D. Della Porta and S. Tarrow, eds.) 203–226. Rowman & Littlefield Publishers, Inc., Lanham, MD, USA.

BESSERVE, M., LOGOTHETIS, N. K. and SCHÖLKOPF, B. (2013). Statistical analysis of coupled time series with kernel cross-spectral density operators. In *Advances in Neural Information Processing Systems* (*NIPS*).

BOWMAN-GRIEVE, L. (2009). Exploring "Stormfront": A virtual community of the radical right. *Stud. Confl. Terrorism* **32** 989–1007.

BOX, G. E. P. and TIAO, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *J. Amer. Statist. Assoc.* **70** 70–79. MR0365957

BRESSLER, S. L. and SETH, A. K. (2011). Wiener–Granger causality: A well established methodology. *NeuroImage* **58** 323–329.

BRODERSEN, K. H., GALLUSSER, F., KOEHLER, J., REMY, N. and SCOTT, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *Ann. Appl. Stat.* **9** 247–274. MR3341115 https://doi.org/10.1214/14-AOAS788

BURNAP, P. and WILLIAMS, M. L. (2014). Hate speech, machine classification and statistical modelling of information flows on Twitter: Interpretation and communication for policy decision making. In *Proceedings of the Internet*, *Policy and Politics Conference* (*IPP*).

BURNAP, P. and WILLIAMS, M. L. (2016). Us and them: Identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data Sci.* **5** 1–15.

Key words and phrases. Time series, event series, statistical association, independence test, event coincidence analysis, event synchronization, causal inference, hate speech, social media, terrorism.

BURNAP, P., WILLIAMS, M. L., SLOAN, L., RANA, O., HOUSLEY, W., EDWARDS, A., KNIGHT, V., PROCTER, R. and VOSS, A. (2014). Tweeting the terror: Modelling the social media reaction to the Woolwich terrorist attack. *Soc. Netw. Anal. Min.* **4** 1–14.

ČEHAJIĆ, S. and BROWN, R. (2008). Not in my name: A social psychological study of antecedents and consequences of acknowledgment of in-group atrocities. *Genocide Stud. Prev.* **3** 195–211.

CHANDRASEKHARAN, E., PAVALANATHAN, U., SRINIVASAN, A., GLYNN, A., EISTENSTEIN, J. and GILBERT, E. (2017). You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proc. ACM Human-Comput. Interact.* **1** 31:1–31:22.

CHWIALKOWSKI, K. and GRETTON, A. (2014). A kernel independence test for random processes. In *Proceedings of the* 31*st International Conference on Machine Learning* (*ICML*).

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. *Springer Series in Statistics*. Springer London, Ltd., London. MR1932132 https://doi.org/10.1007/978-1-4471-3675-0

CUNHA, T. O., WEBER, I. and PAPPA, G. L. (2017). A warm welcome matters! The link between social feedback and weight loss in /r/loseit. In *Proceedings of the* 26*th International Conference on World Wide Web* (*WWW*).

DAVIDSON, H. (2014). 'Not in my name' campaign organiser warns of danger of political rhetoric. *The Guardian*. Available at https://www.theguardian.com/australia-news/2014/oct/03/not-in-my-name-campaign-organiser-warns-of-danger-of-political-rhetoric, 2014-10-03.

DAVIDSON, T., WARMSLEY, D., MACY, M. and WEBER, I. (2017). Automated hate speech detection and the problem of offensive language. In *Proceedings of the* 11*th International Conference on Web and Social Media* (*ICWSM*).

DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. *Cambridge Series in Statistical and Probabilistic Mathematics* **1**. Cambridge Univ. Press, Cambridge. MR1478673 https://doi.org/10.1017/CBO9780511802843

DE KOSTER, W. and HOUTMAN, D. (2008). 'Stormfront is like a second home to me.' On virtual community formation by right-wing extremists. *Inf. Commun. Soc.* **11** 1155–1176.

DE SMEDT, T., DE PAUW, G. and VAN OSTAEYEN, P. (2018). Automatic detection of online jihadist hate speech. Technical report, CLiPS Research Center, Univ. Antwerp.

DEWEY, C. (2016). #StopIslam Twitter-trended for all the right reasons. *Washington Post*. Available at https://www.washingtonpost.com/news/the-intersect/wp/2016/03/22/stopislam-twitter-trended-for-all-the-right-reasons/, 2016-03-22.

DJURIC, N., ZHOU, J., MORRIS, R., GRBOVIC, M., RADOSAVLJEVIC, V. and BHAMIDIPATI, N. (2015). Hate speech detection with comment embeddings. In *Proceedings of Th* 24*th International Conference on World Wide Web* (*WWW*).

DONGES, J. F., SCHLEUSNER, C. F., SIEGMUND, J. F. and DONNER, R. V. (2016). Event coincidence analysis for quantifying statistical interrelationships between event time series: On the role of flood events as triggers of epidemic outbreaks. *Eur. Phys. J. Spec. Top.* **487** 471–487.

DUDOIT, S. and VAN DER LAAN, M. J. (2008). *Multiple Testing Procedures with Applications to Genomics*. *Springer Series in Statistics*. Springer, New York. MR2373771 https://doi.org/10.1007/978-0-387-49317-6

ELSHERIEF, M., NILIZADEH, S., NGUYEN, D., VIGNA, G. and BELDING, E. (2018a). Peer to peer hate: Hate speech instigators and their targets. In *Proceedings of the* 12*th International AAAI Conference on Web and Social Media* (*ICWSM*).

ELSHERIEF, M., KULKARNI, V., NGUYEN, D., WANG, W. Y. and BELDING, E. (2018b). Hate lingo: A target-based linguistic analysis of hate speech in social media. In *Proceedings of the* 12*th International AAAI Conference on Web and Social Media* (*ICWSM*).

NATIONAL CONSORTIUM FOR THE STUDY OF TERRORISM (2018). Global terrorism database [2014–2017]. Available at https://www.start.umd.edu/gtd.

FRAHM, G., JUNKER, M. and SCHMIDT, R. (2005). Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance Math. Econom.* **37** 80–100. MR2156598 https://doi.org/10.1016/j.insmatheco.2005.05.008

GAGLIARDONE, I., GAL, D., ALVES, T. and MARTINEZ, G. (2015). *Countering Online Hate Speech*. UNESCO, Paris, France.

GRANGER, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37** 424–438.

HEMMINGS, C. (2016). From #StopIslam to Allison Pearson and Katie Hopkins, the social media response to Brussels has been shocking. *The Independent*. Available at https://www.independent.co.uk/voices/from-stopislam-to-allison-pearson-and-katie-hopkins-the-social-media-response-to-brussels-has-been-a6946116.html, 2016-03-22.

KRISHNAN, M. (2017). 'Not in my name'—Indians protest vigilante attacks on Muslims. *Deutsche Welle*. Available at https://p.dw.com/p/2fYJP, 2017-06-29.

KWOK, I. and WANG, Y. (2013). Locate the hate: Detecting tweets against blacks. In *Proceedings of the* 27*th AAAI Conference on Artificial Intelligence* (*AAAI*).

MAGDY, W., DARWISH, K. and ABOKHODAIR, N. (2015). Quantifying public response towards Islam on Twitter after Paris attacks. Preprint. Available at arXiv:1512.04570.

MATIAS, J. N., JOHNSON, A., BOESEL, W. E., KEEGAN, B., FRIEDMAN, J. and DETAR, C. (2015). *Reporting, Reviewing, and Responding to Harassment on Twitter*. Women, Action, and the Media.

MEDDAUGH, P. M. and KAY, J. (2009). Hate speech or "Reasonable racism?" The other in stormfront. *J. Mass Media Ethics* **24** 251–268.

MONDAL, M., SILVA, L. A. and BENEVENUTO, F. (2017). A measurement study of hate speech in social media. In *Proceedings of the* 28*th ACM Conference on Hypertext and Social Media* (*HT*).

MÜLLER, K. and SCHWARZ, C. (2019). Fanning the flames of hate: Social media and hate crime. *SSRN Electron. J.* 1–84. https://doi.org/10.2139/ssrn.3082972

OLTEANU, A., CASTILLO, C., BOY, J. and VARSHNEY, K. R. (2018). The effect of extremist violence on hateful speech online. In *Proceedings of the* 12*th International AAAI Conference on Web and Social Media* (*ICWSM*).

QUIROGA, R. Q., KREUZ, T. and GRASSBERGER, P. (2002). Event synchronization: A simple and fast method to measure synchronicity and time delay patterns. *Phys. Rev. E* (3) **66** 041904, 9. MR1935193 https://doi.org/10.1103/PhysRevE.66.041904

RIBEIRO, M. H., CALAIS, P. H., SANTOS, Y. A., ALMEIDA, V. A. F. and MEIRA, W. (2018). Characterizing and detecting hateful users on Twitter. In *Proceedings of the* 12*th International AAAI Conference on Web and Social Media* (*ICWSM*).

SAHA, K., CHANDRASEKHARAN, E. and DE CHOUDHURY, M. (2019). Prevalence and psychological effects of hateful speech in online college communities. In *Proceedings of the* 11*th ACM Conference on Web Science* (*WebSci*).

SCHARWÄCHTER, E. and MÜLLER, E. (2020). Two-sample testing for event impacts in time series. In *Proceedings of the SIAM International Conference on Data Mining* (*SIAM SDM*). https://doi.org/10.1137/1.9781611976236.2

SCHARWÄCHTER, E. and MÜLLER, E. (2020). Supplement to "Does terrorism trigger online hate speech? On the association of events and time series." https://doi.org/10.1214/20-AOAS1338SUPP

SCHLEUSSNER, C.-F., DONGES, J. F., DONNER, R. V. and SCHELLNHUBER, H. J. (2016). Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized countries. *Proc. Natl. Acad. Sci. USA* **113** 9216–9221.

SCHMIDT, A. and WIEGAND, M. (2017). A survey on hate speech detection using natural language processing. In *Proceedings of the* 5*th International Workshop on Natural Language Processing for Social Media* (*SocialNLP*).

SCHREIBER, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* **85** 461–464.

SIEGMUND, J. F., SANDERS, T. G. M., HEINRICH, I., VAN DER MAATEN, E., SIMARD, S., HELLE, G. and DONNER, R. V. (2016). Meteorological drivers of extremes in daily stem radius variations of beech, oak, and pine in northeastern Germany: An event coincidence analysis. *Front. Plant Sci.* **7** 733. https://doi.org/10.3389/fpls.2016.00733

SILVA, L., MONDAL, M., CORREA, D., BENEVENUTO, F. and WEBER, I. (2016). Analyzing the targets of hate in online social media. In *Proceedings of the* 10*th International Conference on Web and Social Media* (*ICWSM*).

TORMEY, S. (2006). 'Not in my name': Deleuze, Zapatismo and the critique of representation. *Parliam. Aff.* **59** 138–154.

WARNER, W. and HIRSCHBERG, J. (2012). Detecting hate speech on the World Wide Web. In *Proceedings of the* 2*nd Workshop on Language in Social Media* (*LSM*).

WILLIAMS, M. L. and BURNAP, P. (2016). Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *Br. J. Criminol.* **56** 211–238.

YAN, X., WU, Q. and ZHANG, W. (2019). Cross-sectional learning of extremal dependence among financial assets. In *Neural Information Processing Systems*.

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1304–1325
https://doi.org/10.1214/20-AOAS1346

PTEM: A POPULARITY-BASED TOPICAL EXPERTISE MODEL FOR COMMUNITY QUESTION ANSWERING

BY HOHYUN JUNG[1,*], JAE-GIL LEE[2], NAMGIL LEE[3] AND SUNG-HO KIM[1,†]

[1]*Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology,* [*]*hhjung@kaist.ac.kr;* [†]*sung-ho.kim@kaist.edu*

[2]*Graduate School of Knowledge Service Engineering, Korea Advanced Institute of Science and Technology, jaegil@kaist.ac.kr*

[3]*Department of Information Statistics, Kangwon National University, namgil.lee@kangwon.ac.kr*

Community Question Answering (CQA) websites are widely used in sharing knowledge, where users can ask questions, reply answers and evaluate answers. So far, the evaluation of answers has been explained by the contents of answers through the investigation of users' topics of interest and expertise levels. In this paper we focus on modeling the user's evaluation behavior, in that users can see the answerer's profile as well as the answer content before evaluating the quality of the answer. We propose a model called Popularity-based Topical Expertise Model (PTEM), a generative model to analyze the rich-get-richer phenomenon that popular user's answers are more recommended. We can simultaneously estimate the topical expertise of each user and the strength of the rich-get-richer effect through the EM algorithm combined with collapsed Gibbs sampling. Experiments are performed on the StackExchange data, and the results demonstrate a rich-get-richer phenomenon in the community. We further discuss the superiority and usefulness of the proposed model through analysis in the discipline of philosophy.

REFERENCES

ASLAY, Ç., O'HARE, N., AIELLO, L. M. and JAIMES, A. (2013). Competition-based networks for expert finding. In *Proceedings of the* 36*th International ACM SIGIR Conference on Research and Development in Information Retrieval* 1033–1036. ACM.

BLEI, D. M., NG, A. Y. and JORDAN, M. I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3** 993–1022.

BOUGUESSA, M., DUMOULIN, B. and WANG, S. (2008). Identifying authoritative actors in question-answering forums: The case of yahoo! answers. In *Proceedings of the* 14*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 866–874. ACM.

CAI, Y. and CHAKRAVARTHY, S. (2013). Expertise ranking of users in QA community. In *International Conference on Database Systems for Advanced Applications* 25–40. Springer.

CAO, X., CONG, G., CUI, B. and JENSEN, C. S. (2010). A generalized framework of exploring category information for question retrieval in community question answer archives. In *Proceedings of the* 19*th International Conference on World Wide Web* 201–210. ACM.

GILKS, W. R. and WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **41** 337–348.

GRIFFITHS, T. L. and STEYVERS, M. (2004). Finding scientific topics. *Proc. Natl. Acad. Sci. USA* **101** 5228–5235.

HOFMANN, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* 289–296. Morgan Kaufmann Publishers Inc.

JUNG, H., LEE, J.-G., LEE, N. and KIM, S.-H. (2018). Comparison of fitness and popularity: Fitness-popularity dynamic network model. *J. Stat. Mech. Theory Exp.* **12** 123403, 15. MR3898969 https://doi.org/10.1088/1742-5468/aaeb40

JUNG, H., LEE, J.-G., LEE, N. and KIM, S.-H. (2020). Supplement to "PTEM: A popularity-based topical expertise model for community question answering." https://doi.org/10.1214/20-AOAS1346SUPPA, https://doi.org/10.1214/20-AOAS1346SUPPB

Key words and phrases. Topic analysis, EM algorithm, community question answering, rich get richer, user behavior.

JURCZYK, P. and AGICHTEIN, E. (2007). Discovering authorities in question answer communities by using link analysis. In *Proceedings of the* 16*th ACM Conference on Information and Knowledge Management* 919–922. ACM.

KONDOR, D., PÓSFAI, M., CSABAI, I. and VATTAY, G. (2014). Do the rich get richer? An empirical analysis of the Bitcoin transaction network. *PLoS ONE* **9** e86197. https://doi.org/10.1371/journal.pone.0086197

LIU, J., SONG, Y.-I. and LIN, C.-Y. (2011). Competition-based user expertise score estimation. In *Proceedings of the* 34*th International ACM SIGIR Conference on Research and Development in Information Retrieval* 425–434. ACM.

LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **44** 226–233. MR0676213

MA, Z., SUN, A., YUAN, Q. and CONG, G. (2015). A tri-role topic model for domain-specific question answering. In *Proceedings of the* 29*th AAAI Conference on Artificial Intelligence*.

MERTON, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science* **159** 56–63.

MOVSHOVITZ-ATTIAS, D., MOVSHOVITZ-ATTIAS, Y., STEENKISTE, P. and FALOUTSOS, C. (2013). Analysis of the reputation system and user contributions on a question answering website: Stackoverflow. In *Proceedings of the* 2013 *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* 886–893. ACM.

PAL, A., FARZAN, R., KONSTAN, J. A. and KRAUT, R. E. (2011). Early detection of potential experts in question answering communities. In *International Conference on User Modeling, Adaptation, and Personalization* 231–242. Springer.

PAPADIMITRIOU, C. H., RAGHAVAN, P., TAMAKI, H. and VEMPALA, S. (2000). Latent semantic indexing: A probabilistic analysis. *J. Comput. System Sci.* **61** 217–235. MR1802556 https://doi.org/10.1006/jcss.2000.1711

PATRA, B. (2017). A survey of community question answering. arXiv preprint arXiv:1705.04009.

PAUL, S. A., HONG, L. and CHI, E. H. (2012). Who is authoritative? Understanding reputation mechanisms in quora. arXiv preprint arXiv:1204.3724.

PERC, M. (2014). The Matthew effect in empirical data. *J. R. Soc. Interface* **11** 20140378.

SRBA, I. and BIELIKOVA, M. (2016). A comprehensive survey and classification of approaches for community question answering. *ACM Trans. Web* **10** 18.

TAUSCZIK, Y. R. and PENNEBAKER, J. W. (2011). Predicting the perceived quality of online mathematics contributions from users' reputations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 1885–1888. ACM.

VAN DE RIJT, A., KANG, S. M., RESTIVO, M. and PATIL, A. (2014). Field experiments of success–breeds–success dynamics. *Proc. Natl. Acad. Sci. USA* **111** 6934–6939. https://doi.org/10.1073/pnas.1316836111

VER HOEF, J. M. and BOVENG, P. L. (2007). Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology* **88** 2766–2772.

WANG, X., HUANG, C., YAO, L., BENATALLAH, B. and DONG, M. (2018). A survey on expert recommendation in community question answering. *J. Comput. Sci. Tech.* **33** 625–653.

XU, F., JI, Z. and WANG, B. (2012). Dual role model for question recommendation in community question answering. In *Proceedings of the* 35*th International ACM SIGIR Conference on Research and Development in Information Retrieval* 771–780. ACM.

YANG, B. and MANANDHAR, S. (2014). Exploring user expertise and descriptive ability in community question answering. In *Proceedings of the* 2014 *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* 320–327. IEEE Press.

YANG, L., QIU, M., GOTTIPATI, S., ZHU, F., JIANG, J., SUN, H. and CHEN, Z. (2013). Cqarank: Jointly model topics and expertise in community question answering. In *Proceedings of the* 22*nd ACM International Conference on Information and Knowledge Management* 99–108. ACM.

ZHANG, J., ACKERMAN, M. S. and ADAMIC, L. (2007). Expertise networks in online communities: Structure and algorithms. In *Proceedings of the* 16*th International Conference on World Wide Web* 221–230. ACM.

ZHOU, G., ZHAO, J., HE, T. and WU, W. (2014). An empirical study of topic-sensitive probabilistic model for expert finding in question answer communities. *Knowl.-Based Syst.* **66** 136–145.

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1326–1341
https://doi.org/10.1214/20-AOAS1349

THE JENSEN EFFECT AND FUNCTIONAL SINGLE INDEX MODELS: ESTIMATING THE ECOLOGICAL IMPLICATIONS OF NONLINEAR REACTION NORMS

BY ZI YE[1,*], GILES HOOKER[1,†] AND STEPHEN P. ELLNER[2]

[1]*Department of Statistical Science, Cornell University,* [*]*zy234@cornell.edu;* [†]*gjh27@cornell.edu*
[2]*Department of Ecology and Evolutionary Biology, Cornell University,* *spe2@cornell.edu*

This paper develops tools to characterize how species are affected by environmental variability, based on a functional single index model relating a response such as growth rate to environmental conditions. In ecology the curvature of such responses are used, via Jensen's inequality, to determine whether environmental variability is harmful or beneficial, and differing nonlinear responses to environmental variability can contribute to the coexistence of competing species.

Here, we address estimation and inference for these models with observational data on individual responses to environmental conditions. Because nonparametric estimation of the curvature (second derivative) in a nonparametric functional single index model requires unrealistic sample sizes, we instead focus on directly estimating the effect of the nonlinearity by comparing the average response to a variable environment with the response at the expected environment, which we call the *Jensen Effect*. We develop a test statistic to assess whether this effect is significantly different from zero. In doing so we reinterpret the SiZer method of Chaudhuri and Marron (*J. Amer. Statist. Assoc.* **94** (1999) 807–823) by maximizing a test statistic over smoothing parameters. We show that our proposed method works well both in simulations and on real ecological data from the long-term data set described in Drake (*Proc. R. Soc. Lond., B Biol. Sci.* **272** (2005) 1823–1827).

REFERENCES

CHAUDHURI, P. and MARRON, J. S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.* **94** 807–823. MR1723347 https://doi.org/10.2307/2669996

CHEN, D., HALL, P. and MÜLLER, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39** 1720–1747. MR2850218 https://doi.org/10.1214/11-AOS882

CHESSON, P. (1994). Multispecies competition in variable environments. *Theor. Popul. Biol.* **45** 227–276.

CHESSON, P. (2000a). Mechanisms of maintenance of species diversity. *Ann. Rev. Ecolog. Syst.* **31** 343–366.

CHESSON, P. (2000b). General theory of competitive coexistence in spatially-varying environments. *Theor. Popul. Biol.* **58** 211–237.

CHESSON, P. L. and WARNER, R. R. (1981). Environmental variability promotes coexistence in lottery competitive systems. *Amer. Nat.* **117** 923–943. MR0647805 https://doi.org/10.1086/283778

COHEN, D. (1966). Optimizing reproduction in a randomly varying environment. *J. Theoret. Biol.* **12** 119–129.

DRAKE, J. M. (2005). Population effects of increased climate variation. *Proc. R. Soc. Lond., B Biol. Sci.* **272** 1823–1827.

ELLNER, S. (1987). Alternate plant life history strategies and coexistence in randomly varying environments. In *Theory and Models in Vegetation Science* 199–208. Springer, Berlin.

HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. MR1212171 https://doi.org/10.1214/aos/1176349020

HRISTACHE, M., JUDITSKY, A. and SPOKOINY, V. (2001). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.* **29** 595–623. MR1865333 https://doi.org/10.1214/aos/1009210681

HUTCHINSON, G. E. (1961). The paradox of the plankton. *Amer. Nat.* **95** 137–145.

ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58** 71–120. MR1230981 https://doi.org/10.1016/0304-4076(93)90114-K

Key words and phrases. Functional data analysis, reaction norm, Jensen Effect, single index model, SiZer.

KOONS, D. N., METCALF, C. J. E. and TULJAPURKAR, S. (2008). Evolution of delayed reproduction in uncertain environments: A life-history perspective. *Amer. Nat.* **172** 797–805.

KOONS, D. N., PAVARD, S., BAUDISCH, A., METCALF, J. E. et al. (2009). Is life-history buffering or lability adaptive in stochastic environments? *Oikos* **118** 972–980.

LEWONTIN, R. C. and COHEN, D. (1969). On population growth in a randomly varying environment. *Proc. Natl. Acad. Sci. USA* **62** 1056–1060. MR0269328 https://doi.org/10.1073/pnas.62.4.1056

LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. With discussion and a rejoinder by the author. MR1137117

LI, J., HUANG, C. and ZHU, H. (2017). A functional varying-coefficient single-index model for functional response data. *J. Amer. Statist. Assoc.* **112** 1169–1181. MR3735368 https://doi.org/10.1080/01621459.2016.1195742

MA, S. (2016). Estimation and inference in functional single-index models. *Ann. Inst. Statist. Math.* **68** 181–208. MR3440219 https://doi.org/10.1007/s10463-014-0488-3

MARRON, J. S. and ZHANG, J.-T. (2005). SiZer for smoothing splines. *Comput. Statist.* **20** 481–502. MR2236619 https://doi.org/10.1007/BF02741310

RAMSAY, J. O. (2006). *Functional Data Analysis*. Wiley, New York.

RAMSAY, J. O., HOOKER, G. and GRAVES, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer, Berlin.

RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression*. *Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge Univ. Press, Cambridge. MR1998720 https://doi.org/10.1017/CBO9780511755453

SONDEREGGER, D. L., WANG, H., CLEMENTS, W. H. and NOON, B. R. (2009). Using SiZer to detect thresholds in ecological data. *Front. Ecol. Environ.* **7** 190–195.

TELLER, B. J., ADLER, P. B., EDWARDS, C. B., HOOKER, G. and ELLNER, S. P. (2016). Linking demography with drivers: Climate and competition. *Methods Ecol. Evol.* **7** 171–183.

VASSEUR, D. A. and MCCANN, K. S. (2007). *The Impact of Environmental Variability on Ecological Systems*. Springer, Berlin.

WOOD, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 413–428. MR1749600 https://doi.org/10.1111/1467-9868.00240

YE, Z. and HOOKER, G. (2018). Local quadratic estimation of the curvature in a functional single index model. Preprint. Available at arXiv:1803.09321.

YE, Z., HOOKER, G. and ELLNER, S. P. (2020). Supplement to "The Jensen effect and functional single index models: Estimating the ecological implications of nonlinear reaction norms." https://doi.org/10.1214/20-AOAS1349SUPPA, https://doi.org/10.1214/20-AOAS1349SUPPB

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1342–1358
https://doi.org/10.1214/20-AOAS1355

CLIMATE EXTREME EVENT ATTRIBUTION USING MULTIVARIATE PEAKS-OVER-THRESHOLDS MODELING AND COUNTERFACTUAL THEORY

BY ANNA KIRILIOUK[1] AND PHILIPPE NAVEAU[2]

[1]*Department of Business Administration & Namur Institute for Complex Systems, Université de Namur, anna.kiriliouk@unamur.be*

[2]*Laboratoire des Sciences du Climat et de l'Environnement, (IPSL, CNRS, CEA, UVSQ), naveau@lsce.ipsl.fr*

Numerical climate models are complex and combine a large number of physical processes. They are key tools in quantifying the relative contribution of potential anthropogenic causes (e.g., the current increase in greenhouse gases) on high-impact atmospheric variables like heavy rainfall. These so-called climate extreme event attribution problems are particularly challenging in a multivariate context, that is, when the atmospheric variables are measured on a possibly high-dimensional grid.

In this paper we leverage two statistical theories to assess causality in the context of multivariate extreme event attribution. As we consider an event to be extreme when at least one of the components of the vector of interest is large, extreme-value theory justifies, in an asymptotical sense, a multivariate generalized Pareto distribution to model joint extremes. Under this class of distributions, we derive and study probabilities of necessary and sufficient causation as defined by the counterfactual theory of Pearl. To increase causal evidence, we propose a dimension reduction strategy based on the optimal linear projection that maximizes such causation probabilities. Our approach is tested on simulated examples and applied to weekly winter maxima precipitation outputs of the French CNRM from the recent CMIP6 experiment.

REFERENCES

ANGÉLIL, O., STONE, D. and WEHNER, M. (2017). An independent assessment of anthropogenic attribution statements for recent extreme temperature and rainfall events. *J. Climate* **30** 5–16.

BALKEMA, A. A. and DE HAAN, L. (1974). Residual life time at great age. *Ann. Probab.* **2** 792–804. MR0359049 https://doi.org/10.1214/aop/1176996548

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes*: *Theory and Applications*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. With contributions from Daniel De Waal and Chris Ferro. MR2108013 https://doi.org/10.1002/0470012382

BERNARD, E., NAVEAU, P., VRAC, M. and MESTRE, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *J. Climate* **26** 7929–7937.

CARREAU, J., NAVEAU, P. and NEPPEL, L. (2017). Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. *Water Resour. Res.* **53** 4407–4426.

CHEN, Y., MOUFOUMA-OKIA, W., MASSON-DELMOTTE, V., ZHAI, P. and PIRANI, A. (2018). Recent progress and emerging topics on weather and climate extremes since the fifth assessment report of the intergovernmental panel on climate change. *Annual Review of Environment and Resources* **43** 35–59. https://doi.org/10.1146/annurev-environ-102017-030052

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. *Springer Series in Statistics*. Springer London, Ltd., London. MR1932132 https://doi.org/10.1007/978-1-4471-3675-0

COOLEY, D., HUNTER, B. D. and SMITH, R. L. (2019). Univariate and multivariate extremes for the environmental sciences. In *Handbook of Environmental and Ecological Statistics*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 153–180. CRC Press, Boca Raton, FL. MR3889895

DAVISON, A. C. and HUSER, R. (2015). Statistics of extremes. *Annu. Rev. Stat. Appl.* **2** 203–235. https://doi.org/10.1146/annurev-statistics-010814-020133

Key words and phrases. Multivariate generalized Pareto distribution, necessary and sufficient causation, heavy rainfall, climate change.

DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B* **52** 393–442. With discussion and a reply by the authors. MR1086795

DE FONDEVILLE, R. and DAVISON, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika* **105** 575–592. MR3842886 https://doi.org/10.1093/biomet/asy026

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events: For Insurance and Finance. Applications of Mathematics (New York)* **33**. Springer, Berlin. MR1458613 https://doi.org/10.1007/978-3-642-33483-2

ENGELKE, S., DE FONDEVILLE, R. and OESTING, M. (2019). Extremal behaviour of aggregated data with an application to downscaling. *Biometrika* **106** 127–144. MR3912387 https://doi.org/10.1093/biomet/asy052

FALK, M. and GUILLOU, A. (2008). Peaks-over-threshold stability of multivariate generalized Pareto distributions. *J. Multivariate Anal.* **99** 715–734. MR2406079 https://doi.org/10.1016/j.jmva.2007.03.009

FERREIRA, A. and DE HAAN, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20** 1717–1737. MR3263087 https://doi.org/10.3150/13-BEJ538

FISCHER, E. M. and KNUTTI, R. (2015). Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nat. Clim. Change* **5** 560–564.

HAMMERLING, D., KATZFUSS, M. and SMITH, R. (2019). Climate change detection and attribution. In *Handbook of Environmental and Ecological Statistics. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 789–817. CRC Press, Boca Raton, FL. MR3889921

HANNART, A. and NAVEAU, P. (2018). Probabilities of causation of climate changes. *J. Climate* **31** 5507–5524.

HANNART, A., PEARL, J., OTTO, F., NAVEAU, P. and GHIL, M. (2016). Causal counterfactual theory for the attribution of weather and climate-related events. *Bull. Am. Meteorol. Soc.* **97** 99–110.

HUSER, R., DAVISON, A. C. and GENTON, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes* **19** 79–103. MR3454032 https://doi.org/10.1007/s10687-015-0230-4

KATZ, R. W., PARLANGE, M. B. and NAVEAU, P. (2002). Statistics of extremes in hydrology. *Adv. Water Resour.* **25** 1287–1304.

KATZFUSS, M., HAMMERLING, D. and SMITH, R. L. (2017). A Bayesian hierarchical model for climate change detection and attribution. *Geophys. Res. Lett.* **11** 5720–5728.

KAUFMAN, L. and ROUSSEEUW, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. A Wiley-Interscience Publication. MR1044997 https://doi.org/10.1002/9780470316801

KEW, S. F., PHILIP, S. Y., JAN VAN OLDENBORGH, G., VAN DER SCHRIER, G., OTTO, F. E. L. and VAUTARD, R. (2019). The exceptional summer heat wave in southern Europe 2017. *Bull. Am. Meteorol. Soc.* **100** S49–S53. https://doi.org/10.1175/BAMS-D-18-0109.1

KIM, Y.-H., MIN, S.-K., ZHANG, X., ZWIERS, F., ALEXANDER, L. V., DONAT, M. G. and TUNG, Y.-S. (2016). Attribution of extreme temperature changes during 1951–2010. *Clim. Dyn.* **46** 1769–1782.

KING, A. D. (2017). Attributing changing rates of temperature record breaking to anthropogenic influences. *Earth's Future* **5** 1156–1168.

KIRILIOUK, A. and NAVEAU, P. (2020). Supplement to "Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory." https://doi.org/10.1214/20-AOAS1355SUPP

KIRILIOUK, A., SEGERS, J. and WARCHOŁ, M. (2016). Nonparametric estimation of extremal dependence. In *Extreme Value Modeling and Risk Analysis* 353–375. CRC Press, Boca Raton, FL. MR3644321

KIRILIOUK, A., ROOTZÉN, H., SEGERS, J. and WADSWORTH, J. L. (2019). Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics* **61** 123–135. MR3933664 https://doi.org/10.1080/00401706.2018.1462738

LUU, L. N., VAUTARD, R., YIOU, P., VAN OLDENBORGH, G. J. and LENDERINK, G. (2018). Attribution of extreme rainfall events in the South of France using EURO-CORDEX simulations. *Geophys. Res. Lett.* **45** 6242–6250.

NATIONAL ACADEMIES OF SCIENCES, ENGINEERING AND MEDICINE (2016). *Attribution of Extreme Weather Events in the Context of Climate Change*. The National Academies Press, Washington, DC. https://doi.org/10.17226/21852

NAVEAU, P., HANNART, A. and RIBES, A. (2020). Statistical methods for extreme event attribution in climate science. *Annu. Rev. Stat. Appl.* **7** 89–110. https://doi.org/10.1146/annurev-statistics-031219-041314

OTTO, F. E., PHILIP, S., KEW, S., LI, S., KING, A. and CULLEN, H. (2018). Attributing high-impact extreme events across timescales—a case study of four different types of events. *Clim. Change* **149** 399–412.

PACIOREK, C. J., STONE, D. A. and WEHNER, M. F. (2018). Quantifying statistical uncertainty in the attribution of human influence on severe weather. *Weather Clim. Extrem.* **20** 69–80. https://doi.org/10.1016/j.wace.2018.01.002

PEARL, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge Univ. Press, Cambridge. MR1744773

PICKANDS, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131. MR0423667

RAGONE, F., WOUTERS, J. and BOUCHET, F. (2018). Computation of extreme heat waves in climate models using a large deviation algorithm. *Proc. Natl. Acad. Sci. USA* **115** 24–29. MR3748914 https://doi.org/10.1073/pnas.1712645115

REICH, B. J., SHABY, B. A. and COOLEY, D. (2014). A hierarchical model for serially-dependent extremes: A study of heat waves in the western US. *J. Agric. Biol. Environ. Stat.* **19** 119–135. MR3257905 https://doi.org/10.1007/s13253-013-0161-y

ROOTZÉN, H., SEGERS, J. and WADSWORTH, J. L. (2018a). Multivariate peaks over thresholds models. *Extremes* **21** 115–145. MR3764613 https://doi.org/10.1007/s10687-017-0294-4

ROOTZÉN, H., SEGERS, J. and WADSWORTH, J. L. (2018b). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *J. Multivariate Anal.* **165** 117–131. MR3768756 https://doi.org/10.1016/j.jmva.2017.12.003

ROOTZÉN, H. and TAJVIDI, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12** 917–930. MR2265668 https://doi.org/10.3150/bj/1161614952

SHABY, B. A. and REICH, B. J. (2012). Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. *Environmetrics* **23** 638–648. MR3019056 https://doi.org/10.1002/env.2178

SHOOTER, R., ROSS, E., TAWN, J. and JONATHAN, P. (2019). On spatial conditional extremes for ocean storm severity. *Environmetrics* **30** e2562, 18. MR4009977 https://doi.org/10.1002/env.2562

STOTT, P. A., STONE, D. A. and ALLEN, M. R. (2004). Human contribution to the European heatwave of 2003. *Nature* **432** 610–614. https://doi.org/10.1038/nature03089

STOTT, P. A., CHRISTIDIS, N., OTTO, F. E., SUN, Y., VANDERLINDEN, J.-P., VAN OLDENBORGH, G. J., VAUTARD, R., VON STORCH, H., WALTON, P. et al. (2016). Attribution of extreme weather and climate-related events. *Wiley Interdiscip. Rev.: Clim. Change* **7** 23–41.

TAJVIDI, N. (1996). Characterisation and some statistical aspects of univariate and multivariate generalized Pareto distributions. Ph.D. thesis, Department of Mathemtics, Chalmers, Göteborg.

ZSCHEISCHLER, J., WESTRA, S., VAN DEN HURK, B. J. J. M., SENEVIRATNE, S. I., WARD, P. J., PITMAN, A., AGHAKOUCHAK, A., BRESCH, D. N., LEONARD, M. et al. (2018). Future climate risk from compound events. *Nat. Clim. Change* **8** 469–477. https://doi.org/10.1038/s41558-018-0156-3

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1359–1378
https://doi.org/10.1214/20-AOAS1347

SPATIOTEMPORAL PROBABILISTIC WIND VECTOR FORECASTING OVER SAUDI ARABIA

BY AMANDA LENZI* AND MARC G. GENTON†

Statistics Program, King Abdullah University of Science and Technology, *lenzi.amanda88@gmail.com; †marc.genton@kaust.edu.sa*

Saudi Arabia has recently begun promoting renewable energy as a potential alternative to fossil fuels for domestic power generation. In order to efficiently connect wind energy to the existing power grids, reliable wind forecasts and an accurate way of quantifying the uncertainties of these forecasts are required. Motivated by a data set of hourly wind speeds from 28 stations in Saudi Arabia, we build spatiotemporal models for short-term probabilistic forecasts of wind vectors. Traditionally, wind speed and wind direction have been considered independently, without taking dependencies into account. However, in many situations, for example, energy management, it is essential to have information on the bivariate nature of the wind. We compare a coregionalization model for the wind vector with a univariate spatiotemporal model for the transformed wind speed in terms of sharpness and calibration. In both cases the linear predictor is a function of covariates, a smooth function to capture the daily seasonality in the wind and a latent Gaussian field to model the spatial and temporal dependencies. Substantial improvements in reliability are observed when modelling the full bivariate structure instead of only considering speed. Furthermore, the bivariate model has the advantage of also producing forecasts for the wind direction. A Bayesian framework is used to obtain forecasts that are accurate and reliable, even at stations without observations, with relatively low computational cost. Simulated high-resolution data from a computer model are used to validate spatiotemporal forecasts. A detailed analysis on this case study shows how increasing the number of locations can improve the forecast performance.

REFERENCES

ACKERMANN and THOMAS (2005). *Wind Power in Power Systems*. Wiley, New York.

ALEXIADIS, M., DOKOPOULOS, P. and SAHSAMANOGLOU, H. (1999). Wind speed and power forecasting based on spatial correlation models. *IEEE Trans. Energy Convers.* **14** 836–842.

BAKKA, H., RUE, H., FUGLSTAD, G.-A., RIEBLER, A., BOLIN, D., ILLIAN, J., KRAINSKI, E., SIMPSON, D. and LINDGREN, F. (2018). Spatial modeling with R-INLA: A review. *Wiley Interdiscip. Rev.: Comput. Stat.* **10** e1443, 24. MR3873676 https://doi.org/10.1002/wics.1443

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. MR3362184

BREMNES, J. B. (2004). Probabilistic wind power forecasts using local quantile regression. *Wind Energy* **7** 47–54.

CAMELETTI, M., LINDGREN, F., SIMPSON, D. and RUE, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Adv. Stat. Anal.* **97** 109–131. MR3045763 https://doi.org/10.1007/s10182-012-0196-3

CHEN, W., CASTRUCCIO, S., GENTON, M. G. and CRIPPA, P. (2018). Current and future estimates of wind energy potential over Saudi Arabia. *J. Geophys. Res., Atmos.* **123** 6443–6459.

CRESSIE, N. (1992). Statistics for spatial data. *Terra Nova* **4** 613–617.

CRESSIE, N. and WIKLE, C. K. (2015). *Statistics for Spatio-Temporal Data*. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2848400

DAWID, A. P. (1984). Present position and potential developments: Some personal views statistical theory the prequential approach. *J. R. Stat. Soc., A* **147** 278–290.

Key words and phrases. Integrated nested Laplace approximation, linear model of coregionalization, probabilistic forecast, spatiotemporal modeling, wind vector.

DIEBOLD, F. X., GUNTHER, T. A. and TAY, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *Internat. Econom. Rev.* **39** 863–883.

DOWELL, J. and PINSON, P. (2016). Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Trans. Smart Grid* **7** 763–770.

DOWELL, J., WEISS, S., HILL, D. and INFIELD, D. (2014). Short-term spatiotemporal prediction of wind speed and direction. *Wind Energy* **17** 1945–1955.

EIDE, S. S., BREMNES, J. B. and STEINSLAND, I. (2017). Bayesian model averaging for wind speed ensemble forecasts using wind speed and direction. *Weather Forecast.* **32** 2217–2227.

EZZAT, A. A., JUN, M. and DING, Y. (2018). Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting. *IEEE Trans. Sustain. Energy* **9** 1437–1447.

FUGLSTAD, G.-A., SIMPSON, D., LINDGREN, F. and RUE, H. (2015). Does non-stationary spatial data always require non-stationary random fields? *Spat. Stat.* **14** 505–531. MR3431054 https://doi.org/10.1016/j.spasta.2015.10.001

FUGLSTAD, G.-A., SIMPSON, D., LINDGREN, F. and RUE, H. (2019). Constructing priors that penalize the complexity of Gaussian random fields. *J. Amer. Statist. Assoc.* **114** 445–452. MR3941267 https://doi.org/10.1080/01621459.2017.1415907

GELFAND, A. E., SCHMIDT, A. M., BANERJEE, S. and SIRMANS, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *TEST* **13** 263–312. MR2154003 https://doi.org/10.1007/BF02595775

GENTON, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics* **18** 681–695. MR2408938 https://doi.org/10.1002/env.854

GENTON, M. G. and KLEIBER, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statist. Sci.* **30** 147–163. MR3353096 https://doi.org/10.1214/14-STS487

GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 243–268. MR2325275 https://doi.org/10.1111/j.1467-9868.2007.00587.x

GRIMIT, E. P., GNEITING, T., BERROCAL, V. and JOHNSON, N. A. (2006). The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Q. J. R. Meteorol. Soc.* **132** 2925–2942.

HASLETT, J. and RAFTERY, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **38** 1–21.

HERING, A. S. and GENTON, M. G. (2010). Powering up with space-time wind forecasting. *J. Amer. Statist. Assoc.* **105** 92–104. MR2757195 https://doi.org/10.1198/jasa.2009.ap08117

HERING, A. S. and GENTON, M. G. (2011). Comparing spatial predictions. *Technometrics* **53** 414–425. MR2850473 https://doi.org/10.1198/TECH.2011.10136

HIJMANS, R. J. (2017). raster: Geographic data analysis and modeling. R package version 2.6-7.

INGEBRIGTSEN, R., LINDGREN, F. and STEINSLAND, I. (2014). Spatial models with explanatory variables in the dependence structure. *Spat. Stat.* **8** 20–38. MR3326819 https://doi.org/10.1016/j.spasta.2013.06.002

JONES, R. H. and ZHANG, Y. (1997). Models for continuous stationary space-time processes. In *Modelling Longitudinal and Spatially Correlated Data* 289–298. Springer, Berlin.

KOUROUNIS, D., FUCHS, A. and SCHENK, O. (2018). Toward the next generation of multiperiod optimal power flow solvers. *IEEE Trans. Power Syst.* **33** 4005–4014.

KRAINSKI, E. T. (2018). Statistical analysis of space-time data: New models and applications. Ph.D. thesis, Norwegian Univ. Science and Technology (NTNU).

KRAINSKI, E. T., GÓMEZ-RUBIO, V., BAKKA, H., LENZI, A., CASTRO-CAMILO, D., SIMPSON, D., LINDGREN, F. and RUE, H. (2019). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. CRC Press, Boca Raton. Github version www.r-inla.org/spde-book.

LAU, A. and MCSHARRY, P. (2010). Approaches for multi-step density forecasts with application to aggregated wind power. *Ann. Appl. Stat.* **4** 1311–1341. MR2758330 https://doi.org/10.1214/09-AOAS320

LENZI, A. and GENTON, M. G. (2020). Supplement to "Spatiotemporal probabilistic wind vector forecasting over Saudi Arabia." https://doi.org/10.1214/20-AOAS1347SUPP

LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. With discussion and a reply by the authors. MR2853727 https://doi.org/10.1111/j.1467-9868.2011.00777.x

MANDIC, D. P., GOH, S. L. and AIHARA, K. (2005). Sequential data fusion via vector spaces: Complex modular neural network approach. In *Machine Learning for Signal Processing*, 2005 *IEEE Workshop on* 147–151.

MATHERON, G. (1982). Pour une analyse krigeante des données régionalisées. Report N-732, Fontainebleau, Centre de Géostatistique.

MOHANDES, M. A., REHMAN, S. and HALAWANI, T. O. (1998). A neural networks approach for wind speed prediction. *Renew. Energy* **13** 345–354.

MOHANDES, M. A., HALAWANI, T. O., REHMAN, S. and HUSSAIN, A. A. (2004). Support vector machines for wind speed prediction. *Renew. Energy* **29** 939–947.

MØLLER, J. K., NIELSEN, H. A. and MADSEN, H. (2008). Time-adaptive quantile regression. *Comput. Statist. Data Anal.* **52** 1292–1303. MR2418566 https://doi.org/10.1016/j.csda.2007.06.027

MUNOZ, F., PENNINO, M. G., CONESA, D., LÓPEZ-QUÍLEZ, A. and BELLIDO, J. M. (2013). Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. *Stoch. Environ. Res. Risk Assess.* **27** 1171–1180.

PINSON, P. (2012). Adaptive calibration of (u, v)-wind ensemble forecasts. *Q. J. R. Meteorol. Soc.* **138** 1273–1284.

PINSON, P. and KARINIOTAKIS, G. (2010). Conditional prediction intervals of wind power generation. *IEEE Trans. Power Syst.* **25** 1845–1856.

PINSON, P. and MADSEN, H. (2009). Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy* **12** 137–155.

PINSON, P. and MADSEN, H. (2012). Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models. *J. Forecast.* **31** 281–313. MR2924797 https://doi.org/10.1002/for.1194

PINSON, P., MCSHARRY, P. and MADSEN, H. (2010). Reliability diagrams for non-parametric density forecasts of continuous variables: Accounting for serial correlation. *Q. J. R. Meteorol. Soc.* **136** 77–90.

REHMAN, S. and AHMAD, A. (2004). Assessment of wind energy potential for coastal locations of the Kingdom of Saudi Arabia. *Energy* **29** 1105–1115.

REHMAN, S., EL-AMIN, I., AHMAD, F., SHAAHID, S., AL-SHEHRI, A. and BAKHASHWAIN, J. (2007). Wind power resource assessment for Rafha, Saudi Arabia. *Renew. Sustain. Energy Rev.* **11** 937–950.

RIENECKER, M. M., SUAREZ, M. J., GELARO, R., TODLING, R., BACMEISTER, J., LIU, E., BOSILOVICH, M. G., SCHUBERT, S. D., TAKACS, L. et al. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Climate* **24** 3624–3648.

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

SCHMIDT, A. M. and GELFAND, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *J. Geophys. Res., Atmos.* **108**.

SCHUHEN, N., THORARINSDOTTIR, T. L. and GNEITING, T. (2012). Ensemble model output statistics for wind vectors. *Mon. Weather Rev.* **140** 3204–3219.

SHAAHID, S., AL-HADHRAMI, L. M. and RAHMAN, M. (2014). Potential of establishment of wind farms in western province of Saudi Arabia. *Energy Proc.* **52** 497–505.

SIMPSON, D., RUE, H., RIEBLER, A., MARTINS, T. G. and SØRBYE, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statist. Sci.* **32** 1–28. MR3634300 https://doi.org/10.1214/16-STS576

SLOUGHTER, J., GNEITING, T. and RAFTERY, A. E. (2013). Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Mon. Weather Rev.* **141** 2107–2119.

TAGLE, F., CASTRUCCIO, S. and GENTON, M. G. (2020). A hierarchical bi-resolution spatial skew-t model. *Spat. Stat.* **35** 100398.

VAN NIEKERK, J., BAKKA, H., RUE, H. and SCHENK, L. (2019). New frontiers in Bayesian modeling using the INLA package in R. Preprint. Available at arXiv:1907.10426.

VISION2030 (2018). Available at http://vision2030.gov.sa/en/node/87. Accessed: 2018-07-23.

WEISSER, D. and FOXON, T. (2003). Implications of seasonal and diurnal variations of wind velocity for power output estimation of a turbine: A case study of Grenada. *Int. J. Energy Res.* **27** 1165–1179.

YIP, C. M. A. (2018). Statistical characteristics and mapping of near-surface and elevated wind resources in the Middle East. Ph.D. thesis, King Abdullah Univ. Science and Technology (KAUST).

YIP, C. M. A., GUNTURU, U. B. and STENCHIKOV, G. L. (2016). Wind resource characterization in the Arabian Peninsula. *Appl. Energy* **164** 826–836.

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1379–1408
https://doi.org/10.1214/19-AOAS1321

QUANTIFYING TIME-VARYING SOURCES IN MAGNETOENCEPHALOGRAPHY—A DISCRETE APPROACH

BY ZHIGANG YAO[1,*], ZENGYAN FAN[1,†], MASAHITO HAYASHI[2] AND WILLIAM F. EDDY[3]

[1]*Department of Statistics and Applied Probability, National University of Singapore,* [*]*zhigang.yao@nus.edu.sg;* [†]*stafz@nus.edu.sg*

[2]*Graduate School of Mathematics, Nagoya University, masahito@math.nagoya-u.ac.jp*

[3]*Department of Statistics, Carnegie Mellon University, bill@stat.cmu.edu*

We study the distribution of brain source from the most advanced brain imaging technique, Magnetoencephalography (MEG) which measures the magnetic fields outside of the human head produced by the electrical activity inside the brain. Common time-varying source localization methods assume the source current with a time-varying structure and solve the MEG inverse problem by mainly estimating the source moment parameters. These methods use the fact that the magnetic fields linearly depend on the moment parameters of the source and work well under the linear dynamic system. However, magnetic fields are known to be nonlinearly related to the location parameters of the source. The existing work on estimating the time-varying unknown location parameters is limited. We are motivated to investigate the source distribution for the location parameters based on a dynamic framework, where the posterior distribution of the source is computed in a closed form discretely. The new framework allows us not only to directly approximate the posterior distribution of the source current, where sequential sampling methods may suffer from slow convergence due to the large volume of measurement, but also to quantify the source distribution at any time point from the entire set of measurements reflecting the distribution of the source, rather than using only the measurements up to the time point of interest. Both a dynamic procedure and a switch procedure are pro- posed for the new discrete approach, balancing estimation accuracy and computational efficiency when multiple sources are present. In both simulation and real data, we illustrate that the new method is able to provide comprehensive insight into the time evolution of the sources at different stages of the MEG and EEG experiment.

REFERENCES

ARULAMPALAM, M. S., MASKELL, S., GORDON, N. and CLAPP, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **50** 174–188.

BAILLET, S. and GARNERO, L. (1997). A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. *IEEE Trans. Biomed. Eng.* **44** 374–385.

BAILLET, S., MOSHER, J. C. and LEAHY, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Process. Mag.* **18** 14–30.

BOTO, E., HOLMES, N., LEGGETT, J., ROBERTS, G., SHAH, V., MEYER, S. S., MUÑOZ, L. D., MULLINGER, K. J., TIERNEY, T. M. et al. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. *Nature* **555** 657.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537

DESTEXHE, A., CONTRERAS, D. and STERIADE, M. (1999). Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. *J. Neurosci.* **19** 4595–4608.

FELZENSZWALB, P. F., HUTTENLOCHER, D. P. and KLEINBERG, J. M. (2004). Fast algorithms for large-state-space HMMs with applications to web usage analysis. In *Advances in Neural Information Processing Systems* 409–416.

Key words and phrases. MEG inverse problem, discrete posterior distribution, expectation-maximization, spatiotemporal model, source localization.

FUKUSHIMA, M., YAMASHITA, O., KNÖSCHE, T. R. and SATO, M. (2015). MEG source reconstruction based on identification of directed source interactions on whole-brain anatomical networks. *NeuroImage* **105** 408–427.

HÄMÄLÄINEN, M. S. and ILMONIEMI, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Med. Biol. Eng. Comput.* **32** 35–42. https://doi.org/10.1007/bf02512476

HÄMÄLÄINEN, M. S., HARI, R., ILMONIEMI, R. J., KNUUTILA, J. and LOUNASMAA, O. V. (1993). Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Modern Phys.* **65** 413–497. https://doi.org/10.1103/revmodphys.65.413

LAMUS, C., HÄMÄLÄINEN, M. S., TEMEREANCA, S., BROWN, E. N. and PURDON, P. L. (2012). A spatiotemporal dynamic distributed solution to the MEG inverse problem. *NeuroImage* **63** 894–909.

LIN, F.-H., WITZEL, T., AHLFORS, S. P., STUFFLEBEAM, S. M., BELLIVEAU, J. W. and HÄMÄLÄINEN, M. S. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage* **31** 160–171.

LIU, J. S. and CHEN, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93** 1032–1044. MR1649198 https://doi.org/10.2307/2669847

LIU, S., POH, J.-H., KOH, H. L., NG, K. K., LOKE, Y. M., LIM, J. K. W., CHONG, J. S. X. and ZHOU, J. (2018). Carrying the past to the future: Distinct brain networks underlie individual differences in human spatial working memory capacity. *NeuroImage* **176** 1–10.

LONG, C. J., PURDON, P. L., TEMEREANCA, S., DESAI, N. U., HÄMÄLÄINEN, M. S. and BROWN, E. N. (2011). State-space solutions to the dynamic magnetoencephalography inverse problem using high performance computing. *Ann. Appl. Stat.* **5** 1207–1228. MR2849772 https://doi.org/10.1214/11-AOAS483

MOSHER, J. C., LEWIS, P. S. and LEAHY, R. M. (1992). Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Trans. Biomed. Eng.* **39** 541–557.

OU, W., HÄMÄLÄINEN, M. S. and GOLLAND, P. (2009). A distributed spatio-temporal EEG/MEG inverse solver. *NeuroImage* **44** 932–946.

PASCUAL-MARQUI, R. D., MICHEL, C. M. and LEHMANN, D. (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* **18** 49–65. https://doi.org/10.1016/0167-8760(84)90014-x

RABINER, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77** 257–286.

SARVAS, J. (1984). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.* **32** 11–12.

TRUJILLO-BARRETO, N. J., AUBERT-VÁZQUEZ, E. and PENNY, W. D. (2008). Bayesian M/EEG source reconstruction with spatio-temporal priors. *NeuroImage* **39** 318–335.

UUTELA, K., HÄMÄLÄINEN, M. and SOMERSALO, E. (1999). Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage* **10** 173–180.

VAIDA, F. (2005). Parameter convergence for EM and MM algorithms. *Statist. Sinica* **15** 831–840. MR2233916

VEEN, B., JOSEPH, J. and HECOX, K. (1992). Localization of intra-cerebral sources of electrical activity via linearly constrained minimum variance spatial filtering. In *Proceedings of IEEE Workshop on Statistical Signal and Array Processing* 1 526–529.

YAO, Z. and EDDY, W. F. (2014). A statistical approach to the inverse problem in magnetoencephalography. *Ann. Appl. Stat.* **8** 1119–1144. MR3262548 https://doi.org/10.1214/14-AOAS716

YAO, Z., ZHANG, Y., BAI, Z. and EDDY, W. F. (2018). Estimating the number of sources in magnetoencephalography using spiked population eigenvalues. *J. Amer. Statist. Assoc.* **113** 505–518. MR3832204 https://doi.org/10.1080/01621459.2017.1341411

YAO, Z., FAN, Z., HAYASHI, M. and EDDY, W. F. (2020). Supplement to "Quantifying time-varying sources in magnetoencephalography—A discrete approach." https://doi.org/10.1214/19-AOAS1321SUPP

ZHANG, J. and LIU, C. (2015). On linearly constrained minimum variance beamforming. *J. Mach. Learn. Res.* **16** 2099–2145. MR3450503 https://doi.org/10.1631/jzus.a1400270

ZHANG, J. and SU, L. (2015). Temporal autocorrelation-based beamforming with MEG neuroimaging data. *J. Amer. Statist. Assoc.* **110** 1375–1388. MR3449033 https://doi.org/10.1080/01621459.2015.1054488

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1409–1431
https://doi.org/10.1214/20-AOAS1356

DOUBLY ROBUST TREATMENT EFFECT ESTIMATION WITH MISSING ATTRIBUTES

BY IMKE MAYER[1], ERIK SVERDRUP[2,*], TOBIAS GAUSS[3,‡], JEAN-DENIS MOYER[3,§], STEFAN WAGER[2,†] AND JULIE JOSSE[4]

[1]*Centre d'Analyse et de Mathématique Sociales, École des Hautes Études en Sciences Sociales, imke.mayer@ehess.fr*
[2]*Graduate School of Business, Stanford University, [*]erikcs@stanford.edu; [†]swager@stanford.edu*
[3]*Department of Anesthesia and Intensive Care, Beaujon Hospital, [‡]tgauss@protonmail.com; [§]jean-denis.moyer@aphp.fr*
[4]*Centre de Mathématiques Appliquées École Polytechnique, Institut Polytechnique de Paris, julie.josse@polytechnique.edu*

Missing attributes are ubiquitous in causal inference, as they are in most applied statistical work. In this paper we consider various sets of assumptions under which causal inference is possible despite missing attributes and discuss corresponding approaches to average treatment effect estimation, including generalized propensity score methods and multiple imputation. Across an extensive simulation study, we show that no single method systematically outperforms others. We find, however, that doubly robust modifications of standard methods for average treatment effect estimation with missing data repeatedly perform better than their nondoubly robust baselines; for example, doubly robust generalized propensity score methods beat inverse-weighting with the generalized propensity score. This finding is reinforced in an analysis of an observational study on the effect on mortality of tranexamic acid administration among patients with traumatic brain injury in the context of critical care management. Here, doubly robust estimators recover confidence intervals that are consistent with evidence from randomized trials, whereas nondoubly robust estimators do not.

REFERENCES

ABADIE, A. and IMBENS, G. W. (2016). Matching on the estimated propensity score. *Econometrica* **84** 781–807. MR3481379 https://doi.org/10.3982/ECTA11293

ATHEY, S., IMBENS, G. W. and WAGER, S. (2018). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 597–623. MR3849336 https://doi.org/10.1111/rssb.12268

ATHEY, S., TIBSHIRANI, J. and WAGER, S. (2019a). Generalized random forests. *Ann. Statist.* **47** 1148–1178. MR3909963 https://doi.org/10.1214/18-AOS1709

ATHEY, S. and WAGER, S. (2017). Efficient policy learning. ArXiv preprint. Available at arXiv:1702.02896.

ATHEY, S. and WAGER, S. (2019). Estimating treatment effects with causal forests: An application. Observational Studies 5.

BLAKE, H. A., LEYRAT, C., MANSFIELD, K. E., TOMLINSON, L. A., CARPENTER, J. and WILLIAMSON, E. J. (2020). Estimating treatment effects with partially observed covariates using outcome regression with missing indicators. *Biom. J.* **62** 428–443. https://doi.org/10.1002/bimj.201900041

BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.

CANDÈS, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proc. IEEE* **98** 925–936.

CAP, A. P. (2019). CRASH-3: A win for patients with traumatic brain injury. *Lancet*.

CARPENTER, J. and KENWARD, M. (2013). *Multiple Imputation and Its Application*. Wiley, Chichester, West Sussex, UK.

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

CRUMP, R. K., HOTZ, V. J., IMBENS, G. W. and MITNIK, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96** 187–199. MR2482144 https://doi.org/10.1093/biomet/asn055

Key words and phrases. Missing data, causal inference, potential outcomes, observational data, propensity score estimation, incomplete confounders, major trauma, public health.

D'AGOSTINO, R. B. JR and RUBIN, D. B. (2000). Estimating and using propensity scores with partially missing data. *J. Amer. Statist. Assoc.* **95** 749–759.

DALKEY, N. and HELMER, O. (1963). An experimental application of the Delphi method to the use of experts. *Manage. Sci.* **9** 458–467.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537

FARRELL, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *J. Econometrics* **189** 1–23. MR3397349 https://doi.org/10.1016/j.jeconom.2015.06.017

HASTIE, T. and MAZUMDER, R. (2015). softImpute: Matrix Completion via Iterative Soft-Thresholded SVD. R package version 1.4.

HAY, S. I., ABAJOBIR, A. A., ABATE, K. H., ABBAFATI, C., ABBAS, K. M., ABD-ALLAH, F., ABDULKADER, R. S., ABDULLE, A. M., ABEBO, T. A. et al. (2017). Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: A systematic analysis for the global burden of disease study 2016. *Lancet* **390** 1260–1344.

HIJAZI, N., FANNE, R. A., ABRAMOVITCH, R., YAROVOI, S., HIGAZI, M., ABDEEN, S., BASHEER, M., MARAGA, E., CINES, D. B. et al. (2015). Endogenous plasminogen activators mediate progressive intracerebral hemorrhage after traumatic brain injury in mice. *Blood* **125** 2558–2567.

HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189. MR1995826 https://doi.org/10.1111/1468-0262.00442

IMBENS, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econ. Stat.* **86** 4–29.

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge Univ. Press, Cambridge.

JIANG, W. (2019). misaem: Logistic Regression with Missing Covariates. R package version 0.9.1.

JIANG, W., JOSSE, J., LAVIELLE, M. and GROUP, T. (2020). Logistic regression with missing covariates—parameter estimation, model selection and prediction within a joint-modeling framework. *Comput. Statist. Data Anal.* **145** 106907. MR4053706 https://doi.org/10.1016/j.csda.2019.106907

JONES, J. and HUNTER, D. (1995). Consensus methods for medical and health services research. *BMJ* **311** 376–380. https://doi.org/10.1136/bmj.311.7001.376

JOSSE, J., PROST, N., SCORNET, E. and VAROQUAUX, G. (2019). On the consistency of supervised learning with missing values. ArXiv preprint.

KALLUS, N., MAO, X. and UDELL, M. (2018). Causal inference with noisy and missing covariates via matrix factorization. In *Advances in Neural Information Processing Systems* 6921–6932.

KINGMA, D. P. and WELLING, M. (2014). Stochastic gradient VB and the variational auto-encoder. In *Second International Conference on Learning Representations*, *ICLR*.

KITAGAWA, T. and TETENOV, A. (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica* **86** 591–616. MR3783340 https://doi.org/10.3982/ECTA13288

LEDERER, D. J., BELL, S. C., BRANSON, R. D., CHALMERS, J. D., MARSHALL, R., MASLOVE, D. M., OST, D. E., PUNJABI, N. M., SCHATZ, M. et al. (2019). Control of confounding and reporting of results in causal inference studies. Guidance for authors from editors of respiratory, sleep, and critical care journals. *Ann. Amer. Thorac. Soc.* **16** 22–28.

LEYRAT, C., SEAMAN, S. R., WHITE, I. R., DOUGLAS, I., SMEETH, L., KIM, J., RESCHE-RIGON, M., CARPENTER, J. R. and WILLIAMSON, E. J. (2019). Propensity score analysis with partially observed covariates: How should multiple imputation be used? *Stat. Methods Med. Res.* **28** 3–19. MR3894510 https://doi.org/10.1177/0962280217713032

LI, F., MORGAN, K. L. and ZASLAVSKY, A. M. (2018). Balancing covariates via propensity score weighting. *J. Amer. Statist. Assoc.* **113** 390–400. MR3803473 https://doi.org/10.1080/01621459.2016.1260466

LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1925014 https://doi.org/10.1002/9781119013563

LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist.* **44** 713–742. MR3476615 https://doi.org/10.1214/15-AOS1384

MATTEI, A. (2009). Estimating and using propensity score in presence of missing background data: An application to assess the impact of childbearing on wellbeing. *Stat. Methods Appl.* **18** 257–273. MR2515482 https://doi.org/10.1007/s10260-007-0086-0

MAYER, I., SVERDRUP, E., GAUSS, T., MOYER, J.-D, WAGER, S. and JOSSE, J. (2020). Supplement to "Doubly robust treatment effect estimation with missing attributes." https://doi.org/10.1214/20-AOAS1356SUPPA, https://doi.org/10.1214/20-AOAS1356SUPPB

PEARL, J. (1995). Causal diagrams for empirical research. *Biometrika* **82** 669–710. MR1380809 https://doi.org/10.1093/biomet/82.4.669

QU, Y. and LIPKOVICH, I. (2009). Propensity score estimation with missing values using a multiple imputation missingness pattern (MIMP) approach. *Stat. Med.* **28** 1402–1414. MR2667534 https://doi.org/10.1002/sim.3549

R CORE TEAM (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

RICHARDSON, T. S. and ROBINS, J. M. (2013). Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality Technical Report Center for Statistics and the Social Sciences, Univ. Washington.

ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

ROBINS, J. M. and WANG, N. (2000). Inference for imputation estimators. *Biometrika* **87** 113–124. MR1766832 https://doi.org/10.1093/biomet/87.1.113

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

ROSENBAUM, P. R. and RUBIN, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *J. Amer. Statist. Assoc.* **79** 516–524.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196 https://doi.org/10.1093/biomet/63.3.581

RUBIN, D. B. (1978). Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. In *Proceedings of the Survey Research Methods Section of the American Statistical Association* **1** 20–34. Amer. Statist. Assoc..

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR0899519 https://doi.org/10.1002/9780470316696

RUBIN, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys*. *Wiley Classics Library*. Wiley Interscience, Hoboken, NJ. MR2117498

SCHAFER, J. L. (1997). *Analysis of Incomplete Multivariate Data*. *Monographs on Statistics and Applied Probability* **72**. CRC Press, London. MR1692799 https://doi.org/10.1201/9781439821862

SCORNET, E., BIAU, G. and VERT, J.-P. (2015). Consistency of random forests. *Ann. Statist.* **43** 1716–1741. MR3357876 https://doi.org/10.1214/15-AOS1321

SEAMAN, S. and WHITE, I. (2014). Inverse probability weighting with missing predictors of treatment assignment or missingness. *Comm. Statist. Theory Methods* **43** 3499–3515. MR3239323 https://doi.org/10.1080/03610926.2012.700371

SHAKUR, H., ROBERTS, I., BAUTISTA, R., CABALLERO, J., COATS, T., DEWAN, Y., EL-SAYED, H., GOGICHAISHVILI, T., GUPTA, S. et al. (2010). CRASH-2 trial collaborators. Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): A randomised, placebo-controlled trial. *Lancet* **376** 23–32.

TEXTOR, J., HARDT, J. and KNÜPPEL, S. (2011). DAGitty: A graphical tool for analyzing causal diagrams. *Epidemiology* **22** 745. https://doi.org/10.1097/EDE.0b013e318225c2be

TIBSHIRANI, J., ATHEY, S., FRIEDBERG, R., HADAD, V., HIRSHBERG, D., MINER, L., SVERDRUP, E., WAGER, S. and WRIGHT, M. (2020). grf: Generalized Random Forests. R package version 1.1.0.

TWALA, B., JONES, M. and HAND, D. J. (2008). Good methods for coping with missing data in decision trees. *Pattern Recogn. Lett.* **29** 950–956.

VAN BUUREN, S. (2018). *Flexible Imputation of Missing Data*, 2nd ed. CRC Press/CRC, Boca Raton, FL.

VAN BUUREN, S. and GROOTHUIS-OUDSHOORN, K. (2011). mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **45** 1–67.

VAN DER LAAN, M. J. and ROSE, S. (2011). *Targeted Learning*: *Causal Inference for Observational and Experimental Data*. *Springer Series in Statistics*. Springer, New York. MR2867111 https://doi.org/10.1007/978-1-4419-9782-1

WAGER, S. and WALTHER, G. (2015). Adaptive concentration of regression trees, with application to random forests. ArXiv preprint. Available at arXiv:1503.06388.

YANG, S., WANG, L. and DING, P. (2019). Causal inference with confounders missing not at random. *Biometrika* **106** 875–888. MR4031203 https://doi.org/10.1093/biomet/asz048

ZHAO, Y., ZENG, D., RUSH, A. J. and KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107** 1106–1118. MR3010898 https://doi.org/10.1080/01621459.2012.695674

ZUBIZARRETA, J. R. (2012). Using mixed integer programming for matching in an observational study of kidney failure after surgery. *J. Amer. Statist. Assoc.* **107** 1360–1371. MR3036400 https://doi.org/10.1080/01621459.2012.703874

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1432–1448
https://doi.org/10.1214/19-AOAS1314

CAUSAL INFERENCE FROM OBSERVATIONAL STUDIES WITH CLUSTERED INTERFERENCE, WITH APPLICATION TO A CHOLERA VACCINE STUDY

BY BRIAN G. BARKLEY[1], MICHAEL G. HUDGENS[2], JOHN D. CLEMENS[3], MOHAMMAD ALI[4] AND MICHAEL E. EMCH[5]

[1]*Department of Marketing Analytics, Kohl's, Inc., barkleybg@outlook.com*

[2]*Department of Biostatistics, University of North Carolina at Chapel Hill, mhudgens@email.unc.edu*

[3]*Office of the Executive Director, icddr,b, jclemens@icddrb.org*

[4]*Department of International Health, Johns Hopkins University, mali25@jhu.edu*

[5]*Department of Geography, University of North Carolina at Chapel Hill, emch@unc.edu*

Understanding the population-level effects of vaccines has important public health policy implications. Inferring vaccine effects from an observational study is challenging because participants are not randomized to vaccine (i.e., treatment). Observational studies of infectious diseases present the additional challenge that vaccinating one participant may affect another participant's outcome, that is, there may be interference. In this paper recent approaches to defining vaccine effects in the presence of interference are considered, and new causal estimands designed specifically for use with observational studies are proposed. Previously defined estimands target counterfactual scenarios in which individuals independently choose to be vaccinated with equal probability. However, in settings where there is interference between individuals within clusters, it may be unlikely that treatment selection is independent between individuals in the same cluster. The proposed causal estimands instead describe counterfactual scenarios which allow for within-cluster dependence in the individual treatment selections. These estimands may be more relevant for policy-makers or public health officials who desire to quantify the effect of increasing the proportion of vaccinated individuals in a population. Inverse probability-weighted estimators for these estimands are proposed. The large-sample properties of the estimators are derived, and a simulation study demonstrating the finite-sample performance of the estimators is presented. The proposed methods are illustrated by analyzing data from a study of cholera vaccination in over 100,000 individuals in Bangladesh.

REFERENCES

ALI, M., EMCH, M. E., VON SEIDLEIN, L., YUNUS, M., SACK, D. A., RAO, M., HOLMGREN, J. and CLEMENS, J. D. (2005). Herd immunity conferred by killed oral cholera vaccines in Bangladesh: A reanalysis. *Lancet* **366** 44–49.

ALI, M., EMCH, M. E., YUNUS, M. and CLEMENS, J. D. (2009). Modeling spatial heterogeneity of disease risk and evaluation of the impact of vaccination. *Vaccine* **27** 3724–3729.

ALI, M., KIM, D. R., KANUNGO, S., SUR, D., MANNA, B., DIGILIO, L., DUTTA, S., MARKS, F., BHATTACHARYA, S. K. et al. (2018). Use of oral cholera vaccine as a vaccine probe to define the geographical dimensions of person-to-person transmission of cholera. *Int. J. Infect. Dis.* **66** 90–95. https://doi.org/10.1016/j.ijid.2017.11.020

BARKLEY, B. G. (2019). clusteredinterference: Causal effects from observational stud- ies with clustered interference. R package version 1.0.1. Available at https://CRAN.R-project.org/package=clusteredinterference.

BARKLEY, B. G., HUDGENS, M. G., CLEMENS, J. D., ALI, M. and EMCH, M. E. (2020). Supplement to "Causal inference from observational studies with clustered interference, with application to a cholera vaccine study." https://doi.org/10.1214/19-AOAS1314SUPP

Key words and phrases. Causal inference, interference, inverse probability-weight, observational study, propensity score, spillover effects.

BATES, D., MÄCHLER, M., BOLKER, B. and WALKER, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67** 1–48. https://doi.org/10.18637/jss.v067.i01

CHEN, T., ZENG, D. and WANG, Y. (2015). Multiple kernel learning with random effects for predicting longitudinal outcomes and data integration. *Biometrics* **71** 918–928. MR3436717 https://doi.org/10.1111/biom.12343

COX, D. R. (1958). *Planning of Experiments. A Wiley Publication in Applied Statistics*. Wiley, New York. MR0095561

DÍAZ, I. D. and VAN DER LAAN, M. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics* **68** 541–549. MR2959621 https://doi.org/10.1111/j.1541-0420.2011.01685.x

ECK, D. J., MOROZOVA, O. and CRAWFORD, F. W. (2018). Randomization for the direct effect of an infectious disease intervention in a clustered study population. Preprint. Available at arXiv:1808.05593.

FITZMAURICE, G. M., LAIRD, N. M. and WARE, J. H. (2011). *Applied Longitudinal Analysis*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2830137

HALLORAN, M. E. and STRUCHINER, C. J. (1995). Causal inference in infectious diseases. *Epidemiology* **6** 142–151.

HONG, G. and RAUDENBUSH, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *J. Amer. Statist. Assoc.* **101** 901–910. MR2324091 https://doi.org/10.1198/016214506000000447

HUDGENS, M. G. and HALLORAN, M. E. (2008). Toward causal inference with interference. *J. Amer. Statist. Assoc.* **103** 832–842. MR2435472 https://doi.org/10.1198/016214508000000292

LIU, L., HUDGENS, M. G. and BECKER-DREPS, S. (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika* **103** 829–842. MR3620442 https://doi.org/10.1093/biomet/asw047

MCCULLOCH, C. E. and NEUHAUS, J. M. (2011). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statist. Sci.* **26** 388–402. MR2917962 https://doi.org/10.1214/11-STS361

PAPADOGEORGOU, G., MEALLI, F. and ZIGLER, C. M. (2019). Causal inference with interfering units for cluster and population level treatment allocation programs. *Biometrics* **75** 778–787. MR4012083 https://doi.org/10.1111/biom.13049

PEREZ-HEYDRICH, C., HUDGENS, M. G., HALLORAN, M. E., CLEMENS, J. D., ALI, M. and EMCH, M. E. (2014). Assessing effects of cholera vaccination in the presence of interference. *Biometrics* **70** 731–741. MR3261790 https://doi.org/10.1111/biom.12184

SAUL, B. C. and HUDGENS, M. G. (2017). A recipe for inferference: Start with causal inference. Add interference. Mix well with R. *J. Stat. Softw.* **82** 1–21. https://doi.org/10.18637/jss.v082.i02

SOBEL, M. E. (2006). What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *J. Amer. Statist. Assoc.* **101** 1398–1407. MR2307573 https://doi.org/10.1198/016214506000000636

STEFANSKI, L. A. and BOOS, D. D. (2002). The calculus of M-estimation. *Amer. Statist.* **56** 29–38. MR1939394 https://doi.org/10.1198/000313002753631330

TCHETGEN, E. J. and COULL, B. A. (2006). A diagnostic test for the mixing distribution in a generalised linear mixed model. *Biometrika* **93** 1003–1010. MR2285086 https://doi.org/10.1093/biomet/93.4.1003

TCHETGEN TCHETGEN, E. J. and VANDERWEELE, T. J. (2012). On causal inference in the presence of interference. *Stat. Methods Med. Res.* **21** 55–75. MR2867538 https://doi.org/10.1177/0962280210386779

VANDERWEELE, T. J. and SHPITSER, I. (2011). A new criterion for confounder selection. *Biometrics* **67** 1406–1413. MR2872391 https://doi.org/10.1111/j.1541-0420.2011.01619.x

VANDERWEELE, T. J. and TCHETGEN TCHETGEN, E. J. (2011). Effect partitioning under interference in two-stage randomized vaccine trials. *Statist. Probab. Lett.* **81** 861–869. MR2793754 https://doi.org/10.1016/j.spl.2011.02.019

WESTREICH, D. (2017). From patients to policy: Population intervention effects in epidemiology. *Epidemiology* **28** 525–528. https://doi.org/10.1097/EDE.0000000000000648

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1449–1470
https://doi.org/10.1214/20-AOAS1334

A BAYESIAN HIERARCHICAL MODEL FOR EVALUATING FORENSIC FOOTWEAR EVIDENCE

BY NEIL A. SPENCER[1] AND JARED S. MURRAY[2]

[1]*Department of Statistics and Data Science, and Machine Learning Department, Carnegie Mellon University,* *nspencer@andrew.cmu.edu*

[2]*Department of Information, Risk, and Operations Management and Department of Statistics and Data Science, University of Texas at Austin,* *jared.murray@mccombs.utexas.edu*

When a latent shoeprint is discovered at a crime scene, forensic analysts inspect it for distinctive patterns of wear such as scratches and holes (known as accidentals) on the source shoe's sole. If its accidentals correspond to those of a suspect's shoe, the print can be used as forensic evidence to place the suspect at the crime scene. The strength of this evidence depends on the random match probability—the chance that a shoe chosen at random would match the crime scene print's accidentals. Evaluating random match probabilities requires an accurate model for the spatial distribution of accidentals on shoe soles. A recent report by the President's Council of Advisors in Science and Technology criticized existing models in the literature, calling for new empirically validated techniques. We respond to this request with a new spatial point process model (code and synthetic data is available as Supplementary Material) for accidental locations, developed within a hierarchical Bayesian framework. We treat the tread pattern of each shoe as a covariate, allowing us to pool information across large heterogeneous databases of shoes. Existing models ignore this information; our results show that including it leads to significantly better model fit. We demonstrate this by fitting our model to one such database.

REFERENCES

ADAIR, T. W., LEMAY, J., MCDONALD, A., SHAW, R. and TEWES, R. (2007). The Mount Bierstadt study: An experiment in unique damage formation in footwear. *J. Forensic Identif.* **57** 199–205.

BODZIAK, W. J. (2017). *Forensic Footwear Evidence*. CRC Press, Boca Raton, FL.

BODZIAK, W. J., HAMMER, L., JOHNSON, G. M. and SCHENCK, R. (2012). Determining the significance of outsole wear characteristics during the forensic examination of footwear impression evidence. *J. Forensic Identif.* **62** 254–278.

BROOKS, S., GELMAN, A., JONES, G. L. and MENG, X.-L., eds. (2011). *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR2742422 https://doi.org/10.1201/b10905

CHAMPOD, C., EVETT, I. W. and JACKSON, G. (2004). Establishing the most appropriate databases for addressing source level propositions. *Sci. Justice* **44** 153–164.

CHEN, M.-H., DEY, D. K., MÜLLER, P., SUN, D. and YE, K., eds. (2010). *Frontiers of Statistical Decision Making and Bayesian Analysis*: *In Honor of James O. Berger*. Springer, New York. MR2766461 https://doi.org/10.1007/978-1-4419-6944-6

CSAFE (2019). Longitudinal shoe outsole impression study. Center for Statistics and Applications in Forensic Evidence. Available at http://forensicstats.org/shoeoutsoleimpressionstudy/.

DALEY, D. J. and VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes*, *Vol. II*: *General Theory and Structure*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR2371524 https://doi.org/10.1007/978-0-387-49835-5

DAMARY, N. K., MANDEL, M., WIESNER, S., YEKUTIELI, Y., SHOR, Y. and SPIEGELMAN, C. (2018). Dependence among randomly acquired characteristics on shoeprints and their features. *Forensic Sci. Int.* **283** 173–179.

Key words and phrases. Random match probability, forensic statistics, forensic footwear analysis, random measures, hierarchical Bayes, point processes, spatial statistics.

EDWARDS, H. T. and GOTSONIS, C. (2009). Strengthening forensic science in the United States: A path forward. Statement before the United States Senate Committee on the Judiciary.

ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. MR1340510

EVETT, I. W., LAMBERT, J. A. and BUCKLETON, J. S. (1998). A Bayesian approach to interpreting footwear marks in forensic casework. *Sci. Justice* **38** 241–247.

FACEY, O. E., HANNAH, I. D. and ROSEN, D. (1992). Shoe wear patterns and pressure distribution under feet and shoes, determined by image analysis. *J. Forensic Sci. Soc.* **32** 15–25. https://doi.org/10.1016/s0015-7368(92)73042-4

FOTI, N. J. and WILLIAMSON, S. A. (2015). A survey of non-exchangeable priors for Bayesian nonparametric models. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 359–371.

FRUCHTENICHT, T. L., HERZIG, W. P. and BLACKLEDGE, R. D. (2002). The discrimination of two-dimensional military boot impressions based on wear patterns. *Sci. Justice* **42** 97–104.

GRIFFIN, J. E. and LEISEN, F. (2017). Compound random measures and their use in Bayesian non-parametrics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 525–545. MR3611758 https://doi.org/10.1111/rssb.12176

GRIFFIN, J. and LEISEN, F. (2018). Modelling and computation using NCoRM mixtures for density regression. *Bayesian Anal.* **13** 897–916. MR3807871 https://doi.org/10.1214/17-BA1072

HERMAN, M. (2016). Measurements and scoring procedures for footwear impression comparisons. National Institute for Standards in Technology.

HJORT, N. L., HOLMES, C., MÜLLER, P. and WALKER, S. G. (2010). *Bayesian Nonparametrics. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

KINGMAN, J. F. C. (1967). Completely random measures. *Pacific J. Math.* **21** 59–78. MR0210185

KONG, B., RAMANAN, D. and FOWLKES, C. (2017). Cross-domain forensic shoeprint matching. In *British Machine Vision Conference* (*BMVC*).

MURRAY, I., ADAMS, R. and MACKAY, D. (2010). Elliptical slice sampling. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* 541–548.

NEAL, R. M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. MR1994729 https://doi.org/10.1214/aos/1056562461

PCAST (2016). Report to the President, Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods. President's Council of Advisors on Science and Technology.

PETRACO, N. D. K., GAMBINO, C., KUBIC, T. A., OLIVIO, D. and PETRACO, N. (2010). Statistical discrimination of footwear: A method for the comparison of accidentals on shoe outsoles inspired by facial recognition techniques. *J. Forensic Sci.* **55** 34–41. https://doi.org/10.1111/j.1556-4029.2009.01209.x

RASMUSSEN, C. E. (2000). The infinite Gaussian mixture model. In *Advances in Neural Information Processing Systems* 554–560.

RICHETELLI, N., LEE, M. C., LASKY, C. A., GUMP, M. E. and SPEIR, J. A. (2017a). Classification of footwear outsole patterns using Fourier transform and local interest points. *Forensic Sci. Int.* **275** 102–109.

RICHETELLI, N., NOBEL, M., BODZIAK, W. J. and SPEIR, J. A. (2017b). Quantitative assessment of similarity between randomly acquired characteristics on high quality exemplars and crime scene impressions via analysis of feature size and shape. *Forensic Sci. Int.* **270** 211–222. https://doi.org/10.1016/j.forsciint.2016.10.008

SHEETS, H. D., GROSS, S., LANGENBURG, G., BUSH, P. J. and BUSH, M. A. (2013). Shape measurement tools in footwear analysis: A statistical investigation of accidental characteristics over time. *Forensic Sci. Int.* **232** 84–91. https://doi.org/10.1016/j.forsciint.2013.07.010

SHOR, Y., WIESNER, S., TSACH, T., GUREL, R. and YEKUTIELI, Y. (2018). Inherent variation in multiple shoe-sole test impressions. *Forensic Sci. Int.* **285** 189–203. https://doi.org/10.1016/j.forsciint.2017.10.030

SKERRETT, J., NEUMANN, C. and MATEOS-GARCIA, I. (2011). A Bayesian approach for interpreting shoemark evidence in forensic casework: Accounting for wear features. *Forensic Sci. Int.* **210** 26–30. https://doi.org/10.1016/j.forsciint.2011.01.030

SPEIR, J. A., RICHETELLI, N., FAGERT, M., HITE, M. and BODZIAK, W. J. (2016). Quantifying randomly acquired characteristics on outsoles in terms of shape and position. *Forensic Sci. Int.* **266** 399–411. https://doi.org/10.1016/j.forsciint.2016.06.012

SPENCER, N. A. and MURRAY, J. S. (2020). Supplement to "A Bayesian hierarchical model for evaluating forensic footwear evidence." https://doi.org/10.1214/20-AOAS1334SUPPA, https://doi.org/10.1214/20-AOAS1334SUPPB

SRIHARI, S. N. and SU, C. (2011). Generative models and probability evaluation for forensic evidence. In *Pattern Recognition, Machine Intelligence and Biometrics* 533–559. Springer, Berlin.

SRIHARI, S. N. and TANG, Y. (2014). Computational methods for the analysis of footwear impression evidence. In *Computational Intelligence in Digital Forensics: Forensic Investigation and Applications* 333–383. Springer, Berlin.

STONE, R. S. (2006). Footwear examinations: Mathematical probabilities of theoretical individual characteristics. *J. Forensic Identif.* **56** 577–599.

THOMPSON, W. C. and NEWMAN, E. J. (2015). Lay understanding of forensic statistics: Evaluation of random match probabilities, likelihood ratios, and verbal equivalents. *Law Hum. Behav.* **39** 332–349. https://doi.org/10.1037/lhb0000134

TOKDAR, S. T. and KASS, R. E. (2010). Importance sampling: A review. *Wiley Interdiscip. Rev.: Comput. Stat.* **2** 54–60.

WILSON, H. D. (2012). Comparison of the individual characteristics in the outsoles of thirty-nine pairs of Adidas Supernova Classic shoes. *J. Forensic Identif.* **62** 194–203.

WYATT, J. M., DUNCAN, K. and TRIMPE, M. A. (2005). Aging of shoes and its effect on shoeprint impressions. *J. Forensic Identif.* **55** 181–188.

YEKUTIELI, Y., SHOR, Y., WIESNER, S. and TSACH, T. (2012). Expert assisting computerized system for evaluating the degree of certainty in 2d shoeprints. Technical Report TP-3211, National Institute of Justice.

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1471–1492
https://doi.org/10.1214/20-AOAS1354

A BAYESIAN MODEL OF MICROBIOME DATA FOR SIMULTANEOUS IDENTIFICATION OF COVARIATE ASSOCIATIONS AND PREDICTION OF PHENOTYPIC OUTCOMES

BY MATTHEW D. KOSLOVSKY[1,*], KRISTI L. HOFFMAN[2], CARRIE R. DANIEL[3] AND MARINA VANNUCCI[1,†]

[1]*Department of Statistics, Rice University,* [*]*mkoslovsky12@gmail.com;* [†]*marina@rice.edu*
[2]*Alkek Center for Metagenomics & Microbiome Research, Baylor College of Medicine, Kristi.Hoffman@bcm.edu*
[3]*Department of Epidemiology, The University of Texas MD Anderson Cancer Center, CDaniel@mdanderson.org*

One of the major research questions regarding human microbiome studies is the feasibility of designing interventions that modulate the composition of the microbiome to promote health and to cure disease. This requires extensive understanding of the modulating factors of the microbiome, such as dietary intake, as well as the relation between microbial composition and phenotypic outcomes, such as body mass index (BMI). Previous efforts have modeled these data separately, employing two-step approaches that can produce biased interpretations of the results. Here, we propose a Bayesian joint model that simultaneously identifies clinical covariates associated with microbial composition data and predicts a phenotypic response using information contained in the compositional data. Using spike-and-slab priors, our approach can handle high-dimensional compositional as well as clinical data. Additionally, we accommodate the compositional structure of the data via balances and overdispersion typically found in microbial samples. We apply our model to understand the relations between dietary intake, microbial samples and BMI. In this analysis we find numerous associations between microbial taxa and dietary factors that may lead to a microbiome that is generally more hospitable to the development of chronic diseases, such as obesity. Additionally, we demonstrate on simulated data how our method outperforms two-step approaches and also present a sensitivity analysis.

REFERENCES

AGUS, A., DENIZOT, J., THEVENOT, J., MARTINEZ-MEDINA, M., MASSIER, S., SAUVANET, P., BERNALIER-DONADILLE, A., DENIS, S., HOFMAN, P. et al. (2016). Western diet induces a shift in microbiota composition enhancing susceptibility to adherent-invasive *E. coli* infection and intestinal inflammation. *Sci. Rep.* **6** Art. ID 19032.

AITCHISON, J. (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability*. CRC Press, London. MR0865647 https://doi.org/10.1007/978-94-009-4109-0

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

ANTONELLI, J., PARMIGIANI, G. and DOMINICI, F. (2019). High-dimensional confounding adjustment using continuous spike and slab priors. *Bayesian Anal.* **14** 825–848. MR3960772 https://doi.org/10.1214/18-BA1131

ARGIENTO, R., BIANCHINI, I. and GUGLIELMI, A. (2015). A priori truncation method for posterior sampling from homogeneous normalized completely random measure mixture models. Preprint. Available at arXiv:1507.04528.

BARBIERI, M. M. and BERGER, J. O. (2004). Optimal predictive model selection. *Ann. Statist.* **32** 870–897. MR2065192 https://doi.org/10.1214/009053604000000238

BOWYER, R. C. E., JACKSON, M. A., PALLISTER, T., SKINNER, J., SPECTOR, T. D., WELCH, A. A. and STEVES, C. J. (2018). Use of dietary indices to control for diet in human gut microbiota studies. *Microbiome* **6** Art. ID 77. https://doi.org/10.1186/s40168-018-0455-y

Key words and phrases. Bayesian statistics, joint modeling, multivariate count data, prediction, variable selection.

BROWN, P. J., VANNUCCI, M. and FEARN, T. (1998). Multivariate Bayesian variable selection and prediction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 627–641. MR1626005 https://doi.org/10.1111/1467-9868.00144

BRUNO, F., GRECO, F. and VENTRUCCI, M. (2016). Non-parametric regression on compositional covariates using Bayesian P-splines. *Stat. Methods Appl.* **25** 75–88. MR3460487 https://doi.org/10.1007/s10260-015-0339-2

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed. *Monographs on Statistics and Applied Probability* **105**. CRC Press/CRC, Boca Raton, FL. MR2243417 https://doi.org/10.1201/9781420010138

CHATFIELD, C. (1995). Model uncertainty, data mining and statistical inference. *J. R. Stat. Soc., Ser. A, Stat. Soc.* **158** 419–466.

CHEN, J. and LI, H. (2013). Variable selection for sparse Dirichlet-multinomial regression with an application to microbiome data analysis. *Ann. Appl. Stat.* **7** 418–442. MR3086425 https://doi.org/10.1214/12-AOAS592

CHEN, J., ZHANG, X. and LI, S. (2017). Multiple linear regression with compositional response and covariates. *J. Appl. Stat.* **44** 2270–2285. MR3670303 https://doi.org/10.1080/02664763.2016.1157145

DAO, M. C., EVERARD, A., ARON-WISNEWSKY, J., SOKOLOVSKA, N., PRIFTI, E., VERGER, E. O., KAYSER, B. D., LEVENEZ, F., CHILLOUX, J. et al. (2016). *Akkermansia muciniphila* and improved metabolic health during a dietary intervention in obesity: Relationship with gut microbiome richness and ecology. *Gut* **65** 426–436.

DAVID, L. A., MAURICE, C. F., CARMODY, R. N., GOOTENBERG, D. B., BUTTON, J. E., WOLFE, B. E., LING, A. V., DEVLIN, A. S., VARMA, Y. et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505** 559–563.

EDDELBUETTEL, D. and SANDERSON, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Comput. Statist. Data Anal.* **71** 1054–1063. MR3132026 https://doi.org/10.1016/j.csda.2013.02.005

EDDELBUETTEL, D., FRANÇOIS, R., ALLAIRE, J., USHEY, K., KOU, Q., RUSSEL, N., CHAMBERS, J. and BATES, D. (2011). Rcpp: Seamless R and C++ integration. *J. Stat. Softw.* **40** Issue 8.

EDGAR, R. C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10** 996–998. https://doi.org/10.1038/nmeth.2604

EGOZCUE, J. J. and PAWLOWSKY-GLAHN, V. (2005). Groups of parts and their balances in compositional data analysis. *Math. Geol.* **37** 795–828. MR2183639 https://doi.org/10.1007/s11004-005-7381-9

GLOOR, G. B., MACKLAIM, J. M., PAWLOWSKY-GLAHN, V. and EGOZCUE, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **8** Art. ID 2224.

EGOZCUE, J. J., PAWLOWSKY-GLAHN, V., MATEU-FIGUERAS, G. and BARCELÓ-VIDAL, C. (2003). Isometric logratio transformations for compositional data analysis. *Math. Geol.* **35** 279–300. MR1986165 https://doi.org/10.1023/A:1023818214614

FALONY, G., JOOSSENS, M., VIEIRA-SILVA, S., WANG, J., DARZI, Y., FAUST, K., KURILSHIKOV, A., BONDER, M. J., VALLES-COLOMER, M. et al. (2016). Population-level analysis of gut microbiome variation. *Science* **352** 560–564.

FAN, X., PETERS, B. A., JACOBS, E. J., GAPSTUR, S. M., PURDUE, M. P., FREEDMAN, N. D., ALEKSEYENKO, A. V., WU, J., YANG, L. et al. (2018). Drinking alcohol is associated with variation in the human oral microbiome in a large study of American adults. *Microbiome* **6** Art. ID 59.

FIŠEROVÁ, E. and HRON, K. (2011). On the interpretation of orthonormal coordinates for compositional data. *Math. Geosci.* **43** Art. ID 455.

GARCIA, T. P., MÜLLER, S., CARROLL, R. J. and WALZEM, R. L. (2013). Identification of important regressor groups, subgroups and individuals via regularization methods: Application to gut microbiome data. *Bioinformatics* **30** 831–837.

GELFAND, A. E. (1996). Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice. Interdiscip. Statist.* 145–161. CRC Press, London. MR1397969

GEORGE, E. I. and MCCULLOCH, R. E. (1997). Approaches for Bayesian variable selection. *Statist. Sinica* 339–373.

GOODSON, J. M., GROPPO, D., HALEM, S. and CARPINO, E. (2009). Is obesity an oral bacterial disease? *J. Dent. Res.* **88** 519–523.

GOPALAKRISHNAN, V., SPENCER, C. N., NEZI, L., REUBEN, A., ANDREWS, M. C., KARPINETS, T. V., PRIETO, P. A., VICENTE, D., HOFFMAN, K. et al. (2018). Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* **359** 97–103.

GURRY, T., GIBBONS, S. M., KEARNEY, S. M., ANANTHAKRISHNAN, A., JIANG, X., DUVALLET, C., KASSAM, Z., ALM, E. J. et al. (2018). Predictability and persistence of prebiotic dietary supplementation in a healthy human cohort. *Sci. Rep.* **8** Art. ID 12699.

HAFFAJEE, A. D. and SOCRANSKY, S. S. (2009). Relation of body mass index, periodontitis and *Tannerella forsythia*. *J. Clin. Periodontol.* **36** 89–99. https://doi.org/10.1111/j.1600-051X.2008.01356.x

HANSEN, T. H., KERN, T., BAK, E. G., KASHANI, A., ALLIN, K. H., NIELSEN, T., HANSEN, T. and PEDERSEN, O. (2018). Impact of a vegan diet on the human salivary microbiota. *Sci*. *Rep*. **8** Art. ID 5847. https://doi.org/10.1038/s41598-018-24207-3

ERCOLINI, D., FRANCAVILLA, R., VANNINI, L., DE FILIPPIS, F., CAPRIATI, T., DI CAGNO, R., IACONO, G., DE ANGELIS, M. and GOBBETTI, M. (2015). From an imbalance to a new imbalance: Italian-style gluten-free diet alters the salivary microbiota and metabolome of African celiac children. *Sci*. *Rep*. **5** Art. ID 18571.

HOFFMAN, K. L., HUTCHINSON, D. S., FOWLER, J., SMITH, D. P., AJAMI, N. J., ZHAO, H., SCHEET, P., CHOW, W.-H., PETROSINO, J. F. et al. (2018). Oral microbiota reveals signs of acculturation in Mexican American women. *PLoS ONE* **13** Art. ID e0194100.

HRON, K., FILZMOSER, P. and THOMPSON, K. (2012). Linear regression with compositional explanatory variables. *J*. *Appl*. *Stat*. **39** 1115–1128. MR2914515 https://doi.org/10.1080/02664763.2011.644268

HUTTENHOWER, C., GEVERS, D., KNIGHT, R., ABUBUCKER, S., BADGER, J. H., CHINWALLA, A. T., CREASY, H. H., EARL, A. M., FITZGERALD, M. G. et al. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486** 207–214.

JAMES, L. F., LIJOI, A. and PRÜNSTER, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand*. *J*. *Stat*. **36** 76–97. MR2508332 https://doi.org/10.1111/j.1467-9469.2008.00609.x

KATO, I., VASQUEZ, A., MOYERBRAILEAN, G., LAND, S., DJURIC, Z., SUN, J., LIN, H.-S. and RAM, J. L. (2017). Nutritional correlates of human oral microbiome. *J*. *Am*. *Coll*. *Nutr*. **36** 88–98.

KNIGHTS, D., PARFREY, L. W., ZANEVELD, J., LOZUPONE, C. and KNIGHT, R. (2011). Human-associated microbial signatures: Examining their predictive value. *Cell Host Microbe* **10** 292–296.

KOMAROFF, A. L. (2017). The microbiome and risk for obesity and diabetes. *J*. *Am*. *Med*. *Assoc*. **317** 355–356.

KOSLOVSKY, M. D., HOFFMAN, K. L., DANIEL, C. R. and VANNUCCI, M. (2020). Supplement to "A Bayesian model of microbiome data for simultaneous identification of covariate associations and prediction of phenotypic outcomes." https://doi.org/10.1214/20-AOAS1354SUPPA, https://doi.org/10.1214/20-AOAS1354SUPPB

KOVATCHEVA-DATCHARY, P., NILSSON, A., AKRAMI, R., LEE, Y. S., DE VADDER, F., ARORA, T., HALLEN, A., MARTENS, E., BJÖRCK, I. et al. (2015). Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of *Prevotella*. *Cell Metab*. **22** 971–982.

LA ROSA, P. S., BROOKS, J. P., DEYCH, E., BOONE, E. L., EDWARDS, D. J., WANG, Q., SODERGREN, E., WEINSTOCK, G. and SHANNON, W. D. (2012). Hypothesis testing and power calculations for taxonomic-based human microbiome data. *PLoS ONE* **7** Art. ID e52078.

LEITE, A. Z., RODRIGUES, N. D. C., GONZAGA, M. I., PAIOLO, J. C. C., DE SOUZA, C. A., STEFANUTTO, N. A. V., OMORI, W. P., PINHEIRO, D. G., BRISOTTI, J. L. et al. (2017). Detection of increased plasma interleukin-6 levels and prevalence of *Prevotella copri* and *Bacteroides vulgatus* in the feces of type 2 diabetes patients. *Front*. *Immunol*. **8** Art. ID 1107.

LI, H. (2015). Microbiome, metagenomics, and high-dimensional compositional data analysis. *Annu*. *Rev*. *Stat*. *Appl*. **2** 73–94.

LI, J., QUINQUE, D., HORZ, H.-P., LI, M., RZHETSKAYA, M., RAFF, J. A., HAYES, M. G. and STONEKING, M. (2014). Comparative analysis of the human saliva microbiome from different climate zones: Alaska, Germany, and Africa. *BMC Microbiol*. **14** Art. ID 316.

LIN, W., SHI, P., FENG, R. and LI, H. (2014). Variable selection in regression with compositional covariates. *Biometrika* **101** 785–797. MR3286917 https://doi.org/10.1093/biomet/asu031

LIN, D., PETERS, B. A., FRIEDLANDER, C., FREIMAN, H. J., GOEDERT, J. J., SINHA, R., MILLER, G., BERNSTEIN, M. A., HAYES, R. B. et al. (2018). Association of dietary fibre intake and gut microbiota in adults. *Br*. *J*. *Nutr*. **120** 1014–1022.

MARCOBAL, A., BARBOZA, M., SONNENBURG, E. D., PUDLO, N., MARTENS, E. C., DESAI, P., LEBRILLA, C. B., WEIMER, B. C., MILLS, D. A. et al. (2011). Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* **10** 507–514.

MARTIN-FERNANDEZ, J. A., BARCELÓ-VIDAL, C. and PAWLOWSKY-GLAHN, V. (2000). Zero replacement in compositional data sets. In *Data Analysis*, *Classification*, *and Related Methods* 155–160. Springer, Berlin.

MARTÍN-FERNÁNDEZ, J.-A., HRON, K., TEMPL, M., FILZMOSER, P. and PALAREA-ALBALADEJO, J. (2015). Bayesian-multiplicative treatment of count zeros in compositional data sets. *Stat*. *Model*. **15** 134–158. MR3325750 https://doi.org/10.1177/1471082X14535524

MARTINEZ-MEDINA, M., DENIZOT, J., DREUX, N., ROBIN, F., BILLARD, E., BONNET, R., DARFEUILLE-MICHAUD, A. and BARNICH, N. (2014). Western diet induces dysbiosis with increased *E. coli* in CEABAC10 mice, alters host barrier function favouring AIEC colonisation. *Gut* **63** 116–124.

MARUVADA, P., LEONE, V., KAPLAN, L. M. and CHANG, E. B. (2017). The human microbiome and obesity: Moving beyond associations. *Cell Host Microbe* **22** 589–599.

MERT, M. C., FILZMOSER, P., ENDEL, G. and WILBACHER, I. (2018). Compositional data analysis in epidemiology. *Stat*. *Methods Med*. *Res*. **27** 1878–1891. MR3803272 https://doi.org/10.1177/0962280216671536

MILLEN, A. E., MIDTHUNE, D., THOMPSON, F. E., KIPNIS, V. and SUBAR, A. F. (2005). The National Cancer Institute diet history questionnaire: Validation of pyramid food servings. *Am. J. Epidemiol.* **163** 279–288.

MORTON, J. T., SANDERS, J., QUINN, R. A., MCDONALD, D., GONZALEZ, A., VÁZQUEZ-BAEZA, Y., NAVAS-MOLINA, J. A., SONG, S. J., METCALF, J. L. et al. (2017). Balance trees reveal microbial niche differentiation. *mSystems* **2** Art. ID e00162-16.

NEWTON, M. A., NOUEIRY, A., SARKAR, D. and AHLQUIST, P. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* **5** 155–176.

PETERS, B. A., MCCULLOUGH, M. L., PURDUE, M. P., FREEDMAN, N. D., UM, C. Y., GAPSTUR, S. M., HAYES, R. B. and AHN, J. (2018). Association of coffee and tea intake with the oral microbiome: Results from a large cross-sectional study. *Cancer Epidemiol. Biomark. Prev.* **27** 814–821.

PINTO, J. R., EGOZCUE, J. J., PAWLOWSKY-GLAHN, V., PAREDES, R., NOGUERA-JULIAN, M. and CALLE, M. L. (2017). Balances: A new perspective for microbiome analysis. bioRxiv 219386. https://doi.org/10.1101/219386

PIOMBINO, P., GENOVESE, A., ESPOSITO, S., MOIO, L., CUTOLO, P. P., CHAMBERY, A., SEVERINO, V., MONETA, E., SMITH, D. P. et al. (2014). Saliva from obese individuals suppresses the release of aroma compounds from wine. *PLoS ONE* **9** Art. ID e85611.

POLSON, N. G., SCOTT, J. G. and WINDLE, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *J. Amer. Statist. Assoc.* **108** 1339–1349. MR3174712 https://doi.org/10.1080/01621459.2013.829001

RICHARDSON, S., BOTTOLO, L. and ROSENTHAL, J. S. (2011). Bayesian models for sparse regression analysis of high dimensional data. In *Bayesian Statistics* 9 539–568. Oxford Univ. Press, Oxford. MR3204018 https://doi.org/10.1093/acprof:oso/9780199694587.003.0018

RIDAURA, V. K., FAITH, J. J., REY, F. E., CHENG, J., DUNCAN, A. E., KAU, A. L., GRIFFIN, N. W., LOMBARD, V., HENRISSAT, B. et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341** 1241214.

SAVITSKY, T., VANNUCCI, M. and SHA, N. (2011). Variable selection for nonparametric Gaussian process priors: Models and computational strategies. *Statist. Sci.* **26** 130–149. MR2849913 https://doi.org/10.1214/11-STS354

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

SHETTY, S. A., HUGENHOLTZ, F., LAHTI, L., SMIDT, H. and DE VOS, W. M. (2017). Intestinal microbiome landscaping: Insight in community assemblage and implications for microbial modulation strategies. *FEMS Microbiol. Rev.* **41** 182–199. https://doi.org/10.1093/femsre/fuw045

SHI, P., ZHANG, A. and LI, H. (2016). Regression analysis for microbiome compositional data. *Ann. Appl. Stat.* **10** 1019–1040. MR3528370 https://doi.org/10.1214/16-AOAS928

SILVERMAN, J. D., WASHBURNE, A. D., MUKHERJEE, S. and DAVID, L. A. (2017). A phylogenetic transform enhances analysis of compositional microbiota data. *eLife* **6** Art. ID e21887.

SONNENBURG, J. L. and BÄCKHED, F. (2016). Diet-microbiota interactions as moderators of human metabolism. *Nature* **535** 56–64. https://doi.org/10.1038/nature18846

SONNENBURG, E. D., ZHENG, H., JOGLEKAR, P., HIGGINBOTTOM, S. K., FIRBANK, S. J., BOLAM, D. N. and SONNENBURG, J. L. (2010). Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141** 1241–1252.

STINGO, F. C., CHEN, Y. A., VANNUCCI, M., BARRIER, M. and MIRKES, P. E. (2010). A Bayesian graphical modeling approach to microRNA regulatory network inference. *Ann. Appl. Stat.* **4** 2024–2048. MR2829945 https://doi.org/10.1214/10-AOAS360

SUBAR, A. F., THOMPSON, F. E., KIPNIS, V., MIDTHUNE, D., HURWITZ, P., MCNUTT, S., MCINTOSH, A. and ROSENFELD, S. (2001). Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: The Eating at America's Table Study. *Am. J. Epidemiol.* **154** 1089–1099.

TADESSE, M. G., IBRAHIM, J. G., GENTLEMAN, R., CHIARETTI, S., RITZ, J. and FOA, R. (2005). Bayesian error-in-variable survival model for the analysis of GeneChip arrays. *Biometrics* **61** 488–497. MR2140921 https://doi.org/10.1111/j.1541-0420.2005.00313.x

TAKESHITA, T., MATSUO, K., FURUTA, M., SHIBATA, Y., FUKAMI, K., SHIMAZAKI, Y., AKIFUSA, S., HAN, D.-H., KIM, H.-D. et al. (2014). Distinct composition of the oral indigenous microbiota in South Korean and Japanese adults. *Sci. Rep.* **4** Art. ID 6990.

MAO, J., CHEN, Y. and MA, L. (2017). Bayesian graphical compositional regression for microbiome data. Preprint. Available at arXiv:1712.04723.

TANG, Y., MA, L. and NICOLAE, D. L. (2018). A phylogenetic scan test on a Dirichlet-tree multinomial model for microbiome data. *Ann. Appl. Stat.* **12** 1–26. MR3773384 https://doi.org/10.1214/17-AOAS1086

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

TURNBAUGH, P. J. (2017). Microbes and diet-induced obesity: Fast, cheap, and out of control. *Cell Host Microbe* **21** 278–281. https://doi.org/10.1016/j.chom.2017.02.021

VALDES, A. M., WALTER, J., SEGAL, E. and SPECTOR, T. D. (2018). Role of the gut microbiota in nutrition and health. *BMJ* **361** Art. ID k2179. https://doi.org/10.1136/bmj.k2179

VEHTARI, A., GELMAN, A. and GABRY, J. (2016). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 0.1.6.

VEHTARI, A., GELMAN, A. and GABRY, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat*. *Comput*. **27** 1413–1432. MR3647105 https://doi.org/10.1007/s11222-016-9696-4

VERSACE, F., KYPRIOTAKIS, G., BASEN-ENGQUIST, K. and SCHEMBRE, S. M. (2015). Heterogeneity in brain reactivity to pleasant and food cues: Evidence of sign-tracking in humans. *Soc*. *Cogn*. *Affect*. *Neurosci*. **11** 604–611.

WADSWORTH, W. D., ARGIENTO, R., GUINDANI, M., GALLOWAY-PENA, J., SHELBURNE, S. A. and VANNUCCI, M. (2017). An integrative Bayesian Dirichlet-multinomial regression model for the analysis of taxonomic abundances in microbiome data. *BMC Bioinform*. **18** Art. ID 94. https://doi.org/10.1186/s12859-017-1516-0

WANG, T. and ZHAO, H. (2017a). A Dirichlet-tree multinomial regression model for associating dietary nutrients with gut microorganisms. *Biometrics* **73** 792–801. MR3713113 https://doi.org/10.1111/biom.12654

WANG, T. and ZHAO, H. (2017b). Constructing predictive microbial signatures at multiple taxonomic levels. *J. Amer. Statist*. *Assoc*. **112** 1022–1031. MR3735357 https://doi.org/10.1080/01621459.2016.1270213

WASHBURNE, A. D., SILVERMAN, J. D., LEFF, J. W., BENNETT, D. J., DARCY, J. L., MUKHERJEE, S., FIERER, N. and DAVID, L. A. (2017). Phylogenetic factorization of compositional data yields lineage-level associations in microbiome datasets. *PeerJ* **5** Art. ID e2969.

WILLETT, W. (1998). Implications of total energy intake for epidemiologic analyses. In *Nutritional Epidemiology* 273–301. Oxford Univ. Press, Oxford.

WU, G. D., CHEN, J., HOFFMANN, C., BITTINGER, K., CHEN, Y.-Y., KEILBAUGH, S. A., BEWTRA, M., KNIGHTS, D., WALTERS, W. A. et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334** 105–108.

XIA, Y. and SUN, J. (2017). Hypothesis testing and statistical analysis of microbiome. *Genes Dis*. **4** 138–148.

XU, Z. and KNIGHT, R. (2015). Dietary effects on human gut microbiome diversity. *Br. J. Nutr.* **113** S1–S5.

YANG, Y., CAI, Q., ZHENG, W., STEINWANDEL, M., BLOT, W. J., SHU, X.-O. and LONG, J. (2019). Oral microbiome and obesity in a large study of low-income and African-American populations. *J. Oral Microbiol*. **11** Art. ID 1650597.

ZEIGLER, C. C., PERSSON, G. R., WONDIMU, B., MARCUS, C., SOBKO, T. and MODÉER, T. (2012). Microbiota in the oral subgingival biofilm is associated with obesity in adolescence. *Obesity* **20** 157–164.

ZHANG, Z., SCHWARTZ, S., WAGNER, L. and MILLER, W. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput*. *Biol*. **7** 203–214.

ZHANG, H., DIBAISE, J. K., ZUCCOLO, A., KUDRNA, D., BRAIDOTTI, M., YU, Y., PARAMESWARAN, P., CROWELL, M. D., WING, R. et al. (2009). Human gut microbiota in obesity and after gastric bypass. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **106** 2365–2370.

ZHANG, Y., ZHOU, H., ZHOU, J. and SUN, W. (2017). Regression models for multivariate count data. *J. Comput*. *Graph*. *Statist*. **26** 1–13. MR3610402 https://doi.org/10.1080/10618600.2016.1154063

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1493–1515
https://doi.org/10.1214/20-AOAS1360

ADAPTIVE LOG-LINEAR ZERO-INFLATED GENERALIZED POISSON AUTOREGRESSIVE MODEL WITH APPLICATIONS TO CRIME COUNTS

BY XIAOFEI XU[1], YING CHEN[2], CATHY W. S. CHEN[3] AND XIANCHENG LIN[4]

[1]*Department of Mathematics, National University of Singapor*
[2]*Risk Management Institute, National University of Singapore*
[3]*Department of Statistics, Feng Chia University, chenws@mail.fcu.edu.tw*
[4]*Department of Statistics, University of California, Davis*

This research proposes a comprehensive ALG model (Adaptive Log-linear zero-inflated Generalized Poisson integer-valued GARCH) to describe the dynamics of integer-valued time series of crime incidents with the features of autocorrelation, heteroscedasticity, overdispersion and excessive number of zero observations. The proposed ALG model captures time-varying nonlinear dependence and simultaneously incorporates the impact of multiple exogenous variables in a unified modeling framework. We use an adaptive approach to automatically detect subsamples of local homogeneity at each time point of interest and estimate the time-dependent parameters through an adaptive Bayesian Markov chain Monte Carlo (MCMC) sampling scheme. A simulation study shows stable and accurate finite sample performances of the ALG model under both homogeneous and heterogeneous scenarios. When implemented with data on crime incidents in Byron, Australia, the ALG model delivers a persuasive estimation of the stochastic intensity of criminal incidents and provides insightful interpretations on both the dynamics of intensity and the impacts of temperature and demographic factors for different crime categories.

REFERENCES

AGOSTO, A., CAVALIERE, G., KRISTENSEN, D. and RAHBEK, A. (2016). Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX). *J. Empir. Finance* **38** 640–663.

ANDERSON, C. A. (2001). Heat and violence. *Curr. Dir. Psychol. Sci.* **10** 33–38. https://doi.org/10.1111/1467-8721.00109

ANDERSON, C. A., ANDERSON, K. B., DORR, N., DENEVE, K. M. and FLANAGAN, M. (2000). Temperature and aggression. In *Advances in Experimental Social Psychology* **32** 63–133.

ANDERSON, J., MORGAN, A., BOXALL, H. and LINDEMAN, K. (2012). *Effective Crime Prevention Interventions for Implementation by Local Government*. *Research and Public Policy Series* **120**. Australian Institute of Criminology, Canberra. https://aic.gov.au/publications/rpp/rpp120.

BELOMESTNY, D. and SPOKOINY, V. (2007). Spatial aggregation of local likelihood estimates with applications to classification. *Ann. Statist.* **35** 2287–2311. MR2363972 https://doi.org/10.1214/009053607000000271

BERENTSEN, G. D., BULLA, J., MARUOTTI, A. and STØVE, B. (2018). Modelling corporate defaults: A Markov-switching Poisson log-linear autoregressive model. arXiv preprint arXiv:1804.09252.

CHEN, C. W. S. and KHAMTHONG, K. (2019). Bayesian modelling of nonlinear negative binomial integer-valued GARCHX models. *Stat. Model.* https://doi.org/10.1177/1471082X19845541

CHEN, C. W. S., KHAMTHONG, K. and LEE, S. (2019). Markov switching integer-valued generalized autoregressive conditional heteroscedastic models for Dengue counts. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **68** 963–983. MR4002379 https://doi.org/10.1111/rssc.12344

CHEN, C. W. S. and LEE, S. (2016). Generalized Poisson autoregressive models for time series of counts. *Comput. Statist. Data Anal.* **99** 51–67. MR3473081 https://doi.org/10.1016/j.csda.2016.01.009

CHEN, C. W. S. and LEE, S. (2017). Bayesian causality test for integer-valued time series models with applications to climate and crime data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 797–814. MR3670418 https://doi.org/10.1111/rssc.12200

Key words and phrases. Bayesian, integer-valued GARCH model, excess zeros, overdispersion, nonstationarity, MCMC.

CHEN, C. W. S., SO, M. K. P., LI, J. C. and SRIBOONCHITTA, S. (2016). Autoregressive conditional negative binomial model applied to over-dispersed time series of counts. *Stat. Methodol.* **31** 73–90. MR3477729 https://doi.org/10.1016/j.stamet.2016.02.001

CHEN, Y. and LI, B. (2017). An adaptive functional autoregressive forecast model to predict electricity price curves. *J. Bus. Econom. Statist.* **35** 371–388. MR3663780 https://doi.org/10.1080/07350015.2015.1092976

CHIB, S. (1995). Marginal likelihood from the Gibbs output. *J. Amer. Statist. Assoc.* **90** 1313–1321. MR1379473

COHN, E. G. (1990). Weather and crime. *Br. J. Criminol.* **30** 51–64.

CONSUL, P. and FAMOYE, F. (1992). Generalized Poisson regression model. *Comm. Statist. Theory Methods* **21** 89–109.

CUI, Y., LI, Q. and ZHU, F. (2020). Flexible bivariate Poisson integer-valued GARCH model. *Ann. Inst. Statist. Math.* https://doi.org/10.1007/s10463-019-00732-4

CUI, Y. and ZHU, F. (2018). A new bivariate integer-valued GARCH model allowing for negative cross-correlation. *TEST* **27** 428–452. MR3799147 https://doi.org/10.1007/s11749-017-0552-4

DAVIS, R. A., DUNSMUIR, W. T. M. and STREETT, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika* **90** 777–790. MR2024757 https://doi.org/10.1093/biomet/90.4.777

DAVIS, R. A., LEE, T. C. M. and RODRIGUEZ-YAM, G. A. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101** 223–239. MR2268041 https://doi.org/10.1198/016214505000000745

FAMOYE, F. and SINGH, K. P. (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *J. Data Sci.* **4** 117–130.

FERLAND, R., LATOUR, A. and ORAICHI, D. (2006). Integer-valued GARCH process. *J. Time Series Anal.* **27** 923–942. MR2328548 https://doi.org/10.1111/j.1467-9892.2006.00496.x

FOKIANOS, K. and FRIED, R. (2010). Interventions in INGARCH processes. *J. Time Series Anal.* **31** 210–225. MR2676152 https://doi.org/10.1111/j.1467-9892.2010.00657.x

FOKIANOS, K. and FRIED, R. (2012). Interventions in log-linear Poisson autoregression. *Stat. Model.* **12** 299–322. MR3179504 https://doi.org/10.1177/1471082X1201200401

FOKIANOS, K., RAHBEK, A. and TJØSTHEIM, D. (2009). Poisson autoregression. *J. Amer. Statist. Assoc.* **104** 1430–1439. With electronic supplementary materials available online. MR2596998 https://doi.org/10.1198/jasa.2009.tm08270

FOKIANOS, K. and TJØSTHEIM, D. (2011). Log-linear Poisson autoregression. *J. Multivariate Anal.* **102** 563–578. MR2755016 https://doi.org/10.1016/j.jmva.2010.11.002

FOKIANOS, K. and TJØSTHEIM, D. (2012). Nonlinear Poisson autoregression. *Ann. Inst. Statist. Math.* **64** 1205–1225. MR2981620 https://doi.org/10.1007/s10463-012-0351-3

FOKIANOS, K., STØVE, B., TJØSTHEIM, D. and DOUKHAN, P. (2020). Multivariate count autoregression. *Bernoulli* **26** 471–499. MR4036041 https://doi.org/10.3150/19-BEJ1132

FRANKE, J., KIRCH, C. and KAMGAING, J. T. (2012). Changepoints in times series of counts. *J. Time Series Anal.* **33** 757–770. MR2969909 https://doi.org/10.1111/j.1467-9892.2011.00778.x

GUPTA, P. L., GUPTA, R. C. and TRIPATHI, R. C. (1996). Analysis of zero-adjusted count data. *Comput. Statist. Data Anal.* **23** 207–218.

HÄRDLE, W., MÜLLER, M., SPERLICH, S. and WERWATZ, A. (2012). *Nonparametric and Semiparametric Models*. Springer Science & Business Media. MR2061786 https://doi.org/10.1007/978-3-642-17146-8

HIRD, C. and RUPAREL, C. (2007). *Seasonality in Recorded Crime*: *Preliminary Findings* **2**. Home Office London. 07.

HSIANG, S. M., BURKE, M. and MIGUEL, E. (2013). Quantifying the influence of climate on human conflict. *Science* **341** 1235367. https://doi.org/10.1126/science.1235367

JAZI, M. A., JONES, G. and LAI, C.-D. (2012). First-order integer valued AR processes with zero inflated Poisson innovations. *J. Time Series Anal.* **33** 954–963. MR2991911 https://doi.org/10.1111/j.1467-9892.2012.00809.x

KANG, J. and LEE, S. (2014). Parameter change test for Poisson autoregressive models. *Scand. J. Stat.* **41** 1136–1152. MR3277042 https://doi.org/10.1111/sjos.12088

KARLIS, D. and PEDELI, X. (2013). Flexible bivariate INAR(1) processes using copulas. *Comm. Statist. Theory Methods* **42** 723–740. MR3211946 https://doi.org/10.1080/03610926.2012.754466

KRISTENSEN, K., NIELSEN, A., BERG, C. W., SKAUG, H. and BELL, B. (2016). TMB: Automatic differentiation and Laplace approximation. *J. Stat. Softw.* **70** 1–21.

LEE, S., LEE, Y. and CHEN, C. W. S. (2016). Parameter change test for zero-inflated generalized Poisson autoregressive models. *Statistics* **50** 540–557. MR3506657 https://doi.org/10.1080/02331888.2015.1083020

MARES, D. M. (2009). Civilization, economic change, and trends in interpersonal violence in western societies. *Theor. Criminol.* **13** 419–449.

MARES, D. M. and MOFFETT, K. W. (2016). Climate change and interpersonal violence: A "global" estimate and regional inequities. *Clim. Change* **135** 297–310.

McDowall, D., Loftin, C. and Pate, M. (2012). Seasonal cycles in crime, and their variability. *J. Quant. Criminol.* **28** 389–410.

Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models. *Ann. Statist.* **32** 577–602. MR2060170 https://doi.org/10.1214/009053604000000102

Pereira, D. V., Andresen, M. A. and Mota, C. M. (2016). A temporal and spatial analysis of homicides. *J. Environ. Psychol.* **46** 116–124.

Ranson, M. (2014). Crime, weather, and climate change. *J. Environ. Econ. Manage.* **67** 274–302.

Rotton, J. and Cohn, E. G. (2000). Violence is a curvilinear function of temperature in Dallas: A replication. *J. Pers. Soc. Psychol.* **78** 1074–1081.

Sellers, K. F. and Shmueli, G. (2010). A flexible regression model for count data. *Ann. Appl. Stat.* **4** 943–961. MR2758428 https://doi.org/10.1214/09-AOAS306

Spokoiny, V. (2009). Multiscale local change point detection with applications to value-at-risk. *Ann. Statist.* **37** 1405–1436. MR2509078 https://doi.org/10.1214/08-AOS612

Stacey, M., Carbone-López, K. and Rosenfeld, R. (2011). Demographic change and ethnically motivated crime: The impact of immigration on anti-Hispanic hate crime in the United States. *J. Contemp. Crim. Justice* **27** 278–298.

Xu, X., Chen, Y., Chen, C. W. S. and Lin, X. (2020a). Supplement to "Adaptive log-linear zero-inflated generalized Poisson autoregressive model with applications to crime counts." https://doi.org/10.1214/20-AOAS1360SUPPA

Xu, X., Chen, Y., Chen, C. W. S. and Lin, X. (2020b). Source code for "Adaptive log-linear zero-inflated generalized Poisson autoregressive model with applications to crime counts." https://doi.org/10.1214/20-AOAS1360SUPPB

Yan, Y. Y. (2004). Seasonality of property crime in Hong Kong. *Br. J. Criminol.* **44** 276–283.

Zhu, F. (2011). A negative binomial integer-valued GARCH model. *J. Time Series Anal.* **32** 54–67. MR2790672 https://doi.org/10.1111/j.1467-9892.2010.00684.x

Zhu, F. (2012a). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *J. Math. Anal. Appl.* **389** 58–71. MR2876481 https://doi.org/10.1016/j.jmaa.2011.11.042

Zhu, F. (2012b). Zero-inflated Poisson and negative binomial integer-valued GARCH models. *J. Statist. Plann. Inference* **142** 826–839. MR2863870 https://doi.org/10.1016/j.jspi.2011.10.002

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1516–1534
https://doi.org/10.1214/20-AOAS1358

IDENTIFYING OVERLAPPING TERRORIST CELLS FROM THE NOORDIN TOP ACTOR–EVENT NETWORK

BY SAVERIO RANCIATI[1], VERONICA VINCIOTTI[2] AND ERNST C. WIT[3]

[1]*Department of Statistical Sciences, University of Bologna, saverio.ranciati2@unibo.it*
[2]*Department of Mathematics, Brunel University London, veronica.vinciotti@brunel.ac.uk*
[3]*Institute of Computational Science, Università della Svizzera italiana, wite@usi.ch*

Actor–event data are common in sociological settings, whereby one registers the pattern of attendance of a group of social actors to a number of events. We focus on 79 members of the Noordin Top terrorist network, who were monitored attending 45 events. The attendance or nonattendance of the terrorist to events defines the social fabric, such as group coherence and social communities. The aim of the analysis of such data is to learn about the affiliation structure. Actor–event data is often transformed to actor–actor data in order to be further analysed by network models, such as stochastic block models. This transformation and such analyses lead to a natural loss of information, particularly when one is interested in identifying, possibly overlapping, subgroups or communities of actors on the basis of their attendances to events. In this paper we propose an actor–event model for overlapping communities of terrorists which simplifies interpretation of the network. We propose a mixture model with overlapping clusters for the analysis of the binary actor–event network data, called `manet`, and develop a Bayesian procedure for inference. After a simulation study, we show how this analysis of the terrorist network has clear interpretative advantages over the more traditional approaches of affiliation network analysis.

REFERENCES

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

AITKIN, M., VU, D. and FRANCIS, B. (2017). Statistical modelling of a terrorist network. *J. Roy. Statist. Soc. Ser. A* **180** 751–768. MR3660159 https://doi.org/10.1111/rssa.12233

CELEUX, G., FORBES, F., ROBERT, C. P. and TITTERINGTON, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Anal.* **1** 651–673. MR2282197 https://doi.org/10.1214/06-BA122

CLAESKENS, G. and HJORT, N. L. (2008). *Model Selection and Model Averaging. Cambridge Series in Statistical and Probabilistic Mathematics* **27**. Cambridge Univ. Press, Cambridge. MR2431297 https://doi.org/10.1017/CBO9780511790485

DAUDIN, J.-J., PICARD, F. and ROBIN, S. (2008). A mixture model for random graphs. *Stat. Comput.* **18** 173–183. MR2390817 https://doi.org/10.1007/s11222-007-9046-7

DOREIAN, P., BATAGELJ, V. and FERLIGOJ, A. (2004). Generalized blockmodeling of two-mode network data. *Soc. Netw.* **26** 29–53.

DURANTE, D. and DUNSON, D. B. (2014). Nonparametric Bayes dynamic modelling of relational data. *Biometrika* **101** 883–898. MR3286923 https://doi.org/10.1093/biomet/asu040

EVERTON, S. F. (2012). *Disrupting Dark Networks. Structural Analysis in the Social Sciences* **34**. Cambridge Univ. Press, Cambridge.

FRANK, O. and STRAUSS, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. MR0860518

GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E., AIROLDI, E. M. et al. (2010). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.

HANDCOCK, M. S., RAFTERY, A. E. and TANTRUM, J. M. (2007). Model-based clustering for social networks. *J. Roy. Statist. Soc. Ser. A* **170** 301–354. MR2364300 https://doi.org/10.1111/j.1467-985X.2007.00471.x

Key words and phrases. Bayesian modeling, mixture models, MCMC algorithm, network, overlapping clusters.

HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262 https://doi.org/10.1198/016214502388618906

HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

INTERNATIONAL CRISIS GROUP (2009). Indonesia: Noordin Top's support base. Asia Briefing 95. Available at http://www.refworld.org/docid/4a968a982.html.

KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models. Springer Series in Statistics*. Springer, New York. MR2724362 https://doi.org/10.1007/978-0-387-88146-1

LEGER, J. B. (2015). Blockmodels: Latent and stochastic block model estimation by a V-EM algorithm. R package version 1.

NEAL, Z. (2014). The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Soc. Netw.* **39** 84–97.

NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255 https://doi.org/10.1198/016214501753208735

PLUMMER, M., BEST, N., COWLES, K. and VINES, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6** 7–11.

RAFTERY, A. E., NIU, X., HOFF, P. D. and YEUNG, K. Y. (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *J. Comput. Graph. Statist.* **21** 901–919. MR3005803 https://doi.org/10.1080/10618600.2012.679240

RANCIATI, S., VINCIOTTI, V. and WIT, E. C. (2020). Supplement to "Identifying overlapping terrorist cells from the Noordin Top actor–event network." https://doi.org/10.1214/20-AOAS1358SUPP

RANCIATI, S., VIROLI, C. and WIT, E. C. (2017). Mixture model with multiple allocations for clustering spatially correlated observations in the analysis of ChIP-Seq data. *Biom. J.* **59** 1301–1316. MR3731217 https://doi.org/10.1002/bimj.201600131

RAND, W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. Amer. Statist. Assoc.* **66** 846–850.

ROBINS, G., SNIJDERS, T., WANG, P., HANDCOCK, M. and PATTISON, P. (2007). Recent developments in exponential random graph (p*) models for social networks. *Soc. Netw.* **29** 192–215.

SEWELL, D. K. and CHEN, Y. (2017). Latent space approaches to community detection in dynamic networks. *Bayesian Anal.* **12** 351–377. MR3620737 https://doi.org/10.1214/16-BA1000

SIGNORELLI, M. and WIT, E. C. (2018). A penalized inference approach to stochastic block modelling of community structure in the Italian Parliament. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **67** 355–369. MR3758770 https://doi.org/10.1111/rssc.12234

SKVORETZ, J. and FAUST, K. (1999). Logit models for affiliation networks. *Sociol. Method.* **29** 253–280.

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 583–639. MR1979380 https://doi.org/10.1111/1467-9868.00353

STEPHENS, M. (2000). Dealing with label switching in mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 795–809. MR1796293 https://doi.org/10.1111/1467-9868.00265

WANG, Y. J. and WONG, G. Y. (1987). Stochastic blockmodels for directed graphs. *J. Amer. Statist. Assoc.* **82** 8–19. MR0883333

WANG, P., SHARPE, K., ROBINS, G. L. and PATTISON, P. E. (2009). Exponential random graph (p*) models for affiliation networks. *Soc. Netw.* **31** 12–25.

WASSERMAN, S. and FAUST, K. (1994). *Social Network Analysis: Methods and Applications. Structural Analysis in the Social Sciences* **8**. Cambridge Univ. Press, Cambridge.

WASSERMAN, S. and PATTISON, P. (1996). Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and p. *Psychometrika* **61** 401–425. MR1424909 https://doi.org/10.1007/BF02294547

XING, E. P., FU, W. and SONG, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *Ann. Appl. Stat.* **4** 535–566. MR2758639 https://doi.org/10.1214/09-AOAS311

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1535–1556
https://doi.org/10.1214/20-AOAS1357

LOG-CONTRAST REGRESSION WITH FUNCTIONAL COMPOSITIONAL PREDICTORS: LINKING PRETERM INFANTS' GUT MICROBIOME TRAJECTORIES TO NEUROBEHAVIORAL OUTCOME

BY ZHE SUN[1,*], WANLI XU[2,‡], XIAOMEI CONG[2,§], GEN LI[3] AND KUN CHEN[1,†]

[1]*Department of Statistics, University of Connecticut,* [*]*zhe.sun@uconn.edu;* [†]*kun.chen@uconn.edu*
[2]*School of Nursing, University of Connecticut,* [‡]*wanli.xu@uconn.edu;* [§]*xiaomei.cong@uconn.edu*
[3]*Department of Biostatistics, Columbia University, gl2521@cumc.columbia.edu*

The neonatal intensive care unit (NICU) experience is known to be one of the most crucial factors that drive preterm infants' neurodevelopmental and health outcome. It is hypothesized that stressful early life experience of very preterm neonate is imprinting gut microbiome by the regulation of the so-called brain-gut axis, and, consequently, certain microbiome markers are predictive of later infant neurodevelopment. To investigate, a preterm infant study was conducted; infant fecal samples were collected during the infants' first month of postnatal age, resulting in functional compositional microbiome data, and neurobehavioral outcomes were measured when infants reached 36–38 weeks of postmenstrual age. To identify potential microbiome markers and estimate how the trajectories of gut microbiome compositions during early postnatal stage impact later neurobehavioral outcomes of the preterm infants, we innovate a sparse log-contrast regression with functional compositional predictors. The functional simplex structure is strictly preserved, and the functional compositional predictors are allowed to have sparse, smoothly varying and accumulating effects on the outcome through time. Through a pragmatic basis expansion step, the problem boils down to a linearly constrained sparse group regression, for which we develop an efficient algorithm and obtain theoretical performance guarantees. Our approach yields insightful results in the preterm infant study. The identified microbiome markers and the estimated time dynamics of their impact on the neurobehavioral outcome shed lights on the linkage between stress accumulation in early postnatal stage and neurodevelpomental process of infants.

REFERENCES

AITCHISON, J. (1982). The statistical analysis of compositional data. *J. Roy. Statist. Soc. Ser. B* **44** 139–177. With discussion. MR0676206

AITCHISON, J. (2003). *The Statistical Analysis of Compositional Data*. Blackburn Press, NJ. MR0865647 https://doi.org/10.1007/978-94-009-4109-0

AITCHISON, J. and BACON-SHONE, J. (1984). Log-contrast models for experiments with mixtures. *Biometrika* **71** 323–330.

AITCHISON, J. and EGOZCUE, J. J. (2005). Compositional data analysis: Where are we and where should we be heading? *Math. Geol.* **37** 829–850. MR2183640 https://doi.org/10.1007/s11004-005-7383-7

BOMAR, L., MALTZ, M., COLSTON, S. and GRAF, J. (2011). Directed culturing of microorganisms using metatranscriptomics. *mBio* **2** e00012-11. https://doi.org/10.1128/mBio.00012-11

BOULANGÉ, C. L., NEVES, A. L., CHILLOUX, J., NICHOLSON, J. K. and DUMAS, M.-E. (2016). Impact of the gut microbiota on inflammation, obesity, and metabolic disease. *Gen. Med.* **8** 42. https://doi.org/10.1186/s13073-016-0303-2

BOYD, S., PARIKH, N., CHU, E., PELEATO, B. and ECKSTEIN, J. (2011). *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Foundations and Trends in Machine Learning* **3**, 1. https://doi.org/10.1561/2200000016

CLARKE, S. F., MURPHY, E. F., NILAWEERA, K., ROSS, P. R., SHANAHAN, F., O'TOOLE, P. W. and COTTER, P. D. (2012). The gut microbiota and its relationship to diet and obesity: New insights. *Gut Microbes* **3** 186–202. https://doi.org/10.4161/gmic.20168

Key words and phrases. Constrained optimization, longitudinal data, simplex, group selection.

CONG, X. (2016). Gut microbiome developmental patterns in early life of preterm infants: Impacts of feeding and gender. *PLoS ONE* **11** 1–19.

CONG, X., JUDGE, M., XU, W. and DIALLO, A. (2017). Influence of feeding type on gut microbiome development in hospitalized preterm infants. *Nursing Research* **66** 123–133.

DINAN, T. G. and CRYAN, J. F. (2012). Regulation of the stress response by the gut microbiota: Implications for psychoneuroendocrinology. *Psychoneuroendocrinology* **37** 1369–1378.

FAN, Y. and TANG, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 531–552. MR3065478 https://doi.org/10.1111/rssb.12001

FANAROFF, A. A., HACK, M. and MC, W. (2003). The NICHD neonatal research network: Changes in practice and outcomes during the first 15 years. *Semin. Perinatol.* **27** 281–287.

FELDMAN, R. and EIDELMAN, A. I. (2003). Direct and indirect effects of breast milk on the neurobehavioral and cognitive development of premature infants. *Dev. Psychobiol.* **43** 109–119.

FRANZ, C. M., HUCH, M., ABRIOUEL, H., HOLZAPFEL, W. H. and GÁLVEZ, A. V. (2011). Enterococci as probiotics and their implications in food safety. *Int. J. Food Microbiol.* **151** **2** 125–40.

HALTALIN, K. C. (1967). Neonatal shigellosis: Report of 16 cases and review of the literature. *JAMA Pediatrics* **114** 603–611.

LI, H. (2015). Microbiome, metagenomics, and high-dimensional compositional data analysis. *Annu. Rev. Stat. Appl.* **2** 73–94.

LIN, W., SHI, P., FENG, R. and LI, H. (2014). Variable selection in regression with compositional covariates. *Biometrika* **101** 785–797. MR3286917 https://doi.org/10.1093/biomet/asu031

LOUNICI, K., PONTIL, M., VAN DE GEER, S. and TSYBAKOV, A. B. (2011). Oracle inequalities and optimal inference under group sparsity. *Ann. Statist.* **39** 2164–2204. MR2893865 https://doi.org/10.1214/11-AOS896

MORRIS, J. S. (2015). Functional regression. *Annu. Rev. Stat. Appl.* **2** 321–359. https://doi.org/10.1146/annurev-statistics-010814-020413

MWANIKI, M. K., ATIENO, M., LAWN, J. E. and NEWTON, C. R. (2012). Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: A systematic review. *Lancet* **379** 445–452.

NEU, J. and RUSHING, J. M. (2011). Cesarean versus vaginal delivery: Long-term infant outcomes and the hygiene hypothesis. *Clin. Perinatol.* **38** **2** 321–31.

R CORE TEAM (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

SHI, P., ZHANG, A. and LI, H. (2016). Regression analysis for microbiome compositional data. *Ann. Appl. Stat.* **10** 1019–1040. MR3528370 https://doi.org/10.1214/16-AOAS928

STOLL, B. J., HANSEN, N. I. and BELL, E. F. (2010). Neonatal outcomes of extremely preterm infants from the NICHD neonatal research network. *Pediatrics* **126** 443–456.

STONE, M. (1974). Cross-validation and multinomial prediction. *Biometrika* **61** 509–515. MR0415896 https://doi.org/10.1093/biomet/61.3.509

SUN, Z. and CHEN, K. (2020). Compack: Regression with compositional covariates. R package version 0.1.0.

SUN, Z., XU, W., CONG, X., LI, G. and CHEN, K. (2020). Supplement to "Log-contrast regression with functional compositional predictors: Linking preterm infants' gut microbiome trajectories to neurobehavioral outcome." https://doi.org/10.1214/20-AOAS1357SUPP

TAYLOR, J. and TIBSHIRANI, R. J. (2015). Statistical learning and selective inference. *Proc. Natl. Acad. Sci. USA* **112** 7629–7634. MR3371123 https://doi.org/10.1073/pnas.1507583112

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

WEI, F. and HUANG, J. (2010). Consistent group selection in high-dimensional linear regression. *Bernoulli* **16** 1369–1384. MR2759183 https://doi.org/10.3150/10-BEJ252

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574 https://doi.org/10.1111/j.1467-9868.2005.00532.x

The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1557–1580
https://doi.org/10.1214/20-AOAS1361

INFERRING A CONSENSUS PROBLEM LIST USING PENALIZED MULTISTAGE MODELS FOR ORDERED DATA

BY PHILIP S. BOONSTRA[1] AND JOHN C.KRAUSS[2]

[1]*Department of Biostatistics, University of Michigan, philb@umich.edu*
[2]*Division of Hematology Oncology, University of Michigan, jkrauss@med.umich.edu*

A patient's medical problem list describes his or her current health status and aids in the coordination and transfer of care between providers. Because a problem list is generated once and then subsequently modified or updated, what is not usually observable is the provider-effect. That is, to what extent does a patient's problem in the electronic medical record actually reflect a consensus communication of that patient's current health status? To that end, we report on and analyze a unique interview-based design in which multiple medical providers independently generate problem lists for each of three patient case abstracts of varying clinical difficulty. Due to the uniqueness of both our data and the scientific objectives of our analysis, we apply and extend so-called multistage models for ordered lists and equip the models with variable selection penalties to induce sparsity. Each problem has a corresponding nonnegative parameter estimate, interpreted as a relative log-odds ratio, with larger values suggesting greater importance and zero values suggesting unimportant problems. We use these fitted penalized models to quantify and report the extent of consensus. We conduct a simulation study to evaluate the performance of our methodology and then analyze the motivating problem list data. For the three case abstracts, the proportions of problems with model-estimated nonzero log-odds ratios were 10/28, 16/47 and 13/30. Physicians exhibited consensus on the highest ranked problems in the first and last case abstracts but agreement quickly deteriorated; in contrast, physicians broadly disagreed on the relevant problems for the middle—and most difficult—case abstract.

REFERENCES

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (*Tsahkadsor*, 1971) 267–281. MR0483125

BENTER, W. et al. (2008). Computer-based horse race handicapping and wagering systems: A report. In *Efficiency of Racetrack Betting Markets* (D. B. Hausch, V. S. Y. Lo and W. T. Ziemba, eds.) 183–198. World Scientific, Singapore.

BOONSTRA, P. S and KRAUSS, JOHN C. (2020). Supplement to "Inferring a consensus problem list using penalized multistage models for ordered data." https://doi.org/10.1214/20-AOAS1361SUPPA, https://doi.org/10.1214/20-AOAS1361SUPPB.

BOULESTEIX, A.-L. and SLAWSKI, M. (2009). Stability and aggregation of ranked gene lists. *Brief. Bioinform.* **10** 556–568.

DECONDE, R. P., HAWLEY, S., FALCON, S., CLEGG, N., KNUDSEN, B. and ETZIONI, R. (2006). Combining results of microarray experiments: A rank aggregation approach. *Stat. Appl. Genet. Mol. Biol.* **5** Art. 15, 25. MR2240846 https://doi.org/10.2202/1544-6115.1204

DENG, K., HAN, S., LI, K. J. and LIU, J. S. (2014). Bayesian aggregation of order-based rank data. *J. Amer. Statist. Assoc.* **109** 1023–1039. MR3265678 https://doi.org/10.1080/01621459.2013.878660

DICKER, L., HUANG, B. and LIN, X. (2013). Variable selection and estimation with the seamless-L_0 penalty. *Statist. Sinica* **23** 929–962. MR3086661

DWORK, C., KUMAR, R., NAOR, M. and SIVAKUMAR, D. (2001). Rank aggregation methods for the web. In *Proceedings of the* 10*th International Conference on World Wide Web* 613–622. ACM, New York.

FLIGNER, M. A. and VERDUCCI, J. S. (1988). Multistage ranking models. *J. Amer. Statist. Assoc.* **83** 892–901. MR0963820

Key words and phrases. Conditional multinomial, L_0 penalty, ranked lists, variable selection.

GORMLEY, I. C. and MURPHY, T. B. (2008a). Exploring voting blocs within the Irish electorate: A mixture modeling approach. *J. Amer. Statist. Assoc.* **103** 1014–1027. MR2528824 https://doi.org/10.1198/016214507000001049

GORMLEY, I. C. and MURPHY, T. B. (2008b). A mixture of experts model for rank data with applications in election studies. *Ann. Appl. Stat.* **2** 1452–1477. MR2655667 https://doi.org/10.1214/08-AOAS178

HURVICH, C. M. and TSAI, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* **76** 297–307. MR1016020 https://doi.org/10.1093/biomet/76.2.297

KENDALL, M. G. (1962). Ranks and measures. *Biometrika* **49** 133–137. MR0138175 https://doi.org/10.1093/biomet/49.1-2.133

KRAUSS, J. C., BOONSTRA, P. S., VANTSEVICH, A. V. and FRIEDMAN, C. P. (2016). Is the problem list in the eye of the beholder? An exploration of consistency across physicians. *J. Am. Med. Inform. Assn.* **23** 859–865.

LI, X., WANG, X. and XIAO, G. (2019). A comparative study of rank aggregation methods for partial and top ranked lists in genomic applications. *Brief. Bioinform.* **20** 178–189.

LI, X., CHOUDHARY, P. K., BISWAS, S. and WANG, X. (2018a). A Bayesian latent variable approach to aggregation of partial and top-ranked lists in genomic studies. *Stat. Med.* **37** 4266–4278. MR3879427 https://doi.org/10.1002/sim.7920

LI, H., XU, M., LIU, J. S. and FAN, X. (2018b). ExtMallows: An extended Mallows model and its hierarchical version for ranked data aggregation. R package version 0.1.0.

LI, H., XU, M., LIU, J. S. and FAN, X. (2020). An extended mallows model for ranked data aggregation. *J. Amer. Statist. Assoc.* **115** 730–746. MR4107676 https://doi.org/10.1080/01621459.2019.1573733

LIN, S. (2010). Space oriented rank-based data integration. *Stat. Appl. Genet. Mol. Biol.* **9** Art. 20, 25. MR2721700 https://doi.org/10.2202/1544-6115.1534

LIN, S. and DING, J. (2009). Integration of ranked lists via cross entropy Monte Carlo with applications to mRNA and microRNA studies. *Biometrics* **65** 9–18. MR2665841 https://doi.org/10.1111/j.1541-0420.2008.01044.x

LUCE, R. D. (1959). *Individual Choice Behavior*: *A Theoretical Analysis*. Wiley, New York. MR0108411

MALLOWS, C. L. (1957). Non-null ranking models. I. *Biometrika* **44** 114–130. MR0087267 https://doi.org/10.1093/biomet/44.1-2.114

MARDEN, J. I. (1995). *Analyzing and Modeling Rank Data*. *Monographs on Statistics and Applied Probability* **64**. CRC Press, London. MR1346107

MEYER, A. N., PAYNE, V. L., MEEKS, D. W., RAO, R. and SINGH, H. (2013). Physicians' diagnostic accuracy, confidence, and resource requests: A vignette study. *JAMA Intern. Med.* **173** 1952–1958.

MOLLICA, C. and TARDELLA, L. (2017). Bayesian Plackett–Luce mixture models for partially ranked data. *Psychometrika* **82** 442–458. MR3665302 https://doi.org/10.1007/s11336-016-9530-0

NEUWIRTH, E. (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.

NOMBEKELA, S. W., MURPHY, M. R., GONYOU, H. W. and MARDEN, J. I. (1994). Dietary preferences in early lactation cows as affected by primary tastes and some common feed flavors. *J. Dairy. Sci.* **77** 2393–2399.

PLACKETT, R. L. (1975). The analysis of permutations. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **24** 193–202. MR0391338 https://doi.org/10.2307/2346567

R CORE TEAM (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

SCHIMEK, M. G., BUDINSKÁ, E., KUGLER, K. G., ŠVENDOVÁ, V., DING, J. and LIN, S. (2015). TopKLists: A comprehensive R package for statistical inference, stochastic aggregation, and visualization of multiple omics ranked lists. *Stat. Appl. Genet. Mol. Biol.* **14** 311–316. MR3353637 https://doi.org/10.1515/sagmb-2014-0093

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

SPEARMAN, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.* **15** 72–101.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

WEBBER, W., MOFFAT, A. and ZOBEL, J. (2010). A similarity measure for indefinite rankings. *ACM Trans. Inf. Syst.* **28** Article 20.

WEED, L. L. (1968). Special article: Medical records that guide and teach. *N. Engl. J. Med.* **278** 593–600.

WICKHAM, H. (2017). tidyverse: Easily install and load the "tidyverse". R package version 1.2.1.

The Annals of Applied Statistics

Next Issues

Mining events with declassified diplomatic documents
YUANJUN GAO, JACK GOETZ, MATTHEW CONNELLY AND RAHUL MAZUMDER

Monotonic effects of characteristics on returns
JARED DALE FISHER, DAVID WALKER PUELZ AND CARLOS MARINHO CARVALHO

Hawkes binomial topic model with applications to coupled conflict-Twitter data
GEORGE MOHLER, ERIN MCGRATH, CODY BUNTAIN AND GARY LAFREE

Effective model calibration via sensible variable identification and adjustment, with application to composite fuselage simulation
YAN WANG, XIAOWEI YUE, RUI TUO, JEFFREY HUNT AND JIANJUN SHI

The statistical performance of matching-adjusted indirect comparisons
DAVID CHENG, RAJEEV AYYAGARI AND JAMES SIGNOROVITCH

A nearest-neighbor based nonparametric test for viral remodeling in heterogeneous single cell proteomic data . . . . TRAMBAK BANERJEE, BHASWAR BHATTACHARYA AND GOURAB MUKHERJEE

Identifying main effects and interactions among exposures using Gaussian processes
FEDERICO FERRARI

Integrative statistical methods for exposure mixtures and health
BRIAN JAMES REICH, YAWEN GUAN, DENIS FOURCHES, JOSHUA L. WARREN, STEFANIE E. SARNAT AND HOWARD H. CHANGE

Measuring timeliness of annual reports filing by jump additive models . . . . YICHENG KANG

Bayesian inference for multi-strain epidemics with application to Escherichia coli O157:H7 in feedlot cattle . . . . PANAYIOTA TOULOUPOU, BARBEL FINKENSTADT, THOMAS E. BESSER, NIGEL P. FRENCH AND SIMON E. F. SPENCER

Late 19th-century navigational uncertainties and their influence on sea surface temperature estimates . . . . CHENGUANG DAI, DUO CHAN, PETER HUYBERS AND NATESH PILLAI

RNDClone: Tumor subclone reconstruction based on integrating DNA and RNA sequence data . . . . TIANJIAN ZHOU, SUBHAJIT SENGUPTA, PETER MÜLLER AND YUAN JI

Incorporating compositional heterogeneity into Lie Markov models for phylogenetic inference . . . . NAOMI ELIZABETH HANNAFORD, SARAH ELIZABETH HEAPS, TOM MATTHEW WILLIAM NYE, TOM ANTHONY WILLIAMS AND T. MARTIN EMBLEY

Nonparametric Bayesian multi-armed bandits for single cell experiment design
FEDERICO CAMERLENGHI, BIANCA DUMITRASCU, FEDERICO FERRARI, BARBARA ENGELHARDT AND STEFANO FAVARO

Data fusion model for speciated nitrogen to identify environmental drivers and improve estimation of nitrogen in lakes . . . . ERIN M. SCHLIEP, SARAH M. COLLINS, SHIRLEY ROJAS-SALAZAR, NOAH R. LOTTIG AND EMILY H. STANLEY

Discovering interactions using covariate informed random partition models
GARRITT L. PAGE, FERNANDO A. QUINTANA AND GARY L. ROSNER

Estimating scale discrepancy in bayesian model calibration for ChemCam on the Mars Curiosity rover . . . . K. SHAM BHAT, KARY MYERS, EARL LAWRENCE, JAMES COLGAN AND ELIZABETH JUDGE

Continued

The Annals of Applied Statistics

Next Issues–*Continued*